RELIABLE CAUSAL INFERENCE UNDER UNRELIABLE ASSUMPTIONS: MACHINE LEARNING METHODS FOR OBSERVATIONAL, QUASI-EXPERIMENTAL, AND STRUCTURED DATA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Miruna Oprescu

May 2026

RELIABLE CAUSAL INFERENCE UNDER UNRELIABLE ASSUMPTIONS: MACHINE LEARNING METHODS FOR OBSERVATIONAL, QUASI-EXPERIMENTAL, AND STRUCTURED DATA

Miruna Oprescu, Ph.D.

Cornell University 2026

While modern machine learning has enabled flexible, high-dimensional modeling, reliable causal inference remains constrained by assumptions that are often unreliable in the settings where we most want answers: observational studies with unmeasured confounding, quasi-experiments with weak instruments and imperfect compliance, and structured systems with dependence across space, time, or networks. This dissertation develops **machine learning methods for reliable causal inference under unreliable assumptions**: methods that remain informative when identification is fragile, nuisance components must be learned flexibly, or the data-generating process departs from idealized models. Across settings, the recurring technical themes are orthogonalization and debiasing, sharp characterization of uncertainty (often through partial identification), and the careful use of structure to recover credible causal information.

We start in Part I with observational settings where average-effect estimands can miss important distributional features and where unobserved confounding can undermine causal identification. Chapter 2 introduces a model-agnostic, debiased pseudo-outcome approach for learning rich causal functionals beyond averages, including conditional distributional treatment effects. Building on this machinery, Chapter 3 derives sharp, quasi-oracle bounds on heterogeneous effects under explicit sensitivity constraints. Chapter 4 ex-

tends partial-identification ideas to sequential decision-making by characterizing sharp bounds and efficient estimators for off-policy policy value in robust Markov decision processes.

Next, Part II addresses causal learning from quasi-experiments, where identification is driven by imperfect experimental variation and compliance is often sparse or heterogeneous. Chapter 5 combines weak instrumental variation with observational data to estimate heterogeneous treatment effects, using observational structure to model heterogeneity while the instrument anchors identification. Chapter 6 then studies adaptive experimentation under noncompliance, introducing sequential encouragement designs that minimize asymptotic variance and support robust estimation with anytime-valid inference for safe monitoring and early stopping.

Lastly, Part III turns to structured data with spatiotemporal and network dependence, where interference and time-varying confounding are central challenges. Chapter 7 introduces a neural framework for spatiotemporal causal inference with time-varying confounding. Chapters 8 and 9 further study causal inference under spatial and network dependence, developing interference-aware deconfounding methods and partial identification guarantees when exposure mappings may be misspecified.

Taken together, this dissertation advances a unified perspective: causal inference in modern applications should not rely on idealized assumptions, but instead explicitly confront their limitations. By combining machine learning with orthogonal estimation, adaptive design, and partial identification, this work provides principled tools for reliable cause-and-effect analysis in observational, quasi-experimental, and structured domains.

BIOGRAPHICAL SKETCH

Miruna Oprescu grew up in Constanța, Romania. She discovered her love of mathematics, physics, and astronomy early, and before college she earned three silver medals and one gold medal at the International Olympiad on Astronomy and Astrophysics (2007–2010). In high school, she attended summer programs in the United States—including the Summer Science Program (SSP) and Stanford University Mathematics Camp (SUMaC)—which sparked a lasting interest in scientific research.

Miruna received an A.B. in Physics and Mathematics from Harvard University, with a secondary field in Computer Science. While initially intending to pursue a Ph.D. in physics, she took a course in machine learning and was drawn to its mix of mathematical structure and real-world impact. After college, she joined Microsoft as a software engineer and later moved to Microsoft Research in 2017 as a Data and Applied Scientist. There, she worked on the ALICE (Automated Learning and Intelligence for Causation and Economics) team and was a core contributor to EconML [Battocchi et al., 2019], an open-source Python library for causal machine learning. Her work helped design and implement methods for estimating treatment effects from observational and experimental data, supporting decision-making in large-scale industry applications.

After her role at Microsoft Research convinced her that the interesting applications start where the assumptions end, Miruna began her Ph.D. in Computer Science at Cornell University in 2021, supported by a Department of Energy Computational Science Graduate Fellowship. During her PhD, she completed research internships at Netflix and at Brookhaven National Laboratory, and her work focuses on developing trustworthy machine learning methods for causal inference and decision-making under imperfect assumptions.

To my father (1967–2016), for the spark.

To my mother, for the backbone.

To my husband, Eric, for the team.

And to my sons, Maxwell and Nicholas, for the wonder.

ACKNOWLEDGEMENTS

I am grateful to my advisor, Nathan Kallus, for his guidance at key moments and the space he gave me to shape this dissertation's direction—even when it involved a few ambitious detours. His encouragement helped me trust my own judgment more, take intellectual risks, and grow into the independent researcher I am today. I also appreciate his flexibility and support throughout my Ph.D., especially in accommodating the constraints of family life.

I would also like to thank Shinjae Yoo, who mentored me during my 2024 internship at Brookhaven National Laboratory and has continued to advise and support me since. He saw potential in my work early on, and his combination of excitement and steady encouragement helped me approach this dissertation with renewed momentum. Beyond the research, I am thankful for his support in my post-Ph.D. job search and career planning.

I am very grateful to Uri Shalit for his mentorship early in my Ph.D., including during periods when Nathan was away. His feedback substantially set the tone of this dissertation, especially the work on Chapter 3, and I am thankful as well for his support during my post-Ph.D. transition.

I am thankful to my committee members, Emma Pierson, Sarah Dean, and Peter Frazier, for their guidance and support. Emma was especially instrumental in my first year: even though our early research explorations did not ultimately become part of this dissertation, her enthusiasm and confidence helped me keep perspective and momentum during those crucial times. I also appreciate Sarah's and Peter's thoughtful input along the way, and I am lucky to have had them on my committee.

I have been very lucky to work with an exceptional set of collaborators throughout my Ph.D. I am especially grateful to Andrew Bennett, Kaiwen

Wang, Brian Cho, Andrew Jesson, Marah Ghoummaid, Jacob Dorn, Ayush Khot, and Maresa Schroder, as well as to my Brookhaven collaborators David Keetae Park, Ai Kagawa, and Xihaier Luo: they are thoughtful, engaged, and refreshingly straightforward to work with, and our conversations consistently made the research better. It was an honor to build this work alongside them.

I am also grateful to my mentors at Netflix. In particular, I thank Aish Fenton and Sudeep Das for their guidance and support during my 2022 internship, and Yonatan Gur and David Hubbard for hosting and mentoring me during my 2025 internship. I appreciate the opportunities to work on real causal questions at scale, and the feedback and perspective that came from being embedded in practice. I am thankful for their time, feedback, and encouragement.

My path into causal inference was shaped in large part by my prior experience at Microsoft Research. I am especially thankful to Vasilis Syrgkanis, Greg Lewis, and Lester Mackey for their mentorship and support. They helped me make sense of how research actually works—how to ask good questions, develop them, and navigate the research community—and they made me believe I could do it, too. I also thank the broader EconML community and collaborators at MSR for the ideas, engineering craft, and countless conversations that shaped how I think about machine learning and causal inference. I still find myself trying to live up to their example—scientifically and as a colleague.

I gratefully acknowledge funding support for this work. This material is based upon work supported by U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE-SC0023112. I thank the Department of Energy Computational Science Graduate Fellowship for the freedom to pursue this research and for the community that helped me grow as a scientist.

Finally, the biggest thank you to my family and friends. Above all, thank you to my husband, Eric. He supported me through the hardest stretches of this Ph.D.—with late-night pep talks, steady patience, and the kind of practical help that matters: taking over with the kids when deadlines hit, feeding me when I was running on fumes, and reminding me (often stubbornly) to keep going when I wanted to quit. I truly could not have done this without him.

To my sons, Maxwell and Nicholas: thank you for the joy, the perspective, and the daily reminder that curiosity is the point. To my mother, thank you for your love and support through all of it. My father, whom I think about often, inspired my love of science and would have loved this. I am also grateful to Eric's family here on Long Island for showing up in so many ways—helping with the kids, logistics, and the everyday support that kept our lives running. And to my friends: thank you for the pep talks, the check-ins, and the humor at exactly the right times. I could not have done this without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Modern machine learning excels at prediction. With enough data and computation, we can forecast outcomes and detect patterns with remarkable accuracy. Yet many scientific and policy questions are fundamentally causal: not what is likely to happen, but what would happen under alternative actions. Answering such counterfactual questions is especially difficult in the settings where they matter most, because the assumptions needed for causal identification are often only approximately satisfied in practice. In observational studies, important confounders may be unmeasured. In quasi-experiments, instruments may be weak and compliance imperfect. In structured scientific systems—such as spatial, temporal, or networked environments—interference and dependence violate assumptions that underlie much of classical causal methodology.

This dissertation studies how to perform causal inference in such regimes. It develops machine learning methods for *reliable causal inference under unreliable assumptions*. Here, reliability has two complementary meanings: when point identification is plausible, methods should remain statistically efficient while allowing flexible nuisance estimation; when point identification is not plausible, methods should characterize explicitly what can still be learned, ideally through sharp bounds or other estimands that make the remaining uncertainty explicit. Across these settings, a common methodological strategy emerges. The methods developed here combine orthogonal estimation, uncertainty-aware causal estimands, and structure-aware learning to produce credible causal conclusions even when classical assumptions hold only approximately.

Three settings, one perspective. The dissertation is organized around three settings in which causal inference must contend with unreliable assumptions.

- Part I studies observational data, where hidden confounding can undermine point identification and where average effects may fail to capture the causal quantities most relevant for decision-making.
- Part II studies quasi-experimental data, where identification is anchored by exogenous variation induced by an instrumental variable, but that variation may be weak, heterogeneous, or mediated by imperfect compliance.
- Part III studies structured scientific data, where treatments, outcomes, and confounders evolve over space, time, or networks, making interference and dependence central features of the problem rather than secondary complications.

Although these settings differ, they are linked by a common perspective: flexible machine learning should be paired with inference procedures that remain valid under imperfect assumptions. We now summarize each part in turn.

1.1 Causal Inference from Observational Data: Beyond Average Effects and Toward Confounding Robustness

A central goal of causal inference is to understand the effect of an intervention from data: for example, how a treatment changes an outcome on average, how that effect varies across individuals, or how it changes the distribution of possible outcomes. The gold standard for answering such questions is a randomized controlled trial, because randomization severs the link between treatment assignment and the many other factors that may also affect outcomes. In many important settings, however, such experiments are infeasible, costly, or unethical. We cannot randomly assign people to smoke in order to study lung cancer, withhold educational resources simply to measure their long-term effects, or re-

run large-scale public policies under controlled experimental conditions. As a result, researchers often rely on *observational data*: data that are passively generated by clinical practice, policy decisions, user behavior, or other real-world processes rather than by controlled randomization.

Observational data create two distinct challenges. First, even when the assumptions needed for identification are considered plausible, the causal quantities of practical interest often extend beyond average treatment effects. Many decisions depend not only on mean shifts, but also on how an intervention changes risk, tail behavior, or the dispersion of outcomes. A treatment with a favorable average effect may still increase the probability of severe adverse outcomes; conversely, an intervention with modest average benefit may create substantial gains in the upper tail. Chapter 2 addresses this problem by developing a model-agnostic approach to learning causal functionals beyond averages, including conditional distributional treatment effects. Methodologically, the chapter shows how debiased pseudo-outcomes and orthogonal estimation can be used to separate the target causal functional from the nuisance estimation problem, so that flexible machine learning methods can be used while preserving valid inference.

Second, observational data may be confounded by important variables that are unmeasured. This is the problem of *unobserved confounding*: treatment assignment and outcomes may both depend on latent factors that are not recorded in the data. For example, if one studies the effect of a health intervention using medical records, patients who receive the intervention may systematically differ from those who do not in motivation, baseline health, or physician judgment, even after conditioning on observed covariates. In such settings, the causal effect is generally no longer *point identified*, meaning that the observed data do not

determine a unique causal estimand without stronger assumptions. Because unobserved confounding cannot in general be tested away from the data, a reliable analysis should not proceed as if point identification still held.

Chapter 3 addresses this challenge through partial identification. Rather than forcing a single potentially misleading estimate, it derives sharp, quasi-oracle bounds on heterogeneous treatment effects under explicit sensitivity constraints. These bounds characterize exactly what can still be learned once hidden confounding is allowed, and sharpness is essential: the intervals should be no wider than the assumptions and data require. In practice, such bounds can still support meaningful conclusions. They may exclude qualitatively important possibilities, reveal whether an effect is robust to moderate confounding, or show that a substantive conclusion depends on assumptions stronger than the data alone can justify.

Chapter 4 extends this perspective to sequential decision-making. There, the target is not a one-step treatment effect but the value of a policy evaluated from observational data, possibly under distribution shift or latent confounding. The chapter develops sharp bounds and efficient estimators for off-policy policy value in robust Markov decision processes, showing how the logic of partial identification and orthogonal estimation carries over to dynamic settings.

Taken together, the chapters in this part develop an observational perspective on reliable causal inference: when working with observational data, the relevant causal questions often extend beyond averages alone to distributional risks, heterogeneity, and tail behavior; when the assumptions needed for point identification are not credible, one should represent explicitly what remains learnable rather than conceal uncertainty behind a point estimate.

1.2 Causal Inference from Quasi-Experiments: Weak Instruments and Imperfect Compliance

Part II studies causal learning from quasi-experiments, where direct randomization of the treatment is infeasible, but some externally induced variation in treatment uptake is still available. In such settings, researchers often turn to *instrumental variables*: variables that shift treatment uptake without directly affecting the outcome except through treatment. These often take the form of randomized encouragements, recommendations, interface changes, or eligibility rules rather than direct treatment assignment. For example, a platform may randomize which recommendation or interface a user sees, a clinician may encourage a patient to take a medication, or a policy may alter eligibility or access without controlling individual uptake. Such instrumental variables can restore point identification even in the presence of hidden confounding, making them an appealing alternative when observational assumptions are too strong and direct randomized experiments are unavailable.

Their usefulness, however, depends on how strongly they actually shift treatment uptake. When compliance is perfect, the problem effectively reduces to a randomized trial. When compliance is absent, the instrument carries essentially no identifying information. The practically important regime lies in between: compliance is positive but weak, heterogeneous across subpopulations, and sometimes effectively zero for the groups one most wants to study. In these settings, standard instrumental-variable methods can become unstable, statistically inefficient, or simply uninformative.

Chapter 5 addresses this problem by combining weak instrumental-variable data with potentially biased observational data. The key idea is to exploit the

complementary strengths of the two sources: observational data are used to learn how treatment effects vary across covariates, while instrumental variation is used to correct for confounding where compliance provides identifying power. This makes it possible to estimate heterogeneous treatment effects even when compliance is sparse, highly uneven across subpopulations, or effectively absent in parts of the covariate space.

Chapter 6 then studies adaptive experimentation under noncompliance. In many ongoing experiments, assignment does not guarantee treatment uptake, but the assignment itself can still be randomized and updated over time. This chapter develops sequential encouragement designs that allocate instrumental variation where it is most informative, together with a multiply robust treatment effect estimator and anytime-valid inference procedures for safe monitoring and early stopping. More broadly, it provides a framework for online causal learning under imperfect compliance, where both assignment and inference must adapt sequentially as data accrue.

Taken together, the chapters in this part develop a quasi-experimental perspective on reliable causal inference: when direct experimentation is not fully available, instrumental variables can still support point identification, but only if design and estimation are tailored to the weakness, heterogeneity, and imperfect compliance that arise in practice.

1.3 Causal Inference in Structured Data: Spatiotemporal and Network Dependence

Part III turns to structured scientific data, where treatments, outcomes, and confounders evolve over space, time, or networks. Such settings arise naturally in

environmental health, climate and Earth systems, epidemiology, and social networks, where interventions are embedded in systems with strong dependence across units and over time. In these problems, the main challenge is not merely that the data are high-dimensional, but that the causal structure itself is shaped by spatial, temporal, or relational interactions.

This creates several difficulties for causal inference. Treatments applied at one location may affect outcomes elsewhere through spillovers or transport. Covariates may evolve over time in response to past interventions and, in turn, influence future treatment assignment, creating time-varying confounding. In network settings, even the mechanism by which neighbors' treatments affect a unit's outcome may be uncertain, so that the causal estimand itself depends on assumptions about exposure and interference. As a result, dependence is not a nuisance to be removed, but a defining feature of the causal problem.

Chapter 7 addresses these challenges in spatiotemporal settings with time-varying confounding. It develops a neural framework that combines representation learning with iterative G-computation in order to estimate intervention effects from structured observational data. The chapter shows how causal adjustment and modern spatiotemporal learning architectures can be integrated so that flexible prediction does not come at the expense of causal interpretability.

Chapter 8 then studies spatial causal interference in the presence of latent confounding. A key insight of this chapter is that interference can sometimes be informative rather than purely problematic: the way neighboring treatments affect outcomes can itself reveal otherwise hidden spatial confounding structure. Building on this idea, the chapter develops a deconfounding approach that uses interference patterns to recover credible causal information from spatial data.

Finally, Chapter 9 considers network settings in which the exposure map-

ping is itself uncertain or misspecified. Rather than committing to a single possibly incorrect representation of how neighbors' treatments matter, the chapter develops a partial identification framework that quantifies what remains learnable when the mechanism of interference is only approximately known. This extends the uncertainty-aware perspective of earlier parts of the dissertation to structured network settings.

Taken together, the chapters in this part develop a structure-aware perspective on causal inference: when data are spatial, temporal, or networked, reliable estimation requires methods that model dependence explicitly while preserving clear causal interpretation and honest uncertainty quantification.

1.4 Reading guide

The dissertation is organized so that each part can be read independently. Readers primarily interested in observational causal inference can begin with Part I; those interested in quasi-experimental methods and noncompliance can begin with Part II; and those interested in spatiotemporal or networked data can begin with Part III. Each chapter includes its own problem setup, assumptions, and estimands, so the dissertation may be read selectively as well as sequentially.

Part I

Causal Inference from Observational Data: Beyond Average Effects and Toward Confounding Robustness

CHAPTER 2

ROBUST AND AGNOSTIC LEARNING OF CONDITIONAL DISTRIBUTIONAL TREATMENT EFFECTS

This chapter is based on Kallus and Oprescu [2023b].

The conditional average treatment effect (CATE) is the best measure of individual causal effects given baseline covariates. However, the CATE only captures the (conditional) average, and can overlook risks and tail events, which are important to treatment choice. In aggregate analyses, this is usually addressed by measuring the distributional treatment effect (DTE), such as differences in quantiles or tail expectations between treatment groups. Hypothetically, one can similarly fit conditional quantile regressions in each treatment group and take their difference, but this would not be robust to misspecification or provide agnostic best-in-class predictions. We provide a new robust and model-agnostic methodology for learning the conditional DTE (CDTE) for a class of problems that includes conditional quantile treatment effects, conditional super-quantile treatment effects, and conditional treatment effects on coherent risk measures given by f-divergences. Our method is based on constructing a special pseudo-outcome and regressing it on covariates using any regression learner. Our method is model-agnostic in that it can provide the best projection of CDTE onto the regression model class. Our method is robust in that even if we learn these nuisances nonparametrically at very slow rates, we can still learn CDTEs at rates that depend on the class complexity and even conduct inferences on linear projections of CDTEs. We investigate the behavior of our proposal in simulations, as well as in a case study of 401(k) eligibility effects on wealth.

2.1 Introduction

Measuring treatment-effect heterogeneity along observed covariates is an important tool for interpreting the results of A/B tests on online platforms, program evaluations in social science whether experimental or observational, and clinical trials in medicine. These analyses can help diagnose how the treatment works, understand for whom it does and does not work, and assess fairness [Heckman et al., 1997, Crump et al., 2008, Kent et al., 2010, Kallus, 2023]. On the other hand, measuring distributional treatment effects (DTEs) such as quantile treatment effects (QTEs) is an important tool for understanding the impact of interventions beyond the mean, especially when outcomes are naturally very skewed, like income or platform usage [Bitler et al., 2006, Firpo, 2007, Belloni et al., 2017a, Kallus et al., 2024].

Motivated by the need to assess *both* heterogeneity *and* distributional impact, in this chapter we study flexible, agnostic, and robust machine learning tools to estimate *conditional* DTE (CDTE) functions. Recent advances in causal machine learning have offered new methods for assessing effect heterogeneity by learning conditional *average* treatment effects (CATEs) [Imai and Ratkovic, 2013, Athey and Imbens, 2016, Wager and Athey, 2018a, Künzel et al., 2019, Kennedy, 2023a, Nie and Wager, 2021]. These works have highlighted the importance of learning CATEs *directly*, rather than learning conditional-average outcomes by treatment arm and taking their difference (also known as the plug-in approach). One issue with the plug-in approach is that it can wash out the effect signal (not robust): *e.g.*, many variables strongly predict baselines but only a few modulate the effect. Another issue is that it fails to give best-in-class predictions (not agnostic): *e.g.*, taking the difference of the best linear predictions of outcome by arm does not yield the best linear prediction of treatment effect.

We tackle the same challenges for CDTEs. We consider CDTEs for a very rich class of distributional metrics that includes quantiles, super-quantiles, and other coherent risk measures. Given any distributional metric in our class, we construct a pseudo-outcome that combines an initial guess for the CDTE along with a debiasing term. Our algorithm is then to regress this pseudo-outcome on covariates, using any given blackbox learner. In the case of the CATE, our method recovers the DR-Learner [Kennedy, 2023a]. We show that this procedure is robust in the sense that the blackbox regression mimics having used the pseudo-outcome with the *true* CDTE as the "initial guess", thus removing any bias or noise from fitting the baselines. We further show that our method is model-agnostic in the sense that, if the blackbox is not a universal approximator, we still get the best approximation to the CDTE function offered by the blackbox. Lastly, we show that our estimating procedure allows for valid statistical inference on the best linear projection of CDTE, thus enabling interpretable analyses of distributional effect heterogeneity. We demonstrate in a comprehensive simulation study that we obtain uniformly better performance than the plug-in approach for several types of CDTEs. Finally, we apply our method on a real-world study of 401(k) eligibility and its impact on financial wealth.

2.2 Related Work

2.2.1 Learning Conditional Average Treatment Effects

CATE estimation is a central problem in causal learning and finds uses in both the analysis of causal interventions and decision support for personalization. Going beyond fully-parametric models, early advances relied on semiparametric models that imposed structure on the CATE function [Robins et al., 1992, Van der Laan and Robins, 2003, Vansteelandt and Joffe, 2014]. Recently there

has been a surge of interest in leveraging machine learning for CATE estimation. These flexible methods either employ specific machine learning models such as Bayesian regression trees [Hill, 2011, Hahn et al., 2020], random forests (RFs) [Wager and Athey, 2018a, Oprescu et al., 2019], neural networks [Johansson et al., 2016, Atan et al., 2018, Shi et al., 2019], or allow for arbitrary blackbox meta-learners [Künzel et al., 2019, Nie and Wager, 2021] by leveraging efficient influence functions [Robins et al., 2017, Kennedy, 2023a, Curth et al., 2020] and Neyman orthogonality [Chernozhukov et al., 2018a, Foster and Syrgkanis, 2023a]. This work is closest to the works on efficient influence functions and blackbox meta-learners, and we add to this literature by considering distributional statistics beyond averages.

2.2.2 Double Machine Learning and Orthogonal Statistical Learning

Another vein of related literature is learning with nuisances. Both our work and the above methods based on regressing efficient influence functions may be phrased within the wider framework of orthogonal statistical learning [Foster and Syrgkanis, 2023a], which extends double machine learning (DML) [Chernozhukov et al., 2018a] from estimation to minimizing loss functions involving unknown nuisances subject to Neyman orthogonality [Neyman, 1959]. Neyman orthogonal losses arise naturally from efficient influence functions [Ichimura and Newey, 2022], a fact that has been leveraged by Foster and Syrgkanis [2023a], Kennedy [2023a], Curth et al. [2020], Athey and Wager [2021] to tackle both CATE and policy learning. These methods have learning rates that adapt to the complexity of the target rather than that of the nuisances, but they largely focus on conditional averages. Our work builds on this line of research and is most similar to [Kennedy, 2023a] in that we propose nuisance agnostic CDTE estimators with guarantees beyond those given by Neyman orthogonality.

2.2.3 Distributional Treatment Effects

The literature on DTEs can be split into two categories: (i) estimating cumulative distribution functions of potential outcomes and (ii) directly estimating distributional parameters of interest. In the first category, the main approach is to model conditional counterfactual distributions using distribution regression [Chernozhukov et al., 2013, 2020] or fully flexible approaches such as neural networks [Ge et al., 2020, Zhou et al., 2022] and mean kernel embeddings [Park et al., 2021]. Since these methods rely on plug-in estimation, they can be slow and biased when concerned with a particular DTE. The second category focuses on estimating a particular DTE. Existing orthogonal/efficient methods focus on unconditional DTEs [Firpo, 2007, Belloni et al., 2017a, Kallus et al., 2024]. Existing methods that tackle CDTEs rely on plug-in estimation [Park et al., 2021] or parametric methods [Hohberg et al., 2020]. Our work bridges the gap by proposing robust, agnostic, and flexible CDTE learning.

2.3 Background and Setup

We consider either an experimental or observational dataset with two treatments, denoted by 0 and 1. Each unit in the dataset is a draw from a population of baseline covariates $X \in \mathcal{X}$, treatment indicator $A \in \{0, 1\}$, and observed outcome $Y \in \mathbb{R}$. The dataset consists of n such independent draws, $Z_i = (X_i, A_i, Y_i) \sim Z = (X, A, Y)$, $i = 1, \ldots, n$. We define the propensity score as $e^*(X) = \mathbb{P}(A = 1 \mid X)$. We assume throughout that $e^*(X) \in (0, 1)$ almost surely, known as overlap.

Each unit is additionally associated with two unobserved potential outcomes, $Y(0)$, $Y(1) \in \mathbb{R}$, representing the potential outcome we would observe if (possibly counter to fact) each treatment were applied. We assume we observe

the potential outcome corresponding to the treatment indicator, $Y = Y(A)$, which also encapsulates an assumption of no interference between unit treatments. We assume unconfoundedness (ignorability) throughout: $Y(a) \perp\!\!\!\perp A \mid X$. For experimental data this is ensured by design via random assignment of A (often with covariate-agnostic assignment, $A \perp\!\!\!\perp X$). For observational data, this is an assumption that all potential sources of confounding are captured in X. For our purposes, the only difference between the two cases is whether the propensity score $e^*(X)$ is known. We are interested in the differences between the conditional distributions of $Y(1)$ and $Y(0)$, given X. In the next section, we describe specific metrics for these differences.

Notation Given a distribution F, we define $\mathbb{E}_F[f(Z)] = \int f(z)dF(z)$. We let $\widehat{\mathbb{E}}_n$ denote the empirical expectation $\widehat{\mathbb{E}}_n f(Z) = \frac{1}{n}\sum_{i=1}^n f(Z_i)$. For a parameter f, we reserve f^* to represent its true value and $\widehat{f}$ a value learned from the data. We let $\|f\| := \mathbb{E}_F[f(z)^2]^{1/2}$ be the L_2 norm of f. We use D to denote directional derivatives: $D_h F(h)|_{h=h'} = \frac{\partial}{\partial a} F(h'+a)|_{a=0}$, whenever this exists. If h is a vector of functions, $D_h F(h) = (D_{h_1} F(h), ..., D_{h_k} F(h))$ and $D_{h_i} F(h)|_{h=h'} = \frac{\partial}{\partial a} F((h'_1, ..., h'_i + a, ..., h'_k))|_{a=0}$. We let conv($\mathcal{S}$) denote the convex hull of $\mathcal{S}$. For two numbers a, b, we take $a \lesssim b$ to mean $a \leq Cb$ for some universal constant C and $a \asymp b$ to mean $cb \leq a \leq Cb$ for some constants c and C. Finally, we let $\overline{1, n}$ denote the set of integers $\{1, ..., n\}$.

2.4 Conditional Distributional Treatment Effects

CDTEs are functions mapping x to a difference in some statistic of the conditional distributions of $Y(1)$ and $Y(0)$, given $X = x$. Examples of such statistics are the mean, yielding the CATE, and the τ-quantile, yielding the CQTE. In this chapter, we handle a very wide range of CDTEs given by statistics defined by *moment equations* [Chamberlain, 1992, Ai and Chen, 2003].

Definition 2.1 (Moment Statistics)**.** Given $\rho : \mathbb{R}^{m+2} \to \mathbb{R}^{m+1}$, we define a statistic of a distribution F on $\mathbb{R}$ as $\kappa^*(F)$ for $(\kappa^*(F), h^*(F)) \in \mathbb{R} \times \mathbb{R}^m$ any solution (if it exists) to the moment equation

$$\mathbb{E}_F[\rho(Y, \kappa, h)] = 0. \tag{2.1}$$

Definition 2.2 (CDTEs)**.** Let $F_{Y(1)|X}$ and $F_{Y(0)|X}$ denote the conditional distributions of $Y(1)$ and $Y(0)$ given X, respectively. Fix a statistic $\kappa^*(F)$ given by Definition 2.1. The corresponding CDTE is given by:

$$\text{CDTE}(X) = \kappa^*(F_{Y(1)|X}) - \kappa^*(F_{Y(0)|X}). \tag{2.2}$$

For brevity, we will define the following functions (scalar-valued and $\mathbb{R}^m$-valued, respectively)

$$\kappa_a^*(X) = \kappa^*(F_{Y(a)|X}), \;\; h_a^*(X) = h^*(F_{Y(a)|X}), \;\; a = 0, 1.$$

We also assume these functions exist in that at least one solution to Eq. (2.1) exists for $F_{Y(a)|X}$ (see Remark 2.8 regarding multiplicity). In the examples below, we give specific names for κ or h (*e.g.*, $q(F; \tau)$ for the τ-quantile of F), in which case we use analogous abbreviations (*e.g.*, $q_a(X; \tau)$).

Note that with unconfoundedness and overlap, $F_{Y(a)|X}$ is the same as $F_{Y|X,A=a}$, the conditional distribution of Y given X and $A = a$. Therefore, κ_a^*, h_a^*, and the CDTE are all identifiable from the data. That is, despite being defined in terms of potential outcomes, the CDTE depends only on the distribution of observed data (X, A, Y) and not on the unobserved data $(X, Y(0), Y(1))$, under our assumptions. The question we address in this chapter is *how* to learn CDTEs from data.

We next review some important examples of CDTEs that fit into our framework. First note that CATE fits into this setup by setting $\rho(y, \kappa) = y - \kappa$ with $m = 0$ in Eq. (2.1) (no h's). The power of our framework is that it captures many CDTEs of interest beyond averages.

2.4.1 Example 1: Conditional Quantile Treatment Effects

Our first example is the **conditional quantile treatment effect** (CQTE). The quantile at level $\tau \in (0,1)$ of the distribution F is defined as $q(F;\tau) = \inf\{y : F(y) \geq \tau\}$ (where F is the cumulative distribution function). Whenever F has a positive derivative at $q(F;\tau)$, it is given by Definition 2.1 with $m = 0$ (no h's) and

$$\rho(y, q) = \tau - \mathbb{I}[y \leq q]. \tag{2.3}$$

The corresponding CDTE is called the CQTE at level τ.

Non-conditional QTEs are an important tool for quantifying the effects of treatments throughout the outcome distribution [Firpo, 2007, Belloni et al., 2017a, Kallus et al., 2024]. This is especially important when we suspect that the outcome distribution might be skewed or heavy-tailed (*e.g.*, income).

CQTEs offer an opportunity to assess such effects at the individual level. In particular, the CQTE can capture the potential increase in an individual's chance to have very poor outcomes due to treatment, even if the average effects are good or neutral. That is, if $A = 1$ denotes an intervention, then CQTEs answer the prediction question: given an individual's covariates, how would intervening affect the 10%-worst possible outcomes.

A related quantity is the difference between the τ and $1-\tau$ conditional quantiles across treatments: $\omega(X;\tau) = q(F_{Y(1)|X};\tau) - q(F_{Y(0)|X};1-\tau)$. This quantity can bound the (unobservable) individual treatment effect:

$$\mathbb{P}\left(\omega(X;\tau) \leq Y(1) - Y(0) \leq \omega(X;1-\tau) \mid X\right) \geq 1 - 4\tau.$$

For the sake of brevity and uniform treatment, we focus on CDTEs (*i.e.*, using the same statistic for both potential outcome distributions), but our results readily extend to quantities like this as well.

2.4.2 Example 2: Conditional Super-Quantile Treatment Effects

Our next example is the **conditional super-quantile treatment effect** (CSQTE). Given $\tau \in (0, 1)$, the super-quantile (also known as conditional value-at-risk or tail expectation) at level τ of a distribution F is defined as

$$\begin{aligned} \mu(F; \tau) &= \inf_\beta \beta + \frac{1}{1-\tau} \int_\beta^\infty (1 - F(y))\, dy \\ &= \inf_\beta \mathbb{E}_F[\beta + (1-\tau)^{-1} \max\{y - \beta, 0\}]. \end{aligned} \tag{2.4}$$

The corresponding CDTE is called the CSQTE at level τ.

Note that a β realizing the above infimization is the quantile at level τ, $q(F; \tau)$. Therefore, given positive density at the quantile, the super-quantile is given in Definition 2.1 by setting $m = 1$ and

$$\rho(y, \mu, q) = \left((1-\tau)^{-1} y\, \mathbb{I}[y \geq q] - \mu,\ \tau - \mathbb{I}[y \leq q]\right). \tag{2.5}$$

The super-quantile $\mu(F; \tau)$ is the largest-possible subpopulation average among all subpopulations comprising a $1 - \tau$ fraction of the population of values described by F. When $F(q(F; \tau)) = \tau$, this is the average of values above the τ quantile. If we are interested in the left tail (average below a quantile), we can simply consider the negative CSQTE in the negative outcomes. Unlike the quantile, the super-quantile is a coherent risk measure [Artzner et al., 1999].

In particular, while the CQTE captures the impact on the outcomes at a single probability level, they can fail to provide a full picture of the risk profile as they are indifferent to anything beyond the threshold of the quantile. In contrast, the CSQTE captures effects beyond quantile breakpoint and quantifies the impact on the *average*, say, 10%-worst (or, best) outcomes. Crucially, since super-quantiles are a coherent risk measure, making individual decisions based on CSQTEs (*e.g.*, intervene when the CSQTE is positive) is rational with respect to the coherent-risk axioms.

2.4.3 Example 3: Conditional f-Risk Treatment Effects

Our third example is a whole class of coherent risk measures. Coherent risk measures quantify how bad/good a random loss/reward is while satisfying certain axioms (monotonicity, sub-additivity, homogeneity, and translational invariance). A key result is that coherent risk measures are equivalent to distributionally robust optimization [Ruszczyński and Shapiro, 2006]. In this section, we focus on the **conditional f-risk treatment effect** (CfRTE), a family of coherent risk measures generated by f-divergences.

Given a convex $f : \mathbb{R} \to \mathbb{R}$ with $f(1) = 0$, we define the conditional f-risk at level $\delta \geq 0$ as:

$$R^f(F;\delta) = \sup_{G \ll F,\, D^f(G\|F) \leq \delta} \mathbb{E}_G[Y],$$

where $D^f(G\|F) = \mathbb{E}_F[f(dG/dF)]$. For example, the f-divergence for $f(x) = x \log x$ is the Kullback–Leibler divergence, and the f-risk is known as entropic value-at-risk (EVaR) which upper bounds the super-quantile at level $1 - e^{-\delta}$ [Ahmadi-Javid, 2012]. The f-risk admits a dual formulation given by the following convex optimization problem [Rockafellar, 1974]:

$$R^f(F;\delta) = \inf_{\beta \geq 0, \lambda \in \mathbb{R}} \mathbb{E}_F\left[m(Y,\beta,\lambda;\delta)\right],$$
$$m(y,\beta,\lambda;\delta) = \delta\beta + \lambda + \beta f^*\left(\beta^{-1}(y-\lambda)\right),$$

where $f^*(x^*) = \sup_{x \in \mathbb{R}} \; xx^* - f(x)$ is the convex conjugate of f. Therefore, under appropriate regularity, $R^f(F;\delta)$ is given by Definition 2.1 with

$$\rho(y,R,\beta,\lambda) = \left(m(y,\beta,\lambda;\delta) - R^f,\; \frac{\partial}{\partial\beta} m(Z,\beta,\lambda;\delta),\; \frac{\partial}{\partial\lambda} m(Z,\beta,\lambda;\delta)\right). \tag{2.6}$$

The corresponding CDTE is called the CfRTE at level δ. Frequently employed in actuarial science and finance, coherent risk measures are a fundamental tool

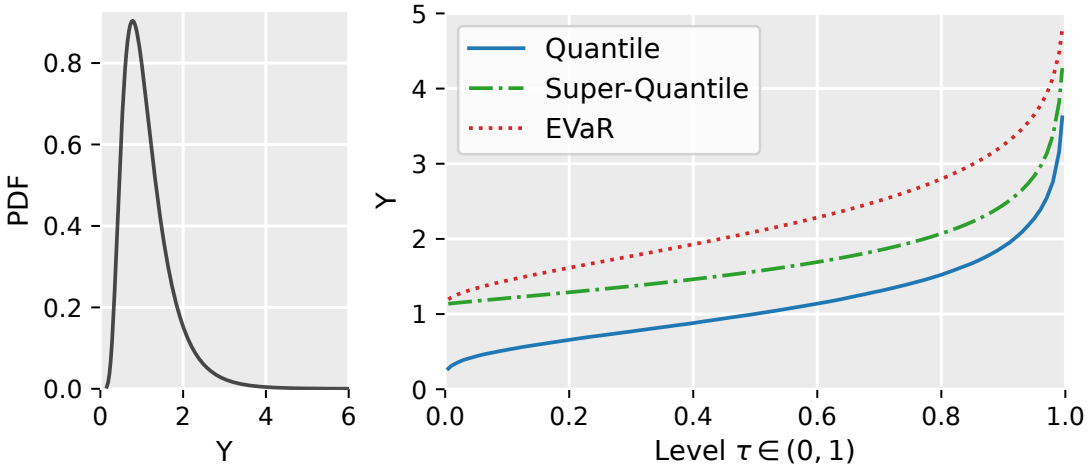


Figure 2.1: Comparison of quantiles, super-quantiles, and EVaRs for a right-truncated ($Y \leq 6$) Lognormal($\mu = 0,\ \sigma = 0.5$) at different risk levels $\tau \in (0, 1)$. *Note:* Level τ corresponds to $\delta = -\log(1 - \tau)$ for EVaR.

for assessing risk. CfRTE can therefore be used to perform risk-benefit analyses on the treatment effect profiles of affected groups.

Remark 2.3. A natural question is which risk measure to use in practice. Ultimately, the appropriate risk measure (and level) will be application-dependent and it is up to the practitioner to select a measure that best reflects the desired risk profile. For illustration purposes, in Figure 2.1 we show how our three examples (quantiles, super-quantiles and EVaRs) compare for a heavy-tailed distribution.

2.5 Pseudo-Outcome Regression For CDTEs

In this section, we propose an algorithm for learning CDTEs from data. As a first step, consider the following naive, but straightforward estimation procedure:

$$\text{CDTE}^{\text{Plugin}}(X) = \widehat{\kappa}_1(X) - \widehat{\kappa}_0(X). \tag{2.7}$$

where $\widehat{\kappa}_a(\cdot)$ are estimates for $\kappa_a(\cdot)$ (see Section 2.5.2 regarding how these may be constructed). This estimator is known as a "plug-in" estimator or, in the

context of CATE learning, a "T-learner" [Künzel et al., 2019]. Unfortunately, this approach has several drawbacks. One concern is that the treatment effect signal can be easily masked by noise in the baseline predictors. In particular, while many variables may strongly predict baseline response, only a few strongly modulate effect. The effect function may often be simpler, sparser, and/or smoother than each baseline function. Therefore, the plug-in estimator can suffer from excessive bias inherent in fitting baseline estimators in high-dimensions or using flexible models. If, on the other hand, we seek to use a simple model such as a linear fit, we will find that differencing the best linear predictors of baseline outcomes does not yield the best linear predictor of effect. For these reasons, it is imperative to learn CDTEs in a direct, model-agnostic, and robust way.

To address the short-comings of the plug-in estimator, we consider the plug-in prediction on each data point $\text{CDTE}^{\text{Plugin}}(X_i)$, debias it using the observed action and outcome A_i, Y_i, and finally regress the debiased prediction on X again. Our first task is to propose a debiased pseudo-outcome for CDTE learning.

Definition 2.4 (CDTE Pseudo-Outcome)**.** Fix a statistic in Definition 2.1. Let $\nu_a^* = (\kappa_a^*, h_a^*)$ and

$$\alpha_a^*(X) = (J_a^*(X))_1^{-1}, \quad \text{where } J_a^*(X) = D_{\nu_a}\{\mathbb{E}[\rho(Y, \nu_a) \mid X, A = a]\}\Big|_{\nu_a = \nu_a^*}$$

provided that $(J_a^*(X))^{-1}$ exists. Here, $(J_a^*(X))_1^{-1}$ denotes the first row of the inverse Jacobian. Given some e, α, ν serving as stand-ins for e^*, α^*, ν^*, we define the CDTE pseudo-outcome by

$$\psi(Z, e, \alpha, \nu) = \kappa_1(X) - \kappa_0(X) - \frac{A - e(X)}{e(X)(1 - e(X))} \alpha_A(X)^T \rho(Y, \nu_A(X)). \tag{2.8}$$

We refer to e^*, α^*, ν^* as *nuisance functions*, as they are unknown functions needed to construct our pseudo-outcome. One initial motivation for Eq. (2.8)

is that, by iterated expectations, we have $\mathbb{E}[\psi(Z, e^*, \alpha^*, \nu^*) \mid X] = \text{CDTE}(X)$. That is, if we had plugged in the true nuisances, then our pseudo-outcome is an unbiased regression target for the CDTE. This is, however, not so surprising because if we plug in $\nu = \nu^*$, the last term in Eq. (2.8) just has zero conditional expectation, given X, so we are left with $\kappa_1^*(X) - \kappa_0^*(X)$. The reason why Eq. (2.8) is special is that, as we will show, if we make small errors in the nuisances, the impact on the conditional expectation of our pseudo-outcome is even smaller, leading to robustness guarantees. This is in contrast to the plug-in approach, where errors in $\widehat{\kappa}_a(X)$ propagate directly to $\text{CDTE}^{\text{Plugin}}$, which is just their difference.

The origin of Eq. (2.8) is that $\psi(Z, e^*, \alpha^*, \nu^*)$ is in fact the efficient influence function for the estimand $\mathbb{E}[\text{CDTE}(X)]$ (for a review of influence functions see Kennedy [2024], Ichimura and Newey [2022]). In particular, if our statistic is simply the mean ($\rho(y, \kappa) = y - \kappa$) then $\psi(Z, e^*, \alpha^*, \nu^*)$ reduces to the familiar doubly-robust influence function, $\mathbb{E}[Y \mid X, A = 1] - \mathbb{E}[Y \mid X, A = 0] + \frac{A - e^*(X)}{e^*(X)(1 - e^*(X))}(Y - \mathbb{E}[Y \mid X, A])$ that produces the CATE when regressed on X (as studied by Kennedy [2023a]). As we never use the fact that $\psi(Z, e^*, \alpha^*, \nu^*)$ is the efficient influence function, we do not prove this or impose the necessary regularity conditions for this to actually hold precisely. Instead, we simply use a perturbation argument to essentially guess the form of Eq. (2.8), given which we directly prove our robustness and inference guarantees.

2.5.1 The CDTE Learning Algorithm

We now describe our algorithm, which is summarized in Algorithm 2.1. We first split the data into K even folds. We then construct a pseudo-outcome for each data point by plugging in estimates of the nuisances into Eq. (2.8). The nuisances at a point are fit on data excluding the fold that the data point belongs to. This ensures that the data point and the nuisance estimates are independent

Algorithm 2.1 CDTE Learner

Input: Data $\{(X_i, A_i, Y_i) : i \in \overline{1,n}\}$, folds $K \geq 2$, nuisance estimators, regression learner

for $k \in \overline{1,K}$ **do**
  Use data $\{(X_i, A_i, Y_i) : i \neq k-1 \pmod K\}$ to
  construct nuisance estimates $\hat{e}^{(k)}, \hat{\alpha}^{(k)}, \hat{\nu}^{(k)}$
  for $i = k-1 \pmod K$ **do**
    Set $\widehat{\psi}_i = \psi(Z_i, \hat{e}^{(k)}, \hat{\alpha}^{(k)}, \hat{\nu}^{(k)})$
  end for
end for
return $\widehat{\mathrm{CDTE}}(x) = \widehat{\mathbb{E}}_n[\widehat{\psi} \mid X = x]$

without splitting the data into two and instead only using parts of it for each task. Finally, we regress the pseudo-outcome on X using a given regressor. We use $\widehat{\mathbb{E}}_n[W \mid X = x]$ to denote the function learned by the given regression method when regressing W on X given n data points (X_i, W_i), $i \in \overline{1,n}$. This notation affords us significant generality. For example, the regression method may be to minimize the sum of squared errors over some function class (*e.g.*, linear or neural nets) or it may be given by local polynomial regression or RFs.

2.5.2 Nuisance Estimation

Algorithm 2.1 requires nuisance estimators as inputs. Exactly how these nuisances are estimated may depend on the particular scenario. Let us first discuss the propensity $e^*(x)$. If it is known, as in experimental settings, we may simply set it as our estimate. Otherwise, we can estimate it using probabilistic classification (*e.g.*, logistic regression or neural nets with softmax output).

Next, we discuss $\nu_a^*(x)$. Most generally, since it is defined by solving the conditional moment restriction $\mathbb{E}[\rho(Y, \nu_a^*(X)) \mid X] = 0$, this nuisance may be learned by employing methods made for solving such models [Ai and Chen, 2003, Chen and Pouzo, 2009, Bennett et al., 2019, Bennett and Kallus, 2023, Athey et al., 2019, Khosravi et al., 2022, Dikkala et al., 2020]. However, in some specific examples,

more direct methods may be applicable. For both CQTE and CSQTE, $\nu_a^*(x)$ includes a conditional quantile function, which can be learned using any quantile regression method, whether minimizing the check loss [Koenker and Bassett Jr, 1978] or using forests [Meinshausen and Ridgeway, 2006]. For CSQTE, we additionally need to fit the conditional super-quantile. Per Eq. (2.5), that nuisance is given by the regression $\mathbb{E}[(1-\tau)^{-1}Y\mathbb{I}[Y \geq q_a(X;\tau)] \mid X, A = a]$. Therefore, one possibility is to split the training data (being all data excluding the k^{th} fold) into two halves, fit a conditional quantile estimate $\hat{q}_a^{(k)}(x;\tau)$ on one, set $\omega_i = (1-\tau)^{-1}Y_i\mathbb{I}[Y_i \geq \hat{q}_a^{(k)}(X_i;\tau)]$ on the other, and return $\hat{\mu}_a^{(k)}(x;\tau) = \widehat{\mathbb{E}}_{(1-1/k)n/2}[\omega \mid X = x, A = a]$. In particular, a given X-regression method can simply be applied once to $A = 0$ and once to $A = 1$. Appendix A of Dorn et al. [2025a] provides guarantees for this procedure when the regression learner minimizes squared error over a class with bracketing entropy and Olma [2021] when using local linear regression.

Lastly, we discuss $\alpha_a^*(x)$. In some cases, it is a known function that need not be estimated. For CfRTE, it is equal to $(-1, 0, 0)$. In other cases, it is given directly by other nuisances. For CSQTE, it is equal to $(-1, (1-\tau)^{-1}q_a^*(x;\tau))$ so we can simply re-use the estimate we constructed for ν_a^*. In yet other cases, it is another nuisance that must be estimated. For CQTE, it is equal to $1/f_{Y|X=x,A=a}(q_a^*(x;\tau))$, the reciprocal of the density of $Y \mid X = x, A = a$ at the conditional quantile. One way to fit this suggested by Leqi and Kennedy [2021] is to split the training data into two halves, fit a conditional quantile estimate $\hat{q}_a^{(k)}(x;\tau)$ on one, set $\omega_i = K((Y_i - q_a^{(k)}(X_i;\tau))/b_n)/b_n$ on the other, where $K(u)$ is a kernel function such as the standard normal density, and return $\hat{\alpha}_a^{(k)}(x) = \widehat{\mathbb{E}}_{(1-1/k)n/2}[\omega \mid X = x, A = a]$.

2.6 Guarantees for Learning

In this section, we study the finite sample error rates for Algorithm 2.1 with arbitrary first- and second-stage estimators. For this and the next section, let us fix some CDTE with pseudo-outcome as in Definition 2.4 and estimation procedures input to Algorithm 2.1. We require the following boundedness conditions.

Assumption 2.5 (Boundedness)**.** For a nuisance realization set Ξ, there exist $c_1 > 0, c_2 \geq 0, c_3 \geq 0, c_4 > 0, c_5 \geq 0$ and matrices $G, H \in \{0,1\}^{(m+1)\times(m+1)}$ such that $\forall (e, \alpha, \nu_a) \in \Xi,\ i, j, l \in \overline{1, m+1},\ \bar{\nu}_a \in \text{conv}\{(\nu_a^*, \nu_a)\}$,

- $e^*(X), e(X) \in [c_1,\ 1 - c_1]$
- $\left| D_{\nu_{a,j}} \mathbb{E}[\rho_i(Y, \nu_a) \mid X, A = a|_{\nu_a = \bar{\nu}_a} \right| \leq c_2 G_{ij}$
- $\left| D_{\nu_{a,l}} D_{\nu_{a,j}} \mathbb{E}[\rho_i(Y, \nu_a) | X = x, A = a]|_{\nu_a = \bar{\nu}_a} \right| \leq c_3 H_{jl}$
- $\det(D_{\nu_a}\{\mathbb{E}[\rho(Y, \nu_a) \mid X = x, A = a]\}\big|_{\nu_a = \bar{\nu}_a}) > c_4$
- $|\rho_i(Y, \bar{\nu}_a)| \leq c_5$

The first condition in Assumption 2.5 ensures that both treatments and controls can be observed for any X with some fixed probability. This is guaranteed in a randomized trial if $e^*(X)$ is a constant and otherwise is a standard assumption in observational studies. The other conditions encode the cross-term structure of the derivatives of our moments. In most examples, the requirement of ρ being bounded amounts to requiring Y to be bounded. Similar boundedness assumptions are often made in debiased machine learning for ATE and CATE to control remainder terms.

We can then prove the following *conditional* Neyman orthogonality for our pseudo-outcomes, which is the key step for learning guarantees.

Theorem 2.6 (Conditional Neyman Orthogonality)**.** *Suppose Assumption 2.5 holds and let* $(e, \alpha, \nu_a) \in \Xi$*. Then,*

$$\begin{aligned}
&\left\|\mathbb{E}\left[\psi(Z, e, \alpha, \nu) - \psi(Z, e^*, \alpha^*, \nu^*) \mid X\right]\right\| \lesssim \mathcal{E}(e, \alpha, \nu), \\
\mathcal{E}(e, \alpha, \nu) = \sum_{a=0}^{1} \Big(&\|\kappa_a - \kappa_a^*\| \, \|e - e^*\| \\
&+ \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} G_{ij} \|\alpha_{a,i} - \alpha_{a,i}^*\| \, \|\nu_{a,j} - \nu_{a,j}^*\| \\
&+ \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} H_{ij} \|\nu_{a,i} - \nu_{a,i}^*\| \, \|\nu_{a,j} - \nu_{a,j}^*\| \Big),
\end{aligned} \tag{2.9}$$

where the $\lesssim$ *absorbs the dependence on* c_1, c_2, c_3, c_4*.*

The result shows that whether we use the pseudo-outcome with oracle nuisances or estimated nuisances as a regression target, the difference is bounded by a *quadratic* form in the nuisance errors, wherein G, H govern which pairwise error products appear. Thus, even if nuisances are estimated slowly, the impact is marginal, as the error is squared.

We will show that, up to the error term $\mathcal{E} = \sum_{k=1}^{K} \mathcal{E}(\hat{e}^{(k)}, \hat{\alpha}^{(k)}, \hat{\nu}^{(k)})$, our estimate $\widehat{\text{CDTE}}$ produced by Algorithm 2.1 behaves as if the regression step were applied to the oracle pseudo-outcome, $\widetilde{\text{CDTE}}(x) = \widehat{E}_n[\psi(Z, e^*, \alpha^*, \nu^*) \mid X = x]$. In particular, if $\mathcal{E}$ is smaller than the regression error $\|\widetilde{\text{CDTE}} - \text{CDTE}\|$, then $\widehat{\text{CDTE}}$ has the same leading behavior as $\widetilde{\text{CDTE}}$. The exact form of this guarantee depends on the choice of last-stage regression method.

The significance is that Algorithm 2.1 will behave like regressing an unbiased observation of CDTE on X, even though we used estimated nuisances. First, this means we can learn CDTE at rates that match the complexity of that function. Second, this implies that we can directly approximate CDTE and get model-agnostic best-in-class guarantees. For example, if we use linear regression, we

would get the *best* linear approximation for CDTE at a rate of $n^{-1/2}$. This is especially important if we seek an interpretable model. In the next section, we show this also enables *inference*.

Remark 2.7. As shown in Appendix A.2, in many examples, Eq. (2.9) will contain only a few of the product terms, because many G, H entries are 0 and/or $\widehat{\alpha}(x, \widehat{\nu}) = \alpha^*(x, \widehat{\nu})$. This enables us to trade off slower rates in one nuisance for faster rates in another while maintaining the same rate for $\mathcal{E}$.

Remark 2.8. We never explicitly assume that the conditional moment restrictions identify $\nu_a^*(x)$ uniquely, only that they exist. While uniqueness is not necessary, the right-hand side of Eq. (2.9) will not vanish unless $\widehat{\nu}_a^{(k)}(x)$ converges to a *single*, *non-random* limit point $\nu_a^*(x)$. This is certainly possible even under solution multiplicity [Imbens et al., 2021, Kallus and Mao, 2022]. At the same time, usually $\nu_a^*(x)$ will be unique under very mild assumptions such as having continuous distributions.

First, we consider an empirical risk minimization algorithm (nonparametric least squares): given a class $\mathcal{F} \subset [\mathcal{X} \to \mathbb{R}]$,

$$\widehat{\mathbb{E}}_n[W \mid X = \cdot] \in \operatorname*{argmin}_{f \in \mathcal{F}} \widehat{\mathbb{E}}_n (W - f(X))^2. \tag{2.10}$$

Theorem 2.9. *Suppose Assumption 2.5 holds almost surely for* $(\hat{e}^{(k)}, \hat{\alpha}^{(k)}, \hat{\nu}^{(k)}) \in \Xi$ *with* $k \in \overline{1, K}$. *Let* $\widehat{\mathbb{E}}_n[\cdot \mid X = x]$ *be as in Eq.* (2.10) *and suppose* $\mathcal{F}$ *is convex, closed and has bracketing entropy* $\log N_{[]}(\mathcal{F}, \epsilon) \lesssim \epsilon^{-r}$ *with* $0 < r < 2$ *and that* $|f(x)| \leq c_5 \; \forall f \in \mathcal{F}, x \in \mathcal{X}$. *Then,*

$$\|\widehat{\text{CDTE}} - \text{CDTE}\| \lesssim O_p(n^{-1/(2+r)}) + \mathcal{E}. \tag{2.11}$$

The rate $O_p(n^{-1/(2+r)})$ is generally the rate for regressing a *known* target using nonparametric least squares over $\mathcal{F}$ with such bracketing entropy. For example,

if $\mathcal{F}$ is Hölder functions of smoothness β in d-dimensional inputs then it satisfies the entropy condition with $r = d/\beta$ [van der Vaart and Wellner, 1996, cor. 2.7.2] and $n^{-\beta/(2\beta+d)}$ is the optimal rate for regression in such a class [Stone, 1982]. Thus, if $\mathcal{E} = O_p(n^{-1/(2+r)})$, our regression behaves as though we supplied it with the oracle nuisances.

Obtaining such a rate for $\mathcal{E}$ is generally lax because it is *quadratic* in nuisance-estimation errors. For example, if nuisances are estimated at the much slower rate $O_p(n^{-1/(4+2r)})$, the condition is ensured. The special structure in Theorem 2.6 further permits some trade-off between the rates of different nuisances. In Appendix A.2, we give the pseudo-outcome for each of the examples in Section 2.10 and instantiate Theorem 2.6 to exactly characterize the trade-offs.

Going beyond empirical risk minimizers, leveraging Theorem 2.6 and invoking a result of Kennedy [2023a] [1] we can characterize the behavior when using a last-stage regression method satisfying certain stability properties.

Theorem 2.10. *Suppose Assumption 2.5 holds almost surely for* $(\hat{e}^{(k)}, \hat{\alpha}^{(k)}, \hat{\nu}^{(k)}) \in \Xi$ *with* $k \in \overline{1, K}$*, and that, for any targets* Y *and* W, $\widehat{\mathbb{E}}_n[Y \mid X = x] + c = \widehat{\mathbb{E}}_n[Y + c \mid X = x]$ *and* $\|\widehat{\mathbb{E}}_n[W \mid X] - \mathbb{E}[W \mid X]\| \asymp \|\widehat{\mathbb{E}}_n[Y \mid X] - \mathbb{E}[Y \mid X]\|$ *whenever* $\mathbb{E}[Y \mid X = x] = \mathbb{E}[W \mid X = x]$*. Then,*

$$\|\widehat{\mathrm{CDTE}} - \mathrm{CDTE}\| \lesssim \|\widetilde{\mathrm{CDTE}} - \mathrm{CDTE}\| + \mathcal{E}. \tag{2.12}$$

2.7 Guarantees for Inference

We next study how Algorithm 2.1 can be used for inference on the best linear projection of CDTE:

$$\gamma^* \in \operatorname*{argmin}_{\gamma \in \mathbb{R}^p} \|\,\mathrm{CDTE} - \gamma^T \phi(\cdot)\|,$$

[1] The result appears as Theorem 1 in v2 of the arxiv preprint.

for a given $\phi : \mathcal{X} \to \mathbb{R}^p$. In particular, ϕ may be subset just some of the features. Linear projections offer an interpretable view into distributional-effect heterogeneity. In the previous section, we showed Algorithm 2.1 can perform well at learning linear projections. Next, we show that we can further conduct inference on the coefficients, which can facilitate interpretation and credible conclusions.

Theorem 2.11 (Asymptotic Normality of Linear Projections)**.** *Suppose Assumption 2.5 holds and almost surely* $(\hat{e}^{(k)}, \hat{\alpha}^{(k)}, \hat{\nu}^{(k)}) \in \Xi$ *for* $k \in \overline{1, K}$*. Let* $\widehat{\gamma}$ *be the coefficient vector returned by Algorithm 2.1 when using ordinary least squares (OLS) on* $\phi(X)$ *as the regression blackbox for the final stage. Furthermore, assume* $\|\widehat{e}^{(k)} - e^*\| = o_p(1)$, $\|\widehat{\kappa}_a^{(k)} - \kappa_a^*\|\|\widehat{e}^{(k)} - e^*\| = o_p(n^{-1/2})$, $\|\widehat{\alpha}_{a,i}^{(k)} - \alpha_{a,i}^*\| = o_p(n^{-1/4})$, $\|\widehat{\nu}_{a,i}^{(k)} - \nu_{a,i}^*\| = o_p(n^{-1/4})$, $\forall i \in \overline{1, m+1}$, $k \in \overline{1, K}$*. Then,* $\widehat{\gamma}$ *satisfies*

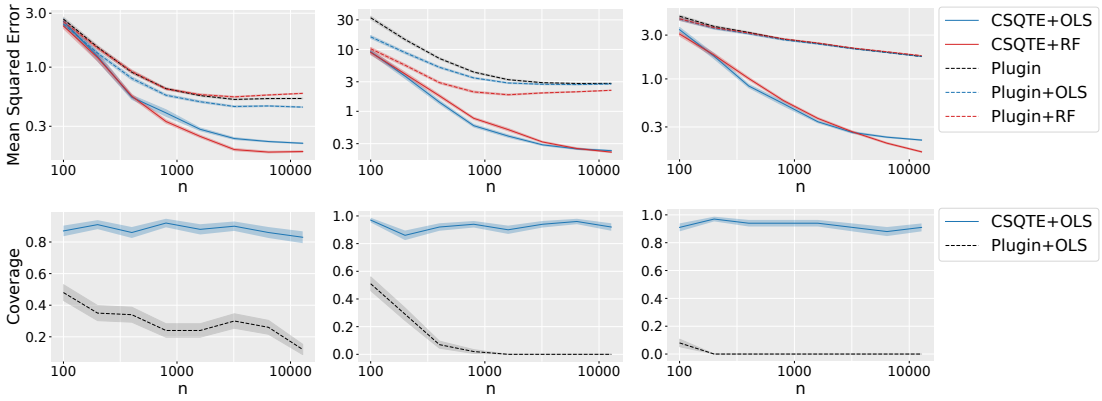
$$\sqrt{n}(\widehat{\gamma} - \gamma^*) \rightsquigarrow \mathcal{N}(0, \Sigma^*), \tag{2.13}$$

where Σ^* *is the asymptotic covariance matrix for the linear regression of* $\psi(Z, e^*, \alpha^*, \nu^*)$ *on* $\phi(X)$*.*

Theorem 2.11 implies that we can just use out-of-the-box OLS with the built-in OLS inference as the final stage regression in Algorithm 2.1. The inference results would remain valid, provided nuisances are estimated slowly but not too slowly. In particular, we should generally use robust (aka sandwich or Huber–White) standard errors, as we do not expect the projection to have homoskedastic errors.

2.8 Empirical Results

In this section, we demonstrate our method and theoretical results by applying Algorithm 2.1 to learn CSQTEs (Section 2.4.2). We first benchmark its performance in simulated data and then illustrate its use in a study of 401(k) eligi-



(a) Flexible $\hat{\mu}$ (b) Misspecified $\hat{\mu}$ (c) Slow $\hat{\mu}$

Figure 2.2: Mean squared error (MSE) and 95% confidence interval coverage for different conditional super-quantile treatment effect (CSQTE) learners in the synthetic experiment. Shaded regions show plus/minus one standard error over 100 simulations.

bility and its effect on wealth accumulation. We provide additional results for CQTEs and CfRTEs in Appendix A.3. Replication code is publicly available at `https://github.com/CausalML/CDTE`.

2.8.1 Simulation Study

We sample from the following data generating process:

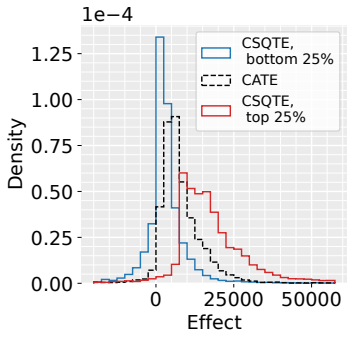
$$X \sim \text{Unif}([0,1]^{10}), \quad A \sim \text{Bernoulli}(\sigma(6X_0 - 3)),$$
$$Y \mid X, A \sim \text{Lognormal}(X_0 + AX_1, 0.2),$$

where σ is the logistic sigmoid. The coefficients in the expression for A are chosen such that the true propensity lies in the range $[0.05, 0.95]$. We seek to measure CSQTE at level $\tau = 0.75$, *i.e.*, the conditional average of values above the 0.75 conditional quantile. For this DGP, the true CSQTE at $\tau = 0.75$ is given by $\mu_1(X;\tau) - \mu_0(X;\tau) = 1.29(e^{X_0+X_1} - e^{X_0})$, which is heterogeneous in X.

We estimate $e(X)$ using logistic regression and $q_a(X;\tau)$ using a quantile random forest (QRF) [Meinshausen and Ridgeway, 2006]. We consider three op-

tions for estimating $\mu_a(X;\tau)$. First, we consider a *flexible* learner we term the superquantile RF (SQRF). SQRF uses a RF to calculate weights (like QRF) and then computes Eq. (2.4), replacing the expectation by the weighted average over the data. Second, we consider doing the same using a Gaussian kernel for calculating weights, choosing the bandwidth by Silverman's rule [Silverman, 2018]. We refer to this as the *slow* learner because it suffers badly from the curse of dimension. Third, we consider estimating $\mu_a(X;\tau)$ using ordinary least squares for the regression $\frac{1}{1-\tau}\mathbb{E}[Y\mathbb{I}[Y \geq \widehat{q}_a(x;\tau)] \mid X = x]$ using the estimated quantile $\widehat{q}_a$. We term this estimator a *misspecified* learner since it is unlikely to fully capture the complexity of the superquantile (which is an exponential function of the features). For the final stage of Algorithm 2.1, we use either an ordinary least squares model (CSQTE+OLS) or a RF (CSQTE+RF). We set $K = 5$ as the number of folds required by Algorithm 2.1. All forests use `scikit-learn` [Pedregosa et al., 2011] defaults except for the minimum leaf size which is set to $n/20$ to control overfitting.

We compare the out-of-sample mean squared error (MSE) of the CSQTE estimator with that of the naive plug-in estimator from Eq. 2.7. To account for possible smoothing by the last stage regressor, we also construct Plugin+OLS and Plugin+RF given by running an additional OLS/RF model on the cross-fitted plug-in predictions. And, when the second stage algorithm is OLS, we check whether the 95%-confidence interval OLS returns for the X_1 coefficient contains the coefficient from the true projection. The results are shown in Figure 2.2. We run 100 simulations for each $n = 100, 200, \ldots, 12800$ and evaluate MSE over a fixed set of 500 random X values. Our CSQTE learner provides uniformly strong MSE performance, and the results show this is not just a consequence of the second-stage regression. For inference, we achieve good coverage whereas



Coefficient	CSQTE Bottom 25%	CATE	CSQTE Top 25%
Intercept	−0.021	−0.95	−2.07
($10,000)	(−1.06, 1.02)	(−2.42, 0.51)	(−7.04, 2.90)
Income	0.25**	0.21	−0.05
	(0.08, 0.43)	(−0.08, 0.50)	(−1.12, 1.01)
Age	105	232**	513
	(−75, 286)	(24, 441)	(−182, 1210)
Education	−801**	16	1340
	(−1440, −164)	(−1050, 1090)	(−2490, 5180)

Figure 2.3: Estimated effect of 401(k) eligibility on financial wealth in the 401(k) application, measured for the bottom 25%, top 25%, and average of the conditional outcome distribution. The left panel shows the distribution of estimated CSQTEs and CATEs when using a random forest last stage. The right panel reports OLS projections of the CSQTE and CATE on income, age, and education. "**" indicates statistical significance at level 0.05 (p-value < 0.05).

plug-in approaches yield little to no coverage. These findings confirm our theoretical results: the quadratic dependence on nuisance errors enables oracle rates when the nuisances are estimated slowly or are misspecified (Theorem 2.6, Theorem 2.9), and we can obtain valid inference when projecting onto linear spaces (Theorem 2.11).

2.8.2 Impact of 401(k) Eligibility on Financial Wealth

We apply the CSQTE estimator to study the impact of 401(k) eligibility on net financial assets. We use the dataset from Chernozhukov and Hansen [2004], which is based on the 1991 Survey of Income and Program Participation. The data contains 9,915 observations with 9 covariates such as age, income, education, family size, marital status, IRA participation, *etc.* While the eligibility for 401(k) (the treatment A) is not assigned at random, [Poterba et al., 1994, Chernozhukov and Hansen, 2004] argue that unconfoundedness can be assumed conditional on the observed covariates. The outcome of interest (Y) is the net financial assets of an individual, defined as the sum of 401(k) balance, bank ac-

counts and interest-earning assets minus non-mortgage debt.

We apply the CSQTE estimator to understand effect heterogeneity beyond the conditional mean. Specifically, we estimate treatment effects on the bottom and top 25% of financial asset holders, conditional on covariates. We analyze the results alongside CATE estimates given by a DR-Learner [Kennedy, 2023a] as implemented by Battocchi et al. [2019]. For nuisance estimation, we use RF models with hyperparameters as in Chernozhukov et al. [2018a]. The superquantile model is SQRF described above, the quantile model is QRF, the outcome learner for CATE is an RF regression, and the propensity model is an RF classifier.

We consider two options for second-stage regressions. First, we consider using an RF model using all 9 covariates. We train the three estimators (CATE and upper/lower CSQTEs), and plot the distribution of predicted conditional effects on the 9,915 observations in Figure 2.3. We observe that 401(k) eligibility has a much higher financial impact on the high end of the conditional net worth distributions as compared to those on the lower end. While the CATE distribution reassuringly lies in between the bottom 25% and top 25% distributions, it fails to capture this disparity in effects.

To understand the heterogeneity in CSQTEs and CATEs, we use an OLS projection in the final stage. We choose the top three features that the RF last stage picked up as most important (see Figure A.3 in Appendix A.3): income, age and education. The resulting coefficients and 95% confidence intervals are depicted in Figure 2.3. For the bottom 25%, income and education are the main (statistically significant) drivers of the average effects. The positive income coefficient suggests that 401(k) eligibility provides more gains to those who potentially have the funds to invest. Likewise, the negative education coefficient means

that the effects are higher among less educated earners. We hypothesize that higher educated earners might have a more comprehensive financial education and would save and invest regardless of 401(k) eligibility, whereas 401(k) availability provides lower educated earners a default investment option. For the top 25% asset holders, higher age and education rather than income have the largest impact on 401(k) eligibility returns. This group also displays more variability as none of the coefficients are statistically significant. The CATEs provide middle-of-the-road estimates that miss out on trends at the two ends of the spectrum. Thus, CSQTEs are a powerful tool for uncovering different behaviours across the conditional distribution that a simple average would obscure.

2.9 Conclusion

We provide new tools to assess distributional effect heterogeneity through agnostic and robust learning of CDTEs in a generic framework that includes quantiles, super-quantiles, and f-risk measures. These tools can be used to analyze treatments as well as to support personalized decisions that take risk into account. We provide strong guarantees for both learning and inference, and demonstrate our methods in both synthetic and real data. We further discuss some limitations of our work in Appendix A.4.

CHAPTER 3

B-LEARNER: QUASI-ORACLE BOUNDS ON HETEROGENEOUS CAUSAL EFFECTS UNDER HIDDEN CONFOUNDING

This chapter is based on Oprescu et al. [2023].

Estimating heterogeneous treatment effects from observational data is a crucial task across many fields, helping policy and decision-makers take better actions. There has been recent progress on robust and efficient methods for estimating the conditional average treatment effect (CATE) function, but these methods often do not take into account the risk of hidden confounding, which could arbitrarily and unknowingly bias any causal estimate based on observational data. We propose a meta-learner called the B-Learner, which can efficiently learn sharp *bounds* on the CATE function under limits on the level of hidden confounding. We derive the B-Learner by adapting recent results for sharp and valid bounds of the average treatment effect [Dorn et al., 2025a] into the framework given by Kallus and Oprescu [2023a] for robust and model-agnostic learning of conditional distributional treatment effects. The B-Learner can use any function estimator such as random forests and deep neural networks, and we prove its estimates are valid, sharp, efficient, and have a quasi-oracle property with respect to the constituent estimators under more general conditions than existing methods. Semi-synthetic experimental comparisons validate the theoretical findings, and we use real-world data to demonstrate how the method might be used in practice.

3.1 Introduction

Using data to estimate the causal effect of actions is a fundamental task in medicine, economics, education research, and more. For instance, we might

wish to use patient data to estimate which patients react well to a certain medication and which patients should avoid it. In many cases, due to economic and ethical considerations, the data available for these tasks is *observational data*, i.e. data that was not collected as part of a randomized experiment. Using such data carries the risk of *unobserved confounding*: correlations between the observed interventions and outcomes that are not accounted for in the available data. For example, patients with more social support might tend to receive certain interventions over others. If the level of a patient's social support is not recorded in the data, the estimated effect of the intervention will be biased due to not observing the confounder of social support. Unobserved confounding cannot be detected from data, and its presence can lead to arbitrary and unknown bias in causal effect estimates. Such bias can in turn lead to unreliable decisions and potentially harmful interventions.

In this work we are concerned with estimating causal effects on an individual level in the presence of a limited degree of unobserved confounding. Specifically, we give a method for effectively learning upper and lower bounds on the conditional average treatment effect (CATE) function that allows for flexible nuisance estimation and high-dimensional conditioning sets like patient medical records. The degree of allowed hidden confounding can be set by domain knowledge; alternatively, we can estimate what degree of hidden confounding is needed to significantly change our understanding of the CATE for any particular instance or sub-population.

We pursue the desirable treatment effect bound properties of validity, sharpness, efficiency, and robustness. A bound is called valid if it contains the true value of the causal estimand. A sharp bound is a valid bound that contains *only* those values of the causal estimand that could emerge from a plausible data

generation process that could have produced the observed data [Ho and Rosen, 2017]. Therefore, sharp bounds are the smallest possible bounds accounting for both observational data and domain knowledge (in the form of the degree of hidden confounding), a property which is important for precise decision making under hidden confounding. In contrast, a valid bound could in principle contain extraneous values, leading to overly cautious decision making. Efficient bounds converge to their target values using as little data as possible. Typically, efficiency at best corresponds to quasi-oracle performance, where only slowly-consistent first-stage estimates are needed to achieve the same error bounds as we would obtain with access to oracle knowledge [Nie and Wager, 2021]. Finally, a robust bound will be insensitive (within limits) to biases in the constituent estimators. We formalize these properties in Section 3.3.

In this chapter, we present the B-Learner, for "bound-learner", a scalable and flexible meta-learner for estimating *bounds* on the CATE function. The B-Learner uses a partially double-robust, Neyman-orthogonal estimating equation for the valid CATE bound characterization of Dorn et al. [2025a]. For unconfounded CATE estimation, there are several well-known meta-learners such as the X-Learner [Künzel et al., 2019], DR-Learner [Kennedy, 2023a], and R-Learner [Nie and Wager, 2021]. These methods allow the user to combine essentially arbitrary learning procedures—including random forests, linear models, and deep neural networks—to estimate the CATE function efficiently. In addition to this flexibility, some of these methods enjoy desirable rate and quasi-oracle properties. The B-Learner offers analogous flexibility, rate, and quasi-oracle guarantees for CATE *bounds* estimation, together with novel guarantees on bound validity and sharpness under appropriate assumptions. We study CATE bounds under Tan's marginal sensitivity model (MSM) [Tan, 2006], which quantifies the de-

gree of unobserved confounding through odds ratios. These properties of Tan's MSM allow the B-Learner to achieve validity under notably weak assumptions.

We evaluate the B-Learner using synthetic and semi-synthetic experiments. In the synthetic experiments, the B-Learner displays quasi-oracle efficiency, requiring only a moderate amount of data for it to perform near-identically with estimated and oracle first-stage nuisances. The B-Learner also performs at least comparably to existing methods with analogous nuisances and can perform better with a well-tailored choice of second-stage regression function. In semi-synthetic experiments, we find the B-Learner is at least as effective as existing state-of-the-art models on a previously proposed benchmark. Finally, we illustrate the use of the B-Learner using real data to estimate the effect of 401(k) eligibility on financial wealth.

3.2 Related Work

To the best of our knowledge, existing methods for CATE sensitivity analysis do not simultaneously achieve all four of the properties targeted by our proposed B-Learner: validity, sharpness, efficiency, and flexibility. In particular, Kallus et al. [2019], Jesson et al. [2021], and Yin et al. [2022] provide methods that achieve validity and, to some extent, rate guarantees. These approaches begin with estimators that perform well under unconfoundedness and then optimize estimated CATE or average treatment effect (ATE) bounds subject to a subset of the constraints implied by Tan's MSM. Because they do not enforce all implications of the MSM, the resulting bounds are generally not sharp except in knife-edge cases. They also do not explicitly exploit Neyman orthogonality, and therefore do not obtain the same rate guarantees available under unconfoundedness. Finally, these methods are tailored to specific learners, rather than

offering the flexibility of a meta-learner.

Related work has also studied bounds under alternative sensitivity models. Yadlowsky et al. [2022] exploit Neyman orthogonality to obtain rate guarantees for CATE estimation and root-n guarantees for ATE estimation under Rosenbaum [2002]'s model, and show that the resulting bounds are sharp under certain outcome-symmetry conditions. Chernozhukov et al. [2022a] develop a method with root-n consistency for *average* potential outcomes, treatment effects, and derivative bounds under limits on variance and covariance, and show that their bounds are sharp provided they do not violate implications of the observed data distribution.

More broadly, there is a rich literature on sensitivity analysis for ATEs, ranging from early work such as Cornfield et al. [1959] and Rosenbaum and Rubin [1983] to more recent developments such as Colnet et al. [2022]. A full review of that literature is beyond the scope of this chapter.

3.3 Background and Setup

We work in an observational data setting using the Neyman-Rubin potential outcomes framework. We assume data is drawn from an unobservable distribution P_{full} over $(X, A, Y(1), Y(0), U)$, where $A \in \{0, 1\}$ is a binary treatment, X is a set of baseline covariates in $\mathbb{R}^d$, $Y(1)$ and $Y(0)$ are the real-valued treated and untreated potential outcomes, respectively, and $U \in \mathbb{R}^k$ is an unobserved confounder. However, we face the fundamental problem of causal inference and only observe n draws from the coarsened distribution P over the observed variables $Z = (X, A, Y)$, where we assume that $Y = Y(A)$, i.e. (causal) consistency.

We are interested in the conditional average treatment effect (CATE):

$$\tau(x) = \mathbb{E}_{P_{\text{full}}}[Y(1) - Y(0) \mid X = x].$$

The average treatment effect (ATE) is $\mathbb{E}[\tau(X)]$. When the (untestable) unconfoundedness assumption holds, formally $A \perp\!\!\!\perp Y(1), Y(0) \mid X$, then the CATE is equivalent to the difference in expected observed potential outcomes: $\tau(x) = \mathbb{E}_P[Y \mid X = x, A = 1] - \mathbb{E}_P[Y \mid X = x, A = 0]$. With the additional assumption of positivity, the CATE can be estimated with standard tools. However, the unconfoundedness assumption is untestable and often unrealistic, as we often have at least some degree of confounding unaccounted for by the observed covariates X. Therefore, we will assume unconfoundedness only holds with the addition of an unobserved $U \in \mathbb{R}^k$ for some k, such that $A \perp\!\!\!\perp Y(1), Y(0) \mid X, U$. In this case, it is possible to bound $\tau(x)$ pointwise, for example, by assuming that unobserved confounding induces only a limited divergence between P and P_{full}.

We proceed under Tan's Marginal Sensitivity Model (MSM) [Tan, 2006]:

Assumption 3.1. Let $e(x, u) = P_{\text{full}}(A = 1 \mid X = x, U = u)$ and $e(x) = P(A = 1 \mid X = x)$ be the full and observed propensity scores, respectively. We assume $e(x), e(x, u) \in (0, 1)$ and that there exists $\Lambda \geq 1$ such that the following holds almost surely under P_{full}:

$$\Lambda^{-1} \leq \frac{e(x, u)}{1 - e(x, u)} \Big/ \frac{e(x)}{1 - e(x)} \leq \Lambda.$$

The MSM imposes a bound on ratio between the full odds of treatment $e(x, u)/(1-e(x, u))$ and the observed odds of treatment $e(x)/(1-e(x))$. (The MSM is sometimes equivalently described using the log odds ratio bound $\log(\Lambda)$.) When $\Lambda = 1$, Assumption 3.1 is equivalent to the classic assumption of unconfoundedness with respect to the observed X. As Λ grows away from 1, greater unobserved confounding is allowed under the MSM and we can generally only

estimate bounds on the CATE. In this chapter, our goal is to characterize these bounds, which describe a notion of "causal" uncertainty in the CATE estimate.

Remark 3.2. The sensitivity parameter Λ is a user-defined hyper-parameter as it specifies how much confounding to allow for. Choosing a suitable Λ is an ongoing area of study. Hsu and Small [2013] propose a procedure where we assess Λ values that correspond to dropping observed covariates and using domain knowledge to judge whether we omitted variables as important as these. Inversely, as our intervals increase with Λ, we can seek the Λ where a conclusion or decision would be overturned and judge whether the implied confounding is plausible. Ultimately, choosing Λ is domain-specific.

Notation We now define the main notation, with a more detailed notation table available in Appendix B.1. To unify the analysis for upper (largest plausible CATE) and lower (smallest plausible CATE) bounds, we employ the convention that $+, -$ indicators symbolize upper and lower bounds, respectively. For nuisance functions (e.g. quantiles), these signs also encode the dependence on $\alpha = \Lambda/(\Lambda + 1)$ (and Λ) which we otherwise generally suppress in the remainder of the chapter. We define the conditional outcome quantile and shorthand quantile notation:

$$q_c^*(x, a) = \inf\{\beta : F(\beta \mid x, a) \geq c\}$$

$$q_+^*(x, a) = q_\alpha^*(x, a), q_-^*(x, a) = q_{1-\alpha}^*(x, a).$$

The $\pm$ and $\mp$ symbols signal that an equation should be read twice, once with $\pm = +, \mp = -$ and once with $\pm = -, \mp = +$ (see example in Appendix B.1, Table B.1). For conciseness and clarity, we focus our main discussion on CATE upper bounds. In Appendix B.2, we provide a similar analysis of CATE lower bounds.

3.3.1 Properties of bound estimates

Our goal is to estimate the *identified set*: the set of CATEs that can be obtained in the unobserved distribution P_{full} generating the observed distribution P and satisfying the requirements of Assumption 3.1.

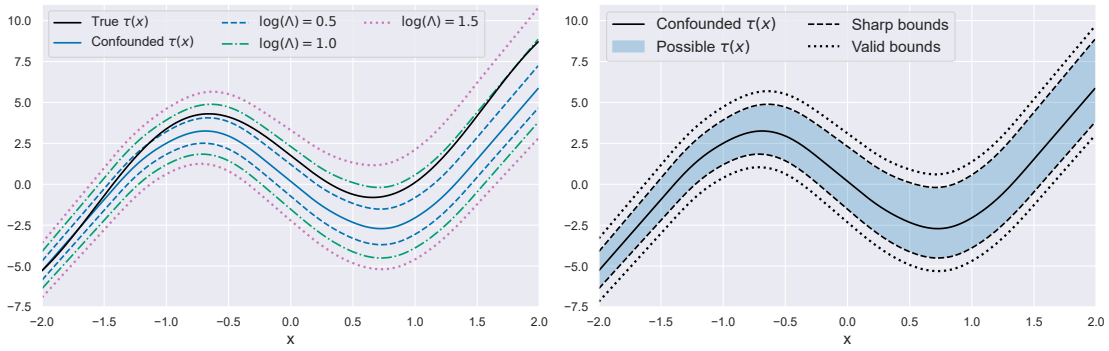
Definition 3.3. The *identified set* of estimands under Assumption 3.1 is the set of estimands that can be obtained for a distribution Q over $(X, A, Y(1), Y(0), U)$ such that the distribution of (X, A, Y) under Q matches the observed distribution P and $\Lambda^{-1} \leq \frac{Q(A=1|X=x,U=u)}{Q(A=0|X=x,U=u)} \Big/ \frac{e^*(x)}{1-e^*(x)} \leq \Lambda$ almost surely. Let $\mathcal{M}(\Lambda)$ be the set of distributions Q that the observed data $Z = (X, Y, A)$ and Assumption 3.1 cannot rule out. Then, the *sharp* (upper) bounds on the identified set of conditional average potential outcomes and CATEs for a given point x are given by:

$$Y^+(x, a) \equiv \sup_{Q \in \mathcal{M}(\Lambda)} \mathbb{E}_Q[Y(a) \mid X = x]$$

$$\tau^+(x) \equiv \sup_{Q \in \mathcal{M}(\Lambda)} \mathbb{E}_Q[Y(1) - Y(0) \mid X = x].$$

Lower bounds follow symmetrically by replacing the suprema with infima. We note that the requirements of Assumption 3.1 decouple across x and are convex, so finding the identified set reduces to finding pointwise bounds. As we will see in Section 3.3.2, the CATE upper bounds $\tau^+(x)$ depend only on the observed distribution of data Z and the sensitivity parameter Λ. We can therefore ask what good properties we might want estimates $\widehat{\tau}^+(x)$ to have. We suggest four desirable properties for bound estimation, of which the last two are closely linked:

Valid Estimates If $\widehat{\tau}^+(x) < \tau^+(x) - o_p(1)$, then our estimated bounds fail to cover the identified set and rule out plausible CATEs even asymptotically, which would be undesirable. Conversely, bound characterizations $\bar{\tau}$ satisfying



(a) CATE bounds for different values of Λ (b) Sharp and valid bounds

Figure 3.1: Example of a CATE function under hidden confounding, with true odds ratio Λ^* given by $\log(\Lambda^*) = 1.0$. The true $\tau(x)$ is the unobserved CATE in the full distribution, $\mathbb{E}_{P_{\text{full}}}[Y(1) - Y(0) \mid X = x]$. The confounded $\tau(x)$ is the biased estimand under assumed uncondoundedness, $\mathbb{E}_P[Y \mid X = x, A = 1] - \mathbb{E}_P[Y \mid X = x, A = 0]$. Panel (3.1a) shows sharp CATE bounds for different values of Λ; panel (3.1b) illustrates the difference between *valid* and *sharp* bounds.

$\bar{\tau}^+(x) \geq \tau^+(x)$ are called "valid" in the partial identification literature, since Assumption 3.1 implies $\tau^+(x) \geq \mathbb{E}[Y(1) - Y(0) \mid X = x]$ Ho and Rosen [2017]. Valid bounds (illustrated in Figure 3.1) give us some but not all information from our assumptions: every value they rule out is implausible, but some values they do not rule out may be implausible as well. We relax the notation and say that bound estimates $\widehat{\tau}$ are *valid* if $\widehat{\tau}^+(x) \geq \tau^+(x) - o_p(1)$.

Sharp Estimates If $\widehat{\tau}^+(x) > \tau^+(x) + o_p(1)$, then our estimated bounds would fail to rule out impossible CATEs asymptotically under our assumptions. Exact characterizations of the identified sets are called "sharp" in the partial identification literature Ho and Rosen [2017]. Sharpness is a stronger property than validity. We use lax notation to say that bound estimates $\widehat{\tau}$ are *sharp* if $\widehat{\tau}^+(x) = \tau^+(x) + o_p(1)$.

Efficient and Robust Estimates We would like our bound estimates to converge to their limits at desirable rates and have multiple chances at sharp or valid limits. Ideally, we would be able to learn CATE bounds at the same rate as

we could obtain under unconfoundedness. These properties relate to "double robust" estimators and may require constructing Neyman-orthogonal characterizations of valid, and ideally sharp, bounds.

3.3.2 Identification and Estimation of Sharp Bounds

In this section, we use results from Dorn et al. [2025a] to show how we can identify and estimate sharp CATE bounds from the observed data distribution, P. In order to express the sharp bounds, we introduce the following pseudo-outcomes from Dorn et al. [2025a] that will correspond to the Conditional Value at Risk and the unobserved outcome bounds under Assumption 3.1:

$$H_{\pm}(z, \bar{q}) = \bar{q}(x, a) + \frac{1}{1-\alpha} \{y - \bar{q}(x, a)\}_{\pm}$$

$$R_{\pm}(z, \bar{q}) = \Lambda^{-1} y + (1 - \Lambda^{-1}) H_{\pm}(z, \bar{q})$$

$$\rho^*_{\pm}(x, a, \bar{q}) = \mathbb{E}[R_{\pm}(z, \bar{q}) \mid X = x, A = a].$$

We use the shorthand $\rho^*_{\pm}(x, a) = \rho^*_{\pm}(x, a, q^*_{\pm})$ to write the $\rho^*_{\pm}$ function evaluated at the true conditional quantiles $q^*_{\pm}$ (which will end up corresponding to sharp bounds). The quantity $\text{CVaR}_{\pm}(x, a) := \mathbb{E}[H_{\pm}(z, q^*_{\pm}) \mid X = x, A = a]$ is known as the Conditional Value at Risk [Artzner et al., 1999, Kallus, 2023]. In the distribution $Y \mid X = x, A = a$, $\text{CVaR}_{+}(x, a)$ is the expectation above the $(1 - \alpha)$ quantile, whereas $\text{CVaR}_{-}(x, a)$, is the expectation below the α quantile. Hence, the pseudo-outcomes H and R correspond to the Conditional Value at Risk and conditional unobserved potential outcome, respectively.

Let $\mu^*(x, a) = \mathbb{E}[Y \mid X = x, A = a]$ be the conditional outcome regression in the observed data. Note that we can write the conditional potential outcome under Q as $\mathbb{E}_Q[Y(a) \mid X = x] = P[A = a \mid X = x]\mu^*(x, a) + P[A = 1 - a \mid X = x]\mathbb{E}_Q[Y(1 - a) \mid X = x, A = a]$ since Q must be consistent with the observed distribution P. Thus, it suffices to bound the conditional unobserved potential

outcome $\mathbb{E}_Q[Y(1-a) \mid X = x, A = a]$, which leads to the following result in terms of $\rho^*_\pm(x,a) = \rho^*_\pm(x,a,q^*_\pm)$:

Result 3.4 (Sharp bounds, Dorn et al. [2025a])**.** *The conditional average unobserved potential outcome $\mathbb{E}_Q[Y(1-a) \mid X = x, A = a]$ has sharp upper and lower bounds under Assumption 3.1 given by $\rho^*_+(x,a)$ and $\rho^*_-(x,a)$, respectively. Thus, the sharp bounds on the conditional average potential outcomes can be written as:*

$$Y^+(x,1) = e^*(x)\mu^*(x,1) + (1 - e^*(x))\rho^*_+(x,1)$$

$$Y^-(x,0) = (1 - e^*(x))\mu^*(x,0) + e^*(x)\rho^*_-(x,0).$$

The sharp CATE upper bound is further given by $\tau^+(x) = Y^+(x,1) - Y^-(x,0)$.

Thus, the bounds are a convex combination of the conditional outcome function $\mu^*(x,a)$ and the corresponding conditional CVaR terms, all of which can be estimated from P. As Λ grows, both the weight on the CVaR term in ρ^* grows and the CVaR term itself become more extreme. If the wrong putative quantile $\bar{q}$ is used instead of the true q^*, the CVaR term moves the bound in a conservative yet valid direction. Finally, the difference between sharp conditional average potential outcome bounds $\tau^+(x)$ clearly yields valid CATE bounds; those bounds are shown to be sharp by arguments outside the scope of this chapter [Dorn and Guo, 2022]. As we will see, this characterization of sharp and valid bounds alone will be insufficient for quasi-oracle estimation.

Pseudo-outcome Regression for Quasi-Oracle Estimation The expression of $\tau^+(x)$ suggests a plug-in strategy. We can estimate e, μ, and ρ through classification and regression and obtain bound estimates. However, such plug-in estimators are known to suffer from excessive bias due to the estimated nuisances Kennedy [2023a], Kallus and Oprescu [2023a], especially when the nuisance functions are more complex than the CATE bounds. We follow the Kallus

and Oprescu [2023a] strategy and derive an efficient pseudo-outcome for the bounds based on the relevant influence function; we then regress that pseudo-outcome on X. We build on this literature to similarly provide an estimator for sharp CATE bounds with desirable properties beyond those of the plug-in estimators implied by Result 3.4.

Henceforth we will refer to e, q, ρ as nuisances as they will need to be estimated from data.

3.4 B-Learner: Pseudo-Outcome Regression for Doubly Robust Sharp CATE Bounds

We propose a debiased learning procedure that consists of regressing a carefully constructed and nuisance-debiasing pseudo-outcome on covariates Kennedy [2023a], Kallus and Oprescu [2023a].

Definition 3.5 (Pseudo-outcome for CATE bounds)**.**
Let $\widehat{\eta} = (\widehat{e}, \widehat{q}_-(\cdot, 0), \widehat{q}_+(\cdot, 1), \widehat{\rho}_-(\cdot, 0), \widehat{\rho}_+(\cdot, 1)) \in \Xi$ be a set of estimated nuisances. We define the pseudo-outcome corresponding to the bounds for $Y^+(x, 1)$, $Y^-(x, 0)$, and $\tau^+(x)$ from Result 3.4 by

$$\phi_1^+(Z, \widehat{\eta}) = AY + (1 - A)\widehat{\rho}_+(X, 1) + \frac{(1 - \widehat{e}(X))\, A}{\widehat{e}(X)} \cdot \left(R_+(Z, \widehat{q}_+(X, 1)) - \widehat{\rho}_+(X, 1)\right),$$

$$\phi_0^-(Z, \widehat{\eta}) = (1 - A)Y + A\widehat{\rho}_-(X, 0) + \frac{\widehat{e}(X)(1 - A)}{(1 - \widehat{e}(X))} \cdot \left(R_-(Z, \widehat{q}_-(X, 0)) - \widehat{\rho}_-(X, 0)\right),$$

$$\phi_\tau^+(Z, \widehat{\eta}) = \phi_1^+(Z, \widehat{\eta}) - \phi_0^-(Z, \widehat{\eta}).$$

The expressions in Definition 3.5 depend purely on the observed data distribution P, and so can be viewed as statistical estimands to be learned from the observed distribution.

Algorithm 3.1 The B-Learner (detailed in Appendix B.5)

Input: Data $\{(X_i, A_i, Y_i) : i \in \{1, \ldots, n\}\}$, folds $K \geq 2$, nuisance estimators, regression learner $\widehat{\mathbb{E}}_n$

1: **for** $k \in \{1, \ldots, K\}$ **do**
2: Use data $\{(X_i, A_i, Y_i) : i \not\equiv k - 1 \pmod{K}\}$ to construct nuisance estimates

$$\widehat{\eta}^{(k)} = (\widehat{e}^{(k)}, \widehat{q}^{(k)}, \widehat{\rho}^{(k)}).$$

3: **for** each i such that $i \equiv k - 1 \pmod{K}$ **do**
4: Set $\widehat{\phi}^{+}_{\tau,i} = \phi^{+}_{\tau}(Z_i, \widehat{\eta}^{(k)})$
5: **end for**
6: **end for**

Output: $\widehat{\tau}^{+}(x) = \widehat{\mathbb{E}}_n[\widehat{\phi}^{+}_{\tau} \mid X = x]$

When $\Lambda = 1$ and unconfoundedness holds, the expression for $\phi^{+}_{\tau}(Z, \widehat{\eta})$ reduces to the familiar doubly-robust pseudo-outcome for CATE estimation, $\widehat{\mu}(X, 1) - \widehat{\mu}(X, 0) + \frac{A - \widehat{e}(X)}{\widehat{e}(X)(1 - \widehat{e}(X))}(Y - \widehat{\mu}(X, A))$ [Kennedy, 2023a, Knaus, 2022].

The pseudo-outcome is based on the efficient influence function of the estimand $\mathbb{E}[\tau^{+}(X)]$, so as we will see, small errors in the nuisance estimation lead to "doubly small" (second-order) errors in the $\widehat{\tau}^{+}(x)$ estimates. This special structure orthogonalizes the $\widehat{\rho}_{+}$ estimation error in the plug-in bound estimand $AY + (1 - A)\widehat{\rho}_{+}$ using the added term $\frac{(1 - \widehat{e}(X))A}{\widehat{e}(X)}(R_{+} - \widehat{\rho}_{+})$ that debiases $\widehat{\rho}_{+}$ estimation error. The weighted CVaR terms $\rho^{*}_{\pm}(X, a, \bar{q})$ involve an objective which is sharpest when $\bar{q}_{\pm} = q^{*}_{\pm}$ and which turns out to have a second-order dependence on $\bar{q}_{\pm} - q^{*}_{\pm}$. Thus, quantile regression errors will move the pseudo-outcome in a conservative but still valid direction and consistent quantile regression errors will have favorable rate properties.

B-Learner We call our full two-stage estimation procedure the *B-Learner*. Our procedure is summarized in Algorithm 3.1 (see Appendix B.5 for a detailed version). In the first stage, we estimate the nuisances (outcome regression, propensity score, CVaR) with K-fold cross-fitting and construct Neyman-Orthogonal

pseudo-outcome estimates based on Definition 3.5. In the second stage, we regress the estimated pseudo-outcomes on our covariates X, resulting in an estimated CATE bound function. As we will now see, the properties of this function depend on both the choice of nuisance estimators and the second-stage model.

Nuisance Estimation The propensity score $e^*(x)$ can be estimated using any standard probabilistic binary classifier. The quantiles $q^*_\pm$ can be likewise estimated using any of several standard quantile regression methods Yu and Jones [1998], Meinshausen and Ridgeway [2006], Athey et al. [2019]. The modified outcome regression $\rho^*_\pm(x,a) = \Lambda^{-1}\mu^*(x,a) + (1-\Lambda)^{-1}CVaR_\pm(x,a)$ is less standard, but it can be learned by either treating the CVaR pseudo-outcome $R_\pm$ as an outcome, or separately learning the μ^* and $\text{CVaR}_\pm$ components of $\mathbb{E}[R_\pm \mid X = x, A = a]$. In the first approach, where we plug in the estimated quantiles into the expression for $R_\pm(Z, \bar{q})$ and then regress $R_\pm$ onto X using any standard regressor, further sample splitting is theoretically required for estimating q^* and ρ^*. In the second approach, we can learn the μ^* and CVaR components on the same sample and then weight them accordingly to obtain estimates of ρ^*. The outcome regression $\mu^*(x,a)$ can be estimated via any regression learner and $\text{CVaR}_\pm$ can be likewise estimated using several existing approaches Athey et al. [2019], Kallus and Oprescu [2023a].

3.5 Theoretical properties of the B-Learner

We now describe the theoretical properties of our estimator. All proofs are in Appendix B.4. In Section 3.5.1, we use Kallus and Oprescu [2023a]'s generic approach and Dorn et al. [2025a]'s validity results to study the bias of the pseudo-outcome with first-stage nuisances. The pointwise bias from the **sharp** bounds is on the order of $|\widehat{e} - e^*||\widehat{\rho} - \rho^*| + (\widehat{q} - q^*)^2$. When the quantiles are inconsistent,

$(\widehat{q}_+ - q_+^*)^2$ and $(\widehat{q}_- - q_-^*)^2$ do not vanish. The pseudo-outcome bounds still remain **valid** in expectation, and any bias in the direction of failing to cover the identified CATE set disappears at a rate on the order of $|\widehat{e} - e^*||\widehat{\rho} - \rho^*(\cdot, \widehat{q})|$. In Section 3.5.2, we characterize the second-stage regression and we show that we can learn CATE bounds at a **rate** dominated by the complexity of the target class. As a result, the estimator has robustness properties from the product-of-errors bias, with two chances at **sharp** bounds in L_2 norm and two chances at **valid** bounds on average. Our main text focuses on ERM-based second stage estimators with L_2 sharp bound guarantees. We show similar guarantees hold pointwise for linear smoother second-stage estimators in Appendix B.3.2.

3.5.1 Pseudo-outcome properties

We first analyze the bias in our proposed pseudo-outcomes.

Definition 3.6 (Conditional Pseudo-outcome Bias)**.** Take $\widehat{\eta} \in \Xi$ be a set of estimated nuisances and let $\diamond \in \{0, 1, \tau\}$. We define the signed conditional pseudo-outcome bias:

$$\mathcal{E}_\diamond^+(x; \widehat{\eta}) = \mathbb{E}\left[\phi_\diamond^+(Z, \widehat{\eta}) - \phi_\diamond^+(Z, \eta^*) \mid X = x\right], \text{ and}$$
$$\mathcal{E}_\diamond^-(x; \widehat{\eta}) = \mathbb{E}\left[\phi_\diamond^-(Z, \widehat{\eta}) - \phi_\diamond^-(Z, \eta^*) \mid X = x\right].$$

It immediately follows from Definition 3.6 that $\mathcal{E}_\tau^+(x; \widehat{\eta}) = \mathcal{E}_1^+(x; \widehat{\eta}) - \mathcal{E}_0^-(x; \widehat{\eta})$ and $|\mathcal{E}_\tau^+(x; \widehat{\eta})| \leq |\mathcal{E}_1^+(x; \widehat{\eta})| + |\mathcal{E}_0^-(x; \widehat{\eta})|$. The pseudo-outcome bias can be understood as the error incurred when performing pseudo-outcome regression with estimated nuisances rather than oracle nuisances. While any bias is undesirable, bias in one direction is worse. When $\mathcal{E}_\tau > 0$, the pseudo-outcomes are biased in a conservative but still valid direction. When $\mathcal{E}_\tau < 0$, the expected pseudo-outcomes are too aggressive and in expectation exclude plausible CATEs.

Our pseudo-outcomes fit into the framework of Kallus and Oprescu [2023a] since the estimands and the nuisances are the solutions of conditional moment restrictions (see Proof of Theorem 3.8). Thus, under mild boundedness conditions, we can leverage their results to upper bound $|\mathcal{E}_\diamond^+|$.

Assumption 3.7 (Boundedness)**.** Let $\widehat{\eta} \in \Xi$ be a set of estimated nuisances, and take $\overline{\eta} \in \text{conv}\{(\eta^*, \widehat{\eta})\}$.

(i) $P(\epsilon \leq e^*(x), \widehat{e}(x) \leq 1 - \epsilon) = 1$ for some $\epsilon > 0$.

(ii) $Y, \overline{q}_+(\cdot, 1), \overline{q}_-(\cdot, 0), \overline{\rho}_+(\cdot, 1), \overline{\rho}_-(\cdot, 0), f(\overline{q}_+(x, 1) \mid x, 1), f(\overline{q}_-(x, 0) \mid x, 0)$ are all uniformly bounded.

The first condition in Assumption 3.7 is a standard requirement known as positivity, ensuring that both treatments and controls can be observed for any X with non-zero probability. The second condition is a common boundedness assumption often made in debiased machine learning for ATE and CATE in order to control the growth of $|\mathcal{E}_\tau^+|$. We now state the conditional Neyman orthogonality result we require, which we derive using the tools from Kallus and Oprescu [2023a] and Dorn et al. [2025a].

Theorem 3.8 (Pseudo-Outcome Conditional Neyman Orthogonality)**.** *Suppose Assumption 3.7 holds. Then a Neyman-orthogonal characterization of the conditional outcome moment* $\mathbb{E}[AY + (1 - A)\rho_+^*(X, 1) - Y^+(X, 1) \mid X] = 0$ *has the form of* ϕ_1^+ *from Definition 3.5, and the symmetric result holds for* ϕ_0^-*. The absolute bias of the CATE upper bound has the product of rates bound:*

$$\left|\mathcal{E}_\tau^+(x; \widehat{\eta})\right| \lesssim |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_+(x, 1) - \rho_+^*(x, 1)| + |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_-(x, 0) - \rho_-^*(x, 0)| + (\widehat{q}_+(x, 1) - q_+^*(x, 1))^2 + (\widehat{q}_-(x, 0) - q_-^*(x, 0))^2.$$

The undesirable direction of bias has the more favorable bound in terms of $\rho^*(x, a, \widehat{q})$:

$$\mathcal{E}_\tau^+(x; \widehat{\eta}) \gtrsim -|\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_+(x, 1) - \rho_+^*(x, 1, \widehat{q}_+)| \\ - |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_-(x, 0) - \rho_-^*(x, 0, \widehat{q}_-)|.$$

Theorem 3.8 lets us characterize the pseudo-outcome biases.

Sharp Pseudo-outcome Bias An immediate implication is that the pseudo-outcome bias for the CATE bound is pointwise "doubly sharp" Dorn et al. [2025a]: its bias tends to zero if $\widehat{q}_\pm$ and one of $\widehat{e}$ or $\widehat{\rho}_\pm$ are consistent, and the bias converges faster than the individual nuisances if all nuisances are consistent.

Valid Pseudo-outcome Bias In some cases it may be difficult to estimate quantiles consistently or at a sufficient rate for the quantile error $(\widehat{q} - q^*)^2$ to vanish faster than $|\widehat{e} - e^*|\,|\widehat{\rho} - \rho^*|$. If so, the absolute value of pseudo-outcome bias relative to sharp bounds might be relevant to the second-stage estimates, but the level of bias in the direction of failing to cover the identified set still disappears at a product rate $|\widehat{e} - e^*|\,|\widehat{\rho} - \rho^*(\cdot, \widehat{q})|$. The pseudo-outcome estimator is therefore "doubly valid" Dorn et al. [2025a]: its undesirable bias tends to zero if one of $\widehat{e}$ or $\widehat{\rho}_\pm$ is consistent, and the **rate** of bias goes to zero faster than the individual nuisances if both are consistent.

Next, we leverage these results to illustrate the quasi-oracle properties of our B-Learner.

3.5.2 ERM-based Estimators

We consider Algorithm 3.1 with an empirical risk minimization (ERM) algorithm as the second-stage estimator. In other words, given a class of functions

$\mathcal{F} \subset [\mathcal{X} \to \mathbb{R}]$, the regression learner $\widehat{\mathbb{E}}_n$ satisfies:

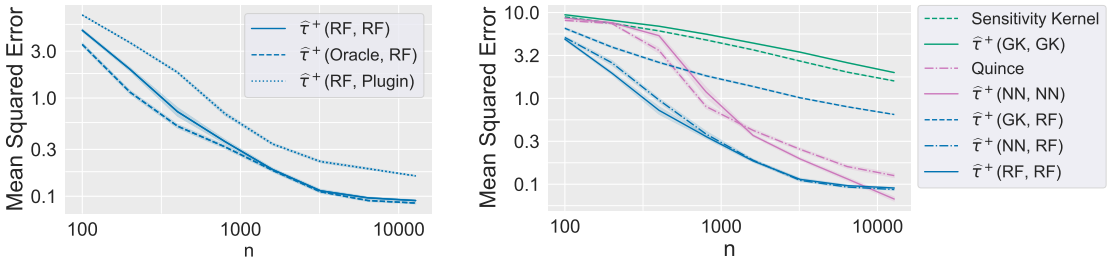
$$\widehat{\mathbb{E}}_n\left[\widehat{\phi}_\tau^+ \mid X = \cdot\right] \in \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\phi}_{\tau,i}^+ - f(X_i)\right)^2. \tag{3.1}$$

In this scenario, the error rates of our estimation procedure depend on the complexity of the class $\mathcal{F}$. These were studied in the context of learning with nuisance components in several works including Foster and Syrgkanis [2023a], Kallus and Oprescu [2023a]. The implication of Theorem 3.8 is we can immediately apply Kallus and Oprescu [2023a]'s Theorem 2 in our setting, employing bracketing entropy as a class complexity measure. We note that bracketing entropy is a *global* technique, with guarantees on the L_2 loss over the support of the estimand, in contrast with the *local* methods presented in Appendix B.3.2 which enable pointwise guarantees.

Corollary 3.9 (Rates for ERM Estimators, Theorem 2 from Kallus and Oprescu [2023a])**.** *Suppose Assumption 3.7 holds for* $\widehat{\eta}^{(k)} \in \Xi, k \in \{1, ..., K\}$. *Let* $\mathcal{E}_\tau^+(x) := \sum_{k=1}^K \mathcal{E}_\tau^+(x; \widehat{\eta}^{(k)})$ *and let* $\widehat{\mathbb{E}}_n[\cdot \mid X = x]$ *be as in Eq.* (3.1). *Further, suppose* $\mathcal{F}$ *is convex and closed and has bracketing entropy* $\log N_{[]}(\mathcal{F}, \epsilon) \lesssim \epsilon^{-r}$ *with* $0 < r < 2$ *and that* $|f(x)|$ *is bounded* $\forall f \in \mathcal{F}, x \in \mathcal{X}$. *Then,*

$$\|\widehat{\tau}^+(x) - \tau^+(x)\| \lesssim O_p(n^{-1/(2+r)}) + \|\mathcal{E}_\tau^+(x)\|.$$

Second-stage Sharp Consistency and Robustness When $\|\mathcal{E}_\tau^+(x)\| = o_p(1)$ and the conditions above hold, Corollary 3.9 shows that ERM estimates are L_2 consistent for the sharp CATE bounds. Learners satisfying the conditions of Corollary 3.9 include sparse linear models, neural networks, kernel classes Foster and Syrgkanis [2023a], and Besov, Sobolev, Hölder-type function classes Nickl and Pötscher [2007]. L_2 consistency of the pseudo-outcome bias follows if $\widehat{q}$ and one of $\widehat{e}$ or $\widehat{\rho}$ are L_2 consistent.



(a) B-Learner with its oracle and plugin variants

(b) B-Learner, *Sensitivity Kernel*, and *Quince*

Figure 3.2: Mean squared error (MSE) for different learners of the upper CATE bound $\widehat{\tau}^+$ in the synthetic hidden-confounding experiment. Shaded regions show plus/minus one standard error over 50 simulations.

Second-stage Sharp Rates If $\|\mathcal{E}_\tau^+(x)\| = o_p\left(n^{-1/(2+r)}\right)$ and the conditions of Corollary 3.9 hold, the pseudo-outcome bias has a negligible contribution to the CATE bounds estimation error. Thus, the estimation error is equivalent to the error as if the nuisances were known, a result known as the "quasi-oracle property" [Nie and Wager, 2021]. Because the pseudo-outcome bias involves the product of rates, it will be sufficient to ask all pseudo-outcome nuisances to be consistent at an $o_p(n^{-1/4})$ rate. We give an example of sufficient conditions for our estimator to be oracle efficient (the property we synonymously call "quasi-oracle" in our main text) in Appendix B.3.1.

Second-stage Validity When the quantile estimates are inconsistent, we cannot apply Corollary 3.9 directly. Still, we will have two chances to derive CATE bound estimates that are valid on average. In Appendix B.3.2, we show that linear smoothers can yield stronger pointwise validity guarantees.

Corollary 3.10 (ERM Validity on Average)**.** *Assume the conditions of Corollary 3.9 are satisfied and for all $f \in \mathcal{F}$ and $c \in \mathbb{R}$ we have $f + c \in \mathcal{F}$. If $\|\widehat{q}_+(\cdot, 1) - \bar{q}_+(\cdot, 1)\| = o_p(1)$ and $\|\widehat{q}_-(\cdot, 0) - \bar{q}_-(\cdot, 0)\| = o_p(1)$ for a (potentially inconsistent) putative quantile function $\bar{q}$ and either $\|\widehat{e} - e^*\| = o_p(1)$ or both $\|\widehat{\rho}_+(\cdot, 1) - \rho_+^*(\cdot, 1, \bar{q}_+)\| = o_p(1)$ and*

$\|\widehat{\rho}_-(\cdot, 0) - \rho^*_-(\cdot, 0, \bar{q}_-)\| = o_p(1)$*, then the estimated CATE bounds are valid on average in the sense that* $\frac{1}{n}\sum_{i=1}^n \widehat{\tau}^+(X_i) - \tau^+(X_i) \geq -o_p(1)$.

3.6 Experiments

In this section, we demonstrate our method on synthetic and semi-synthetic datasets, as well as on a real-world case study. We first benchmark the B-Learner using a synthetic example similar to that in Kallus et al. [2019]. We then illustrate how CATE bound estimators can be used for treatment deferral by using the hidden confounding variant of the IHDP dataset introduced by Jesson et al. [2021]. For both sets of experiments, we compare with state-of-the-art methods proposed by Kallus et al. [2019] (*Sensitivity Kernel*) and Jesson et al. [2021] (*Quince*[1]). We illustrate the usage of the B-Learner with real data through a case study of 401(k) eligibility effects on wealth. While we have focused our discussion on CATE upper bounds, our real data experiments also require estimating the CATE lower bounds we discuss in Appendix B.2. Details about the data generation processes, specific model implementation, hyperparameter selection and validation procedures used are given in Appendix B.6. We provide replication code at `https://github.com/CausalML/BLearner`.

While the *Sensitivity Kernel* approach uses Gaussian kernels and the *Quince* model uses Bayesian neural networks, the B-Learner (Algorithm 3.1) is flexible in the types of estimators allowed for both the first- and second-stage learners. We therefore compare three classes of nuisance and second-stage estimators: Random Forests (RF), Gaussian Kernels (GK), and Bayesian Neural Networks (NN). Whenever possible, we use the same hyperparameters and validation routine across models. For example, the B-Learner with NN nuisances uses

[1]Jesson et al. [2021] train an ensemble of several models, which is a computationally intensive task. For the purposes of this section, we do not ensemble any of the compared methods.

the exact same neural networks as *Quince*.

We denote the upper bound given by the B-Learner output (Algorithm 3.1) by $\widehat{\tau}^{+}(\{1^{\text{st}} \text{ stage}\}, \{2^{\text{nd}} \text{ stage}\})$ (e.g. $\widehat{\tau}^{+}(RF, RF)$) to indicate the type of first- and second-stage learners used. For insight into the theoretical properties of our estimator, we also provide an oracle first-stage estimator $\widehat{\tau}^{+}(\text{Oracle}, \{2^{\text{nd}} \text{ stage}\})$ which uses the true nuisances in the pseudo-outcome calculation, as well as a "plug-in" estimator $\widehat{\tau}^{+}(\{1^{\text{st}} \text{ stage}\}, \text{Plugin})$ which plugs in the estimated nuisances into the expressions from Result 3.4.

3.6.1 Simulated Data

Our synthetic dataset is sampled as follows:

$$X \sim \text{Unif}([-2, 2]^5), \quad A \mid X \sim \text{Bern}(\sigma(0.75X_0 + 0.5)),$$
$$Y \sim \mathcal{N}((2A - 1)(X_0 + 1) - 2\sin((4A - 2)X_0), 1),$$

where σ is the sigmoid function. We wish to provide an estimate $\widehat{\tau}^{+}(x)$ for the CATE upper bound under a level of confounding given by $\log \Lambda = 1$. With this simulation, it is straightforward to obtain the true nuisances e^*, μ^*, ρ^*. These, along with Result 3.4, allow us to determine the true value $\tau^{+}(x)$ of the upper bound. We run 50 simulations for sample sizes $n = 100, 200, 400, ..., 12800$ and evaluate the different models on a fixed test set of 400 data points initially drawn at random. We compare the mean squared error (MSE) performance of each estimator with respect to the true bound and depict our findings in Figure 3.2.

In Figure 3.2a, we study the MSE convergence rates of the $\widehat{\tau}^{+}(RF, RF)$ estimator, along with its oracle and plug-in variants. The convergence rate of our estimator matches the rate of the oracle estimator. That is, Algorithm 3.1 with more than a few hundred observations performs essentially as well as if the estimator had access to the true, oracle nuisances. This confirms our theoretical

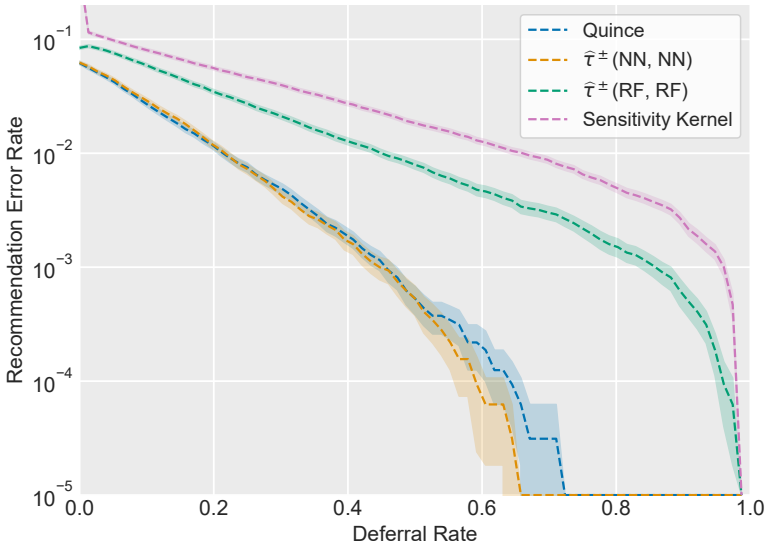


Figure 3.3: IHDP hidden-confounding experiment: treatment recommendation error rate as a function of the deferral rate. The x-axis reflects the fraction of units for which the method defers treatment recommendation rather than making a possibly incorrect recommendation.

results from Corollary 3.9 in that small errors in the nuisance estimation lead to second-order errors in $\widehat{\tau}(x)$. Moreover, we see that the simple plug-in estimator suffers from so-called plug-in bias for every value of n, as anticipated. The B-Learner MSE improvement slows for large n, which we expect reflects our use of rules-of-thumb to extrapolate hyperparameters to large samples.

In Figure 3.2b, we benchmark our estimator against *Sensitivity Kernel* and *Quince* for various first- and second- stage combinations. We see that using the same nuisances (GKs and NNs, respectively) leads to our method performing comparably with competitors. However, the B-Learner with NN or GK first stages and with RF second stage learners performs better than the state-of-the-art methods. This result underscores the importance of flexibility in choosing nuisance estimators, a key property of our method.

3.6.2 IHDP Hidden Confounding

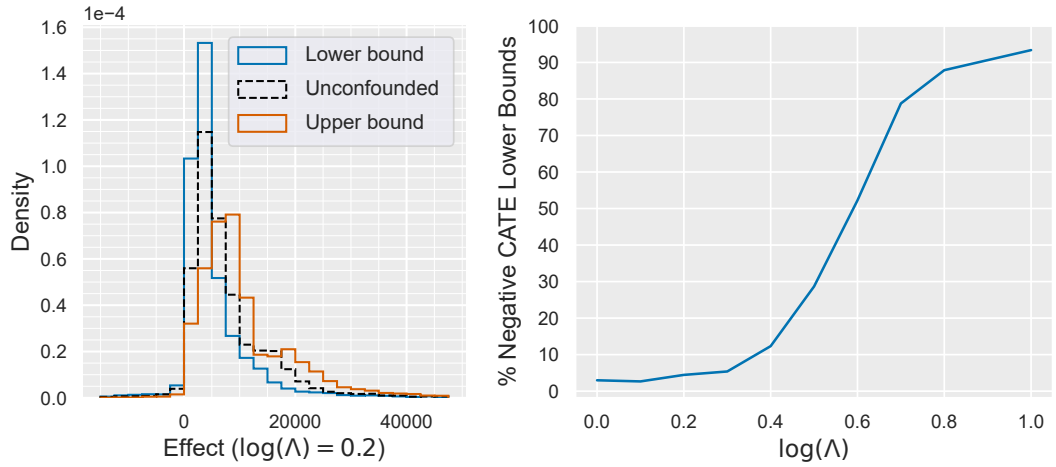
We now show how the B-Learner can be used for other causal inference tasks, such as informing *deferral policies* for treatment recommendations. We replicate the experiment from Jesson et al. [2021] on IHDP Hidden Confounding. The dataset is multi-dimensional, has low overlap, and has hidden confounding due to a single covariate being hidden from the training models. The dataset contains synthetic potential outcomes generated according to the response surface B described by Hill [2011]. We use the same deferral policy as in Jesson et al. [2021], namely, the policy either recommends treatment or defers to an expert. We make a treatment recommendation (either $A = 0$ or $A = 1$, according to the sign of CATE estimate) if and only if the predicted CATE interval excludes zero.

We measure model performance in terms of recommendation error rate across multiple deferral rates. The deferral rate is the fraction of observations for which we defer the action decision to the expert. The error rate is the percentage of observations for which we recommend the wrong treatment, among those in which we did not defer. Note that in this experiment we know the best treatment for each unit since we simulate both potential outcomes, although these effects do not correspond to a sharp bound under Assumption 3.1.

We compare two different variants of B-Learner: $\widehat{\tau}^{\pm}(RF, RF)$ and $\widehat{\tau}^{\pm}(NN, NN)$ with *Sensitivity Kernel* and *Quince*. We see in Figure 3.3 that the RF B-Learner outperforms the GK-based *Sensitivity Kernel* method, and that the best performing methods are the NN B-Learner and *Quince* which perform very similarly.

3.6.3 Impact of 401(k) Eligibility on Wealth Distribution

We apply the B-Learner to illustrate the impact of hidden confounding in a study of 401(k) eligibility and its effects on financial wealth. We use the real-



(a) (b)

Figure 3.4: Estimated bounds on the effect of 401(k) eligibility on financial wealth under hidden confounding. Panel (3.4a) shows the distribution of lower and upper CATE bounds for $\log \Lambda = 0.2$. Panel (3.4b) shows the fraction of lower bounds that are negative as $\log(\Lambda)$ varies from 0.1 to 1.0.

world dataset from Chernozhukov and Hansen [2004] that draws on the 1991 Survey of Income and Program Participation. The treatment of interest is 401(k) eligibility, while the target outcome is the net financial assets of an individual (taken as the aggregate of 401(k) balance, bank accounts and interest-earning assets minus non-mortgage debt).

This 401(k) eligibility dataset has been used in many analyses Poterba et al. [1994], Chernozhukov and Hansen [2004], often assuming unconfoundedness holds given observed covariates and finding a strong positive effect. However, unconfoundedness is an untestable assumption, so here we explore the uncertainty in the (conditional) treatment effects under varying degrees of hidden confounding. To that end, we apply the B-Learner algorithm repeatedly for different settings of Λ: $\log \Lambda = 0.1, 0.2, ..., 1.0$. The nuisances are all estimated using Random Forest models with hyperparameters as in [Chernozhukov et al., 2018a]. We also estimate the CATE under assumed unconfoundedness ($\log \Lambda = 0$ which corresponds to the DR-Learner Kennedy [2023a]).

In Figure 3.4, we plot the distribution of predicted conditional effects on the 9,915 observations for $\log \Lambda = 0.2$ as well as the fraction of negative lower bound effects (frequency of $\mathbb{I}(\widehat{\tau^-}(x) \leq 0)$) as we vary Λ. For lower values of Λ, the majority of lower bounds are still positive, which means that under those levels of confounding, most true conditional treatment effects are still positive. However, as we increase Λ, more and more of the true effects could be negative as the lower bound is comprised of mostly negative effects. For example, at $\log(\Lambda) = 0.6$, about half of the CATE lower bounds are negative which is to be interpreted as: if the data were truly confounded at this level, 50% of the effects measured as positive could in reality have been negative due to unobserved confounders. Regardless of what Λ level is most appropriate here, we see the B-Learner is a powerful tool for practitioners who wish to conduct what-if experiments for potential unobserved confounding.

3.7 Conclusion

We presented the B-Learner, a meta-learner for estimating bounds on the CATE function. The B-Learner can use any learning method as its base learners, including random forests and neural nets. We showed that the B-Learner provides bound estimates that are valid, sharp, robust, and have quasi-oracle rate properties, making it (to the best of our knowledge) the first CATE sensitivity analysis method with all these properties. Experiments validate our theoretical findings, show that the B-Learner is comparable in performance to existing state-of-the-art methods, and demonstrate it can be used with real-world data to gain insight into the uncertainty of estimated causal effects.

CHAPTER 4

EFFICIENT AND SHARP OFF-POLICY EVALUATION IN ROBUST MARKOV DECISION PROCESSES

This chapter is based on Bennett et al. [2025], developed jointly with Kaiwen Wang. My contributions focused on deriving the efficient influence function and analyzing the orthogonal and efficient estimator for the robust policy value. The chapter fits the dissertation's broader theme of reliable causal inference under unreliable assumptions.

We study the evaluation of a policy under best- and worst-case perturbations to a Markov decision process (MDP), using transition observations from the original MDP, whether they are generated under the same or a different policy. This is an important problem when there is the possibility of a shift between historical and future environments, *e.g.* due to unmeasured confounding, distributional shift, or an adversarial environment. We propose a perturbation model that allows changes in the transition kernel densities up to a given multiplicative factor or its reciprocal, extending the classic marginal sensitivity model (MSM) for single time-step decision-making to infinite-horizon RL. We characterize the sharp bounds on policy value under this model – *i.e.*, the tightest possible bounds based on transition observations from the original MDP – and we study the estimation of these bounds from such transition observations. We develop an estimator with several important guarantees: it is semiparametrically efficient, and remains so even when certain necessary nuisance functions, such as worst-case Q-functions, are estimated at slow, nonparametric rates. Our estimator is also asymptotically normal, enabling straightforward statistical inference using Wald confidence intervals. Moreover, when certain nuisances are estimated inconsistently, the estimator still provides valid, albeit possibly not sharp, bounds on the policy value. We validate these properties in numerical

simulations. The combination of accounting for environment shifts from train to test (robustness), being insensitive to nuisance-function estimation (orthogonality), and addressing the challenge of learning from finite samples (inference) together leads to credible and reliable policy evaluation.

4.1 Introduction

Offline policy evaluation (OPE) from historical data is crucial in domains where active, on-policy experimentation is costly, risky, unethical, or otherwise operationally infeasible. Relevant domains range from medicine to finance and recommendation systems. However, whenever historical data is used to study future behavior, there is a concern of non-stationarity – shift between the environment generating the data (training environment) and the environment in which a policy will be deployed (test environment). This may occur, *e.g.*, due to general distributional shifts in the environment over time, unobserved confounding in the observed historical data, or adversarial elements of the environment (such as other agents) that may react when the agent is deployed. While standard OPE in offline reinforcement learning (ORL) accounts for the change between the logging and evaluation policies, it may overlook the fact that the Markov decision process (MDP) too has changed. While this issue is particularly critical in high-stakes domains, it is broadly appealing to understand how value shifts across different environments in any application domain.

Robust MDPs [Iyengar, 2005, Nilim and El Ghaoui, 2005] model unknown environments by allowing an adversary to choose from any one environment in a set. Therefore, they offer a natural model for unknown environment shifts by simply considering all environments to which we could possibly shift. A variety of work addresses questions such as planning in a known robust MDP

[Wiesemann et al., 2013, Mannor et al., 2016, Goyal and Grand-Clement, 2023] as well as online learning Badrinath and Kalathil [2021], Wang and Zou [2021]. Here we focus on a purely statistical estimation question: given observations of transitions from some unknown transition kernel, we wish to estimate the worst-case (or best-case) value of a given evaluation policy in a robust MDP, defined by a set of MDPs whose transition functions are centered around the observed transition kernel.

This setting captures the previously studied unconfounded robust OPE problem [Wang et al., 2024a], where the observed transition kernel corresponds to an MDP, and the observed transitions are the result of applying some logging policy within this MDP. In such cases, the goal is to estimate policy values that are robust to future changes in the MDP dynamics. However, our setting is more general in that it also captures problems where the observed transitions are confounded by some unobserved variables, in which case they do *not* correspond to observations from the transition kernel of an MDP. In this case, the robust MDP and the robust policy value estimates are designed to account for worst-case (or best-case) impact of this confounding bias. In either case, as in ORL, we emphasize that we do *not* know the observational MDP, and can only access it via a sample of transitions. Furthermore, even in the simple case with no unmeasured confounding, in a notable departure from standard ORL, the problem can be difficult even if the logging and evaluation policies are the same (the usually easy on-policy setting), since the policy can induce very different visitation distributions in the original and perturbed MDPs.

Such robust offline evaluation from transition data was considered in recent work [Panaganti et al., 2022, Bruns-Smith and Zhou, 2023]. We build on this recent work by focusing the question of statistically *efficient* and *robust* estima-

tion of the *sharp* bounds (*i.e.*, the tightest possible given the data). Previous work focused on evaluation using only the Q-function under the worst-case environment (in some cases under a relaxation of the adversary, leading to loose bounds). Thus, any error in its estimation translates directly to error in evaluation. In other words, estimating this function flexibly and nonparametrically can yield bounds that converge slowly and are not semiparametrically efficient. Moreover, without a clear characterization of the estimator's variability, we cannot construct valid confidence bands, which risks producing bounds that are too tight.

We address these challenges by developing an orthogonalized estimation method that combines several nuisance functions: the worst-case Q-function, the state-visitation frequency in the worst-case environment, and a threshold function characterizing the worst-case transition kernel. Our first key result is that, to first order, our estimator behaves as a sample average using the true values of these functions without having to estimate them at all, provided we just estimate them at certain slow nonparametric rates. This ensures that we obtain a $\sqrt{n}$-rate of estimation even when the nuisance functions are estimated more slowly, and that our estimator is asymptotically normal. This allows for the construction of confidence bands on the bounds, providing assurance that the true bound is captured. We further show that our asymptotic variance is in fact the minimum variance among all regular and asymptotically linear (RAL) estimators, ensuring semiparametric efficiency. Our second key result is that even if we do not estimate some of the nuisance functions correctly, we are still consistent to sharp or valid bounds. That is, even when we are biased due to misestimation of nuisances, our bias (if any) only enlarges our bounds, so they remain valid. We illustrate these guarantees numerically. Collectively, these

guarantees lend substantial credibility to the bounds generated by our method.

Our contributions are summarized as follows:

1. We provide novel algorithms and analysis for learning robust Q-functions (Section 4.4) and robust visitation density ratios (Section 4.5) under the function approximation setting.
2. We derive the sharp and efficient estimator for the robust policy value, which is optimal in the local-minimax sense and is the gold standard in semiparametric estimation (Section 4.6).
3. We empirically validate the efficiency and sharpness of our approach (Section 4.7).

4.2 Related Work

4.2.1 Unobserved Confounding in Sequential Decision-Making

OPE in robust MDPs is related to OPE bounds in confounded MDPs, where the behavior policy and the transition kernel are influenced by unobserved confounders. The constraint Eq. (4.1) that defines our target robust MDP aligns with the Marginal Sensitivity Model (MSM) [Tan, 2006] employed in sensitivity analysis for causal inference. Yet, unlike the MSM, which limits the ratio of policy densities, our approach directly constrains the ratio of the transition kernels. Our formulation can be viewed as a generalization of the MSM from traditional two-action no-horizon causal effects (where the constrains coincide) to multi-action infinite-horizon discounted MDPs, where the next state is the "potential outcome". In that sense, our model essentially serves as an outcome-based sensitivity model [Bonvini et al., 2022]. This distinction is crucial as it enables our model to subsume the policy-based MSM in cases where the pol-

icy is confounded. Nonetheless, the reverse does not hold, and the policy-based MSM does not imply a transition kernel-based MSM for $A > 2$. This point is further corroborated by Bruns-Smith and Zhou [2023], who explore the policy-based MSM within confounded MDPs and obtain *non-sharp* identification bounds when $A > 2$. In contrast, our approach yields *sharp* identification in general, regardless of the number of actions and without placing assumptions on the behavior policy, which may or may not be confounded.

Bruns-Smith [2021] also considered an MSM-like model in the transition kernel but their formulation assumes $A = 2$. Kallus and Zhou [2020] operates under the setting of Bruns-Smith and Zhou [2023] and required tabular states. We note that all these works including ours considers *i.i.d.* confounders at each step, which translates to a robust MDP with (s, a)-rectangularity and ensures that the worst-case problem is still an MDP rather than a POMDP. The importance of this assumption was verified by Namkoong et al. [2020], who showed that without it, the non-memoryless confounder can create exponential-in-horizon changes in value.

4.2.2 Neyman Orthogonality and Semiparametric Efficient Estimation

We leverage a body of research focusing on learning with nuisances functions (e.g., Q-functions) that we need to estimate from data but are not the primary target (e.g., policy value). Much of this research [Chernozhukov et al., 2018a, Foster and Syrgkanis, 2023b, van der Laan et al., 2011, Semenova and Chernozhukov, 2021, Belloni et al., 2017b, Chernozhukov et al., 2018b, among others] aims to identify Neyman-orthogonal estimators, which are first order orthogonal (insensitive) to nuisance errors. This literature is tightly linked to the semiparametric efficient estimation literature since Neyman-orthogonal scores can arise naturally from efficient influence functions [Ichimura and Newey, 2022,

Schick, 1986]. Going beyond the no-horizon causal inference setting, some explore such estimators in off-policy sequential-decisions contexts [Kallus and Uehara, 2020, Lewis and Syrgkanis, 2021, Chernozhukov et al., 2022b, Kennedy, 2019, Laan and Robins, 2003]. Notably, Kallus and Uehara [2022] derive efficient influence functions and orthogonal estimation for standard, non-robust OPE in infinite-horizon RL, which corresponds to our unconfounded, no-uncertainty case ($\Lambda = 1$).

Moving beyond point-identified settings, some works explore orthogonality and efficiency for partial identification and sensitivity analysis. In the causal inference literature, efficient/orthogonal estimation in the no-horizon setting has been studied extensively under several sensitivity models [Dorn et al., 2025a, Bonvini et al., 2022, Chernozhukov et al., 2022a, Oprescu et al., 2023]. Closest to our work is Dorn et al. [2025a] who provide an orthogonal estimator and convergence rates under the MSM [Tan, 2006], which coincides with our setting under $\gamma = 1$. In the sequential setting, [Namkoong et al., 2020] considers confounding at a single time step under the MSM, but their estimator is not orthogonal when the quantile function is unknown. Bruns-Smith and Zhou [2023] provide a fitted-Q-iteration learner with an orthogonalized loss function, but not orthogonal/efficient estimates of worst-case policy value.

4.3 Background and Setup

We consider an MDP with state space $\mathcal{S}$, action space $\mathcal{A}$, transition kernel $P(s' \mid s, a)$, reward function $r(s, a) \in [0, 1]$ and initial state distribution $d_1 \in \Delta(\mathcal{S})$. We do not require $\mathcal{S}$ or $\mathcal{A}$ to be finite. We assume r and d_1 are known for simplicity, and it is standard to extend our analysis to when they are unknown. We are given a dataset $\mathcal{D}$ of n *i.i.d.* tuples (s_i, a_i, r_i, s_i') such that $(s_i, a_i) \sim \nu$, $s_i' \sim P(\cdot \mid s, a)$ and

$r_i = r(s_i, a_i)$, where ν is an arbitrary data-generating distribution. For discount factor $\gamma \in [0, 1)$, let the Q function be the discounted cumulative rewards under a policy $\pi : \mathcal{S} \to \mathcal{A}$, $Q_{\pi,P}(s, a) = \mathbb{E}_{\pi,P}[\sum_{t=0}^{\infty} \gamma^t r_t(s_t, a_t) \mid s_1 = s, a_1 = a]$. Similarly, define the value function as $V_{\pi,P}(s) = Q_{\pi,P}(s, \pi)$, where we use the notation $f(s, \pi) := \mathbb{E}_{a\sim\pi(s)}[f(s, a)]$ for any function $f : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$.

We are interested in estimating the value of a fixed target policy π_t (a.k.a. evaluation policy) in an unobserved MDP with a feasible perturbed transition kernel U. We say U is a feasible perturbation of the observed, nominal kernel P if for all s, a, s': we have

$$\Lambda^{-1}(s, a) \leq \frac{\mathrm{d}U(s' \mid s, a)}{\mathrm{d}P(s' \mid s, a)} \leq \Lambda(s, a) \tag{4.1}$$

where $\Lambda(s, a) \in [1, \infty)$ is a sensitivity parameter chosen by the practitioner. On the extremes, $\Lambda = 1$ corresponds to no-confounding (*i.e.*, classic OPE setting) and $\Lambda = \infty$ corresponds to maximal-confounding (*i.e.*, worst or best outcome). We denote the set of all feasible perturbations of P by $\mathcal{U}(P)$, which is an s, a-rectangular set [Mannor et al., 2016]. We define the best- and worst-case Q functions of π_t as

$$Q^+(s, a) := \sup_{U\in\mathcal{U}(P)} Q_{\pi_t,U}(s, a); \qquad Q^-(s, a) := \inf_{U\in\mathcal{U}(P)} Q_{\pi_t,U}(s, a). \tag{4.2}$$

Thus, the goal in this chapter is to estimate the best- and worst-case value of π_t at the initial state,

$$V^{\pm}_{d_1} := (1 - \gamma)\mathbb{E}_{s_1\sim d_1}[V^{\pm}(s_1)]. \tag{4.3}$$

where $V^{\pm}(s) = \mathbb{E}_{a\sim\pi_t(s)}[Q^{\pm}(s, a)]$ and the $\pm$ symbol signals that an equation should be read twice, once with $\pm = +$ and once with $\pm = -$. For clarity, we focus the discussion in the main text on estimating the worst-case policy value, $V^-_{d_1}$. We provide a similar analysis for policy values under best-case perturbations ($V^+_{d_1}$) in Appendix C.2.

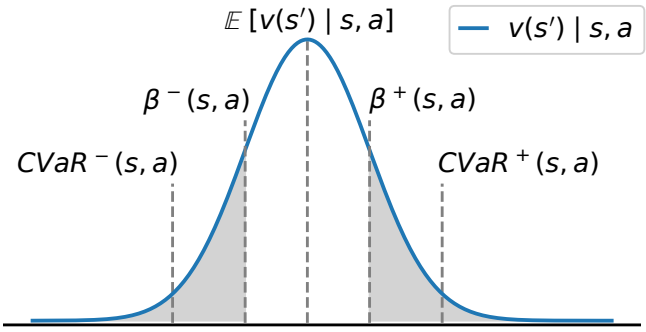


Figure 4.1: Lower and upper conditional value-at-risk (CVaR) and corresponding quantiles β for the conditional distribution $v(s') \mid s, a$. The figure illustrates the lower-tail and upper-tail risk functionals used in the robust off-policy evaluation framework.

Compared to standard OPE, robust OPE is more challenging since the best- and worst-case transition kernels $U^{\pm}$ are unobserved as our dataset $\mathcal{D}$ is generated under P. For example, standard OPE is easy in the on-policy case *i.e.*, if $\mathcal{D}$ were generated by π_t, but our problem is still "off-data" and non-trivial.

Discounted Visitation Distributions For any transition kernel U, define the discounted visitation distribution of π_t under U as: $d^{\pi_t,\infty}_{d_1,U}(s) := (1-\gamma)\sum_{h=1}^{\infty}\gamma^{h-1}d^{\pi_t,h}_{d_1,U}(s)$, where $d^{\pi_t,h}_{d_1,U}(s)$ is the probability of reaching state s in the Markov chain induced by U and policy π_t starting from $d_1(\cdot)$. We use $d^{-,\infty}$ as shorthand for $d^{\pi_t,\infty}_{d_1,U^-}$, where U^- denotes the worst-case kernel in $\mathcal{U}(P)$.

Bellman-type Operators. For any function $f : \mathcal{S}\times\mathcal{A} \to \mathbb{R}$ and transition kernel U, recall the Bellman operator is defined as $\mathcal{T}_U f(s,a) := r(s,a) + \gamma\mathbb{E}_U[f(s',\pi_t) \mid s,a]$. For robust OPE, we define the following robust analog $\mathcal{T}^+_{\text{rob}} f(s,a) := r(s,a) + \gamma \sup_{U\in\mathcal{U}(P)} \mathbb{E}_U[f(s',\pi_t) \mid s,a]$ and $\mathcal{T}^-_{\text{rob}} f(s,a) := r(s,a) + \gamma \inf_{U\in\mathcal{U}(P)} \mathbb{E}_U[f(s',\pi_t) \mid s,a]$. Moreover, we define $\mathcal{J}_U f(s,a) := \gamma\mathbb{E}_U[f(s',\pi_t) \mid s,a] - f(s,a)$. For any linear operator $\mathcal{T}$, also let $\mathcal{T}'$ denote its adjoint: that is, for all $f, g \in L_2(\nu)$, $\langle f, \mathcal{T} g\rangle = \langle \mathcal{T}' f, g\rangle$, where $\langle\cdot,\cdot\rangle$ is the inner product in $L_2(\nu)$.

Conditional Value-at Risk (CVaR) For a random variable X, its upper/lower CVaRs at level $\tau \in [0, 1]$ is defined as the average outcome of the upper/lower τ-fraction of cases, and are formally defined as follows [Rockafellar and Uryasev, 2002]:

$$\mathrm{CVaR}_\tau^+(X) := \min_{b\in\mathbb{R}}\Big\{b + \tau^{-1}\mathbb{E}[(X-b)_+]\Big\},$$
$$\mathrm{CVaR}_\tau^-(X) := \max_{b\in\mathbb{R}}\Big\{b + \tau^{-1}\mathbb{E}[(X-b)_-]\Big\},$$

where $y_+ := \max(0, y)$ and $y_- := \min(0, y)$ for $y \in \mathbb{R}$. The optima are attained at the upper/lower τ-th quantile of X which we denote as $\beta_\tau^+(X)/\beta_\tau^-(X)$, *i.e.*,

$$\mathrm{CVaR}_\tau^+(X) := \beta_\tau^+(X) + \tau^{-1}\mathbb{E}[(X-\beta_\tau^+(X))_+],$$
$$\mathrm{CVaR}_\tau^-(X) := \beta_\tau^-(X) + \tau^{-1}\mathbb{E}[(X-\beta_\tau^-(X))_-].$$

If X has a cumulative distribution function (CDF) which is differentiable at $\beta_\tau^\pm(X)$, its CVaRs simplify to $\mathrm{CVaR}_\tau^+(X) = \mathbb{E}[X \mid X \geq \beta_\tau^+(X)]$ and $\mathrm{CVaR}_\tau^-(X) = \mathbb{E}[X \mid X \leq \beta_\tau^-(X)]$. In this chapter, τ will often be set to $(\Lambda + 1)^{-1} \in [0, 0.5]$.

Notation We use $x \lesssim y$ to mean that $x \leq Cy$ holds for some universal constant C. The indicator function $\mathbb{I}[p]$ takes value 1 if p is true and 0 otherwise. For a measure μ, we let $\|f\|_\mu := (\mathbb{E}_\mu|f(X)|^2)^{1/2}$ denote the L_2 norm of f, provided it exists. When μ is clear from context, we also use $\|f\|_p := (\mathbb{E}|f(X)|^p)^{1/p}$ to denote the L_p norm of f and $\|f\|_{p,n} := (\mathbb{E}_n|f(X)|^p)^{1/p}$ to denote the empirical analog. For a data sample of size n, we define the empirical mean as $\mathbb{E}_n[f(X)] = \frac{1}{n}\sum_{i=1}^n f(x_i)$. For a nuisance function f, we reserve f^* as its true value and $\widehat{f}$ as the learned value from data. Moreover, we employ + and − to denote functions corresponding to best- and worst-case bounds, respectively. See Appendix C.1 for a comprehensive notation table.

4.3.1 Background: Non-robust OPE

We provide a quick primer on the double RL (DRL) estimator for classic OPE in non-robust MDPs [Kallus and Uehara, 2020], which combines estimates of the Q-function and density ratio w to achieve orthogonality, double robustness and semiparametric efficiency. This sets the stage for our orthogonal estimator in Section 4.6, which generalizes DRL to robust MDPs by incorporating the robust Q-function and density ratio in the worst-case MDP, as described in Section 4.4 and Section 4.5 respectively.

The DRL estimator involves two nuisances: (1) q, for which the oracle (true value) is the Q-function of the target policy Q^{π_t}, and (2) w, for which the oracle is the density ratio of the target policy's visitation distribution and the data distribution $w^{\pi_t} = \mathrm{d}d^{\pi_t,\infty}_{d_1,P}/\mathrm{d}\nu$. In this section, let $\eta = (w, q)$ denote the DRL nuisances (outside this section, we use η to denote our robust estimator's nuisances) and let $\eta^\star = (w^{\pi_t}, Q^{\pi_t})$ denote their true values, then the recentered efficient influence function (EIF) of $V^{\pi_t}_{d_1}$ in non-robust MDPs is given by:

$$\psi^{\mathsf{DRL}}(s, a, s'; w, q) = V^{\pi_t}_{d_1} + w(s, a) \cdot (r(s, a) + \gamma q(s', \pi_t) - q(s, a)).$$

The DRL estimator uses cross-fitting to learn nuisances $\widehat{\eta}^{[k]}$ on all data excluding the k-th fold $\mathcal{D}^k$, for $k = 1, 2, \ldots, K$ and estimates the OPE value via:

$$\widehat{V}^{\mathsf{DRL}}_{d_1} = \frac{1}{n} \sum_{k=1}^{K} \sum_{(s,a,s') \in \mathcal{D}^k} \psi^{\mathsf{DRL}}(s, a, s'; \widehat{\eta}^{[k]}).$$

As we will see, this paves the way for the EIF of the robust value (Theorem 4.9) and our orthogonal estimator (Algorithm 4.3). There are two main guarantees for DRL: double robustness and semiparametric efficiency. Let r^w_n and r^q_n be rate functions depending on $n = |\mathcal{D}|$ such that $\|\widehat{q}^{[k]} - Q^{\pi_t}\|_2 \leq r^q_n$ and $\|\widehat{w}^{[k]} - w^{\pi_t}\|_2 \leq r^w_n$. Then, DRL enjoys $|\widehat{V}^{\mathsf{DRL}}_{d_1} - V^{\pi_t}_{d_1}| \leq O_p(n^{-1/2} + r^w_n r^q_n)$, which confers the algorithm

double robustness properties. Moreover, if Σ^{ope} is the efficiency bound (*i.e.*, minimum achievable asymptotic variance among RAL estimators in nonparametric models for (s, a, s')), then $\sqrt{n}(\widehat{V}_{d_1}^{\mathsf{DRL}} - V_{d_1}^{\pi_t}) \xrightarrow{d} \mathcal{N}(0, \Sigma^{\mathsf{ope}})$. We seek similar guarantees for our orthogonal robust estimator.

4.4 Robust Q-Function Estimation with Fitted-Q Evaluation

In this section, we identify the robust Q-function using the robust Bellman equation and then derive convergence rates for iteratively minimizing the robust Bellman error.

4.4.1 Identification of the worst-case Q-function

The robust worst-case Q-function of π_t, denoted as Q^-, satisfies the robust Bellman equation $Q^-(s, a) = \mathcal{T}_{\mathsf{rob}}^- Q^-(s, a), \forall s, a$ since the uncertainty set $\mathcal{U}(P)$ factorizes over s, a [Iyengar, 2005]. While these equations may seem intractable due to the inf in the definition of $\mathcal{T}_{\mathsf{rob}}^-$, Bruns-Smith and Zhou [2023] showed that $\mathcal{T}_{\mathsf{rob}}^-$ has a closed form solution in terms of the CVaR under the *observed* kernel P.

Lemma 4.1. *Set $\tau(s, a) = (\Lambda(s, a) + 1)^{-1}$. Then, for any $q : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$,*

$$\mathcal{T}_{\mathsf{rob}}^- q(s, a) = r(s, a) + \gamma \Lambda^{-1}(s, a) \mathbb{E}[v(s') \mid s, a] + \gamma(1 - \Lambda^{-1}(s, a)) \operatorname{CVaR}_{\tau(s,a)}^-[v(s') \mid s, a],$$

where $v(s') = \mathbb{E}_{a' \sim \pi_t(s')}[q(s', a')]$, and $\mathbb{E}, \operatorname{CVaR}_\tau$ are under the observed kernel $P(\cdot \mid s, a)$.

Lemma 4.1 implies that Q^- is identified via the following equation of observable distributions:

$$\begin{aligned} Q^-(s, a) = r(s, a) &+ \gamma \Lambda^{-1}(s, a) \mathbb{E}[Q^-(s', \pi_t) \mid s, a] \\ &+ \gamma(1 - \Lambda^{-1}(s, a)) \operatorname{CVaR}_{\tau(s,a)}^-[Q^-(s', \pi_t) \mid s, a]. \end{aligned}$$

Under no confounding ($\Lambda(s, a) = 1$), this recovers the classic Bellman equation.

Algorithm 4.1 RobustFQE: Iterative fitting for estimating Q^- and β_τ^-.

1: **Input:** Number of iterations M, dataset $\mathcal{D}$ of size n, Q-function class $\mathcal{Q}$.
2: Initialize $\widehat{v_0^-}(s) = 0$.
3: **for** $i = 1, 2, \ldots, M$ **do**
4: Set $\mathcal{D}_i = \mathcal{D}[ni/M : n(i+1)/M]$.
5: Split $\mathcal{D}_i$ into two halves, $\mathcal{D}_i^{(1)}$ and $\mathcal{D}_i^{(2)}$.
6: Using $\mathcal{D}_i^{(1)}$ and the quantile regression oracle QR, estimate the $\tau(s,a)$-lower quantile of $\widehat{v_{i-1}^-}(s')$, $s' \sim P(\cdot \mid s, a)$, and denote it by $\widehat{\beta_i^-}(s,a)$.
7: For $(s, a, s') \in \mathcal{D}_i^{(2)}$, define the pseudo-outcome

$$\begin{aligned} &y^-(s,a,s') \\ &= r(s,a) + \gamma\Lambda^{-1}(s,a)\widehat{v_{i-1}^-}(s') \\ &\quad + \gamma(1-\Lambda^{-1}(s,a))\Big(\widehat{\beta_i^-}(s,a) + \tau^{-1}(s,a)\mathbb{E}_{a'\sim\pi_{\mathrm{t}}(s')}[(\widehat{q_i^-}(s',a') - \widehat{\beta_i^-}(s,a))_-]\Big). \end{aligned}$$

8: Estimate $\widehat{q_i^-}$ by solving

$$\widehat{q_i^-} \leftarrow \operatorname*{argmin}_{q\in\mathcal{Q}} \frac{1}{|\mathcal{D}_i^{(2)}|} \sum_{(s,a,s')\in\mathcal{D}_i^{(2)}} (y^-(s,a,s') - q(s,a))^2.$$

9: **end for**
10: **Output:** $\widehat{q_M^-}, \widehat{\beta_M^-}$.

4.4.2 Estimating the Robust Q-Function with Robust FQE

In this section, we estimate Q^- via an iterative fitting algorithm based on fitted Q-evaluation (FQE) [Munos and Szepesvári, 2008]. Our algorithm RobustFQE (Algorithm 4.1) proceeds for M iterations with two main steps in each iteration i. First, in line 6, we estimate the lower-quantile of $\widehat{v}_{i-1}(s') \mid s, a$. Here, we assume access to an oracle QR for quantile regression, which is a well-established problem, allowing for the use of various existing algorithms. Second, in line 7, we solve the tractable robust Bellman equation in Lemma 4.1 with the CVaR term estimated by its orthogonal estimating equation with the learned quantiles [Olma, 2021]. By orthogonally estimating CVaR, we achieve second-order dependence on the quantile estimation errors from the first step. Next, we minimize the mean squared error using a general function class,

$\mathcal{Q} \subset \mathcal{S} \times \mathcal{A} \mapsto [0, (1-\gamma)^{-1}]$.

To enable convergence guarantees, we make two assumptions. First, we assume a specific convergence rate for the quantile regression oracle, which can be guaranteed under certain smoothness conditions [Bhattacharya and Gangopadhyay, 1990, Takeuchi et al., 2006, Meinshausen and Ridgeway, 2006, El Ghouch and Genton, 2009, Cevid et al., 2022, Racine and Li, 2017, Elie-Dit-Cosaque and Maume-Deschamps, 2022]. Distributional RL may also be modified to learn quantiles of the next state value and have shown benefits in practice [Dabney et al., 2018b,a] and in theory [Wang et al., 2023b, 2024c, 2025, Ayoub et al., 2024].

Assumption 4.2 (QR Oracle)**.** For any $v : \mathcal{S} \mapsto [0, (1-\gamma)^{-1}]$, let the true $\tau(s,a)$-quantile of $v(s'), s' \sim P(s,a)$ be denoted by $\beta_\tau^v(s,a)$. Given a dataset $\mathcal{D}_{\mathsf{QR}}$, we assume QR outputs estimates $\widehat{\beta}_v$ with bounded ℓ_∞ error: for any δ, w.p. $1-\delta$, $\|\widehat{\beta}_q - \beta_\tau^q\|_\infty < \mathrm{err}_{\mathsf{QR}}(|\mathcal{D}_{\mathsf{QR}}|, \delta)$.

The second assumption is completeness under the robust Bellman $\mathcal{T}_{\mathsf{rob}}^-$. Completeness is a standard assumption in algorithms based on temporal-difference learning and without it, fitted-Q can diverge or converge to suboptimal fixed points [Tsitsiklis and Van Roy, 1996, Kolter, 2011].

Assumption 4.3 (Completeness)**.** For all $q \in \mathcal{Q}$, we have $\mathcal{T}_{\mathsf{rob}}^- q \in \mathcal{Q}$.

We note that the current proofs of Panaganti et al. [2022], Bruns-Smith and Zhou [2023] require a stronger completeness: $\mathcal{T}_\beta q \in \mathcal{Q}$ for all $q \in \mathcal{Q}$ and feasible β. We circumvent the need for the stronger "all-β" completeness by bounding model misspecification of least squares regression with second order error in the quantile regression.

Finally, we express our bounds with the critical radius $\varepsilon_n^{\mathcal{Q}}$, a standard tool for deriving fast rates in statistics; see Appendix C.4.2 for a summary. Also, we

denote the standard concentrability coefficient with $C_{d_1}^- := \left\| \mathrm{d}d_\mu^{-,\infty} / \mathrm{d}d_1 \right\|_\infty$, a standard and necessary quantity for OPE.

Theorem 4.4. *Let $\varepsilon_n^{\mathcal{Q}}$ denote the critical radius of $\mathcal{Q}$. Under Assumptions 4.2 and 4.3,* RobustFQE *ensures that for any $\delta \in (0,1)$, w.p. $1-\delta$,*

$$\|\widehat{q_M^-} - Q^-\|_{d_1} \lesssim (1-\gamma)^{-2}\left(\sqrt{C_{d_1}^-} \cdot \varepsilon_n^{\mathcal{Q}} + \mathrm{err}_{\mathsf{QR}}^2(n/2M, \delta/2M)\right), \quad \textit{and}$$
$$\left|(1-\gamma)\mathbb{E}_{d_1}[\widehat{v_M^-}(s_1)] - V_{d_1}^-\right| \lesssim \gamma^M + (1-\gamma)^{-1}\left(\sqrt{C_{d_1}^-} \cdot \varepsilon_n^{\mathcal{Q}} + \mathrm{err}_{\mathsf{QR}}^2(n/2M, \delta/2M)\right).$$

For parametric classes (*e.g.*, finite or linear), the critical radius converges at the standard $\widetilde{O}(n^{-1/2})$ rate. Due to orthogonal estimation of CVaR, we benefit from a favorable second-order dependence on $\mathrm{err}_{\mathsf{QR}}$, which allows quantile regression to converge at slower rates without compromising the overall rate. However, if $\mathcal{Q}$ is nonparametric with metric entropy of order $1/t^2$, then $\varepsilon_n^{\mathcal{Q}}$ converges at rate $\widetilde{O}(n^{-1/4})$ [Van der Vaart, 2000], which is slow enough to degrade overall convergence to sub-$\sqrt{n}$. In Section 4.6, we present an orthogonal estimator that tolerates these slower rates while achieving semiparametric efficiency.

4.5 Robust w-Function Estimation with Minimax Learning

Before we present our orthogonal estimator, we study another essential nuisance function: the robust visitation density ratio, *i.e.*, the robust w-function [Kallus and Uehara, 2022, Amortila et al., 2024]. In this section, we first identify the worst-case transition kernel U^- in our uncertainty set $\mathcal{U}(P)$. Then, we propose a minimax estimator [Uehara et al., 2021] for the robust w-function, an important nuisance function for our orthogonal estimator in Section 4.6.

Identification of U^- The robust transition kernel U^- is defined as the feasible perturbed kernel that achieves the inf in the robust Bellman equation $Q^-(s,a) =$

$\mathcal{T}^-_{\mathsf{rob}} Q^-(s,a)$. Let $F^-(y \mid s,a) = P(V^-(s') \leq y \mid s,a)$ be the next-state pushforward measure of the robust value function V^-. Then, U^- is a convex combination of the nominal kernel P and a reweighting of P by an indicator function.

Lemma 4.5. *Suppose $F^-(\beta^-_\tau(s,a) \mid s,a) = \tau$, where $\beta^-_\tau(s,a)$ is the lower τ-th quantile of $F^-(\cdot \mid s,a)$. Then,*

$$U^-(s' \mid s,a)/P(s' \mid s,a) = \Lambda^{-1}(s,a) + (1-\Lambda^{-1})\tau(s,a)^{-1}\mathbb{I}[(V^-(s') - \beta^-_\tau(s,a)) \leq 0]. \quad (4.4)$$

The proof strategy decomposes U^- into its nominal and perturbed components, leveraging the primal solution of CVaR_τ [Ang et al., 2018]; we formalize this in Appendix C.5.

Identification of w^- Using the identification of U^- in Lemma 4.5, we can now identify the robust w-function based on the Bellman flow equations in the worst-case MDP. The Bellman flow in the robust MDP is given by $d^{-,\infty}(s) = (1-\gamma)d_1(s) + \gamma \mathbb{E}_{\widetilde{s}\sim d^{-,\infty}, \widetilde{a}\sim\pi_{\mathsf{t}}(\widetilde{s})} U^-(s \mid \widetilde{s}, \widetilde{a})$. where $d^{-,\infty}(s)$ was defined in Section 4.3. Thus, the robust visitation density, defined as $w^-(s) := {\mathrm{d}d^{-,\infty}(s)}/{\mathrm{d}\nu(s)}$, satisfies the following moment condition for all $f : \mathcal{S} \mapsto \mathbb{R}$:

$$\mathbb{E}[w^-(s)f(s)] = (1-\gamma)\mathbb{E}_{d_1}[f(s_1)] + \gamma \mathbb{E}[w^-(s,a)\mathbb{E}_{s'\sim U^-(s,a)}[f(s')]], \quad (4.5)$$

where we relaxed notation and defined $w^-(s,a) := w(s) \cdot \pi_{\mathsf{t}}(a \mid s)/\nu(a \mid s)$. As before, when there is no unobserved confounding ($\Lambda = 1$), this result recovers the classic Bellman flow.

4.5.1 Estimating w^- with Robust Minimax Indirect Learning

We now propose a penalized minimax estimator for w^- that generalizes the Minimax Indirect Learning (MIL) of Uehara et al. [2021] to our robust MDP setting. Our estimator, RobustMIL (Algorithm 4.2), leverages a general function

Algorithm 4.2 RobustMIL: Minimax estimation of $w^{\pm}$ with a stabilizer

1: **Input:** Dataset $\mathcal{D}$, prior stage estimate $\widetilde{\zeta}$, function classes $\mathcal{W}, \mathcal{F}$, stabilizer weight $\lambda > 0$.
2: **Output:**

$$\widehat{w^{-}} = \operatorname*{argmin}_{w \in \mathcal{W}} \max_{f \in \mathcal{F}} \mathbb{E}_n[w(s,a)(\gamma \xi^{-}(s,a,s') f(s', \pi_t) - f(s,a)) + (1-\gamma)\mathbb{E}_{d_1} f(s_1, \pi_t)] - \lambda \left\| \gamma \xi^{-}(s,a,s'; \widetilde{\zeta}) f(s', \pi_t) - f(s,a) \right\|_{2,n}^2. \tag{4.6}$$

class $\mathcal{W} \subset \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}_+$ to approximately solve the moment equation in Eq. (4.5). It does so by minimizing the difference between the left- and right-hand sides of the equation across a sufficiently large set of adversaries f in a discriminator class $\mathcal{F} \subset \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$. Since U^{-} is unknown, we approximate it via Eq. (4.4) by plugging in a threshold $\widetilde{\zeta}(s,a,s')$ in the indicator function to approximate the true threshold $\zeta^{-}(s,a,s') := V^{-}(s') - \beta^{-}_{\tau(s,a)}(s,a)$. This yields the minimax objective in Eq. (4.6), where we also allow for an optional regularization of the adversary's norm which can be useful for obtaining fast convergence rates.

We make the following assumptions for MIL [Uehara et al., 2021]. The first is a regularity condition that (i) our function class has bounded outputs and (ii) ζ is continuously distributed around the threshold.

Assumption 4.6 (Regularity)**.** (i) $\sup_{w \in \mathcal{W} \cup \{w^{-}\}} \|w\|_\infty < \infty$; (ii) the marginal CDF of $V^{-}(s') - \beta^{-}(s,a)$, *i.e.*, $F(y) = P(V^{-}(s') - \beta^{-}_{\tau(s,a)}(s,a) \leq y)$, is boundedly differentiable around 0.

If next-value distribution is discrete, we can use the discrete form of CVaR and (ii) can be removed. The second assumption is that the adversary class is rich enough to capture all projected errors under the adjoint of the operator $\mathcal{J}_{U^{-}} f(s,a) := \gamma \mathbb{E}_{U^{-}}[f(s', \pi_t) \mid s, a] - f(s,a)$.

Assumption 4.7 (w^{-}-realizability and completeness)**.** $w^{-} \in \mathcal{W}$ and $\mathcal{J}'_{U^{-}}(\mathcal{W} -$

$w^-) \subset \mathcal{F}$.

We note that Assumption 4.7 is monotone in the function class size and can be satisfied by making the function class more expressive, *e.g.*, increasing size of the neural net. Our algorithms are also robust to violations in Assumption 4.7, which we show in Appendix C.7.

We are now ready to state the main estimation result for w^- in terms of the critical radius (Appendix C.4.2) of the function class.

Theorem 4.8. *Let $\varepsilon_n^{\mathcal{W}}$ denote the maximum critical radii of the following classes:*

$$\mathcal{G}_1 = \{(s,a,s') \mapsto (f(s,a) - \gamma f(s', \pi_t)), f \in \mathcal{F}\},$$
$$\mathcal{G}_2 = \{(s,a,s') \mapsto (w(s,a) - w^-(s,a))(\gamma f(s', \pi_t) - f(s,a)), f \in \mathcal{F}, w \in \mathcal{W}\}.$$

Under Assumptions 4.6 and 4.7, RobustMIL *ensures that for any δ, w.p. $1-\delta$,*

$$\left\|\mathcal{J}'_{U^-}(\widehat{w} - w^-)\right\|_2 \lesssim \varepsilon_n^{\mathcal{W}} + \|\widetilde{\zeta}^- - \zeta^-\|_\infty + \sqrt{\log(1/\delta)/n}.$$

As before, the critical radius $\varepsilon_n^{\mathcal{W}}$ converges at an $\widetilde{O}(n^{-1/2})$ rate for parametric classes. Notably, our bounds degrade linearly w.r.t. the ℓ_∞ error in $\widetilde{\zeta}^-$ for estimating ζ^-. For example, if $\widetilde{\zeta}(s,a,s') = \widehat{v}(s') - \widehat{\beta}(s,a)$ where $\widehat{v}, \widehat{\beta}$ are estimated with RobustFQE, then the ζ-error can be bounded by $O(\|\widehat{v} - v^-\|_\infty + \|\widehat{\beta} - \beta^-\|_\infty)$. We present the full proof in Appendix C.7, where we also present a more general result that is robust to misspecifications to realizability and completeness (Assumption 4.7).

4.6 Orthogonal and Efficient Estimator for Robust Policy Value

In this section, we propose an orthogonal estimator that is robust against errors in the nuisances (exhibiting only second-order sensitivity), achieves semiparametric efficiency, and enables inference. Our estimator is based on the efficient

influence function (EIF) of $V^-_{d_1}$, which is the canonical gradient of a statistical estimand [Tsiatis, 2006]. The adoption of EIFs for developing efficient estimators is a broadly employed technique in causal inference [Chernozhukov et al., 2018a, Kennedy, 2023b] and reinforcement learning [Jiang and Li, 2016, Kallus and Uehara, 2022].

We define the collection of nuisance parameters by $\eta^- = (w^-, q^-, \beta^-)$. The notation $\widehat{\eta}$ indicates that these functions are estimated from data, while the notation η denotes their true values.

Theorem 4.9 ((Recentered) Efficient Influence Function)**.** *The (R)EIF of $V^-_{d_1}$ is given by:*

$$\psi(s,a,s';\eta^-) = V^-_{d_1} + w^-(s,a)(r(s,a) + \gamma\rho^-(s,a,s';v^-,\beta^-) - q^-(s,a)), \quad \textit{where}$$

$$\rho^-(s,a,s';v^-,\beta^-) = \Lambda(s,a)^{-1}v^-(s') + (1-\Lambda(s,a)^{-1})(\beta^-(s,a) + \tau^{-1}(v^-(s') - \beta^-(s,a))_-).$$

Remark 4.10. When $\Lambda = 1$, there is no shift in the target environment, and the CVaR weight is zero. The (R)EIF then reduces to the (R)EIF in Kallus and Uehara [2022] for regular OPE with an infinite horizon. As $\Lambda \to \infty$, the CVaR term becomes predominant, with the quantile $\beta^-(s,a)$ taking extreme values. This yields the (novel) (R)EIF for the problem in Du et al. [2022], where the expected value term is replaced solely by a CVaR component in the Bellman equation.

The (R)EIF forms the basis of our orthogonal estimator. First, we note that $\mathbb{E}[\psi(s,a,s';\eta^-)]$ is an unbiased estimator of $V^-_{d_1}$. Furthermore, the expression for $\psi(s,a,s';\eta^-)$ depends only on quantities w^-, q^-, β^- which can be estimated from data. Thus, we can cast the expression $\mathbb{E}[\psi(s,a,s';\eta^-)]$ as a statistical estimand to be learned from the observed sample. This suggests a natural two-stage estimator that we summarize in Algorithm 4.3. In the first stage, we estimate the nuisances $\widehat{\eta}$ from data with K-fold cross-fitting; in the second stage, these estimates

Algorithm 4.3 Orthogonal Estimator for $V_{d_1}^-$

Input: Dataset $\mathcal{D}$, number of splits K.
for $k = 1, 2, \ldots, K$ **do**
    Use data $\mathcal{D} \setminus \mathcal{D}_k$ to learn $(q^{-,[k]}, \beta^{-,[k]})$ with Algorithm 4.1 and $w^{-,[k]}$ with Algorithm 4.2.
    for $i = \lfloor (k-1)n/K \rfloor, \ldots, \lfloor kn/K \rfloor - 1$ **do**
        $\psi_i^- = \psi(s_i, a_i, s_i', \widehat{\eta}^-)$.
    end for
end for
Output: $\widehat{V}_{d_1}^- = \frac{1}{n} \sum_{i=1}^n \psi_i^-$.

are incorporated into the (R)EIF expression and we calculate the empirical average using the observed data. We summarize our procedure in Algorithm 4.3.

The nuisance estimation is detailed in Sections 4.4.2 and 4.5.1. The reliance on the EIF confers our estimator desirable statistical properties including a second order bias due to the nuisances, meaning the bias has a product structure with respect to the nuisance errors. Thus, this special structure orthogonalizes away the dependency on $\widehat{Q}^-$ errors which now only appear in second order. Furthermore, our estimator is semiparametrically efficient in the sense that under mild consistency assumptions, it achieves minimum variance among all regular and asymptotically linear (RAL) estimators. We provide theoretical justifications for these properties in the next section.

4.6.1 Theoretical Guarantees of the Orthogonal Estimator

We now characterize the theoretical properties of our orthogonal estimator. We consider the K-fold cross-fitted estimator in Algorithm 4.3 given by

$$\widehat{V}_{d_1}^- = \frac{1}{n} \sum_{k=1}^{K} \sum_{(s,a,s') \in \mathcal{D}^k} \psi(s, a, s'; \widehat{\eta}^{[k]}),$$

where nuisances $\widehat{\eta}^{[k]}, k \in [K]$ are trained on all data excluding the k^{th} fold $\mathcal{D}^k$. The following theorem outlines the theoretical guarantees of this estimator:

Theorem 4.11 (Efficiency of $\widehat{V}_{d_1}^{-}$)**.** *Let $r_{n,p}^{w}, r_{n,p}^{q}, r_{n,p}^{\beta}$ be functions of the same size $n = |\mathcal{D}|$ such that $\|\mathcal{J}'_{U^-}(\widehat{w}^{-,[k]} - w)\|_p \leq r_{n,p}^{w}$, $\|\widehat{q}^{-,[k]} - q\|_p \leq r_{n,p}^{q}$, and $\|\beta^{-,[k]} - \beta\|_p \leq r_{n,p}^{\beta}$ for any $k \in [K]$. Furthermore, assume that the regularity conditions in Assumption 4.6 hold. Then:*

$$|\widehat{V}_{d_1}^{-} - V_{d_1}^{-}| \lesssim O_p(n^{-1/2}) + O_p(r_{n,2}^{w} r_{n,2}^{q} + (r_{n,\infty}^{q})^2 + (r_{n,\infty}^{\beta})^2) \qquad \text{(Rates)}$$

Furthermore, if $r_{n,2}^{w} \vee r_{n,2}^{q} = o_p(1)$, $r_{n,2}^{w} r_{n,2}^{q} = o_p(n^{-1/2})$, $r_{n,\infty}^{q} = o_p(n^{-1/4})$, and $r_{n,\infty}^{\beta} = o_p(n^{-1/4})$, then $\widehat{V}_{d_1}^{-}$ satisfies:

$$\sqrt{n}(\widehat{V}_{d_1}^{-} - V_{d_1}^{-}) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad \Sigma = \mathrm{Var}(\psi(s, a, s'; \eta^{-})). \qquad \text{(Normality \& Efficiency)}$$

Moreover, Σ is the minimum achievable asymptotic variance among RAL estimators in the nonparametric model for (s, a, s') (the efficiency bound).

We provide the intuition along with a detailed proof in Appendix C.8. The first part of Theorem 4.11 implies that as long as we estimate the nuisances at rates faster that $n^{-1/4}$, then we can learn $\widehat{V}_{d_1}^{-}$ at parametric rates. The second part of Theorem 4.11 states that under mild consistency assumptions, our estimator attains the efficiency bound and is asymptotically normal. That means, for example, we can construct asymptotically valid lower 95%-confidence bound on $\widehat{V}_{d_1}^{-}$ by simply subtracting 1.64 times $\widehat{\mathrm{se}} = \frac{1}{n}(\sum_{k=1}^{K} \sum_{(s,a,s') \in \mathcal{D}^k} (\psi(s, a, s'; \widehat{\eta}^{[k]}) - \widehat{V}_{d_1}^{-})^2)^{1/2}$. Then, we can be sure to have a bound on the worst-case RL policy value, accounting *both* for potential environment shift and finite data. Finally, in Appendix C.10, we describe two settings when our orthogonal estimator remains valid even if some nuisances are *inconsistent*, which is a desirable guarantee for sensitivity analysis [Dorn and Guo, 2023].

Bringing it all together We can instantiate Theorem 4.11 with the nuisance estimators from the previous sections. First, use RobustFQE to estimate $\widehat{q}^{-}$ and $\widehat{\beta}^{-}$,

ensuring $\|\widehat{q}^{-} - Q^{-}\|_2 \leq O(\varepsilon_n^Q + \mathrm{err}_{\mathsf{QR}}^2)$. Under smoothness conditions (Lemma C.3), the L_2 guarantee for $\widehat{q}^{-}$ implies an L_∞ guarantee for $\widehat{q}^{-}$, which also ensures an L_∞ guarantee for $\widehat{\beta}^{-}$. This ensures $\max(\|\widehat{q}^{-} - Q^{-}\|_\infty, \|\widehat{\beta}^{-} - \beta^{-}\|_\infty)$ is well-controlled. Then, we can set $\widetilde{\zeta}^{-}(s,a,s') = \widehat{q}^{-}(s', \pi_{\mathrm{t}}) - \widehat{\beta}^{-}(s,a)$ and run RobustMIL for estimating $\widehat{w}^{-}$. By Theorem 4.8, its projected-L_2 error is $O(\varepsilon_n^{\mathcal{W}} + \|\widehat{q}^{-} - Q^{-}\|_\infty + \|\widehat{\beta}^{-} - \beta^{-}\|_\infty)$. Therefore, the final rate via Theorem 4.11 is $O((\varepsilon_n^Q + \mathrm{err}_{\mathsf{QR}}^2) \cdot \varepsilon_n^{\mathcal{W}} + \|\widehat{q}^{-} - Q^{-}\|_\infty^2 + \|\widehat{\beta}^{-} - \beta^{-}\|_\infty^2)$.

4.7 Empirical Evaluation

We now provide a proof-of-concept empirical investigation to validate our theoretical findings. We experiment with our proposed methodology in a simple synthetic environment. First, we discuss our environment, followed by our approach for solving for the nuisances functions η^{-}. Then, we provide empirical results for our orthogonal estimator, and compare its performance to weighted or direct estimators using the Q^{-} or w^{-} nuisances only. The code for our experiments is open-sourced and available at `https://github.com/CausalML/adversarial-ope/`.

Experimental Setup We consider a synthetic MDP with a one-dimensional state and two actions, modeled after a simple control problem with non-deterministic dynamics. The task is to estimate the worst-case policy value $V_{d_1}^{-}$ of a fixed candidate policy π_{t}, across four different constant values of the sensitivity parameter: $\Lambda(s,a) \in \{1, 2, 4, 8\}$.

We considered three methods for estimating the robust value $V_{d_1}^{-}$:

1. **Q** (RobustFQE): Direct method using the estimated robust quality function $\widehat{Q}^{-}$ only.
2. **W** (RobustMIL): Importance-sampling method using the estimated robust

density ratio $\widehat{w}^-$ only.

3. **Orth**: Our orthogonal estimator which combines the former two, as described in Algorithm 4.3.

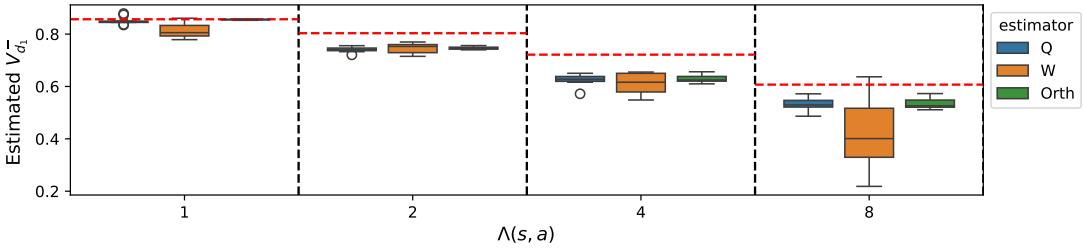
We performed 10 replications of our experimental procedure, where for each replication we: (1) sampled a dataset of 20,000 tuples using a different fixed logging policy π_b; (2) fit the nuisance functions Q^-, β^-, and w^- following the method outlined in Algorithms 4.1 and 4.2 for each Λ; and (3) estimated the corresponding robust policy value $V^-_{d_1}$ for all estimators using the fitted nuisances.

Results We summarize our results in Figure 4.2. We note that all of our estimators are consistently valid for all values of Λ in our experiment. Notably, **Orth** consistently has the lowest mean squared error for the true worst-case policy value. In particular, incorporating the robust importance-sampling weights improves the RobustFQE estimator **Q**, even though these importance-sampling weights by themselves (as in **W**) are much noisier estimators. This is consistent with our theory that the orthogonal estimator is semiparametrically efficient and insensitive to errors in the nuisance functions.

Full experimental details, including our MDP, target and logging policies, methodology for computing the true robust policy values $V^-_{d_1}$, and nuisance estimation, are provided in Appendix C.11. Finally, we also performed an empirical evaluation on the real-world problem of sepsis management using the MIMIC-III dataset [Johnson et al., 2016]. We detail these results in Appendix C.12.

4.8 Conclusion

We consider the problem of infinite-horizon OPE in RL settings when there can be unknown, but bounded, shifts in the transition distribution compared to the



Λ	Mean squared error (MSE) to true worst-case policy value		
	Q	**W**	**Orth**
1	.000240 ± .000170	.002722 ± .002266	**.000005 ± .000006**
2	.004053 ± .001329	.003584 ± .002311	**.003244 ± .000600**
4	.009799 ± .005172	.013862 ± .009228	**.008721 ± .002543**
8	.006247 ± .003980	.052643 ± .050839	**.005713 ± .002730**

Figure 4.2: Synthetic robust off-policy evaluation experiment. We show results for our three estimators on all four Λ values, over our 10 experiment replications. **Above:** Box plot summarizing range of policy value estimates for each combination of estimator and Λ, with Horizontal red dashed lines showing the true worst-case policy values $V_{d_1}^-$. **Below:** Table summarizing the corresponding MSE of these estimators for the true worst-case policy value, along with one standard deviation errors.

transition distribution generating the data. This can arise due to unobserved confounding, where observed transitions do not reflect the true causal ones, non-stationarity in the environment, or adversarial environments. We propose a sensitivity model for such transition kernel shifts analogous to the classic MSM for static decision making, and provide theoretical guarantees for identifying and estimating the sharp (*i.e.*, tightest possible) bounds on the best/worst-case policy value, as well as the corresponding robust Q-function and state density ratio functions. Our estimator for the best/worst-case policy value is orthogonal (insensitive to how the nuisance functions are estimated) and achieves semiparametric efficiency (attaining the best possible asymptotic variance). Finally, our estimator also supports inference, ensuring we can derive reliable bounds for the robust policy value even with finite data.

Part II

Causal Inference from Quasi-Experiments: Weak Instruments and Imperfect Compliance

CHAPTER 5

ESTIMATING HETEROGENEOUS TREATMENT EFFECTS BY COMBINING WEAK INSTRUMENTS AND OBSERVATIONAL DATA

This chapter is based on Oprescu and Kallus [2025].

Accurately predicting conditional average treatment effects (CATEs) is crucial in personalized medicine and digital platform analytics. Since the treatments of interest often cannot be directly randomized, observational data is leveraged to learn CATEs, but this approach can incur significant bias from unobserved confounding. One strategy to overcome these limitations is to leverage instrumental variables (IVs) as latent quasi-experiments, such as randomized intent-to-treat assignments or randomized product recommendations. This approach, on the other hand, can suffer from low compliance, *i.e.*, IV weakness. Some subgroups may even exhibit zero compliance, meaning we cannot instrument for their CATEs at all. In this work, we develop a novel approach to combine IV and observational data to enable reliable CATE estimation in the presence of unobserved confounding in the observational data and low compliance in the IV data, including no compliance for some subgroups. We propose a two-stage framework that first learns *biased* CATEs from the observational data, and then applies a compliance-weighted correction using IV data, effectively leveraging IV strength variability across covariates. We characterize the convergence rates of our method and validate its effectiveness through a simulation study. Additionally, we demonstrate its utility with real data by analyzing the heterogeneous effects of 401(k) plan participation on wealth.

5.1 Introduction

The use of observational data for individual-level causal analyses is becoming increasingly common in personalized medicine, online platforms, and any setting where understanding individualized responses is crucial and/or presents an opportunity for personalization. The key quantity for such analyses is the conditional average treatment effect (CATE), which captures how treatment effects vary according to baseline covariates (features). This measure provides insight into effect heterogeneity and enables personalization.

Using observational data can nonetheless introduce bias from unobserved confounding, where the observed relationship between outcomes and interventions is influenced not only by treatment effects, but also by variables that affect both outcome and treatment, such as socioeconomic status, health, user mood, *etc.*, which are not captured by baseline covariates. These biases can skew causal effect estimates, resulting in unreliable analyses or harmful policy decisions.

Randomized trials are the gold standard for causal inference, but they are often infeasible. For instance, digital services cannot force users to view or buy a product, and clinical trials cannot require invasive treatments. A common alternative is to randomize the *encouragement* of certain actions, such as recommending a product or treatment. These encouragements can serve as instrumental variables (IVs) which, under certain conditions, enable unbiased estimation of treatment effects [Angrist et al., 1996].

Identification of CATEs using IVs hinges on the premise that compliance—the relationship between the treatment received and the intent/encouragement—is nonzero across all baseline covariate values. When compliance is nonzero but small, IV-based estimates tend to have high variance, making them unreliable [Andrews et al., 2019]. In practice, the assumption of strong compliance is

often violated. For example, users on digital platforms may ignore recommendations entirely or reject certain types of content, while participants on mobile health platforms may disregard prompts (*e.g.*taking 250 steps per hour) due to time constraints or lack of interest.

To address the challenge of estimating unbiased CATEs in the presence of unobserved confounding and low IV compliance, we introduce a two-stage framework. In the first stage, we estimate a biased, confounded CATE from observational data. Then, in the second stage, we utilize an IV to learn the confounding bias by weighting the samples according to their compliance levels. By assuming only that the bias can be extrapolated, this approach extends treatment effect adjustments even to groups minimally influenced by the IV, employing a transfer learning approach that leverages varying instrument strengths across covariate groups.

This framework mirrors strategies in causal inference that combine randomized trials with observational data to address low covariate overlap. Building on this body of work, we introduce two methodologies for extrapolating confounding bias within the observational dataset: a parametric estimation approach, assuming the confounding bias adheres to a parametric form, and a transfer learning strategy that assumes a shared representation between the true and biased CATE. We study the properties of our CATE estimators in finite samples and validate our approaches through comprehensive empirical studies.

5.2 Related Work

We briefly overview related work here; for a more comprehensive discussion, refer to Appendix D.1.

5.2.1 Heterogeneous Treatment Effect Estimation from Observational Data

Recent advances in machine learning have expanded the use of observational data to estimate CATEs using diverse techniques such as random forests [Wager and Athey, 2018b], Bayesian algorithms [Hill, 2011], deep learning [Shi et al., 2019], and meta-learners [Künzel et al., 2019]. However, these methods often unrealistically assume an absence of confounding, limiting their real-world applicability. Efforts to account for unobserved confounding either construct *bounds* on treatment effects [Oprescu et al., 2023, Frauen et al., 2024] or use latent variable models and multiple/sequential treatments to debias CATE estimates [Louizos et al., 2017, Wang and Blei, 2019, Bica et al., 2020a], but they frequently depend on unverifiable assumptions or require accurate proxy data, reducing their practical utility.

5.2.2 Heterogeneous Treatment Effect Estimation Using IVs

Integrating machine learning with instrumental variable (IV) methods enhances CATE estimation flexibility over traditional approaches. Techniques range from advanced two-stage least squares (2SLS) that incorporate complex feature mappings via kernel methods [Singh et al., 2019] and deep learning [Xu et al., 2021] to neural networks for conditional density estimation [Hartford et al., 2017] and moment conditions for IV estimation [Bennett et al., 2019]. Yet, these rely on the consistent relevance of instruments across covariate groups, which is not guaranteed with weak instruments.

5.2.3 Treatment Effect Estimation with Weak Instruments

Traditional IV methods like 2SLS can be unreliable when instruments are weak, leading to biased, high-variance estimates. Recent advancements include novel estimators such as bias-adjusted 2SLS, limited information maximum likeli-

hood, and jackknife IV estimators (see Huang et al. [2021] and references therein). Other techniques attempt to reduce variance by exploiting first-stage heterogeneity (variation in compliance) [Coussens and Spiess, 2021, Abadie et al., 2024]. Some approaches also combine multiple weak instruments into robust composites, useful in settings like genetic studies [Kang et al., 2016]. Our approach extends Coussens and Spiess [2021], Abadie et al. [2024] by leveraging compliance weighting to estimate heterogeneous effects and address weak instruments using additional observational data.

5.2.4 Combining Observational and Randomized Data

Increasing research focuses on integrating observational datasets with randomized control trial (RCT) data to mitigate observational bias. Strategies include imposing structural assumptions, such as strong parametric constraints [Kallus et al., 2018], or assuming a shared structure between biased and unbiased CATE functions [Hatt et al., 2022], as well as optimizing dual estimators from both data types for improved bias correction [Yang and Ding, 2019]. Our work aligns with efforts to debias treatment effects using both observational and experimental data, but also addresses challenges such as low IV compliance, the need to debias the overall effect function rather than individual outcome functions, and the complexity of estimating CATEs from IV data using a ratio estimator.

Where Our Work Lies To the best of our knowledge, no current estimation technique effectively combines an IV study, particularly one with weak instruments or low compliance, with an observational study to derive robust and unbiased CATE estimates. We bridge this gap by introducing two robust and consistent CATE estimation techniques, building upon previous work on combining RCT and observational data [Kallus et al., 2018, Hatt et al., 2022], as

well as work that addresses the complexities associated with weak instruments [Coussens and Spiess, 2021, Abadie et al., 2024].

5.3 Background and Setup

We consider the standard setting of causal inference where the goal is to estimate the conditional average treatment effect of a binary treatment $A \in \{0, 1\}$ on an outcome $Y \in \mathbb{R}$ in the presence of covariates $X \in \mathcal{X} \subseteq \mathbb{R}^m$. Our approach is grounded in Rubin's potential outcomes framework, wherein each unit is associated with two potential outcomes $Y(0), Y(1)$ of which only $Y = Y(A)$ is observed (causal consistency). Our objective is to learn the CATE function, given by:

$$\tau(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x]. \tag{5.1}$$

However, we only have access to n_O i.i.d. samples from an observational dataset $O = (X_i^O, A_i^O, Y_i^O)_{i=1}^{n_O} \sim (X^O, A^O, Y^O)$. Thus, we face the fundamental problem of causal inference: only the outcome under the administered treatment is observed, while the counterfactual remains unobserved. Without further assumptions, there exists the possibility of unobserved confounding, leading to a situation where

$$\tau^O(x) = \mathbb{E}[Y^O \mid A^O = 1, X^O = x] - \mathbb{E}[Y^O \mid A^O = 0, X^O = x] \neq \tau(x), \tag{5.2}$$

which indicates a persistent bias in the observed treatment effects that does not diminish even with an increasing sample size. We denote this bias by $b(x)$:

$$b(x) = \tau(x) - \tau^O(x).$$

Assuming this bias is induced by a set of unobserved confounders $U \subseteq \mathbb{R}^k$, the discrepancy arises because the selection into treatment in the observational

population is influenced by U, which also impacts the outcome Y^O. Our goal is to mitigate this bias by leveraging additional data.

Alongside the observational dataset, we have n_E i.i.d. samples from an experimental, intent-to-treat dataset $E = (X_i^E, Z_i^E, A_i^E, Y_i^E)_{i=1}^{n_E} \sim (X^E, Z^E, A^E, Y^E)$ where Z^E is a binary instrument taking values in $\{0, 1\}$. We let $X^E \in \mathcal{X}$ and assume the $p_{X^E}(x) = p_{X^O}(x)$, where p_X denotes the density of the random variable X. Moreover, we assume that the joint distribution of covariates and unobserved confounders (X, U) is consistent across both datasets. As before, we use $Y^E(A, Z)$ to denote the potential outcome given treatment A and instrument Z. Additionally, let $A^E(Z)$ denote the potential treatment under instrument Z, and define the compliance and defiance indicators C and D by $C := \mathbb{I}[A^E(1) > A^E(0)]$ and $D := \mathbb{I}[A^E(1) < A^E(0)]$, respectively. We assume that this dataset follows standard IV assumptions on the data generating process:

Assumption 5.1 (Standard IV Assumptions)**.** We assume the following properties hold: (*Exclusion*) $Y^E(A^E, Z^E) = Y^E(A^E)$, i.e. the instrument affects the outcome only through the treatment; (*Independence*) $Z \perp\!\!\!\perp U \mid X$ for any unobserved confounder U; and (*Relevance*) there exists a subset $\mathcal{X}' \subseteq \mathcal{X}$ with non-zero measure such that $Z^E \not\perp\!\!\!\perp A^E \mid X^E$ for $X^E \in \mathcal{X}'$.

Assumption 5.2 (Unconfounded Compliance [Wang and Tchetgen Tchetgen, 2018])**.** The individual treatment effect is independent of the compliance status given covariates: $Y^E(1) - Y^E(0) \perp\!\!\!\perp (A^E(1) - A^E(0)) \mid X^E$.

We note that the relevance assumption in Assumption 5.1 is a weaker version of the standard IV assumptions since we allow for arbitrarily weak instruments in some regions of the covariate spaces. With Assumption 5.1 and Assump-

tion 5.2, we can identify the CATE for $x \in \mathcal{X}'$ as:

$$\tau^E(x) = \frac{\mathbb{E}[Y^E \mid Z^E = 1, X^E = x] - \mathbb{E}[Y^E \mid Z^E = 0, X^E = x]}{\mathbb{E}[A^E \mid Z^E = 1, X^E = x] - \mathbb{E}[A^E \mid Z^E = 0, X^E = x]} := \frac{\delta_Y(x)}{\gamma(x)} = \tau(x). \quad (5.3)$$

We provide the proof of Equation 5.3 in Appendix D.2. Here, $\gamma(x)$ denotes heterogeneous compliance, a measure of instrument strength, given by $\gamma(x) = P(C = 1 \mid X^E = x) - P(D = 1 \mid X^E = x)$ under Assumption 5.2. A *strong* instrument ($\gamma(x) \to 1$) indicates high adherence to the recommended treatment, with $\gamma(x) = 1$ signifying perfect compliance, similar to a randomized controlled trial. Conversely, a *weak* instrument ($\gamma(x) \to 0$) suggests little influence on treatment uptake, with $\gamma(x) = 0$ indicating no compliance and a confounded selection into treatment. The relevance assumption in Assumption 5.1 ensures $\gamma(x) \neq 0$ for $x' \in \mathcal{X}'$, validating the estimation procedure in Equation 5.3. However, small $\gamma(x)$ values lead to estimates of $\tau(x)$ with high asymptotic variance. Moreover, we wish to extend the estimation of $\tau(x)$ from $\mathcal{X}'$ to $\mathcal{X}$, the population of interest.

Thus, relying solely on observational data results in biased $\tau(x)$ estimates, while experimental data alone can yield high variance or invalid estimates for $x \in \mathcal{X}$ with low compliance. This work addresses these challenges by strategically combining the strengths of both datasets to provide a robust CATE estimation technique.

Notation We denote the L_2 norm of a function f as $\|f\|_{L_2} := \mathbb{E}_F[f(X)^2]^{1/2}$, and the L_2 Euclidean norm of a vector $\theta \in \mathbb{R}^d$ as $\|\theta\|_2$. The notation $\widehat{f}$ represents the estimated value of a parameter or function, where f is the true value. We omit the distribution subscript when clear from context; *e.g.*, $\mathbb{E}[X^E]$ and $\mathbb{E}[X^O]$ denote expectations over experimental and observational samples, respectively.

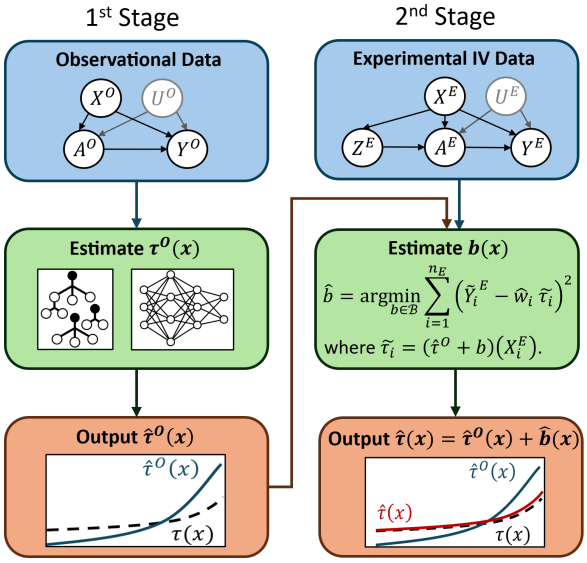


Figure 5.1: Illustration of the two-stage procedure for combining observational and instrumental-variable data. In the first stage, the method learns a biased observational CATE; in the second stage, it uses IV data and estimated compliance to correct that bias.

5.4 Estimation Method

To obtain robust estimates of the CATE function for the population of interest $\mathcal{X}$, we propose a two-step framework that integrates information from both the observational data and the IV study. First, we estimate the confounded CATE function $\widehat{\tau}^O(x)$ using the observational data $(X_i^O, A_i^O, Y_i^O)_{i=1}^{n_O}$. This is a well-established problem in both causal inference and machine learning, and it can be addressed using various existing techniques, including meta-learners ([Künzel et al., 2019]), random forests ([Wager and Athey, 2018b]), and neural networks ([Shi et al., 2019]).

Next, we wish to approximate the bias function $b(x) = \tau(x) - \tau^O(x)$ using the learned $\widehat{\tau}^O(x)$. Without oracle access to the true CATE function $\tau(x)$, we instead rely on samples from the experimental (IV) study $(X_i^E, Z_i^E, A_i^E, Y_i^E)_{i=1}^{n_E}$ for which we can estimate an unbiased, though potentially high variance, CATE for $x \in \mathcal{X}'$, as

given in Equation 5.3. Our approach hinges on the following lemma:

Lemma 5.3. *[CATE Estimation with IVs] Let $\pi_Z(x) := P(Z^E = 1 \mid X^E = x)$ be the instrument propensity. Then, the following identity holds for every $x \in \mathcal{X}'$:*

$$\mathbb{E}\left[\frac{Y^E Z^E}{\pi_Z(x)\gamma(x)} - \frac{Y^E(1-Z^E)}{(1-\pi_Z(x))\gamma(x)}\middle| X^E = x\right] = \tau(x)$$

We note that in the case of randomized instrument assignment, the instrument propensity is known and often given by a constant, *i.e.*, $\pi_Z(x) = \pi_Z > 0$. By defining $V_Z(x) := \pi_Z(x)(1 - \pi_Z(x))$, Lemma 5.3 shows that the bias function $b(x)$ can be expressed in terms of observable quantities as $b(x) = \mathbb{E}\left[\frac{Y^E Z^E(1-\pi_Z(X^E)) - Y^E(1-Z^E)\pi_Z(X^E)}{V_Z(X^E)\gamma(X^E)} - \tau^O(x) \mid X^E = x\right]$ for $x \in \mathcal{X}'$. This formulation suggests a practical estimation strategy. We estimate $\gamma(x)$ and, if necessary, $\pi_Z(x)$ from data, and define the pseudo-outcome

$$\frac{\widetilde{Y}^E}{\widehat{V}_Z(X^E)\widehat{\gamma}(X^E)} := \frac{Y^E Z^E(1-\widehat{\pi}_Z(X^E)) - Y^E(1-Z^E)\widehat{\pi}_Z(X^E)}{\widehat{V}_Z(X^E)\widehat{\gamma}(X^E)}.$$

We then combine this pseudo-outcome with the estimated $\widehat{\tau}^O(x)$ in a subsequent regression step to obtain an unbiased and consistent estimate of $b(x)$ for $x \in \mathcal{X}'$, provided π_Z, γ, and $\widehat{\tau}^O$ are estimated consistently. However, such an estimator only provides estimates for $\mathcal{X}'$ where $\gamma(x) \neq 0$. Additionally, for small values of $\gamma(x)$, $\pi_Z(x)$, and $1 - \pi_Z(x)$, this method may result in high variance in the estimates $\widehat{b}(x)$, especially for certain parametric function classes. To address these challenges, we weight the data samples by the inverse variance of $\widetilde{Y}^E/(\widehat{\gamma}(x)\widehat{V}_Z(x))$ given by $\mathrm{Var}(\widetilde{Y}^E|X^E = x)^{-1}\widehat{\gamma}^2(x)\widehat{V}_Z^2(x)$. This approach is frequently used in generalized least squares methods (GLS, [Agresti, 2015]) to confer the algorithm asymptotic efficiency. While $\mathrm{Var}(\widetilde{Y}^E|X^E = x)$ can be estimated from data using machine learning methods, it is generally preferable to weight the estimator solely by compliance and instrument propensity to avoid issues with small values of $\mathrm{Var}(\widetilde{Y}^E|X^E = x)$. Assuming the bias function belongs to a class of functions

$\mathcal{B}$, our proposed algorithm can be described by the following weighted empirical risk minimization (ERM) procedure.

$$\widehat{b} = \arg\min_{b \in \mathcal{B}} \sum_{i=1}^{n_E} \left(\widetilde{Y}_i^E - \widehat{\gamma}(X_i^E)\widehat{V}_Z(X_i^E)\widehat{\tau}^O(X_i^E) - \widehat{\gamma}(X_i^E)\widehat{V}_Z(X_i^E)b(X_i^E) \right)^2 \tag{5.4}$$

where the factor $\widehat{\gamma}^2(x)\widehat{V}_Z^2(x)$ was used for weighting the squared loss. This estimator automatically extrapolates to all of $\mathcal{X}$ since we assign weights of 0 when $\widehat{\gamma}(x) = 0$. Moreover, this method places higher emphasis on lower-variance pseudo-outcomes, thereby minimizing the risk of overfitting to data points with high variance. This weighting technique is commonly employed in other IV estimation tasks, such as local *average* treatment effect estimation (LATE), where weighting data points by compliance yields estimators with lower variance ([Coussens and Spiess, 2021, Abadie et al., 2024]).

The weighting scheme in Equation 5.4 creates a weighted distribution, $\tilde{p}_{X^E}(x)$, for optimizing the ERM procedure. Since $\tilde{p}_{X^E}(x)$ differs from the target distribution $p_{X^E}(x)$, this introduces a transfer learning problem. Without additional constraints on the function class $\mathcal{B}$, the minimization in Equation 5.4 may yield many possible solutions. To ensure a unique or limited solution set, $\mathcal{B}$ must have low complexity or require further structural assumptions. We explore two function classes $\mathcal{B}$: a parametric class defined by $b(x) = \theta^T\phi(x), \theta \in \mathbb{R}^d$ with a known mapping $\phi : \mathcal{X} \to \mathbb{R}^d$, and a second parametric class where $b(x) = \nu^T\phi(x)$, with $\nu \in \mathbb{R}^d$ and $\phi \in \Phi$ being a learned representation common to both the observational and IV datasets.

5.4.1 Integrating Observational and Experimental Data via Parametric Extrapolation

We consider a parametric class $\mathcal{B}_\phi = \{\theta^T\phi(x) : \theta \in \mathbb{R}^d\}$ for a known mapping $\phi : \mathcal{X} \to \mathbb{R}^d$. Since the compliance factor $\gamma(x)$, instrument propensity $\pi_Z(x)$, and

Algorithm 5.1 CATE Estimation with Parametric Extrapolation

Input: Observational dataset $O = (X_i^O, A_i^O, Y_i^O)_{i=1}^{n_O}$ and IV dataset $E = (X_i^E, Z_i^E, A_i^E, Y_i^E)_{i=1}^{n_E}$; $\tau^O(x)$ estimator $\mathcal{T}$; $\gamma(x)$ estimator $\mathcal{G}$; $\pi_Z(x)$ estimator $\mathcal{P}$; known mapping $\phi : \mathcal{X} \to \mathbb{R}^d$.
Learn $\widehat{\tau}^O(x)$ using $\mathcal{T}$ on O.
Initialize $\widetilde{\mathbf{Y}} \in \mathbb{R}^{n_E}$ and $\widetilde{\mathbf{X}} \in \mathbb{R}^{n_E \times d}$.
for $k = 1, 2, \ldots, K$ **do**
    Set $\mathcal{I}_k = \{i \in \{1, \ldots, n_E\} : i \equiv k - 1 \pmod K\}$.
    Using $\{(X_i^E, Z_i^E, A_i^E, Y_i^E) \in E : i \notin \mathcal{I}_k\}$, learn $\widehat{\gamma}^{(k)}(x)$ with $\mathcal{G}$ and $\widehat{\pi}_Z^{(k)}(x)$ with $\mathcal{P}$.
    for $i \in \mathcal{I}_k$ **do**
        Set $\widetilde{\mathbf{Y}}_i = Y_i^E Z_i^E (1 - \widehat{\pi}_Z^{(k)}(X_i^E)) - Y_i^E (1 - Z_i^E)\widehat{\pi}_Z^{(k)}(X_i^E) - \widehat{w}^{(k)}(X_i^E)\widehat{\tau}^O(X_i^E)$.
        for $j = 1, \ldots, d$ **do**
            Set $\widetilde{\mathbf{X}}_{ij} = \widehat{w}^{(k)}(X_i^E)\phi_j(X_i^E)$.
        end for
    end for
end for
Output: $\widehat{\theta} = (\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^\top \widetilde{\mathbf{Y}}$.

the parameter of interest θ^T are learned from the same dataset E, we propose the following K-fold cross-fitted estimation procedure:

$$\widehat{\theta} = \arg\min_{\theta \in \mathbb{R}^d} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k^E} \left(\widetilde{Y}_i^E - \widehat{w}^{(k)}(X_i^E)\widehat{\tau}^O(X_i^E) - \theta^T \widehat{w}^{(k)}(X_i^E)\phi(X_i^E) \right)^2 \tag{5.5}$$

where $\widehat{w}^{(k)}(X_i^E) := \widehat{\gamma}^{(k)}(X_i^E)\widehat{V}_Z^{(k)}(X_i^E)$, and the compliance factor $\widehat{\gamma}^{(k)}$ and instrument propensity $\widehat{\pi}_Z^{(k)}$, $k \in [K]$ are trained on E excluding the k^{th} fold containing indices $\mathcal{I}_k^E$. K-fold cross-fitting is crucial because it ensures that the weights are learned from data distinct from that used in the ERM algorithm. This separation is essential for maintaining desirable theoretical properties as we remain methodologically agnostic to the techniques used for learning γ and π_Z.

The compliance factor $\gamma(x) = \mathbb{E}[A^E \mid Z^E = 1, X^E = x] - \mathbb{E}[A^E \mid Z^E = 0, X^E = x]$ can be estimated using standard machine learning classification algorithms, either by training separate classifiers for $A^E \mid Z^E = 1, X^E = x$ and $A^E \mid Z^E = 0, X^E = x$ or by using one classifier with Z^E as an additional feature. Similarly, instrument propensity estimation is a straightforward classification task with Z^E as the tar-

get. Given estimates $\widehat{\tau}^O$, $\widehat{\gamma}^{(k)}$, and $\widehat{\pi}_Z^{(k)}$, the result in Equation 5.5 is obtained by performing an OLS procedure with the targets $\widetilde{Y}_i^E - \widehat{w}^{(k)}(X_i^E)\widehat{\tau}^O(X_i^E)$ and the design matrix $\widetilde{\mathbf{X}} = W(X^E)\Phi(X^E)$. Here, $W(X^E) = \text{diag}(\widehat{w}^{(k)}(X_i^E), \ldots, \widehat{w}^{(k)}(X_{n_E}^E))$, and $\Phi(X^E) = (\phi(X_1^E), \ldots, \phi(X_{n_E}^E))^T$. Algorithm 5.1 details the two-step procedure.

Next, we provide theoretical guarantees for our parametric extrapolation approach. We begin by describing the regularity assumptions that enable the consistency of our estimator.

Assumption 5.4 (Regularity Assumptions)**.** The following claims are true:

1. (Treatment Positivity in O) $\epsilon \leq P(A^O = 1 \mid X^O = x) \leq 1 - \epsilon$ for some $\epsilon > 0$.
2. (Instrument Positivity in E) $\epsilon \leq \pi_Z(X^E), \widehat{\pi}_Z(X^E) \leq 1 - \epsilon$ for some $\epsilon > 0$.
3. (Boundedness) Y^E, Y^O, $\|X^E\|_2$, $\|\phi(X^E)\|_2$, $\widehat{\tau}^O(x)$, $\widehat{\gamma}(x)$ are uniformly bounded.
4. (Realizability of $b(x)$) $b(x) \in \mathcal{B}_\phi$, *i.e.*$\tau(x) - \tau^O(x) = \theta^T\phi(x)$ for some $\theta \in \mathbb{R}^d$.
5. (Identifiability of θ) $\mathbb{E}[\phi(X^E)\phi(X^E)^T]$ is invertible.

The first two conditions in Assumption 5.4 are standard in causal inference, ensuring that both treatments (or instruments) and controls are observable for every $x \in \mathcal{X}$, enabling CATE estimation. The third condition imposes a common boundedness assumption to control the growth of estimands. The fourth condition ensures our model for the bias function $b(x)$ is well-specified given $\mathcal{B}_\phi$. The final condition requires that the design matrix has rank d, ensuring we can learn the parameter θ from data. Given Assumption 5.4, we present the following theoretical result:

Theorem 5.5 (Estimator Consistency for Parametric Extrapolation)**.** *Let $r_\gamma(n)$, $r_{\pi_Z}(n)$, and $r_{\tau^O}(n)$ be $o_p(1)$ functions of $n \in \mathbb{N}$ such that $\|\gamma - \widehat{\gamma}^{(k)}\|_{L_2} \leq r_\gamma(n_E)$, $\|\pi_Z - \widehat{\pi}_Z^{(k)}\|_{L_2} \leq r_{\pi_Z}(n_E)$, and $\|\tau^O - \widehat{\tau}^O\|_{L_2} \leq r_{\tau^O}(n_O)$. Furthermore, assume the conditions*

of Assumption 5.1, Assumption 5.2, and Assumption 5.4 hold. Then, the parameter $\widehat{\theta}$ *returned by Algorithm 5.1 is consistent and satisfies*

$$\left\|\widehat{\theta} - \theta\right\|_2 = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O) + 1/\sqrt{n_E}\right).$$

Moreover, $\widehat{\tau}$ *is consistent on* $\mathcal{X}$ *with convergence rate given by*

$$\|\widehat{\tau} - \tau\|_{L_2} = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O) + 1/\sqrt{n_E}\right).$$

We include the proof of Theorem 5.5 in Appendix D.2. The core insight is that weighted OLS remains consistent as long as the estimates for $\widehat{\gamma}$, $\widehat{\pi}_Z$, and $\widehat{\tau}^O$ are themselves consistent. However, the overall convergence rate is constrained by the slowest of these rates. In most cases, π_Z is assumed to be known, meaning the convergence rate is primarily dictated by the rates of $\widehat{\gamma}$ and $\widehat{\tau}^O$. This result highlights the trade-off involved in leveraging both datasets to achieve accurate effect estimation for the target population.

Remark 5.6 (Impact of Realizability Violations). When realizability does not hold, *i.e.*$b(x) \notin \mathcal{B}$, our estimator may be inconsistent, with asymptotic bias proportional to the deviation of the true function from $\mathcal{B}$. Nonetheless, conducting this analysis might still be valuable, as the resulting bias may be smaller than confounding bias in observational estimates or the variance from low compliance in IV studies. Thus, under realizability violations, our method may provide more accurate CATE estimates by effectively balancing bias and variance.

5.4.2 Integrating Observational and Experimental Data via a Common Representation

Without expert knowledge, the mapping $\phi(x)$ may not be known a priori. In this section, we introduce a method to jointly learn both the unbiased CATE function and the mapping $\phi(x)$ (hereafter referred to as the *representation*), based

on the assumption that the true CATE $\tau(x)$ and the biased CATE $\tau^O(x)$ share a common representation. This approach leverages machine learning techniques that assume a common structure across tasks, such as multi-task and transfer learning. In causal inference, it has been suggested that a shared representation can be assumed between treatment arms [Shalit et al., 2017, Shi et al., 2019] or between randomized data and confounded observational data [Hatt et al., 2022]. This framework enables us to learn the bias function $b(x)$ even when the mapping $\phi(x) \in \Phi$ is otherwise unknown.

We consider a class Φ of representations $\phi(x) : \mathcal{X} \to \mathbb{R}^d$ and assume that there exists a shared representation $\phi \in \Phi$ between the true and biased CATEs. Specifically, there exist linear hypotheses $h, h^O \in \mathbb{R}^d$ such that $\tau(x) = h^T\phi(x)$ and $\tau^O(x) = (h^O)^T\phi(x)$, resulting in the bias function $b(x) = (h - h^O)^T\phi(x) := \nu^T\phi(x)$. For simplicity, we focus on linear-in-representation classes, but more complex hypotheses h with $\tau(x) = h(\phi(x))$ can be considered – see [Shalit et al., 2017, Hatt et al., 2022]. Thus, $b(x) \in \mathcal{B}_\phi$ for the unknown ϕ, with $\mathcal{B}_\phi$ defined in Section 5.4.1. Suppose there exists an ERM algorithm $\mathcal{T}$ that can jointly learn $\phi(x)$ and h^O from the observational data, O. Our learning algorithm proceeds as follows: first, we use $\mathcal{T}$ to learn $\widehat{\phi}(x)$ and $\widehat{h}^O$ from O, alongside estimates $\widehat{\gamma}^{(k)}(x)$ and $\widehat{V}_Z^{(k)}(x)$ from E as described in Section 5.4.1. In the second stage, we apply the following ERM procedure to estimate the parameter ν:

$$\widehat{\nu} = \arg\min_{\nu \in \mathbb{R}^d} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k^E} \left(\widetilde{Y}_i^E - (\widehat{h}^O)^T \widehat{w}^{(k)}(X_i^E)\widehat{\phi}(X_i^E) - \nu^T \widehat{w}^{(k)}(X_i^E)\widehat{\phi}(X_i^E) \right)^2 . \tag{5.6}$$

This procedure is detailed in Algorithm 5.2. Finally, we recover $\widehat{\tau}(x)$ by setting $\widehat{\tau}(x) = (\widehat{h}^O + \widehat{\nu})^T\widehat{\phi}(x)$.

Example 5.7 (Representation learning with neural networks)**.** Let Φ be a class of feed-forward neural networks. Then $\widehat{\phi}(x), \widehat{h}^O$ and $\widehat{\tau}^O(x)$ can be jointly learned

Algorithm 5.2 CATE Estimation with Representation Learning

1: **Input:** Observational dataset $O = (X_i^O, A_i^O, Y_i^O)_{i=1}^{n_O}$ and IV dataset $E = (X_i^E, Z_i^E, A_i^E, Y_i^E)_{i=1}^{n_E}$; (ϕ, h^O) estimator $\mathcal{T}$, $\gamma(x)$ estimator $\mathcal{G}$, $\pi_Z(x)$ estimator $\mathcal{P}$.
2: Learn $\widehat{\phi}(x)$ and $\widehat{h^O}$ using $\mathcal{T}$ on O.
3: Call Algorithm 5.1 with $\phi = \widehat{\phi}$ and $\widehat{\tau}^O(x) = (\widehat{h^O})^T\widehat{\phi}(x)$. Let $\widehat{\nu}$ be its output.
4: **Output:** $\widehat{\nu}$.

by composing Φ with two linear output heads for $Y^O \mid A^O = 1, X^O = x$ and $Y^O \mid A^O = 0, X^O = x$, respectively. By taking the difference between the two output heads, we can reconstruct $\widehat{\tau}^O(x)$, assuming that $\mathbb{E}[Y^O \mid A^O = 1, X^O = x]$ and $\mathbb{E}[Y^O \mid A^O = 0, X^O = x]$ are also linear in ϕ (see [Shalit et al., 2017, Shi et al., 2019]). Without this assumption, we can learn $\tau^O(x)$ directly by composing Φ with one linear output layer and considering the pseudo-outcome $\frac{Y^O A^O}{\pi_A(X^O)} - \frac{Y^O(1-A^O)}{(1-\pi_A(X^O))}$. Here, $\pi_A(X^O) = P(A^O = 1 \mid X^O)$ is the treatment propensity in O and can be learned using any black-box machine learning classifier.

With this setup, we obtain theoretical results similar to those in Theorem 5.5:

Theorem 5.8 (Estimator Consistency for Shared Representation Learning)**.** *Let $r_\gamma(n)$, $r_{\pi_Z}(n)$, and $r_\phi(n)$ be $o_p(1)$ functions of $n \in \mathbb{N}$ such that $\|\gamma - \widehat{\gamma}^{(k)}\|_{L_2} \leq r_\gamma(n_E)$, $\|\pi_Z - \widehat{\pi}_Z^{(k)}\|_{L_2} \leq r_{\pi_Z}(n_E)$, and $\left\|\phi - \widehat{\phi}\right\|_{L_2} \leq r_\phi(n_O)$. Additionally, assume $\left\|\widehat{\phi}\right\|_2$ is bounded and $\mathbb{E}[\widehat{\phi}(X)\widehat{\phi}(X)^T]$ is invertible. Let us also consider the conditions specified in Assumption 5.1 and Assumption 5.2 to be satisfied. Moreover, assume that $\tau^O(x) = (h^O)^T\phi(x)$ for some ϕ that is realizable within the representation class Φ and let Assumption 5.4 hold for ϕ. Under these conditions, the parameter $\widehat{\nu}$ returned by Algorithm 5.2 is consistent and satisfies*

$$\|\widehat{\nu} - \nu\|_2 = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_\phi(n_O) + 1/\sqrt{n_E} + 1/\sqrt{n_O}\right).$$

Moreover, $\widehat{\tau}$ is consistent on $\mathcal{X}$ with convergence rate given by

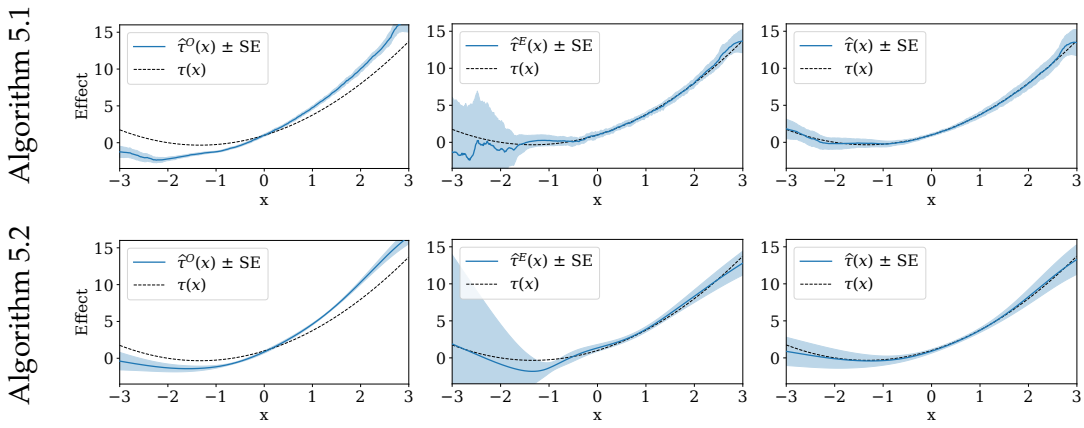
$$\|\widehat{\tau} - \tau\|_{L_2} = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_\phi(n_O) + 1/\sqrt{n_E} + 1/\sqrt{n_O}\right).$$

We provide the proof of Theorem 5.8 in Appendix D.2. This result hinges on the realizability assumption in Φ and the linear-in-representation structure of both τ and τ^O. In Example 5.7, $r_\phi(n)$ bounds the generalization error for feed-forward neural networks. For ReLU activations and bounded outputs, $r_\phi(n) = C\sqrt{WL\log W\log n}/\sqrt{n}$, where W is the total number of weights, L is the number of layers, and C is a constant independent of n and W [Yarotsky, 2017, Farrell et al., 2021]. While this rate is parametric, it scales linearly with W, which becomes problematic for over-parameterized networks. For 1-Lipschitz activations and bounded weights, Golowich et al. [2018] derive a rate of $r_\phi(n) = C\sqrt{\Pi_{l=1}^{L} M(l)}/n^{1/4}$, where $M(l)$ bounds the Frobenius norm of layer l's weight matrix.

Practical Guidance in High Dimensions When $\phi(x)$ is high-dimensional, controlling the complexity of $\mathcal{B}_\phi$ through regularization is crucial, especially since the bias function $b(x)$ is used to extrapolate the CATE into low-variance regions where compliance is low and the risk of overfitting is high. In the parametric extrapolation approach (Section 5.4.1), applying L_1 or L_2 regularization via Lasso or Ridge regression in the final step is effective for controlling model complexity. In the shared representation approach (Example 5.7), regularization not only helps control the parameters h^O and ν but also prevents over-parametrization in the neural network ϕ. The choice between L_1 and L_2 regularization, and how they are applied, should be aligned with the data-generating process and the specific characteristics of the model.

5.5 Experimental Results

We evaluate our method on both simulated and real-world data. We begin with the confounded synthetic example of Kallus et al. [2018], and construct a corre-



(a) $\widehat{\tau}^O(x)$ (b) $\widehat{\tau}^E(x)$ (c) $\widehat{\tau}(x)$

Figure 5.2: Means and standard errors of CATE estimates from 100 simulated observational/IV dataset pairs (O, E) using Random Forest learners (top row) or Neural Network learners (bottom row). (5.2a): Biased observational CATE $\tau^O(x)$. (5.2b): High-variance IV-based CATE estimate from Equation 5.3. (5.2c): Final debiased CATE estimate from Algorithm 5.2 using parametric extrapolation (top row) or representation learning (bottom row).

sponding IV study using a similar data-generating process that preserves the same confounding structure and treatment effects. We use this setup to assess Algorithms 5.1 and 5.2 for unbiased CATE estimation by integrating the two datasets. We then apply our estimators to a real-world dataset on the effect of 401(k) participation on financial wealth. Additional experiments, along with implementation details, hyperparameter selection, and validation procedures, are provided in Appendix D.3. Replication code is available at `https://github.com/CausalML/Weak-Instruments-Obs-Data-CATE`.

5.5.1 Simulation Studies

We generate the observational dataset $O = (X^O, A^O, Y^O)$ as follows (see [Kallus et al., 2018]):

$$X \sim \mathcal{N}(0,1), \quad A \sim \text{Bern}(0.5), \quad U \mid X, A \sim \mathcal{N}\left(X\left(A - 0.5\right), 0.75\right)$$
$$Y = 1 + A + X + 2AX + 0.5X^2 + 0.75AX^2 + U + 0.5\epsilon_Y, \quad \epsilon_Y \sim \mathcal{N}(0,1) \tag{5.7}$$

In this DGP [1], the true CATE is given by $\tau(x) = 0.75x^2 + 2x + 1$, whereas the biased observational CATE is represented by $\tau^O(x) = 0.75x^2 + 3x + 1$. This results in a bias $b(x) = -x$, which is linear in x. We modify this DGP to generate the experimental IV dataset $E = (X^E, Z^E, A^E, Y^E)$ as follows:

$$X \sim \mathcal{N}(0,1), \quad Z \sim \text{Bern}(0.5), \quad A^* \sim \text{Bern}(0.5)$$

$$\gamma(X) = \sigma(2X), \quad C \sim \text{Bern}(\gamma(X)), \quad A = C \cdot Z + (1 - C) \cdot A^*$$

$$U \mid X, A, C \sim C \cdot \mathcal{N}(0,1) + (1 - C) \cdot \mathcal{N}(X\,(A - 0.5)\,, 0.75)$$

where C is the (unknown) compliance indicator, σ is the logistic sigmoid and we keep the same outcome function as in Equation 5.7. In this modified DGP, the randomized instrument has compliance sharply determined by X, with low X values indicating almost no compliance and high X values indicating near-perfect compliance.

We generate 100 observational and IV datasets, each with 5,000 samples, from the proposed DGP. We first apply Algorithm 5.1 to each dataset. With a randomized instrument, $\pi_Z(x) = 0.5$. We estimate $\gamma(x)$ as the difference between Random Forest (RF) classifiers, one trained on the subset with $Z^E = 0$ and the other on the subset with $Z^E = 1$, using X^E as features and A^E as target. The biased observational CATE is modeled using the T-learner approach [Künzel et al., 2019], with RF regressors trained on $X^O, Y^O \mid A^O = 0$ and $X^O, Y^O \mid A^O = 1$. For comparison, we implement a CATE estimator for the experimental data using Equation 5.3. We compute $\delta_Y(x)$ as the difference between RF regressors trained on $X^E, Y^E \mid Z^E = 0$ and $X^E, Y^E \mid Z^E = 1$, then divide by $\widehat{\gamma}(x)$, clipping the compliance score at 0.1. We calculate $\widehat{\gamma}(x)$, $\widehat{\tau}^O(x)$, and $\widehat{\tau}^E(x)$ for each dataset pair and proceed with the second step of Algorithm 5.1 by setting $\phi(x) = x$.

[1] For experimental results using a higher-dimensional version of this DGP, refer to Appendix D.3.

In Figure 5.2 (top row), we depict the means and standard errors of our estimators across 100 simulations. The first two plots illustrate the learned observational CATE $\widehat{\tau}^O(x)$ and the learned IV CATE $\widehat{\tau}^E(x)$. As expected, $\widehat{\tau}^O(x)$ shows clear bias, while $\widehat{\tau}^E(x)$ has high variance despite aggressive compliance score clipping. The third plot presents the results from Algorithm 5.1, showing that the resulting $\widehat{\tau}(x)$ is both unbiased and has low variance across $\mathcal{X}$. These findings demonstrate that our two-stage estimation procedure effectively leverages the strengths of both datasets to capture the true CATE and address the limitations of each individual study design.

We note that in our DGP, $\tau(x)$, $\tau^O(x)$, and $b(x)$ are linear in the polynomial representation (x, x^2). Thus, we next apply Algorithm 5.2 with Example 5.7 to learn the true CATE and the common representation from the generated dataset. For consistency, we employ feed-forward neural networks (NNs) to estimate all quantities. The estimator for $\widehat{\gamma}$ uses a NN with a sigmoid activation in the output layer, trained on X^E with the pseudo-outcome $2A^EZ^E - 2A^E(1 - Z^E)$. The representation $\phi(x)$ and the biased CATE $\tau^O(x)$ are learned using a representation network with two output heads for learning $Y^O \mid X^O, A^O = 0$ and $Y^O \mid X^O, A^O = 1$. A similar dual-head approach is used to learn $\delta_Y(x)$, by modeling $Y^E \mid X^E, Z^E = 0$ and $Y^E \mid X^E, Z^E = 1$ simultaneously. When calculating $\delta_Y(x)/\gamma(x)$, we clip the compliance score at 0.1. Unlike Algorithm 5.1, we don't guarantee the polynomial representation will be fully captured by the chosen representation class, but we expect a sufficiently flexible Φ to adequately represent these relationships.

The means and standard errors of our estimators from 100 simulations using neural networks and Algorithm 5.2 are shown in Figure 5.2 (bottom row). As before, $\widehat{\tau}^O(x)$ shows bias, while $\widehat{\tau}^E(x)$ has high variance in low-compliance regions, despite compliance score clipping. However, Figure 5.2c shows that

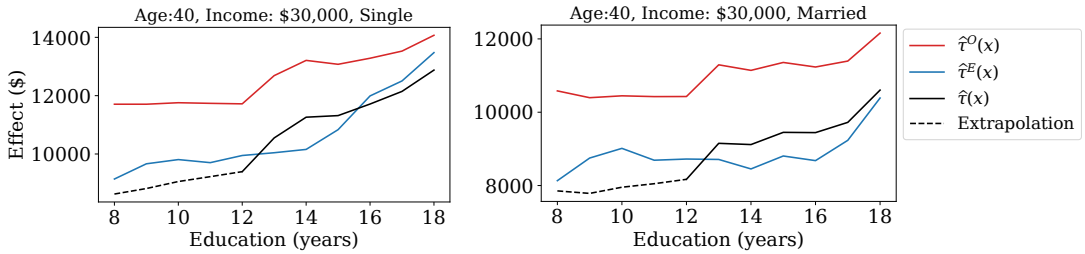


Figure 5.3: Estimated effect of 401(k) participation on net worth by education level in the 401(k) application. Age, income, and binary covariates are held fixed while education and marital status vary. The black line shows the final debiased CATE estimate from Algorithm 5.1, the dashed segment indicates extrapolation into the no-compliance region, the blue line shows the biased observational CATE, and the orange line shows the IV-based CATE estimate without artificial noncompliance.

the $\widehat{\tau}$ returned by Algorithm 5.2 remains unbiased with relatively low variance across $\mathcal{X}$. This demonstrates that combining observational and IV data, where the biased and true CATE share a representation, allows us to reliably learn both the representation and the unbiased CATE, overcoming the limitations of each individual study.

5.5.2 Impact of 401(k) Participation on Financial Wealth

We demonstrate our method's effectiveness with a real-world case study on the impact of 401(k) participation on financial wealth, using data from the 1991 Survey of Income and Program Participation [Chernozhukov and Hansen, 2004]. The dataset includes 9,915 respondents with nine covariates: age, income, education, family size, marital status, two-earner status, pension status, IRA participation, and home ownership. The primary variable of interest is 401(k) participation (A), with eligibility (Z) as the instrumental variable. Although 401(k) eligibility is not randomly assigned, it is argued to maintain conditional independence given observed features [Poterba et al., 1994, Chernozhukov and Hansen, 2004]. We assume 401(k) eligibility influences net worth only through 401(k)

participation, characterizing this as an IV study with one-sided non-compliance, where non-eligible individuals cannot participate ($A^E(0) = 0$). The target variable (Y) is net financial assets, calculated as the total of 401(k) balance, bank account balances, and interest-earning assets, minus non-mortgage debt.

To replicate the scenario in this work, we split the dataset into two halves: one for the IV study and the other for the observational study (where we intentionally remove the instrument information). Our goal is to use these datasets, along with the parametric extrapolation approach in Algorithm 5.1, to recover the unbiased CATEs. Due to one-sided non-compliance, the estimated compliance factor $\widehat{\gamma}(x)$ is high ($0.49 - 0.90$, see Appendix D.3). To show the utility of our method, we introduce artificial non-compliance by setting $\gamma(x)$ to 0 for individuals with less than 12 years of education (13% of the population). In the first stage of Algorithm 5.1, we use RF regressors and classifiers to estimate $\tau^O(x)$, $\gamma(x)$, and $\pi_Z(x)$, with hyperparameters set based on other related work on this dataset [Chernozhukov et al., 2018a]. In the second stage, we define the mapping $\phi(x)$ with an intercept term, the 9 covariates, and their interactions (46 features total). We apply a mild L_1 regularization in the final linear regression due to the large number of resulting features.

In Figure 5.3, we study how the CATE function from Algorithm 5.1 varies with education. We focus on education as it is selected as a top feature by the compliance model, while the outcome models do not rank it as highly significant (see Appendix D.3). To explore this relationship, we vary education and marital status, holding age and income at their median values and setting all binary variables to zero. Since compliance in the IV study is high, we consider the estimate $\widehat{\tau}^E(x)$ *without* the artificial non-compliance as a reference estimate for comparison. Our analysis shows that observational data treatment effects are

upwardly biased, likely due to unobserved confounders such as financial literacy. The $\widehat{\tau}(x)$ from Algorithm 5.1, shown with a dashed line for non-compliance regions, closely aligns with $\widehat{\tau}^E(x)$ (excluding the artificial non-compliance). This demonstrates that combining IV and observational data can effectively estimate unbiased CATEs in real-world settings, offering a robust solution for causal inference even in the presence of low compliance and unobserved confounding.

5.6 Conclusion

This study introduces a method that combines observational and instrumental variable (IV) data to address unobserved confounders in observational studies and low compliance in IV studies. Our two-stage framework first estimates biased CATEs from the observational data, then corrects them using compliance-weighted IV samples. We explore two variations of our procedure: one that models confounding bias parametrically, and another that leverages a shared representation between the true and biased CATEs. Both methods are shown to be consistent, validated through simulations and real-world applications. Our approach holds significant promise for applications in digital platforms, personalized medicine, and economics, offering a robust tool for deriving reliable, actionable insights from complex data. We discuss limitations of our work in Appendix D.4.

CHAPTER 6

EFFICIENT ADAPTIVE EXPERIMENTATION WITH NONCOMPLIANCE

This chapter is based on Oprescu et al. [2025].

We study the problem of estimating the average treatment effect (ATE) in adaptive experiments where treatment can only be encouraged—rather than directly assigned—via a binary instrumental variable. Building on semiparametric efficiency theory, we derive the efficiency bound for ATE estimation under arbitrary, history-dependent instrument-assignment policies, and show it is minimized by a variance-aware allocation rule that balances outcome noise and compliance variability. Leveraging this insight, we introduce AMRIV—an **A**daptive, **M**ultiply-**R**obust estimator for **I**nstrumental-**V**ariable settings with variance-optimal assignment. AMRIV pairs (i) an online policy that adaptively approximates the optimal allocation with (ii) a sequential, influence-function–based estimator that attains the semiparametric efficiency bound while retaining multiply-robust consistency. We establish asymptotic normality, explicit convergence rates, and anytime-valid asymptotic confidence sequences that enable sequential inference. Finally, we demonstrate the practical effectiveness of our approach through empirical studies, showing that adaptive instrument assignment, when combined with the AMRIV estimator, yields improved efficiency and robustness compared to existing baselines.

6.1 Introduction

Adaptive experimentation enables efficient estimation of treatment effects in sequential settings by adjusting assignment strategies based on accumulating data. Compared to traditional randomized controlled trials (RCTs), adaptive designs can reduce estimation variance, thus accelerating discovery and limit-

ing exposure to ineffective or harmful interventions. These methods are now widely used across domains—from medicine to online platforms—and have been formally endorsed by the U.S. Food and Drug Administration [GUIDANCE, 2018], driving both practical adoption and theoretical advances.

In many such settings, however, **direct treatment assignment is not feasible or ethical**. The treatment must instead be *encouraged* through a randomized recommendation or design choice—often referred to as an instrumental variable (IV)—leaving the final decision to the participant. This gives rise to *noncompliance*, where the assigned encouragement and the actual treatment may differ, and where self-selection based on unobserved factors introduces confounding that biases standard estimators. For instance, in a real TripAdvisor experiment, users were exposed to different premium sign-up interfaces that encouraged membership enrollment [Syrgkanis et al., 2019]. The actual treatment—whether a user subscribed—could not be enforced, but the interface (the instrument) could be randomized or adaptively adjusted. Similarly, in clinical trials, a physician may recommend a new medication but cannot compel adherence: the recommendation can be assigned, yet treatment uptake remains endogenous. Related challenges arise in recommender systems and public health interventions, where engagement or behavioral uptake is voluntary.

In such applications, reducing estimator variance is not merely a statistical preference; it determines how quickly and confidently a study can reach conclusions. Because high variance delays both the detection of harmful effects and the confirmation of beneficial ones, adaptive designs that learn to allocate instruments in variance-minimizing ways can mitigate these risks, yielding tighter confidence sequences and enabling earlier, statistically valid stopping decisions. Despite extensive work on adaptive designs for settings where treatment can be

directly enforced [Hahn et al., 2011, Kato et al., 2020, Cook et al., 2024], adaptive experimentation under noncompliance, where treatment is voluntary but encouragement can be adaptively controlled, remains largely unexplored, even though it describes many real-world scenarios.

This work addresses this gap. We study the problem of estimating the average treatment effect (ATE) in a sequential experiment where the experimenter can assign only a binary instrument, while the treatment itself is determined endogenously. Building on the semiparametric framework of Wang and Tchetgen Tchetgen [2018], which identifies the ATE under an unconfounded compliance assumption and provides a multiply robust, efficient influence-function–based estimator, we extend this framework to the adaptive setting. Specifically:

- We derive the **semiparametric efficiency bound** and characterize the **variance-optimal adaptive policy** that minimizes it through covariate-dependent instrument assignment.
- **We introduce AMRIV**, an **A**daptive, **M**ultiply **R**obust estimator for **IV** settings, which applies a sequential, plug-in version of the efficient influence function evaluated under the adaptive policy.
- We establish **strong theoretical guarantees**, including asymptotic normality, explicit convergence rates, multiply robust consistency, and time-uniform asymptotic confidence sequences for valid inference at arbitrary stopping times.
- We demonstrate **practical effectiveness** through both synthetic and semi-synthetic studies, showing improved efficiency and robustness over non-adaptive baselines and alternatives.

In contrast to prior work on adaptive design with instruments [Gupta et al., 2021, Chandak et al., 2024, Ailer et al., 2024, Zhao et al., 2024], our method focuses on point estimation of the ATE, achieves semiparametric efficiency, and supports multiply robust inference under adaptation. To our knowledge, this is the first method to bring the full suite of modern semiparametric tools—efficient influence functions, adaptive policy learning, robust plug-in estimation, and anytime-valid inference—to the adaptive IV setting under noncompliance.

6.2 Related Work

We provide a brief overview of related work here, with a more detailed discussion in Appendix E.1.

6.2.1 Instrumental Variables ATE Estimation

ATE identification in IV settings has traditionally relied on structural equation models (SEMs) that impose parametric assumptions on the outcome and treatment assignment mechanisms. More recent work has proposed flexible alternatives—such as DeepIV [Hartford et al., 2017], kernel IV [Singh et al., 2019], and orthogonal moment methods [Syrgkanis et al., 2019, Bennett et al., 2019]—that enable conditional effect estimation in high-dimensional or nonlinear settings. However, these approaches do not directly target robustness or semiparametric efficiency for the ATE. We instead build on the framework of Wang and Tchetgen Tchetgen [2018], which establishes point identification of the ATE via an unconfounded compliance assumption without requiring SEMs. Their influence-function–based estimator achieves semiparametric efficiency and is multiply robust, remaining consistent under partial nuisance misspecification. We extend this framework to the adaptive setting and use it as the

foundation for our estimator.

6.2.2 Adaptive Experimentation for ATE Estimation

A large and growing literature studies adaptive designs where the treatment itself can be directly assigned, with the goal of minimizing estimator variance or its regret analogue. Early two-stage designs asymptotically achieve the semiparametric efficiency bound [Hahn et al., 2011], and fully sequential approaches such as A2IPW attain variance-optimal Neyman allocation [Kato et al., 2020]. Subsequent extensions learn the allocation policy online [Kato et al., 2021] and add principled policy truncation with the first anytime-valid confidence sequences [Cook et al., 2024]. Parallel work from an online-learning perspective achieves sublinear or logarithmic "Neyman regret" via clipped or optimistic algorithms [Dai et al., 2023, Neopane et al., 2025b,c], matches finite-sample lower bounds under low-switching policies [Li et al., 2024], and extends to covariate-adaptive and off-policy settings [Kato et al., 2024, Lee and Ma, 2024]. Together, these methods form a mature toolkit—adaptive nuisance learning, cross-fitting, policy truncation, regret-style allocation, and time-uniform inference—but all assume the experimenter can randomize the *treatment* directly. Our work generalizes these efficiency guarantees to the less explored regime where only an *instrument* can be assigned and treatment uptake is endogenous.

6.2.3 Adaptive Experimentation with Instrumental Variables

A small but growing literature explores adaptive design in settings where only an instrument, rather than the treatment, can be assigned. Closest to our work is Chandak et al. [2024], who propose a practical influence-function–based procedure to reduce prediction error in nonparametric IV regression. However, they focus on prediction accuracy and do not address semiparametric efficiency or

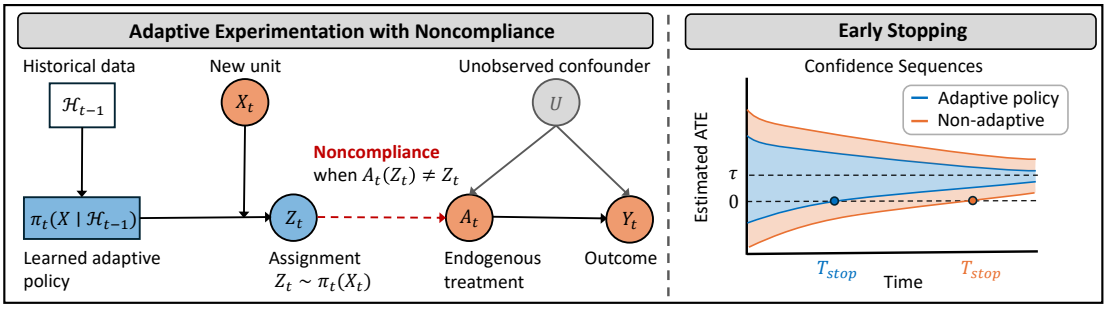


Figure 6.1: Adaptive experimentation with noncompliance. *Left:* causal structure of the sequential experiment, where Blue elements denote learned or assigned quantities (π_t, Z_t), orange elements represent observed variables (X_t, A_t, Y_t), and the dashed red arrow indicates noncompliance ($A_t(Z_t) \neq Z_t$). *Right:* schematic comparison of confidence-sequence contraction under adaptive vs. non-adaptive assignment, illustrating earlier stopping with the adaptive policy.

robustness to nuisances. Other approaches focus on partial identification [Ailer et al., 2024], adaptive data acquisition [Gupta et al., 2021], or regret minimization in bandit-style settings with endogeneity [Zhao et al., 2024, Della Vecchia and Basu, 2025]. In contrast, our goal is to enable efficient and robust adaptive ATE estimation under noncompliance, providing the first efficient and multiply robust estimator for this setting.

Additional related work on semiparametric inference, multiply robust estimation, and confidence sequences is discussed in detail in Appendix E.1.

6.3 Background and Setup

We consider the problem of estimating the average treatment effect (ATE) of a binary treatment $A \in \{0, 1\}$ on a real-valued outcome $Y \in \mathbb{R}$, in the presence of unobserved confounding, within an adaptive experimental setting. We adopt the potential outcomes framework, where each unit is associated with two potential outcomes, $Y(0)$ and $Y(1)$, corresponding to the outcomes under control and treatment, respectively. However, only the realized outcome $Y = Y(A)$, cor-

responding to the treatment actually received, is observed.

Each unit is also associated with covariates $X \in \mathbb{R}^m$, and we assume that the random variables $(X, Y(0), Y(1))$ are jointly distributed according to some unknown distribution P. Our goal is to estimate the ATE given by:

$$\tau := \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)].$$

Because direct treatment assignment may be infeasible, we rely on a binary instrumental variable $Z \in \{0, 1\}$ that influences treatment uptake. The instrument can be interpreted as a recommendation or encouragement—something the experimenter can assign, unlike the treatment itself. We denote by $Y(a, z)$ the potential outcome under treatment level a and instrument value z, and by $A(z)$ the potential treatment that would be taken under instrument assignment z. *Noncompliance* arises when the potential treatment does not follow the instrument, i.e., $A(z) \neq z$, reflecting endogenous treatment selection. Equivalently, a unit complies if its treatment adheres to the assigned instrument ($A(z) = z$).

The experiment proceeds over $T \in \mathbb{N}$ rounds. At each round t, a new unit with covariates X_t is drawn from P. The experimenter observes X_t and selects an instrument value $Z_t \sim \pi_t(\cdot \mid X_t, \mathcal{H}_{t-1})$, where π_t is an **adaptive policy** that depends on the current covariates X_t and past observations

$$\mathcal{H}_{t-1} := \{(X_1, Z_1, A_1, Y_1), \ldots, (X_{t-1}, Z_{t-1}, A_{t-1}, Y_{t-1})\}.$$

This allows the instrument-assignment policy to evolve over time based on accumulated data. Following the instrument assignment Z_t, the treatment $A_t = A_t(Z_t)$ is realized, and the outcome $Y_t = Y(A_t, Z_t)$ is observed. The full observation at time t is thus (X_t, Z_t, A_t, Y_t). After T rounds, the experimenter estimates the ATE from accumulated data $\mathcal{H}_T = \{(X_i, Z_i, A_i, Y_i)\}_{i=1}^T$. We emphasize that, unlike in standard adaptive experiments for ATE estimation, the experimenter in

our setting *cannot assign the treatment* directly. Instead, they can only assign the instrument Z, and treatment uptake is determined endogenously.

To identify the ATE under endogenous treatment selection (that may be influenced by unobserved confounders), we adopt standard instrumental variable assumptions, as well as the unconfounded compliance condition introduced in Wang and Tchetgen Tchetgen [2018]. We summarize these below.

Assumption 6.1 (Standard IV Assumptions)**.** The following properties hold: (*Exclusion*) $Y(a, z) = Y(a)$—the instrument affects the outcome only through the treatment; (*Independence*) $Z \perp\!\!\!\perp U \mid X$—the instrument is independent of unobserved confounders U given covariates; and (*Relevance*) $\text{Cov}(Z, A \mid X) \neq 0$—the instrument has an effect on treatment uptake for almost every X.

Assumption 6.2 (Unconfounded Compliance, from [Wang and Tchetgen Tchetgen, 2018])**.** The treatment effect is independent of compliance status given covariates: $Y(1) - Y(0) \perp A(1) - A(0) \mid X$.

Assumption 6.1 is standard in the IV literature and ensures instrument validity. The independence assumption can often be satisfied by design, e.g., via randomization. While sufficient for identifying the local average treatment effect (LATE), these assumptions do not identify the ATE under effect heterogeneity or treatment endogeneity. To enable ATE identification, we invoke Assumption 6.2 from Wang and Tchetgen Tchetgen [2018], which assumes the treatment effect is mean-independent of compliance type given covariates, ruling out interactions with unobserved confounding.

Remark 6.3 (Interpretation under violations of Assumption 6.2). If Assumption 6.2 does not hold, the ATE is no longer point-identified. The estimand then shifts to the average conditional local average treatment effect (ACLATE), which

averages treatment effects among instrument-responsive individuals (compliers) across covariate strata. The ACLATE remains identified and interpretable, capturing how causal effects vary among compliers when compliance is confounded and cannot be fully controlled.

With Assumption 6.1 and 6.2, the ATE can be point-identified. For notational convenience, we define the instrument-induced outcome and treatment models $\mu^Y(z, X) := \mathbb{E}[Y \mid Z = z, X]$ and $\mu^A(z, X) := \mathbb{E}[A \mid Z = z, X]$ for $z \in \{0, 1\}$. The ATE can then be expressed as (Theorem 1 from [Wang and Tchetgen Tchetgen, 2018]):

$$\tau = \mathbb{E}_X \left[\frac{\mu^Y(1, X) - \mu^Y(0, X)}{\mu^A(1, X) - \mu^A(0, X)} \right] := \mathbb{E}_X \left[\frac{\delta^Y(X)}{\delta^A(X)} \right], \tag{6.1}$$

where $\delta^Y(X)$ and $\delta^A(X)$ denote the instrument-induced shifts in outcome and treatment, respectively. We refer to $\delta^A(X)$ as the *compliance score*, representing the instrument's effect on treatment uptake.

In the non-adaptive setting, where the instrument assignment policy is fixed over time—*i.e.*, $\pi_t(1 \mid X, \mathcal{H}_{t-1}) \equiv \pi(X)$—Wang and Tchetgen Tchetgen [2018] (Theorem 5) derive the efficient influence function (EIF) for the ATE estimator in Equation 6.1. Let $\pi(x), \eta(x) := \{\mu^Y(0, x), \mu^A(0, x), \delta^A(x), \delta(x)\}$ denote the nuisances, where $\delta(X) := \delta^Y(X)/\delta^A(X)$. The (Recentered) EIF is then given by

$$\begin{aligned} &\phi(X, Z, A, Y; \pi, \eta) \\ &= \frac{2Z - 1}{Z\pi(X) + (1 - Z)(1 - \pi(X))} \frac{1}{\delta^A(X)} \left[Y - A\delta(X) - \mu^Y(0, X) + \mu^A(0, X)\delta(X) \right] + \delta(X). \end{aligned} \tag{6.2}$$

The corresponding estimator—known as the multiply robust instrumental variables estimator (MRIV) [Wang and Tchetgen Tchetgen, 2018, Frauen and Feuerriegel, 2023]—uses plug-in estimates of nuisance functions within the recentered efficient influence function. It attains the semiparametric efficiency bound when all nuisances are correctly specified and remains consistent under partial misspecification.

We extend this framework to the adaptive setting, where the instrument assignment policy π_t evolves with accumulating data. We characterize the optimal adaptive policy that minimizes asymptotic variance and develop **AMRIV**, an **A**daptive **M**ultiply **R**obust **IV** estimator that combines adaptive policy learning with sequential influence-function–based estimation. Our method achieves semiparametric efficiency, ensures multiply robust consistency, and enables valid time-uniform inference.

Notation We write $\pi_t(X_t \mid \mathcal{H}_{t-1}) := \pi_t(1 \mid X_t, \mathcal{H}_{t-1})$ for the probability of assigning $Z_t = 1$ given covariates and history. The L_2 norm of a function f is $\|f\|_2 := \mathbb{E}_P[f(X)^2]^{1/2}$, and $\widehat{f_t}$ denotes an estimate of f based on t samples. We use $\widehat{\mathbb{E}}$ to denote empirical expectations computed from data. Notation used throughout the chapter is summarized in Appendix E.2 (Table E.1).

6.4 Efficiency Bounds and Optimal Instrument Assignment

To guide optimal experiment design under the IV setting, we derive the semiparametric efficiency bound for ATE estimation under a fixed instrument policy $\pi(X)$. This characterizes the variance-minimizing allocation strategy and motivates our adaptive estimator.

Theorem 6.4 (Semiparametric Efficiency Bound)**.** *Under Assumption 6.1 and 6.2, the semiparametric efficiency bound for estimating the ATE τ is given by*

$$V_{\mathit{eff}}(\pi) := \mathbb{E}\left[\frac{1}{\delta^A(X)^2}\left(\frac{\sigma^2(1,X)}{\pi(X)} + \frac{\sigma^2(0,X)}{1-\pi(X)}\right) + (\delta(X) - \tau)^2\right], \tag{6.3}$$

where $\sigma^2(z, X) = \mathrm{Var}(Y - A\delta(X) \mid Z = z, X)$.

Corollary 6.5 (Optimal Instrument Assignment)**.** *The assignment policy $\pi^*(X)$ that*

minimizes the efficiency bound in Theorem 6.4 is given by

$$\pi^*(X) = \frac{\sqrt{\sigma^2(1,X)}}{\sqrt{\sigma^2(1,X)} + \sqrt{\sigma^2(0,X)}}. \tag{6.4}$$

Proofs of Theorem 6.4 and Corollary 6.5 are included in Appendix E.5.

Drivers of Efficient Allocation From Corollary 6.5, the optimal assignment policy $\pi^*(X)$ allocates more weight to the arm z with higher residual variance $\mathrm{Var}(Y - A\delta(X) \mid Z = z, X)$, where

$$\begin{aligned}\mathrm{Var}(Y - A\delta(X) \mid Z = z, X) = \mathrm{Var}(Y \mid Z = z, X) + \delta(X)^2 \,\mathrm{Var}(A \mid Z = z, X) \\ - 2\delta(X)\,\mathrm{Cov}(Y, A \mid Z = z, X).\end{aligned}$$

This expression reveals that, unlike in standard adaptive ATE estimation, the residual variance depends jointly on both outcome noise ($\mathrm{Var}(Y \mid Z = z, X)$) and compliance noise ($\mathrm{Var}(A \mid Z = z, X)$). When compliance is more uncertain in one arm, the estimator becomes noisier in that region, leading the optimal policy to allocate more probability mass to that arm to compensate.

Connection to Neyman Allocation In the special case of perfect compliance (when $A(Z) = Z$), the treatment is fully determined by the (conditionally) randomized instrument and our setting becomes the classical adaptive ATE estimation scenario. In this setting, $\mathrm{Var}(Y - A\delta(X) \mid Z = z, X) = \mathrm{Var}(Y - A\delta(X) \mid A = z, X) = \mathrm{Var}(Y \mid A = z, X)$ and thus the optimal allocation reduces to $\frac{\sqrt{\mathrm{Var}(Y|A=1,X)}}{\sqrt{\mathrm{Var}(Y|A=0,X)} + \sqrt{\mathrm{Var}(Y|A=1,X)}}$ which exactly matches the classical Neyman allocation for minimizing the variance of a difference-in-means estimator [Neyman, 1992, Kato et al., 2020]. This highlights that our policy generalizes Neyman allocation to settings with noncompliance and endogenous treatment, adjusting for both outcome and compliance-driven noise.

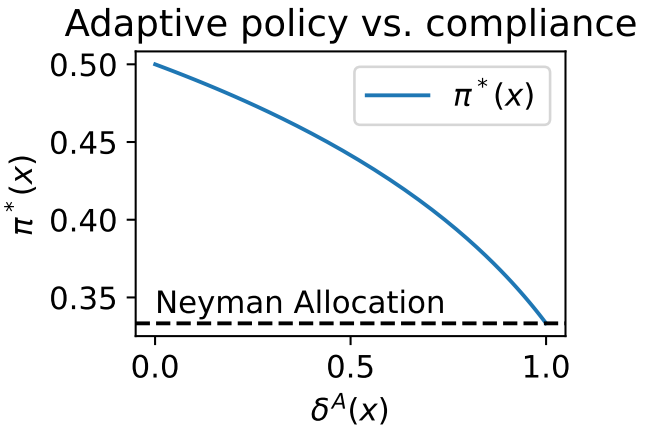


Figure 6.2: Variance-optimal instrument-assignment policy $\pi^*(X)$ as a function of the compliance score $\delta^A(X)$. The figure compares the adaptive IV allocation rule with classical Neyman allocation as compliance varies.

Motivating Illustration Consider an example with *one-sided noncompliance*—treatment is only accessible to those who receive the instrument, so $\mu^A(0, X) = 0$. This captures scenarios such as vaccine access, product rollouts, or behavioral nudges. Let compliance vary with $X \in \mathbb{R}$ via $\delta^A(x) = \mu^A(1, x) = \sigma(-2x)$, and let outcomes follow $Y = f(A, X) + uA + \epsilon_A$, where u is a fixed unobserved confounder and $\epsilon_A \sim \mathcal{N}(0, A + 4(1 - A))$, with higher variance in the control arm. As shown in Figure 6.2, the optimal policy $\pi^*(X)$ approaches Neyman allocation when $\delta^A(X) \to 1$, but shifts toward uniform allocation when $\delta^A(X) \to 0$. This reflects a key design intuition: *under low compliance, assigning more units to $Z = 1$ compensates for scarce treatment uptake*, helping preserve estimator efficiency.

6.5 Adaptive Estimation of the Average Treatment Effect

We propose an adaptive framework for estimating the ATE in sequential experiments with a binary instrument. Our goal is to minimize the semiparametric efficiency bound from Theorem 6.4 by combining: (1) **instrument assignment** via a data-driven policy $\pi_t(X_t \mid \mathcal{H}_{t-1})$ that approximates the optimal allocation

$\pi^*(X)$; and (2) **treatment effect estimation** using an adaptive plug-in version of the multiply robust estimator from Equation 6.2. Although the theory assumes per-round updates, the method also applies in batch settings with fewer updates. We detail both components below.

6.5.1 Adaptive Instrument Assignment

To stabilize nuisance estimation, we begin with a burn-in phase of $T_0 < T$ rounds using a fixed policy $\pi_{\text{init}}(X)$, such as uniform randomization. From round $T_0 + 1$ onward, instruments are assigned via a data-driven policy $\pi_t(X \mid \mathcal{H}_{t-1})$ that approximates the optimal allocation in Corollary 6.5. Specifically, we compute a plug-in estimate $\widetilde{\pi}_t(X \mid \mathcal{H}_{t-1})$ as

$$\widetilde{\pi}_t(X \mid \mathcal{H}_{t-1}) := \frac{\sqrt{\widehat{\sigma}^2_{t-1}(1, X)}}{\sqrt{\widehat{\sigma}^2_{t-1}(0, X)} + \sqrt{\widehat{\sigma}^2_{t-1}(1, X)}}, \tag{6.5}$$

where $\widehat{\sigma}^2_{t-1}(z, X)$ is an estimate of the conditional residual variance $\text{Var}(Y - A\delta(X) \mid Z = z, X)$ based on data in $\mathcal{H}_{t-1}$. We then apply a truncation step to $\widetilde{\pi}_t(X \mid \mathcal{H}_{t-1})$ (see below) to obtain the final assignment policy $\pi_t(X \mid \mathcal{H}_{t-1})$ used at time t.

Residual–variance estimation One option for estimating $\widehat{\sigma}^2_{t-1}(z, X)$ is through the decomposition

$$\text{Var}(Y - A\delta(X) \mid Z = z, X) = \mathbb{E}[(Y - A\delta(X))^2 \mid Z = z, X] - (\mu^Y(0, X) - \mu^A(0, X)\,\delta(X))^2.$$

We proceed in two stages: (i) fit $\widehat{\mu}^Y_{t-1}(0, X)$, $\widehat{\mu}^A_{t-1}(0, X)$ and $\widehat{\delta}_{t-1}(X)$ using $\mathcal{H}_{t-1}$; (ii) form residuals $\widehat{R}_{t-1} = Y - A\,\widehat{\delta}_{t-1}(X)$ and regress $\widehat{R}^2_{t-1}$ on (Z, X) to obtain $\widehat{s}_{t-1}(z, X) := \widehat{\mathbb{E}}[\widehat{R}^2_{t-1} \mid Z = z, X]$.

Unbiased two-stage estimation via cross-fitting To mitigate finite-sample bias in estimating $\widehat{\sigma}^2_{t-1}(z, X)$, we apply the sequential cross-fitting scheme of Waudby-Smith et al. [2024a]. Thus, we split $\mathcal{H}_{t-1}$ into two temporal folds

$\mathcal{H}_{t-1}^{(j)} = \{(X_i, Z_i, A_i, Y_i) : i \in [t-1], i \mod 2 = j\}$, $j \in \{0,1\}$, fit $\widehat{\delta}_{t-1}$ on one and compute residuals $\widehat{R}_{t-1}^2$ on the other, and *vice-versa*. The combined residuals are used to regress $\widehat{R}_{t-1}^2$ on (Z, X) to estimate $\widehat{s}_{t-1}(z, X)$. Since $\widehat{\mu}_{t-1}^Y$ and $\widehat{\mu}_{t-1}^A$ do not depend on other nuisances, they can be learned on the full history $\mathcal{H}_{t-1}$.

Nuisance learners Any sequentially consistent nonparametric regressor can be used for $\widehat{\mu}_{t-1}^Y, \widehat{\mu}_{t-1}^A$, and $\widehat{s}_{t-1}$, *e.g.* k-NN [Yang and Zhu, 2002], kernel smoothers [Qian and Yang, 2016], random forests [Wager and Athey, 2018b], or neural nets [Schmidt-Hieber, 2020]. $\widehat{\mu}_{t-1}^A$ may also be estimated via these methods or a classifier such as logistic regression. We compute $\widehat{\delta}_{t-1}(X) = \widehat{\delta}_{t-1}^Y(X)/\widehat{\delta}_{t-1}^A(X)$, where $\widehat{\delta}_{t-1}^Y$ is estimated via either a difference of regressions $\widehat{\mu}_{t-1}^Y(1, X) - \widehat{\mu}_{t-1}^Y(0, X)$ or a direct IPW-style regression $\widehat{\mathbb{E}}\left[\frac{YZ}{\pi_{t-1}(X)} - \frac{Y(1-Z)}{1-\pi_{t-1}(X)} \mid X\right]$. An estimate of $\widehat{\delta}_{t-1}^A(X)$ is obtained analogously by replacing Y with A.

To guarantee non-negativity of the estimated variances, we define $\widehat{\sigma}_{t-1}^2(z, X)$ as

$$\widehat{\sigma}_{t-1}^2(z, X) = \begin{cases} \left\{\widehat{s}_{t-1}(z, X) - (\widehat{\mu}_{t-1}^Y(0, X) - \widehat{\mu}_{t-1}^A(0, X)\,\widehat{\delta}_{t-1}(X))^2\right\} & \text{if } \{\cdots\} > 0 \\ \varepsilon & \text{otherwise} \end{cases} \tag{6.6}$$

for a small constant $\varepsilon > 0$.

Remark 6.6 (Choice of nuisance and variance estimators). Sequential cross-fitting removes the need for restrictive conditions (e.g., Donsker)—standard nonparametric convergence rates suffice for the guarantees in our main theorems. In practice, any consistent learner (e.g., k-NN, random forests, or neural networks) can be used, depending on data structure and sample size (see Appendix E.3). The variance estimator in Eq. (6.6) ensures non-negativity via a small floor ε, though fully nonnegative alternatives such as self-normalized kernel estimators may also be used. Appendix E.3 further discusses these options and outlines online or streaming implementations that avoid full data storage.

Truncation for Finite-Sample Stability Following recent work on adaptive ATE estimation without endogenous treatment assignment (*e.g.*, [Cook et al., 2024, Dai et al., 2023, Neopane et al., 2025b]), we apply a truncation scheme to stabilize the assignment policy $\widetilde{\pi}_t(X \mid \mathcal{H}_{t-1})$. After computing the raw plug-in policy $\widetilde{\pi}_t(X \mid \mathcal{H}_{t-1})$ via Eq. (6.5), we define the truncated policy $\pi_t(X \mid \mathcal{H}_{t-1})$ as

$$\pi_t(X \mid \mathcal{H}_{t-1}) := \min\left\{1 - \frac{1}{k_t}, \max\left\{\frac{1}{k_t}, \widetilde{\pi}_t(X \mid \mathcal{H}_{t-1})\right\}\right\}, \tag{6.7}$$

where $k_t \in [2, \infty)$ is a truncation parameter satisfying $k_t \to \infty$ as $t \to \infty$. Truncation ensures that the instrument assignment probabilities remain bounded away from 0 and 1, thereby improving finite-sample stability and leading to better theoretical guarantees.

6.5.2 AMRIV: Adaptive Multiply Robust Estimation of the ATE

We now introduce our estimator, **AMRIV**, which adaptively estimates the ATE using the recentered efficient influence function in Eq. (6.2) evaluated on sequentially updated plug-in estimates of nuisance functions. The estimator is defined as

$$\widehat{\tau}_T^{\text{AMRIV}} := \frac{1}{T}\sum_{t=1}^{T} \phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t), \tag{6.8}$$

where $\widehat{\eta}_t = \{\widehat{\mu}^Y_{t-1}(0, X), \widehat{\mu}^A_{t-1}(0, X), \widehat{\delta}^A_{t-1}(X), \widehat{\delta}_{t-1}(X)\}$ denotes plug-in estimates of the nuisance functions at time t, constructed solely from the past data $\mathcal{H}_{t-1}$ (Note: the instrument assignment policy $\pi_t(X \mid \mathcal{H}_{t-1})$, defined by the experimenter based on the estimated optimal rule from Section 6.5.1, is treated as known and does not require further estimation from data). This construction confers the estimator $\widehat{\tau}_T^{\text{AMRIV}}$ a *near-martingale* structure, that is, it can be written as the sum of a true martingale difference sequence and a remainder term of order $o_p(T^{-1/2})$, enabling, as we will show in Section 6.6, valid asymptotic inference under sequential dependence.

Algorithm 6.1 AMRIV: Adaptive Multiply Robust IV Estimation

Input: Burn-in period T_0; initial policy $\pi_{\text{init}}(X)$; regression/classification learners for $\mu^Y(z,X)$, $\mu^A(z,X)$, $\delta(X)$, $\delta^A(X)$, $s(z,X)$

for $t = 1, \ldots, T$ **do**
  Observe covariates X_t
  if $t \le T_0$ **then**
    Assign $Z_t \sim \text{Bern}(\pi_{\text{init}}(X_t))$
  else
    Estimate nuisance functions $\widehat{\mu}^Y_{t-1}(0,X)$, $\widehat{\mu}^A_{t-1}(0,X)$, $\widehat{\delta}_{t-1}(X)$, and $\widehat{s}_{t-1}(z,X)$ from $\mathcal{H}_{t-1}$ using cross-fitting
    Compute $\widehat{\sigma}^2_{t-1}(z,X)$ using Eq. (6.6)
    Compute plug-in assignment probability

$$\widetilde{\pi}_t(X \mid \mathcal{H}_{t-1}) = \frac{\sqrt{\widehat{\sigma}^2_{t-1}(1,X)}}{\sqrt{\widehat{\sigma}^2_{t-1}(0,X)} + \sqrt{\widehat{\sigma}^2_{t-1}(1,X)}}$$

    Apply truncation to obtain

$$\pi_t(X \mid \mathcal{H}_{t-1}) := \min\left\{1 - \frac{1}{k_t},\ \max\left\{\frac{1}{k_t},\ \widetilde{\pi}_t(X \mid \mathcal{H}_{t-1})\right\}\right\}$$

    Assign $Z_t \sim \text{Bern}(\pi_t(X_t \mid \mathcal{H}_{t-1}))$
  end if
  Observe instrumented treatment $A_t = A(Z_t)$ and outcome $Y_t = Y(A_t)$
  Construct $\widehat{\eta}_t = \{\widehat{\mu}^Y_{t-1}(0,X), \widehat{\mu}^A_{t-1}(0,X), \widehat{\delta}^A_{t-1}(X), \widehat{\delta}_{t-1}(X)\}$
  Compute $\phi_t = \phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t)$ using Eq. (6.9)
end for
return $\widehat{\tau}^{\text{AMRIV}}_T = \frac{1}{T}\sum_{t=1}^T \phi_t$

Nuisance Estimation The nuisances $\widehat{\mu}^Y_{t-1}(0,X)$, $\widehat{\mu}^A_{t-1}(0,X)$, $\widehat{\delta}^A_{t-1}(X)$, $\widehat{\delta}_{t-1}(X)$ can be estimated using any flexible nonparametric regression method applied to the historical data $\mathcal{H}_{t-1}$. To reduce computational overhead, we can reuse the estimates of $\widehat{\mu}^Y_{t-1}(0,X)$ and $\widehat{\mu}^A_{t-1}(0,X)$ previously obtained for instrument assignment in Section 6.5.1. Similarly, the estimate of $\widehat{\delta}_{t-1}(X)$ can be formed by averaging the cross-fitted estimates $\widehat{\delta}^{(0)}_{t-1}$ and $\widehat{\delta}^{(1)}_{t-1}$, trained on the two data folds $\mathcal{H}^{(0)}_{t-1}$ and $\mathcal{H}^{(1)}_{t-1}$, respectively. The only remaining component is $\widehat{\delta}^A_{t-1}(X)$, which must be esti-

mated separately if it was not already computed as part of the $\widehat{\delta}_{t-1}(X)$ estimation pipeline.

For completeness, the final estimate of the (R)EIF at time t is given by

$$\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) = \frac{2Z_t - 1}{Z_t \pi_t(X_t \mid \mathcal{H}_{t-1}) + (1 - Z_t)(1 - \pi_t(X_t \mid \mathcal{H}_{t-1}))} \frac{1}{\widehat{\delta}^A_{t-1}(X_t)} \cdot \left[Y_t - A_t \widehat{\delta}_{t-1}(X_t) - \widehat{\mu}^Y_{t-1}(0, X_t) + \widehat{\mu}^A_{t-1}(0, X_t) \widehat{\delta}_{t-1}(X_t) \right] + \widehat{\delta}_{t-1}(X_t) \quad (6.9)$$

where all quantities are constructed from $\mathcal{H}_{t-1}$. The full procedure is summarized in Algorithm 6.1. Unlike prior adaptive ATE methods without IVs, the estimator $\widehat{\tau}^{\text{AMRIV}}_T = \frac{1}{T} \sum_{t=1}^{T} \phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t)$ *cannot* be written as a martingale difference sequence. Hence, standard MDS central limit theorems do not apply directly, and we must instead decompose the estimator to recover a suitable martingale structure.

6.6 Theoretical Guarantees

This section provides theoretical guarantees for the AMRIV estimator. We establish its asymptotic normality, characterize its convergence rates, and demonstrate its multiply-robust consistency. Furthermore, in Appendix E.4, we consider the sequential inference setting and derive asymptotically-valid, time-uniform *confidence sequences* for the AMRIV estimator.

6.6.1 Efficiency and Asymptotic Normality of the AMRIV Estimator

We start by establishing the asymptotic properties of the AMRIV estimator $\widehat{\tau}^{\text{AMRIV}}_T$. We first introduce the following assumption:

Assumption 6.7 (Bounded Outcomes and Nuisances)**.** The potential outcomes and nuisance function estimates are uniformly bounded. That is, there exists a

constant $C > 0$ such that, for all t and x,

$$|Y_t(0)|,\ |Y_t(1)| \le C, \qquad |\widehat{\mu}_t^Y(0,x)|,\ |\widehat{\mu}_t^A(0,x)|,\ |\widehat{\delta}_t(x)|,\ |\widehat{\delta}_t^A(x)|^{-1} \le C.$$

This boundedness assumption is standard in influence-function-based ATE estimation and ensures stability of the estimator. With this assumption in place, we now state our main result on the asymptotic efficiency of the AMRIV estimator.

Theorem 6.8 (Asymptotic Normality of the AMRIV Estimator)**.** *Suppose Assumptions 6.1, 6.2 and 6.7 hold and there exists a non-adaptive policy $\pi(X) \in [\epsilon, 1-\epsilon]$ for some $\epsilon > 0$ such that the nuisances estimates $\widehat{\eta}_t$ and the adaptive assignment policy $\pi_t(X \mid \mathcal{H}_{t-1})$ are L_2-consistent relative to the truncation schedule, i.e. $k_t\|\widehat{f}_{t-1} - f\|_2 = o_p(1)$ and $k_t\|\pi_t - \pi\|_2 = o_p(1)$ for $f \in \{\mu^Y(0,\cdot), \mu^A(0,\cdot), \delta(\cdot), \delta^A(\cdot)\}$. Furthermore, assume $\|\widehat{\delta}_{t-1} - \delta\|_2 \|\widehat{\delta}_{t-1}^A - \delta^A\|_2 = o_p(t^{-1/2})$. Then, the AMRIV estimator is asymptotically normal:*

$$\sqrt{T}\left(\widehat{\tau}_T^{AMRIV} - \tau\right) \xrightarrow{d} \mathcal{N}\left(0,\ V_{\mathit{eff}}(\pi)\right), \tag{6.10}$$

where V_{eff} is defined in Theorem 6.4. In particular, if we have $\pi(X) = \pi^*(X)$, then $\widehat{\tau}_T^{AMRIV}$ is semiparametrically efficient.

The key insight behind Theorem 6.8 is that the AMRIV estimator admits the following near-martingale decomposition: $\sqrt{T}(\widehat{\tau}_T^{\mathrm{AMRIV}} - \tau) = \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^T z_t\right) + \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^T m_t\right)$, where $z_t = \phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta) - \tau$ is a martingale difference sequence (MDS), and $m_t = \phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) - \phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta)$ is an asymptotically vanishing term that captures the impact of estimating the nuisance functions. The first term satisfies a central limit theorem for MDS under standard Lindeberg-type conditions [Zhang et al., 2021], while the second is controlled by the L_2-consistency of the nuisance estimates. We formalize this in Appendix E.6. Importantly, this result holds under mild assumptions: we only require L_2 convergence (no pointwise convergence or Donsker conditions [Bickel et al., 1993]),

bounded outcomes and nuisance estimates, and L_2-consistency of the nuisance components w.r.t. the truncation schedule. This allows AMRIV to accommodate flexible, data-dependent policies and sequential nuisance estimation.

The truncation schedule k_t plays a central role by ensuring positivity of $\pi_t(X)$—crucial for variance control—while still allowing π_t to approach an optimal policy π^* as $k_t \to \infty$. For this to hold, we must ensure $\lim_{t\to\infty} k_t > \sup_X \frac{1}{\pi^*(X)}$, so the truncation threshold does not constrain the optimal allocation in the limit and semiparametric efficiency can be achieved (see last line in Theorem 6.8). This mirrors tradeoffs in efficient ATE estimation [Cook et al., 2024], where adaptive truncation stabilizes estimation without distorting the estimator asymptotically. In practice, when the plug-in policy $\tilde{\pi}_t$ is uniformly bounded away from 0 and 1, truncation becomes unnecessary: setting $k_t = 1/\min_X\{\tilde{\pi}_t(X), 1 - \tilde{\pi}_t(X)\}$ ensures $\pi_t = \tilde{\pi}_t$ for all t.

6.6.2 Consistency Guarantees under Partial Nuisance Misspecification

As shown in Theorem 6.8, the convergence rate of AMRIV is primarily governed by the estimation error of $\widehat{\delta}(X)$ and $\widehat{\delta}^A(X)$. This reflects its *multiply robust* property: AMRIV remains consistent even when other nuisance components are misspecified. This robustness goes beyond prior work on IV methods in adaptive settings, where such guarantees were not established. The next two results formalize AMRIV's convergence rate and multiply robust consistency.

Theorem 6.9 (Convergence Rate of the AMRIV Estimator)**.** *Suppose Assumptions 6.1, 6.2 and 6.7 hold, and that there exists a non-adaptive policy $\pi(X) \in [\epsilon, 1-\epsilon]$ for some $\epsilon > 0$ such that the adaptive policies π_t satisfy $k_t\|\pi_t - \pi\|_2 = o_p(1)$. Let $\widetilde{\eta} = \{\widetilde{\mu}^Y(0,\cdot), \widetilde{\mu}^A(0,\cdot), \widetilde{\delta}(\cdot), \widetilde{\delta}^A(\cdot)\}$ denote a possibly misspecified limit of the nuisance functions, and suppose that $k_t\|\widehat{f}_{t-1} - \widetilde{f}\|_2 = o_p(1)$ for each $\widetilde{f} \in \widetilde{\eta}$. Then the AMRIV*

estimator satisfies

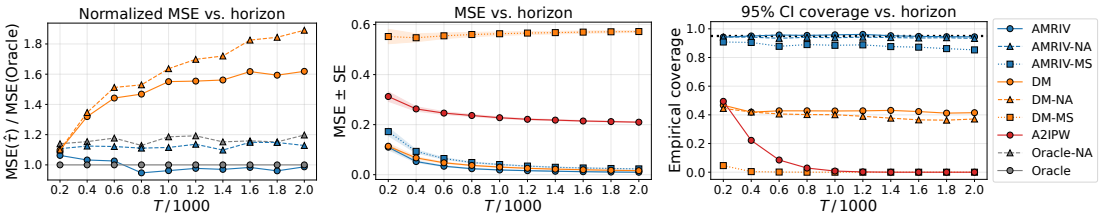
$$\left|\widehat{\tau}_T^{\text{AMRIV}} - \tau\right| \;=\; O_p(T^{-1/2}) + O_p\left(\|\widehat{\delta}_T^A - \delta^A\|_2 \|\widehat{\delta}_T - \delta\|_2\right). \tag{6.11}$$

Corollary 6.10 (Multiply Robust Consistency Guarantees)**.** *Under the conditions of Theorem 6.9, if either $\widehat{\delta}_t$ or $\widehat{\delta}_t^A$ is L_2-consistent, then the AMRIV estimator $\widehat{\tau}_T^{\text{AMRIV}}$ is consistent for τ.*

We provide a proof of Theorem 6.9 and Corollary 6.10 in Appendix E.7. An immediate consequence of Theorem 6.9 is that if both $\widehat{\delta}(X)$ and $\widehat{\delta}^A(X)$ converge at rate $o_p(T^{-1/4})$, AMRIV achieves the parametric $O_p(T^{-1/2})$ rate. This is usually attainable under mild regularity conditions, even with flexible nonparametric models. Furthermore, Corollary 6.10 shows that AMRIV inherits the *multiply robust* property from its static counterpart [Wang and Tchetgen Tchetgen, 2018, Frauen and Feuerriegel, 2023]. In the static setting, the MRIV converges if either (i) $\widehat{\mu}^Y(0,\cdot)$, $\widehat{\mu}^A(0,\cdot)$, $\widehat{\delta}$ are correctly specified, (ii) both $\widehat{\pi}$ and $\widehat{\delta}^A$ are, or (iii) $\widehat{\pi}$ and $\widehat{\delta}$ are. However, in the adaptive setting, we can establish a stronger result: even if the outcome-related nuisance functions $\widehat{\mu}_t^Y(0,\cdot)$ and $\widehat{\mu}_t^A(0,\cdot)$ are misspecified, AMRIV is still consistent as long as *one* of $\widehat{\delta}_t^A$ or $\widehat{\delta}_t$ converges. This is due to the adaptive setting design where we control the instrument assignment $\pi_t(X_t \mid \mathcal{H}_{t-1})$ which confers robustness to misspecification in $\widehat{\mu}_t^Y(0,\cdot)$ and $\widehat{\mu}_t^A(0,\cdot)$. Thus, our adaptive generalization preserves the multiple robustness property, making it particularly well-suited for practice where some nuisance components may be difficult to estimate reliably.

6.7 Experimental Results

We demonstrate the practical effectiveness of our approach in both synthetic and semi-synthetic studies. In each setting, we compare our estimator (**AMRIV**)



(a) Efficiency (b) Consistency (c) Coverage

Figure 6.3: Performance of AMRIV and baseline estimators as a function of sample size T in the synthetic adaptive-IV experiment. **(a)** Efficiency: Normalized MSE relative to an oracle benchmark. **(b)** Consistency: MSE ± standard errors. **(c)** Coverage: Empirical coverage of nominal 95% confidence intervals.

to its non-adaptive counterpart (**AMRIV-NA**), which assigns the instrument uniformly at random; the plug-in direct method from Eq. (6.1) (**DM**) and its non-adaptive version (**DM-NA**); the **A2IPW** estimator of [Kato et al., 2020]; and two oracle baselines: a fully oracle-efficient estimator using the true nuisance functions (**Oracle**) and a non-adaptive version (**Oracle-NA**). To assess robustness, we also evaluate misspecified variants of AMRIV and DM—denoted **AMRIV-MS** and **DM-MS**—in which the $\delta(X)$ estimator is deliberately misspecified.

Across both experiments, we evaluate three desiderata: (i) *efficiency*, measured by normalized MSE relative to the Oracle estimator; (ii) *consistency*, assessed via MSE decay with sample size T; and (iii) *coverage*, computed from empirical 95% confidence intervals. We implement all estimators using Random Forests (RF, [Breiman, 2001]) and update the nuisance estimates in mini-batches for efficiency. Further implementation details, including model hyperparameters and ablations, are provided in Appendix E.8. The replication code is available at `https://github.com/CausalML/Adaptive-IV`.

6.7.1 Simulation Studies with Synthetic Data

We construct a synthetic environment with one-sided noncompliance, where the treatment A is only accessible to those who receive the instrument $Z = 1$. At

each time t, we sample covariates X_t, assign the instrument $Z_t \sim \pi_t(X_t \mid \mathcal{H}_{t-1})$, and realize $A_t = C_t Z_t$, where C_t is a latent compliance indicator sampled from Bern($\delta^A(X_t)$). The outcome Y_t depends on A_t, X_t, an unobserved confounder U_t, and heteroskedastic noise. The full data-generating process is detailed in Appendix E.8.1.

We set $T = 2000$, $T_0 = 200$, and run 1000 trajectories. All estimators are updated in batches of size $b = 200$ and implemented using Random Forests (RFs) when applicable. For the adaptive estimators, we use the truncated optimal policy in Eq. (6.7), with truncation schedule $k_t = 2/0.999^t$. AMRIV uses RF classifiers for $\delta^A(X)$ (clipped at 0.01) and RF regressors for $\mu^Y(z, X)$, while $\delta(X)$ is computed via the plug-in ratio. A2IPW follows Kato et al. [2020] with Neyman allocation and RF regressors. To induce misspecification, we replace $\widehat{\mu}^Y(1, X)$ with the constant $\widehat{\mathbb{E}}[\mu^Y(1, X)]$, flattening outcome heterogeneity. Figure 6.3 summarizes the experimental results.

Adaptivity As shown in panel (a), adaptive design consistently improves the efficiency of all estimators. AMRIV approaches the oracle benchmark despite using estimated nuisances, while AMRIV-NA exhibits a constant efficiency gap due to suboptimal allocation. This illustrates how adaptivity enables more effective data collection: by dynamically allocating instruments to regions of high uncertainty, AMRIV concentrates sampling effort where it contributes most to precision. The effect is particularly evident under one-sided noncompliance, where asymmetries in both outcome and compliance variance make uniform allocation especially inefficient (Theorem 6.4). Panel (a) also confirms that AMRIV and AMRIV-NA converge at the expected $O_p(T^{-1/2})$ rate (Theorem 6.9), whereas DM and DM-NA converge more slowly, as their normalized MSE grows with the sample size T.

Consistency Panel (b) confirms that AMRIV, AMRIV-NA, and DM converge to the true τ, with AMRIV variants achieving lower error due to variance-aware allocation. In contrast, A2IPW is biased and fails to converge, as expected, since it does not correct for unobserved confounding in treatment selection. The comparison further highlights the importance of robust nuisance estimation: DM-MS diverges when $\delta(X)$ is misspecified, while AMRIV-MS remains consistent. This robustness reflects the multiply-robust property formalized in Theorem 6.9, which ensures consistency as long as either the compliance model or the outcome model is estimated correctly—an especially desirable feature in practice when some nuisance components are difficult to learn reliably.

Coverage In panel (c), we evaluate the empirical coverage of 95% asymptotic confidence intervals. Only AMRIV and AMRIV-NA achieve nominal coverage, consistent with our theoretical guarantees (Theorem 6.8). The misspecified and plug-in methods under-cover, with DM and A2IPW performing particularly poorly as T grows, owing to finite-sample bias and unaddressed confounding bias, respectively. AMRIV-MS provides partial correction but still falls short of nominal coverage, consistent with the requirement that $\delta(X)$ be estimated consistently for asymptotic validity.

6.7.2 Simulation Studies with Semi-Synthetic Data

We also evaluate AMRIV on a semi-synthetic dataset based on the TripAdvisor customer simulator from Syrgkanis et al. [2019], where we use customer features as covariates X, a simulated signup prompt as the instrument Z, and subscription revenue as the outcome Y. The DGP and oracle nuisances are described in Appendix E.8.2. Results are consistent with the synthetic setting: adaptive instrument assignment improves efficiency, AMRIV achieves superior coverage

and consistency, and robustness holds under partial misspecification.

Overall, these findings confirm that pairing adaptive design with the AMRIV estimator improves efficiency, enhances robustness, and yields more reliable inference than non-adaptive baselines.

6.8 Conclusion

We develop AMRIV, an adaptive, multiply robust estimator for ATE estimation in experiments where treatment can only be encouraged via a binary instrument. Our approach (i) derives the semiparametric efficiency bound and optimal assignment policy, (ii) constructs a sequential estimator that attains the bound under adaptive allocation, (iii) provides asymptotic normality, convergence rates, and multiply robust consistency, and (iv) supports valid inference through time-uniform confidence sequences. Empirical results on synthetic and semi-synthetic data confirm that adaptive instrument assignment improves both efficiency and robustness over non-adaptive baselines. This work represents a step toward principled, data-efficient experimentation in real-world settings where compliance is optional and uncertainty is unavoidable. We discuss limitations and broader impacts in Appendix E.9.

Part III

Causal Inference in Structured Data: Spatiotemporal and Network Dependence

CHAPTER 7

GST-UNET: A NEURAL FRAMEWORK FOR SPATIOTEMPORAL CAUSAL INFERENCE WITH TIME-VARYING CONFOUNDING

This chapter is based on Oprescu et al. [2025].

Estimating causal effects from spatiotemporal observational data is essential in public health, environmental science, and policy evaluation, where randomized experiments are often infeasible. Existing approaches, however, either rely on strong structural assumptions or fail to handle key challenges such as interference, spatial confounding, temporal carryover, and *time-varying confounding*—where covariates are influenced by past treatments and, in turn, affect future ones. We introduce the **GST-UNet** (**G**-computation **S**patio-**T**emporal **UNet**), a theoretically grounded neural framework that combines a U-Net-based spatiotemporal encoder with regression-based iterative G-computation to estimate location-specific potential outcomes under complex intervention sequences. GST-UNet explicitly adjusts for time-varying confounders and captures non-linear spatial and temporal dependencies, enabling valid causal inference from a *single* observed trajectory in data-scarce settings. We validate its effectiveness in synthetic experiments and in a real-world analysis of wildfire smoke exposure and respiratory hospitalizations during the 2018 California Camp Fire. Together, these results position GST-UNet as a **principled and ready-to-use framework** for spatiotemporal causal inference, advancing reliable estimation in policy-relevant and scientific domains.

7.1 Introduction

Environmental hazards, public health interventions, and socio-economic policies often require understanding complex cause-and-effect relationships across

space and time [Reid et al., 2016b, Papadogeorgou et al., 2019b, Song et al., 2020]. For instance, evaluating the health impacts of air quality regulations requires assessing how interventions influence both immediate outcomes and downstream effects across regions. Such applications demand robust tools for estimating causal effects from observational spatiotemporal data.

However, causal inference in spatiotemporal settings poses unique challenges. Outcomes are influenced not only by local covariates and interventions but also by those of neighboring regions (spatial confounding and interference). Effects may persist and accumulate over time (temporal carryover), and covariates often evolve in response to past interventions while simultaneously affecting future ones (time-varying confounding). For example, air quality regulations are often implemented in reaction to recent pollution levels and hospitalizations, which themselves shape future exposures and health outcomes—creating feedback loops that violate standard independence assumptions. These complexities induce bias in naive estimators and are especially challenging in single-trajectory settings, where replication across units or time is infeasible.

Existing approaches offer limited solutions: classical methods rely on rigid structural assumptions or user-defined exposure mappings, while recent neural models emphasize predictive accuracy over causal identification. Many assume independent time series or model only spatial correlations, leaving a gap in methods that can jointly address interference, temporal dependencies, and evolving confounding within a principled causal framework (see Section 7.2).

To bridge this gap, we introduce the **GST-UNet** (**G**-computation **S**patio-**T**emporal **UNet**), a theoretically grounded neural framework for estimating location-specific potential outcomes in spatiotemporal settings with time-varying confounding. GST-UNet builds on formal identification and consis-

tency results derived under a representation-based time-invariance assumption, showing how causal effects can be recovered from a single observed trajectory. We then instantiate this theory in a practical neural architecture: a U-Net encoder with ConvLSTM and attention modules coupled to an iterative G-computation procedure that performs recursive causal adjustment over time. To ensure stable estimation over long horizons, we design a curriculum-based training strategy that gradually refines recursive pseudo-outcomes, enabling effective learning even in data-scarce regimes. Unlike existing approaches, GST-UNet requires no user-specified structural models and can be directly deployed in real-world spatiotemporal applications.

Our contributions are threefold: (1) We develop the first unified framework that couples theoretical identification and consistency guarantees with an end-to-end neural implementation for spatiotemporal causal inference; (2) We demonstrate through controlled simulations that GST-UNet robustly handles interference, temporal carryover, and time-varying confounding; and (3) We illustrate its practical value via a real-world analysis of wildfire smoke exposure and respiratory hospitalizations during the 2018 California Camp Fire.

In summary, GST-UNet provides a **principled and ready-to-use framework** for causal inference from spatiotemporal data, combining formal guarantees with a flexible neural implementation. By abstracting away model-specific assumptions, GST-UNet makes spatiotemporal causal estimation both **accessible and reliable** for applied scientific and policy domains.

7.2 Related Work

We summarize the most relevant prior work here, with a more detailed discussion in Appendix F.1.

Classical Spatiotemporal Causal Inference Early approaches to spatiotemporal causal inference (e.g., spatial econometrics [Anselin, 2013], difference-in-differences [Keele and Titiunik, 2015], synthetic controls [Ben-Michael et al., 2022]) rely on strong assumptions such as parallel trends and no interference. More recent methods incorporate time-varying confounding using inverse propensity weighting (IPW) and marginal structural models [Papadogeorgou et al., 2022, Zhou et al., 2024], but cannot address interference unless via user-specified exposure mappings or hyper-local assumptions [Wang, 2021, Christiansen et al., 2022, Zhang and Ning, 2023]. As noted by Zhou et al. [2024], the literature remains sparse, especially in settings with rich feedback dynamics.

Machine Learning for Spatiotemporal Modeling Deep learning models for prediction–e.g., CNNs and RNNs [Shi et al., 2015, Zhang et al., 2017], graph-based methods [Li et al., 2018, Wu et al., 2019], and video transformers [Bertasius et al., 2021, Liu et al., 2022]–capture complex spatial-temporal patterns but do not incorporate causal adjustments, and thus cannot estimate counterfactuals or adjust for time-varying confounders.

Longitudinal Causal Inference Causal methods for longitudinal data include marginal structural models [Robins et al., 2000], iterative G-computation [Robins and Hernan, 2008], and recent ML-based extensions using recurrent networks, transformers, or meta-learners [Bica et al., 2020b, Seedat et al., 2022, Melnychuk et al., 2022, Li et al., 2021, Frauen et al., 2025, Hess et al., 2024]. However, these assume access to independent time series (*e.g.*across patients) and cannot model cross-unit interactions in spatiotemporal settings.

Neural-Based Spatiotemporal Causal Inference Tec et al. [2023] propose a UNet-based model that adjusts for non-local spatial confounding but focuses on static exposures and does not address interference or time-varying effects.

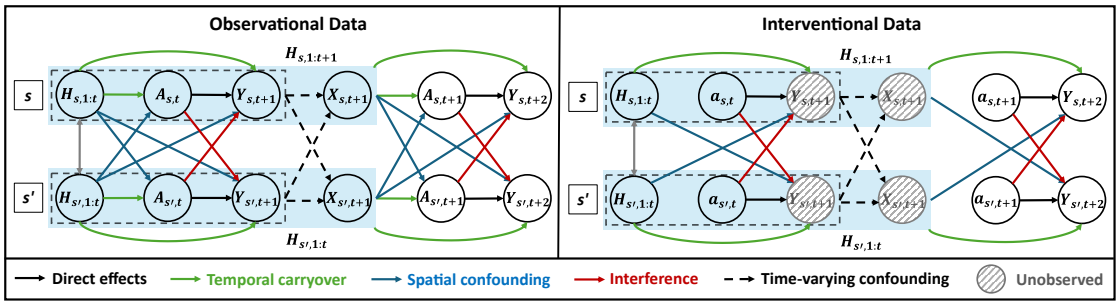


Figure 7.1: Causal structure of observational data (left) versus interventional data (right) for a spatiotemporal horizon $\tau = 2$ across two locations (s, s'). **Green** arrows indicate temporal carryover, **blue** arrows show spatial confounding, and **red** arrows depict interference; dashed arrows denote time-varying confounding, and dashed circles represent unobserved variables at inference time. Under the intervention (right), treatments are set independently of confounders, and the full history is not observed for the entire horizon.

Most similar to our work, [Ali et al., 2024] presents a climate-focused model that shares certain architectural similarities but emphasizes prediction rather than causal adjustment, leaving causal identification under time-varying confounding largely unaddressed.

Positioning of Our Work Our work bridges these threads by uniting a theoretically grounded G-computation framework with a neural architecture for spatiotemporal data. Unlike prior time-series methods that assume independent units or spatial models that overlook confounding feedback, GST-UNet is the first end-to-end approach that (i) establishes identification and consistency under explicit assumptions for a *single* spatiotemporal trajectory, and (ii) implements this theory in a practical neural model capable of handling interference, spatial confounding, and time-varying dynamics.

7.3 Background and Setup

Spatiotemporal Data We model observed data as random variables on a discrete spatial domain represented by an $N_X \times N_Y$ lattice: $\mathcal{S} = \{(i, j) \mid i \in [N_X], j \in$

$[N_Y]\}$, where $[N] = \{1, \ldots, N\}$ denotes the index set. Time is indexed by $t \in [T]$. At each spatial location $s = (i, j)$ at time t, we observe a tuple $(\mathbf{X}_{s,t}, A_{s,t}, Y_{s,t})$, where $A_{s,t} \in \{0, 1\}$ represents a binary treatment (or intervention), $Y_{s,t} \in \mathbb{R}$ is a continuous outcome of interest, and $\mathbf{X}_{s,t} \in \mathbb{R}^{d_X}$ is a vector of time-varying covariates (*e.g.* local weather conditions, pollution levels, or socioeconomic indicators). Additionally, each location s is associated with static features $V_s \in \mathbb{R}^{d_V}$ (*e.g.*geographical characteristics and socioeconomic indicators). While we focus on binary interventions for clarity, the methods generalize to more complex treatments. Conceptually, each variable forms a 3D spatiotemporal tensor of size $T \times N_X \times N_Y$, though in practice observations may be incomplete. Missing data can be accommodated using masking during downstream modeling.

To streamline notation, we use boldface symbols for random variables defined over the entire spatial domain. For $U \in \{X, A, Y\}$, let $\mathbf{U}_t$ denote its value at time t, and let $\mathbf{U}_{t:t+\tau} = (\mathbf{U}_t, \ldots, \mathbf{U}_{t+\tau})$ denote its value over a time interval. For a specific location s, we write $U_{s,t:t+\tau} = (U_{s,t}, \ldots, U_{s,t+\tau})$. The history up to time t is denoted by $\mathbf{H}_{1:t} = (\mathbf{X}_{1:t}, \mathbf{A}_{1:t-1}, \mathbf{Y}_{1:t}, \mathbf{V})$ for the entire spatial domain and $H_{s,1:t} = (X_{s,1:t}, A_{s,1:t-1}, Y_{s,1:t}, V_s)$ for a specific location s. Specific instantiations of these random variables are denoted using lowercase letters (e.g., $u \in \{x, a, y, h\}$).

Quantities of Interest Our primary goal is to estimate location-specific Conditional Average Potential Outcomes (CAPOs) for a sequence of future spatiotemporal interventions, conditioned on observed history. Our approach builds on Rubin's potential outcomes framework [Rubin, 1978, Robins and Hernan, 2008, Robins et al., 2000], which we extend to accommodate spatiotemporal settings. More concretely, we consider a future time horizon of length $\tau \geq 1$ and a predetermined interventional sequence $\mathbf{a}_{t:t+\tau-1}$ applied across the spatial domain starting at time t. Our goal is to estimate the potential outcomes at time $t + \tau$,

denoted as $\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}]$. In particular, we aim to compute:

$$\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}] \tag{7.1}$$

which represents the CAPOs at time $t + \tau$ under the given treatment sequence. Given two different interventional sequences $\mathbf{a}_{t:t+\tau}$ and $\mathbf{a}'_{t:t+\tau}$, a related secondary goal is to estimate the location specific Conditional Average Treatment Effect (CATE), given by:

$$\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] - \mathbf{Y}_{t+\tau}[\mathbf{a}'_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}]$$

Although we focus primarily on CAPOs, CATEs and other effect measures can be derived similarly.

Prefix Data in a Single Spatiotemporal Chain The conditional expectations defining the CAPOs in Eq. (7.1) cannot be directly estimated from a single observed spatiotemporal realization, since the empirical averages would contain only one sample of each future outcome $\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}]$. To obtain a workable regression-based estimator, we therefore reorganize the single observed trajectory into overlapping "prefixes" of varying lengths. For each $t \in \{1, \ldots, T - \tau\}$, we define

$$\mathbf{P}_t^\tau \;=\; (\mathbf{X}_{1:t+\tau}, \mathbf{A}_{1:t+\tau}, \mathbf{Y}_{1:t+\tau}, \mathbf{V}),$$

which represents the observed history up to time $t + \tau$ along with all covariates, treatments, and outcomes. When $T \gg \tau$, this construction yields $T - \tau$ segments that partially overlap in time, providing additional training samples in this intrinsically data-scarce, single-chain setting.

However, these prefixes are *not* independent: successive segments share overlapping histories, so standard i.i.d. assumptions do not apply. In the next section, we introduce conditions under which these prefixes can be treated as

conditionally exchangeable given an appropriate learned embedding. This enables regression-based estimation of CAPOs by pooling information across time without violating the dependence structure of the original process.

7.4 Identification and Estimation of CAPOs in Spatiotemporal Settings

Identification of CAPOs from observational data relies on standard causal inference assumptions. In our setting, these must be complemented by additional structure to handle the fact that we observe only a *single* spatiotemporal trajectory. Building on the prefix construction introduced above, we impose conditions that render these overlapping segments *conditionally exchangeable*, enabling principled pooling of information across time.

Assumption 7.1 (Causal Inference Assumptions)**.** We assume: *(Consistency)* $\mathbf{Y}_{t+\tau} = \mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}]$ whenever the observed sequence of treatments $\mathbf{A}_{t:t+\tau-1}$ satisfies $\mathbf{A}_{t:t+\tau-1} = \mathbf{a}_{t:t+\tau-1}$; (*Positivity*) $P(A_{s,t} = a_{s,t} \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}) > 0$ for any $a_{s,t} \in \{0, 1\}$ and feasible realization of history $\mathbf{h}_{1:t}$; (*Sequential Unconfoundedness*) $\mathbf{Y}_{t+1:T}[\mathbf{a}_{t+1:T}] \perp \mathbf{A}_t \mid \mathbf{H}_{1:t}$, $\forall \mathbf{a}_{t+1:T} \in \{0, 1\}^{T-t}$, *i.e.*at each time step t, the treatment assignment is independent of future potential outcomes.

Assumption 7.2 (Representation-Based Time Invariance)**.** There exists a function (or embedding) $\phi : \mathcal{H} \times \mathcal{A} \to \mathcal{Z} \subseteq \mathbb{R}^h$ that maps $(\mathbf{H}_{1:t}, \mathbf{A}_t)$ to a finite-dimensional representation such that once we condition on $z = \phi(\mathbf{H}_{1:t}, \mathbf{A}_t)$, the distribution $(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1})$ does not explicitly depend on t. Formally, for any $t, t' \in \{1, \ldots, T\}$ and $z \in \mathcal{Z}$, we have:

$$p(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1} \mid \phi(\mathbf{H}_{1:t}, \mathbf{A}_t) = z) = p(\mathbf{X}_{t'+1}, \mathbf{Y}_{t'+1} \mid \phi(\mathbf{H}_{1:t'}, \mathbf{A}_{t'}) = z).$$

Assumption 7.1 is a standard set of requirements in longitudinal causal inference settings (e.g., [Robins et al., 2000, Robins and Hernan, 2008, Bica et al., 2020b, Li et al., 2021, Melnychuk et al., 2022, Hess et al., 2024]). Assumption 7.2 is specific to the single-time series setting, where pooling information across time is essential to enable estimation. We note that the single time-series setting frequently arises in causal inference, where assumptions such as stationarity or strict time homogeneity enable consistent estimation [Bojinov and Shephard, 2019, Papadogeorgou et al., 2022, Zhou et al., 2024]. In contrast, our representation-based time invariance is *weaker*: rather than requiring $\mathbf{X}_t, \mathbf{Y}_t$ themselves to have a time-invariant distribution, we only assume that, once the history is summarized by $\phi(\mathbf{H}_{1:t}, \mathbf{A}_t)$, the transition to $(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1})$ follows a single shared mechanism. This approach aligns with modern time-series causal inference that learn time-invariant latent embeddings to pool information across time steps [Lim, 2018, Li et al., 2021, Hess et al., 2024], thus leveraging more data for a single, stable representation rather than time-dependent parameters.

Under Assumption 7.2, conditioning on $\phi(\mathbf{H}_{1:t}, \mathbf{A}_t)$ removes explicit dependence on t, such that

$$\mathbb{E}_{\mathbf{P}}[\mathbf{Y}_{t+\tau} \mid \phi(\mathbf{H}_{1:t}, \mathbf{A}_t)]$$

represents a shared conditional expectation across all prefix segments. In this view, t indexes the segment's position rather than a distinct distribution. Pooling over t thus yields $T - \tau$ approximately exchangeable segments from a single trajectory, enabling regression-based estimation of future outcomes from embedded histories.

7.4.1 Identification via Representation-Based G-Computation

Given $\mathbf{P}_t^\tau$, we next show how to identify CAPOs from observational data. For horizons $\tau \geq 2$, *future* covariates and outcomes (*i.e.*$\mathbf{X}_{t+1:t+\tau-1}, \mathbf{Y}_{t+1:t+\tau-1}$) can influence subsequent treatments, inducing time-varying confounding [Coston et al., 2020]. Such feedback violates standard "condition-on-history" adjustments and leads to biased estimates. Figure 7.1 illustrates these dependencies by contrasting observational data (left) and hypothetical interventions (right) for $\tau = 2$. By contrast, when $\tau = 1$, conditioning on $\mathbf{H}_{1:t}$ is sufficient under standard assumptions, as no future confounders intervene between $\mathbf{A}_t$ and $\mathbf{Y}_{t+1}$. Formally, the following naive identification fails to hold for $\tau > 1$:

$$\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}] \neq \mathbb{E}[\mathbf{Y}_{t+\tau} \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_{t:t+\tau-1} = \mathbf{a}_{t:t+\tau-1}] \tag{7.2}$$

To correct this bias, we adapt *regression-based iterative G-computation* [Bang and Robins, 2005, Robins and Hernan, 2008] to the spatiotemporal setting, yielding a principled adjustment procedure for evolving confounders and valid CAPO estimation. We formalize this connection in the following result:

Theorem 7.3 (Identification with G-Computation)**.** *Assume that Assumption 7.1 and Assumption 7.2 hold. Further, let* $\mathbf{H}^{\mathbf{a}}_{1:t+k} := (\mathbf{X}_{1:t+k}, [\mathbf{A}_{1:t-1}, \mathbf{a}_{t:t+k-1}], \mathbf{Y}_{1:t+k})$ *denote the history where observed treatments from time t onward are replaced by* $\mathbf{a}_{t:t+k-1}$*. Define recursively:*

$$Q_\tau(\mathbf{H}_{1:t+\tau-1}, \mathbf{A}_{t+\tau-1}) = \mathbb{E}_{\mathbf{P}}[\mathbf{Y}_{t+\tau} \mid \phi(\mathbf{H}_{1:t+\tau-1}, \mathbf{A}_{t+\tau-1})]$$

$$Q_{\tau-1}(\mathbf{H}_{1:t+\tau-2}, \mathbf{A}_{t+\tau-2}) = \mathbb{E}_{\mathbf{P}}[Q_\tau(\mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1}) \mid \phi(\mathbf{H}_{1:t+\tau-2}, \mathbf{A}_{t+\tau-2})]$$

$$\dots$$

$$Q_1(\mathbf{H}_{1:t}, \mathbf{A}_t) = \mathbb{E}_{\mathbf{P}}[Q_2(\mathbf{H}^{\mathbf{a}}_{1:t+1}, \mathbf{a}_{t+1}) \mid \phi(\mathbf{H}_{1:t}, \mathbf{A}_t)]$$

Then $\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}] = Q_1(\mathbf{h}_{1:t}, \mathbf{a}_t)$.

We provide a proof of Theorem 7.3 in Appendix F.2. This result naturally motivates a recursive regression approach for spatiotemporal CAPO estimation, fitting each $Q_k(\cdot)$ in reverse order and substituting interventional treatments where required.

7.4.2 Estimation via Iterative G-Computation

While Theorem 7.3 motivates a recursive regression algorithm for each Q_k ($k = 1, \ldots, \tau$), only Q_τ can be directly estimated from the prefix data. At the next step, $Q_{\tau-1}$ depends on $Q_\tau(\mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1})$—where the observed treatments $\mathbf{A}_{t:t+\tau-1}$ are replaced by $\mathbf{a}_{t:t+\tau-1}$—but such substituted outcomes are not observed in the prefix data. Therefore, for $k < \tau$, we propose a procedure where we generate *pseudo-outcomes* by predicting with the previously learned $\widehat{Q}_{k+1}$. Going forward, we use $\widehat{F}$ to denote any quantity F estimated from data. Formally, let $\phi \in \Phi$ be an embedding satisfying Assumption 7.2, and let $\mathcal{Q}$ be our function class for Q_k. We learn the sequence $\widehat{Q}_\tau, \ldots, \widehat{Q}_1$ from prefix data $\{\mathbf{P}^\tau_t : t = 1, \ldots, T - \tau\}$, via:

1. **Initialization.** Fit $\widehat{Q}_\tau$ by predicting $\mathbf{Y}_{t+\tau}$ from the prefix embedding $\phi(\mathbf{H}_{1:t+\tau-1}, \mathbf{A}_{t+\tau-1})$.
2. **Backward recursion.** For $k = \tau - 1, \ldots, 1$:
   (a) *Substitute interventions.* For each prefix $\mathbf{P}^\tau_t$, replace $\mathbf{A}_{t+k}$ by the interventional $\mathbf{a}_{t+k}$ to form the modified history $\mathbf{H}^{\mathbf{a}}_{1:t+k}$.
   (b) *Generate pseudo-outcomes.* Let $\widetilde{Y}_{t+k+1} = \widehat{Q}_{k+1}(\mathbf{H}^{\mathbf{a}}_{1:t+k}, \mathbf{a}_{t+k})$, where $\widehat{Q}_{k+1}$ was learned in the previous step. These $\widetilde{Y}_{t+k+1}$ act as surrogates for $\mathbf{Y}_{t+k+1}$ in the prefix data.
   (c) *Fit $\widehat{Q}_k$.* Regress $\widetilde{Y}_{t+k+1}$ on the current embedding $\phi(\mathbf{H}_{1:t+k-1}, \mathbf{A}_{t+k-1})$ to learn $\widehat{Q}_k \in \mathcal{Q}$.
3. **Final step.** Given a new history $\mathbf{h}_{1:t}$ and an interventional path $\mathbf{a}_{t:t+\tau-1}$, we

predict

$$\widehat{\mathbb{E}}_{\mathbf{P}}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \phi(\mathbf{H}_{1:t}, \mathbf{a}_t) = \phi(\mathbf{h}_{1:t}, \mathbf{a}_t)] = \widehat{Q}_1(\mathbf{h}_{1:t},\ \mathbf{a}_t).$$

The iterative regression procedure yields consistent CAPO estimates provided each stage Q_k is estimated consistently from data [Laan and Robins, 2003]. Informally, if the learned embedding $\widehat{\phi}$ converges to the true time-invariant representation ϕ, and small perturbations in ϕ or $\widehat{Q}_k$ lead to proportionally small changes in predictions, then the overall recursive estimator remains consistent. These regularity conditions—formalized through uniform stochastic equicontinuity—are detailed in Appendix F.3. Formally, we state the following theorem:

Theorem 7.4 (Consistency of Iterative G-Computation in Spatiotemporal Settings)**.** *Assume Assumptions 7.1 and 7.2 and that (a) the learned embedding $\widehat{\phi}$ is L_2-consistent for ϕ, and (b) each regression head $\widehat{Q}_k$ consistently estimates Q_k and is uniformly well-behaved*[1] *on* $\operatorname{Im}\phi$ *(intuitively, small input perturbations induce small output changes). Let $\mathbf{Z}_k := (\mathbf{H}_{1:t+k}, \mathbf{A}_{t+k})$ denote the history–action pair at step k. Then*

$$\left\|\widehat{Q}_1(\mathbf{Z}_0; \widehat{\phi}) - Q_1(\mathbf{Z}_0; \phi)\right\|_2 = o_p(1),$$

so the recursive estimator $\widehat{Q}_1$ of the CAPO is probabilistically consistent.

We provide a proof of Theorem 7.4 in Appendix F.3. In the following section, we instantiate this procedure in our **GST-UNet** architecture, illustrating how to incorporate spatial dependencies and interference into ϕ and each Q_k, and implement a streamlined, end-to-end training strategy that unifies history embeddings and outcome predictions.

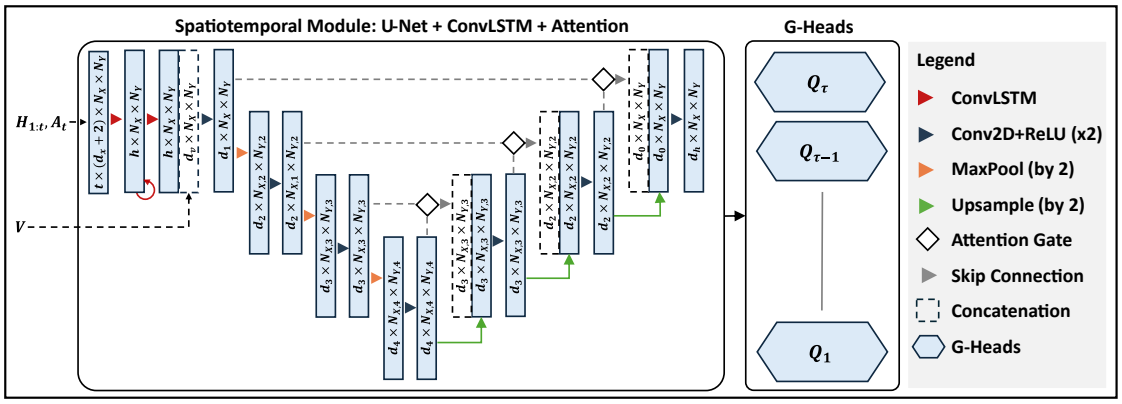


Figure 7.2: Overview of the GST-UNet architecture for spatiotemporal causal inference. The spatiotemporal learning module (left) is a U-Net augmented with a ConvLSTM layer and attention gates. Its final feature map is passed to a set of *G-heads* (right), where each G-head Q_k implements iterative G-computation (see Algorithm 7.1).

7.5 GST-UNet Implementation

The theoretical results above establish how CAPOs can be identified and consistently estimated from a single spatiotemporal trajectory. We now provide a concrete neural implementation of this procedure. **GST-UNet** instantiates the iterative G-computation framework with a spatiotemporal deep architecture that embeds strong inductive biases—locality, translation invariance, and temporal smoothness—well suited to data-scarce settings. While alternative backbones could be employed, our U-Net with ConvLSTM and attention offers a natural choice for learning stable, history-invariant representations that satisfy Assumption 7.2. We now describe the architecture and the training procedure that realizes the GST-UNet (Algorithm 7.1).

7.5.1 Model Architecture

The **GST-UNet** consists of two main components:

[1]We formalize "well-behaved" via uniform stochastic equicontinuity and continuity in Appendix F.3.

1. **Spatiotemporal Learning Module:** a U-Net-based network augmented with ConvLSTM and attention gates for spatiotemporal processing.
2. **Neural Causal Module:** τ G-computation heads, each mapping the spatiotemporal features to final outcome predictions in the iterative procedure.

We illustrate the GST-UNet architecture in Figure 7.2 and describe its main components below.

Spatiotemporal Learning Module While our framework is agnostic to the choice of spatiotemporal learning module, we use a U-Net with ConvLSTM and attention because it performs well in data-scarce regimes.

Spatial module. To efficiently process high-dimensional spatial data, we use U-Net [Ronneberger et al., 2015a], a fully convolutional architecture originally developed for biomedical image segmentation. U-Net follows an encoder–decoder design with skip connections. The encoder progressively downsamples the spatial grid through convolution and pooling, while the decoder upsamples back to the original resolution and merges encoder features at each scale.

Temporal module. A standard U-Net is limited in its ability to capture temporal structure. To address this, we integrate a Convolutional Long Short-Term Memory (ConvLSTM) layer [Shi et al., 2015] into the encoder. This layer maintains a hidden state across time steps while aggregating spatial information through convolutions.

After computing the final ConvLSTM state, we append the static (time-invariant) covariates $\mathbf{V}$ as additional feature channels. This gives the subsequent encoder–decoder direct access to both temporal dynamics and static location-specific information. In the decoder, we incorporate attention gates [Oktay et al., 2018] to selectively emphasize relevant spatial regions and refine the skip

connections. The resulting embedding module produces a d_h-dimensional feature map of size $N_X \times N_Y$, capturing essential spatiotemporal context—including interference, spatial confounding, and static covariates—for downstream G-computation.

Neural Causal Module We attach τ *G-computation heads* to the U-Net's final feature maps, corresponding to the Q_k estimators in the iterative procedure (see Section 7.4.2). Each head can be a small convolutional module or a simple feed-forward network, depending on how much spatial structure remains to be captured. The information flow at the G-computation heads proceeds as follows: each head Q_k ($k = 1, \ldots, \tau$) receives the $d_h \times N_X \times N_Y$ U-Net embedding $\widehat{\phi}(\mathbf{H}_{1:t+k-1}, \mathbf{A}_{t+k-1})$ (encompassing spatiotemporal and static context) and outputs an $N_X \times N_Y$ prediction for that time step. We refer to this as the *supervision step*, since Q_τ compares its predictions to the *real* observed outcomes $\mathbf{Y}_{t+\tau}$, anchoring the model in genuine data, while each $Q_{k<\tau}$ compares its predictions to pseudo-outcomes $\widetilde{\mathbf{Y}}_{t+k+1}$ provided by $\widehat{Q}_{k+1}$. These pseudo-outcomes arise in a subsequent *generation step*, wherein Q_{k+1} processes the intervened history $\widehat{\phi}(\mathbf{H}^{\mathbf{a}}_{1:t+k}, \mathbf{a}_{t+k})$ in a *detached* forward pass (so $\widehat{Q}_{k+1}$ is not updated by Q_k's loss), thereby creating surrogate targets for Q_k. This procedure realizes the iterative G-computation logic from Section 7.4.2, enabling GST-UNet to estimate future outcomes under various counterfactual treatments. By separating the spatiotemporal embedding from the G-heads, we maintain a common representation for all prefix data (see Assumption 7.2) and flexibly capture interference and spatial confounding. Each G-head enforces the proper temporal adjustments to yield bias-free counterfactual inference.

Algorithm 7.1 GST-UNet Training and Inference

Input: Horizon τ, prefixes $\{\mathbf{P}_t^\tau\}_{t=1}^{T-\tau}$, interventions $\mathbf{a}_{t:t+\tau-1}$, curriculum $\alpha_k^{(e)}$, total epochs E.
Initialize: parameters θ (U-Net embedding + G-heads).
for $e = 1 \dots E$ **do**
  for $k = \tau \dots 1$ **do**
    (Supervision) For each prefix i, predict the head-specific outcomes $\widehat{Y}_{t+k}^{(i)} = Q_k(\phi(\mathbf{H}_{1:t+k-1}^{(i)}, \mathbf{A}_{t+k-1}^{(i)}); \theta)$.
    (Generation, detached) For each prefix i, generate pseudo-outcomes

$$\widetilde{Y}_{t+k+1}^{(i)} = \begin{cases} Q_{k+1}\Big(\phi((\mathbf{H}^{\mathbf{a}}_{1:t+k})^{(i)}, \mathbf{a}_{t+k}^{(i)}); \theta\Big), & k < \tau, \\ Y_{t+\tau}^{(i)}, & k = \tau, \end{cases}$$

    where the observed $\mathbf{A}_{t:t+k-1}$ are replaced with $\mathbf{a}_{t:t+k-1}$ in $\mathbf{H}^{\mathbf{a}}_{1:t+k}$.
  end for
  (Loss aggregation) Compute the MSE loss

$$\mathcal{L}(\theta; e) = \frac{1}{\tau} \sum_{k=1}^{\tau} \alpha_k^{(e)} \sum_i (\widehat{Y}_{t+k}^{(i)} - \widetilde{Y}_{t+k+1}^{(i)})^2.$$

  (Backward pass) Update θ by backpropagation.
end for
(Inference) Given a $\mathbf{h}_{1:t}$, return $Q_1(\phi(\mathbf{h}_{1:t}, \mathbf{a}_t); \widehat{\theta})$.

7.5.2 Training and Inference

While each G-head Q_k could be trained sequentially–from Q_τ down to Q_1–by passing pseudo-outcomes backward through time, this creates a conflict when all heads share the same U-Net embedding ϕ. Specifically, each Q_k may push ϕ toward optimizing its own objective, resulting in misaligned training signals and unstable learning.

Joint Loss and Multi-Task Training To address this issue, we employ a *joint* (or *multi-task*) training approach [Caruana, 1997, Evgeniou and Pontil, 2004] by aggregating the loss terms from all G-heads into a single objective, then back-propagating once per batch. Concretely, for each head Q_k, let $\widetilde{Y}_{t+k+1}$ be the *real* outcomes if $k = \tau$ or *pseudo-outcomes* (generated by $\widehat{Q}_{k+1}$) if $k < \tau$. Our head-

specific loss is a mean squared error (MSE) over all prefix samples:

$$\mathcal{L}_k(\theta) \;=\; \sum_{i=1}^{T-\tau} \Big[Q_k(\phi(\mathbf{H}^{(i)}_{1:t+k-1}, \mathbf{A}^{(i)}_{t+k-1}); \theta) - \widetilde{Y}^{(i)}_{t+k+1} \Big]^2,$$

where θ encompasses *all* model parameters (the shared U-Net embedding ϕ and the G-heads Q_k).

Let $\alpha_k^{(e)}$ denote a *head-weight* for epoch e. We then form the overall training objective at epoch e by

$$\mathcal{L}(\theta; e) \;=\; \frac{1}{\tau} \sum_{k=1}^{\tau} \alpha_k^{(e)} \; \mathcal{L}_k(\theta). \tag{7.3}$$

By summing the losses and performing a single backward pass, we learn a common embedding $\widehat{\phi}$ that balances the needs of all G-heads, rather than fitting each head separately.

Curriculum Training A naive implementation of Eq. (7.3)–where each G-head is given equal weight–can be suboptimal: early in training, Q_τ (which sees real data) is inaccurate, and the pseudo-outcomes generated for $Q_{k<\tau}$ are effectively noise. Consequently, $Q_1, \ldots, Q_{\tau-1}$ may overfit to poor targets before Q_τ has converged, leading to suboptimal solutions. To mitigate this, we employ a *curriculum* training approach [Bengio et al., 2009], gradually increasing the loss weight of earlier heads as Q_τ improves.

While many curricula are possible, we adopt a simple scheme controlled by a single hyperparameter e_c (the "curriculum period") so we can readily tune it. Let $p(e) = \min\{\tau, \lceil e/e_c \rceil\}$, which indexes a "phase" based on the current epoch e. We then define

$$\alpha_k^{(e)} \;=\; \begin{cases} 1/p(e), & \text{if } k \in \{\tau,\, \tau-1,\, \ldots,\, \tau-p(e)+1\}, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, during epochs $1 \le e \le e_c$ (phase $p(e) = 1$), only Q_τ is active with $\alpha_\tau^{(e)} = 1$; in the next interval $e_c < e \le 2e_c$ (phase $p(e) = 2$), Q_τ and $Q_{\tau-1}$ each have

weight $1/2$, and so on until all heads are active with uniform weight $1/\tau$. For $e > \tau e_c$, training continues with $\alpha_k^{(e)} = 1/\tau$ for all heads. This schedule ensures Q_τ becomes reasonably accurate before earlier heads rely on its pseudo-outcomes. The hyperparameter e_c controls the pacing, helping prevent early training noise.

We also adopt standard neural network practices, including mini-batch optimization and early stopping, to stabilize training and mitigate overfitting. At *inference* time, given a new history $\mathbf{h}_{1:t}$ and an interventional sequence $\mathbf{a}_{t:t+\tau-1}$, we compute $\widehat{Q}_1(\phi(\mathbf{h}_{1:t}, \mathbf{a}_t); \theta)$ as our target CAPO estimate. We sketch the overall training and inference procedure in Algorithm 7.1.

7.6 Experiments

We evaluate the proposed GST-UNet framework through two applications. First, we simulate synthetic data that incorporates key spatiotemporal causal inference challenges: interference, spatial confounding, temporal carryover, and time-varying confounding. Using this synthetic data generation process (DGP), we compare the GST-UNet algorithm against several baselines. Next, we demonstrate the utility of GST-UNet on a real-world dataset analyzing the impact of wildfire smoke on respiratory hospitalizations during the 2018 California Camp Fire. Additional details—including exact simulation parameters, model architecture and execution setups, hyperparameter selection strategies, and validation procedures—can be found in Appendix F.4. Replication code is available at `https://github.com/moprescu/GSTUNet`.

7.6.1 Synthetic Data

We generate $T = 200$ time steps of a 64×64 ($N_X \times N_Y$) grid of observational data using the following data generating process (DGP):

$$\mathbf{X}_t = \alpha_0 + \alpha_1 \mathbf{X}_{t-1} + \alpha_2 \mathbf{A}_{t-1} + \alpha_3 (K_X * \mathbf{X}_{t-1}) + \epsilon_X,$$

$$\mathbf{A}_t \sim \text{Bern}\Big(\sigma(\beta_1(\beta_0 + \tfrac{1}{L}\sum_{l=0}^{L-1} K_A * \mathbf{X}_{t-l}))\Big),$$

$$\mathbf{Y}_t = \gamma_0 + \gamma_1 (K_{YA} * \mathbf{A}_{t-1}) + \gamma_2 \tfrac{1}{L}\sum_{l=1}^{L} (K_{YX} * \mathbf{X}_{t-l}) + \gamma_3 \mathbf{Y}_{t-1} + \epsilon_Y,$$

where $d_X = 1$, "$*$" denotes a 3×3 spatial convolution over the $N_X \times N_Y$ grid, and $\epsilon_X, \epsilon_Y \sim \mathcal{N}(0, 1)$ are i.i.d. noise. Each kernel K_X, K_A, K_{YA}, K_{YX} encodes a local advection–diffusion process that mimics wind-driven pollutant transport, with interventions $\mathbf{A}_t$ injecting additional emissions that propagate through the same kernel. This physically realistic setup produces **interference**, **spatial confounding**, and **temporal carryover**—the three challenges GST-UNet is designed to address. Each equation is evaluated at every spatial location, so $\mathbf{X}_t$, $\mathbf{A}_t$, and $\mathbf{Y}_t$ are $N_X \times N_Y$ matrices. Here, $\mathbf{X}_t$ acts as a *time-varying confounder*: its past influences both $\mathbf{A}_t$ and $\mathbf{Y}_t$, while current interventions $\mathbf{A}_t$ affect future $\mathbf{X}_{t+1}$. For example, $\mathbf{A}_t$ may represent regulatory actions, $\mathbf{X}_t$ air quality, and $\mathbf{Y}_t$ health outcomes—capturing feedback from policy to exposure, outcome, and back to future policy.

We vary β_1 to control time-varying confounding: when $\beta_1 = 0$, $\mathbf{X}_t$ does not affect $\mathbf{A}_t$, eliminating confounding; larger values increase its strength. For each β_1, we generate 50 test trajectories from random initial states, fix their histories, and simulate 100 τ-step counterfactual futures to estimate true CAPOs, with $\tau \in \{5, 10\}$. We compare GST-UNet against three baselines: (i) **UNet+**, which uses a U-Net + ConvLSTM + Attention backbone with A_t as an input channel but performs no iterative adjustment; (ii) **STCINet** [Ali et al., 2024], which estimates direct and indirect effects without modeling time-varying confound-

Table 7.1: RMSE ± standard deviation across test trajectories in the GST-UNet synthetic experiment. Columns correspond to different levels of time-varying confounding β_1, and rows compare GST-UNet, baselines, and ablations. Bold indicates the lowest error per column; color shows improvement (RMSE **decrease** or **increase**) over the best baseline, excluding ablations.

τ	Model	$\beta_1 = 0.0$	$\beta_1 = 0.5$	$\beta_1 = 1.0$	$\beta_1 = 1.5$	$\beta_1 = 2.0$
5	UNet+	**0.28 ± 0.00**	0.36 ± 0.00	0.54 ± 0.01	0.71 ± 0.01	0.81 ± 0.01
	STCINet	0.29 ± 0.00	0.38 ± 0.01	0.62 ± 0.01	0.80 ± 0.01	0.90 ± 0.01
	IPWUNet	0.60 ± 0.01	0.58 ± 0.01	0.58 ± 0.01	0.59 ± 0.01	0.59 ± 0.01
	GST-UNet w/o Attention	0.50 ± 0.00	0.46 ± 0.00	0.51 ± 0.00	0.45 ± 0.01	0.47 ± 0.01
	GST-UNet w/o Curriculum	0.69 ± 0.00	0.64 ± 0.00	0.63 ± 0.00	0.61 ± 0.01	0.61 ± 0.01
	GST-UNet	0.33 ± 0.00	**0.35 ± 0.00**	**0.40 ± 0.00**	**0.44 ± 0.00**	**0.40 ± 0.01**
		(+17.9%)	(-2.7%)	(-21.6%)	(-25.4%)	(-32.2%)
10	UNet+	**0.28 ± 0.00**	0.61 ± 0.00	1.18 ± 0.00	1.45 ± 0.00	1.71 ± 0.01
	STCINet	0.31 ± 0.00	0.68 ± 0.00	1.25 ± 0.00	1.47 ± 0.01	1.60 ± 0.01
	IPWUNet	0.78 ± 0.01	0.80 ± 0.01	0.96 ± 0.01	1.19 ± 0.02	1.08 ± 0.01
	GST-UNet w/o Attention	0.42 ± 0.00	0.60 ± 0.00	0.61 ± 0.00	0.79 ± 0.01	1.07 ± 0.01
	GST-UNet w/o Curriculum	0.62 ± 0.00	0.88 ± 0.00	1.02 ± 0.00	1.08 ± 0.01	1.12 ± 0.01
	GST-UNet	0.38 ± 0.00	**0.55 ± 0.00**	**0.68 ± 0.00**	**0.73 ± 0.01**	**0.85 ± 0.01**
		(+35.7%)	(-9.8%)	(-29.2%)	(-38.7%)	(-21.3%)

ing; and (iii) **IPWUNet**, an inverse-propensity-weighting variant that reweights pseudo-outcomes using a UNet-style propensity estimator but cannot correct for spatial interference (details in Appendix F.4). We also test ablations of GST-UNet without curriculum or attention.

Table 7.1 shows that when $\beta_1 = 0$, UNet+ performs best—G-computation is unnecessary and adds noise. As β_1 increases, UNet+ and STCINet degrade sharply, while GST-UNet remains stable. IPWUNet shows some benefit but is biased even at $\beta_1 = 0$ due to uncorrected interference. GST-UNet consistently outperforms all baselines, demonstrating the value of iterative G-computation.

Curriculum training substantially improves performance across horizons, while attention yields modest gains—consistent with our predominantly local dynamics. Additional ablation analyses, including neighbor aggregation experiments, are reported in Appendix F.4.

7.6.2 Impact of Wildfires on Respiratory Health

Wildfire smoke has been linked to short-term respiratory harms [Reid et al., 2016b,a, Cascio, 2018, Cleland et al., 2021, Letellier et al., 2025], with older adults especially vulnerable [DeFlorio-Barker et al., 2019]. At the time this work was conducted (January 2025), a series of 14 destructive wildfires affected the Los Angeles metropolitan area and San Diego County in California, underscoring the urgency of understanding the health impacts of such events. In this study, we focus on a previous large-scale episode: the 2018 California wildfire season [Wikipedia, 2025], which included the *Carr Fire* (July–August) and the *Camp Fire* (November) and significantly worsened air quality.

We use daily county-level data from Letellier et al. [2025] (see Appendix F.4.2), including $PM_{2.5}$, respiratory/cardiovascular hospitalizations, weather variables (temperature, precipitation, humidity, radiation, wind), and population estimates from the California Department of Finance. Each weather variable can be a *time-varying confounder*: weather conditions affect future smoke levels and health outcomes, while also being influenced by prior smoke levels.

We focus on weeks 20–48 (May 18–December 2, 2018), covering the Carr and Camp fires. Following standard practice, we label a county as "treated" on days with mean $PM_{2.5} > 10\ \mu g/m^3$ and use raw hospitalization counts (rather than per-10,000 incidence, which can be unstable for small counties). We interpolate daily county-level data (treatment, outcome, five covariates) onto a

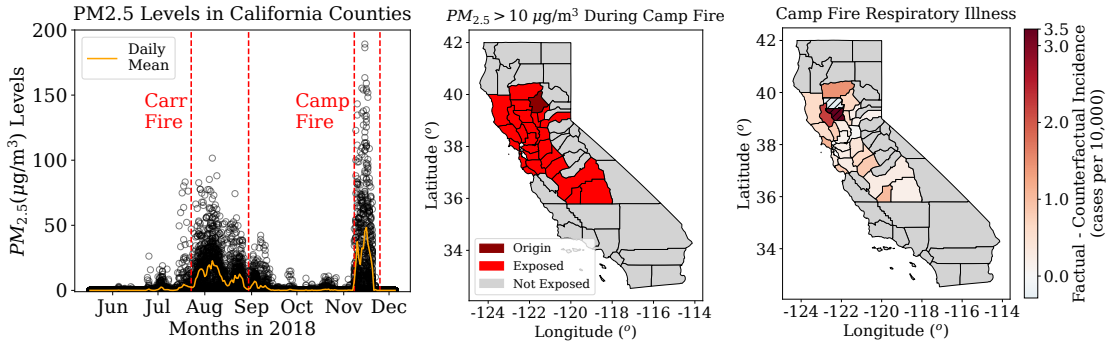


Figure 7.3: Wildfire-smoke application in California during 2018. **(Left)** Daily county-level $PM_{2.5}$ levels across California from May to December 2018, with red lines marking the Carr and Camp fires. **(Center)** Counties exposed to average $PM_{2.5} > 10\,\mu g/m^3$ during the Camp Fire (red), origin county in dark red. **(Right)** Estimated increase in daily respiratory admissions during the Camp Fire, computed as factual minus GST-UNet CAPO-predicted counterfactual admissions under no wildfire-smoke event. Hashed areas indicate small-population counties (< 30,000).

40×44 latitude–longitude grid, discarding cells outside California, yielding a spatiotemporal tensor of size $203 \times 7 \times 40 \times 44$. Interpolation ensures each grid cell approximates the region it overlaps (area-weighted), enabling the model to capture spatial gradients in $PM_{2.5}$, weather, and hospitalizations. We train GST-UNet with horizon $\tau = 10$, using the Carr Fire period (June–July) for validation, and generate counterfactual predictions for the Camp Fire peak, November 8–17. See Appendix F.4.2 for preprocessing and masking details.

Figure 7.3 (left) shows the rise in PM2.5 during the mid-late 2018 wildfire season; (center) highlights counties with daily PM2.5 $> 10, \mu g/m^3$. Using GST-UNet, we estimate daily CAPOs had the Camp Fire not occurred (i.e., setting $PM_{2.5} \leq 10, \mu g/m^3$ statewide). Figure 7.3 (right) compares these to factual daily incidence (hospitalizations per 10,000 residents). To reduce small-sample variability, we exclude counties with population below 30,000 (vs. >70,000 for others), marking them with hatching (see Appendix F.4.2). Over November 8–17, GST-UNet predicts **approximately 4,650 excess respira-**

tory hospitalizations (465/day) attributable to the Camp Fire, with the highest incidence near the fire source. This aligns with a 95% bootstrap confidence interval of $[1888, 6535]$. UNet+ yields a lower mean and higher uncertainty (3,981; $[-899, 5202]$), STCINet produces highly variable near-zero estimates (88; $[-3077, 3281]$), and IPWUNet gives implausibly high, near-constant values (~20,500), reflecting limitations of weighting under rare-event support. These results underscore GST-UNet's improved stability and accuracy in counterfactual estimation. Our findings are qualitatively consistent with Letellier et al. [2025], who report 259 excess daily cases averaged over a longer, lower-intensity window (Nov 8-Dec 5). Overall, the GST-UNet captures spatiotemporal variation in smoke exposure and health outcomes, illustrating its promise for real-world causal inference in domains such as environmental health and policy.

7.7 Conclusion

We presented **GST-UNet**, a neural framework for spatiotemporal causal inference that combines U-Net–based representation learning with iterative G-computation to adjust for time-varying confounders. GST-UNet addresses key challenges such as interference, spatial confounding, temporal carryover, and time-varying feedback. We establish theoretical identification and consistency guarantees, validate performance in synthetic settings with controlled confounding, and demonstrate practical utility in estimating the impact of wildfire smoke exposure during the 2018 Camp Fire. Together, these results position GST-UNet as a **ready-to-use tool for practitioners**, offering reliable, interpretable causal estimates in complex spatiotemporal environments. We discuss limitations and broader impacts in Appendix F.5.

CHAPTER 8

SPATIAL DECONFOUNDER: INTERFERENCE-AWARE DECONFOUNDING FOR SPATIAL CAUSAL INFERENCE

This chapter is based on Khot et al. [2025], developed jointly with Ayush Khot and collaborators. My contributions focused on the core problem formulation, methodological setup, theoretical development, and experimental design. The chapter fits the dissertation's broader theme of reliable causal inference under unreliable assumptions.

Causal inference in spatial domains faces two intertwined challenges: (1) unmeasured spatial factors, such as weather, air pollution, or mobility, that confound treatment and outcome, and (2) interference from nearby treatments that violate standard no-interference assumptions. While existing methods typically address one by assuming away the other, we show they are deeply connected: *interference reveals structure* in the latent confounder. Leveraging this insight, we propose the **Spatial Deconfounder**, a two-stage method that reconstructs a substitute confounder from local treatment vectors using a conditional variational autoencoder (C-VAE) with a spatial prior, then estimates causal effects via a flexible outcome model. We show that this approach enables nonparametric identification of both direct and spillover effects under weak assumptions—without requiring multiple treatment types or a known model of the latent field. Empirically, we extend `SpaCE`, a benchmark suite for spatial confounding, to include treatment interference, and show that the Spatial Deconfounder consistently improves effect estimation across real-world datasets in environmental health and social science. By turning interference into a multi-cause signal, our framework bridges spatial and deconfounding literatures to advance robust causal inference in structured spatial data.

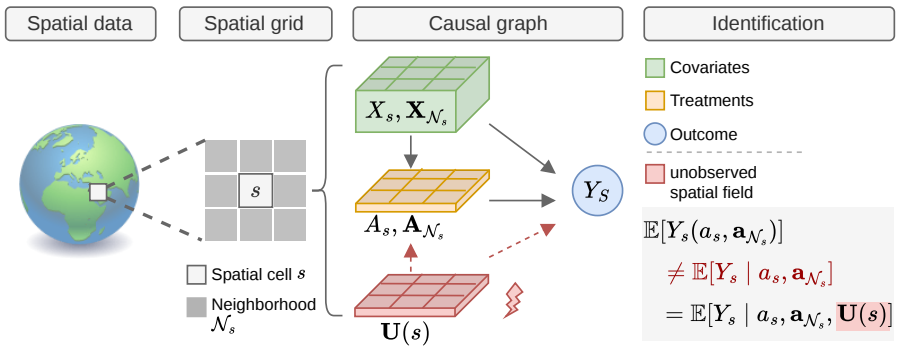


Figure 8.1: **Schematic of spatial interference/confounding**. Spatial data is represented in **geographical cells** indexed by site s with neighborhood $\mathcal{N}_s$. The **outcome** at s (e.g., mortality rate) is affected by the **treatments** (e.g., air quality) and **observed confounders** (e.g., demographic information) at both s and $\mathcal{N}_s$. However, **unobserved latent factors** (e.g., humidity) can confound the relationship, rendering causal effects unidentifiable.

8.1 Introduction

Causal inference in spatial settings is critical for science and policy, from estimating the health effects of pollution to evaluating land use, climate interventions, and the spread of infectious disease. Most data in these domains are observational, since large-scale interventions are typically infeasible or unethical, so robust methodology is needed to draw valid conclusions. Yet observational studies in these settings face two fundamental challenges that standard methods rarely address together: (1) *spillover (interference)*, where the treatment at one site affects outcomes at nearby sites, violating the Stable Unit Treatment Value Assumption (SUTVA), and (2) *spatially structured unobserved confounding*, where latent fields such as weather or socioeconomic context jointly drive treatment exposures and outcomes. Both are pervasive, and ignoring either leads to biased conclusions.

Consider air quality and health: respiratory mortality rates depend on local pollution and on neighboring regions' pollution due to transport and mobility,

while latent meteorological factors such as temperature and humidity confound both. Any method that neglects interference or hidden confounders risks misleading the decisions policy-makers rely on for regulation and public health.

Existing approaches for spatial causal inference typically address either interference *or* unobserved spatial confounding, but rarely both. (i) Methods developed for interference model spillovers through exposure mappings or spatial/auto-regressive dependencies, enabling estimation of direct and spillover effects under the assumption that all relevant confounders are observed [Hudgens and Halloran, 2008, Forastiere et al., 2021]. When important confounders are unobserved, these estimators can be biased. (ii) A separate line of work targets spatial treatment effect estimation under unobserved confounding using confounding-adjustment strategies such as splines, matching, or instrumental variables [Dupont et al., 2022, Papadogeorgou et al., 2019a, Papadogeorgou and Samanta, 2023]. These approaches typically rely on explicit structure—e.g., smooth latent-field priors, parametric outcome models, or exclusion restrictions—and often assume away interference or incorporate it only implicitly (e.g., via spatial trends), limiting their ability to identify and interpret spillover effects when interference is present and making results sensitive to model misspecification.

In an orthogonal literature stream, the *deconfounder* framework [Wang and Blei, 2019] shows that when each unit receives multiple causes, their joint distribution can reveal latent confounders. However, this method is designed for i.i.d. data with simultaneous treatments—not spatial domains with localized interactions. Overall, no method can non-parametrically estimate treatment effects under both interference and unobserved confounding.

We close this gap with the **Spatial Deconfounder**. Our key insight is that

interference *creates* the very multi-cause structure that deconfounders require: each unit receives its own treatment together with those of its neighbors, all shaped by the same latent spatial field. Rather than a nuisance, *interference becomes a source of signal for recovering hidden confounders*. Building on this, we develop a non-parametric and model-agnostic two-stage framework that first reconstructs a smooth substitute confounder using a conditional variational autoencoder (C-VAE) with a spatial prior[1], then estimates direct and spillover effects via any flexible outcome model (e.g., U-Net, GNN). This enables causal identification without requiring multiple treatment types, explicit latent-field models, or parametric outcome model specification. Our **contributions** can be summarized as follows:

1. We introduce the **Spatial Deconfounder**, a novel *non-parametric and model-agnostic* framework to *jointly* address spatial interference and unmeasured confounding by treating neighborhood treatment exposures as multi-cause signals.
2. We prove *identification* of direct and spillover effects under localized interference and a weak latent-field sufficiency assumption, without requiring a parametric model for the hidden process.
3. We extend the `SpaCE` benchmark to include structured interference and show, across climate-, health-, and social-science datasets, that our method consistently reduces bias relative to spatial autoregressive, matching, and spline-based baselines.

By leveraging interference as a lens into the hidden structure, the Spatial Deconfounder bridges spatial causal inference and multi-cause deconfounding,

[1]We use a C-VAE instantiation in this work, but the first-stage can be implemented with any suitable factor model that captures shared latent spatial structure.

opening a path to robust causal estimation in complex geographic systems.

8.2 Related Work

We give a brief overview of the related literature (see Appendix G.1 for a comprehensive survey and discussion). Our work sits at the intersection of three main areas: (i) spatial causal inference under interference and spatially structured confounding, (ii) deconfounding in general average treatment effect (ATE) estimation, and (iii) deep learning for spatial and latent structure modeling.

Classical Spatial Causal Inference Design- and model-based approaches assume exchangeability after conditioning on *observed* covariates (given an exposure mapping) [e.g., Hudgens and Halloran, 2008, Anselin, 1988, Hanks et al., 2015, Forastiere et al., 2021, Tchetgen Tchetgen et al., 2021]. They capture spatial dependence (splines/RSR, SAR, GNNs; simulators for domain physics) but do not address *unobserved* spatial confounding.

Spatial Confounding and Bias-Adjustment Methods Bias from *unmeasured* spatial structure is mitigated using latent spatial effects, orthogonalization methods such as S2SLS and SPATIAL+, proximity-based matching, instrumental variables, or Bayesian priors [e.g., Hodges and Reich, 2010, Dupont et al., 2022, Papadogeorgou et al., 2019a, Angrist et al., 1996]. These methods rely on explicit smooth-field models, IV assumptions, or strong priors; none can nonparametrically reconstruct the hidden confounder.

ATE Estimation Under Unobserved Confounding With unmeasured confounding, point identification typically fails. Sensitivity analysis instead yields assumption-indexed bounds, trading point identification for robustness [e.g.,

VanderWeele et al., 2015, Frauen et al., 2023]. Another approach is to reconstruct the unobserved confounder via the *deconfounder* framework, which fits a factor model to multiple causes to infer a substitute for the latent confounder and thereby restore point identification [Wang and Blei, 2019, Bica et al., 2020a]. However, existing deconfounder methods require many simultaneous causes and assume no interference. We invert this logic: interference itself yields multi-cause treatment vectors, enabling latent-field recovery even with a single treatment type.

Deep Learning for Spatial Representation and Latent Structure Deep learning architectures such as U-Nets, GNNs, and patch-wise transformers capture multi-scale and long-range spatial structure [e.g., Ronneberger et al., 2015b, Kipf and Welling, 2017, Liu et al., 2021], while C-VAEs and related deep generative models can recover latent factors from data [Kingma and Welling, 2013, Sohn et al., 2015]. However, these methods are typically predictive rather than causal. We combine these perspectives in a spatial causal setting: interference induces a multi-cause treatment signal that allows a deep latent-variable model to nonparametrically reconstruct a smooth latent confounder, enabling identification of both direct and spillover effects without requiring a specified latent field.

Positioning of Our Work Most methods for spatial causal inference under interference ignore unmeasured confounders or rely on strong priors, while "deconfounder" methods are not adapted to spatial settings. We close this gap by using interference as a multi-cause signal to nonparametrically reconstruct latent confounders, identifying direct and spillover effects without specifying a latent-field model.

8.3 Background and Setup

Notation We use uppercase letters (e.g., X) for random variables and lowercase letters (e.g., x) for realizations. Bold symbols denote vectors. We write the distribution of X as P_X, and omit subscripts when the meaning is clear.

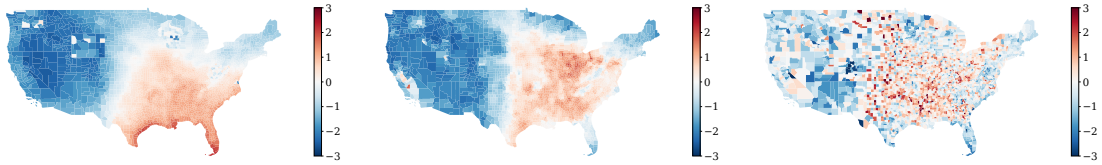
Data Structure: Lattice, Neighborhoods, and Observed Variables We consider a rectangular lattice $\mathcal{S} = \{(i, j) \mid i \in [N_x],\ j \in [N_y]\}$, where each site $s = (i, j)$ indexes a geographic cell. For a fixed radius $r > 0$, we define the neighborhood of s using the ℓ_∞ metric,

$$\mathcal{N}_s = \{s' \in \mathcal{S} : \|s' - s\|_\infty \le r,\ s' \neq s\}, \tag{8.1}$$
$$\text{where } \|s' - s\|_\infty = \max\{|i' - i|, |j' - j|\}.$$

Thus $\mathcal{N}_s$ is the $(2r+1) \times (2r+1)$ square centered at s, excluding s itself. We take r to be in *pixels* (multiples of the cell size), though it may also be specified as a physical distance and mapped to the grid resolution. Other shapes (e.g., ℓ_2 balls) are possible, but we use the square ℓ_∞ ball by default for computational convenience.

At each site s we observe covariates $\mathbf{X}_s \in \mathbb{R}^{d_x}$, a binary treatment $A_s \in \{0, 1\}$, and an outcome $Y_s \in \mathbb{R}$. For a neighborhood $\mathcal{N}_s$, we write $\mathbf{X}_{\mathcal{N}_s} = \{\mathbf{X}_{s'} : s' \in \mathcal{N}_s\}$, and analogously $A_{\mathcal{N}_s}$ and $Y_{\mathcal{N}_s}$. Realizations are denoted in lowercase, e.g., $\boldsymbol{x}_s$, a_s, y_s, and $\boldsymbol{x}_{\mathcal{N}_s} = \{\boldsymbol{x}_{s'} : s' \in \mathcal{N}_s\}$. For clarity, we focus on binary treatments, but the framework extends to continuous or multi-valued treatments via standard generalizations of the potential outcomes framework.

Potential Outcomes and Interference We adopt Rubin's potential outcomes framework [Rubin, 2005]. Standard causal inference relies on SUTVA, which rules out interference, i.e., one unit's outcome cannot depend on others' treatments. In spatial settings, this assumption is often violated, since treatment



(a) Humidity ($U(s)$) (b) $PM_{2.5}$ (A_s) (c) Mortality (Y_s)

Figure 8.2: Example spatial distributions of an unobserved confounder, treatment, and outcome in a real-world environmental-health application. The confounder $U(s)$ (summer humidity) varies smoothly across space, while the treatment A_s ($PM_{2.5}$) shows more local heterogeneity. The outcome Y_s (respiratory and cardiovascular mortality) reflects broader spatial health patterns.

exposures spill over. We assume *localized interference*: the potential outcome at site s depends only on its own treatment and those of its neighbors,

$$Y_s(\mathbf{a}) = Y_s(a_s, \mathbf{a}_{\mathcal{N}_s}), \tag{8.2}$$

where $\mathbf{a}$ is the full treatment vector, a_s the treatment at s, and $\mathbf{a}_{\mathcal{N}_s} = \{a_{s'} : s' \in \mathcal{N}_s\}$. The observed data contain only the realized outcome $Y_s = Y_s(A_s, \mathbf{A}_{\mathcal{N}_s})$ under the assigned intervention.

Causal Estimands Let $\mathbf{a}^{(1)}_{\mathcal{N}_s}$ and $\mathbf{a}^{(0)}_{\mathcal{N}_s}$ be two realizations of the neighbor treatments. Our targets are (i) the *average direct effect*, which varies the unit's own treatment while holding neighbors fixed,

$$\tau_{\text{dir}} = \mathbb{E}[Y_s(1, \mathbf{a}_{\mathcal{N}_s}) - Y_s(0, \mathbf{a}_{\mathcal{N}_s})], \tag{8.3}$$

and (ii) the *average spillover effect*, which varies neighbors' treatments while holding the unit fixed,

$$\tau_{\text{spill}} = \mathbb{E}[Y_s(a, \mathbf{a}^{(1)}_{\mathcal{N}_s}) - Y_s(a, \mathbf{a}^{(0)}_{\mathcal{N}_s})], \quad a \in \{0, 1\}, \tag{8.4}$$

with expectations taken over the observed joint distribution of $(\mathbf{X}_s, A_{\mathcal{N}_s})$.

Unobserved Spatial Confounding To identify the treatment effects in Eqs. (8.3) and (8.4), one typically assumes *ignorability*: potential outcomes

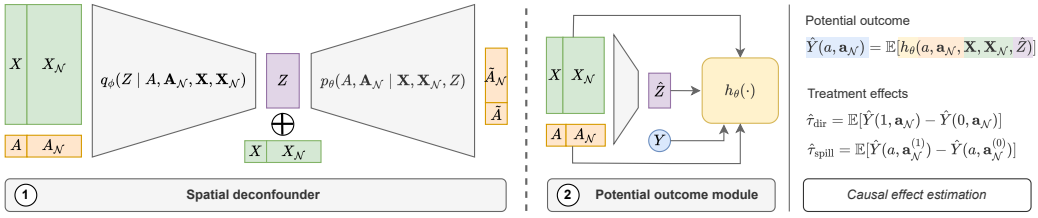


Figure 8.3: **Architecture of the spatial deconfounder & estimation framework.** Stage ①: The C-VAE takes **treatments** and **observed confounders** as input to learn the latent **substitute confounder**. Stage ②: We employ the reconstructed confounder together with the observed variables (now including the **outcome**) to train the **potential outcome** estimation module.

$Y_s(a_s, \mathbf{a}_{\mathcal{N}_s})$ are independent of treatment assignment given observed covariates $(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$. This assumption is not testable, and violations induce biased causal estimates. In practice, important drivers of both treatment exposure and outcomes often remain unobserved. We therefore posit an unobserved spatial field $U : \mathcal{S} \to \mathbb{R}^{d_U}$ that captures latent influences such as topography, wind patterns, or socioeconomic context. Since $U(s)$ may affect both treatment and outcomes, ignorability generally fails:

$$\text{Cov}(A_s, U(s)) \neq 0 \quad \text{and} \quad \text{Cov}(Y_s(a, \mathbf{a}_{\mathcal{N}_s}), U(s)) \neq 0, \tag{8.5}$$

where the covariances are understood component-wise when $U(s)$ is vector-valued. Thus, ignorability fails when conditioning only on $\mathbf{X}_s$ and $\mathbf{X}_{\mathcal{N}_s}$. In Section 8.5, we show that identification can nevertheless be recovered under mild smoothness assumptions on U together with our deconfounding procedure, by reconstructing a substitute latent field from treatment patterns.

Motivating Example Consider real environmental health data on a $0.25° \times 0.25°$ grid covering the continental United States. At each grid cell s, the treatment A_s indicates whether fine particulate matter ($PM_{2.5}$) exceeds the WHO guideline of 10 $\mu g/m^3$. Neighbor assignments are defined by a radius of one to two grid cells (roughly 25–50 km). The outcome Y_s is the rate of respiratory

and cardiovascular mortality aggregated from hospital records. Latent factors can confound this relationship; for example, a meteorological driver such as humidity varies smoothly across space and may jointly influence both pollution exposures and health outcomes. Figure 8.2 illustrates treatment, outcome, and such a confounder for this dataset. This example captures the type of smoothly varying, spatially shared latent structure our method targets: large-scale meteorological drivers such as humidity form a latent field $U(s)$ that jointly affects $PM_{2.5}$ exposures and mortality across neighboring counties, while any purely local one-off factors are captured in $(X_s, X_{\mathcal{N}_s})$ or assumed negligible. We formalize this as a latent-field sufficiency assumption in Section 8.5.

The remainder of this chapter shows how the joint vector $(A_s, \mathbf{A}_{\mathcal{N}_s})$—a "multiple-cause" analogue supplied for free by interference—can be harnessed to reconstruct $U(s)$ and obtain unbiased estimates of Eqs. (8.3) and (8.4).

8.4 Methodology

As illustrated in Algorithm 8.1, our approach proceeds in two stages. First, we reconstruct a smooth substitute confounder from the joint distribution of local and neighbor treatments, using a conditional variational autoencoder (C-VAE) that leverages interference as a multi-cause signal. Second, we feed the reconstructed confounder into a flexible outcome model for estimation. This separation follows standard practice in deconfounding to prevent mediators from being inadvertently learned into the substitute confounder, which would compromise identification of treatment effects.

Algorithm 8.1 Spatial Deconfounder

Input: Spatial covariates $\{\mathbf{X}_s\}_{s\in\mathcal{S}}$, treatments $\{A_s\}_{s\in\mathcal{S}}$, outcomes $\{Y_s\}_{s\in\mathcal{S}}$, neighborhood radius r, grid Laplacian L

Stage ①: Confounder reconstruction (C-VAE)

Define encoder $q_\phi(Z_s \mid A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}) = \mathcal{N}(\mu_\phi, \operatorname{diag}\sigma_\phi^2)$, decoder $p_\psi(A_s \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s)$, and prior $p_\theta(Z) = \mathcal{N}(\mathbf{0}, \tau^{-1}(L+\epsilon I)^{-1})$.

Minimize

$$\mathcal{L}_A = \sum_s \mathbb{E}_{q_\phi}[-\log p_\psi(A_s \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s)] + \sum_s \mathrm{KL}(q_\phi \,\|\, p_\theta).$$

Set substitute confounder $\hat{Z}_s \leftarrow \mathbb{E}_{q_\phi}[Z_s]$ for all s.

Stage ②: Potential outcome module

Choose a spatial model h (e.g., U-Net) for $\mathbb{E}[Y \mid \cdot]$ given the observed variables and substitute confounder, and fit it by minimizing

$$\mathcal{L}_Y = \sum_s \left(Y_s - h(A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s)\right)^2.$$

Estimate effects by plug-in contrasts using (8.11).

Stage ①: Confounder Reconstruction We model the assignment of treatments $\{A_s\}_{s\in\mathcal{S}}$ using an interference-aware C-VAE. The encoder

$$q_\phi(Z_s \mid A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}) = \mathcal{N}(\mu_\phi(\cdot), \operatorname{diag}\sigma_\phi^2(\cdot)) \tag{8.6}$$

maps the local treatment and neighborhood treatments, together with local and neighborhood covariates $(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$, into a latent embedding Z_s of the unobserved spatial field $U(s)$. The decoder

$$p_\psi(A_s \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s) = \sigma(f_\psi(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s)) \tag{8.7}$$

predicts A_s given covariates and the latent. To encode smoothness, we impose a Gaussian–Markov random-field (GMRF) prior $p_\theta(Z) = \mathcal{N}(\mathbf{0}, \tau^{-1}(L+\epsilon I)^{-1})$ with grid Laplacian L, or equivalently a deterministic penalty $\lambda Z^\top L Z$.

Formally, our generative model for the treatment field is

$$p_\theta(Z) = \mathcal{N}(\mathbf{0}, \tau^{-1}(L + \epsilon I)^{-1}),$$
$$p(A \mid X, Z) = \prod_{s \in \mathcal{S}} p_\psi(A_s \mid X_s, X_{\mathcal{N}_s}, Z_s),$$

with $A_s \mid X_s, X_{\mathcal{N}_s}, Z_s \sim \text{Bernoulli}(\sigma(f_\psi(X_s, X_{\mathcal{N}_s}, Z_s)))$. Thus, conditional independence of treatments holds across sites given (Z, X), and spatial dependence is encoded entirely via the GMRF prior on Z. The "multi-cause" structure of $(A_s, A_{\mathcal{N}_s})$ enters on the inference side through the encoder $q_\phi(Z_s \mid A_s, A_{\mathcal{N}_s}, X_s, X_{\mathcal{N}_s})$, which uses local treatment patterns (plus covariates) to infer a substitute confounder for the local value of the spatial latent field.

This C-VAE is trained by minimizing

$$\mathcal{L}_A(\phi, \psi) = \sum_s \mathbb{E}_{q_\phi}[-\log p_\psi(A_s \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s)] + \beta \sum_s \text{KL}(q_\phi \| p_\psi), \tag{8.8}$$

with KL warm-up ($\beta \uparrow 1$). After convergence, we set $\hat{Z}_s = \mathbb{E}_{q_\phi}[Z_s]$ as the reconstructed confounder.

Our C-VAE differs from standard C-VAE-type models in two ways tailored to the spatial–interference setting: (i) the encoder explicitly conditions on $(A_s, A_{\mathcal{N}_s}, X_s, X_{\mathcal{N}_s})$, using neighbor treatments as a multi-cause signal, and (ii) the latent field Z is given a GMRF prior with grid Laplacian L, enforcing spatial dependence consistent with our latent-field sufficiency assumption (Assumption 8.5 below).

Stage ②: Potential Outcome Module Given $\hat{Z}_s$, we estimate outcomes using a flexible function h:

$$\hat{Y}_s = \hat{\mathbb{E}}[Y \mid A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s] = h(A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s) \tag{8.9}$$

by minimizing the squared error loss

$$\mathcal{L}_Y = \sum_s \Big(Y_s - h(A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s)\Big)^2. \tag{8.10}$$

This module can be instantiated with any spatial model capable of handling interference and spatial confounding. For example, a U-Net architecture [Ronneberger et al., 2015b] captures multi-scale spatial dependencies through an encoder–decoder with skip connections. Notably, Oprescu et al. [2025], Ali et al. [2024] use a U-Net to account for interference and spatial confounding in spatiotemporal settings. Other options include graph neural networks, patch-wise transformers, or classical spatial regression models, depending on the data modality.

Effect estimation proceeds by plug-in contrasts: the *direct effect* is

$$\hat{\tau}_{\text{dir}} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \Big[h(1, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s) - h(0, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s) \Big], \tag{8.11}$$

and analogously for spillover effects by varying $\mathbf{A}_{\mathcal{N}_s}$. By drawing multiple $\hat{Z}_s$ from the full posterior q_ϕ instead of the mean, we can obtain uncertainty bands on $\hat{Z}_s$. We can then obtain uncertainty bands (with respect to the substitute confounder) by evaluating Eq. 8.11 on different draws of $\hat{Z}_s$.

Remark 8.1 (End-to-end Variant). One may train a single network by minimizing $\mathcal{L}_A + \gamma \mathcal{L}_Y$ while *blocking* gradients from $\mathcal{L}_Y$ into the C-VAE. This preserves mediator avoidance while making the overall implementation and training more straightforward. This separation ensures that the C-VAE is used only to reconstruct a substitute confounder, not to perform outcome estimation end-to-end.

Predictive Checks Following Rubin [1984], we assess whether the substitute confounder adequately explains the treatment assignment through posterior

predictive checks. On a held-out validation set, we draw M replicated treatment vectors $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(M)}$ from the decoder p_ψ and compare them against the observed assignment $\mathbf{a}$. Specifically, we compute the predictive p-value

$$p = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\Big\{T(\mathbf{a}^{(m)}) < T(\mathbf{a})\Big\}, \tag{8.12}$$

where $T(\mathbf{a})$ is a discrepancy statistic measuring model fit. Following Wang and Blei [2019], we use

$$T(\mathbf{a}) = \mathbb{E}_{Z \sim q_\phi}\Big[\log p_\psi(\mathbf{a} \mid \mathbf{X}, Z)\Big], \tag{8.13}$$

the marginal log-likelihood of the observed assignment under the posterior distribution of Z. A value of p close to 0.5 indicates that the C-VAE reproduces the treatment assignment distribution well, whereas extreme values signal model misspecification. In our experiments, we only consider C-VAE models with $0.25 < p < 0.75$.

8.5 Theoretical Properties of the Spatial Deconfounder

We now provide conditions under which the Spatial Deconfounder establishes causal identifiability of the direct and spillover effects in Eqs. (8.3) and (8.4). Our argument separates two steps: (i) an *identification* step showing that if a substitute confounder from observed neighborhood exposures exists, then direct and spillover effects are identified; and (ii) an *estimation* step stating conditions under which our Stage ① procedure (a C-VAE instantiation of a conditional factor model) consistently recovers this target. We begin with assumptions on consistency, positivity, and interference structure.

Assumption 8.2 (Spatial consistency)**.** The observed outcome is the potential outcome under the realized individual and neighborhood treatments. That is,

$$Y_s = Y_s(a_s, \mathbf{a}_{\mathcal{N}_s})$$

if a site s receives treatment a_s and its neighborhood $\mathcal{N}_s$ receives the vector of treatments $\mathbf{a}_{\mathcal{N}_s}$.

Assumption 8.3 (Spatial positivity)**.** For any site s, covariates $(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$, and treatment exposures $(a_s, \mathbf{a}_{\mathcal{N}_s})$, the probability of assignment is strictly positive: $0 < \Pr(a_s, \mathbf{a}_{\mathcal{N}_s} \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}) < 1$. Furthermore, we require *latent positivity* conditional on Z, i.e., $0 < \Pr(a_s, \mathbf{a}_{\mathcal{N}_s} \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \mathbf{Z}_s) < 1$ if $\Pr(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \mathbf{Z}_s) > 0$.

Assumption 8.4 (Localized interference)**.** The potential outcome at site s depends only on its own treatment and those of its neighbors $\mathcal{N}_s$, not on treatments outside $\mathcal{N}_s$.

Assumptions 8.2–8.4 are standard in the causal inference literature [e.g., Chen et al., 2024b, Forastiere et al., 2021] and ensure that the potential outcomes and the direct/spillover estimands in Eqs. (8.3) and (8.4) are well-defined under localized interference. Identification additionally requires assumptions on the confounding structure. Classical approaches for spatial treatment effects assume ignorability of the joint neighborhood exposure given observed covariates; we relax this and allow unobserved confounding driven by a shared latent spatial field $U : \mathcal{S} \to \mathbb{R}^{d_U}$ spanning the grid, while requiring that confounders affecting purely local variation are observed in $(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$.

Assumption 8.5 (Latent Field Sufficiency)**.** All confounders that act only on a single site are observed in $(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$. Any remaining unobserved confounding is mediated through a shared spatial latent field $U : \mathcal{S} \to \mathbb{R}^{d_U}$ that affects treatment assignments across multiple sites. In particular, there is no additional unobserved confounder $\tilde{U}$ that changes $(A_s, A_{\mathcal{N}_s}, Y_s(a, \mathbf{a}_{\mathcal{N}_s}))$ at some site s without also influencing treatments at other sites s'.

Assumption 8.5 is the spatial analogue of the "no single-cause confounders"

assumption in the deconfounder literature [e.g., Wang and Blei, 2019, Bica et al., 2020a]: all purely local confounders are observed, and any remaining unobserved confounding arises from a shared latent field U that induces dependence across sites. This is precisely the regime in which the neighborhood exposure $(A_s, A_{\mathcal{N}_s})$ can act as a multi-cause signal: multiple components of exposure are jointly shaped by the same latent spatial structure. Under a factor-model representation of the joint exposure, Proposition 5 of Wang and Blei [2019] implies that there exists a *population* substitute confounder $Z_s^\star$ (measurable with respect to $(A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$) such that the joint assignment $(A_s, A_{\mathcal{N}_s})$ is ignorable given $(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s^\star)$.

Finally, we connect this population target to what our Stage ① model learns.

Assumption 8.6 (Recoverable Substitute Confounder and Stage ① consistency)**.** There exists a population substitute confounder $Z_s^\star$ that is a deterministic function of the observed neighborhood exposure and covariates,

$$Z_s^\star = f_\phi(A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}),$$

such that conditioning on $(\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s^\star)$ renders the joint exposure $(A_s, A_{\mathcal{N}_s})$ ignorable as in Def. G.1. Moreover, the fitted Stage ① model yields an estimator

$$\hat{Z}_s = f_{\hat{\phi}}(A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$$

that converges to $Z_s^\star$ (e.g., $q_{\hat{\phi}}(Z_s \mid A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$ concentrates at $Z_s^\star$) as the sample size grows.

Assumption 8.6 has two parts. First, an *identification* requirement: there exists a population substitute confounder $Z_s^\star$, measurable with respect to $(A_s, A_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})$, that restores ignorability. Second, an *estimation* requirement: the chosen Stage ① factor model consistently recovers $Z_s^\star$. We do not claim that

C-VAEs are identifiable in full generality; rather, this assumption should be read as a well-specification or consistency condition on the chosen conditional factor-model class. In practice, we encourage stable recovery by incorporating spatial priors and can further regularize toward identifiability using objectives such as the IMA-regularized loss of Reizinger et al. [2022].

Intuition Under interference, each site's treatment is observed together with those of its neighbors. Because both A_s and $A_{\mathcal{N}_s}$ are influenced by the same latent field U, they provide multiple noisy "views" of the underlying spatial structure. By fitting a factor model to the joint distribution of own and neighbor treatments, we target the population substitute confounder $Z_s^\star$ and estimate it with $\hat{Z}_s$. Conditioning on this substitute confounder together with observed covariates restores ignorability, enabling estimation of direct and spillover effects.

Sensitivity to Proxy Error In Appendix G.2, we show that if the outcome regression is Lipschitz in Z, then using $\hat{Z}$ instead of $Z^\star$ induces $O(\mathbb{E}\|\hat{Z} - Z^\star\|)$ error in the direct and spillover treatment effects.

For notational simplicity, we write Z_s for the population target $Z_s^\star$ in what follows.

Theorem 8.7 (Causal identifiability)**.** *Suppose Assumptions 8.2–8.6 hold. Let Z_s be a piecewise constant function of the assigned neighborhood exposure and covariates $(a, \mathbf{a}_\mathcal{N}, \boldsymbol{x}, \boldsymbol{x}_\mathcal{N})$ and let the outcome be a separable function of the observed and unobserved variables:*

$$\begin{aligned}\mathbb{E}_Y[Y_s(a, \mathbf{a}_\mathcal{N}) \,|\mathbf{X}_s = \boldsymbol{x}, \mathbf{X}_{\mathcal{N}_s} = \boldsymbol{x}_\mathcal{N}, Z_s = z] \\ = f_1(a, \mathbf{a}_\mathcal{N}, \boldsymbol{x}, \boldsymbol{x}_\mathcal{N}) + f_2(z),\end{aligned} \tag{8.14}$$

$$\begin{aligned}\mathbb{E}_Y[Y_s \mid A_s = a, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_\mathcal{N}, \mathbf{X}_s = \boldsymbol{x}, \mathbf{X}_{\mathcal{N}_s} = \boldsymbol{x}_\mathcal{N}, Z_s = z] \\ = f_3(a, \mathbf{a}_\mathcal{N}, \boldsymbol{x}, \boldsymbol{x}_\mathcal{N}) + f_4(z),\end{aligned} \tag{8.15}$$

for continuously differentiable functions f_1, f_2, f_3, f_4. Consequently, the direct and spillover effects are identifiable as

$$\tau_{\text{dir}} = \mathbb{E}_{\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z}\Big[\mathbb{E}_Y[Y_s \mid A_s = 1, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s] - \mathbb{E}_Y[Y_s \mid A_s = 0, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]\Big], \tag{8.16}$$

$$\tau_{\text{spill}} = \mathbb{E}_{\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z}\Big[\mathbb{E}_Y[Y_s \mid a, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}^{(1)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s] - \mathbb{E}_Y[Y_s \mid a, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}^{(0)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]\Big]. \tag{8.17}$$

Proof. The proof is provided in Appendix G.2. □

Remark 8.8. Our identifiability result applies to settings with separable structural equations, a standard assumption in related work [e.g., Wang and Blei, 2019, Papadogeorgou and Samanta, 2023]. In spatial applications, this can capture latent factors that shift outcomes but are not fully observed, such as persistent baseline differences in respiratory risk due to long-run pollution exposure, chronic disease burden, or regional variation in care-seeking and reporting. Systematic measurement error in outcomes can be viewed similarly.

8.6 Experiments

We evaluate the Spatial Deconfounder on semi-synthetic datasets from the `SpaCE` benchmark [Tec et al., 2024], modified to incorporate both local interference and spatial confounding on real-world environmental data. To simulate unobserved confounding, we mask key covariates after data generation, i.e., we completely remove them from the dataset. We then compare different instantiations of our method against a range of spatial baselines under both local and spatial confounding scenarios. The section proceeds as follows: we describe the `SpaCE` environment and our data generation process, introduce the baselines and evaluation metrics, and finally interpret the results.

Additional details—including data generation, residual sampling, packages, hyperparameter tuning, and validation procedures—can be found in Appendix G.3. Replication code is available at `https://github.com/moprescu/Spatial-CI`.

Datasets and SpaCE Benchmark We build on the `SpaCE` benchmark [Tec et al., 2024], which provides semi-synthetic spatial datasets for causal inference under unobserved confounding. In its original form, `SpaCE` simulates causal effects by masking important covariates in real-world environmental and health data, but it assumes independent treatments and does not account for interference between neighboring units. This makes it inadequate for evaluating methods, such as ours, that explicitly address both unobserved spatial confounding and localized spillovers.

To address this, we extend the `SpaCE` data generation process in two ways. First, we project the raw environmental data onto a uniform $0.25^\circ \times 0.25^\circ$ latitude–longitude grid, allowing convolutional architectures to exploit spatial locality while preserving large-scale patterns. Second, we incorporate *interference* into the potential outcome model by allowing outcomes to depend not only on local treatment A_s but also on neighbor treatments $A_{\mathcal{N}_s}$ within radius r_d. We generate outcomes under two confounding regimes:

$$\text{(Local confounding)} \quad \hat{Y}_s = f(A_s, A_{\mathcal{N}_s}, X_s) + R_s, \tag{8.18}$$

$$\text{(Spatial confounding)} \quad \hat{Y}_s = f(A_s, A_{\mathcal{N}_s}, X_s, X_{\mathcal{N}_s}) + R_s, \tag{8.19}$$

where f is a predictive function learned from the observed data, X_s are observed covariates, and R_s are exogenous residuals. The local setting restricts confounding to site-level variables, while the spatial setting also allows neighborhood covariates to act as confounders.

Semi-Synthetic Data Generation To construct $\hat{Y}_s$, we proceed in four steps: (1) fit f using ensembles of machine learning models to predict observed outcomes Y_s, (2) compute residuals $\hat{R}_s = Y_s - f(\cdot)$ and estimate their spatial distribution P_R, (3) replace endogenous residuals with exogenous noise $R_s \sim P_R$, and (4) generate counterfactuals by varying local and neighbor treatments while holding confounders and residuals fixed. To simulate hidden confounding, we identify influential covariates by measuring the change in predictive performance when each is removed, then mask the most important ones at training and testing.

Raw Datasets From the full `SpaCE` suite, we focus in the main text on two collections:

Air Pollution and Mortality: County-level data for the mainland US in 2010, including elderly mortality (CDC), fine particulate matter ($PM_{2.5}$) treatment exposure [Di et al., 2019], behavioral risk factors (BRFSS) [Centers for Disease Control and Prevention, 2010], and Census demographics [U.S. Census Bureau, 2010]. We study the effect of $PM_{2.5}$ exposure (treatment) on mortality ($PM_{2.5} \rightarrow m$), with different masked confounders.

$PM_{2.5}$ Components: High-resolution (1×1 km) gridded data on total $PM_{2.5}$ [Di et al., 2019] and its chemical composition [Amini et al., 2022], using annual averages for 2000. We focus on the effect of sulfate on overall $PM_{2.5}$ ($SO_4 \rightarrow PM_{2.5}$), with key latent drivers such as *ammonium* (NH_4) and *organic carbon* (OC) masked.

The datasets are complementary: the first captures socioeconomic and demographic confounding, while the second reflects atmospheric chemistry. Additional datasets and hidden-confounder variants are described in Appendix G.4.

Baselines and Model Variants We benchmark against classical and modern spatial methods: S2SLS [Anselin, 1988] with outcome autoregression; spline-based SPATIAL and residualized SPATIAL+ [Dupont et al., 2022]; GCNN [Kipf and Welling, 2017] for non-linear neighbor aggregation; DAPSM [Papadogeorgou et al., 2019a] for proximity-based matching; and UNET [Ronneberger et al., 2015b], which can capture spillovers via neighbor treatments but does not adjust for hidden confounding.

For the *Spatial Deconfounder*, we instantiate the potential outcome module differently by setting the head to SPATIAL+ under local confounding (to ensure fairness) and to UNET under spatial confounding (to flexibly capture multi-scale structure). We also vary the neighborhood radius $r \in \{1, 2\}$ considered by the model and the latent confounder dimension in the C-VAE ($d_Z \in \{1, 2, 4, 8, 16, 32\}$).

Evaluation Metrics We assess performance on the direct (DIR) and spillover (SPILL) effects. As standard in causal inference [Hill, 2011, Shi et al., 2019, Cheng et al., 2022], we report standardized absolute bias, $\sigma_y^{-1} |\hat{\tau} - \tau|$, with true effect τ, estimate $\hat{\tau}$, and outcome standard deviation σ_y.

Results Tables 8.1 and 8.2 report performance under local and spatial confounding across different masked confounders (e.g., humidity, population density, ammonium, organic carbon). Across environments, the Spatial Deconfounder (C-VAE) variants consistently achieve lower bias on direct effects than existing spatial baselines. Even with non-smooth unobserved confounders like population density (ρ_{pop}), our framework still achieves lower bias. Importantly, unlike most benchmarks, both the Spatial Deconfounder and UNET can recover spillover effects, with the Spatial Deconfounder generally providing more accurate estimates.

Table 8.1: Performance of the Spatial Deconfounder and baselines under *local confounding* in the semi-synthetic spatial benchmark. Results are averaged over 10 runs with 95% confidence intervals. Here, r_d denotes the neighborhood radius used in data generation, and r denotes the neighborhood radius used by the deconfounder. Lower values of DIR and SPILL indicate lower bias; p is the predictive-check p-value, with values closer to 0.5 indicating better fit.

Setup	Conf.	Method	DIR	SPILL	p
$PM_{2.5}$	q_{summer}	C-VAE-SPATIAL+ ($r = 1$)	**0.04 ± 0.01**	**0.42 ± 0.08**	0.37 ± 0.07
↓		C-VAE-SPATIAL+ ($r = 2$)	**0.04 ± 0.01**	0.44 ± 0.09	0.36 ± 0.04
m		DAPSM	0.30 ± 0.03	n/a	n/a
($r_d = 1$)		GCNN	0.41 ± 0.03	n/a	n/a
		S2SLS-LAG1	0.20 ± 0.00	n/a	n/a
		SPATIAL+	0.13 ± 0.04	n/a	n/a
		SPATIAL	0.10 ± 0.07	n/a	n/a
$PM_{2.5}$	ρ_{pop}	C-VAE-SPATIAL+ ($r = 1$)	0.05 ± 0.02	**0.15 ± 0.05**	0.34 ± 0.04
↓		C-VAE-SPATIAL+ ($r = 2$)	**0.04 ± 0.03**	0.24 ± 0.06	0.35 ± 0.04
m		DAPSM	0.16 ± 0.01	n/a	n/a
($r_d = 2$)		GCNN	0.18 ± 0.03	n/a	n/a
		S2SLS-LAG1	0.07 ± 0.00	n/a	n/a
		SPATIAL+	0.10 ± 0.02	n/a	n/a
		SPATIAL	0.17 ± 0.03	n/a	n/a
SO_4	NH_4	C-VAE-SPATIAL+ ($r = 1$)	**0.07 ± 0.03**	0.64 ± 0.10	0.38 ± 0.04
↓		C-VAE-SPATIAL+ ($r = 2$)	**0.07 ± 0.03**	**0.16 ± 0.06**	0.39 ± 0.06
$PM_{2.5}$		DAPSM	1.44 ± 0.00	n/a	n/a
($r_d = 1$)		GCNN	0.52 ± 0.16	n/a	n/a
		S2SLS-LAG1	0.09 ± 0.00	n/a	n/a
		SPATIAL+	0.11 ± 0.03	n/a	n/a
		SPATIAL	0.08 ± 0.02	n/a	n/a
SO_4	OC	C-VAE-SPATIAL+ ($r = 1$)	**0.06 ± 0.03**	**0.18 ± 0.09**	0.43 ± 0.03
↓		C-VAE-SPATIAL+ ($r = 2$)	0.12 ± 0.06	0.35 ± 0.08	0.43 ± 0.04
$PM_{2.5}$		DAPSM	1.24 ± 0.01	n/a	n/a
($r_d = 2$)		GCNN	0.30 ± 0.10	n/a	n/a
		S2SLS-LAG1	0.21 ± 0.00	n/a	n/a
		SPATIAL+	0.13 ± 0.07	n/a	n/a
		SPATIAL	0.29 ± 0.01	n/a	n/a

Additional experiments in Appendix G.4 confirm these trends in broader settings. In a few cases where classical baselines perform comparably or slightly better, the scenarios involve very weak or highly smooth confounding—conditions under which stronger parametric assumptions may help. Overall, the results show that leveraging interference as a multi-cause signal sub-

Table 8.2: Performance of the Spatial Deconfounder and baselines under *spatial confounding* in the semi-synthetic spatial benchmark. Results are averaged over 10 runs with 95% confidence intervals. Here, r_d denotes the neighborhood radius used in data generation, and r denotes the neighborhood radius used by the deconfounder. Lower values of DIR and SPILL indicate lower bias; p is the predictive-check p-value, with values closer to 0.5 indicating better fit.

Setup	Conf.	Method	DIR	SPILL	p
$PM_{2.5}$	ρ_{pop}	C-VAE-UNET ($r = 1$)	0.05 ± 0.01	0.22 ± 0.06	0.34 ± 0.03
↓		C-VAE-UNET ($r = 2$)	**0.04 ± 0.02**	**0.12 ± 0.06**	0.36 ± 0.06
m		DAPSM	0.20 ± 0.01	n/a	n/a
($r_d = 1$)		GCNN	0.17 ± 0.06	n/a	n/a
		S2SLS-LAG1	0.05 ± 0.00	n/a	n/a
		SPATIAL+	0.27 ± 0.18	n/a	n/a
		SPATIAL	0.06 ± 0.06	n/a	n/a
		UNET	0.06 ± 0.01	0.17 ± 0.04	n/a
SO_4	OC	C-VAE-UNET ($r = 1$)	0.06 ± 0.02	0.09 ± 0.04	0.44 ± 0.03
↓		C-VAE-UNET ($r = 2$)	0.06 ± 0.02	0.18 ± 0.06	0.45 ± 0.03
$PM_{2.5}$		DAPSM	1.57 ± 0.00	n/a	n/a
($r_d = 1$)		GCNN	0.42 ± 0.15	n/a	n/a
		S2SLS-LAG1	0.13 ± 0.00	n/a	n/a
		SPATIAL+	0.06 ± 0.05	n/a	n/a
		SPATIAL	**0.04 ± 0.01**	n/a	n/a
		UNET	0.07 ± 0.02	**0.05 ± 0.02**	n/a

stantially improves both direct and spillover effect estimation. These findings support the core premise of the Spatial Deconfounder: interference can be exploited, rather than treated as a nuisance, to improve causal inference under unobserved confounding.

Stress Tests We run several stress and robustness tests. For example, our theory hinges on Assumption 8.5, which rules out purely local ("single-cause") unobserved confounders. We therefore run a targeted stress test that injects an additional site-specific latent confounder, violating latent-field sufficiency; performance degrades as this violation strengthens, and baselines that do not rely on multi-cause structure can become competitive in the extreme. As an additional check, we evaluate cases with high treatment sparsity. The p-value sub-

stantially declines relative to our values in Table G.3, demonstrating that this metric can effectively flag poor reconstructions. See Appendix G.5 for details.

8.7 Conclusion

We introduce the **Spatial Deconfounder**, a framework that jointly addresses interference and unobserved spatial confounding by treating neighborhood treatments as a multi-cause signal. A C-VAE with a spatial prior reconstructs a substitute confounder, enabling estimation of direct and spillover effects with flexible outcome models. We prove identification under assumptions on the latent spatial field and outcome structure.

More broadly, our results suggest a shift in perspective: rather than treating interference solely as a nuisance, it can provide signal about hidden structure. While our guarantees rely on idealized assumptions, our semi-synthetic experiments on minimally modified environmental-health data show consistent bias reductions relative to strong spatial and deep-learning baselines, supporting the practical value of interference-driven multi-cause representations. Future work includes extensions to spatiotemporal settings, continuous treatments, and larger-scale deployments.

CHAPTER 9

CAUSAL INFERENCE ON NETWORKS UNDER MISSPECIFIED EXPOSURE MAPPINGS: A PARTIAL IDENTIFICATION FRAMEWORK

This chapter is based on Schröder et al. [2026], developed jointly with Maresa Schröder. I contributed to the initial problem framing and to the theoretical development. The version included here emphasizes the conceptual setup and theoretical results most closely connected to the dissertation's themes.

Network interference complicates causal inference because a unit's outcome may depend on treatments assigned to other units in the network. A standard approach is to summarize neighbors' treatments through an exposure mapping, but this mapping is typically imposed by the analyst and may be misspecified. When that occurs, conventional estimators of direct and spillover effects can be substantially biased.

This chapter develops a partial identification framework for causal inference under misspecified exposure mappings. We model misspecification through sensitivity bounds on the ratio between the exposure propensity induced by the analyst-specified mapping and that induced by the true, unknown interference mechanism. Under this formulation, we derive sharp bounds on conditional and average potential outcomes, and hence on direct, spillover, and overall treatment effects. We further develop an orthogonal estimation strategy for these bounds and establish guarantees showing robustness to nuisance estimation error, recovery of the sharp identified set under suitable consistency conditions, and asymptotic validity under weaker misspecification. The resulting framework extends sensitivity analysis and partial identification ideas to network settings where uncertainty about how interference is summarized is itself a central inferential challenge.

9.1 Introduction

Interference is one of the main obstacles to causal inference in structured data. In network settings, the treatment assigned to one unit may affect not only its own outcome but also the outcomes of connected units, so the usual no-interference formulation is no longer appropriate. This makes both identification and estimation substantially more difficult, since in principle each unit's potential outcome may depend on a high-dimensional treatment configuration over the network.

A common response is to impose an *exposure mapping* that compresses neighbors' treatments into a low-dimensional summary. Examples include the proportion of treated neighbors, whether treated exposure exceeds a threshold, or treatment within a fixed-radius neighborhood. Such summaries make causal analysis tractable and are now standard in the literature on network interference [Aronow and Samii, 2017, Forastiere et al., 2021, 2022, Ogburn et al., 2024]. But they also introduce a strong structural assumption: namely, that the chosen summary correctly captures how spillovers operate. In many applications, that assumption is difficult to justify. Social influence may vary with tie strength, spatial spillovers may decay with distance in unknown ways, and effects may extend beyond the neighborhood encoded by the analyst.

This work asks what can still be learned when the exposure mapping is not trusted to be exactly correct. Rather than treating misspecification as a nuisance to be ignored, we model it explicitly. Our starting point is that a misspecified exposure mapping changes the induced exposure propensity, and that this shift can be bounded in a sensitivity-analysis-style formulation. This leads naturally to a partial identification perspective: instead of a single point estimate built on a fragile exposure assumption, we derive identified intervals that quantify

what remains learnable under controlled deviations from the analyst-specified mapping in network settings.

The chapter develops this idea in three steps. First, we formulate a general sensitivity model for exposure-mapping misspecification and derive sharp bounds on conditional and average potential outcomes. Second, we show how this framework specializes to several common forms of misspecification, including weighted neighborhood exposure, threshold misspecification, and omitted higher-order spillovers. Third, we develop an orthogonal estimation strategy for the resulting bounds and summarize corresponding guarantees on robustness, sharpness, and asymptotic validity.

The role of this chapter within the dissertation is primarily conceptual and theoretical. It extends the broader theme of reliable causal inference under unreliable assumptions to network settings, where the fragile assumption is not only unconfoundedness or overlap, but also the analyst's representation of interference itself.

9.2 Background and Setup

We consider causal inference on a known network under interference. Our goal is to estimate potential outcomes and treatment effects when a unit's outcome may depend on both its own treatment and a summary of its neighbors' treatments. The key challenge is that this summary, usually encoded through an exposure mapping, is rarely known with certainty. We therefore begin by formalizing the network setting and the causal estimands of interest.

Notation We use capital letters, such as X, to denote random variables, and lowercase letters, such as x, to denote realizations. The distribution of X is writ-

ten as $\mathbb{P}_X$, though we omit the subscript when the meaning is clear from context. For discrete variables, we write the probability mass function as $P(x) = P(X = x)$; for continuous variables, we write the density as $p(x)$. Throughout, we work in the potential outcomes framework [Rubin, 2005]. A summary of notation is deferred to Appendix H.1.

9.2.1 Network Setting

We follow the standard setting for causal inference on networks [Chen et al., 2024c, Forastiere et al., 2021]. We consider an undirected network of known structure with node set $\mathcal{N}_{\mathcal{G}}$, where $|\mathcal{G}| = N$, and edge set $\mathcal{E}$, where $(i, j) = (j, i)$ for $i, j \in \mathcal{G}$. For each node i, we define the partition $(i, \mathcal{N}_i, \mathcal{N}_{-i})$, where $\mathcal{N}_i$ denotes the *neighborhood* of node i, that is, the set of nodes j such that $(i, j) \in \mathcal{E}$, and $\mathcal{N}_{-i}$ denotes its complement in $\mathcal{G}$. We write $|\mathcal{N}_i| = n_i$ for the *degree* of node i, and omit the subscript when it is clear from context.

Each unit i is associated with a binary treatment $T_i \in \{0, 1\}$, covariates $X_i \in \mathcal{X}^d$, and an outcome $Y_i \in \mathcal{Y}$. We allow the treatment assignment to depend either on the unit's own covariates, so that $\mathbf{X}_i = X_i$, or on both the unit's and its neighbors' covariates, so that $\mathbf{X}_i = (X_i, X_{\mathcal{N}_i})$. In the latter case, we additionally assume that every node has the same degree n. The treatment assignment of unit i is assumed independent of the treatment assignments of other units given $\mathbf{X}_i$. We denote the unit-level propensity score by

$$\pi^t(\mathbf{x}) := P(T = t \mid \mathbf{X} = \mathbf{x}). \tag{9.1}$$

A standard device in network causal inference is the *exposure mapping*, a function $g : [0, 1]^{n_i} \to \mathcal{Z}$ that summarizes the treatments assigned to the neigh-

bors of unit i. Writing

$$z_i := g(t_{\mathcal{N}_i}), \tag{9.2}$$

the exposure mapping reduces the potentially high-dimensional vector of neighboring treatments to a scalar or low-dimensional exposure variable. Under this representation, the potential outcome is written as $Y_i(t_i, z_i)$ and depends on both the unit's own treatment and the summarized neighborhood exposure. We denote the corresponding network propensity by

$$\pi^g(z \mid \mathbf{x}) := p(g(t_{\mathcal{N}_i}) = z \mid \mathbf{X}_i = \mathbf{x}). \tag{9.3}$$

9.2.2 Causal Estimands Under Interference

We are interested in the *average potential outcome* (APO) under individual treatment $T = t$ and neighborhood exposure $Z = z$, defined as

$$\psi(t, z) := \mathbb{E}[Y(t, z)], \tag{9.4}$$

and the *conditional average potential outcome* (CAPO),

$$\mu(t, z, \mathbf{x}) := \mathbb{E}[Y(t, z) \mid \mathbf{X} = \mathbf{x}]. \tag{9.5}$$

Under interference, total causal effects decompose naturally into direct and spillover components.

Definition 9.1 (Direct Effects (ADE / IDE))**.** The average direct effect (ADE) and individual direct effect (IDE) between individual treatment assignments $T = t$ and $T = t'$ while holding the neighborhood exposure $Z = z$ fixed are

$$\tau_d^{(t,z),\,(t',z)} := \psi(t, z) - \psi(t', z), \tag{9.6}$$

$$\tau_{d_i}^{(t,z),\,(t',z)}(\mathbf{x}_i) := \mu(t, z, \mathbf{x}_i) - \mu(t', z, \mathbf{x}_i). \tag{9.7}$$

Definition 9.2 (Spillover Effects (ASE / ISE))**.** The average spillover effect (ASE) and individual spillover effect (ISE) between neighborhood treatment levels $Z = z$ and $Z = z'$ while holding the individual treatment $T = t$ fixed are

$$\tau_s^{(t,z),\,(t,z')} := \psi(t,z) - \psi(t,z'), \tag{9.8}$$

$$\tau_{s_i}^{(t,z),\,(t,z')}(\mathbf{x}_i) := \mu(t,z,\mathbf{x}_i) - \mu(t,z',\mathbf{x}_i). \tag{9.9}$$

Definition 9.3 (Overall Effects (AOE / IOE))**.** The average overall effect (AOE) and individual overall effect (IOE) between treatment-exposure pairs (t,z) and (t',z') are

$$\tau_o^{(t,z),\,(t',z')} := \psi(t,z) - \psi(t',z'), \tag{9.10}$$

$$\tau_{o_i}^{(t,z),\,(t',z')}(\mathbf{x}_i) := \mu(t,z,\mathbf{x}_i) - \mu(t',z',\mathbf{x}_i). \tag{9.11}$$

As in prior work on causal inference with network interference [e.g., Chen et al., 2024a, Forastiere et al., 2021], we assume consistency, interference through a summary exposure, unconfoundedness, and positivity, adapted to the network setting.

Assumption 9.4 (Network Consistency)**.** The potential outcome equals the observed outcome under the realized individual and neighborhood treatment assignments; that is, $y_i = y_i(t_i, t_{\mathcal{N}_i})$ if unit i receives treatment t_i and its neighbors receive treatment vector $t_{\mathcal{N}_i}$.

Assumption 9.5 (Network Interference)**.** A unit's treatment affects its own outcome and possibly the outcomes of other units through a summary function g^*. Specifically, for all $t_{\mathcal{N}_i}, t'_{\mathcal{N}_i}$ satisfying $g^*(t_{\mathcal{N}_i}) = g^*(t'_{\mathcal{N}_i})$, it holds that

$$y_i(t_i, t_{\mathcal{N}_i}) = y_i(t_i, t'_{\mathcal{N}_i}). \tag{9.12}$$

Assumption 9.6 (Network unconfoundedness)**.** Given the relevant covariates, the potential outcome is independent of the individual and neighborhood treatment assignments:

$$y_i(t, t_{\mathcal{N}_i}) \perp\!\!\!\perp T_i, t_{\mathcal{N}_i} \mid \mathbf{x}_i \tag{9.13}$$

for all $t, t_{\mathcal{N}_i}$. If the summary function g^* is correctly specified, then

$$y_i(t, g^*(t_{\mathcal{N}_i})) \perp\!\!\!\perp T_i, g^*(t_{\mathcal{N}_i}) \mid \mathbf{x}_i \tag{9.14}$$

for all $t, t_{\mathcal{N}_i}$.

Assumption 9.7 (Network positivity)**.** Given the individual and neighborhood covariates, every treatment pair (t, z) is observed with positive probability: $0 < p(t, z \mid \mathbf{x}) < 1$ for all $\mathbf{x}, t, z$.

Under Assumptions 9.4 to 9.7 and a correctly specified exposure mapping g^*, the potential outcomes are identified from observational data. However, if the analyst uses an exposure mapping $g \neq g^*$, then the key assumptions underlying point identification may fail. We therefore move to partial identification.

9.3 Partial Identification Under Exposure-Mapping Misspecification

When the exposure mapping is misspecified, direct point identification is generally unavailable. We therefore model misspecification explicitly through a sensitivity formulation and derive identified intervals for potential outcomes and treatment effects.

9.3.1 Sensitivity Model for Misspecified Exposure Mappings

We formalize misspecification as a distribution shift in the *exposure propensity*

$$\pi^g(z \mid \mathbf{x}) := p(g(t_N) = z \mid \mathbf{x}) \tag{9.15}$$

between the analyst-specified mapping g and the true but unknown mapping g^*. For a given level of misspecification, we introduce lower and upper bounds $b^-(z, \mathbf{x}) \leq b^+(z, \mathbf{x})$, with $b^-(z, \mathbf{x}) \in (0, 1]$ and $b^+(z, \mathbf{x}) \in [1, \infty)$, such that

$$b^-(z, \mathbf{x}) \leq \frac{p(g^*(t_N) = z \mid \mathbf{x})}{p(g(t_N) = z \mid \mathbf{x})} \leq b^+(z, \mathbf{x}) \tag{9.16}$$

for all $(z, \mathbf{x})$.

This sensitivity model does not require parametric assumptions on the data-generating process. Its interpretation depends on the form of the exposure mapping and the way misspecification is modeled. The next subsection develops general sharp bounds under Eq. (9.16); concrete examples of how to construct $b^{\pm}$ are deferred to Section 9.4.

9.3.2 Sharp Bounds on Potential Outcomes

We now derive sharp upper and lower bounds on the CAPO under the misspecification model above. In Appendix H.2, we translate these into corresponding bounds for direct, spillover, and overall effects. We then develop an orthogonal estimator for these bounds in Section 9.5 and summarize its theoretical guarantees in Section 9.6. All proofs are deferred to Appendix H.3.

Definition 9.8 (Sharp bounds)**.** Let $\tilde{\mathbb{P}}$ denote a distribution on $(\mathbf{X}, T, Z, Y(T, Z))$ such that: (i) $\tilde{\mathbb{P}}$ agrees with the observed distribution $\mathbb{P}$ on $(\mathbf{X}, T, Z, Y)$, and (ii) the corresponding conditional distribution $\tilde{\pi}^g(z \mid \mathbf{x})$ satisfies $b^-(z, \mathbf{x}) \leq \frac{\tilde{\pi}^g(z|\mathbf{x})}{\pi^g(z|\mathbf{x})} \leq b^+(z, \mathbf{x})$ almost surely. Let $\mathcal{M}$ denote the set of all such distributions $\tilde{\mathbb{P}}$. Then the sharp

upper and lower bounds on the CAPO are

$$\mu^+(t,z,\mathbf{x}) = \sup_{\tilde{\mathbb{P}}\in\mathcal{M}} \mathbb{E}_{\tilde{\mathbb{P}}}[Y(t,z) \mid \mathbf{X}=\mathbf{x}], \tag{9.17}$$

$$\mu^-(t,z,\mathbf{x}) = \inf_{\tilde{\mathbb{P}}\in\mathcal{M}} \mathbb{E}_{\tilde{\mathbb{P}}}[Y(t,z) \mid \mathbf{X}=\mathbf{x}]. \tag{9.18}$$

To understand these bounds, note that the problem is to bound the conditional mean

$$\mathbb{E}[Y \mid T=t, Z=z, \mathbf{X}=\mathbf{x}] = \int_y y\, p(y \mid t,z,\mathbf{x})\,\mathrm{d}y. \tag{9.19}$$

A naive construction based directly on Eq. (9.16) yields valid but generally non-sharp bounds. To obtain sharp bounds, we follow ideas from causal sensitivity analysis [Dorn et al., 2025a, Frauen et al., 2023] and characterize the optimal redistribution of mass through a cutoff on the outcome distribution.

Let $F_Y(y) := F_Y(y \mid t,z,\mathbf{x})$ denote the conditional cumulative distribution function of Y, and define

$$\alpha^\pm = \frac{(1-b^\mp(z,\mathbf{x}))b^\pm(z,\mathbf{x})}{b^\pm(z,\mathbf{x}) - b^\mp(z,\mathbf{x})}. \tag{9.20}$$

We then define the conditional quantile

$$Q^\pm(t,z,\mathbf{x}) := \begin{cases} \inf\{y \mid F_Y(y) \geq \alpha^\pm\}, & \text{if } b^- < 1 < b^+, \\ \inf\left\{y \mid F_Y(y) \geq \frac{1}{2}\right\}, & \text{if } b^- = b^+. \end{cases} \tag{9.21}$$

Using this quantile as the cutoff, the sharp bounds admit a closed-form representation.

Theorem 9.9. *[Sharp Bounds] Let $Q^\pm(t,z,\mathbf{x})$ be defined as in Eq.* (9.21)*, and let $(u)_+ = \max\{u,0\}$. Then the sharp CAPO upper and lower bounds are*

$$\begin{aligned} \mu^\pm(t,z,\mathbf{x}) = Q^\pm(t,z,\mathbf{x}) &+ \frac{1}{b^\mp(z,\mathbf{x})}\mathbb{E}\left[(Y - Q^\pm(t,z,\mathbf{x}))_+ \mid t,z,\mathbf{x}\right] \\ &- \frac{1}{b^\pm(z,\mathbf{x})}\mathbb{E}\left[(Q^\pm(t,z,\mathbf{x}) - Y)_+ \mid t,z,\mathbf{x}\right]. \end{aligned} \tag{9.22}$$

Remark 9.10 (Limits of the Sensitivity Model). If $b^-(z, \mathbf{x}) = b^+(z, \mathbf{x}) = 1$, then the identified set collapses and

$$\mu^{\pm}(t, z, \mathbf{x}) = \mathbb{E}[Y \mid t, z, \mathbf{x}]. \tag{9.23}$$

As $b^+(z, \mathbf{x}) \to \infty$ with $b^-(z, \mathbf{x})$ fixed, the upper bound concentrates on the upper $b^-(z, \mathbf{x})$-tail of $Y \mid (t, z, \mathbf{x})$, and the lower bound concentrates on the corresponding lower tail. In the extreme limit $b^-(z, \mathbf{x}) \to 0$ and $b^+(z, \mathbf{x}) \to \infty$, the bounds become vacuous and converge to the conditional support:

$$\mu^+(t, z, \mathbf{x}) \to \operatorname{ess\,sup}(Y \mid t, z, \mathbf{x}), \tag{9.24}$$

$$\mu^-(t, z, \mathbf{x}) \to \operatorname{ess\,inf}(Y \mid t, z, \mathbf{x}). \tag{9.25}$$

Implications for Direct, Spillover, and Overall Effects The bounds in Theorem 9.9 immediately induce bounds on treatment effects. Since direct, spillover, and overall effects are all contrasts of potential outcomes, upper and lower effect bounds are obtained by combining upper and lower endpoint bounds on the relevant potential outcomes.

For example, for a generic contrast between two treatment-exposure pairs (a, b), the upper bound takes the form

$$\tau^+ = f^+(a) - f^-(b), \tag{9.26}$$

while the lower bound takes the form

$$\tau^- = f^-(a) - f^+(b), \tag{9.27}$$

where f denotes either μ or ψ depending on whether the estimand is conditional or marginal. Applying this principle yields corresponding identified intervals for direct, spillover, and overall effects. We defer the explicit formulas to Appendix H.2.

9.4 Examples of Exposure-Mapping Misspecification

We now instantiate the general sensitivity model for several canonical forms of exposure-mapping misspecification.

9.4.1 Weighted Neighborhood Exposure

Suppose the analyst specifies the exposure mapping as the proportion of treated neighbors,

$$g(t_{\mathcal{N}}) := \sum_{j \in \mathcal{N}} \frac{t_j}{n} = \frac{N_T}{n}, \tag{9.28}$$

where N_T denotes the number of treated neighbors and n denotes the neighborhood size. Now suppose the true exposure mapping $g^*(t_{\mathcal{N}})$ is instead a weighted proportion of treated neighbors, where each weight may differ from $1/n$ by at most ε, with $0 \le \varepsilon \le 1/n$. Then one can construct the sensitivity bounds

$$b^-(z, \mathbf{x}) = \inf_{s \in \mathcal{Z}} \frac{P\left(\frac{ns}{1-\varepsilon n} \le N_T \le \frac{nz}{1+\varepsilon n} \mid \mathbf{x}\right)}{P(ns \le N_T \le nz \mid \mathbf{x})}, \quad b^+(z, \mathbf{x}) = \sup_{s \in \mathcal{Z}} \frac{P\left(\frac{ns}{1+\varepsilon n} \le N_T \le \frac{nz}{1-\varepsilon n} \mid \mathbf{x}\right)}{P(ns \le N_T \le nz \mid \mathbf{x})}. \tag{9.29}$$

These characterize the degree to which using an unweighted exposure may distort the induced exposure propensity relative to the true weighted exposure.

9.4.2 Threshold Misspecification

Next suppose the analyst specifies a threshold exposure mapping. Let $h(t_{\mathcal{N}}) := \sum_{j \in \mathcal{N}} \frac{t_j}{n}$ denote the proportion of treated neighbors, and define $g(t_{\mathcal{N}}) := \mathbf{1}_{[h(t_{\mathcal{N}}) \ge c]}$. Then

$$P(g(t_{\mathcal{N}}) = 1 \mid \mathbf{x}) = P(N_T \ge nc \mid \mathbf{x}).$$

If the true threshold is instead $c^* \in [c - \varepsilon, c + \varepsilon]$, so that $g^*(t_{\mathcal{N}}) := \mathbf{1}_{[h(t_{\mathcal{N}}) \ge c^*]}$, then the ratio in Eq. (9.16) can be bounded by

$$\frac{P(N_T \ge n(c+\varepsilon) \mid \mathbf{x})}{P(N_T \ge nc \mid \mathbf{x})} \le \frac{P(g^*(t_{\mathcal{N}}) = 1 \mid \mathbf{x})}{P(g(t_{\mathcal{N}}) = 1 \mid \mathbf{x})} \le \frac{P(N_T \ge n(c-\varepsilon) \mid \mathbf{x})}{P(N_T \ge nc \mid \mathbf{x})}, \tag{9.30}$$

with the corresponding bounds for $z = 0$ obtained by complement probabilities.

9.4.3 Higher-Order Spillovers

Finally, suppose the analyst truncates interference to direct neighbors, but the true exposure depends also on treatments assigned to additional units $t_U \subset t_{\mathcal{N}_{-i}}$. In that case, the exposure mapping is misspecified because the true exposure depends on $g(t_{\mathcal{N}\cup U})$, rather than only on $g(t_{\mathcal{N}})$. This induces an unobserved distortion of the exposure summary and can be treated analogously to unobserved confounding. Following the sensitivity analysis literature [Dorn and Guo, 2022, Frauen et al., 2023], we assume user-specified functions $b^{\pm}$ satisfying

$$b^{-}(z, \mathbf{x}) \leq \frac{p(g(t_{\mathcal{N}\cup U}) = z \mid \mathbf{x})}{p(g(t_{\mathcal{N}}) = z \mid \mathbf{x})} \leq b^{+}(z, \mathbf{x}). \tag{9.31}$$

This formulation allows for a broad class of higher-order spillover misspecifications and can be combined with different baseline exposure mappings g.

9.5 Orthogonal Estimation of the Bounds

The characterization in Theorem 9.9 suggests a natural plug-in strategy based on estimating the cutoff $Q^{\pm}(t, z, \mathbf{x})$ and the corresponding conditional tail expectations. However, plug-in estimators can be highly sensitive to nuisance estimation error and may suffer substantial finite-sample bias, especially in flexible nonparametric settings where the nuisance functions are more complex than the bound function itself Kennedy [2023a]. To obtain a more robust procedure, we therefore apply orthogonalization strategies [Dorn et al., 2025a, Oprescu et al., 2023] and develop orthogonal pseudo-outcomes for the bounds whose first-order sensitivity to nuisance estimation error vanishes.

9.5.1 Orthogonal Pseudo-outcomes

Define the conditional tail moments

$$\gamma_u^{\pm}(t,z,\mathbf{x}) := \mathbb{E}\left[(Y - Q^{\pm}(\cdot))_+ \mid T = t, Z = z, \mathbf{X} = \mathbf{x}\right], \tag{9.32}$$

$$\gamma_l^{\pm}(t,z,\mathbf{x}) := \mathbb{E}\left[(Q^{\pm}(\cdot) - Y)_+ \mid T = t, Z = z, \mathbf{X} = \mathbf{x}\right]. \tag{9.33}$$

We refer to $(\pi^t, \pi^g, Q^{\pm}, \gamma_u^{\pm}, \gamma_l^{\pm})$ collectively as *nuisance functions*.

Recall that $T \in \{0,1\}$, while Z may be discrete or continuous. When Z is discrete, the target functional $\mu^{\pm}(t,z,\mathbf{x})$ is pathwise differentiable for fixed (t,z), allowing efficient influence-function-based construction. When Z is continuous, point evaluation at $Z = z$ is not pathwise differentiable, so we instead localize using a kernel $K_h(Z - z)$ with bandwidth h.

For ease of presentation, we state the pseudo-outcome for the upper bound. The corresponding lower-bound construction is deferred to Appendix H.2.

Theorem 9.11. *Let $S = (\mathbf{X}, Y, T, Z)$ and fix (t,z). Define*

$$\omega_{z,h}(Z) := \begin{cases} \mathbf{1}_{[Z=z]}, & \text{if } Z \text{ is discrete}, \\ K_h(Z - z), & \text{if } Z \text{ is continuous}, \end{cases}$$

and let $\pi^g(Z \mid \mathbf{X})$ denote the conditional pmf when Z is discrete and the conditional density when Z is continuous. Let

$$\widehat{\eta} = (\widehat{\pi}^t, \widehat{\pi}^g, \widehat{Q}^+, \widehat{\gamma}_u^+, \widehat{\gamma}_l^+) \tag{9.34}$$

be estimates of the nuisance functions. Then an orthogonal pseudo-outcome for the CAPO upper bound $\mu^+(t,z,\mathbf{x})$ is

$$\begin{aligned} \phi_{t,z}^+(S;\widehat{\eta}) = \widehat{Q}^+(t,z,\mathbf{X}) &+ \frac{\widehat{\gamma}_u^+(t,z,\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\widehat{\gamma}_l^+(t,z,\mathbf{X})}{b^+(z,\mathbf{X})} \\ &+ \frac{\mathbf{1}_{[T=t]}\omega_{z,h}(Z)}{\widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(Z \mid \mathbf{X})}\Bigg[\frac{(Y - \widehat{Q}^+(t,Z,\mathbf{X}))_+ - \widehat{\gamma}_u^+(t,Z,\mathbf{X})}{b^-(Z,\mathbf{X})} \\ &\qquad - \frac{(\widehat{Q}^+(t,Z,\mathbf{X}) - Y)_+ - \widehat{\gamma}_l^+(t,Z,\mathbf{X})}{b^+(Z,\mathbf{X})}\Bigg]. \end{aligned} \tag{9.35}$$

Algorithm 9.1 Orthogonal Estimator for Sharp Bounds

Input: data $\{S_i = (\mathbf{X}_i, Y_i, T_i, Z_i)\}_{i=1}^n$, target (t, z), bandwidth h (if Z is continuous), folds $\{\mathcal{I}_k\}_{k=1}^K$, nuisance estimators, regression learner $\widehat{\mathbb{E}}_n$
for $k = 1, \ldots, K$ **do**
  Fit nuisances $\widehat{\eta}^{(-k)}$ on $\{S_i : i \notin \mathcal{I}_k\}$
  for $i \in \mathcal{I}_k$ **do**
    $\widehat{\phi}^+_{t,z,i} \leftarrow \phi^+_{t,z}(S_i; \widehat{\eta}^{(-k)})$
  end for
end for
$\widehat{\psi}^+(t, z) \leftarrow \frac{1}{n} \sum_{i=1}^n \widehat{\phi}^+_{t,z,i}$
$\widehat{\mu}^+(t, z, \mathbf{x}) \leftarrow \widehat{\mathbb{E}}_n[\phi^+_{t,z}(S; \widehat{\eta}) \mid \mathbf{X} = \mathbf{x}]$
Output: $\widehat{\mu}^+(t, z, \cdot)$, $\widehat{\psi}^+(t, z)$

Moreover, when $\widehat{\eta} = \eta$, the pseudo-outcome is unbiased for its target bound functional.

Remark 9.12 (Unbiasedness of the pseudo-outcome). When $\widehat{\eta} = \eta$ and Z is discrete, we have

$$\mathbb{E}\left[\phi^+_{t,z}(S; \eta) \mid \mathbf{X} = \mathbf{x}\right] = \mu^+(t, z, \mathbf{x}), \quad \mathbb{E}\left[\phi^+_{t,z}(S; \eta)\right] = \psi^+(t, z) \tag{9.36}$$

When Z is continuous, the kernel-localized pseudo-outcome targets a bandwidth indexed functional (μ^+_h, ψ^+_h). Under standard smoothness in z, $\mu^+_h(t, z, \mathbf{x}) \to \mu^+(t, z, \mathbf{x})$ and $\psi^+_h(t, z) \to \psi^+(t, z)$ as $h \downarrow 0$.

9.5.2 Estimation Strategy

Motivated by Theorem 9.11, we estimate the bounds using a *two-stage procedure* (Algorithm 9.1). We first estimate the nuisance functions, then evaluate the orthogonal pseudo-outcome, and finally estimate the bound functional by either regression or sample averaging. Specifically, we obtain the CAPO upper bound

$$\widehat{\mu}^+(t, z, \mathbf{x}) := \widehat{\mathbb{E}}_n[\phi^+_{t,z}(S; \widehat{\eta}) \mid \mathbf{X} = \mathbf{x}] \tag{9.37}$$

by regressing the pseudo-outcome on $\mathbf{X}$, and the APO upper bound

$$\widehat{\psi}^+(t, z) := \widehat{\mathbb{E}}_n[\phi^+_{t,z}(S; \widehat{\eta})] \tag{9.38}$$

by sample averaging.

To mitigate overfitting bias and enable standard orthogonalization guarantees, we use K-fold cross-fitting [Chernozhukov et al., 2018a]. Each pseudo-outcome $\widehat{\phi}^{+}_{t,z,i}$ is computed using nuisance estimates trained on data that do not contain observation i.

9.6 Theoretical Guarantees

Theorem 9.9 establishes that the identified CAPO bounds $\mu^{\pm}(t, z, \mathbf{x})$ are sharp. We now summarize three additional guarantees for the orthogonal estimator from Theorem 9.11. First, orthogonality yields second-order sensitivity to nuisance estimation error, implying quasi-oracle rates for the CAPO bounds and, when Z is discrete, root-n inference for the APO bounds. Second, if $Q^{\pm}$ is consistently estimated and either the propensity block (π^t, π^g) or the moment block $(\gamma^{\pm}_u, \gamma^{\pm}_l)$ is consistently estimated, then the estimated endpoints converge to the sharp bounds. Third, even when $Q^{\pm}$ is misspecified, the resulting intervals remain asymptotically valid, though potentially conservative, provided one nuisance block is learned consistently. We present the discrete-Z results in the main text and defer the continuous-Z case to Appendix H.2.2. All proofs are deferred to Appendix H.3.

9.6.1 Second-order Robustness to Nuisance Estimation

The key advantage of orthogonalization is that nuisance estimation errors enter only through second-order products. We formalize this under the following regularity condition.

Assumption 9.13 (Regularity and overlap)**.** There exist constants $\varepsilon > 0$ and $M < \infty$ such that, almost surely:

(i) $\varepsilon \le \pi^t(\mathbf{X}), \widehat{\pi}^t(\mathbf{X}) \le 1 - \varepsilon$;

(ii) if Z is discrete, then $\varepsilon \le \pi^g(z \mid \mathbf{X}), \widehat{\pi}^g(z \mid \mathbf{X})$ for all $(z, \mathbf{X})$; if Z is continuous, then there exists a neighborhood $\mathcal{N}_z$ of z such that for all $u \in \mathcal{N}_z$,

$$\varepsilon \le \pi^g(u \mid \mathbf{X}), \widehat{\pi}^g(u \mid \mathbf{X}) \le M; \tag{9.39}$$

(iii) $|Y|$, $|\widehat{\gamma}_u^{\pm}|$, $|\widehat{\gamma}_l^{\pm}|$, and $|\widehat{Q}^{\pm}|$ are all bounded by M.

Theorem 9.14. *Assume Z is discrete and Assumption 9.13 holds. Let $\widehat{\eta} = (\widehat{\pi}^t, \widehat{\pi}^g, \widehat{Q}^+, \widehat{\gamma}_u^+, \widehat{\gamma}_l^+)$ be the cross-fitted nuisances used in $\phi_{t,z}^+(S; \widehat{\eta})$ from Theorem 9.11. Define $r_{n,\pi} := \|\widehat{\pi}^t - \pi^t\|_2 + \|\widehat{\pi}^g - \pi^g\|_2$, $r_{n,Q} := \|\widehat{Q}^+ - Q^+\|_2$, and $r_{n,\gamma} := \|\widehat{\gamma}_u^+ - \gamma_u(\widehat{Q}^+; \cdot)\|_2 + \|\widehat{\gamma}_l^+ - \gamma_l(\widehat{Q}^+; \cdot)\|_2$, where*

$$\gamma_u(\widehat{Q}^+; \mathbf{X}) := \mathbb{E}[(Y - \widehat{Q}^+(\mathbf{X}))_+ \mid T = t, Z = z, \mathbf{X}], \tag{9.40}$$

$$\gamma_l(\widehat{Q}^+; \mathbf{X}) := \mathbb{E}[(\widehat{Q}^+(\mathbf{X}) - Y)_+ \mid T = t, Z = z, \mathbf{X}]. \tag{9.41}$$

Then

$$\left\| \mathbb{E}\left[\phi_{t,z}^+(S; \widehat{\eta}) - \phi_{t,z}^+(S; \eta) \mid \mathbf{X} \right] \right\|_2 = O_p\left(r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2 \right). \tag{9.42}$$

Thus, the nuisance contribution enters only through second-order terms. To translate this into a rate for the final CAPO estimator, we impose a generic condition on the second-stage regression learner.

Assumption 9.15 (Second-stage Regression Rate)**.** Fix (t, z) and let $\widehat{\phi}_{t,z,i}^+$ denote the cross-fitted pseudo-outcome. Let $m_{t,z}^+(\mathbf{x}) := \mathbb{E}[\widehat{\phi}_{t,z}^+ \mid \mathbf{X} = \mathbf{x}]$. Assume the regression learner used to form $\widehat{\mu}^+(t, z, \cdot)$ satisfies

$$\|\widehat{\mu}^+(t, z, \cdot) - m_{t,z}^+(\cdot)\|_2 = O_p(\delta_n) \tag{9.43}$$

for some possibly model-dependent rate δ_n.

Remark 9.16 (Second-Stage Regression Assumption). Assumption 9.15 treats the final-stage regression step as a black box: it assumes that, when regressing the cross-fitted pseudo-outcomes on $\mathbf{X}$, the learner attains an L_2 error rate δ_n uniformly over the admissible nuisance estimates $\widehat{\eta} \in \Xi$. A broad class of learners satisfy this, including nonparametric least-squares/ERM estimators over a bounded function class $\mathcal{F}$ with bracketing entropy $\log N_{[]}(\mathcal{F}, \epsilon) \lesssim \epsilon^{-r}$ $(0 < r < 2)$, which yields the usual regression rate $\delta_n \asymp n^{-1/(2+r)}$ (up to approximation error); in particular, for d-dimensional Hölder(β) classes, $\delta_n = n^{-\beta/(2\beta+d)}$. More generally, black-box regressors satisfying standard stability/oracle-inequality properties (e.g., linear smoothers) also fit this template [Kennedy, 2023b]. We therefore state our results in terms of δ_n, which separates orthogonalization from the choice of final-stage regression method.

Corollary 9.17. *Suppose Assumptions 9.13 and 9.15 hold, and let $r_{n,\pi}, r_{n,\gamma}, r_{n,Q}$ be as in Theorem 9.14. For the CAPO upper-bound estimator,*

$$\|\widehat{\mu}^+(t,z,\cdot) - \mu^+(t,z,\cdot)\|_2 = O_p\Big(\delta_n + r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2\Big). \tag{9.44}$$

In particular, if $r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2 = o_p(\delta_n)$, then $\|\widehat{\mu}^+(t,z,\cdot) - \mu^+(t,z,\cdot)\|_2 = O_p(\delta_n)$.
For the APO upper-bound estimator $\widehat{\psi}^+(t,z) = \mathbb{E}_n[\widehat{\phi}^+_{t,z}]$,

$$\left|\widehat{\psi}^+(t,z) - \psi^+(t,z)\right| = O_p\Big(n^{-1/2} + r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2\Big). \tag{9.45}$$

If moreover $r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2 = o_p(n^{-1/2})$, then

$$\sqrt{n}\left(\widehat{\psi}^+(t,z) - \psi^+(t,z)\right) \rightsquigarrow \mathcal{N}(0, V^+(t,z)), \tag{9.46}$$

where $V^+(t,z) := \mathrm{Var}(\phi^+_{t,z}(S;\eta))$.

Corollary 9.17 establishes a *quasi-oracle* property: if the nuisance estimators converge at rate $o_p(\delta_n^{1/2})$ for CAPO bounds (or $o_p(n^{-1/4})$ for APO), the estimator

achieves the oracle rate $O_p(\delta_n)$, as if the nuisances were known. This follows, since nuisance errors enter only through the second-order remainder $r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2$. For APOs, this additionally enables valid and tight inference.

9.6.2 Sharpness of the Estimated Bounds

The previous subsection controlled the effect of nuisance estimation error on the estimator. We now give conditions under which the estimated intervals converge to the sharp identified bounds from Theorem 9.9.

Proposition 9.18. *Assume the conditions of Corollary 9.17 hold. Suppose $\delta_n = o_p(1)$ and $r_{n,Q} = o_p(1)$, and in addition either $r_{n,\pi} = o_p(1)$ or $r_{n,\gamma} = o_p(1)$. Then*

$$\|\widehat{\mu}^{\pm}(t, z, \cdot) - \mu^{\pm}(t, z, \cdot)\|_2 = o_p(1) \tag{9.47}$$

and

$$|\widehat{\psi}^{\pm}(t, z) - \psi^{\pm}(t, z)| = o_p(1). \tag{9.48}$$

Consequently, the estimated CAPO and APO intervals converge to the sharp identified intervals.

Proposition 9.18 shows that consistency of the cutoff together with consistency of either the propensity block or the conditional-moment block is sufficient for the orthogonal estimator to recover the sharp bounds.

9.6.3 Validity Under Misspecified Cutoffs

Sharpness requires consistent estimation of $Q^{\pm}$. We now show that even when the cutoff is misspecified, the resulting intervals remain asymptotically valid, though potentially conservative, provided that one nuisance block is consistently estimated.

Corollary 9.19. *Assume the conditions of Corollary 9.17 hold. Let $\overline{Q}^{\pm}(t, z, \mathbf{x})$ be any measurable cutoff and define the induced bounds*

$$\overline{\mu}^{\pm}(t, z, \mathbf{x}; \overline{Q}^{\pm}) = \overline{Q}^{\pm}(t, z, \mathbf{x}) + \frac{1}{b^{\mp}(z, \mathbf{x})}\mathbb{E}\left[(Y - \overline{Q}^{\pm}(t, z, \mathbf{x}))_+ \mid t, z, \mathbf{x}\right] - \frac{1}{b^{\pm}(z, \mathbf{x})}\mathbb{E}\left[(\overline{Q}^{\pm}(t, z, \mathbf{x}) - Y)_+ \mid t, z, \mathbf{x}\right], \tag{9.49}$$

and analogously define $\overline{\psi}^{\pm}(t, z) := \mathbb{E}[\overline{\mu}^{\pm}(t, z, \mathbf{X})]$. Then

$$[\overline{\mu}^{-}(t, z, \mathbf{x}), \overline{\mu}^{+}(t, z, \mathbf{x})] \tag{9.50}$$

is a valid, though not necessarily sharp, CAPO interval, and likewise

$$[\overline{\psi}^{-}(t, z), \overline{\psi}^{+}(t, z)] \tag{9.51}$$

is a valid APO interval. Moreover, if $\widehat{Q}^{\pm} \to \overline{Q}^{\pm}$ in L_2 and either (i) $(\widehat{\pi}^t, \widehat{\pi}^s)$ is consistent, or (ii) $(\widehat{\gamma}_u^{\pm}, \widehat{\gamma}_l^{\pm})$ is consistent for the tail-moment targets induced by $\overline{Q}^{\pm}$, then the resulting estimated CAPO and APO intervals converge to

$$[\overline{\mu}^{-}, \overline{\mu}^{+}] \quad \textit{and} \quad [\overline{\psi}^{-}, \overline{\psi}^{+}], \tag{9.52}$$

respectively, and are asymptotically valid, though potentially conservative. If $\overline{Q}^{\pm} = Q^{\pm}$, then the bounds coincide with the sharp bounds.

Remark 9.20 (Continuous Z). When Z is continuous, evaluation at a point z requires kernel localization, leading to the usual bias-variance tradeoff in the bandwidth. We defer the corresponding rates and inference results to Appendix H.2.2.

9.7 Conclusion

This chapter develops a partial identification framework for causal inference on networks when the exposure mapping may be misspecified. The central

point is that the choice of exposure mapping is itself a consequential modeling assumption, and when it is uncertain, inference should reflect that uncertainty rather than rely on brittle point identification.

By modeling misspecification through sensitivity bounds on the induced exposure propensity, we derived sharp identified intervals for conditional and average potential outcomes, and hence for direct, spillover, and overall treatment effects. We further developed an orthogonal estimation strategy for these bounds and summarized the resulting guarantees on robustness to nuisance estimation error, sharpness, and asymptotic validity.

Within the broader dissertation, this chapter extends the theme of reliable causal inference under unreliable assumptions to network settings, where the key fragility lies not only in confounding or overlap assumptions, but also in how interference itself is represented.

Part IV

Appendix

APPENDIX A

APPENDIX FOR CHAPTER 2

A.1 Proofs of Main Theorems

A.1.1 Proof of Theorem 2.6

We first note the following useful identities:

1. (Functional analog of Taylor's expansion - 2nd order). Let $F : \mathcal{F}^d \to \mathbb{R}$, where $\mathcal{F}$ is a vector space of functions. For any $h, h' \in \mathcal{F}^d$, assume $t \mapsto F(t \cdot h + (1-t) \cdot h')$ has second order derivatives in an open interval containing $[0, 1]$, then $\exists\, \overline{h} \in \text{conv}(\{h, h'\})$ such that

$$F(h') = F(h) + \sum_{i=1}^{d} D_{h_i} F(h)(h'_i - h_i) + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} D_{h_i} D_{h_j} F(\overline{h})(h'_i - h_i)(h'_j - h_j)$$

2. For all $i \in \overline{1, m+1}, a \in \{0, 1\}, x \in \mathcal{X}$, the following hold:

$$\mathbb{E}[\rho_i(Y, \nu_a^*) \mid X = x, A = a] = 0 \qquad \text{(Moment equations)}$$

$$\sum_{i=1}^{m+1} \alpha_{a,i}(x, \nu_a^*) D_{\nu_{a,j}} \mathbb{E}[\rho_i(Y, \nu_a^*) \mid X = x, A = a] = \mathbb{I}[j = 1] \qquad \text{(Definition of } \alpha\text{)}$$

Proof. We wish to bound the term:

$$\mathcal{E}(e, \alpha, \nu) = \|\mathbb{E}[\psi(Z, e, \alpha, \nu) \mid X = x] - \mathbb{E}[\psi(Z, e^*, \alpha^*, \nu^*) \mid X = x]\|$$

To simplify our calculations, we write ψ as a difference of two terms $\psi^1 - \psi^0$ given by:

$$\psi^1(Z, e, \alpha, \nu) = \kappa_1(X) - \frac{A}{e(X)} \sum_{i=1}^{m+1} \alpha_{1,i}(X, \nu_{1,i}) \rho(Y, \nu_{1,i})$$

$$\psi^0(Z, e, \alpha, \nu) = \kappa_0(X) - \frac{1 - A}{1 - e(X)} \sum_{i=1}^{m+1} \alpha_{1,i}(X, \nu_{0,i}) \rho(Y, \nu_{0,i})$$

With this notation, we have:

$$\mathcal{E}(e,\alpha,\nu) \lesssim \sum_{a=0}^{1} \left\| \mathbb{E}\left[\psi^a(Z,e,\alpha,\nu) - \psi^a(Z,e^*,\alpha^*,\nu^*) \mid X = x\right] \right\|$$

The inequality above follows from the inequalities $(a+b)^2 \leq 2(a^2+b^2)$ and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ (for non-negative a, b). Without loss of generality, we consider the bound for ψ^1. The proof for ψ^0 is very similar. We have:

$$\begin{aligned}
&\mathbb{E}\Big[\psi^1(Z,e,\alpha,\nu) - \psi^1(Z,e^*,\alpha^*,\nu^*) \mid X = x\Big] \\
&\quad = \sum_{a=0}^{1} \mathbb{E}\Big[\psi^1(Z,e,\alpha,\nu) - \psi^1(Z,e^*,\alpha^*,\nu^*) \mid X = x, A = a\Big] P(A = a \mid X = x) \\
&\quad = \kappa_1(x) - \kappa_1^*(x) - \frac{e^*(x)}{e(x)} \sum_{i=1}^{m+1} \alpha_{1,i}(x,\nu_1) \mathbb{E}[\rho_i(Y,\nu_1) \mid X = x, A = 1] \\
&\quad = \kappa_1(x) - \kappa_1^*(x) - \underbrace{\frac{e^*(x)}{e(x)} \sum_{i=1}^{m+1} (\alpha_{1,i}(x,\nu_1) - \alpha_{1,i}^*(x,\nu_1^*)) \mathbb{E}[\rho_i(Y,\nu_1) \mid X = x, A = 1]}_{\Lambda_1} \\
&\qquad\qquad - \underbrace{\frac{e^*(x)}{e(x)} \sum_{i=1}^{m+1} \alpha_{1,i}^*(x,\nu_1^*) \mathbb{E}[\rho_i(Y,\nu_1) \mid X = x, A = 1]}_{\Lambda_2}
\end{aligned}$$

where in the last equality we subtracted and then added back a term. We now study Λ_1 and Λ_2. For both these quantities, we do a Taylor expansion of $\mathbb{E}[\rho_i(Y,\nu_1) \mid X = x, A = 1]$ around ν_1^*. For Λ_1, it suffices to use the Taylor expansion to first order only. For Λ_2, we perform a second order expansion:

$$\begin{aligned}
&\mathbb{E}[\rho_i(Y,\nu_1) \mid X = x, A = 1] \\
&= \mathbb{E}[\rho_i(Y,\nu_1^*) \mid X = x, A = 1] + \sum_{j=1}^{m+1} D_{\nu_j} \mathbb{E}[\rho_i(Y,\overline{\nu}) \mid X = x, A = 1](\nu_{1,j}(x) - \nu_{1,j}^*(x))
\end{aligned}$$

(first order expansion)

$$\begin{aligned}
&\mathbb{E}[\rho_i(Y,\nu_1) \mid X = x, A = 1] \\
&= \mathbb{E}[\rho_i(Y,\nu_1^*) \mid X = x, A = 1] + \sum_{j=1}^{m+1} D_{\nu_j} \mathbb{E}[\rho_i(Y,\nu_1^*) \mid X = x, A = 1](\nu_{1,j}(x) - \nu_{1,j}^*(x))
\end{aligned}$$

$$+\frac{1}{2}\sum_{j=1}^{m+1}\sum_{l=1}^{m+1} D_{\nu_j} D_{\nu_l} \mathbb{E}[\rho_i(Y,\overline{\nu}) \mid X=x, A=1](\nu_{1,j}(x)-\nu^*_{1,j}(x))(\nu_{1,l}(x)-\nu^*_{1,l}(x))$$

(second order expansion)

Then Λ_1 is given by:

$$\begin{aligned}
&\Lambda_1 \\
&= \frac{e^*(x)}{e(x)}\sum_{i=1}^{m+1}(\alpha_{1,i}(x,\nu_1)-\alpha^*_{1,i}(x,\nu^*_1))\mathbb{E}[\rho_i(Y,\nu_1) \mid X=x, A=1] \\
&= \frac{e^*(x)}{e(x)}\sum_{i=1}^{m+1}(\alpha_{1,i}(x,\nu_1)-\alpha^*_{1,i}(x,\nu^*_1))\Big(\cancelto{0}{\mathbb{E}[\rho_i(Y,\nu^*_1) \mid X=x, A=1]} \\
&\qquad\qquad + \sum_{j=1}^{m+1} D_{\nu_j}\mathbb{E}[\rho_i(Y,\overline{\nu}) \mid X=x, A=1](\nu_{1,j}(x)-\nu^*_{1,j}(x))\Big)
\end{aligned}$$

(first order expansion)

$$\begin{aligned}
&= \frac{e^*(x)}{e(x)}\sum_{i=1}^{m+1}\sum_{j=1}^{m+1} D_{\nu_j}\mathbb{E}[\rho_i(Y,\overline{\nu}) \mid X=x, A=1](\alpha_{1,i}(x,\nu_1)-\alpha^*_{1,i}(x,\nu^*_1))(\nu_{1,j}(x)-\nu^*_{1,j}(x)) \\
\|\Lambda_1\| &\le \frac{c_2}{c_1}\sum_{i=1}^{m+1}\sum_{j=1}^{m+1} G_{ij}\|\alpha_{1,i}-\alpha^*_{1,i}\|\,\|\nu_{1,j}-\nu^*_{1,j}\| \\
&\lesssim \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} G_{ij}\|\alpha_{1,i}-\alpha^*_{1,i}\|\,\|\nu_{1,j}-\nu^*_{1,j}\|
\end{aligned}$$

We proceed in a similar way for Λ_2, using the second order expansion:

$$\begin{aligned}
\Lambda_2 &= \frac{e^*(x)}{e(x)}\sum_{i=1}^{m+1}\alpha^*_{1,i}(x,\nu^*_1)\mathbb{E}[\rho_i(Y,\nu_1) \mid X=x, A=1] \\
&= \frac{e^*(x)}{e(x)}\sum_{i=1}^{m+1} \\
&\quad \alpha^*_{1,i}(x,\nu^*_1)\Big(\cancelto{0}{\mathbb{E}[\rho_i(Y,\nu^*_1) \mid X=x, A=1]} \\
&\qquad + \sum_{j=1}^{m+1} D_{\nu_j}\mathbb{E}[\rho_i(Y,\nu^*_1) \mid X=x, A=1](\nu_{1,j}(x)-\nu^*_{1,j}(x)) \\
&\qquad + \frac{1}{2}\sum_{j=1}^{m+1}\sum_{l=1}^{m+1} D_{\nu_j}D_{\nu_l}\mathbb{E}[\rho_i(Y,\overline{\nu}_1) \mid X=x, A=1](\nu_{1,j}(x)-\nu^*_{1,j}(x))(\nu_{1,l}(x)-\nu^*_{1,l}(x))\Big)
\end{aligned}$$

$$= \frac{e^*(x)}{e(x)} \sum_{j=1}^{m+1} \underbrace{\sum_{i=1}^{m+1} \alpha^*_{1,i}(x, \nu^*_1) D_{\nu_j} \mathbb{E}[\rho_i(Y, \nu^*_1) \mid X = x, A = 1]}_{=\mathbb{I}[j=1]} (\nu_{1,j}(x) - \nu^*_{1,j}(x))$$

(see useful identities)

$$+ \frac{1}{2} \frac{e^*(x)}{e(x)} \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \sum_{l=1}^{m+1} \Bigg(\alpha^*_{1,i}(x, \nu^*_1) D_{\nu_j} D_{\nu_l} \mathbb{E}[\rho_i(Y, \overline{\nu}_1) \mid X = x, A = 1] \cdot (\nu_{1,j}(x) - \nu^*_{1,j}(x))(\nu_{1,l}(x) - \nu^*_{1,l}(x)) \Bigg)$$

$$= \frac{e^*(x)}{e(x)} (\kappa_1(x) - \kappa^*_1(x)) + \frac{1}{2} \frac{e^*(x)}{e(x)} \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \sum_{l=1}^{m+1} \Bigg(\alpha^*_{1,i}(x, \nu^*_1) D_{\nu_j} D_{\nu_l} \mathbb{E}[\rho_i(Y, \overline{\nu}_1) \mid X = x, A = 1] \cdot (\nu_{1,j}(x) - \nu^*_{1,j}(x))(\nu_{1,l}(x) - \nu^*_{1,l}(x)) \Bigg)$$

Adding back the κ_1 terms to Λ_2, we have:

$$\kappa_1(x) - \kappa_1(x) - \Lambda_2 = \frac{1}{e(x)} (e(x) - e^*(x))(\kappa_1(x) - \kappa^*_1(x)) + \frac{1}{2} \frac{e^*(x)}{e(x)} \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \sum_{l=1}^{m+1} \Bigg(\alpha^*_{1,i}(x, \nu^*_1) D_{\nu_j} D_{\nu_l} \mathbb{E}[\rho_i(Y, \overline{\nu}_1) \mid X = x, A = 1] \cdot (\nu_{1,j}(x) - \nu^*_{1,j}(x))(\nu_{1,l}(x) - \nu^*_{1,l}(x)) \Bigg)$$

$$\|\kappa_1(x) - \kappa_1(x) - \Lambda_2\| \leq \frac{1}{c_1} \|\kappa_1 - \kappa^*_1\| \|e - e^*\| + \frac{c_3 c_4}{2 c_1} \sum_{j=1}^{m+1} \sum_{l=1}^{m+1} H_{jl} \|\nu_{1,j} - \nu^*_{1,j}\| \, \|\nu_{1,l} - \nu^*_{1,l}\|$$

$$\lesssim \|\kappa_1 - \kappa^*_1\| \|e - e^*\| + \sum_{j=1}^{m+1} \sum_{l=1}^{m+1} H_{jl} \|\nu_{1,j} - \nu^*_{1,j}\| \, \|\nu_{1,l} - \nu^*_{1,l}\|$$

Constants c_1, c_2, c_3, c_4 and binary matrices H, G are defined in Assumption 2.5.

Putting everything together, we have:

$$\left\| \mathbb{E}\left[\psi^1(Z, e, \alpha, \nu) - \psi^1(Z, e^*, \alpha^*, \nu^*) \mid X = x \right] \right\|$$

$$= \|\kappa_1(x) - \kappa_1(x) - \Lambda_2 - \Lambda_1\|$$

$$\lesssim \|\kappa_1(x) - \kappa_1(x) - \Lambda_2\| + \|\Lambda_1\|$$

$$\lesssim \|\kappa_1 - \kappa_1^*\|\|e - e^*\| + \sum_{j=1}^{m+1}\sum_{l=1}^{m+1} H_{jl}\|\nu_{1,j} - \nu_{1,j}^*\|\,\|\nu_{1,l} - \nu_{1,l}^*\| + \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} G_{ij}\|\alpha_{1,i} - \alpha_{1,i}^*\|\,\|\nu_{1,j} - \nu_{1,j}^*\|$$

Putting the ψ^0 and ψ^1 bounds together, we obtain the desired result:

$$\mathcal{E}(e,\alpha,\nu) \lesssim \sum_{a=1}^{1}\left(\|\kappa_a - \kappa_a^*\|\,\|e - e^*\| + \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} G_{ij}\|\alpha_{a,i} - \alpha_{a,i}^*\|\,\|\nu_{a,j} - \nu_{a,j}^*\| + \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} H_{ij}\|\nu_{a,i} - \nu_{a,i}^*\|\,\|\nu_{a,j} - \nu_{a,j}^*\|\right)$$

□

A.1.2 Proof of Theorem 2.9

Proof. Let $I_k = \{i : i = k - 1 \pmod K\}$ and $I_k^C = \{i : i \neq k - 1 \pmod K\}$ be the k^{th} data fold and its complement, let $n_k = |I_k| \in \{\lfloor n/K \rfloor, \lceil n/K \rceil\}$, and let $\lambda_k = n_k/n$ (note none of I_k, n_k, λ_k are random). Let $\mathbb{P}g(Z)$ denote expectation over Z alone (that is, conditioning on the data, should g depend on it) and $\hat{\mathbb{P}}_k g(Z)$ denote empirical expectation over I_k. Further, let $\mathbb{P}(g(Z) \mid X)$ denote the conditional expectation, $\|g\|_2 = \mathbb{P}g^2$, and $\Pi_{\mathcal{F}}(g) \in \operatorname{argmin}_{f\in\mathcal{F}} \|f - g\|_2$.

Define the pseudo-outcome random variables:

$$\psi = \psi(Z, e^*, \alpha^*, \nu^*),$$
$$\hat{\psi}_k = \psi(Z, \hat{e}^{(k)}, \hat{\alpha}^{(k)}, \hat{\nu}^{(k)}),$$
$$\hat{\psi} = \sum_{k=1}^{K} \lambda_k \hat{\psi}_k.$$

Further, define the squared-error objectives:

$$\hat{R}_k(f) = \hat{\mathbb{P}}_k(f(X) - \hat{\psi}_k)^2,$$
$$\hat{R}(f) = \sum_{k=1}^{K} \lambda_k \hat{R}_k(f),$$

$$\tilde{R}_k(f) = \mathbb{P}(f(X) - \hat{\psi}_k)^2,$$

$$\tilde{R}(f) = \sum_{k=1}^{K} \lambda_k \tilde{R}_k(f),$$

$$\bar{R}(f) = \mathbb{P}(f(X) - \hat{\psi})^2,$$

$$R(f) = \mathbb{P}(f(X) - \psi)^2.$$

Note that $\tilde{R}(f) - \bar{R}(f) = \tilde{R}(f') - \bar{R}(f')$ for any f, f', that is, $\tilde{R}(f)$ and $\bar{R}(f)$ are equal up to additive constants. Finally, we have the predictors:

$$\hat{f} \in \operatorname*{argmin}_{f \in \mathcal{F}} \hat{R}(f),$$

$$\bar{f} = \Pi_{\mathcal{F}}(\mathbb{P}(\hat{\psi} \mid X)) \in \operatorname*{argmin}_{f \in \mathcal{F}} \bar{R}(f) = \operatorname*{argmin}_{f \in \mathcal{F}} \tilde{R}(f),$$

$$f^* = \Pi_{\mathcal{F}}(\mathbb{P}(\psi \mid X)) \in \operatorname*{argmin}_{f \in \mathcal{F}} R(f).$$

Note $\hat{f} = \widehat{\text{CDTE}}$ and $f^* = \text{CDTE}$; we rename them for brevity.

We then have:

$$\begin{aligned}
\left\|\hat{f} - f^*\right\| &\le \left\|\hat{f} - \bar{f}\right\| + \left\|\bar{f} - f^*\right\| \\
&= \left\|\hat{f} - \bar{f}\right\| + \left\|\Pi_{\mathcal{F}}(\mathbb{P}(\hat{\psi} \mid X)) - \Pi_{\mathcal{F}}(\mathbb{P}(\psi \mid X))\right\| \\
&\le \underbrace{\left\|\hat{f} - \bar{f}\right\|}_{\text{Error}_A} + \underbrace{\left\|\mathbb{P}(\hat{\psi} \mid X) - \mathbb{P}(\psi \mid X)\right\|}_{\text{Error}_B},
\end{aligned}$$

where the inequality is because $\mathcal{F}$ is closed convex

We first tackle Error_B:

$$\begin{aligned}
\left\|\mathbb{P}(\hat{\psi} \mid X) - \mathbb{P}(\psi \mid X)\right\| &= \left\|\sum_{k=1}^{K} \lambda_k (\mathbb{P}(\hat{\psi}_k \mid X) - \mathbb{P}(\psi \mid X))\right\| \\
&\le \sum_{k=1}^{K} \lambda_k \left\|\mathbb{P}(\hat{\psi}_k \mid X) - \mathbb{P}(\psi \mid X)\right\|,
\end{aligned}$$

which we bound as $\lesssim \mathcal{E}$ by Theorem 2.6.

Next, we address Error_A. Note that $\left\|\hat{f} - \bar{f}\right\| \ge t$ means that $\tilde{R}(\hat{f}) - \tilde{R}(\bar{f}) \ge t^2$ and $\hat{R}(\hat{f}) - \hat{R}(\bar{f}) \le 0$. By intermediate value theorem and convexity of $\tilde{R}, \hat{R}, \mathcal{F}$, we

have for some $f \in [\hat{f}, \bar{f}]$ that $\tilde{R}(f) - \tilde{R}(\bar{f}) = t^2$ and $\hat{R}(f) - \hat{R}(\bar{f}) \leq 0$. This in turn implies that $\exists f \in \mathcal{F}$ with $\left\|f - \bar{f}\right\| \leq t$ and $(\tilde{R}(f) - \tilde{R}(\bar{f})) - (\hat{R}(f) - \hat{R}(\bar{f})) \geq t^2$. Because $\tilde{R} - \hat{R}$ is average of $\tilde{R}_k - \hat{R}_k$, this in turn implies that for at least one $k \in \overline{1, K}$ we have $\exists f \in \mathcal{F}$ with $\left\|f - \bar{f}\right\| \leq t$ and $(\tilde{R}_k(f) - \tilde{R}_k(\bar{f})) - (\hat{R}_k(f) - \hat{R}_k(\bar{f})) \geq t^2$. Since $|\psi| \leq c_5, \left|\hat{\psi}_k\right| \leq c_5, |f(x)| \leq c_5$, we have $(f(X) - \hat{\psi}_k)^2 - (\bar{f}(X) - \hat{\psi}_k)^2 \leq 4c_5 \left|f(X) - \bar{f}(X)\right|$. Now, consider $g(Z) = \frac{(f(X) - \bar{f}(X))(f(X) + \bar{f}(X) - 2\hat{\psi}_k)}{4c_5}$. Then $(\mathbb{P} - \hat{\mathbb{P}}_k)g \geq t^2/(4c_5)$ and $\|g\| \leq \|f - \bar{f}\| \leq t$ since $\|f(X) + \bar{f}(X) - 2\hat{\psi}_k\| \leq 4c_5$.

Let

$$\mathcal{G}_k = \left\{ \frac{(f - \bar{f})(f + \bar{f} - 2\hat{\psi}_k)}{4c_5} : \ f \in \mathcal{F} \right\}$$

and define the event

$$E_k(t) = \left\{ \exists g \in \mathcal{G}_k : \mathbb{P}g^2 \leq t^2, \ (\mathbb{P} - \hat{\mathbb{P}}_k)g \geq \frac{t^2}{4c_5} \right\}.$$

Therefore, by union bound and iterated expectations,

$$\mathbb{P}\left(\left\|\hat{f} - \bar{f}\right\| \geq t\right) \leq \sum_{k=1}^{K} \mathbb{E}\Big[\mathbb{P}\left(E_k(t) \mid \{Z_i : i \in I_k^C\}\right)\Big].$$

By lemma 3.4.2 of van der Vaart and Wellner [1996] and Markov's inequality, we have

$$\mathbb{P}\left(E_k(t) \mid \{Z_i : i \in I_k^C\}\right) \leq \frac{4c_5 J(1 + c_5 J/(\sqrt{n_k} t^2))}{t^2},$$

where $J = t + \int_0^t \sqrt{\epsilon^{-r}}\, d\epsilon \leq t + \frac{2}{2-r} t^{1-r/2}$. Thus, $\text{Error}_A = O_p(n^{-1/(2+r)})$. □

A.1.3 Proof of Theorem 2.10

Proof. Using the stability conditions for the estimator $\widehat{\mathbb{E}}_n$ outlined in Theorem 2.10, we can immediately apply Theorem 1 from Kennedy [2023a] (as appears in the v2 preprint on arXiv) with the cross-fitted nuisances to obtain:

$$\left\|\widehat{(\text{CDTE})} - (\text{CDTE})\right\|$$

$$
\begin{aligned}
&\lesssim \left\|\widehat{(\widetilde{\text{CDTE}})} - (\text{CDTE})\right\| \\
&\quad + \sum_{k=1}^{K} \left\|\mathbb{E}[\psi(Z, \widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) \mid X = x] - \mathbb{E}[\psi(Z, e^*, \alpha^*, \nu^*) \mid X = x]\right\| \\
&\lesssim \left\|\widehat{(\widetilde{\text{CDTE}})} - (\text{CDTE})\right\| + \mathcal{E}
\end{aligned}
$$

where the last inequality was obtained by applying Theorem 2.6 to nuisance sets $(\widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)})$ in each fold $k \in \overline{1, K}$. □

A.1.4 Proof of Theorem 2.11

Proof. Let $\mathcal{I}_k = \{i : i = k - 1 \pmod K\}$ be the data indices in the k-th data fold, and let $\widehat{\mathbb{E}}_k f(Z) = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} f(Z_i)$ be the empirical expectation of data with indices in $\mathcal{I}_k$. The linear regression parameters $\gamma^*, \widetilde{\gamma}$ and $\widehat{\gamma}$ are given by:

$$
\begin{aligned}
\gamma^* &= \mathbb{E}[\phi(X)\phi(X)^T]^{-1}\mathbb{E}[\text{CDTE}(X)\phi(X)] \\
&= \mathbb{E}[\phi(X)\phi(X)^T]^{-1}\mathbb{E}[\psi(Z, e^*, \alpha^*, \nu^*)\phi(X)] \\
&\qquad\qquad \text{(consistency and iterated expectations)} \\
\widetilde{\gamma} &= \widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1}\widehat{\mathbb{E}}_n[\psi(Z, e^*, \alpha^*, \nu^*)\phi(X)] \\
&= \widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1}\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k[\psi(Z, e^*, \alpha^*, \nu^*)\phi(X)] \\
\widehat{\gamma} &= \widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1}\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k[\psi(Z, \widehat{e}^{(k)}, {}^{(k)}, \widehat{\nu}^{(k)})\phi(X)]
\end{aligned}
$$

We note that we can rewrite $\sqrt{n}(\widehat{\gamma} - \gamma^*)$ as:

$$
\sqrt{n}(\widehat{\gamma} - \gamma^*) = \sqrt{n}(\widehat{\gamma} - \widetilde{\gamma}) + \underbrace{\sqrt{n}(\widetilde{\gamma} - \gamma^*)}_{\rightsquigarrow \mathcal{N}(0, \Sigma^*)} \tag{A.1}
$$

The second term converges in distribution to the desired $\mathcal{N}(0, \Sigma^*)$. Thus, it suffices to show that the first term is $o_p(1)$. We decompose the first term into two components and study them separately:

$$
\sqrt{n}(\widehat{\gamma} - \widetilde{\gamma}) \tag{A.2}
$$

$$= \sqrt{n}\widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1}\frac{1}{K}\sum_{k=1}^{K}\left(\widehat{\mathbb{E}}_k[\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)})\phi(X)] - \widehat{\mathbb{E}}_k[\psi(Z,e^*,\alpha^*,\nu)\phi(X)]\right)$$

$$= \sqrt{n}\widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1}\frac{1}{K}\sum_{k=1}^{K}\underbrace{\left(\mathbb{E}[\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)})\phi(X)] - \mathbb{E}[\psi(Z,e^*,\alpha^*,\nu^*)\phi(X)]\right)}_{\Lambda_{1,k}}$$

$$+ \sqrt{n}\widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1}\frac{1}{K}\sum_{k=1}^{K}\underbrace{(\widehat{\mathbb{E}}_k - \mathbb{E})\left[(\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)} - \psi(Z,e^*,\alpha^*,\nu^*))\phi(X)\right]}_{\Lambda_{2,k}}$$

We now show that $\Lambda_{1,k}$ and $\Lambda_{2,k}$ in the equation above are both $o_p(1/\sqrt{n})$. Let $\Sigma_\phi = \mathbb{E}[\phi(X)\phi(X)^T]$. Then, each element of $\Lambda_{1,k}$ is given by:

$$\begin{aligned}
(\Lambda_{1,k})_i &= \mathbb{E}[\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)})(\phi(X))_i] - \mathbb{E}[\psi(Z,e^*,\alpha^*,\nu^*)\phi(X)_i] \\
&= \mathbb{E}[\mathbb{E}[\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)}) - \psi(Z,e^*,\alpha^*,\nu^*) \mid X]\phi(X)_i] \\
\Rightarrow |(\Lambda_{1,k})_i| &\leq \|\mathbb{E}[\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)}) - \psi(Z,e^*,\alpha^*,\nu^*) \mid X]\| \, \|\phi(X)_i\| \qquad \text{(CS)} \\
&\lesssim \left\{ \sum_{a=1}^{1}\left(\|\widehat{\kappa}_a^{(k)} - \kappa_a^*\| \, \|\widehat{e}^{(k)} - e^*\| + \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} G_{ij}\|\widehat{\alpha}_{a,i}^{(k)} - \alpha_{a,i}^*\| \, \|\widehat{\nu}_{a,j}^{(k)} - \nu_{a,j}^*\| \right.\right. \\
&\qquad \left.\left. + \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} H_{ij}\|\widehat{\nu}_{a,i}^{(k)} - \nu_{a,i}^*\| \, \|\widehat{\nu}_{a,j}^{(k)} - \nu_{a,j}^*\| \right)\right\} \sqrt{(\Sigma_\phi)_{ii}} \qquad \text{(Thm. 2.6)}
\end{aligned}$$

By the theorem's assumptions, we have that $\Lambda_{1,k}$ is $o_p(1/\sqrt{n})$, as desired. We now see how we can control $\Lambda_{2,k}$. By Chebyshev's inequality, $\Lambda_{2,k}$ is

$$O_p\left(n^{-1/2}\sum_{i=1}^{p}\mathbb{E}[(\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)}) - \psi(Z,e^*,\alpha^*,\nu^*))^2(\phi(X)_i)^2]^{1/2}\right) \qquad \text{(A.3)}$$

where we leveraged that $|\mathcal{I}_k| \simeq n/K$ and K is a fixed integer constant that doesn't depend on n. The sum in the expression above further reduces to:

$$\begin{aligned}
&\sum_{i=1}^{p}\mathbb{E}[(\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)}) - \psi(Z,e^*,\alpha^*,\nu^*))^2(\phi(X))_i^2]^{1/2} \\
&\qquad \leq \sum_{i=1}^{p}\|\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)}) - \psi(Z,e^*,\alpha^*,\nu^*)\|\|\phi(X)_i\| \\
&\qquad = \sum_{i=1}^{p}\sqrt{(\Sigma_\phi)_{ii}}\,\|\psi(Z,\widehat{e}^{(k)},\widehat{\alpha}^{(k)},\widehat{\nu}^{(k)}) - \psi(Z,e^*,\alpha^*,\nu^*)\|
\end{aligned}$$

We now study the convergence of $\|\psi(Z, \widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi(Z, e^*, \alpha^*, \nu^*)\|$. We have:

$$\begin{aligned}
\|\psi(Z, \widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi(Z, e^*, \alpha^*, \nu^*)\| &= \sum_{a=0}^{1} \|\psi^a(Z, \widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi^a(Z, e^*, \alpha^*, \nu^*)\| \\
&\leq \sum_{a=0}^{1} \Big(\|\psi^a(Z, \widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi^a(Z, e^*, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)})\| \\
&\qquad + \|\psi^a(Z, e^*, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi^a(Z, e^*, \alpha^*, \nu^*)\| \Big) \\
&\qquad \text{(Cauchy-Schwartz)}
\end{aligned}$$

where the ψ^a's are defined in the proof of Theorem 2.6. Without loss of generality, we study the convergence of ψ^1:

$$\begin{aligned}
&\|\psi^1(Z, \widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi^1(Z, e^*, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)})\| \\
&\quad = \left\| \frac{A(\widehat{e}^{(k)}(X) - e^*(X))}{e^*(X)\widehat{e}^{(k)}(X)} \sum_{i=1}^{m+1} \widehat{\alpha}_{1,i}^{(k)}(X, \widehat{\nu}_1^{(k)}) \rho_i(Y, \widehat{\nu}_1^{(k)}) \right\| \\
&\quad \leq \frac{(m+1)c_4 c_5}{c_1^2} \|\widehat{e}^{(k)} - e^*\| \qquad \text{(from boundedness assumptions)} \\
&\quad \lesssim \|\widehat{e}^{(k)} - e^*\| \\
&\|\psi^1(Z, e^*, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi^1(Z, e^*, \alpha^*, \nu^*)\| \\
&\quad = \left\| \widehat{\kappa}_1^{(k)}(X) - \kappa_1^*(X) - \frac{A}{e^*(X)} \left(\sum_{i=1}^{m+1} \widehat{\alpha}_{1,i}^{(k)}(X, \widehat{\nu}_1^{(k)}) \rho_i(Y, \widehat{\nu}_1^{(k)}) - \sum_{i=1}^{m+1} \alpha_{1,i}^*(X, \nu_1^*) \rho_i(Y, \nu_1^*) \right) \right\| \\
&\quad \leq \|\widehat{\kappa}_1^{(k)} - \kappa_1^*\| + \frac{c_5}{c_1} \sum_{i=1}^{m} \|\widehat{\alpha}_{1,i}^{(k)} - \alpha_{1,i}^*\| \\
&\quad \lesssim \|\widehat{\kappa}_1^{(k)} - \kappa_1^*\| + \sum_{i=1}^{m} \|\widehat{\alpha}_{1,i}^{(k)} - \alpha_{1,i}^*\|
\end{aligned}$$

Therefore:

$$\|\psi(Z, \widehat{e}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\nu}^{(k)}) - \psi(Z, e^*, \alpha^*, \nu^*)\| \lesssim \|\widehat{e}^{(k)} - e^*\| + \sum_{a=0}^{1} \left(\|\kappa_a^{(k)} - \kappa_a^*\| + \sum_{i=1}^{m} \|\widehat{\alpha}_{a,i}^{(k)} - \alpha_{a,i}^*\| \right)$$

Given our assumptions on the nuisance convergence rates, the term above is $o_p(1)$. Putting everything together, we obtain that $\Lambda_{1,k} = o_p(1/\sqrt{n})$ and $\Lambda_{1,k} + \Lambda_{2,k} = o_p(1/\sqrt{n})$. Going back to Eq. A.2, we have that:

$$\sqrt{n}(\widehat{\gamma} - \widetilde{\gamma}) = \sqrt{n}\widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1} \frac{1}{K} \sum_{k=1}^{K} (\Lambda_{1,k} + \Lambda_{2,k})$$

Using the continuous mapping theorem, we have that $\widehat{\mathbb{E}}_n[\phi(X)\phi(X)^T]^{-1} \xrightarrow{P} \Sigma_\phi^{-1}$. Finally, by Slutsky's theorem and the fact that K does not grow with n, we obtain that $\sqrt{n}(\widehat{\gamma} - \widetilde{\gamma})$ is $o_p(1)$ and therefore:

$$\sqrt{n}(\widehat{\gamma} - \gamma^*) \rightsquigarrow \mathcal{N}(0, \Sigma^*).$$

□

A.2 Applications of Theorem 2.6

A.2.1 Pseudo-outcomes and Rates for CQTE

Corollary A.1 (Pseudo-outcomes and rates for CQTE)**.** *Assume the distribution F is continuous. Then, the pseudo-outcome for the conditional quantile treatment at level τ, $CQTE(x;\tau)$, is given by:*

$$\psi^{CQTE}(Z, e, q, f) = q_1(X;\tau) - q_0(X;\tau) + \frac{A - e(X)}{e(X)(1-e(X))} \frac{1}{f_A(x)} (\tau - \mathbb{I}[Y \leq q_A(x;\tau)])$$

where $f_a(x) = f_{Y|X=x,A=a}(q_a(x;\tau))$ is the conditional density of Y evaluated at the conditional quantile. Furthermore, if the conditions of Assumption 2.5 are satisfied, the oracle rate deviation is given by:

$$\mathcal{E} \leq \sum_{k=1}^{K} \sum_{a=0}^{1} \|\widehat{q}_a^{(k)} - q_a^*\| \left(\|\widehat{e}^{(k)} - e^*\| + \left\| \frac{1}{\widehat{f}_a^{(k)}} - \frac{1}{f_a^*} \right\| + \|\widehat{q}_a^{(k)} - q_a^*\| \right)$$

Remark A.2. We note that given Assumption 2.5, f_a and $\widehat{f_a}$ are bounded away from 0, and thus the rate deviation can be written as $\sum_{k=1}^{K} \sum_{a=0}^{1} \|\widehat{q}_a^{(k)} - q_a^*\| \left(\|\widehat{e} - e^*\| + \|\widehat{f}_a^{(k)} - f_a^*\| + \|\widehat{q}_a^{(k)} - q_a^*\| \right)$. When considering medians ($\alpha = 0.5$), this result reduces to the result in Leqi and Kennedy [2021] for median effects.

Proof. For continuous CDFs, the conditional quantile at level τ is the solution to the moment:

$$\mathbb{E}[\tau - \mathbb{I}[Y \leq q_a(x;\tau)] \mid X = x, A = a] = 0$$

The Jacobian is given by $J_a^*(X) = D_{q_a}\{\mathbb{E}[\tau - \mathbb{I}[Y \leq q_a(X;\tau)] \mid X = x, A = a]\}\Big|_{q_a = q_a^*} = -f_a^*(x)$ where $f_a(x) = f_{Y|X=x,A=a}(q_a(x;\tau))$ is the conditional density evaluated at the conditional quantile. By Definition 2.4, $\alpha_a^*(X) = -1/f_a^*(X)$ and the pseudo-outcome is given by:

$$\psi^{\text{CQTE}}(Z, e, q, f) = q_1(X;\tau) - q_0(X;\tau) + \frac{A - e(X)}{e(X)(1-e(X))} \frac{1}{f_A(x)} (\tau - \mathbb{I}[Y \leq q_A(x;\tau)])$$

We apply Theorem 2.6 with $\alpha_a = 1/f_a$, $\nu_a = q_a$ and obtain that the oracle deviation is given by:

$$\mathcal{E} \leq \sum_{k=1}^{K} \sum_{a=0}^{1} \|\widehat{q}_a^{(k)} - q_a^*\| \left(\|\widehat{e}^{(k)} - e^*\| + \left\| \frac{1}{\widehat{f}_a^{(k)}} - \frac{1}{f_a^*} \right\| + \|\widehat{q}_a^{(k)} - q_a^*\| \right)$$

as desired.

□

A.2.2 Pseudo-outcomes and Rates for CSQTE

Corollary A.3 (Pseudo-outcomes and rates for CSQTE)**.** *Assume the distribution F is continuous. Then, the pseudo-outcome for the conditional super-quantile treatment at level τ, CSQTE$(x;\tau)$, is given by:*

$$\begin{aligned}
&\psi^{CSQTE}(Z, e, \mu, q) \\
&= \mu_1(X;\tau) - \mu_0(X;\tau) \\
&\quad + \frac{A - e(X)}{e(X)(1-e(X))} \Big(q_A(X;\tau) + \frac{1}{1-\tau}(Y - q_A(X;\tau))\mathbb{I}[Y \geq q_A(X;\tau)] - \mu_A(X;\tau) \Big)
\end{aligned}$$

Furthermore, if the conditions of Assumption 2.5 are satisfied, the oracle rate deviation is given by:

$$\mathcal{E} \leq \sum_{k=1}^{K} \sum_{a=0}^{1} \left(\|\widehat{\mu}_a^{(k)} - \mu_a^*\| \|\widehat{e}^{(k)} - e^*\| + \|\widehat{q}_a^{(k)} - q_a^*\|^2 \right)$$

Proof. From Example 2.4.2, we have that for continuous CDFs, the conditional super-quantile at level τ is the solution to the conditional moments:

$$\mathbb{E}[(1-\tau)^{-1} Y \mathbb{I}[Y \geq q_a(x;\tau)] - \mu_a(x;\tau) \mid X = x, A = a] = 0$$

$$\mathbb{E}[\tau - \mathbb{I}[Y \le q_a(x;\tau)] \mid X = x, A = a] = 0$$

Thus, the Jacobian $J_a^*(X)$ and its inverse if given by:

$$J_a^*(X) = \begin{pmatrix} -1 & -(1-\tau)^{-1} q_a(X;\tau) f_a(X) \\ 0 & -f_a(X) \end{pmatrix}, \quad (J_a^*(X))^{-1} = \begin{pmatrix} -1 & (1-\tau)^{-1} q_a(X;\tau) \\ 0 & -1/f_a(X) \end{pmatrix}$$

where $f_a(x) = f_{Y|X=x,A=a}(q_a(x;\tau))$ is the conditional density evaluated at the conditional quantile. By Definition 2.4 with $\alpha_a = (-1, -(1-\tau)^{-1} q_a(X;\tau))$, $\nu_a = (\mu_a, q_a)$, the pseudo-outcome is given by:

$$\begin{aligned}
&\psi^{\text{CSQTE}}(Z, e, \mu, q) \\
&= \mu_1(X;\tau) - \mu_0(X;\tau) + \frac{A - e(X)}{e(X)(1-e(X))} \frac{1}{1-\tau} \Big(\\
&\qquad Y \mathbb{I}[Y \ge q_A(X;\tau)] - (1-\tau)\mu_A(X;\tau) - q_A(X;\tau)\,(\tau - \mathbb{I}[Y \le q_A(X;\tau)]) \Big) \\
&= \mu_1(X;\tau) - \mu_0(X;\tau) \\
&\quad + \frac{A - e(X)}{e(X)(1-e(X))} \Big(q_A(X;\tau) + \frac{1}{1-\tau}(Y - q_A(X;\tau)) \mathbb{I}[Y \ge q_A(X;\tau)] - \mu_A(X;\tau) \Big)
\end{aligned}$$

We now apply Theorem 2.6. We first note that the binary matrices G and H are given by:

$$G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Moreover, $\|\widehat{\alpha}_{a,1}^{(k)} - \alpha_{a,1}^*\| = 0$ since $\alpha_{a,1} = -1$ is a constant. Therefore, in the cross-products $\|\widehat{\alpha}_{a,i}^{(k)} - \alpha_{a,i}^*\| \|\widehat{\nu}_{a,j}^{(k)} - \nu_{a,j}^*\|$, the only surviving term is $\|\widehat{\alpha}_{a,2}^{(k)} - \alpha_{a,2}^*\| \|\widehat{\nu}_{a,2}^{(k)} - \nu_{a,2}^*\| = \|\widehat{q}_a^{(k)} - q_a^*\|^2$ due to $G_{21} = 0$. Furthermore, given H, the only surviving term in the cross-products $\|\widehat{\nu}_{a,i}^{(k)} - \nu_{a,i}^*\| \|\widehat{\nu}_{a,j}^{(k)} - \nu_{a,j}^*\|$ is $\|\widehat{\nu}_{a,2}^{(k)} - \nu_{a,2}^*\|^2 = \|\widehat{q}_a^{(k)} - q_a^*\|^2$. Thus, the oracle deviation is given by:

$$\mathcal{E} \le \sum_{k=1}^{K} \sum_{a=0}^{1} \left(\|\widehat{\mu}_a^{(k)} - \mu_a^*\| \|\widehat{e}^{(k)} - e^*\| + \|\widehat{q}_a^{(k)} - q_a^*\|^2 \right)$$

as desired. □

A.2.3 Pseudo-outcomes and Rates for CfRTE

Corollary A.4 (Pseudo-outcomes and rates for CfRTE)**.** *The conditional f-risk treatment effect at level δ, $CfRTE(x;\delta)$, is given by:*

$$\psi^{CfRTE}(Z,e,R,\beta,\lambda) = R_1^f(X;\delta) - R_0^f(X;\delta) + \frac{A-e(X)}{e(X)(1-e(X))}(m(Y,\beta_A,\lambda_A;\delta) - R_A^f(X;\delta))$$

Furthermore, if the conditions of Assumption 2.5 are satisfied, the oracle rate deviation is given by:

$$\mathcal{E} \leq \sum_{k=1}^{K}\sum_{a=0}^{1}\Big(||\widehat{R}_a^{f,(k)} - R_a^{f,*}||||\widehat{e}^{(k)} - e^*|| + ||\widehat{\beta}_a^{(k)} - \beta_a^*||^2 + ||\widehat{\lambda}_a^{(k)} - \lambda_a^*||^2 + ||\widehat{\beta}_a^{(k)} - \beta_a^*||||\widehat{\lambda}_a^{(k)} - \lambda_a^*||\Big)$$

Proof. From Example 2.4.3, we have that the CfRTE is identified by the following conditional moments:

$$\begin{aligned}
&\mathbb{E}[m(Z,\beta_a,\lambda_a;\delta) - R_a^f(x;\delta) \mid X=x, A=a] = 0\\
&\mathbb{E}\left[\frac{\partial}{\partial\beta_a}m(Z,\beta_a,\lambda_a;\delta)\middle| X=x, A=a\right] = 0\\
&\mathbb{E}\left[\frac{\partial}{\partial\lambda_a}m(Z,\beta_a,\lambda_a;\delta)\middle| X=x, A=a\right] = 0
\end{aligned}$$

where it is understood that β_a and λ_a are functions of X. Thus, the Jacobian $J_a^*(X)$ and its inverse if given by:

$$J_a^*(X) = \begin{pmatrix} -1 & 0 & 0\\ 0 & -\mathbb{E}\left[\frac{(Y-\lambda_a(X))^2}{\beta_a(X)^3}(f^*)''\left(\frac{Y-\lambda_1(X)}{\beta_a(X)}\right)\middle| X=x, A=1\right] & \mathbb{E}\left[\frac{Y-\lambda_a(X)}{\beta_a(X)^2}(f^*)''\left(\frac{Y-\lambda_a(X)}{\beta_a(X)}\right)\middle| X=x, A=1\right]\\ 0 & \mathbb{E}\left[\frac{Y-\lambda_a(X)}{\beta_a(X)^2}(f^*)''\left(\frac{Y-\lambda_a(X)}{\beta_a(X)}\right)\middle| X=x, A=1\right] & -\mathbb{E}\left[\frac{1}{\beta_a(X)}(f^*)''\left(\frac{Y-\lambda_a(X)}{\beta_a(X)}\right)\middle| X=x, A=1\right]\end{pmatrix}$$

$$(J_a^*(X))^{-1} = \begin{pmatrix} -1 & 0 & 0\\ 0 & \dots & \dots\\ 0 & \dots & \dots\end{pmatrix}$$

By Definition 2.4 with $\alpha_a = (-1, 0, 0)$, $\nu_a = (R_a^f, \beta_a, \lambda_a)$, the pseudo-outcome is given by:

$$\psi^{\mathrm{C}f\mathrm{RTE}}(Z, e, R, \beta, \lambda) = R_1^f(X;\delta) - R_0^f(X;\delta) + \frac{A - e(X)}{e(X)(1-e(X))}(m(Y, \beta_A, \lambda_A;\delta) - R_A^f(X;\delta))$$

We now apply Theorem 2.6. We first note that the binary matrices G and H are given by:

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Moreover, $\|\widehat{\alpha}_{a,i}^{(k)} - \alpha_{a,i}^*\| = 0$ since the $\alpha_{a,i}$'s are constants. Therefore, the cross-products $\|\widehat{\alpha}_{a,i}^{(k)} - \alpha_{a,i}^*\|\|\widehat{\nu}_{a,j}^{(k)} - \nu_{a,j}^*\|$ vanish. Furthermore, given H, the only surviving terms in the cross-products $\|\widehat{\nu}_{a,i}^{(k)} - \nu_{a,i}^*\|\|\widehat{\nu}_{a,j}^{(k)} - \nu_{a,j}^*\|$ involve only β_a and λ_a. Thus, the oracle deviation is given by:

$$\mathcal{E} \leq \sum_{k=1}^{K}\sum_{a=0}^{1}\left(\|\widehat{R}_a^{f,(k)} - R_a^*\|\|\widehat{e}^{(k)} - e^*\| + \|\widehat{\beta}_a^{(k)} - \beta_a^*\|^2 + \|\widehat{\lambda}_a^{(k)} - \lambda_a^*\|^2 + \|\widehat{\beta}_a^{(k)} - \beta_a^*\|\|\widehat{\lambda}_a^{(k)} - \lambda_a^*\|\right)$$

□

We further provide a specific example of CfRTE for $f(x) = x\log x$ which corresponds to the Kullback–Leibler (KL) divergence. The associated risk is known as the entropic value-at-risk (EVaR). We term this conditional f-risk treatment effect the CKLRTE. The convex conjugate of f is given by $f^*(x^*) = e^{x^*-1}$. From the first-order optimality conditions, we obtain:

$$\beta_a^*(x;\delta) = \arg\inf_{\beta_a(x;\delta)\geq 0} \beta_a(x;\delta)\left(\log \mathbb{E}\left[e^{\frac{Y}{\beta_a(x;\delta)}}\middle| X = x, A = a\right] + \delta\right) \tag{A.4}$$

$$\begin{aligned} R^{KL}(x;\delta) &= \beta_a^*(x;\delta)\left(\log \mathbb{E}\left[e^{\frac{Y}{\beta_a^*(x;\delta)}}\middle| X = x, A = a\right] + \delta\right) \\ \lambda_a^*(x;\delta) &= \beta_a^*(x;\delta)\left(\log \mathbb{E}\left[e^{\frac{Y}{\beta_a^*(x;\delta)}}\middle| X = x, A = a\right] - 1\right) \\ &= R^{KL}(x;\delta) - \beta_a^*(x;\delta)(\delta + 1) \end{aligned} \tag{A.5}$$

Thus, the optimization problem contains only one variable of interest, $\beta_a(x;\delta)$. The pseudo-outcome for CKLRTE at level δ is then given by:

$$\psi^{\text{CKLRTE}}(Z,e,R,\beta,\lambda) = R_1^{KL}(X;\delta) - R_0^{KL}(X;\delta) + \\ + \frac{A-e(X)}{e(X)(1-e(X))}\left(\delta\beta_A(X;\delta) + \lambda_A(X;\delta) + \beta_A(X;\delta)e^{\frac{Y-\lambda_A(X;\delta)}{\beta_A(X;\delta)}-1} - R_A^{KL}(X;\delta)\right)$$

Remark A.5. A practical problem is the estimation of $\beta_a^*(X;\delta)$. We observe that the empirical analog of Eq. (A.4) requires fitting a regression function $\widehat{\mathbb{E}}_n\left[e^{\frac{Y}{\beta_a(x;\delta)}}\middle|X=x,A=a\right]$ for each candidate $\beta_a(x;\delta)$. This can instead be mitigated by learning similarity weights between x and the training data and then replacing $\widehat{\mathbb{E}}_n\left[e^{\frac{Y}{\beta_a(x;\delta)}}\middle|X=x,A=a\right]$ with a weighted sum. We can then use this approximation with any convex optimization method since the argument of Eq. (A.4) is a convex function in β_a.

A.3 Additional Experimental Results

The replication code is distributed under an MIT license. The results in Section 2.8 were obtained using an Amazon Web Services instance with 32 vCPUs and 64 GiB of RAM.

A.3.1 Simulation Study

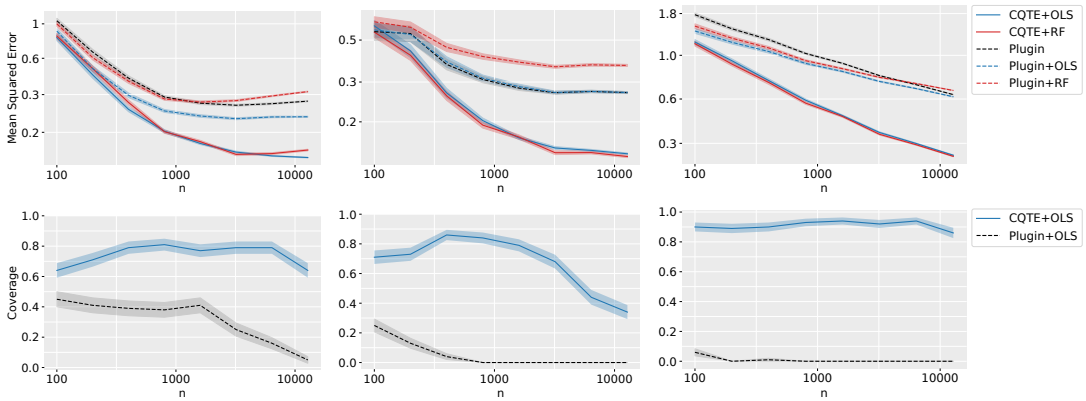
In this section, we provide additional results on simulated data for CQTEs (Section 2.4.1) and CfRTE (Section 2.4.3). Our analysis relies on the same data generating process and setup described in Section 2.8. For nuisance and second-stage estimation, we use Python's `scikit-learn` library [Pedregosa et al., 2011] for logistic and linear regression, random forest regression (RF) and linear quantile regression (LQR). We employ the extension `sklearn-quantile` for quantile random forest regression (QRF). We use the estimators' default hyperparameters except for the RFs where we set the minimum leaf size to $n/20$ to control

overfitting. For each comparison, we run 100 simulations for each sample size $n = 100, 200, \ldots, 12800$ and evaluate the mean squared error (MSE) over a fixed set of 500 random X values.

CQTE Experiments. We measure the conditional quantile treatment effect at level $\tau = 0.75$. For the given DGP, the true CQTE at level τ is given by $q_1(X;\tau) - q_0(X;\tau) \simeq 1.14(e^{X_0+X_1} - e^{X_0})$, a heterogenous function of X. Learning CQTEs requires estimating three nuisances: the propensity score $e(X)$, the conditional quantile $q_a(X;\tau)$, as well as the density at the conditional quantile, $f_a(X)$.

We estimate $e(X)$ using logistic regression. Likewise, we learn $f_a(X)$ using the method described in Section 2.5.2 with a Gaussian kernel (bandwidth $b = 1$) for the estimates ω_i and a random forest (RF) regressor as the final estimator $\widehat{\mathbb{E}}_n[\omega \mid X = x, A = a]$. Finally, we consider three methods for estimating $q_a(X;\tau)$. First, we consider a quantile RF (QRF) [Meinshausen and Ridgeway, 2006] as the *flexible* learner. Then, we consider a "*misspecified*" model such as the linear quantile regressor (LQR) [Koenker, 2005]. Lastly, we consider an estimator that uses a Gaussian kernel for calculating weights and then computes $\widehat{q}_a$ using the weighted version of the moment in Eq. (2.3). We choose the kernel bandwidth by Silverman's rule [Silverman, 2018]. This will be our *slow* estimator as we expect it to suffer from the curse of dimensionality. For the final stage, we use an ordinary least squares model (CQTE+OLS) or an RF (CQTE+RF).

For the performance comparisons, we proceed similarly to Section 2.8. Thus, we compare the out-of-sample MSE of the CQTE estimator with that of the plug-in estimator from Eq. 2.7. We also construct Plugin+OLS and Plugin+RF given by running an additional OLS/RF model on the cross-fitted plug-in predictions. We do so to account for additional smoothing from the last stage regressor. When the second stage algorithm is OLS, we check whether the 95%-confidence



(a) Flexible $\hat{q}$ (b) Misspecified $\hat{q}$ (c) Slow $\hat{q}$

Figure A.1: Mean squared error (MSE) and 95% confidence interval coverage for different conditional quantile treatment effect (CQTE) learners in the synthetic experiment. Shaded regions show plus/minus one standard error over 100 simulations.

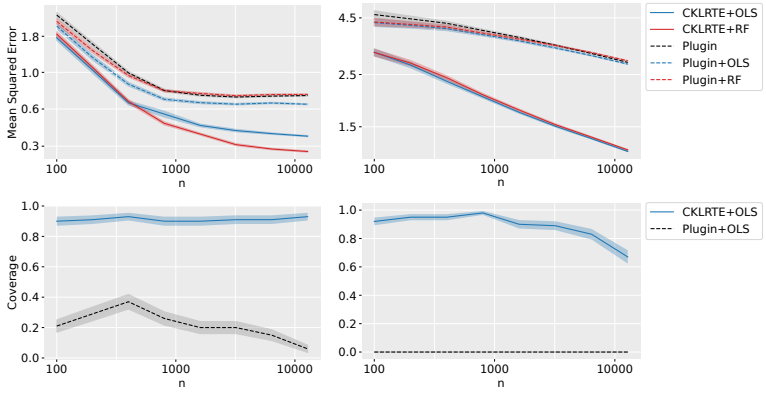
interval OLS returns for the X_1 coefficient contains the coefficient from the true projection. The results are shown in Figure A.1. Our CQTE learner provides uniformly better MSE performance that the plugin counterparts, and the results show this is not just a consequence of the second-stage regression. For inference, we achieve better coverage than the plug-in approaches. However, for the flexible and "misspecified" estimator, the coverage does not reach the desired level of 95%. This is most likely due to misspecifcation in the conditional density learner. *Note:* we put *misspecified* in quotes since the MSE of this estimator is better than that of the *flexible* estimator. We attribute this to the fact that for small t we can approximate $e^t \simeq 1 + t$ which makes the true CQTE close to a linear approximation.

CfRTE Experiments. We now turn to estimating the conditional f-risk treatment effects when we set the f-divergence to be the KL divergence. This f-risk measures the difference between conditional entropic values-at-risk (EVaRs). We call this risk treatment effect the CKLRTE (see Appendix A.2.3). We now

wish to measure the CKLRTE at level $\tau = 0.75$ ($\delta = -\log(1-\tau)$). We note that our DGP does not admit EVaRs as the moment generating function of a lognormal distribution diverges. Thus, we modify the DGP to truncate any values above the 99$^{\text{th}}$ conditional quantile. For the truncated DGP, the true CKLRTE at level δ is given by $R_1^{KL}(X;\delta) - R_0^{KL}(X;\delta) \simeq 1.42(e^{X_0+X_1} - e^{X_0})$, a heterogeneous function of X. Learning CKLRTEs requires estimating the following nuisances: the propensity score $e(X)$, the conditional EVaR $R_a^{KL}(X;\delta)$ and the $\beta_a(X;\delta)$ optimization parameter.

As before, we estimate $e(X)$ using logistic regression. We learn $R_a^{KL}(X;\delta)$ and the optimization parameter $\beta_a(X;\delta)$ jointly via the procedure described in Appendix A.2.3. In particular, we first learn weights for approximating $\widehat{\mathbb{E}}_n\left[e^{Y/\beta_a(x;\delta)} \mid X = x, A = a\right]$ by a weighted average, and then solve the convex optimization problem in Eq. (A.4). For the weights, we use either a random forest (a *flexible* learner) or a Gaussian kernel (a *slow* learner, expected to suffer from the curse of dimensionality), with bandwidth chosen by Silverman's rule [Silverman, 2018]. For the final stage, we use either ordinary least squares (CKLRTE+OLS) or a random forest (CKLRTE+RF).

Similar to our other experimental benchmarks, we compare the out-of-sample MSE of the CKLRTE estimator with that of the plug-in estimator from Eq. 2.7. We also construct Plugin+OLS and Plugin+RF given by running an additional OLS/RF model on the cross-fitted plug-in predictions. This will account for any additional smoothing from the last stage regressor. When the second stage algorithm is OLS, we check whether the 95%-confidence interval for the X_1 coefficient contains the coefficient from the true projection. We display the results in Figure A.2. Our CKLRTE learner provides uniformly better MSE performance than the plugin counterparts, regardless of the second stage re-



(a) Flexible $\widehat{R}^{KL}$ (b) Slow $\widehat{R}^{KL}$

Figure A.2: Mean squared error (MSE) and 95% confidence interval coverage for different conditional KL-risk treatment effect (CKLRTE) learners in the synthetic experiment. Shaded regions show plus/minus one standard error over 100 simulations.

gression. For inference, we achieve good coverage whereas plug-in approaches yield little to no coverage.

A.3.2 Impact of 401(k) Eligibility on Financial Wealth

We provide a detailed description the 401(k) dataset features in Table A.1. We then compare the feature importances given by the random forest final stage of the CSQTE estimators and the DR-Learner. Figure A.3 shows that the features driving the treatment effects for the three estimators have the same importance profile across tasks. For example, income, age and education are the most important features when determining the conditional average treatment effect, as well as the conditional average effects in the left 25% tail and right 25% tail.

A.4 Practical Considerations for CDTE Estimation

One of the limitations of our work is that the interpretation of our estimands as causal effects only hold when the unconfoundedness assumption holds. Whether it holds or does not, our estimands are differences in distributional

Table A.1: Covariates included in the 401(k) dataset used in the Chapter 2 application.

Name	Description	Type
age	age	continuous
inc	income	continuous
educ	years of completed education	continuous
fsize	family size	continuous
marr	marital status	binary
two_earn	whether dual-earning household	binary
db	defined benefit pension status	binary
pira	IRA participation	binary
hown	home ownership	binary
e401	401 (k) eligibility	binary
net_tfa	net financial assets	continuous

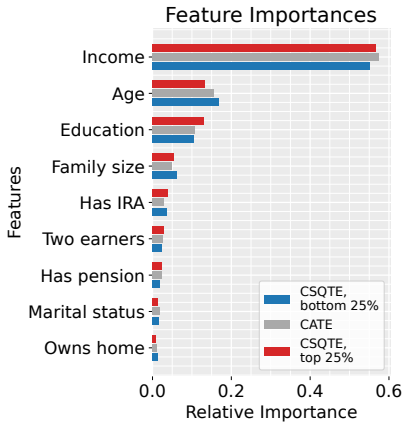


Figure A.3: Feature importances from the final-stage learner in the 401(k) application, for CSQTE on the bottom 25% of financial-asset holders, CATE, and CSQTE on the top 25% of financial-asset holders.

statistics between the distributions of $Y \mid X, A = 1$ and $Y \mid X, A = 0$. If it does hold, this coincides with the differences in distributional statistics between the distributions of $Y(1) \mid X$ and $Y(0) \mid X$. While unconfoundedness can usually be guaranteed in experiments barring any non-compliance, there is no way to test for it in observational data. Thus, the onus is on the researcher to determine whether unconfoundedness is plausible in their data and to interpret the estimands carefully.

Another possible concern in practice is the possibility for biased selection, where the sample may not be representative of the population of interest. To the extent that any unrepresentativeness is explained by X (known as, selection at random), the CDTEs given X will be unaffected. Therefore, we can alleviate this issue in the application of our method by learning CDTEs with a second-stage regression that is a universal approximator (such as nonparametric regression, forests, or deep networks) so it learns the true CDTEs, thus adjusting for X fully. If there is unrepresentativeness not reflected in X (known as, selection *not* at random), then even the true CDTEs given $X = x$ only reflect effects on the data-generating subpopulation with $X = x$. This may be alleviated by considering richer X so as to minimize the discrepancy and/or by acknowledging possible biases. Finally, whether selection is at random or not, any best-in-class prediction guarantees (such as for simple second-stage regressions like OLS) only hold with respect to the data-generating population, and therefore "best-performance" may be unrepresentative of performance in the population of interest. If selection is at random, this is nonetheless easily fixable by reweighting.

Yet another practical concern is whether the outcome we observe truly reflects the appropriate notion of benefit/risk. Otherwise, CDTEs similarly may not reflect the right risk measure, and our estimators could further propagate undesirable biases and choices encoded in the data [Passi and Barocas, 2019].

Lastly, there may be issues of calibrating risk profiles to human preferences, so as to make CDTEs informative for decision making. In particular, there may be conflicting risk preferences between the decision making entity (*e.g.*, doctor, policymaker, product manager) and the individual (*e.g.*, patient, citizen, platform user). If the risk measure employed is in conflict with individual preferences, it is possible to have a negative impact on individuals from their own

perspective. Nonetheless, using the outputs of our models using different risk measures on an individualized basis could potentially be used to make policy decisions for individuals based on their own risk preferences.

APPENDIX B

APPENDIX FOR CHAPTER 3

Note: Throughout the appendix, we use ± notation to encode either upper/ lower bounds results. This allows us to unify upper/ lower results and proofs at the cost of some readability.

B.1 Notation

We summarize the notation we use throughout this work in Table B.1. In addition, note that we use upper case letters (e.g. X) to denote random variables and lower case letters (e.g. x) to refer to specific values of a random variable.

Table B.1: Notation used in Chapter 3.

Symbol	Meaning
X	The observed covariates in $\mathbb{R}^d$
A	A binary treatment ($A \in \{0, 1\}$)
Y	The outcome
Z	(X, A, Y) drawn from an observed distribution P
$Y(1), Y(0)$	Real-valued treated and untreated potential outcomes, respectively
U	The unobserved confounder in $\mathbb{R}^k$
P_{full}	An unobservable distribution over $(X, A, Y(1), Y(0), U)$
α	$\frac{\Lambda}{\Lambda+1} \in [0.5, 1)$ for $\Lambda \geq 1$
$\{b\}_+, \{b\}_-$	$\max\{b, 0\}$ and $\min\{b, 0\}$, respectively, for a real number b
$b \lesssim d$	$b \leq Cd$, for $b, d \in \mathbb{R}$, and for some universal constant C
g^*	The true value of a function g
$\bar{g}$	A putative value of a function g
$\widehat{g}$	An estimated value of a function g from data
$\|\|g\|\| := \mathbb{E}_F[g(z)^2]^{1/2}$	The L_2 norm of g under a probability distribution $F(z)$
$+, -$	Indicators denoting upper and lower bounds, respectively
$\pm, \mp$	Symbols indicating that an equation should be read twice, once with $\pm = +, \mp = -$ and once with $\pm = -, \mp = +$

Continued on next page

Table B.1 continued from previous page

Symbol	Meaning
	E.g., $a^{\pm} = b^{\pm} + c^{\mp}$ encodes two equalities: $a^{+} = b^{+} + c^{-}$ and $a^{-} = b^{-} + c^{+}$
$e(x)$	The observed propensity score $P(A = 1 \mid X = x)$
$e(x, u)$	The full propensity score $P_{\text{full}}(A = 1 \mid X = x, U = u)$
$F(y \mid x, a)$	The conditional outcome distribution, $P(Y \leq y \mid X = x, A = a)$
$f(y \mid x, a)$	The conditional outcome density, $\frac{d}{dy}F(y \mid x, a)$
$\mu^*(x, a)$	$\mathbb{E}[Y \mid X = x, A = a]$, the outcome regression
$q_c^*(x, a)$	$\inf\{\beta : F(\beta \mid x, a) \geq c\}$, the conditional outcome quantile
$q_+^*(x, a)$	$q_\alpha^*(x, a)$, shorthand α^{th} quantile notation
$q_-^*(x, a)$	$q_{1-\alpha}^*(x, a)$, shorthand $(1-\alpha)^{\text{th}}$ quantile notation
$H_{\pm}(z, \bar{q})$	$\bar{q}(x, a) + \frac{1}{1-\alpha}\{y - \bar{q}(x, a)\}_{\pm}$, Conditional Value at Risk pseudo-outcome
$\text{CVaR}_{\pm}(x, a)$	$\mathbb{E}[H_{\pm}(z, q_{\pm}^*) \mid X = x, A = a]$, the Conditional Value at Risk
$\text{CVaR}_{+}(x, a)$	The expectation above the $(1-\alpha)$ quantile
$\text{CVaR}_{-}(x, a)$	The expectation below the α quantile
$R_{\pm}(z, \bar{q})$	$\Lambda^{-1}y + (1 - \Lambda^{-1})H_{\pm}(z, \bar{q})$, pseudo-outcome for the (conditional) unobserved potential outcome
$\rho_{\pm}^*(x, a, \bar{q})$	$\mathbb{E}[R_{\pm}(z, \bar{q}) \mid X = x, A = a]$, the (conditional) expected unobserved potential outcome
$\rho_{\pm}^*(x, a)$	Shorthand for $\rho_{\pm}^*(x, a, q_{\pm}^*)$, i.e., $\rho_{\pm}^*$ evaluated at the true conditional quantiles $q_{\pm}^*$
CATE Bounds Pseudo-Outcomes	
$\phi_1^+(Z, \widehat{\eta})$	$AY + (1-A)\widehat{\rho}_+(X, 1) + \frac{(1-\widehat{e}(X))A}{\widehat{e}(X)}\left(R_+(Z, \widehat{q}_+(X, 1)) - \widehat{\rho}_+(X, 1)\right)$
$\phi_0^-(Z, \widehat{\eta})$	$(1-A)Y + A\widehat{\rho}_-(X, 0) + \frac{\widehat{e}(X)(1-A)}{1-\widehat{e}(X)}\left(R_-(Z, \widehat{q}_-(X, 0)) - \widehat{\rho}_-(X, 0)\right)$
$\phi_\tau^+(Z, \widehat{\eta})$	$\phi_1^+(Z, \widehat{\eta}) - \phi_0^-(Z, \widehat{\eta})$

B.2 Results for CATE Lower Bounds

The results for the CATE lower bound $\tau^-(x)$ can be obtained by interchanging + and − symbols in the nuisances and/or replacing A with $1 - A$. We state them here for completeness.

CATE Lower Bounds Identification (Result 3.4) The sharp CATE lower bound is given by $\tau^-(x) = Y^-(x,1) - Y^+(x,0)$, where the relevant bounds on the conditional average potential outcomes can be expressed as:

$$Y^-(x,1) = e^*(x)\mu^*(x,1) + (1 - e^*(x))\rho^*_-(x,1),$$
$$Y^+(x,0) = (1 - e^*(x))\mu^*(x,0) + e^*(x)\rho^*_+(x,0).$$

Thus, the lower bounds can be expressed as a convex combination of quantities that can be estimated from the observed data, i.e. they are *identifiable* from data.

Pseudo-outcome for CATE Bounds (Definition 3.5) Let

$$\widehat{\eta} = (\widehat{e}, \widehat{q}_+(\cdot,0), \widehat{q}_-(\cdot,1), \widehat{\rho}_+(\cdot,0), \widehat{\rho}_-(\cdot,1)) \in \Xi$$

be a set of nuisances. The pseudo-outcomes for the bounds $Y^-(x,1)$, $Y^+(x,0)$, and $\tau^-(x)$ are given by

$$\phi_1^-(Z,\widehat{\eta}) = AY + (1-A)\widehat{\rho}_-(X,1) + \frac{(1-\widehat{e}(X))A}{\widehat{e}(X)} \cdot \left(R_-(Z,\widehat{q}_-(X,1)) - \widehat{\rho}_-(X,1)\right),$$
$$\phi_0^+(Z,\widehat{\eta}) = (1-A)Y + A\widehat{\rho}_+(X,0) + \frac{\widehat{e}(X)(1-A)}{1-\widehat{e}(X)} \cdot \left(R_+(Z,\widehat{q}_+(X,0)) - \widehat{\rho}_+(X,0)\right),$$
$$\phi_\tau^-(Z,\widehat{\eta}) = \phi_1^-(Z,\widehat{\eta}) - \phi_0^+(Z,\widehat{\eta}).$$

Validity and Sharpness for CATE Lower Bounds We call lower bound estimates $\widehat{\tau^-}(x)$ *valid* if $\widehat{\tau^-}(x) - \tau^-(x) \leq -o_p(1)$. Similarly, the lower bound estimates $\widehat{\tau^-}(x)$ are *sharp* if $\widehat{\tau^-}(x) = \tau^-(x) + o_P(1)$.

Pseudo-outcome Bias for CATE Lower Bounds The absolute bias of the CATE lower bound pseudo-outcome has the form:

$$\left|\mathcal{E}_\tau^-(x;\widehat{\eta})\right| \lesssim |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_-(x,1) - \rho^*_-(x,1)| + |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_+(x,0) - \rho^*_+(x,0)|$$
$$+ (\widehat{q}_-(x,1) - q^*_-(x,1))^2 + (\widehat{q}_+(x,0) - q^*_+(x,0))^2.$$

whereas the signed bias bound is given by:

$$\mathcal{E}_\tau^-(x;\widehat{\eta}) \lesssim |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_-(x,1) - \rho_-^*(x,1,\widehat{q}_-)| \\ - |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_+(x,0) - \rho_+^*(x,0,\widehat{q}_+)|.$$

The proofs of the theorems and corollaries in the chapter (Appendix B.4) are unified across lower/upper bounds by using the ± notation described above. For example, we will write the consolidated pseudo-outcome bias bounds as:

$$\left|\mathcal{E}_a^\pm(x;\widehat{\eta})\right| \lesssim |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_\pm(x,a) - \rho_\pm^*(x,a)| + (\widehat{q}_\pm(x,a) - q_\pm^*(x,a))^2$$

$$\mp\mathcal{E}_a^\pm(x;\widehat{\eta}) \lesssim |\widehat{e}(x) - e^*(x)|\,|\widehat{\rho}_\pm(x,a) - \rho_\pm^*(x,a,\widehat{q}_\pm)|.$$

which, together with $\mathcal{E}_\tau^\pm(x;\widehat{\eta}) = \mathcal{E}_1^\pm(x;\widehat{\eta}) - \mathcal{E}_0^\mp(x;\widehat{\eta})$, yield the bias bounds for the lower and upper CATE bounds.

B.3 More Estimation Results

B.3.1 More ERM Results

Corollary B.1 (Conditions for ERM Oracle Efficiency)**.** *Let $\mathcal{F}$ be a class of β-smooth functions in d dimensions (i.e. Hölder) and let $e, \rho_\pm, q_\pm$ be γ_e, γ_ρ, and γ_q-smooth functions, respectively. Then, the L_2 error rate of Algorithm 3.1 is $O_p(n^{-1/(2+d/\beta)} + n^{-2/(2+d/\gamma_q)} + n^{-(1/(2+d/\gamma_e)+1/(2+d/\gamma_\rho))})$. Furthermore, if $\gamma_q \geq \frac{d/2}{1+d/\beta}$ and $\gamma_\rho\gamma_e \geq \frac{d^2}{4} - \frac{(\gamma_\rho+d/2)(\gamma_e+d/2)}{1+2\beta/d}$, our estimator is oracle efficient in the sense that the leading order error is that of the oracle estimator, $\widehat{\mathbb{E}}_n[\phi_\tau^\pm(Z,\eta^*) \mid X = x]$.*

B.3.2 Doubly Robust-Style Smoothing Estimators

We now study the behavior of Algorithm 3.1 with a DR Learner-style smoothing estimator as the second-stage learner. This technique was introduced in Kennedy [2023a] and includes a wide range of estimators satisfying certain stability conditions, with linear smoothers as the archetype of this class. In this

section, we analyze a generic linear smoother defined as follows:

$$\widehat{\mathbb{E}}_n\left[\widehat{\phi}_\tau^\pm \mid X = x\right] = \frac{1}{n}\sum_{i=1}^n w_i(x)\widehat{\phi}_{\tau,i}^\pm$$

where the $w_i(x)$'s are weights learned on a different sample than $\widehat{\phi}_{\tau,i}^\pm$ (which can be achieved by sample splitting). Under mild regularity assumptions, this estimator can yield stronger guarantees in the form of pointwise error bounds.

Theorem B.2 (Rates for Linear Smoothing Estimators)**.** *Assume the conditions of Assumption 3.7. Then:*

$$\left|\widehat{\tau}^\pm(x) - \tau^\pm(x)\right| \lesssim \left|\widetilde{\tau}^\pm(x) - \tau^\pm(x)\right| + b_n^\pm(x) + O_p\left(\left(\|\widehat{\phi}_\tau^\pm - \phi_\tau^\pm\|_{w^2} + o_p(1)\right)\left(\frac{1}{n^2}\sum_{i=1}^n w_i(x)^2\right)^{1/2}\right)$$

where $\widetilde{\tau}^\pm(x)$ corresponds to the linear smoother procedure with oracle first-stage nuisances, $\|\cdot\|_{w^2}$ is the empirical $w_i(x)^2$-weighted distance of Kennedy [2023c], and the $b_n^\pm(x)$ bias function is of the form:

$$b_n^\pm(x) = \left|\frac{1}{n}\sum_{i=1}^n w_i(x)\mathcal{E}_\tau^\pm(X_i;\widehat{\eta})\right|$$

Second-stage Sharp Consistency and Robustness $\tilde{\tau}$ consistency follows under weak conditions like $\frac{1}{n}\sum_{i=1}^n |w_i(X_i)| \leq C$ Stone [1977]. Thus, we can state corollaries that prove consistent estimation of sharp bounds under either strong restrictions on weights or strong requirements on consistency. We show one such corollary for a wide class of linear smoothers that includes linear and ridge regression, local polynomial and RKHS regression, kernel estimators, and some tree methods Wasserman [2006]. In this corollary, we ask for uniform nuisance consistency to make the bias term $b_n^\pm$ tend to zero.

Corollary B.3 (Pointwise Consistency of Sharp CATE Bounds)**.** *Assume the conditions of Theorem B.2 are satisfied, the $\frac{w_i(x)}{n}$ weighting functions satisfy the requirements of Stone [1977] Theorem 1, and $\frac{1}{n}\sum|w_i(x)| = O_p(1)$. If $\widehat{q}_\pm$ and either $\widehat{e}$ or $\widehat{\rho}$ are uniformly consistent, then $\widehat{\tau}^\pm(x)$ converges to the true pointwise sharp CATE bounds.*

Second-stage Sharp Rates Take $\tau^{\pm}, e, \rho_{\pm}, q_{\pm}$ be γ_q to be Hölder with smoothness β, γ_e, γ_ρ, and γ_q. Then, the pointwise error rate of Algorithm 3.1 is $O_p(n^{-1/(2+d/\beta)} + n^{-2/(2+d/\gamma_q)} + n^{-(1/(2+d/\gamma_e)+1/(2+d/\gamma_\rho)})$ and the estimator will be oracle efficient, though the error bounds here are *pointwise* (local), whereas the ERM-based bounds are L_2 (global).

Second-stage Validity The linear smoothers also have pointwise validity. Unlike in the ERM case where the best model fit to conservative bounds might extrapolate to invalid bounds for some regions of the covariates, the linear smoothers will have pointwise validity guarantees.

Corollary B.4 (Pointwise Validity of Lax CATE Bounds)**.** *Assume the conditions of Theorem B.2 are satisfied, $w_i(x)$ satisfies the requirements of Theorem 1 in Stone [1977], and $\frac{1}{n}\sum_{i=1}^{n}|w_i(x)| = O_p(1)$. If $\widehat{q}_{\pm}$ is uniformly consistent to some limiting quantile $\bar{q}_{\pm}$ and $\widehat{e}$ is uniformly consistent for e^* or $\widehat{\rho}_{\pm}$ is uniformly consistent for $\rho^*_{\pm}(X, A, \bar{q}_{\pm})$ and $f(\bar{q}_{\pm}(x,a) \mid x, a) > 0$. Then the estimated bounds are pointwise valid in the sense that $\pm\left(\widehat{\tau^{\pm}}(x) - \tau^{\pm}(x)\right) \geq -o_P(1)$.*

B.4 Proofs

Note: we assume throughout that X, U, $Y(0)$, and $Y(1)$ to have probability measures absolutely continuous w.r.t. the Lebesgue measure so that we can condition on the event $X = x$.

Proof of Theorem 3.8. We start with the bound for the unsigned bias. Consider the $Y^+(X, 1)$ bound for simplicity. We first show that our problem fits into the framework of Kallus and Oprescu [2023a] since the estimand and the oracle nuisances are the solutions of following conditional moment restrictions:

$$\mathbb{E}[AY + (1 - A)\rho^*_+(X, 1) - Y^+(X, 1) \mid X] = 0 \qquad \text{(Estimand moment)}$$

$$\mathbb{E}[R_+(Z, q_+^*(X,1)) - \rho_+^*(X,1) \mid X, A = 1] = 0 \qquad \text{(Modified outcome moment)}$$

$$\mathbb{E}[\alpha - \mathbb{I}(Y \leq q_+^*(X,1)) \mid X, A = 1] = 0 \qquad \text{(Quantile moment)}$$

Let ν_1 be the nuisance set corresponding to this set of moments (as defined in Kallus and Oprescu [2023a]). Then $\nu_1^*(X) = (Y^+(X,1), \rho_+^*(X,1), q_+^*(X,1))$. This is different from η^* since the propensity does not have an estimating conditional moment. The Jacobian of the moments with respect to ν_1^* is thus given by:

$$J_1^*(X) = \begin{pmatrix} -1 & 1 - e^*(X) & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -f(q_+^*(X,1) \mid X, 1) \end{pmatrix}$$

where $f(y \mid x, 1)$ be the conditional density at a point y for $a = 1$. The first row of the inverse is then given by $\alpha_1^*(X) := (J_1^*(X))_1^{-1} = (-1, e^*(X) - 1, 0)$. Thus, using the pseudo-outcome in Definition 3 of Kallus and Oprescu [2023a], replacing ν_1^*, α_1^* with their estimated counterparts $\widehat{\nu}_1(X) = (\widehat{Y}^+(X,1), \widehat{\rho}_+(X,1), \widehat{q}_+(X,1))$, $\widehat{\alpha}_1(X) = (-1, \widehat{e}(X) - 1, 0)$, and noting that the first moment is conditional only on X, we obtain the pseudo-outcome:

$$\phi_1^+(Z, \widehat{\eta}) = AY + (1 - A)\widehat{\rho}_+(X,1) + \frac{(1 - \widehat{e}(X))\, A}{\widehat{e}(X)} \cdot (R_+(Z, \widehat{q}_+) - \widehat{\rho}_+(X,1))$$

as desired. Therefore, our Assumption 3.7 is a direct application of the boundedness assumption (Assumption 1) in Kallus and Oprescu [2023a] and the bound for the unsigned bias follows largely from their Theorem 1. We first note that the results of Theorem 1 in Kallus and Oprescu [2023a] also hold pointwise (see the proof in their Appendix A). It now remains to calculate the H and G matrices in their Assumption 1:

$$G = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

since G is just a binary mask for $J_1^*(X)$ and H involves second order derivatives of the moments. Plugging these into the bound for the unsigned bias, we obtain:

$$\begin{aligned}\left|\mathcal{E}_1^+(x;\widehat{\eta})\right| &\lesssim \sum_{i=1}^{3}\sum_{j=1}^{3} G_{ij}\left|\widehat{\alpha}_{1,i}(x)-\alpha_{1,i}^*(x)\right|\left|\widehat{\nu}_{1,j}(x)-\nu_{1,j}^*(x)\right| \\ &\quad + \sum_{i=1}^{3}\sum_{j=1}^{3} H_{ij}\left|\widehat{\nu}_{1,i}(x)-\nu_{1,i}^*(x)\right|\left|\widehat{\nu}_{1,j}(x)-\nu_{1,j}^*(x)\right| \\ &\lesssim |\widehat{e}(x)-e^*(x)|\,|\widehat{\rho}_+(x,1)-\rho_+^*(x,1)| + (\widehat{q}_+(x,1)-q_+^*(x,1))^2.\end{aligned}$$

The result for $\mathcal{E}_0^+(x;\widehat{\eta})$ follows from replacing $a=1$ with $a=0$ everywhere. The bound for $\mathcal{E}_a^-(x;\widehat{\eta})$ follows from writing the corresponding conditional moments for $Y^-(X,a)$. We now study the bound for the signed bias. We first take the expectation of $\phi_1^+(Z,\widehat{\eta})$:

$$\begin{aligned}\mathbb{E}\left[\phi_1^+(Z,\widehat{\eta})\mid X\right] &= \mathbb{E}\left[AY+\left(1-\frac{A}{\widehat{e}(X)}\right)\widehat{\rho}_+(X,1)+A\frac{1-\widehat{e}(X)}{\widehat{e}(X)}R_+(Z,\widehat{q}_+)\mid X\right] \\ &= e^*(X)\mu^*(X,1)+\frac{\widehat{e}(X)-e^*(X)}{\widehat{e}(X)}\widehat{\rho}_+(X,1)+\left(\frac{e^*(X)}{\widehat{e}(X)}-e^*(X)\right)\rho_+^*(X,1,\widehat{q}_+) \\ &= e^*(X)\mu^*(X,1)+\left(1-\frac{e^*(X)}{\widehat{e}(X)}\right)(\widehat{\rho}_+(X,1)-\rho_+^*(X,1,\widehat{q}_+)) \\ &\quad + (1-e^*(X))\,\rho_+^*(X,1,\widehat{q}_+)\end{aligned}$$

As a result, we can write write:

$$\begin{aligned}\mathbb{E}\left[\phi_1^+(Z,\widehat{\eta})-\phi_1^+(Z,\eta^*)\mid X\right] &= \frac{\widehat{e}(X)-e^*(X)}{\widehat{e}(X)}(\widehat{\rho}_+(X,1)-\rho_+^*(X,1,\widehat{q}_+)) \\ &\quad + (\rho_+^*(X,1,\widehat{q}_+)-\rho_+^*(X,1))\,(1-e^*(X))\end{aligned}$$

Recall the CVaR property that $\rho_+^*(X,1) = \inf_{\bar{q}}\rho_+^*(X,1,\bar{q})$, so that $\rho_+^*(X,1,\widehat{q}_+) \geq \rho_+^*(X,1)$. Therefore we have:

$$\begin{aligned}-\mathcal{E}_1^+(x;\widehat{\eta}) &= \mathbb{E}\left[\phi_1^+(Z,\eta^*)-\phi_1^+(Z,\widehat{\eta})\mid X=x\right] \\ &= -\frac{\widehat{e}(x)-e^*(x)}{\widehat{e}(x)}(\widehat{\rho}_+(x,1)-\rho_+^*(x,1,\widehat{q}_+))-(\rho_+^*(x,1,\widehat{q}_+)-\rho_+^*(x,1))\,(1-e^*(x)) \\ &\leq -\frac{\widehat{e}(x)-e^*(x)}{\widehat{e}(x)}(\widehat{\rho}_+(x,1)-\rho_+^*(x,1,\widehat{q}_+))\end{aligned}$$

$$\lesssim -\left|\widehat{e}(x) - e^*(x)\right| \left|\widehat{\rho}_+(x,1) - \rho^*_+(x,1;\widehat{q}_+)\right|$$

The result for $Y^+(X,0)$ follows by symmetry. The result for $Y^-(X,a)$ follows by negating Y, applying the argument, and negating the argument. Results for $\tau^\pm(X)$ follow by Result 3.4. □

Proof of Corollary 3.9. Since we showed in the proof of Theorem 3.8 that our pseudo-outcomes fit into the framework of Kallus and Oprescu [2023a] and our Assumption 3.7 maps to their Assumption 1, we can apply their Theorem 2 directly to our setting, yielding the statement of our theorem. □

Proof of Corollary 3.10. Using lax notation, choose an estimated $\widehat{f^\pm} \in \mathcal{F}$ to minimize Equation (3.1) and then an estimated $\widehat{c^\pm} \in \mathbb{R}$ such that $\widehat{f^\pm} + \widehat{c^\pm}$ is a minimizer of Equation (3.1). By construction, we must have $\widehat{f^\pm} + 0$ is an optimizer.

If we differentiate (3.1) with respect to $\widehat{c^\pm}$, evaluate it at 0, and divide by 2, we obtain the requirement on any optimizer that $\frac{1}{n}\sum_{i=1}^n \left(\widehat{\phi}^\pm_{\tau,i} - \widehat{f^\pm}(X_i)\right) = 0$. Thus:

$$\pm\left(\frac{1}{n}\sum_{i=1}^n \widehat{\tau^\pm}(X_i) - \tau^\pm(X_i)\right) = \pm\left(\frac{1}{n}\sum_{i=1}^n \widehat{\phi}^\pm_{\tau,i} - \tau^\pm(X_i)\right)$$

By applying Chebyshev's inequality to the average of zero-meaned bounded random variables $\widehat{\tau^\pm}(X_i) - \tau^\pm(X_i) - \mathcal{E}^\pm_\tau(X_i;\widehat{\eta})$, we can further obtain:

$$\begin{aligned}
\pm\left(\frac{1}{n}\sum_{i=1}^n \widehat{\tau^\pm}(X_i) - \tau^\pm(X_i)\right) &= \pm\frac{1}{n}\sum_{i=1}^n \mathcal{E}^\pm_\tau(X_i;\widehat{\eta}) - O_p(n^{-1/2}) \\
&\geq -\frac{1}{n}\sum O\left(|\widehat{e}(X) - e^*(X)| \sum_a |\widehat{\rho}_\pm(X,a) - \rho^*_\pm(X,a,\widehat{q}_\pm)|\right) \\
&\quad - O_p(n^{-1/2}) \\
&\geq -O\left(\|\widehat{e} - e^*\| \sum_a \|\widehat{\rho}_\pm(\cdot,a) - \rho^*_\pm(\cdot,a,\widehat{q}_\pm)\|\right) - o_p(1) = -o_p(1),
\end{aligned}$$

demonstrating the desired bound. □

Proof of Corollary B.1. The L_2 convergence rate for a Hölder β-smooth functions in d dimension is $O_P(n^{-1/(2+d/\beta)})$. Taking $e, \rho_\pm, q_\pm$ to be γ_e, γ_ρ, and γ_q-Hölder, we have that their convergence rates are $O_P(n^{-1/(2+d/\gamma_e)}), O_P(n^{-1/(2+d/\gamma_\rho)})$, and $O_P(n^{-1/(2+d/\gamma_q)})$ respectively. Thus, the L_2 conditional bias in Theorem 3.8 is bounded above by a term that is $O_p(n^{-2/(2+d/\gamma_q)} + n^{-(1/(2+d/\gamma_e)})$. Applying Corollary 3.9 with a β-smooth function class $\mathcal{F}$, we obtain the desired rate $O_p(n^{-1/(2+d/\beta)} + n^{-2/(2+d/\gamma_q)} + n^{-(1/(2+d/\gamma_e)+1/(2+d/\gamma_\rho)})$. The rest follows by algebraic manipulation. □

Proof of Theorem B.2. We first derive:

$$
\begin{aligned}
&\left|\widehat{\tau^\pm}(x) - \tau^\pm(x)\right| = \\
&\left|\tilde{\tau}^\pm(x) - \tau^\pm(x) + \widehat{\tau^\pm}(x) - \tilde{\tau}^\pm(x)\right| \\
&= \left|\tilde{\tau}^\pm(x) - \tau^\pm(x) + \frac{1}{n}\sum_{i=1}^n w_i(x)\left(\phi_\tau^\pm(Z_i, \widehat{\eta}) - \phi_\tau^\pm(Z_i, \eta^*)\right)\right| \\
&= \left|\tilde{\tau}^\pm(x) - \tau^\pm(x) + \frac{1}{n}\sum_{i=1}^n w_i(x)\left(\mathcal{E}_\tau^\pm(X_i; \widehat{\eta}) + \phi_\tau^\pm(Z_i, \widehat{\eta}) - \phi_\tau^\pm(Z_i, \eta^*) - \mathcal{E}_\tau^\pm(X_i; \widehat{\eta})\right)\right| \\
&\leq \left|\tilde{\tau}^\pm(x) - \tau^\pm(x)\right| + \left|\frac{1}{n}\sum_{i=1}^n w_i(x)\mathcal{E}_\tau^\pm(X_i; \widehat{\eta})\right| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^n w_i(x)\left(\phi_\tau^\pm(Z_i, \widehat{\eta}) - \phi_\tau^\pm(Z_i, \eta^*) - \mathcal{E}_\tau^\pm(X_i; \widehat{\eta})\right)\right|
\end{aligned}
$$

Since $\phi_\tau^\pm(Z_i, \widehat{\eta}) - \phi_\tau^\pm(Z_i, \eta) - \mathcal{E}_\tau^\pm(X_i; \widehat{\eta})$ is zero-meaned conditional on X and nuisances (including weights), we can apply Chebyshev's inequality to randomness in $(A, Y) \mid X$ to obtain:

$$
\begin{aligned}
\left|\widehat{\tau^\pm}(x) - \tau^\pm(x)\right| \leq \left|\tilde{\tau}^\pm(x) - \tau^\pm(x)\right| + \left|\frac{1}{n}\sum_{i=1}^n w_i(x)\mathcal{E}_\tau^\pm(X_i; \widehat{\eta})\right| \\
+ O_p\left(\|\widehat{\phi_\tau^\pm} - \phi_\tau^\pm - \mathcal{E}_\tau^\pm\|_{w^2}\frac{1}{n^2}\sum_{i=1}^n w_i(x)^2\right)
\end{aligned}
$$

Since $\mathcal{E}_\tau^\pm(X; \widehat{\eta}) = \mathbb{E}\left[\widehat{\phi_\tau^\pm} - \phi_\tau^\pm \mid X\right]$, we can take advantage of the weighted L^2 norm and the weak law of large numbers to further bound $\|\widehat{\phi_\tau^\pm} - \phi_\tau^\pm - \mathcal{E}_\tau^\pm\|_{w^2} \leq \|\widehat{\phi_\tau^\pm} -$

$\phi_\tau^\pm\|_{w^2} + o_p(1)$:

$$\left|\widehat{\tau}^\pm(x) - \tau^\pm(x)\right| \leq \left|\tilde{\tau}^\pm(x) - \tau^\pm(x)\right| + b_n^\pm(x) + O_p\left((\|\widehat{\phi}_\tau^\pm - \phi_\tau^\pm\|_{w^2} + o_p(1))\left(\frac{1}{n^2}\sum_{i=1}^n w_i(x)^2\right)^{1/2}\right),$$

which is the desired inequality. □

Proof of Corollary B.3. By Stone [1977] Theorem 1 and since $|\phi_\tau^\pm|$ is bounded, $\tilde{\tau}(x) \to^p \mathbb{E}[\phi_\tau^\pm(Z,\eta) \mid X = x] = \tau^\pm(x)$. For the second term, we use the Theorem 3.8 and the supremum assumptions to derive:

$$\begin{aligned}
|b_n^\pm(x)| &\leq \frac{1}{n}\sum_{i=1}^n |w_i(x)| \sup_x |\mathcal{E}_\tau^\pm(x;\widehat{\eta})\| \\
&\lesssim \left(\sup_x |\widehat{e}(x) - e^*(x)|\right)\left(\sup_{x,a} |\widehat{\rho}_\pm(x,a) - \rho_\pm^*(x,a)|\right) + \sup_{x,a}(\widehat{q}_\pm(x,a) - q_\pm^*(x,a))^2 \\
&= o_p(1)
\end{aligned}$$

For the final term, the sup consistency implies $\|\widehat{\phi}_\tau^\pm - \phi_\tau^\pm\| = o_p(1)$. We also have:

$$\frac{1}{n^2}\sum_{i=1}^n w_i(x)^2 \leq \left(\frac{1}{n}\sum_{i=1}^n |w_i(x)|\right)^2 = O_p(1)$$

So that $O_p\left(\frac{1}{n^2}\sum_{i=1}^n w_i(x)^2\|\widehat{\phi}_\tau^\pm - \phi_\tau^\pm\|_{w^2}\right) = o_p(1)$. □

Proof of Corollary B.4. If we define $\bar{\tau}^\pm(x)$ for the linear smoother estimate that uses $\bar{\eta}$ as first-stage nuisances, we can similarly argue that:

$$\begin{aligned}
\left|\widehat{\tau}^\pm(x) - \bar{\tau}^\pm(x)\right| &= \left|\tilde{\tau}^\pm(x) - \bar{\tau}^\pm(x) + \widehat{\tau}^\pm(x) - \tilde{\tau}^\pm(x)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^n |w_i(x)| \left|\mathcal{E}_\tau^\pm(X_i;\widehat{\eta}) - \mathcal{E}_\tau^\pm(X_i;\bar{\eta})\right| \\
&\quad + O_p\left(\left(\|\phi_\tau^\pm(\cdot,\widehat{\eta}) - \phi_\tau^\pm(\cdot,\bar{\eta})\|_{w^2} + o_p(1)\right)\left(\frac{1}{n^2}\sum_{i=1}^n w_i(x)^2\right)^{1/2}\right) \\
&= o_p(1)
\end{aligned}$$

By Stone [1977] Theorem 1, $\bar{\tau}^{\pm}(x) \to^{p} \mathbb{E}[\phi_{\tau}^{\pm}(Z, \bar{\eta}) \mid X = x]$.

By Theorem 3.8, $\pm\left(\mathbb{E}[\phi_{\tau}^{\pm}(Z, \bar{\eta}) \mid X = x] - \tau^{\pm}(x)\right) \geq 0$.

Therefore $\pm\left(\widehat{\tau}^{\pm}(x) - \tau^{\pm}(x)\right) \geq -o_p(1)$. □

B.5 Detailed Algorithm

We present a more detailed version of the B-Learner pseudocode in Algorithm B.1.

Algorithm B.1 The B-Learner: Detailed

Input: Data $\{(X_i, A_i, Y_i) : i \in \{1, \ldots, n\}\}$, folds $K \geq 2$, sensitivity parameter $\Lambda \geq 1$, nuisance estimators, regression learner $\widehat{\mathbb{E}}_n$

for $k \in \{1, \ldots, K\}$ **do**
  Set $\mathcal{S}_k = \{(X_i, A_i, Y_i) : i \not\equiv k - 1 \pmod{K}\}$
  Using $\mathcal{S}_k$, learn outcome model $\widehat{\mu}^{(k)}(x, a) = \widehat{\mathbb{E}}[Y \mid X = x, A = a]$
  Learn propensity model $\widehat{e}^{(k)}(x) = \widehat{p}(A = 1 \mid X = x)$
  Learn conditional outcome quantile models $\widehat{q}_{+}^{(k)}(x, a) = \inf\{\beta : F(\beta \mid X = x, A = a) \geq \Lambda/(\Lambda + 1)\}$ and $\widehat{q}_{-}^{(k)}(x, a) = \inf\{\beta : F(\beta \mid X = x, A = a) \geq 1/(\Lambda + 1)\}$
  Learn conditional value-at-risk models $\widehat{\mathrm{CVaR}}_{\pm}^{(k)}(x, a) = \widehat{q}_{\pm}^{(k)}(x, a) + (\Lambda + 1)\widehat{\mathbb{E}}\left[\{Y - \widehat{q}_{\pm}^{(k)}(x, a)\}_{\pm} \mid X = x, A = a\right]$
  Set $\widehat{\rho}_{\pm}^{(k)}(x, a) = \Lambda^{-1}\widehat{\mu}^{(k)}(x, a) + (1 - \Lambda^{-1})\widehat{\mathrm{CVaR}}_{\pm}^{(k)}(x, a)$
  for each i such that $i \equiv k - 1 \pmod{K}$ **do**
    Set $R_{\pm,i} = \Lambda^{-1}Y_i + (1 - \Lambda^{-1})\left(\widehat{q}_{\pm}^{(k)}(X_i, A_i) + \frac{1}{1-\alpha}\{Y_i - \widehat{q}_{\pm}^{(k)}(X_i, A_i)\}_{\pm}\right)$
    Set pseudo-outcomes for $Y^{\pm}(X, 1)$:
      $\widehat{\phi}_{1,i}^{\pm} = A_i Y_i + (1 - A_i)\widehat{\rho}_{\pm}^{(k)}(X_i, 1) + \frac{(1-\widehat{e}^{(k)}(X_i))A_i}{\widehat{e}^{(k)}(X_i)}\left(R_{\pm,i} - \widehat{\rho}_{\pm}^{(k)}(X_i, 1)\right)$
    Set pseudo-outcomes for $Y^{\pm}(X, 0)$:
      $\widehat{\phi}_{0,i}^{\pm} = (1 - A_i)Y_i + A_i\widehat{\rho}_{\pm}^{(k)}(X_i, 0) + \frac{\widehat{e}^{(k)}(X_i)(1-A_i)}{1-\widehat{e}^{(k)}(X_i)}\left(R_{\pm,i} - \widehat{\rho}_{\pm}^{(k)}(X_i, 0)\right)$
    Set pseudo-outcomes for CATE: $\widehat{\phi}_{\tau,i}^{\pm} = \widehat{\phi}_{1,i}^{\pm} - \widehat{\phi}_{0,i}^{\mp}$
  end for
end for
Create datasets $\mathcal{T}^{\pm} = \{(X_i, \widehat{\phi}_{\tau,i}^{\pm})\}$
Learn upper- and lower-bound functions $\widehat{\tau}^{\pm}(x) = \widehat{\mathbb{E}}_n\left[\widehat{\phi}_{\tau}^{\pm} \mid X = x\right]$ from the datasets $\mathcal{T}^{\pm}$

Output: $\widehat{\tau}^{\pm}$

B.6 Additional Experimental Details

The replication code for all simulations is distributed under an MIT license.

B.6.1 Simulated Data

The results in Section 3.6 were obtained using an Amazon Web Services instance with 32 vCPUs and 64 GiB of RAM. For the Random Forest (RF) models, we use the `RandomForestRegressor` model from `scikit-learn`. For Gaussian Kernels (GK), we use the `RBF` (radial basis function) method from `scikit-learn`. Finally, for the Bayesian Neural Networks (NN) we use several functions from the `PyTorch` package. The quantile estimators use weights from the nuisance regressors when RFs or GKs are used or are calculated from the sampled outcome distributions when NNs are used. We include the hyperparameters for the different models used with the synthetic data in Table B.2.

Table B.2: Hyperparameters for the synthetic-data models in the Chapter 3 hidden-confounding experiments.

Model	Hyperparameter	Value
Random Forest (`scikit-learn`)	max_depth	6
	min_samples_leaf	0.05
RBF (`scikit-learn`)	length_scale	$0.9 \times n^{-\frac{1}{4+d}}$
Neural Network (`PyTorch`)	hidden units	100
	network depth	4
	negative slope	0.3
	dropout rate	0.2
	batch size	50
	learning rate	$5e$-4

B.6.2 IHDP Dataset

We use Jesson et al. [2021]'s hidden confounding version of the Infant Health and Development Program (IHDP) that was introduced by Hill [2011]. The data comes from an experiment that targeted "low-birth-weight, premature in-

Table B.3: Continuous covariates in the IHDP dataset used in the Chapter 3 semi-synthetic experiment, together with their descriptions.

Covariate	Description
x_1	birth weight
x_2	head circumference
x_3	number of weeks pre-term
x_4	birth order
x_5	"neo-natal health index"
x_6	mom's age

fants, and provided the treatment group with both intensive high-quality child care and home visits from a trained provider" [Hill, 2011]. For the purpose of simulating an observational study, Hill [2011] generates simulated outcomes using the following features: measurements on the child–birth weight, head circumference, weeks born preterm, birth order, firstborn, neonatal health index, sex, twin status–as well as behaviors engaged in during pregnancy–smoked cigarettes, drank alcohol, took drugs–and measurements on the mother at the time she gave birth–age, marital status, educational attainment (did not graduate from high school, graduated from high school, attended some college but did not graduate, graduated from college), whether she worked during pregnancy, whether she received prenatal care, and the site (8 total) in which the family resided at the start of the intervention. A non-random portion of the treatment group, the children of non-white mothers, are excluded from the study in order to mimic confounding in an otherwise randomized trial. Covariates consist of 6 continuous variables and 19 binary variables. We use the covariate descriptions from Jesson et al. [2021] which we replicate in Tables B.3 and B.4 for completeness. The dataset consists of 747 samples, of which 139 are in the treatment group.

Table B.4: Binary covariates in the IHDP dataset used in the Chapter 3 semi-synthetic experiment. Covariates $x_9 - x_{18}$ describe maternal attributes; "College" corresponds to $x_{10} - x_{12}$, and site 8 corresponds to $x_{19} - x_{25} = 0$. The table also reports the frequency of occurrence for each binary covariate, $p(x = 1)$, as well as the adjusted mutual information $I(x; \mathrm{t})$ between the binary covariate and the treatment variable.

Covariate	Description	$I(x; \mathrm{t})$	$p(x = 1)$
x_7	child's gender (female= 1)	0.00	0.51
x_8	is child a twin	0.00	0.09
x_9	married when child born	**0.02**	**0.52**
x_{10}	left High School	0.00	0.36
x_{11}	completed High School	0.00	0.27
x_{12}	some College	0.00	0.22
x_{13}	child is first born	0.00	0.36
x_{14}	smoked cigarettes when pregnant	**0.01**	**0.48**
x_{15}	consumed alcohol when pregnant	0.00	0.14
x_{16}	used drugs when pregnant	0.00	0.96
x_{17}	worked during pregnancy	**0.01**	**0.59**
x_{18}	received any prenatal care	**0.01**	0.96
x_{19}	site 1	0.00	0.14
x_{20}	site 2	**0.01**	0.14
x_{21}	site 3	0.00	0.16
x_{22}	site 4	**0.01**	0.08
x_{23}	site 5	**0.02**	0.07
x_{24}	site 6	**0.01**	0.13
x_{25}	site 7	**0.02**	0.16

Jesson et al. [2021] create the Hidden Confounding of IHDP by hiding the covariate x_9 from models during training, however, the causal model depends on it for the data generation. Following is the data generation process of the Hidden Confounding version of response surface B Hill [2011], we restate the data generation process from Jesson et al. [2021]:

$$\mathrm{u} \coloneqq N_{\mathrm{u}}, \tag{B.1a}$$

$$\mathbf{x} := N_{\mathbf{x}}, \tag{B.1b}$$

$$\mathrm{t} := N_{\mathrm{t}}, \tag{B.1c}$$

$$\mathrm{y} := (\mathrm{t}-1)(\exp(\beta_{\mathbf{x}}(\mathbf{x}+\mathbf{w})+\beta_{\mathrm{u}}(\mathrm{u}+0.5))+N_{\mathrm{Y}^0})+\mathrm{t}(\beta_{\mathbf{x}}\mathbf{x}+\beta_{\mathrm{u}}\mathrm{u}-\omega^s+N_{\mathrm{Y}^1})), \tag{B.1d}$$

where $(N_{\mathrm{u}}, N_{\mathbf{x}}, N_{\mathrm{t}}) \sim p_{\mathcal{D}}(\mathrm{x}_9, \{\mathrm{x}_1, \ldots, \mathrm{x}_8, \mathrm{x}_{10}, \ldots, \mathrm{x}_{25}\}, \mathrm{t})$, $N_{\mathrm{Y}^0} \sim \mathcal{N}(0,1)$, and $N_{\mathrm{Y}^1} \sim \mathcal{N}(0,1)$. The coefficient β_{u} is randomly sampled from $(0.1, 0.2, 0.3, 0.4, 0.5)$ with probabilities $(0.2, 0.2, 0.2, 0.2, 0.2)$, $\beta_{\mathbf{x}}$ is a vector of randomly sampled values $(0.0, 0.1, 0.2, 0.3, 0.4)$ with probabilities $(0.6, 0.1, 0.1, 0.1, 0.1)$, w is a vector with all the coordinates equals 0.5, where ω^s was chosen as in Hill [2011]: "for the s^{th} simulation, it was chosen in the overlap setting, where we estimate the effect of the treatment on the treated, such that CATT equals 4; similarly it was chosen in the incomplete setting, where we estimate the effect of the treatment on the controls so that CATC equals 4".

Following Jesson et al. [2021]'s Hidden Confounding experiment, we generate 400 realizations of the IHDP dataset, such that the seed for each realization is the corresponding index of the realization, where the indices are $0, 1, ..., 400$. Each realization is split into training (n = 470), validation (n = 202), and test (n = 75) subsets. For the B-Learner with NNs, we use the same models and hyperparameters used by *Quince* in Jesson et al. [2021]. For the B-Learner with RF base estimators, we use the `RandomForestRegressor` from `scikit-learn` and `ForestRegressor` from `econml.grf` where we control for forest growth only through the max_depth (= 6) and min_samples_leaf (= 0.01) parameters. As for *Kernel Sensitivity* and *Quince*, to replicate the results from Jesson et al. [2021], we use the same models and hyperparameters they used for the Hidden Confounding IHDP experiment. *Note:* we exclude 5 of the 400 IHDP trials from the original analysis due to poor data quality (e.g. low overlap) that affects the NN training. These issues seem to be mitigated by ensembling which is why they do

not pose a problem for the experiments in Jesson et al. [2021]. We will perform a comparison of the ensembled methods in a future iteration of this work.

B.6.3 401(k) Eligibility Study

The dataset includes 9,915 observations with 9 covariates such as age, income, education, family size, marital status, IRA participation, etc. We describe the features of the 401(k) dataset in Table B.5. In order to replicate the CATEs obtained by Chernozhukov et al. [2018a], we use the same models (`RandomForestRegressor` and `RandomForestClassifier` from `scikit-learn`) and hyperparameters (n_estimators = 100, max_depth = 7, max_features = 3, min_samples_leaf = 10) for our nuisance estimators and second stage models.

Table B.5: Covariates included in the 401(k) dataset used in the Chapter 3 application.

Name	Description	Type
age	age	continuous
inc	income	continuous
educ	years of completed education	continuous
fsize	family size	continuous
marr	marital status	binary
two_earn	whether dual-earning household	binary
db	defined benefit pension status	binary
pira	IRA participation	binary
hown	home ownership	binary
e401	401 (k) eligibility	binary
net_tfa	net financial assets	continuous

APPENDIX C

APPENDIX FOR CHAPTER 4

C.1 Notations

Table C.1: Notation used in Chapter 4.

Notation	Meaning
$\mathcal{S}, \mathcal{A}$	State and action spaces.
$\Delta(S)$	The set of distributions supported on S.
d_1	The initial state distribution.
$\Lambda(s,a)$	Tolerance parameter for kernel shift at (s,a). Takes values in $[1,\infty]$.
$\tau(s,a)$	$\tau(s,a) = \frac{1}{1+\Lambda(s,a)} \in [0, \frac{1}{2}]$.
$V^{\pm}, Q^{\pm}$	Robust value and quality functions of the target policy π_{t}.
$f(s,\pi)$	$f(s,\pi) := \mathbb{E}_{a\sim\pi(s)}[f(s,a)]$.
$U^{\pm}(s' \mid s,a)$	Robust transition kernel attaining the best- or worst-case value.
$\mathcal{T}_U, \mathcal{T}^{\pm}_{\mathsf{rob}}$	Bellman operator under U and the robust Bellman operators.
$\mathcal{J}_U$	$\mathcal{J}_U f(s,a) := \gamma \mathbb{E}_U[f(s',\pi_{\mathrm{t}}) \mid s,a] - f(s,a)$.
$\beta^{\pm}_{\tau}(s,a)$	The upper τ-th quantile of $V^+(s')$ and lower τ-th quantile of $V^-(s')$, where $s' \sim P(s,a)$.
$d^{\pi_{\mathrm{t}},\infty}_{d_1,U}$	The γ-discounted average visitation of π_{t} under the MDP with transition U, starting from d_1.
$d^{\pm,\infty}$	$d^{\pm,\infty} = d^{\pi_{\mathrm{t}},\infty}_{d_1,U^{\pm}}$.
$\nu(s), \nu(s,a)$	Data-generating distribution. $\nu(s)$ marginalizes over actions.
$w^{\pm}$	$w^{\pm} = \mathrm{d}d^{\pm,\infty}/\mathrm{d}\nu$. This is valid both as a function of s or (s,a).
$\omega(s,a)$	$\omega(s,a) = \frac{\pi_{\mathrm{t}}(a\mid s)}{\nu(a\mid s)}$.
x_+, x_-	$\max(0,x)$ and $\min(0,x)$, respectively, for $x \in \mathbb{R}$.
$x \lesssim y$	$x \le Cy$ for some constant C.
$\mathbb{E}_n$	Empirical average over n samples.
$\|f\|_p$	L^p norm, $(\mathbb{E}\|f(X)\|^p)^{1/p}$.
$f^\star$	True (oracle) value of a parameter or function f.
$f, \bar{f}$	Putative value of a parameter or function f.
$\widehat{f}$	Estimated value of a parameter or function f.

C.2 Results for OPE Under Best-Case Perturbations

In this section, we present analogous results for the best-case perturbation under the uncertainty set, corresponding to the supremum case of Eq. (4.2). We derive a similar orthogonal estimator with the properties outlined in Theorem 4.11, following the same reasoning presented in the main text.

Q^+ Identification and Estimation We give the results of Lemma 4.1 for $\mathcal{T}^+_{\mathsf{rob}}$:

$$\mathcal{T}^+_{\mathsf{rob}} q(s,a) = r(s,a) + \gamma \Lambda^{-1}(s,a) \mathbb{E}[v(s') \mid s,a] + \gamma(1 - \Lambda^{-1}(s,a))\, \mathrm{CVaR}^+_{\tau(s,a)}[v(s') \mid s,a].$$

Next, applying Assumption 4.2 and Assumption 4.3 to $\mathcal{T}^+_{\mathsf{rob}}$, we derive from Theorem 4.4 for Q^- that:

$$\|\widehat{q}^+_M - Q^+\|_{d_1} \lesssim (1-\gamma)^{-2}\big(\sqrt{C^+_{d_1}} \cdot \varepsilon^{\mathcal{Q}}_n + \mathrm{err}^2_{\mathsf{QR}}(n/2M, \delta/2M)\big), \text{ and}$$
$$\left|(1-\gamma)\mathbb{E}_{d_1}[\widehat{v}^+_M(s_1)] - V^+_{d_1}\right| \lesssim \gamma^M + (1-\gamma)^{-1}\big(\sqrt{C^+_{d_1}} \cdot \varepsilon^{\mathcal{Q}}_n + \mathrm{err}^2_{\mathsf{QR}}(n/2M, \delta/2M)\big).$$

w^+ Identification and Estimation We first state the identification result for U^- as in Lemma 4.5:

$$U^+(s' \mid s,a)/P(s' \mid s,a) = \Lambda^{-1}(s,a) + (1-\Lambda^{-1})\tau(s,a)^{-1}\mathbb{I}[(V^+(s') - \beta^+_\tau(s,a)) \geq 0].$$

Then, under Assumption 4.6 and Assumption 4.7 formulated for U^+, the minimax rates from Theorem 4.8 are given by:

$$\left\|\mathcal{J}'_{U^+}(\widehat{w} - w^+)\right\|_2 \lesssim \varepsilon^{\mathcal{W}}_n + \|\widetilde{\zeta}^+ - \zeta^+\|_\infty + \sqrt{\log(1/\delta)/n}.$$

Orthogonal and Efficient Estimator for $V^+_{d_1}$. Let the set of nuisance parameters be denoted by $\eta^+ = (w^+, q^+, \beta^+)$. Then, the (recentered) efficient influence function (R)EIF (see Theorem 4.9) for in $V^+_{d_1}$ is formulated as:

$$\psi(s,a,s';\eta^+) = V^+_{d_1} + w^+(s,a)(r(s,a) + \gamma\rho^+(s,a,s';v^+,\beta^+) - q^+(s,a)), \quad \text{where}$$

Algorithm C.1 Orthogonal Estimator for $V_{d_1}^+$

Input: Dataset $\mathcal{D}$, number of splits K.
for $k = 1, 2, \ldots, K$ **do**
    Use data $\mathcal{D} \setminus \mathcal{D}_k$ to learn $(q^{+,[k]}, \beta^{+,[k]})$ with Algorithm 4.1 and $w^{+,[k]}$ with Algorithm 4.2.
    for $i = \lfloor (k-1)n/K \rfloor, \ldots, \lfloor kn/K \rfloor - 1$ **do**
        $\psi_i^+ = \psi(s_i, a_i, s_i', \widehat{\eta}^+)$.
    end for
end for
Output: $\widehat{V}_{d_1}^+ = \frac{1}{n} \sum_{i=1}^n \psi_i^+$.

$$\rho^+(s, a, s'; v^+, \beta^+) = \Lambda(s,a)^{-1} v^+(s') + (1 - \Lambda(s,a)^{-1})(\beta^+(s,a) + \tau^{-1}(v^+(s') - \beta^+(s,a))_+).$$

Using this (R)EIF, the orthogonal estimator for $V_{d_1}^+$ is presented in Algorithm C.1. We now restate Theorem 4.11 for $\widehat{V}_{d_1}^+$:

Theorem C.1 (Efficiency of $\widehat{V}_{d_1}^+$)**.** *Let $r_{n,p}^w, r_{n,p}^q, r_{n,p}^\beta$ be functions of $n = |\mathcal{D}|$ such that $\|\mathcal{J}'_{U^+}(\widehat{w}^{+,[k]} - w^*)\|_p \le r_{n,p}^w$, $\|\widehat{q}^{+,[k]} - q^*\|_p \le r_{n,p}^q$, and $\|\beta^{+,[k]} - \beta^*\|_p \le r_{n,p}^\beta$ for any $k \in [K]$. Furthermore, assume that the regularity conditions in Assumption 4.6 hold. Then:*

$$|\widehat{V}_{d_1}^+ - V_{d_1}| \lesssim O_p(n^{-1/2}) + O_p(r_{n,2}^w r_{n,2}^q + (r_{n,\infty}^q)^2 + (r_{n,\infty}^\beta)^2) \qquad \text{(Rates)}$$

Furthermore, if $r_{n,2}^w \vee r_{n,2}^q = o_p(1)$, $r_{n,2}^w r_{n,2}^q = o_p(n^{-1/2})$, $r_{n,\infty}^q = o_p(n^{-1/4})$, and $r_{n,\infty}^\beta = o_p(n^{-1/4})$, then $\widehat{V}_{d_1}^+$ satisfies:

$$\sqrt{n}(\widehat{V}_{d_1}^+ - V_{d_1}) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad \Sigma = \mathrm{Var}(\psi(s, a, s'; \eta^+)). \qquad \text{(Normality \& Efficiency)}$$

Moreover, Σ is the minimum achievable asymptotic variance among RAL estimators in the nonparametric model for (s, a, s') (the efficiency bound).

C.3 Additional Related Works

Robust MDPs There is a rich literature on Robust MDPs [Iyengar, 2005, Wiesemann et al., 2013, Mannor et al., 2016, Goyal and Grand-Clement, 2023]

with s, a-rectangular uncertainty sets, but these foundational works assumed knowledge of the transition kernel. Recently, learning-based robust MDP algorithms have been proposed for uncertainty sets under the total variation [Panaganti et al., 2022, Kumar et al., 2023] and more generally L_p balls [Kumar et al., 2022]. These L_p uncertainty sets are additive in nature, *i.e.*, the adversary adds or subtracts a vector in the ℓ_p ball to $P(\cdot \mid s, a)$, whereas our uncertainty set is multiplicative in nature, *i.e.*, the adversary can multiply or divide a bounded factor and is more commonly used in causal inference to model unobserved confounding. In the contextual bandit setting, [Kallus et al., 2022] also derived efficiency bounds for robust OPE where both state distribution and reward distributions may shift – their work is however restricted to the one-step bandit setting while our full RL setting is more challenging.

Risk-Sensitive RL Risk-sensitive RL is the problem of optimizing the risk measure of cumulative rewards [Howard and Matheson, 1972] and is tightly related to robust MDPs [Chow et al., 2015]. For example, as we proved in Lemma 4.1, the MSM uncertainty set is indeed equivalent to risk-sensitive RL with the dynamic risk measure $\Lambda\mathbb{E} + (1 - \Lambda)\,\mathrm{CVaR}_\tau$. We note that efficient online RL algorithms have been proposed for similar measures Du et al. [2022], Xu et al. [2023]. Static risk-sensitive RL also modifies the Bellman equations in an augmented MDP [Wang et al., 2023a, 2024b]. Our focus is on deriving the optimal *off-policy evaluation* estimators for the problem, which involves a different set of challenges such as deriving the efficiency bound and ensuring sharpness guarantees even when nuisances are estimated slowly.

C.4 Additional Technical Details

C.4.1 Higher Order Norms via Smoothness

For any $x \in \mathbb{R}^+$, define $\lfloor x \rfloor$ as the greatest integer that is strictly less than x, and let x and $\{x\} = x - \lfloor x \rfloor$ represent the fractional part. Thus, we obtain the distinct decomposition $x = \lfloor x \rfloor + \{x\}$, where $\lfloor x \rfloor \in \mathbb{N}$ and $\{x\} \in (0, 1]$.

Definition C.2 (α-smooth functions)**.** Given $\alpha \in (0, \infty)$ and $\mathcal{X} \subseteq \mathbb{R}^m$, $f : \mathcal{X} \to \mathbb{R}$ is an α-smooth function if (1) the mixed derivatives up to $\lfloor \alpha \rfloor$-order exist and are bounded; and (2) all $\lfloor \alpha \rfloor$-order derivatives are $\{\alpha\}$-Hölder continuous [Leoni, 2017].

Lemma C.3 (L^∞ Bound for α-Smooth Functions)**.** *Let $f : \mathcal{X} \to \mathbb{R}, \mathcal{X} \subseteq \mathbb{R}^m$ be an α-smooth function as in Definition C.2. Then, if $\mathcal{X}$ is $\mathbb{R}^m$, a half-space or a bounded Lipschitz domain in $\mathbb{R}^m$, there exists a constant C such the following inequality holds:*

$$\|f\|_\infty \leq C \|f\|_p^{\frac{p\alpha}{p\alpha+m}}.$$

Proof. This lemma is a direct application of the fractional Gagliardo-Nirenberg interpolation inequality (Theorem 1 in Brezis and Mironescu [2019]) from the functional analysis literature. For a more comprehensive exposition on this result, see Appendix A.1 in Oprescu et al. [2024]. □

C.4.2 Localized Rademacher Complexity and Critical Radius

Here, we recap the localized Rademacher complexity and critical radius which is a standard complexity measure for obtaining fast rates for squared loss [Wainwright, 2019]. Let $\mathcal{G}$ be a class of functions $g : \mathcal{Z} \to \mathbb{R}$. Given n datapoints $z_1, z_2, \ldots, z_n$, the empirical localized Rademacher complexity is:

$$\mathcal{R}_n(\varepsilon, \mathcal{G}) := \mathbb{E}_\sigma \left[\sup_{g \in \mathcal{G} : \|g\|_n \leq \varepsilon} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i g(z_i) \right],$$

where $\mathbb{E}_\sigma$ is expectation over n independent Rademacher random variables $\sigma_1, \sigma_2, \ldots, \sigma_n$, *i.e.*, $\mathbb{E}_\sigma[\cdot] = \frac{1}{2^n}\sum_{\sigma\in\{-1,1\}^n}[\cdot]$. Note that when $\varepsilon = \infty$, there is no localization and $\mathcal{R}_n(\infty, \mathcal{G})$ reduces to the vanilla Rademacher complexity. Let $C := \sup_{g\in\mathcal{G}} \|g\|_\infty$ be the envelope of $\mathcal{G}$. Then, the critical radius of $\mathcal{G}$ with n, called ε_n, is the smallest ε that satisfies $\mathcal{R}_n(\varepsilon, \mathcal{G}) \le \varepsilon^2/C$.

Unless otherwise stated, we will posit that $\mathcal{G}$ is star-shaped: there exists $g_0 \in \mathcal{G}$ such that for all $g \in \mathcal{G}$ and $\alpha \in [0, 1]$, we have $\alpha g_0 + (1-\alpha)g \in \mathcal{G}$. If not, we can replace $\mathcal{G}$ by its star-hull, *i.e.*, the smallest star-shaped set containing $\mathcal{G}$. We will also posit that $\mathcal{G}$ is symmetric for simplicity.

The critical radius is a well-studied quantity in statistics [Wainwright, 2019] and also recently in RL [Duan et al., 2021, Uehara et al., 2021]. For example if $\mathcal{G}$ has d VC-subgraph dimension, then w.p. $1-\delta$, $\varepsilon_n \le O(\sqrt{d \log n/n})$. For nonparametric models with metric entropy at most $1/t^\beta$, the critical radius can also be bounded by $O(n^{-1/(\max(2+\beta, 2\beta))})$ [Uehara et al., 2021], *e.g.*, is $O(n^{-1/4})$ if $\beta = 2$.

C.5 Proofs for Identification Results

Identification of robust Q

Lemma 4.1. *Set $\tau(s,a) = (\Lambda(s,a)+1)^{-1}$. Then, for any $q : \mathcal{S}\times\mathcal{A} \to \mathbb{R}$,*

$$\mathcal{T}^-_{\mathsf{rob}} q(s,a) = r(s,a) + \gamma\Lambda^{-1}(s,a)\mathbb{E}[v(s') \mid s,a] + \gamma(1-\Lambda^{-1}(s,a))\,\mathrm{CVaR}^-_{\tau(s,a)}[v(s') \mid s,a],$$

where $v(s') = \mathbb{E}_{a'\sim\pi_t(s')}[q(s',a')]$, and $\mathbb{E}, \mathrm{CVaR}_\tau$ are under the observed kernel $P(\cdot \mid s,a)$.

Proof. Consider the uncertainty set in $\mathcal{T}_{\mathsf{rob}}$ where the constraint on U (Eq. (4.1)) can be rewritten as:

$$0 \le \frac{U(s'|s,a) - \Lambda^{-1}(s,a)P(s'|s,a)}{P(s'|s,a)} \le \Lambda(s,a) - \Lambda^{-1}(s,a).$$

Therefore, we can write $U(s' \mid s,a) = \Lambda^{-1}(s,a)P(s' \mid s,a) + (1-\Lambda^{-1})G(s' \mid s,a)$ where we define $G(s' \mid s,a) := \frac{U(s'|s,a)-\Lambda^{-1}(s,a)P(s'|s,a)}{1-\Lambda^{-1}(s,a)}$. Thus, the constraints on G are that $G(\cdot \mid s,a) \ll P(\cdot \mid s,a)$ and $\|\frac{\mathrm{d}G(s'|s,a)}{\mathrm{d}P(s'|s,a)}\| \leq \Lambda(s,a)+1$. Setting $\tau(s,a) = \frac{1}{\Lambda(s,a)+1}$, we can apply the primal form of CVaR [Dorn et al., 2025a, Ang et al., 2018] to obtain

$$\inf_{G \ll P : \|\frac{\mathrm{d}G(\cdot|s,a)}{\mathrm{d}P(\cdot|s,a)}\|_\infty \leq \tau^{-1}(s,a)} \mathbb{E}_G[f(s')] = \mathrm{CVaR}^-_{\tau(s,a)}[f(s') \mid s,a].$$

Therefore, the supremum in $\mathcal{T}_{\mathsf{rob}}$ can be expressed as $\Lambda^{-1}(s,a)$ times the expectation under nominal P and $(1-\Lambda^{-1}(s,a))$ times the above CVaR expression, which finishes the proof of the $-$ case.

For the $+$ case, we can simply use sup instead of inf and upper CVaR instead of lower CVaR. □

Identification of robust kernel and visitation

Lemma 4.5. *Suppose $F^-(\beta^-_\tau(s,a) \mid s,a) = \tau$, where $\beta^-_\tau(s,a)$ is the lower τ-th quantile of $F^-(\cdot \mid s,a)$. Then,*

$$U^-(s' \mid s,a)/P(s' \mid s,a) = \Lambda^{-1}(s,a) + (1-\Lambda^{-1})\tau(s,a)^{-1}\mathbb{I}[(V^-(s') - \beta^-_\tau(s,a)) \leq 0]. \quad (4.4)$$

Lemma C.4. *Fix any $v : \mathcal{S} \to \mathbb{R}$ and define the pushforward $F_v(y \mid s,a) = P(v(s') \leq y \mid s,a)$. Suppose $F_v(\beta^\pm_{\tau,F_v(\cdot|s,a)}(s,a) \mid s,a) = \frac{1}{2} \pm (\frac{1}{2} - \tau)$, where $\beta^\pm_{\tau,F_v}$ is the upper/lower τ-quantile of F_v. Then, $\sup_{U \in \mathcal{U}(P)} \mathbb{E}_U[v(s') \mid s,a] = \mathbb{E}_{s' \sim U^+_v(s,a)}[v(s')]$ and $\inf_{U \in \mathcal{U}(P)} \mathbb{E}_U[v(s') \mid s,a] = \mathbb{E}_{s' \sim U^-_v(s,a)}[v(s')]$, where*

$$U^\pm_v(s' \mid s,a)/P(s' \mid s,a) = \Lambda^{-1}(s,a) + (1-\Lambda^{-1})\tau(s,a)^{-1}\mathbb{I}[\pm(v(s') - \beta^\pm_{\tau,F_v(\cdot|s,a)}(s,a)) \geq 0].$$

Proof. We start with some intuitions. First, if the CDF of $v(s')$ is differentiable $\beta^+_\tau(s,a)$, then $\mathrm{CVaR}^+_\tau(v(s') \mid s,a) = \mathbb{E}[v(s') \mid f(s') \geq \beta^+_\tau(s,a), s,a]$ and the result follows immediately from Lemma 4.1 by noticing that the form of U^+ exactly recovers the convex combination of expectation and CVaR. Alternatively, one

can use the closed form solution of the primal CVaR as derived in [Ang et al., 2018] to obtain the result.

We now provide a formal proof. Fix any s, a and let $\tau = \tau(s,a)$. Fix any function $v(s') \in \mathbb{R}$. We want to show that the worst-case $U^+ = \arg\max_{U \in \mathcal{U}(P)} \mathbb{E}_U[v(s') \mid s,a]$ has a closed form expression as shown in line 725. By the proof of Lemma 3.1 above, we can rewrite $U^+(s' \mid s,a) = \Lambda^{-1}(s,a)P(s' \mid s,a) + (1 - \Lambda^{-1}(s,a))G^+(s' \mid s,a)$, where $G^+ = \arg\max_{G \ll P : |dG(\cdot|s,a)/dP(\cdot|s,a)|_\infty \le \tau^{-1}(s,a)} \mathbb{E}_G[v(s')]$. Thus, it suffices to simplify G^+. To do so, we invoke the premise that the CDF of $v(s')$ is differentiable at β^+_τ, i.e. $F_v(\beta^+_{\tau,F_v}(s,a) \mid s,a) = 1 - \tau$. This implies that the CVaR is exactly the conditional expectation of the $1 - \tau(s,a)$-fraction of best outcomes, i.e. $\text{CVaR}^+_\tau(v(s') \mid s,a) = \mathbb{E}[v(s') \mid v(s') \ge \beta^+_\tau(s,a), s, a]$, which in turn is equal to $\tau^{-1}\mathbb{E}[v(s')\mathbb{I}[v(s') \ge \beta^+_\tau(s,a)] \mid s,a]$. Thus, $G^+(s' \mid s,a) = \tau^{-1}P(s' \mid s,a)\mathbb{I}[v(s') \ge \beta^+_\tau(s,a)]$. This concludes the proof for the + case. The proof for the − case follows identical steps. □

C.6 Proofs for Robust FQE

We prove a more general result with approximate completeness, which shows that Theorem 4.4 is robust to approximate completeness.

Assumption C.5 (Approximate Completeness)**.** $\max_{q \in \mathcal{Q}} \min_{g \in \mathcal{Q}} \|g - \mathcal{T}^\pm_{\text{CVaR}} q\|_\nu \le \varepsilon_{\mathsf{QComp}}$.

Theorem C.6. *Assume Assumption C.5. Under the same setup as Theorem 4.4, we have*

$$\left\|\widehat{q}^\pm_K - Q^\pm\right\|_\mu \lesssim \frac{1}{(1-\gamma)^2}\Big(\sqrt{C^\pm_\mu} \cdot \left(\varepsilon^{\mathcal{Q}}_n + \varepsilon_{\mathsf{QComp}}\right) + \text{err}^2_{\mathsf{QR}}(n/2K, \delta/2K)\Big),$$

and

$$\left|V^\pm_{d_1} - (1-\gamma)\mathbb{E}_{d_1}[\widehat{q}^\pm_K(s_1, \pi_t)]\right| \lesssim \gamma^K + \frac{1}{1-\gamma}\left(\sqrt{C^\pm_\mu} \cdot \left(\varepsilon^{\mathcal{Q}}_n + \varepsilon_{\mathsf{QComp}}\right) + \text{err}^2_{\mathsf{QR}}(n/2K, \delta/2K)\right)$$

Proof. Let $U^\pm$ denote the worst-case kernel that satisfies $V^\pm_{d_1} = (1-\gamma)\mathbb{E}_{d_1} V^{\pi_t}_{U^\pm}(s_1)$. Then,

$$\begin{aligned} V^\pm_{d_1} - (1-\gamma)\mathbb{E}_{d_1}[\widehat{q}^\pm_K(s_1,\pi_t)] &= (1-\gamma)\mathbb{E}_{d_1}[V^{\pi_t}_{U^\pm}(s_1) - \widehat{q}_K(s_1,\pi_t)] \\ &= \mathbb{E}_{d^{\pi,\infty}_{U^\pm}}[\mathcal{T}^{\pi_t}_{U^\pm}\widehat{q}_K(s,a) - \widehat{q}_K(s,a)] && \text{(Lemma C.7)} \\ &\le \frac{4}{1-\gamma}\max_{k=1,2,\dots}\left\|\widehat{q}_k - \mathcal{T}^{\pi_t}_{U^\pm}\widehat{q}_{k-1}\right\|_{d^{\pi_t,\infty}_{U^\pm}} + \gamma^{K/2}. && \text{(Lemma C.8)} \end{aligned}$$

Consider any $k = 1, 2, \dots$. By definition of $U^\pm$, we have

$$\left\|\widehat{q}_k - \mathcal{T}^{\pi_t}_{U^\pm}\widehat{q}_{k-1}\right\|_{d^{\pi_t,\infty}_{U^\pm}} = \left\|\widehat{q}_k - \mathcal{T}^{\pm}_{\beta^\star_k}\widehat{q}_{k-1}\right\|_{d^{\pm,\infty}}, \qquad \text{(by def of } U^\pm\text{)}$$

where $\beta^\star_k(s,a)$ is the true quantile of $\widehat{v}_{k-1}(s')$. Denote $q^\star_k := \mathcal{T}^{\pm}_{\mathsf{rob}}\widehat{q}_{k-1}$ and let $\beta^\star_k$ be the true upper/lower quantile of $\widehat{q}_{k-1}$. Recall the population loss function is

$$\begin{aligned} L_k(q,\beta) &:= \mathbb{E}\left[\left(y^\beta_k(s,a,s') - q(s,a)\right)^2\right] \\ y^\beta_k(s,a,s') &= r(s,a) + \gamma\Lambda^{-1}(s,a)\widehat{v}_{k-1}(s') \\ &\quad + \gamma(1-\Lambda^{-1}(s,a))\Big(\beta(s,a) + \tau^{-1}(s,a)(\widehat{v}_{k-1}(s') - \beta(s,a))_\pm\Big). \end{aligned}$$

The empirical loss $\widehat{L}_k(q,\beta)$ is defined by replacing $\mathbb{E}$ with $\mathbb{E}_n$. Note that $\widehat{q}_k = \operatorname{argmin}_{q\in\mathcal{Q}}\widehat{L}_k(q,\widehat{\beta}_k)$.

Nonparametric Least Squares with Model Misspecification We will directly invoke Wainwright [2019, Theorem 13.13], which gives a fast rate for misspecified least squares with general nonparametric classes. We now bound the misspecification. Recall that at the k-th iteration, our regression Bayes-optimal is $\mathbb{E}[y^{\widehat{\beta}_k}_k(s,a,s') \mid s,a] = \mathcal{T}_{\widehat{\beta}_k}\widehat{q}_{k-1}(s,a)$. By Lemma C.14, we know this is close to $\mathcal{T}_{\beta^\star_k}\widehat{q}_{k-1}(s,a)$ with second order errors in β: for any μ, we have

$$\left\|\mathcal{T}^{\pm}_{\widehat{\beta}_k}\widehat{q}_{k-1} - \mathcal{T}^{\pm}_{\beta^\star_k}\widehat{q}_{k-1}\right\|_{d^{\pm,\infty}_\mu} \lesssim \|\widehat{\beta}_k - \beta^\star_k\|^2_\infty.$$

Finally, by approximate completeness (Assumption C.5), there exists $g \in \mathcal{Q}$ such that $\|\mathcal{T}_{\beta^\star_k}\widehat{q}_{k-1}(s,a) - g\| \le \varepsilon_{\mathsf{QComp}}$. Putting this together: for any k, there exists a

$g \in \mathcal{Q}$ such that

$$\begin{aligned}\|g - \mathcal{T}_{\widehat{\beta}_k}\widehat{q}_{k-1}(s,a)\|_{d_\mu^{\pm,\infty}} &\le \|g - \mathcal{T}_{\beta_k^\star}\widehat{q}_{k-1}(s,a)\|_{d_\mu^{\pm,\infty}} + \|\mathcal{T}_{\beta_k^\star}\widehat{q}_{k-1}(s,a) - \mathcal{T}_{\widehat{\beta}_k}\widehat{q}_{k-1}(s,a)\|_{d_\mu^{\pm,\infty}} \\ &\le \sqrt{C_\mu^\pm} \cdot \varepsilon_{\mathsf{QComp}} + \|\widehat{\beta}_k - \beta_k^\star\|_\infty^2.\end{aligned}$$

Therefore, Wainwright [2019, Theorem 13.13], along with concentration of least squares, certifies that:

$$\left\|\widehat{q}_k - \mathcal{T}_{\widehat{\beta}_k}\widehat{q}_{k-1}\right\|_{d^{\pm,\infty}} \lesssim \sqrt{C_\mu^\pm} \cdot \left(\varepsilon_{\mathsf{QComp}} + \varepsilon_n\right) + \|\widehat{\beta}_k - \beta_k^\star\|_\infty^2.$$

Therefore, we have proven:

$$\begin{aligned}\left\|\widehat{q}_k - \mathcal{T}^\pm_{\beta_k^\star}\widehat{q}_{k-1}\right\|_{d_\mu^{\pm,\infty}} &\le \left\|\widehat{q}_k - \mathcal{T}^\pm_{\widehat{\beta}_k}\widehat{q}_{k-1}\right\|_{d_\mu^{\pm,\infty}} + \left\|\mathcal{T}^\pm_{\widehat{\beta}_k}\widehat{q}_{k-1} - \mathcal{T}^\pm_{\beta_k^\star}\widehat{q}_{k-1}\right\|_{d_\mu^{\pm,\infty}} \\ &\lesssim \sqrt{C_\mu^\pm} \cdot \left(\varepsilon_{\mathsf{QComp}} + \varepsilon_n\right) + \|\widehat{\beta}_k - \beta_k^\star\|_\infty^2.\end{aligned}$$

This concludes the proof. □

Lemma C.7 (Performance Difference)**.** *For any π, transition kernel P, and function $f : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, we have*

$$V_P^\pi - \mathbb{E}_{s\sim d_1}[f(s,\pi)] = \frac{1}{1-\gamma}\mathbb{E}_{d_P^{\pi,\infty}}[\mathcal{T}_P^\pi f(s,a) - f(s,a)].$$

Proof. See Lemma C.1 of [Chang et al., 2022]. □

Lemma C.8 (Unrolling)**.** *For any π, transition kernel P, and functions $f_0, f_1, \ldots, f_K : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ satisfying $f_0(s,a) = 0$, we have:*

$$\left\|f_K - \mathcal{T}_P^\pi f_K\right\|_{d_P^{\pi,\infty}} \le \frac{4}{1-\gamma}\max_{k=1,2,\ldots}\left\|f_k - \mathcal{T}_P^\pi f_{k-1}\right\|_{d_P^{\pi,\infty}} + \gamma^{K/2}$$

Proof. See Lemma C.2 of [Chang et al., 2022]. □

C.7 Proofs for Robust Minimax Algorithm

Assumption C.9 (Approximate W-realizability and completeness)**.** Assume the following hold for $\mathcal{W}$ and $\mathcal{F}$:

(A) Approximate realizability: $\min_{w\in\mathcal{W}}\|\mathcal{J}_{U^\pm}(w^\pm - w)\|_2 \le \varepsilon_{\mathsf{WReal}}$;

(B) Approximate completeness: $\max_{w\in\mathcal{W}}\min_{f\in\mathcal{F}}\left\|f - \mathcal{J}'_{U^\pm}(w - w^\pm)\right\|_2 \le \varepsilon_{\mathsf{WComp}}$.

We prove a more general result with approximate realizability and completeness, which implies Theorem 4.8 that is robust to misspecification in its assumptions.

Theorem C.10. *Under Assumption C.9 and the same setup as Theorem 4.8, we have*

$$\left\|\mathcal{J}'_{U^\pm}(\widehat{w} - w^\pm)\right\|_2 \lesssim \varepsilon_n^{\mathcal{W}} + \|\widetilde{\zeta}^\pm - \zeta^\pm\|_\infty + \sqrt{\frac{\log(1/\delta)}{n}} + \varepsilon_{\mathsf{WReal}} + \varepsilon_{\mathsf{WComp}}.$$

Proof. For this proof, we focus on the worst-case kernel $P^\star$ of the form $\frac{P^\star(s'|s,a)}{P(s'|s,a)} = \tau^{-1}(s,a)\mathbb{I}[\zeta^\star(s,a,s') \le 0]$ where $\zeta^\star(s,a,s') = V^-(s') - \beta^-(s,a)$. This corresponds to the pure CVaR case of $\mathcal{T}^-_{\mathsf{rob}}$; the $\mathbb{E}$ part is identical to standard non-robust RL so we omit it. The best-case kernel U^+ can be handled similarly. Let $\widehat{P}(s' \mid s,a)$ denote our estimated robust kernel, which satisfies $\frac{\widehat{P}(s'|s,a)}{P(s'|s,a)} = \tau^{-1}(s,a)\mathbb{I}[\widehat{\zeta}(s,a,s') \le 0]$, where $\widehat{\zeta}(s,a,s')$ is the given prior stage estimate of $\zeta^\star(s,a,s') = V^-(s') - \beta^-(s,a)$.

The key and only difference between our Algorithm 4.2 and the MIL algorithm ($\widehat{w}_{\mathrm{mil}}$) of Uehara et al. [2021] is that our next-state samples are importance weighted with $\xi^\pm(s,a,s')$, which is the density ratio of the estimated robust kernel $\widehat{P}(s' \mid s,a)$ and the nominal kernel $P(s' \mid s,a)$. Note also that $\xi^\pm(s,a,s') \le \tau^{-1}(s,a) < \infty$, and hence $|\mathbb{E}_n[\zeta(s,a,s')f(s')] - \mathbb{E}_{s,a\sim\nu,s'\sim\widehat{P}(s,a)}[f(s')]| \lesssim \sqrt{\log(1/\delta)/n}$ w.p. $1-\delta$. Therefore, up to $O(\sqrt{\log(1/\delta)/n})$ errors, our Algorithm 4.2 can be viewed as MIL applied to the MDP with kernel $\widehat{P}$.

To invoke the result of Uehara et al. [2021, Theorem 6.1] (in MDP with kernel $\widehat{P}$), we need to show that its assumptions are met by bounding the

model misspecification, *i.e.*, Eq. (6) and Appendix C of Uehara et al. [2021]. Note that these misspecifications are w.r.t. the MDP with kernel $\widehat{P}$, since this is the MDP in which we're applying Theorem 6.1 of Uehara et al. [2021]. Specifically, the two errors we need to bound are, (A) approximate realizability: $\varepsilon_A = \min_{w \in \mathcal{W}} \|\mathcal{J}'_{\widehat{P}}(w_{\widehat{P}} - w)\|_2$; and (B) approximate completeness: $\varepsilon_B = \max_{w \in \mathcal{W}} \min_{f \in \mathcal{F}} \|f - \mathcal{J}'_{\widehat{P}}(w - w_{\widehat{P}})\|_2$ where recall that $\mathcal{J}_P$ is the linear operator defined as $\mathcal{J}_P f(s,a) := \gamma \mathbb{E}_P[f(s', \pi_t) \mid s, a] - f(s,a)$ and $\mathcal{J}'_P$ is the adjoint.

Bounding misspecifications by $\|\widehat{\zeta} - \zeta^\star\|_\infty$ Since $\zeta^\star(s,a,s')$ has a marginal CDF that's boundedly differentiable around 0 (*i.e.*, (ii) of Assumption 4.6), Kallus [2022, Lemma 3] implies that $\zeta^\star(s,a,s')$ satisfies a 1-margin (Definition C.13). Hence, Lemma C.14 and the continuity of $\zeta^\star(s,a,s')$ implies that

$$\begin{aligned}
&\Pr\Big(\mathbb{I}[\widehat{\zeta}(s,a,s') \le 0] \neq \mathbb{I}[\zeta^\star(s,a,s') \le 0]\Big) \\
&= \Pr\Big((\mathbb{I}[\widehat{\zeta}(s,a,s') \le 0] \neq \mathbb{I}[\zeta^\star(s,a,s') \le 0]), \zeta^\star(s,a,s') \neq 0\Big) \lesssim \|\widehat{\zeta} - \zeta^\star\|_\infty,
\end{aligned}$$

Thus, for any $v : \mathcal{S} \to \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}\Big|(\mathbb{E}_{\widehat{P}} - \mathbb{E}_{P^\star})[v(s') \mid s,a]\Big| &\le \mathbb{E}[\tau^{-1}(s,a)(\mathbb{I}[\widehat{\zeta}(s,a,s') \le 0] \neq \mathbb{I}[\zeta^\star(s,a,s') \le 0]) \cdot |v(s')|] \\
&\lesssim \|v\|_\infty \cdot \Pr\Big(\mathbb{I}[\widehat{\zeta}(s,a,s') \le 0] \neq \mathbb{I}[\zeta^\star(s,a,s') \le 0]\Big) \\
&\lesssim \|v\|_\infty \|\widehat{\zeta} - \zeta^\star\|_\infty,
\end{aligned}$$

or equivalently

$$\mathbb{E}\|\widehat{P}(\cdot \mid s,a) - P^\star(\cdot \mid s,a)\|_{\mathsf{TV}} \lesssim \|\widehat{\zeta} - \zeta^\star\|_\infty. \tag{C.1}$$

Equipped with Eq. (C.1), we can now bound the following two types of errors: (i) $\langle f, (\mathcal{T}_{P^\star} - \mathcal{T}_{\widehat{P}})g \rangle$, and (ii) $\langle w_{\hat{P}} - w_{P^\star}, h \rangle$, where $f, g : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and $h : \mathcal{S} \to \mathbb{R}$, and $\mathcal{T}_P$ and w_P are the Bellman operator and visitation density of target policy π_t in the MDP with kernel P.

For (i):

$$\begin{aligned}\left|\langle f, (\mathcal{J}_{P^\star} - \mathcal{J}_{\widehat{P}})g\rangle\right| &= \left|\mathbb{E}[f(s,a)\Big(\gamma(\mathbb{E}_{P^\star} - \mathbb{E}_{\widehat{P}})[g(s', \pi_{\mathrm{t}}) \mid s, a]\Big)]\right| \\ &\leq \gamma \|f\|_\infty \mathbb{E}\Big|(\mathbb{E}_{P^\star} - \mathbb{E}_{\widehat{P}})[g(s', \pi_{\mathrm{t}}) \mid s, a]\Big| \\ &\lesssim \gamma \|f\|_\infty \|g(\cdot, \pi_{\mathrm{t}})\|_\infty \|\widehat{\zeta} - \zeta^\star\|_\infty.\end{aligned}$$

For (ii):

$$\begin{aligned}\langle w_{\widehat{P}} - w_{P^\star}, h\rangle &= \mathbb{E}[(w_{\widehat{P}}(s) - w_{P^\star}(s))h(s)] \\ &\leq \|h\|_\infty \|d_{\widehat{P}} - d_{P^\star}\|_{\mathsf{TV}} \\ &\leq \|h\|_\infty \frac{\gamma}{1-\gamma} \mathbb{E}_{d_{P^\star}} \|\widehat{P}(\cdot \mid s, a) - P^\star(\cdot \mid s, a)\|_{\mathsf{TV}} && \text{(Eq. (C.3))} \\ &\lesssim C\|h\|_\infty \frac{\gamma}{1-\gamma} \mathbb{E} \|\widehat{P}(\cdot \mid s, a) - P^\star(\cdot \mid s, a)\|_{\mathsf{TV}} && \text{(Assumption 4.6(i))} \\ &\lesssim C\|h\|_\infty \frac{\gamma}{1-\gamma} \|\widehat{\zeta} - \zeta^\star\|_\infty,\end{aligned}$$

where $C = \|^{\mathrm{d}d^{P^\star}}/_{\mathrm{d}\nu}\|_\infty < \infty$. For approximate realizability (ε_A): for any $w \in \mathcal{W}$, we have

$$\begin{aligned}&\|\mathcal{J}'_{\widehat{P}}(w_{\widehat{P}} - w)\|_2 \\ &\leq \|(\mathcal{J}_{\widehat{P}} - \mathcal{J}_{P^\star})'(w_{\widehat{P}} - w)\|_2 + \|\mathcal{J}'_{P^\star}(w_{\widehat{P}} - w_{P^\star})\|_2 + \|\mathcal{J}'_{P^\star}(w^\star - w)\|_2 \\ &= \langle w_{\widehat{P}} - w, (\mathcal{J}_{\widehat{P}} - \mathcal{J}_{P^\star})g_1\rangle + \langle w_{\widehat{P}} - w_{P^\star}, \mathcal{J}_{P^\star} g_2\rangle + \|\mathcal{J}'_{P^\star}(w^\star - w)\|_2 \\ &\lesssim \|\widehat{\zeta} - \zeta^\star\|_\infty + \|\mathcal{J}'_{P^\star}(w^\star - w)\|_2\end{aligned}$$

where g_1 and g_2 are given by $g_1 = ((\mathcal{J}_{P^\star} - \mathcal{J}_{\widehat{P}})'(w_{\widehat{P}} - w))/\|(\mathcal{J}_{P^\star} - \mathcal{J}_{\widehat{P}})'(w_{\widehat{P}} - w)\|_2$, $g_2 = (\mathcal{J}'_{P^\star}(w_{\widehat{P}} - w_{P^\star}))/\|\mathcal{J}'_{P^\star}(w_{\widehat{P}} - w_{P^\star})\|_2$. The last inequality uses (i) and (ii) with the fact that $\|g_1\|_\infty < \infty$ and $\|g_2\|_\infty < \infty$ as the w terms are bounded by our premise. Therefore, taking min over w and using Assumption C.9, we have $\varepsilon_A \lesssim \|\widehat{\zeta} - \zeta^\star\|_\infty + \varepsilon_{\mathsf{WReal}}$.

For approximate completeness (ε_B): for any $w \in \mathcal{W}$ and $f \in \mathcal{F}$, we have

$$\|f - \mathcal{J}'_{\widehat{P}}(w - w_{\widehat{P}})\|_2$$

$$\leq \|f - \mathcal{J}'_{P^\star}(w - w_{P^\star})\|_2 + \|(\mathcal{J}_{P^\star} - \mathcal{J}_{\widehat{P}})'(w - w_{P^\star})\|_2 + \|\mathcal{J}'_{P^\star}(w_{\widehat{P}} - w_{P^\star})\|_2$$
$$\lesssim \|f - \mathcal{J}'_{P^\star}(w - w_{P^\star})\|_2 + \|\widehat{\zeta} - \zeta^\star\|_\infty,$$

using a similar reasoning as for ε_A. Thus, $\varepsilon_B \lesssim \|\widehat{\zeta} - \zeta^\star\|_\infty + \varepsilon_{\mathsf{WComp}}$.

To sum up, we have shown that the misspecification is at most $O(\|\widehat{\zeta} - \zeta^\star\|_\infty + \varepsilon_{\mathsf{WReal}} + \varepsilon_{\mathsf{WComp}})$. Therefore, Uehara et al. [2021, Theorem 6.1 and Appendix C] ensures that w.p. $1 - \delta$, our learned $\widehat{w}$ satisfies,

$$\left\|\mathcal{J}'_{\widehat{P}}(\widehat{w} - w_{\widehat{P}})\right\|_2 \lesssim \varepsilon_n^{\mathcal{W}} + \|\widehat{\zeta} - \zeta^\star\|_\infty + \varepsilon_{\mathsf{WReal}} + \varepsilon_{\mathsf{WComp}} + \sqrt{\log(1/\delta)/n}.$$

Concluding the proof. The final step is to translate the above guarantee to $\|\mathcal{J}'_{P^\star}(\widehat{w} - w_{P^\star})\|_2$. The following shows that the switching cost is $O(\|\widehat{\zeta} - \zeta^\star\|_\infty)$:

$$\|\mathcal{J}'_{P^\star}(\widehat{w} - w_{P^\star})\|_2$$
$$\leq \|(\mathcal{J}_{P^\star} - \mathcal{J}_{\widehat{P}})'(\widehat{w} - w_{P^\star})\|_2 + \|\mathcal{J}'_{\widehat{P}}(\widehat{w} - w_{\widehat{P}})\|_2 + \|\mathcal{J}'_{\widehat{P}}(w_{\widehat{P}} - w_{P^\star})\|_2$$
$$\lesssim \varepsilon_n^{\mathcal{W}} + \|\widehat{\zeta} - \zeta^\star\|_\infty + \varepsilon_{\mathsf{WReal}} + \varepsilon_{\mathsf{WComp}} + \sqrt{\log(1/\delta)/n}.$$

This concludes the proof. □

Lemma C.11 (Visitation performance-difference)**.** *Let $P, U : \mathcal{S} \to \mathbb{R}_+$ be non-negative measures, which should be thought of as transitions in a discounted Markov chain. Assume U satisfies $\sum_{s'} U(s' \mid s) \leq 1$. Define $d_U = (1-\gamma)\sum_{h=1}^\infty \gamma^{h-1} d_U^h$, where $d_U^h = \int_{s_1, s_2, \ldots, s_{h-1}} d_1(s_1) U(s_2 \mid s_1) \ldots U(s \mid s_{h-1}) \mathrm{d}s_{1:h-1}$. Assume the same for P.*

Let $\mathcal{F} \subset \mathcal{S} \to \mathbb{R}$ be a function class that satisfies $f \in \mathcal{F} \implies g(s) = \mathbb{E}_{s' \sim P(s)}[f(s')] \in \mathcal{F}$, i.e.*, closed under projection with P. Then, define the integral (probability) metric $\|P - U\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |(\mathbb{E}_P - \mathbb{E}_U)[f(s)]|$. Then we have,*

$$\|d_P - d_U\|_{\mathcal{F}} \leq \frac{\gamma}{1-\gamma} \mathbb{E}_{d_U} \|P(\cdot \mid s) - U(\cdot \mid s)\|_{\mathcal{F}}. \tag{C.2}$$

Proof. Recall Bellman's flow, which is $d_P(s) = (1-\gamma)d_1(s) + \gamma \mathbb{E}_{\widetilde{s}\sim d_P} P(s \mid \widetilde{s})$. Fix any $f \in \mathcal{F}$. The initial state distributions cancel, so we have,

$$
\begin{aligned}
&\left|(\mathbb{E}_{d_P} - \mathbb{E}_{d_U})[f(s)]\right| \\
&= \left|\gamma \mathbb{E}_{\widetilde{s}\sim d_P}\mathbb{E}_{s\sim P(\cdot|\widetilde{s})}[f(s)] - \gamma \mathbb{E}_{\widetilde{s}\sim d_U}\mathbb{E}_{s\sim U(\cdot|\widetilde{s})}[f(s)]\right| \\
&\le \left|\gamma \mathbb{E}_{\widetilde{s}\sim d_P}\mathbb{E}_{s\sim P(\cdot|\widetilde{s})}[f(s)] - \gamma \mathbb{E}_{\widetilde{s}\sim d_U}\mathbb{E}_{s\sim P(\cdot|\widetilde{s})}[f(s)]\right| \\
&+ \left|\gamma \mathbb{E}_{\widetilde{s}\sim d_U}\mathbb{E}_{s\sim P(\cdot|\widetilde{s})}[f(s)] - \gamma \mathbb{E}_{\widetilde{s}\sim d_U}\mathbb{E}_{s\sim U(\cdot|\widetilde{s})}[f(s)]\right| \\
&\le \gamma\left|(\mathbb{E}_{\widetilde{s}\sim d_P} - \mathbb{E}_{\widetilde{s}\sim d_U})[\mathbb{E}_{s\sim P(\cdot|\widetilde{s})} f(s)]\right| + \gamma \mathbb{E}_{\widetilde{s}\sim d_U}\left|(\mathbb{E}_{s\sim P(\cdot|\widetilde{s})} - \mathbb{E}_{s\sim U(\cdot|\widetilde{s})})[f(s)]\right|.
\end{aligned}
$$

Thus, taking supremum over $\mathcal{F}$, we have

$$
\begin{aligned}
&\|d_P - d_U\|_{\mathcal{F}} \\
&\le \gamma \sup_{f\in\mathcal{F}}\left|(\mathbb{E}_{\widetilde{s}\sim d_P} - \mathbb{E}_{\widetilde{s}\sim d_U})[\mathbb{E}_{s\sim P(\widetilde{s})} f(s)]\right| + \gamma \mathbb{E}_{\widetilde{s}\sim d_U} \sup_{f\in\mathcal{F}}\left|(\mathbb{E}_{s\sim P(\cdot|\widetilde{s})} - \mathbb{E}_{s\sim U(\cdot|\widetilde{s})})[f(s)]\right| \\
&= \gamma\|d_P - d_U\|_{\mathcal{F}} + \gamma \mathbb{E}_{\widetilde{s}\sim d_U}\|P(\cdot \mid \widetilde{s}) - U(\cdot \mid \widetilde{s})\|_{\mathcal{F}}. \qquad (\mathcal{F} \text{ closed under } P\text{-projection})
\end{aligned}
$$

Rearranging terms finishes the proof. □

If $\mathcal{F}$ is the class of functions with $\|f\|_\infty \le 1$, then this recovers the TV distance, which gives,

$$
\|d_P - d_U\|_{\mathsf{TV}} \le \frac{\gamma}{1-\gamma}\mathbb{E}_{d_U}\|P(\cdot \mid s) - U(\cdot \mid s)\|_{\mathsf{TV}}. \tag{C.3}
$$

This generalizes Lemma E.3 of Agarwal et al. [2023] to infinite horizon.

C.8 Proofs and Details for the Orthogonal Estimator

C.8.1 Intuition for Theorem 4.11

We provide some intuition for the results in Theorem 4.11. Consider the V^- bound and let us decouple the indicator $\mathbb{I}[v(s') - \beta(s,a) \le 0]$ that appears implicitly in the $(v^-(s') - \beta^-(s,a))_-$ notation of Theorem 4.9. We augment the set

of nuisances with $\zeta(s,a,s') = v^-(s') - \beta^-(s,a)$ such that $(v^-(s') - \beta^-(s,a))_- = (v^-(s') - \beta^-(s,a))\mathbb{I}[\zeta(s,a,s') \leq 0]$. We state the following lemma (which we elaborate upon in Lemmas C.15 and C.16 in the Appendix):

Lemma C.12 (Double sharpness with correct $\zeta^\star$)**.** *Let $\mathbb{E}[\psi(s,a,s';q,w,\beta,\zeta^\star)]$ be the expectation of the (R)EIF with an arbitrary nuisance set $\eta = (w,q,\beta)$, but where the indicator $\mathbb{I}[v^-(s') \leq \beta^-(s,a)]$ has been replaced with the correct indicator $\mathbb{I}[\zeta^\star(s,a,s') \leq 0]$. Then:*

$$V_{d_1}^- = \mathbb{E}[\psi(s,a,s';q,w^\star,\beta^\star,\zeta^\star)] = \mathbb{E}[\psi(s,a,s';q^\star,w,\beta^\star,\zeta^\star)]$$

This lemma implies that if $\beta^- = (\beta^*)^-$ and $\zeta = \zeta^*$, then the estimator $\widehat{V}_{d_1}^-$ has a property known as "double-robustness" Kennedy [2023b] or "double-sharpness" Dorn et al. [2025a] in q and w, meaning the bias vanishes when either q or w is consistent. Moreover, the convergence rate would be $O_p(r_{n,2}^w r_{n,2}^q)$. This condition holds provided that β and ζ are correctly specified. However, estimation errors in β introduce an additional $O_p\left((r_{n,\infty}^\beta)^2\right)$ term, reflecting that β is first-order optimal for the CVaR component. Additionally, discrepancies between ζ and ζ^* contribute an extra $O_p\left((r_{n,\infty}^q)^2\right)$ to the error. While this discussion gives some insight into how we achieve the results in Theorem 4.11, we provide a a rigorous analysis in the next section.

Preliminaries

For this proof, our focus will be on $\widehat{V}_{d_1}^-$. The argument for $\widehat{V}_{d_1}^+$ is analogous, following a symmetric approach. To improve the clarity of our exposition, we will omit the $-$ and τ indices, assuming their presence is clear from the context.

For simplicity, we assume that n is a multiple of K such that $n = Kn_K$, where n_K is the size of a fold. We let $\mathbb{E}_n, \mathbb{E}_k$ denote the empirical averages over the

entire sample and the k^{th} fold, respectively. Recall that we use $\widehat{\eta} = (\widehat{w}, \widehat{q}, \widehat{\beta})$ and $\eta^* = (w^*, q^*, \beta^*)$ to denote the estimated and oracle nuisances, respectively.

We further suppress the dependency on s, a in Λ and τ and we write the ρ term in Theorem 4.9 as

$$\rho(s, a, s'; v, \beta) = (1 - \lambda)v(s') + \lambda\Big(\beta(s, a) + \tau^{-1}(v(s') - \beta(s, a))_-\Big). \tag{C.4}$$

We justify this by noting that the analysis holds regardless of whether λ and τ depend on s, a. Sometimes, it will be useful to decouple the indicator $\mathbb{I}[v(s') - \beta(s, a) \leq 0]$ implicit in the definition of ρ. In this case, we augment the set of nuisances with $\zeta(s, a, s') = v(s') - \beta(s, a)$ and write ρ as

$$\rho(s, a, s'; v, \beta, \zeta) = (1 - \lambda)v(s') + \lambda\Big(\beta(s, a) + \tau^{-1}(v(s') - \beta(s, a))\mathbb{I}[\zeta(s, a, s') \leq 0]\Big). \tag{C.5}$$

Similarly define $\psi(\cdot; w, q, \beta, \zeta)$ with the $\rho(\cdot; v, \beta, \zeta)$.

Auxiliary Lemmas

Definition C.13 (Margin Condition)**.** A function $f : \mathcal{X} \to \mathbb{R}$ of some random variable X is said to satisfy the margin condition with sharpness $\alpha \in [0, \infty]$ (or more succinctly, an α-margin) if there exist a fixed constant $c > 0$ such that

$$\forall t > 0 : P(0 < |f(X)| \leq t) \leq ct^{\alpha}.$$

If $f(X)$ is either zero or bounded away from zero almost surely, then f satisfies an infinite margin, *i.e.*, $\alpha = \infty$ [Kallus, 2022, Lemma 2]. If $f(X)$ is continuously distributed in a neighborhood around 0, *i.e.*, its CDF is boundedly differentiable on $(-\varepsilon, 0) \cup (0, \varepsilon)$ for some $\varepsilon > 0$, then f has a 1-margin [Kallus, 2022, Lemma 3].

Lemma C.14 (Margin Guarantees). *For any $f : \mathcal{X} \to \mathbb{R}$ satisfying α-margin (Definition C.13), $p \in [1, \infty]$, and any $g : \mathcal{X} \to \mathbb{R}$, the following statements hold for some constant $C > 0$:*

$$\mathbb{E}[(\mathbb{I}[g(X) \leq 0] - \mathbb{I}[f(X) \leq 0])f(X)] \leq C\|f - g\|_p^{\frac{p(1+\alpha)}{p+\alpha}}, \tag{C.6}$$

$$P[\mathbb{I}[g(X) \leq 0] \neq \mathbb{I}[f(X) \leq 0], f(X) \neq 0] \leq C\|f - g\|_p^{\frac{p\alpha}{p+\alpha}}, \tag{C.7}$$

where $\|\cdot\|_p$ is the L^p norm and we set $\infty t/\infty = t$ in the exponents.

The proof of Eq. (C.6) for any $p \in [1, \infty]$ and of Eq. (C.7) for $p = \infty$ is given in Audibert and Tsybakov [2007, Lemmas 5.1 and 5.2]. The proof of Eq. (C.7) for $p < \infty$ is given in Kallus [2022, Lemma 5].

Lemma C.15 (Sharpness with correct $q^\star$ and $\beta^\star$). $\frac{1}{n}\sum_{(s,a,s')\sim\mathcal{D}} \psi(s, a, s'; w, q, \beta)$ *is an unbiased estimator of $V_{d_1}^\star$ when $q = q^\star, \beta = \beta^\star$,* i.e.,

$$(1 - \gamma)\mathbb{E}_{d_1} v^\star(s_1) = \mathbb{E}[\psi(s, a, s'; w, q^\star, \beta\star)].$$

Proof. Since $q^\star$ and $\beta^\star$ are correct, the robust Bellman equation holds, and so for every s, a,

$$\mathbb{E}\Big[(1 - \lambda)v^\star(s') + \lambda(\beta^\star(s, a) + \tau^{-1}(v^\star(s') - \beta^\star(s, a))_-) \mid s, a\Big] = 0.$$

Thus, multiplying by any w does not change the fact that the debiasing term in ψ has expectation zero. Since we have $v^\star$, the first term in ψ is exactly the estimand, which concludes the proof. □

Lemma C.16 (Sharpness with correct w^* and ζ^*). $\frac{1}{n}\sum_{(s,a,s')\sim\mathcal{D}} \psi(s, a, s'; w, q, \beta, \zeta)$ *is an unbiased estimator of $V_{d_1}^\star$ when $w = w^\star, \zeta = \zeta^\star$,* i.e.,

$$(1 - \gamma)\mathbb{E}_{d_1} v^\star(s_1) = \mathbb{E}[\psi(s, a, s'; q, w^\star, \beta, \zeta^\star)]$$

Proof. Let $P^\star$ denote the robust transition kernel and let $d^\star$ denote the robust visitation measure under π, which satisfies: for all functions f,

$$\mathbb{E}_{d^\star}[f(s,a)] = (1-\gamma)\mathbb{E}_{d_1} f(s,\pi) + \gamma \mathbb{E}_{\widetilde{s,a}\sim d^\star, s\sim P^\star(s,a)}[f(s,\pi)].$$

Since $\zeta^\star$ is correct, for any v, s, a, we have

$$\begin{aligned}
&\mathbb{E}_{s'\sim P(s,a)}\Big[(1-\lambda)v(s') + \lambda\left(\beta(s,a) + \tau^{-1}(v(s') - \beta(s,a))\mathbb{I}[\zeta^\star(s,a,s') \le 0]\right)\Big] \\
&= \mathbb{E}_{s'\sim P(s,a)}\Big[(1-\lambda)v(s') + \lambda\tau^{-1}v(s')\mathbb{I}[\zeta^\star(s,a,s') \le 0]\Big] && (\bigstar) \\
&= \mathbb{E}_{s'\sim P^\star(s,a)}[v(s')], && \text{(Lemma 4.5)}
\end{aligned}$$

where in $\bigstar$ we used $\mathbb{E}_{s'\sim P(s,a)}\Big[\beta(s,a)\Big(1-\tau^{-1}\mathbb{I}[\zeta^\star(s,a,s') \le 0]\Big)\Big] = \beta(s,a)\Big(1-\tau^{-1}\tau\Big) = 0$. That is, for all function f, we have

$$\begin{aligned}
&(1-\gamma)\mathbb{E}_{d_1} v(s_1) + \mathbb{E}\big[w^\star(s,a)(r(s,a) + \gamma\rho(s,a,s';v,\beta,\zeta^\star) - q(s,a))\big] \\
&= (1-\gamma)\mathbb{E}_{d_1} v(s_1) + \mathbb{E}_{s,a\sim d^\star}[r(s,a) + \gamma\rho(s,a,s';v,\beta,\zeta^\star) - q(s,a)] \\
&= \mathbb{E}_{s,a\sim d^\star}[r(s,a)] + (1-\gamma)\mathbb{E}_{d_1} v(s_1) + \mathbb{E}_{s,a\sim d^\star}[\gamma\mathbb{E}_{s'\sim P^\star(s,a)}[v(s')] - q(s,a)] \\
&= \mathbb{E}_{s,a\sim d^\star}[r(s,a)] && \text{(robust Bellman flow)} \\
&= (1-\gamma)\mathbb{E}_{d_1} v^\star(s_1).
\end{aligned}$$

This concludes the proof. □

C.8.2 Proof of Rates

The estimation error is given by:

$$|\widehat{V}_{d_1} - V^*_{d_1}| = \left|\frac{1}{K}\sum_{k=1}^K \mathbb{E}_k[\psi(s,a,s';\widehat{\eta}^{[k]})] - V^*_{d_1}\right| \le \frac{1}{K}\sum_{k=1}^K \left|\mathbb{E}_k[\psi(s,a,s';\widehat{\eta}^{[k]})] - V^*_{d_1}\right|$$

We wish to bound $\left|\mathbb{E}_k[\psi(s,a,s';\widehat{\eta}^{[k]})] - V^*_{d_1}\right|$. We have that:

$$\begin{aligned}
\left|\mathbb{E}_k[\psi(s,a,s';\widehat{\eta}^{[k]})] - V^*_{d_1}\right| &\le \left|\mathbb{E}_k[\psi(s,a,s';\widehat{\eta}^{[k]})] - \mathbb{E}[\psi(s,a,s';\widehat{\eta}^{[k]})]\right| \\
&\quad + \left|\mathbb{E}[\psi(s,a,s';\widehat{\eta}^{[k]})] - V^*_{d_1}\right|
\end{aligned}$$

The first term is $O_p(n^{-1/2})$ by the CLT. We are now interested in bounding the second term:

$$\varepsilon(\widehat{\eta}) := \left|\mathbb{E}[\psi(s,a,s';\widehat{\eta})] - V^*_{d_1}\right|. \tag{C.8}$$

where we dropped the $[k]$ indicator without loss of generality. We further decompose $\varepsilon(\widehat{\eta})$ into two error terms, ε_A and ε_B, as follows:

$$\begin{aligned}
\varepsilon(\widehat{\eta}) &= \left|\mathbb{E}\left[\psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta})\right] - \mathbb{E}\left[\psi(s,a,s';\widehat{q},w^\star,\widehat{\beta},\zeta^\star)\right]\right| && \text{(Lemma C.16)}\\
&\le \left|\mathbb{E}\left[\psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta})\right] - \mathbb{E}\left[\psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta},\zeta^\star)\right]\right| && (\varepsilon^A)\\
&\quad + \left|\mathbb{E}\left[\psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta},\zeta^\star)\right] - \mathbb{E}\left[\psi(s,a,s';\widehat{q},w^\star,\widehat{\beta},\zeta^\star)\right]\right|. && (\varepsilon^B)
\end{aligned}$$

Bounding ε^A: Error from the incorrect indicator ζ

$$\begin{aligned}
\varepsilon_A &= \gamma\lambda\tau^{-1}\mathbb{E}\widehat{w}(s,a)\left(\widehat{v}(s') - \widehat{\beta}(s,a)\right)\left(\mathbb{I}\left[\widehat{v}(s') - \widehat{\beta}(s,a) \le 0\right] - \mathbb{I}[v^\star(s') - \beta^\star(s,a) \le 0]\right)\\
&\le C\gamma\lambda\tau^{-1}\mathbb{E}\left(\widehat{v}(s') - \widehat{\beta}(s,a)\right)\left(\mathbb{I}\left[\widehat{v}(s') - \widehat{\beta}(s,a) \le 0\right] - \mathbb{I}[v^\star(s') - \beta^\star(s,a) \le 0]\right)\\
&&& \text{(Assumption 4.6)}\\
&\lesssim \mathbb{E}\left(\widehat{v}(s') - \widehat{\beta}(s,a)\right)\left(\mathbb{I}\left[\widehat{v}(s') - \widehat{\beta}(s,a) \le 0\right] - \mathbb{I}[v^\star(s') - \beta^\star(s,a) \le 0]\right)
\end{aligned}$$

We break these terms down as follows:

$$\begin{aligned}
&\mathbb{E}\left(\widehat{v}(s') - \widehat{\beta}(s,a)\right)\left(\mathbb{I}\left[\widehat{v}(s') - \widehat{\beta}(s,a) \le 0\right] - \mathbb{I}[v^\star(s') - \beta^\star(s,a) \le 0]\right)\\
&= \mathbb{E}(v^\star(s') - \beta^\star(s,a))\left(\mathbb{I}\left[\widehat{v}(s') - \widehat{\beta}(s,a) \le 0\right] - \mathbb{I}[v^\star(s') - \beta^\star(s,a) \le 0]\right) && (\varepsilon^A_1)\\
&+ \mathbb{E}\left(\widehat{v}(s') - \widehat{\beta}(s,a) - v^\star(s') + \beta^\star(s,a)\right)\left(\mathbb{I}\left[\widehat{v}(s') - \widehat{\beta}(s,a) \le 0\right] - \mathbb{I}[v^\star(s') - \beta^\star(s,a) \le 0]\right).\\
&&& (\varepsilon^A_2)
\end{aligned}$$

We first bound ε^A_1. Assumption 4.6 implies

$$P(0 < |v^\star(s') - \beta^\star(s,a)| \le t) \le c''t,\ \forall t \in [0,c'), \quad P(|v^\star(s') - \beta^\star(s,a)| = 0) = 0,$$

where $c' < 1$ is the min of 1 and the given neighborhood of zero and $c'' \ge 1$ is the max of 1 and the bound on the density in that neighborhood. This implies a margin condition with $\alpha = 1$ and $c = c''/c'$.

We can instantiate the first part of Lemma C.14 with $f(X) = v^\star(s') - \beta^\star(s,a)$, $g(X) = \widehat{v}(s') - \widehat{\beta}(s,a)$ and obtain

$$\begin{aligned}\varepsilon_1^A &\lesssim \left\| v^\star(s') - \beta^\star(s,a) - \widehat{v}(s') + \widehat{\beta}(s,a) \right\|_p^{\frac{2p}{p+1}} \\ &\leq \left\| \widehat{v}(s') - v^\star(s') \right\|_p^{\frac{2p}{p+1}} + \left\| \widehat{\beta}(s,a) - \beta^\star(s,a) \right\|_p^{\frac{2p}{p+1}}.\end{aligned}$$

To bound ε_2^A, first write

$$\begin{aligned}&\left| \mathbb{E}\Big(\widehat{v}(s') - \widehat{\beta}(s,a) - v^\star(s') + \beta^\star(s,a) \Big) \Big(\mathbb{I}\Big[\widehat{v}(s') - \widehat{\beta}(s,a) \leq 0 \Big] - \mathbb{I}[v^\star(s') - \beta^\star(s,a) \leq 0] \Big) \right| \\ &\leq \left\| \widehat{v}(s') - \widehat{\beta}(s,a) - v^\star(s') + \beta^\star(s,a) \right\|_p \\ &\quad \cdot \mathbb{P}\Big(\mathbb{I}\Big[\widehat{v}(s') - \widehat{\beta}(s,a) \leq 0 \Big] \neq \mathbb{I}[v^\star(s') - \beta^\star(s,a) \leq 0] \Big)^{(p-1)/p}. \quad \text{(Holder's inequality)}\end{aligned}$$

We can bound $\mathbb{P}\Big(\mathbb{I}\Big[\widehat{v}(s') - \widehat{\beta}(s,a) \leq 0 \Big] \neq \mathbb{I}[v^\star(s') - \beta^\star(s,a) \leq 0] \Big)$ using the second part of Lemma C.14 such that

$$\begin{aligned}\varepsilon_2^A &\lesssim \left\| \widehat{v}(s') - \widehat{\beta}(s,a) - v^\star(s') + \beta^\star(s,a) \right\|_p \left\| \widehat{v}(s') - \widehat{\beta}(s,a) - v^\star(s') + \beta^\star(s,a) \right\|^{\frac{p-1}{p+1}} \\ &= \left\| \widehat{v}(s') - \widehat{\beta}(s,a) - v^\star(s') + \beta^\star(s,a) \right\|_p^{\frac{2p}{p+1}} \\ &\leq \left\| \widehat{v}(s') - v^\star(s') \right\|_p^{\frac{2p}{p+1}} + \left\| \widehat{\beta}(s,a) - \beta^\star(s,a) \right\|_p^{\frac{2p}{p+1}}.\end{aligned}$$

Putting the ε_1^A and ε_2^A together, we have

$$\begin{aligned}\varepsilon_A &\lesssim \left\| \widehat{v}(s') - v^\star(s') \right\|_p^{\frac{2p}{p+1}} + \left\| \widehat{\beta}(s,a) - \beta^\star(s,a) \right\|_p^{\frac{2p}{p+1}} && \text{(when } p \in [1,\infty)\text{)} \\ &\lesssim \left\| \widehat{v}(s') - v^\star(s') \right\|_\infty^2 + \left\| \widehat{\beta}(s,a) - \beta^\star(s,a) \right\|_\infty^2. && \text{(when } p = \infty\text{)}\end{aligned}$$

Bounding ε^B: Error with correct indicator but wrong nuisances Now we focus on bounding ε^B.

$$\begin{aligned}\varepsilon_B &= \mathbb{E}\Big[\psi(s,a,s'; \widehat{q}, \widehat{w}, \widehat{\beta}, \zeta^\star) \Big] - \mathbb{E}\Big[\psi(s,a,s'; \widehat{q}, w^\star, \widehat{\beta}, \zeta^\star) \Big] \\ &= \mathbb{E}\big(\widehat{w}(s,a) - w^\star(s,a) \big) \Big(r(s,a) + \gamma \rho(s,a,s'; \widehat{v}, \widehat{\beta}, \zeta^\star) - \widehat{q}(s,a) \Big) \\ &= \mathbb{E}\big(\widehat{w}(s,a) - w^\star(s,a) \big) \Big(r(s,a) + \gamma \rho(s,a,s'; \widehat{v}, \widehat{\beta}, \zeta^\star) - \widehat{q}(s,a) \Big)\end{aligned}$$

$$- \mathbb{E}(\widehat{w}(s,a) - w^\star(s,a))(r(s,a) + \gamma\rho(s,a,s';v^\star,\beta^\star) - q^\star(s,a)) \qquad \text{(Lemma C.15)}$$

$$= \mathbb{E}(\widehat{w}(s,a) - w^\star(s,a))\Big(\widehat{q}(s,a) - q^\star(s,a) + \gamma(\rho(s,a,s';\widehat{v},\widehat{\beta},\zeta^\star) - \rho(s,a,s';v^\star,\beta^\star))\Big).$$

In the Lemma C.15 step, we used

$$\begin{aligned} 0 &= (1-\gamma)\mathbb{E}_{d_1} v^\star(s_1) - \mathbb{E}[\psi(s,a,s';q^\star,\widehat{w},\beta^\star)] \\ &= (1-\gamma)\mathbb{E}_{d_1} v^\star(s_1) - \mathbb{E}[\psi(s,a,s';q^\star,w^\star,\beta^\star)]. \end{aligned}$$

Finally, note that

$$\begin{aligned} &\rho(s,a,s';\widehat{v},\widehat{\beta},\zeta^\star) - \rho(s,a,s';v^\star,\beta^\star) \\ &= (1-\lambda)(\widehat{v}(s') - v^\star(s')) + \lambda\tau^{-1}(\widehat{v}(s') - v^\star(s'))\mathbb{I}[\zeta^\star(s,a,s') \le 0] \\ &\quad + \lambda(\widehat{\beta}(s,a) - \beta^\star(s,a))\Big(1 - \tau^{-1}\mathbb{I}[\zeta^\star(s,a,s') \le 0]\Big). \end{aligned}$$

Due to the continuity of the CDF of $v^\star(s')$ at $\beta^\star(s,a)$ for all s,a, we have $\Pr(\zeta^\star(s',s,a) \le 0 \mid s,a) = \tau$ and so the last term vanishes. Thus, we're left with a quantity that is at most $\lesssim (\widehat{v}(s') - v^\star(s'))$. Therefore,

$$\begin{aligned} \varepsilon_B &\lesssim \mathbb{E}(\widehat{w}(s,a) - w^\star(s,a))(\mathcal{J}_{U^\pm}(\widehat{q}(s,a) - q^\star(s,a))) \\ &\le \|\mathcal{J}'_{U^\pm}(\widehat{w} - w^\star)\|_2 \|\widehat{q} - q^\star\|_2. \qquad \text{(Holder's inequality)} \end{aligned}$$

Putting everything together, we obtain the desired rates:

$$\begin{aligned} |\widehat{V}_{d_1} - V^*_{d_1}| &\lesssim O_p(n^{-1/2}) + \|\mathcal{J}'_{U^\pm}(\widehat{w} - w^\star)\|_2 \|\widehat{q} - q^\star\|_2 + \left\|\widehat{v} - v^\star\right\|_p^{\frac{2p}{p+1}} + \left\|\widehat{\beta} - \beta^\star\right\|_p^{\frac{2p}{p+1}} \\ &= O_p(n^{-1/2}) + O_p\Big(r_n^w r_n^q + (r_{n,p}^q)^{\frac{2p}{p+1}} + (r_{n,p}^\beta)^{\frac{2p}{p+1}}\Big) \qquad \text{(when } p \in [1,\infty)\text{)} \\ &\lesssim O_p(n^{-1/2}) + \|\mathcal{J}'_{U^\pm}(\widehat{w} - w^\star)\|_2 \|\widehat{q} - q^\star\|_2 + \left\|\widehat{v} - v^\star\right\|_\infty^2 + \left\|\widehat{\beta} - \beta^\star\right\|_\infty^2 \\ &= O_p(n^{-1/2}) + O_p\Big(r_n^w r_n^q + (r_{n,\infty}^q)^2 + (r_{n,\infty}^\beta)^2\Big). \qquad \text{(when } p = \infty\text{)} \end{aligned}$$

C.8.3 Proof of Normality & Efficiency

In this part of the theorem, we let:

$$\widetilde{V}_{d_1} = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_k[\psi(s,a,s';\eta^*)]$$

Then, we can write the following equality:

$$\sqrt{n}(\widehat{V}_{d_1} - V^*_{d_1}) = \sqrt{n}(\widehat{V}_{d_1} - \widetilde{V}_{d_1}) + \underbrace{\sqrt{n}(\widetilde{V}_{d_1} - V^*_{d_1})}_{\xrightarrow{d} \mathcal{N}(0,\Sigma)}$$

The second term converges in distribution to $\mathcal{N}(0, \Sigma)$ from the CLT and the fact that ψ is the efficient influence function. Thus, it remains to show that the first term is $o_p(1)$. We decompose the first term as follows:

$$\sqrt{n}(\widehat{V}_{d_1} - \widetilde{V}_{d_1}) = \sqrt{n}\frac{1}{K}\sum_{k=1}^{n}\Big(\mathbb{E}[\psi(s,a,s';\widehat{\eta}^{[k]})] - \mathbb{E}[\psi(s,a,s';\eta^*)]\Big) \tag{C.9}$$

$$+ \sqrt{n}\frac{1}{K}\sum_{k=1}^{n}\underbrace{(\mathbb{E}_k - \mathbb{E})[\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*)]}_{\varepsilon_k} \tag{C.10}$$

In Eq. (C.9), we have that $|\mathbb{E}[\psi(s,a,s';\widehat{\eta}^{[k]})] - \mathbb{E}[\psi(s,a,s';\eta^*)]|$ is bounded as in Eq. (Rates). Given the theorem's assumption about the nuisance rates, this term is $o_p(n^{-1/2})$ and Eq. (C.9) is $o_p(1)$. We now seek to control the ε_k term in Eq. (C.10). Letting $\mathcal{D}_k$ represent the samples in the k^{th} fold, we leverage sample splitting to show that the mean of $\varepsilon_k \mid \mathcal{D}_k$ is 0:

$$\begin{aligned}
&\mathbb{E}[\varepsilon_k \mid \mathcal{D}_k] \\
&= \mathbb{E}[\mathbb{E}_k[\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*)] - \mathbb{E}[\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*)] \mid \mathcal{D}_k] \\
&= 0
\end{aligned}$$

where we consider $\widehat{\eta}^{[k]}$ fixed with respect to the second expectation. The result follows from the fact that $\widehat{\eta}^{[k]}$ does not depend on $\mathcal{D}_k$. Then, we can invoke Chebyshev's inequality to obtain the following bound:

$$P\left(\frac{\varepsilon_k}{\text{Var}[\varepsilon_k \mid \mathcal{D}_k]^{1/2}} \geq \epsilon \middle| \mathcal{D}_k\right) \leq \frac{1}{\epsilon^2}, \ \forall \epsilon > 0$$

We have shown that $\varepsilon_k \mid \mathcal{D}_k = O_p(\text{Var}[\varepsilon_k \mid \mathcal{D}_k]^{1/2}) = O_p(n^{-1/2}\mathbb{E}[(\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*))^2 \mid \mathcal{D}_k]^{1/2})$. Here, we used the fact that $n_K = n/K$ (the size of $\mathcal{D}_k$)

and that K is a fixed integer that doesn't grow with n. Moreover, ε_k has 0 conditional mean. For the remainder of the analysis, we leave the conditioning on $\mathcal{D}_k$ implicit for simplicity.

To bound $\mathbb{E}[(\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*))^2 \mid \mathcal{D}_k]^{1/2} = \|\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*)\|_2$, we use similar notation and techniques as in Appendix C.8.2:

$$\begin{aligned}\|\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*)\|_2 \leq \|\psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta}) - \psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta},\zeta^*)\|_2 \qquad (\sigma_1)\\ + \|\psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta},\zeta^*) - \psi(s,a,s';q^*,w^*,\beta^*,\zeta^*)\|_2 \\ (\sigma_2)\end{aligned}$$

where we invoked Cauchy-Schwarz for the L_2 norm. We bound σ_2 as follows:

$$\begin{aligned}\sigma_2 &\leq \|\psi(s,a,s';\widehat{q},\widehat{w},\widehat{\beta}) - \psi(s,a,s';q^*,\widehat{w},\widehat{\beta},\zeta^*)\|_2 && (\sigma_{2a})\\ &\quad + \|\psi(s,a,s';q^*,\widehat{w},\widehat{\beta},\zeta^*) - \psi(s,a,s';q^*,\widehat{w},\beta^*,\zeta^*)\|_2 && (\sigma_{2b})\\ &\quad + \|\psi(s,a,s';q^*,\widehat{w},\beta^*,\zeta^*) - \psi(s,a,s';q^*,w^*,\beta^*,\zeta^*)\|_2 && (\sigma_{2c})\\ &\leq \|\widehat{v} - v^*\|_2 + \gamma(1-\lambda)\|\widehat{w}\|_2\|\widehat{v} - v^*\|_2 + \gamma\lambda\tau^{-1}\|\widehat{w}\|_2\|\widehat{v} - v^*\|_2 + \|\widehat{w}\|_2\|\widehat{q} - q^*\|_2 && (\sigma_{2a})\\ &\quad + \gamma\lambda\|\widehat{w}\|_2\|\widehat{\beta} - \beta^*\|_2 + \gamma\lambda\tau^{-1}\|\widehat{w}\|_2\|\widehat{\beta} - \beta^*\|_2 && (\sigma_{2b})\\ &\quad + \|\widehat{w} - w^*\|_2\left(\|r\|_2 + \gamma(1-\lambda)\|v^*\|_2 + \gamma\lambda\|\beta^*\|_2 + \gamma\lambda\tau^{-1}\|v^* - \beta^*\|_2\right) && (\sigma_{2c})\end{aligned}$$

Given our rate assumptions, our boundedness assumptions for $\widehat{w}$, the implicit boundedness of q^*, v^*, w^*, β^*, as well as the ordering of the L_2 and L_∞ norms, σ_2 is $o_p(1)$. We now bound the σ_1 term:

$$\sigma_2 = \gamma\lambda\tau^{-1}\left\|\widehat{w}(s,a)(\widehat{v}(s') - \widehat{\beta}(s,a))(\mathbb{I}[\widehat{v}(s') \leq \widehat{\beta}(s,a)] - \mathbb{I}[v^*(s') \leq \beta^*(s,a)])\right\|_2$$

There are two cases in which the difference of indicators is non-zero:

$$\begin{cases}\widehat{v}(s') \leq \widehat{\beta}(s,a) \text{ and } v^*(s') > \beta^*(s,a) \Rightarrow \mathbb{I}[\widehat{v}(s') \leq \widehat{\beta}(s,a)] - \mathbb{I}[v^*(s') \leq \beta^*(s,a)] = 1\\ \widehat{v}(s') > \widehat{\beta}(s,a) \text{ and } v^*(s') \leq \beta^*(s,a) \Rightarrow \mathbb{I}[\widehat{v}(s') \leq \widehat{\beta}(s,a)] - \mathbb{I}[v^*(s') \leq \beta^*(s,a)] = -1\end{cases}$$

In the first case, $\widehat{v}(s') - \widehat{\beta}(s,a) \leq 0, \beta^*(s,a) - v^*(s') < 0$ and thus

$$|(\widehat{v}(s') - \widehat{\beta}(s,a))(\mathbb{I}[\widehat{v}(s') \leq \widehat{\beta}(s,a)] - \mathbb{I}[v^*(s') \leq \beta^*(s,a)])| \leq |\widehat{v}(s') - \widehat{\beta}(s,a) + \beta^*(s,a) - v^*(s')|.$$

In the second case, $\widehat{v}(s') - \widehat{\beta}(s,a) > 0, \beta^*(s,a) - v^*(s') \leq 0$ and

$$|(\widehat{v}(s') - \widehat{\beta}(s,a))(\mathbb{I}[\widehat{v}(s') \leq \widehat{\beta}(s,a)] - \mathbb{I}[v^*(s') \leq \beta^*(s,a)])| \leq |\widehat{v}(s') - \widehat{\beta}(s,a) + \beta^*(s,a) - v^*(s')|.$$

Going back to σ_1, we have:

$$\begin{aligned}
\sigma_2 &\leq \gamma\lambda\tau^{-1}\|\widehat{w}\|_2 \|\widehat{v}(s') - \widehat{\beta}(s,a) + \beta^*(s,a) - v^*(s'))\|_2 \\
&\leq \gamma\lambda\tau^{-1}\|\widehat{w}\|_2 (\|\widehat{v} - v^*\|_2 + \|\widehat{\beta} - \beta^*\|_2)
\end{aligned}$$

By out theorem's assumptions, this term is also $o_p(1)$. Putting σ_1 and σ_2 together, we have that $\|\psi(s,a,s';\widehat{\eta}^{[k]}) - \psi(s,a,s';\eta^*)\|_2$ is $o_p(1)$ and $\varepsilon_k \mid \mathcal{D}_k$ is $o_p(n^{-1/2})$. By the bounded convergence theorem, this implies that ε_k is also $o_p(n^{-1/2})$. Then, the term in C.10 is $o_p(1)$, which further means that $\sqrt{n}(\widehat{V}_{d_1} - \widetilde{V}_{d_1}) = o_p(1)$. Our proof is now complete.

C.9 Derivation of the Efficient Influence Function

We use the ε-contamination approach of Hines et al. [2022] to derive an influence function (IF) for our estimand $V^-_{d_1}$. The proof for $V^+_{d_1}$ follows symmetrically. We note that since our tangent space is the whole space as it factorizes in the trivial way (as in [Kallus and Uehara, 2022, Page 54]), the IF we derive is actually the efficient influence function (EIF).

Let $P(s,a,s')$ denote the data distribution. Consider the ε-contamination $P_\varepsilon(s,a,s') = (1-\varepsilon)P(s,a,s') + \varepsilon\delta(\bar{s},\bar{a},\bar{s}')$, where $\delta(\bar{z})$ is the dirac delta at $\bar{z}$, *i.e.*, $\delta(\bar{z})$ has infinite mass at $\bar{z}$ and 0 mass elsewhere. Let V^-_ε denote the robust value function under the transition kernel $P_\varepsilon(s' \mid s,a)$. Omitting the ε subscript means

$\varepsilon = 0$. The IF of $V^-_{d_1}$ is then given by

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}(1-\gamma)\mathbb{E}_{d_1} V^-_\varepsilon(s_1)|_{\varepsilon=0}.$$

We dedicate the rest of this section towards this goal, which will be obtained in Theorem C.21.

Lemma C.17.

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} P_\varepsilon(s' \mid s,a)|_{\varepsilon=0} = \frac{\delta(\bar{s},\bar{a})}{P(s,a)}(\delta(\bar{s}') - P(s' \mid s,a)).$$

Proof. Use the fact $P_\varepsilon(s' \mid s,a) = \frac{P_\varepsilon(s,a,s')}{P_\varepsilon(s,a)} = \frac{(1-\varepsilon)P(s,a,s')+\varepsilon\delta(\bar{s},\bar{a},\bar{s}')}{(1-\varepsilon)P(s,a)+\varepsilon\delta(\bar{s},\bar{a})}$ and take derivative. □

Lemma C.18 (IF of conditional expectation)**.** *For any s, a and f_ε,*

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{E}_{P_\varepsilon}[f_\varepsilon(s') \mid s,a]|_{\varepsilon=0} = \frac{\delta(\bar{s},\bar{a})}{P(s,a)}(f(\bar{s}') - \mathbb{E}_P[f(s') \mid s,a]) + \mathbb{E}_P\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon} f_\varepsilon(s')|_{\varepsilon=0} \mid s,a\right],$$

where $f = f_0$.

Proof.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{E}_{P_\varepsilon}[f_\varepsilon(s') \mid s,a]|_{\varepsilon=0} &= \sum_{s'} f(s')\frac{\mathrm{d}}{\mathrm{d}\varepsilon}P_\varepsilon(s' \mid s,a)|_{\varepsilon=0} + \sum_{s'}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}f_\varepsilon(s')|_{\varepsilon=0}P(s' \mid s,a) \\ &= \frac{\delta(\bar{s},\bar{a})}{P(s,a)}(f_0(\bar{s}') - \mathbb{E}_P[f_0(s') \mid s,a]) + \mathbb{E}_P\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon} f_\varepsilon(s')|_{\varepsilon=0} \mid s,a\right],\end{aligned}$$

□

Lemma C.19 (IF of conditional CVaR)**.** *For any τ, s, a and f_ε,*

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\,\mathrm{CVaR}_{\tau,P_\varepsilon}[f_\varepsilon(s') \mid s,a]|_{\varepsilon=0} \\ &= \frac{\delta(\bar{s},\bar{a})}{P(s,a)}\Big(\beta_\tau(s,a) + \tau^{-1}(f(\bar{s}') - \beta_\tau(s,a))_- - \mathrm{CVaR}_\tau(f(s') \mid s,a)\Big) \\ &\quad + \mathbb{E}_P\left[\tau^{-1}\mathbb{I}[f(s') \le \beta_\tau(s,a)]\frac{\mathrm{d}}{\mathrm{d}\varepsilon} f_\varepsilon(s')|_{\varepsilon=0} \mid s,a\right],\end{aligned}$$

where $f = f_0$ and $\beta_\tau(s,a)$ be the $(1-\tau)$-th quantile of $f(s')$, $s' \sim P(s,a)$.

Proof.

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\,\mathrm{CVaR}_{P_\varepsilon}[f_\varepsilon(s') \mid s,a]|_{\varepsilon=0} \tag{C.11}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\min_b \mathbb{E}_{P_\varepsilon}\Big[b + \tau^{-1}(f_\varepsilon(s') - b)_- \mid s,a\Big]|_{\varepsilon=0} \tag{C.12}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{E}_{P_\varepsilon}\Big[\beta_\tau(s,a) + \tau^{-1}(f_\varepsilon(s') - \beta_\tau(s,a))_- \mid s,a\Big]|_{\varepsilon=0}, \tag{C.13}$$

where the last equality is due to Danskin's theorem and the fact that $\beta_\tau(s,a)$ is the maximizer of the CVaR dual form at $\varepsilon = 0$. Continuing, let $g_\varepsilon(s';s,a) := \beta_\tau(s,a) + \tau^{-1}(f_\varepsilon(s') - \beta_\tau(s,a))_-$, so

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{E}_{P_\varepsilon}[g_\varepsilon(s';s,a) \mid s,a] \\
&= \frac{\delta(\bar{s},\bar{a})}{P(s,a)}(g(\bar{s}';s,a) - \mathbb{E}_P[g(s',s,a) \mid s,a]) + \mathbb{E}_P\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}g_\varepsilon(s';s,a)|_{\varepsilon=0} \mid s,a\right] \qquad \text{(Lemma C.18)} \\
&= \frac{\delta(\bar{s},\bar{a})}{P(s,a)}(g(\bar{s}';s,a) - \mathrm{CVaR}_\tau(f(s') \mid s,a)) \\
&\quad + \mathbb{E}_P\left[\tau^{-1}\mathbb{I}[f(s') \le \beta_\tau(s,a)]\frac{\mathrm{d}}{\mathrm{d}\varepsilon}f_\varepsilon(s')|_{\varepsilon=0} \mid s,a\right].
\end{aligned}$$

This concludes the proof. □

We now prove the key "one-step forward" lemma.

Lemma C.20 (One-Step Forward)**.** *For any state distribution $\nu(s)$, we have*

$$\begin{aligned}
&\mathbb{E}_{s\sim\nu}\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_\varepsilon^-(s)|_{\varepsilon=0}\right] \\
&= \frac{\nu(\bar{s})\pi(\bar{a} \mid \bar{s})}{P(\bar{s},\bar{a})}(r(\bar{s},\bar{a}) + \gamma\Big((1-\lambda)V^-(\bar{s}') + \lambda\Big(\beta_\tau(\bar{s},\bar{a}) + \tau^{-1}(V^-(\bar{s}') - \beta_\tau(\bar{s},\bar{a}))_-\Big)\Big) \\
&\qquad\qquad - Q^-(\bar{s},\bar{a})) \\
&\quad + \gamma\mathbb{E}_{s\sim\nu}\left[\mathbb{E}_{\pi,P}\left[\Big((1-\lambda) + \lambda\tau^{-1}\mathbb{I}[V^-(s') \le \beta_\tau(s,a)]\Big)\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_\varepsilon^-(s')|_{\varepsilon=0} \mid s\right]\right].
\end{aligned}$$

Proof. For any s_1, we have

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_\varepsilon^-(s_1)$$

$$
\begin{aligned}
&= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{E}_{a_1\sim\pi(s_1)}\big[r(s_1,a_1) + \gamma((1-\lambda)\mathbb{E}_{P_\varepsilon}[V_\varepsilon^-(s_2) \mid s_1,a_1] + \lambda\,\mathrm{CVaR}_{\tau,P_\varepsilon}[V_\varepsilon^-(s_2)\mid s_1,a_1]\big]_{\varepsilon=0}\\
&= \gamma\mathbb{E}_{a_1\sim\pi(s_1)}\left[(1-\lambda)\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{E}_{\tau,P_\varepsilon}[V_\varepsilon^-(s_2)\mid s_1,a_1]|_{\varepsilon=0} + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\,\mathrm{CVaR}_{\tau,P_\varepsilon}[V_\varepsilon^-(s_2)\mid s_1,a_1]|_{\varepsilon=0}\right]\\
&= \gamma(1-\lambda)\mathbb{E}_{a_1\sim\pi(s_1)}\left[\frac{\delta(\bar{s},\bar{a})}{P(s_1,a_1)}(V^-(\bar{s}') - \mathbb{E}_P[V^-(s_2)\mid s_1,a_1])\right]\\
&+ \gamma(1-\lambda)\mathbb{E}_{a_1\sim\pi(s_1)}\mathbb{E}_P\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_\varepsilon^-(s_2)|_{\varepsilon=0}\mid s_1,a_1\right]\\
&+ \gamma\lambda\mathbb{E}_{a_1\sim\pi(s_1)}\left[\frac{\delta(\bar{s},\bar{a})}{P(s_1,a_1)}\Big(\beta_\tau(s_1,a_1) + \tau^{-1}(V^-(\bar{s}') - \beta_\tau(s_1,a_1))_- - \mathrm{CVaR}_\tau(V^-(s_2)\mid s_1,a_1)\Big)\right]\\
&+ \gamma\lambda\mathbb{E}_{a_1\sim\pi(s_1)}\mathbb{E}_P\left[\tau^{-1}\mathbb{I}[V^-(s_2)\le\beta_\tau(s_1,a_1)]\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_{\pi,P_\varepsilon}^-(s_2)\right].
\end{aligned}
$$

Taking expectation over $s_1 \sim \nu$, we have

$$
\begin{aligned}
\mathbb{E}_{s\sim\nu}\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_\varepsilon^-(s)|_{\varepsilon=0}\right] = \gamma\frac{\nu(\bar{s})\pi(\bar{a}\mid\bar{s})}{P(\bar{s},\bar{a})}\Big((1-\lambda)V^-(\bar{s}') + \lambda\Big(\beta_\tau(\bar{s},\bar{a}) + \tau^{-1}(V^-(\bar{s}') - \beta_\tau(\bar{s},\bar{a}))_-\Big)\\
- ((1-\lambda)\mathbb{E}[V^-(s')\mid\bar{s},\bar{a}] + \lambda\,\mathrm{CVaR}_\tau(V^-(s')\mid\bar{s},\bar{a}))\Big)\\
+ \gamma\mathbb{E}_{s\sim\nu}\left[\mathbb{E}_{\pi,P}\left[\Big((1-\lambda) + \lambda\tau^{-1}\mathbb{I}[V^-(s')\le\beta_\tau(s,a)]\Big)\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_\varepsilon^-(s')|_{\varepsilon=0}\mid s\right]\right].
\end{aligned}
$$

Finally recall that V^- satisfies the Bellman equation, so

$$
(1-\lambda)\mathbb{E}[V^-(s')\mid\bar{s},\bar{a}] + \lambda\,\mathrm{CVaR}_\tau(V^-(s')\mid\bar{s},\bar{a}) = Q^-(\bar{s},\bar{a}) - r(\bar{s},\bar{a}).
$$

This concludes the proof. □

Equipped with our main one-step lemma, we can now unroll it an infinite number of steps to derive the IF of our estimand.

Theorem C.21 (IF of Estimand)**.** *Let us denote*

$$
g(\bar{s},\bar{a},\bar{s}') := r(\bar{s},\bar{a}) + \gamma\Big((1-\lambda)V^-(\bar{s}') + \lambda\Big(\beta_\tau(\bar{s},\bar{a}) + \tau^{-1}(V^-(\bar{s}') - \beta_\tau(\bar{s},\bar{a}))_-\Big)\Big).
$$

Then, we have

$$
\mathbb{E}_{d_1}\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V_\varepsilon^-(s_1)|_{\varepsilon=0}\right] = \frac{d_{rob}^{\pi,\infty}(\bar{s},\bar{a})}{P(\bar{s},\bar{a})}g(\bar{s},\bar{a},\bar{s}').
$$

Proof. Let d_h denote the h-th step visitation in the robust MDP, with transition P_{rob} satisfying $\frac{P_{\text{rob}}(s'|s,a)}{P(s'|s,a)} = (1-\lambda) + \lambda\tau^{-1}\mathbb{I}[V^-(s') \le \beta_\tau(s,a)]$. Then notice that the final term of Lemma C.20 is exactly $\mathbb{E}_{s\sim\nu}\Big[\mathbb{E}_{\pi,P_{\text{rob}}}\Big[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V^-_\varepsilon(s')|_{\varepsilon=0} \mid s\Big]\Big]$. Therefore,

$$\begin{aligned}
&\mathbb{E}_{d_1}\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V^-_\varepsilon(s_1)|_{\varepsilon=0}\right] \\
&= \frac{d_1(\bar{s})\pi(\bar{a}\mid\bar{s})}{P(\bar{s},\bar{a})}g(\bar{s},\bar{a},\bar{s}') + \gamma\mathbb{E}_{s_2\sim d_2}\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V^-_\varepsilon(s_2)|_{\varepsilon=0}\right] \\
&= \frac{d_1(\bar{s})\pi(\bar{a}\mid\bar{s})}{P(\bar{s},\bar{a})}g(\bar{s},\bar{a},\bar{s}') + \gamma\frac{d_2(\bar{s})\pi(\bar{a}\mid\bar{s})}{P(\bar{s},\bar{a})}g(\bar{s},\bar{a},\bar{s}') + \gamma^2\mathbb{E}_{s_3\sim d_3}\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V^-_\varepsilon(s_3)|_{\varepsilon=0}\right].
\end{aligned}$$

Iterating the process, we have

$$\mathbb{E}_{d_1}\left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}V^-_\varepsilon(s_1)|_{\varepsilon=0}\right] = \sum_{h=1}^{\infty}\gamma^{h-1}\frac{d_h(\bar{s})\pi(\bar{a}\mid\bar{s})}{P(\bar{s},\bar{a})}g(\bar{s},\bar{a},\bar{s}') = \frac{d^{\pi,\infty}_{\text{rob}}(\bar{s},\bar{a})}{P(\bar{s},\bar{a})}g(\bar{s},\bar{a},\bar{s}'),$$

as desired. □

Finally, we can conclude that the IF in Theorem C.21 is in fact the efficient IF (EIF) because it is in the tangent space, as the tangent space is contains all functions [Kallus and Uehara, 2022].

C.10 Additional Validity Guarantees for Orthogonal Estimator

Our orthogonal estimator has additional desirable properties such as *validity* when some nuisances are misspecified. Specifically, the bounds returned by our orthogonal estimator will be asymptotically valid, though possibly loose, when some nuisances are inconsistent, *i.e.*, do not converge to the their true values. Below, we detail conditions under which we achieve validity. To be concise, we focus on the − case as the + case is symmetric.

Validity with correct $Q^\pm$ If $\widehat{Q} = Q^\pm$, we obtain valid bounds even if w, β are inconsistent.

Lemma C.22. *For any w, β, we have $\mathbb{E}[\psi(s,a,s';Q^-,\beta,w)] \le V^-_{d_1}$ with equality when $\beta = \beta^-_\tau$.*

Validity with $Q = \mathcal{T}_\beta^\pm Q$ Even if $\widehat{Q}$ is misspecified, we still have a valid bound if it solves a Bellman-type equation of the dual CVaR form. For a $\beta : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, define:

$$\mathcal{T}_\beta^\pm f(s,a) := r(s,a) + \gamma \Lambda^{-1}(s,a) \mathbb{E}[f(s', \pi_t) \mid s, a] \\ + \gamma(1 - \Lambda^{-1}(s,a)) \mathbb{E}\Big[\beta(s,a) + \tau^{-1}(s,a)(f(s', \pi_t) - \beta(s,a))_\pm \mid s, a\Big].$$

Lemma C.23. *Fix any w, β. If $Q_\beta^\pm = \mathcal{T}_\beta^\pm Q_\beta^\pm$, then $\mathbb{E}[\psi(s,a,s'; Q_\beta^-, \beta, w)] \le V_{d_1}^-$.*

Remark C.24. Lemmas C.22 and C.23 are dual to each other: in Lemma C.22, the plug-in is consistent while the debiasing correction errs in the valid direction (*i.e.*, ≥ 0 for $+$ and ≤ 0 for $-$). In Lemma C.23, the plug-in is valid while the debiasing correction has expectation zero.

C.10.1 Proofs for Validity

Lemma C.22. *For any w, β, we have $\mathbb{E}[\psi(s,a,s'; Q^-, \beta, w)] \le V_{d_1}^-$ with equality when $\beta = \beta_\tau^-$.*

Proof.

$$\mathbb{E}[\psi(s,a,s'; Q^-, \beta, w)] \le (1-\gamma)\mathbb{E}_{d_1}[V_\beta^-(s_1)] + \mathbb{E}[w(s,a)(Q^-(s,a) - \mathcal{T}_{\mathrm{CVaR}}^- Q^-(s,a))] \\ = V_{d_1}^- + 0 = V_{d_1}^-,$$

where the inequality comes from the fact that β is sub-optimal for $\mathbb{E}[\beta(s,a) + \tau^{-1}(V^-(s') - \beta(s,a))_-]$. The same proof applies for Q^+. □

We now prove Lemma C.23. First, we show that the $\mathcal{T}_\beta$ perspective gives rise to a dual definition of $Q^\pm$ (dual to Eq. (4.2)).

Lemma C.25.

$$Q^+(s,a) = \operatorname{argmin}_{\beta: Q_\beta = \mathcal{T}_\beta^+ Q_\beta} Q_\beta(s,a), \quad Q^-(s,a) = \operatorname{argmax}_{\beta: Q_\beta = \mathcal{T}_\beta^- Q_\beta} Q_\beta(s,a).$$

Proof. Unroll $Q^-(s,a) = r(s,a) + \gamma \inf_{U \in \mathcal{U}(P)} \mathbb{E}_U[r(s',a') + \gamma \inf_{U \in \mathcal{U}(P)} \mathbb{E}_U[\dots]]$, replacing each $\inf_{U \in \mathcal{U}(P)}$ with the convex combination of $\mathbb{E}$ and CVaR from Lemma 4.1. Then, write each CVaR using the dual form, *i.e.*, $\max_\beta\{\beta(s,a) + \tau^{-1}(s,a)\mathbb{E}[(\cdots - \beta(s,a))_+]\}$. By s,a-rectangularity, the scalar $\max_\beta$ separates per s,a, so we can pull all the maxes out front as a max over $\beta(s,a)$ functions. Note that not all $\beta(s,a)$ functions have a well-defined infinite sum in this manner, as $\mathcal{T}_\beta$ is not always a contraction. The condition $Q_\beta = \mathcal{T}_\beta^- Q_\beta$ exactly characterizes when this unrolling is well-defined. Thus, Q^- is exactly the minimum Q_β whenever this procedure of unrolling with β is well-defined. This concludes the proof. □

Lemma C.23. *Fix any w, β. If $Q_\beta^\pm = \mathcal{T}_\beta^\pm Q_\beta^\pm$, then $\mathbb{E}[\psi(s,a,s'; Q_\beta^-, \beta, w)] \le V_{d_1}^-$.*

Proof.

$$\mathbb{E}[\psi(s,a,s'; Q_\beta^+, \beta, w)] \ge (1-\gamma)\mathbb{E}_{d_1}[V_\beta^+(s_1)] + 0 \ge V_{d_1}^+,$$
$$\mathbb{E}[\psi(s,a,s'; Q_\beta^-, \beta, w)] = (1-\gamma)\mathbb{E}_{d_1}[V_\beta^-(s_1)] + 0 \le V_{d_1}^-.$$

The first equality is because the correction term is $\mathcal{T}_\beta^- Q_\beta^- - Q_\beta^-$, which is zero since Q_β^- is a fixed point. The inequality is due to Lemma C.25. □

C.11 Additional Details for Main Experiment

Environment We consider a simple MDP with a one-dimensional state space $\mathcal{S} = [0,5]$, a binary action space $\mathcal{A} = \{0,1\}$, reward function

$$r(s,a) = \frac{26 - s^2 - \mathbb{I}[a=1]}{26},$$

which we note takes values in the range $[0, 1]$, and with transitions given by

$$P(\cdot \mid s, a = 0) = \text{UnifClip}[s - 0.2,\ s + 1]$$
$$P(\cdot \mid s, a = 1) = \text{UnifClip}[0.2s - 0.02,\ s + 0.5]\,,$$

where $\text{UnifClip}[a, b]$ denotes a uniform distribution between $\max(a, 0)$ and $\min(b, 5)$. In addition, the environment always starts in initial state $s_0 = 2$. Essentially, this is a simple control environment, where high rewards are obtained by maintaining state as close to zero as possible, the action $a = 1$ is a control action that (in expectation) moves the state closer to zero, and which occurs a small reward cost, and the action $a = 0$ is a passive action that allows the state to freely drift (with an overall drift away from zero).

Target Policy We focus on estimating the worst-case policy value $V_{d_1}^-$ for the simple threshold-based target policy π_t which takes action $a = 1$ when $s \geq 2$, and $a = 0$ whenever $s < 2$.

Logging Policy and Data Sampling Procedure We sample data using an evaluation policy π_b which is an ϵ-smoothed threshold policy similar to π_t. Specifically, π_b takes action $a = 1$ when $s \geq 1.5$ with probability 0.95, and takes action $a = 0$ when $s < 1.5$ with probability 0.95. We obtain a dataset $\{s_i, a_i, s_i', r_i\}$ by first rolling out with π_b for 1000 burn-in time steps, and then sampling the tuple (s, a, s', r) every 10 time steps. For each replication of our experiment, we sample 10,000 tuples in total.

Calculation of True Worst-Case Policy Values A major challenge in studying robust policy value estimation is that, even with ground truth knowledge of the MDP and/or access to a simulator, it may be intractable to estimate the robust policy values $V_{d_1}^{\pm}$. Fortunately, the above environment has the desirable property that we can analytically compute the best/worst-case transition dis-

tributions allowed by our sensitivity model, since no matter what policy π_t the agent is acting with, it always strictly prefers transitions to smaller states. In detail, suppose that for some state, action pair (s, a) we have $P(\cdot \mid s, a) = \text{Unif}[x, y]$, for some $0 \leq x \leq y \leq 5]$. Then, letting $\alpha = 1/(1 + \Lambda(s, a))$, it is easy to verify that the worst case transition kernel is given by

$$U^-(\cdot \mid s, a) = (1 - \Lambda^{-1}(s, a))\text{Unif}[y - \alpha(y - x), y] + \Lambda^{-1}(s, a)\text{Unif}[x, y]\,.$$

That is, the worst case transition kernel is given by a mixture of two uniform distributions. Therefore, we can easily simulate rollouts with the best/worst case transition kernels, and accurately estimate the robust policy values. This allows us to validate our methodology in this synthetic environment. Specifically, for each $\Lambda(s, a)$ we experiment with, we can compute the corresponding ground truth $V^-_{d_1}$ up to arbitrary precision via Monte Carlo sampling, by rolling out trajectories with π_t in the adversarial MDP according to the above worst-case transition kernel.

Note as well that if one wanted to estimate the best-case policy value, analogous reasoning would give us

$$U^+(\cdot \mid s, a) = (1 - \Lambda^{-1}(s, a))\text{Unif}[x, x + \alpha(y - x)] + \Lambda^{-1}(s, a)\text{Unif}[x, y]\,.$$

However, in our experiments we only concern ourselves with worst-case policy value estimation.

Nuisance Estimation We instantiate variations of Algorithms 4.1 and 4.2 using neural nets for the classes $\mathcal{Q}$, $\mathcal{B}$, and $\mathcal{W}$ used for fitting Q^-, β^-, and w^- respectively, and linear sieves for the corresponding critic class $\mathcal{Q}$ that we perform maximization over for the minimax estimation of w^-. Specifically, we grow the linear sieve for the critic class in a data-driven way, as follows: at each step k

of the respective algorithm, we compute the best response $q_k \in \mathcal{Q}$ to the previous iterate solution $w_k \in \mathcal{W}$ by optimizing over a neural net class, and then we append this best-response function to the set of functions in our linear sieve for the corresponding critic class. Full exact nuisance estimation details necessary for reproducibility will be available in our code release.

Estimators We estimate the worst-case policy value using three different estimators:

- **Q**: Direct estimator given by:

$$\widehat{V}_{d_1}^{-} = \widehat{Q}^{-}(s_1, \pi_t(s_1))\,,$$

  where s_1 is the deterministic initial state.

- **W**: Importance sampling-style estimator using $\hat{w}^{-}$, which is given by:

$$\widehat{V}_{d_1}^{-} = \frac{1}{n}\sum_{i=1}^{n} \widehat{w^{-}}(s_i, a_i)\widehat{\xi}_i r_i\,,$$

  where

$$\widehat{\xi_i} = \Lambda^{-1} + (1-\Lambda^{-1})(1+\Lambda)\mathbb{I}\Big[\widehat{V}^{-}(s_i') \leq \widehat{\beta}^{-}(s_i, a_i)\Big]\,.$$

- **Orth**: Our orthogonal estimator using EIF, given by

$$\widehat{V}_{d_1}^{-} = \frac{1}{n}\sum_{i=1}^{n} \psi(s_i, a_i, s_i'; \widehat{Q}^{-}, \widehat{\beta}^{-}, \widehat{w^{-}})\,.$$

Note as well that we used a simpler data splitting procedure rather than the cross-fitting procedure described in Algorithm 4.3. Specifically, we used the first 10,000 tuples for estimating nuisances, and the second 10,000 tuples for the final estimators. This was done for the sake of computational ease in running experiments with many replications, and was performed in the same way for all methods.

In addition, for extra robustness, in each experiment replication we ran the nuisance estimation pipeline 5 times (on the same fixed sampled dataset), and

took the 80th percentile policy value estimates, since the estimators tend to under-estimate the true policy value by design, with greater under-estimation when the nuisance estimates are less well optimized.

C.12 Empirical Investigation on Medical Application

Here, we describe an additional empirical investigation of our methodology on medical data. Specifically, we consider the problem of sepsis management using RL. For all parts of the investigation described below, fully complete details can be obtained from our code release.

Motivation of Investigation Training RL models in simulated environments derived from real-world data is an exciting avenue for leveraging AI towards critical medical use cases. However, doing this obviously has the downside that, unless one undergoes the very risky process of training an RL agent online via real medical interventions, one has to resort to training within simulators, and then has to account for the inevitable "sim-to-real" gap. Therefore, our robust OPE methodology provides an interesting approach for estimating worst-case performance of RL models under potential changes in dynamics when moving to real application.

RL Environment Our RL environment is based on the OpenAI Gym sepsis simulator environment of Kiani et al. [2019]. This RL environment allows for simulation of dynamic sepsis management, which was created by training a blackbox ML model to mimic observed transition dynamics from the real-world electronic health record-based MIMIC-III dataset [Johnson et al., 2016]. This existing sepsis simulator is an episodic environment that continues until the agent either recovers or dies. It has a 46-dimensional state space containing various vital measurements, a discrete action space containing 24 possible actions (where

an action is essentially the Cartesian product of some independent base actions). The reward function in this original simulator gives zero reward whenever an episode has not terminated, a +15 reward at termination when if the patient survives, or a -15 reward at termination if the patient dies. Please see Kiani et al. [2019] and the code release linked therein for additional details.

We built an RL environment for our investigation by creating a simple wrapper around this existing sepsis simulator, in order to make it fit our setup. In particular, we made the following key changes:

1. We made the environment infinite-horizon, by automatically looping to a new random starting state for a fresh patient whenever the episode in the base simulator terminates
2. We normalized the reward function so that it lies in range $[0, 1]$, where:
   (a) $r(s, a) = 0$ if patient dies
   (b) $r(s, a) = 1$ if patient recovers and is discharged
   (c) $r(s, a) = 0.5$ if treatment has not terminated for current patient

In addition, for this environment, we perform all experiments with $\gamma = 0.95$.

Policies for Investigation We constructed RL policies for our empirical investigation by training some deep RL models using the sepsis simulator environment.

In the case of the behavioral policy π_b used to generate the observational offline data, we trained this policy by running Proximal Policy Optimization (PPO: Schulman et al. [2017]) over a relatively large (16,000) number timesteps, in order to emulate a reasonably good "current best practices" model for creating observational data.

In the case of the target policy π_t to be evaluated, we trained this policy using

Deep Q Learning (DQL: Mnih [2013]), over a relatively small (1,600) number of timesteps, in order to emulate a potentially risky new candidate model.

Creating an Offline Dataset Using our behavioral policy π_b which we created as above, we generated a fixed offline dataset consisting of 20,000 observed tuples of state, action, reward, and next state. Unlike with our main empirical investigation in the main chapter, we did not perform any "thinning" on these sampled tuples to make them more independent, so that the observed transitions are sequentially correlated as with real-world medical data.

Nuisance Estimation We perform nuisance estimation almost identically as in our main empirical investigation, with the only change being a slight change to our neural network architectures to better handle the large discrete action space. Specifically, instead of training neural networks that take state as input and produce $|\mathcal{A}|$ outputs (one per action), we train neural networks that take both state and action as inputs, using a learnt low-dimensional encoding of the actions, and produce a single output. Please see our code release for details.

Estimators We consider the same three estimators (**Q**, **W**, and **Orth**) as in our main empirical investigation. As in that investigation, we use these to estimate the worst-case policy value for the given $\Lambda(s,a)$. In addition, as in the main experiments, we consider these estimators for various fixed $\Lambda(s,a)$ that do not depend on s or a. In this case, we consider $\Lambda \in \{1,2,4\}$, as these reflect a reasonable range of possible confounding strength for real application.

Results Below, in Table C.2 we show the estimated policy value for all three estimators for each fixed $\Lambda \in \{1,2,4\}$. Here, we present the median policy value estimate over 5 runs of our estimators from random starting seeds after remov-

Λ	Median Policy Value Estimate		
	Q	**W**	**Orth**
1	.546 ± .003	.386 ± .087	.532 ± .008
2	.454 ± .040	.534 ± .141	.515 ± .036
4	.381 ± .077	.287 ± .106	.338 ± .086

Table C.2: Median policy-value estimates for the sepsis-management experiment in Chapter 4, for each estimator and each value of Λ, over 5 random-seed runs. The ± values denote half the difference between 80th and 20th percentiles.

ing outliers.[1] In addition, we present a ± spread given by half the difference between the 80th and 20th percentiles.

Although for this investigation we cannot analytically compute the ground truth "true" adversarial policy values to evaluate against when $\Lambda > 1$, we can still analyze the trends of these estimators and compare them to those observed in our main synthetic experiment, and we can also compare their accuracy when $\Lambda = 1$.

First, in the case of $\Lambda = 1$, we computed the true policy value of π_t to be within the range 0.532 ± 0.002 with 95% confidence. This is almost exactly equal to the median **Orth** estimator, but far outside the spread of outputs of the **Q** estimator. That is, although the **Q** estimator has somewhat lower variance in outputs over multiple runs for $\Lambda = 1$ compared with **Orth**, it appears to be far more biased.

Next, looking more broadly across all values of Λ, as in our main experiment, the **Q** and **Orth** estimators generally result in similar estimates to each other, and the **W** estimators are very variable. This may reflect the relative difficulty of estimating the w^- nuisance function compared with Q^- and β^-; although both **Orth** and **W** are affected by this difficulty, the **Orth** estimator has a theoretical

[1] Specifically, we exclude policy value estimates that lie outside the possible range of $[0, 1]$, which occasionally occur due to bad optimization from the starting seed.

robustness to the errors of these nuisance functions that the **W** estimator does not, as outlined in our theory.

We also observe that when $\Lambda = 1$ the **Q** estimator is significantly more stable than **Orth**, but when $\Lambda > 1$ the stability of **Orth** is either comparable to or superior to **Q**. In order to understand this, we first note that unlike in our main experiments, here the repetitions are re-runs of the estimators with the same offline sepsis dataset, so these ± spreads reflect potential computational errors rather than statistical errors. Given this, this pattern of errors could be explained by the fact that when $\Lambda = 1$ the **Q** estimation is extremely simple, reducing to standard FQI, whereas when $\Lambda > 1$ it requires a more complex robust FQI estimation with simultaneous estimation of β^{-}. That is, the difference in computational difficulty of estimating **Orth** versus **Q** may be smaller for $\Lambda > 1$.

Overall, although it is hard to definitively compare the accuracy of these estimators for $\Lambda > 1$ given a fundamental lack of ground truth, given both a similar pattern of results as in our synthetic experiments, as well as the far greater accuracy of **Orth** when $\Lambda = 1$, it seems reasonably to believe based on these results that our proposed **Orth** estimator may be more reliable than the existing robust FQI approach of the **Q** estimator.

Finally, we consider the implication of our results for the problem of learning sepsis management policies from simulators. Our **Orth** estimator suggests that there is relatively little sensitivity of this environment to deviations allowed by $\Lambda = 2$, but very significant deviation allowed by $\Lambda = 4$. Indeed, given the reward structure described above, the worst-case results under $\Lambda = 4$ imply an extremely high mortality rate. Whether worst-case deviations of this magnitude are reasonable or not is unclear, and this is something that requires further investigation for future work on RL for sepsis management.

APPENDIX D

APPENDIX FOR CHAPTER 5

D.1 Additional Related Works

Heterogeneous treatment effect estimation from observational data Recently, there has been a significant interest in applying machine learning to estimate CATEs using observational data. This field has seen adaptations of a wide range of machine learning techniques, from random forests [Wager and Athey, 2018b, Oprescu et al., 2019] and Bayesian algorithms *e.g.*[Hill, 2011, Hahn et al., 2020] to deep learning [Johansson et al., 2016, Atan et al., 2018, Shi et al., 2019] and blackbox meta-learners [Künzel et al., 2019, Nie and Wager, 2021] that utilize efficient influence functions [Kennedy, 2023d, Curth et al., 2020] and Neyman orthogonality [Chernozhukov et al., 2018a, Foster and Syrgkanis, 2023b]. Despite these advancements, a significant challenge remains as these methods typically assume the absence of confounding in observational data (ignorability) – an often unrealistic and unverifiable assumption – limiting their real-world applicability. Without ignorability, point identification of effects is impossible, although some studies propose methods to construct *bounds* on treatment effect estimates under assumptions about the structure of unobserved confounding [Rosenbaum et al., 2010, Kallus and Zhou, 2018, Oprescu et al., 2023, Frauen et al., 2024]. Nonetheless, these bounds often have limited practical utility. Other efforts to address confounding bias in CATE estimation rely on latent variable models to recover unobserved confounders from noisy proxies [Louizos et al., 2017, Kuzmanovic et al., 2021] or utilize multiple or sequential treatments [Wang and Blei, 2019, Bica et al., 2020a, Hatt and Feuerriegel, 2024]. However, these methods also have limited practical impact, as

they depend on either the availability of additional accurate proxy data or unverifiable assumptions such as no unobserved single-cause confounders.

Heterogeneous treatment effect estimation using IVs Machine learning techniques have recently been integrated with instrumental variable methods, offering significant advantages over traditional approaches, including the flexible estimation of CATEs. Singh et al. [2019] and Xu et al. [2021] expand on two-stage least squares (2SLS) to incorporate complex feature mappings via kernel methods and deep learning. In the same vein, Hartford et al. [2017] introduced a two-stage neural network for conditional density estimation, while Bennett et al. [2019] applied moment conditions for IV estimation. Syrgkanis et al. [2019] propose novel IV estimators that exhibit Neyman orthogonality. However, these techniques rely on the assumption that instruments are relevant across all covariate groups, a condition that is not consistently met with weak instruments.

Treatment effect estimation with weak instruments Weak instruments compromise the reliability of traditional IV methods like 2SLS, often producing biased, high-variance estimates and undermining causal claims. To mitigate these issues, several approaches have been developed, including bias-adjusted 2SLS estimators, limited information maximum likelihood (LIML), and jackknife instrumental variable (JIVE) estimators (see Huang et al. [2021] and references therein). Recent methods reduce 2SLS estimator variance by exploiting first-stage heterogeneity (variation in compliance) through a weighting scheme, as detailed in Kennedy et al. [2020], Coussens and Spiess [2021], Abadie et al. [2024]. However, these methods do not extend to estimating conditional average treatment effects. Another strand of research focuses on combining multiple weak instruments into a robust composite, showing promise in genetic studies using Mendelian randomization ([Kang et al., 2016, Lin et al., 2024]). These ap-

proaches require access to multiple weak instruments for the same treatment. Our work aligns most closely with Kennedy et al. [2020], Coussens and Spiess [2021], Abadie et al. [2024] in that we leverage compliance heterogeneity and employ compliance weighting to merge IV studies with observational data for efficient confounding bias estimation. Unlike these studies, however, our approach distinctively estimates heterogeneous effects and leverages additional observational data to address challenges posed by weak instruments.

Combining observational and randomized data There has been a proliferation of research in combining observational datasets with randomized control trials – experimental data with *perfect* compliance – to mitigate bias from observational studies. One of the strategies is to impose structural assumptions such as strong parametric assumptions for the confounding bias [Kallus et al., 2018] or a shared structure between the biased and unbiased CATE functions that can be estimated from the two datasets [Hatt et al., 2022]. Other studies advocate for dual estimators from both data types, optimizing bias correction through a weighted average [Yang and Ding, 2019, Cheng and Cai, 2021, Rosenman et al., 2023]. Additionally, approaches like those by Athey et al. [2020] and Imbens et al. [2025] leverage outcomes from different time-steps, such as short-term and long-term effects, to enhance estimation accuracy. Our work is closest to Kallus et al. [2018] and Hatt et al. [2022]. However, our study faces additional complexities: firstly, the CATE estimation techniques differ between the datasets, requiring us to debias the overall effect function rather than just individual outcome functions. Secondly, RCTs may not represent the target population due to their narrow scope, our instrumental variable (IV) study faces representation issues due to minimal or absent compliance in strata that are not known a priori. Thirdly, the CATE estimation in our IV study uses a ratio estimator, which

is highly sensitive to changes in the compliance denominator, adding a layer of complexity to our analysis.

D.2 Proofs of Theorems and Lemmas

D.2.1 Proof of Equation 5.3

The exclusion and independence conditions in Assumption 5.1 imply that the following identification equation holds:

$$\begin{aligned}&\mathbb{E}\left[Y^E\left(A^E(1)\right) - Y^E\left(A^E(0)\right) \mid X^E = x\right] \\ &\quad = \mathbb{E}[Y^E \mid Z^E = 1, X^E = x] - \mathbb{E}[Y^E \mid Z^E = 0, X^E = x].\end{aligned} \tag{D.1}$$

By noting that

$$\begin{cases} Y^E\left(A^E(1)\right) - Y^E\left(A^E(0)\right) = Y(1) - Y(0), & \text{if } A^E(1) = 1, A^E(0) = 0 \text{ (compliers)} \\ Y^E\left(A^E(1)\right) - Y^E\left(A^E(0)\right) = Y(0) - Y(1), & \text{if } A^E(1) = 0, A^E(0) = 1 \text{ (defiers)} \\ Y^E\left(A^E(1)\right) - Y^E\left(A^E(0)\right) = 0, & \text{if } A^E(1) = 1, A^E(0) = 1 \text{ (always-takers)} \\ Y^E\left(A^E(1)\right) - Y^E\left(A^E(0)\right) = 0, & \text{if } A^E(1) = 0, A^E(0) = 0 \text{ (never-takers)} \end{cases}$$

the left-hand side of this equation can further be written as:

$$\begin{aligned}&\mathbb{E}[Y^E(A^E(1)) - Y^E(A^E(0)) \mid X^E = x] && \text{(D.2)} \\ &= \mathbb{E}[(Y^E(1) - Y^E(0))(A^E(1) - A^E(0)) \mid X^E = x] \\ &= \mathbb{E}[Y^E(1) - Y^E(0) \mid X^E = x] \cdot \mathbb{E}[A^E(1) - A^E(0) \mid X^E = x] && \text{(Assumption 5.2)} \\ &= \tau(x) \cdot (\mathbb{E}[A^E \mid Z^E = 1, X^E = x] - \mathbb{E}[A^E \mid Z^E = 0, X^E = x]) && \text{(Assumption 5.1)}\end{aligned}$$

Since the claim of Equation 5.3 holds for $x \in \mathcal{X}'$, we have that $\mathbb{E}[A^E \mid Z^E = 1, X^E = x] - \mathbb{E}[A^E \mid Z^E = 0, X^E = x] \neq 0$ by the relevance condition in Assumption 5.1. From Eqs. D.1 and D.2, we obtain:

$$\tau(x) = \frac{\mathbb{E}[Y^E \mid Z^E = 1, X^E = x] - \mathbb{E}[Y^E \mid Z^E = 0, X^E = x]}{\mathbb{E}[A^E \mid Z^E = 1, X^E = x] - \mathbb{E}[A^E \mid Z^E = 0, X^E = x]}, \quad \forall x \in \mathcal{X}'.$$

D.2.2 Proof of Lemma 5.3

Recall that for any $x \in \mathcal{X}'$, we have that $\gamma(x) \neq 0$ by Assumption 5.2. Then, assuming that $\pi_Z(x) > 0$, we use the law of total expectation as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{Y^E Z^E}{\pi_Z(x)\gamma(x)} - \frac{Y^E(1-Z^E)}{(1-\pi_Z(x))\gamma(x)}\middle| X^E = x\right] \\
&= \mathbb{E}\left[\frac{Y^E Z^E}{\pi_Z(x)\gamma(x)} - \frac{Y^E(1-Z^E)}{(1-\pi_Z(x))\gamma(x)}\middle| Z^E = 1, X^E = x\right] P(Z^E = 1 \mid X^E = x) \\
&\quad + \mathbb{E}\left[\frac{Y^E Z^E}{\pi_Z(x)\gamma(x)} - \frac{Y^E(1-Z^E)}{(1-\pi_Z(x))\gamma(x)}\middle| Z^E = 0, X^E = x\right] P(Z^E = 0 \mid X^E = x) \\
&= \mathbb{E}\left[\frac{Y^E}{\pi_Z(x)\gamma(x)}\middle| Z^E = 1, X^E = x\right]\pi_Z(x) \\
&\quad - \mathbb{E}\left[\frac{Y^E}{(1-\pi_Z(x))\gamma(x)}\middle| Z^E = 0, X^E = x\right](1-\pi_Z(x)) \\
&= \frac{\mathbb{E}\left[Y^E \mid Z^E = 1, X^E = x\right] - \mathbb{E}\left[Y^E \mid Z^E = 0, X^E = x\right]}{\gamma(x)} \\
&= \frac{\mathbb{E}\left[Y^E \mid Z^E = 1, X^E = x\right] - \mathbb{E}\left[Y^E \mid Z^E = 0, X^E = x\right]}{\mathbb{E}\left[A^E \mid Z^E = 1, X^E = x\right] - \mathbb{E}\left[A^E \mid Z^E = 0, X^E = x\right]} = \tau(x) \qquad \text{(Equation 5.3)}
\end{aligned}
$$

where the intermediate steps follow from the definitions of $\pi_Z(x)$ and $\gamma(x)$ and the last equality comes from the identification result in Equation 5.3.

D.2.3 Proof of Theorem 5.5

For simplicity, we omit the E subscripts from X^E, Z^E, A^E, Y^E throughout this proof. Furthermore, assume that n_E is an integer multiple of the number of folds K. Let $\widehat{\mathbb{E}}_k f(Z) = \frac{1}{|\mathcal{I}_k|}\sum_{i\in\mathcal{I}_k} f(Z_i)$, recalling that $\mathcal{I}_k = \{i \in \{1,\ldots,n_E\} : i = k-1 \pmod K\}$, which indexes the subset of data in the k^{th} fold. Then, we can write the estimated parameter $\widehat{\theta}$ as:

$$
\begin{aligned}
\widehat{\theta} = &\left(\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\widehat{w}^{(k)}(X)^2\phi(X)\phi(X)^T\right]\right)^{-1} \\
&\cdot \frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\left(YZ(1-\widehat{\pi}_Z^{(k)}(X)) - Y(1-Z)\widehat{\pi}_Z^{(k)}(X) - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X)\right)\widehat{w}^{(k)}(X)\phi(X)\right]
\end{aligned}
$$

We also define the following quantities:

$$\begin{aligned}
\widetilde{\theta}_{n_E} &= \widehat{\mathbb{E}}_{n_E}\left[w(X)^2\phi(X)\phi(X)^T\right]^{-1} \\
&\quad \cdot \widehat{\mathbb{E}}_{n_E}\left[\left(YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau^O(X)\right)w(X)\phi(X)\right] \\
\widetilde{\theta} &= \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T\right]^{-1} \\
&\quad \cdot \mathbb{E}\left[\left(YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau^O(X)\right)w(X)\phi(X)\right]
\end{aligned}$$

We note that these quantities are well defined because $\mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T\right]$ is invertible. This follows from the first and last conditions of Assumption 5.4, along with the stipulation in Assumption 5.1 that $\gamma(x) \neq 0$ for all x in a set of non-zero measure. Using these definitions, we can write

$$\begin{aligned}
\left\|\widehat{\theta} - \theta\right\|_2 &= \left\|\widehat{\theta} - \widetilde{\theta}_{n_E} + \widetilde{\theta}_{n_E} - \widetilde{\theta} + \widetilde{\theta} - \theta\right\|_2 \\
&\leq \underbrace{\left\|\widehat{\theta} - \widetilde{\theta}_{n_E}\right\|_2}_{\lambda_1} + \underbrace{\left\|\widetilde{\theta}_{n_E} - \widetilde{\theta}\right\|_2}_{\lambda_2} + \underbrace{\left\|\widetilde{\theta} - \theta\right\|_2}_{\lambda_3} \qquad \text{(Triangle Inequality)}
\end{aligned}$$

We study these terms separately. We notice that λ_2 is just linear regression of the modified outcome $YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau^O(X)$ on $\phi(X)$ using weights $w(X)$. Given the regularity conditions in Assumption 5.4 (which subsume the standard regularity conditions of linear regression), we have that λ_2 is $O_p(1/\sqrt{n_E})$. Then, consider the $\tilde{\theta}$ term. We have:

$$\begin{aligned}
\tilde{\theta} &= \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T\right]^{-1} \\
&\quad \cdot \mathbb{E}\left[\left(YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau^O(X)\right)w(X)\phi(X)\right] \\
&= \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T\right]^{-1} \\
&\quad \cdot \mathbb{E}\left[\left(YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau(X) + w(X)\theta^T\phi(X)\right)w(X)\phi(X)\right] \\
&\qquad\qquad \text{(Realizability of } b(X)\text{)} \\
&= \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T \mid \gamma(X) \neq 0\right]^{-1} P(\gamma(X) \neq 0)^{-1} \\
&\quad \cdot \Big(\mathbb{E}\left[\left(YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau(X)\right)w(X)\phi(X) \mid \gamma(X) \neq 0\right]
\end{aligned}$$

$$+ \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T\theta \mid \gamma(X) \neq 0\right]\Big)P(\gamma(X) \neq 0)$$

$$= \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T \mid \gamma(X) \neq 0\right]^{-1}$$

$$\cdot \mathbb{E}\left[(YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau(X))\, w(X)\phi(X) \mid \gamma(X) \neq 0\right] + \theta$$

$$= \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T \mid \gamma(X) \neq 0\right]^{-1}$$

$$\cdot \mathbb{E}\left[\left(\frac{YZ}{\pi_Z(X)\gamma(X)} - \frac{Y(1-Z)}{(1-\pi_Z(X))\gamma(X)} - \tau(X)\right) w(X)^2\phi(X)\middle|\gamma(X) \neq 0\right] + \theta$$

(Since $\gamma(X) \neq 0$ implies $w(X) \neq 0$ by Assumption 5.4)

$$= \theta. \qquad \text{(Lemma 5.3)}$$

Thus, $\tilde{\theta} = \theta$ which implies $\lambda_3 = 0$. We now tackle the λ_1 term. To streamline the exposition, let us introduce the following shorthand notation:

$$\widehat{Y}^{(k)} := YZ(1-\widehat{\pi}_Z^{(k)}(X)) - Y(1-Z)\widehat{\pi}_Z^{(k)}(X)$$

$$\widetilde{Y} := YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X)$$

$$\widehat{\Sigma}_K := \frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\widehat{w}^{(k)}(X)^2\phi(X)\phi(X)^T\right]$$

$$\Sigma_K := \mathbb{E}\left[\widehat{w}^{(k)}(X)^2\phi(X)\phi(X)^T\right]$$

$$\widehat{\Sigma} := \widehat{\mathbb{E}}_{n_E}\left[w(X)^2\phi(X)\phi(X)^T\right]$$

$$\Sigma := \mathbb{E}\left[w(X)^2\phi(X)\phi(X)^T\right]$$

We can then write the $\widehat{\theta} - \widetilde{\theta}_{n_E}$ as follows:

$$\widehat{\theta} - \widetilde{\theta}_{n_E}$$

$$= \widehat{\Sigma}_K^{-1}\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\left(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X)\right)\widehat{w}^{(k)}(X)\phi(X)\right]$$

$$- \widehat{\Sigma}^{-1}\widehat{\mathbb{E}}_{n_E}\left[\left(\widetilde{Y} - w(X)\tau^O(X)\right)w(X)\phi(X)\right]$$

$$= (\widehat{\Sigma}_K^{-1} - \widehat{\Sigma}^{-1})\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\left(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X)\right)\widehat{w}^{(k)}(X)\phi(X)\right] \qquad (\lambda_{1,1})$$

$$+ \widehat{\Sigma}^{-1}\frac{1}{K}\sum_{k=1}^{K}\Big(\mathbb{E}[(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X))\widehat{w}^{(k)}(X)\phi(X)]$$

$$- \mathbb{E}[(\widetilde{Y} - w(X)\tau^O(X))w(X)\phi(X)]\Big) \qquad (\lambda_{1,2})$$

$$+ \widehat{\Sigma}^{-1}\frac{1}{K}\sum_{k=1}^{K}(\widehat{\mathbb{E}}_k - \mathbb{E})\Big[\left(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X)\right)\widehat{w}^{(k)}(X)\phi(X)$$

$$- \left(\widetilde{Y} - w(X)\tau^O(X)\right)w(X)\phi(X)\Big] \qquad (\lambda_{1,3})$$

By Cauchy-Schwartz, we can bound the λ_1 term as

$$\lambda_1 = \left\|\widehat{\theta} - \widetilde{\theta}_{n_E}\right\|_2 \leq \sum_{i=1}^{3}\|\lambda_{1,i}\|_2,$$

where we used the $\lambda_{1,i}$ notation introduced in the preceding equation. We bound each of the $\lambda_{1,i}$'s separately. We let $\|A\|_F$ denote the Frobenius norm of the matrix A. Then, consider $\lambda_{1,1}$:

$$\begin{aligned}
&\|\lambda_{1,1}\|_2 \\
&\leq \left\|\widehat{\Sigma}_K^{-1} - \widehat{\Sigma}^{-1}\right\|_F \left\|\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\left(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X)\right)\widehat{w}^{(k)}(X)\phi(X)\right]\right\|_2 \\
&\qquad\qquad \text{(Cauchy-Schwartz)} \\
&= \left\|\widehat{\Sigma}_K^{-1}(\widehat{\Sigma} - \widehat{\Sigma}_K)\widehat{\Sigma}^{-1}\right\|_F \left\|\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\left(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X)\right)\widehat{w}^{(k)}(X)\phi(X)\right]\right\|_2 \\
&\leq \left\|\widehat{\Sigma}_K^{-1}\right\|_F \left\|\widehat{\Sigma} - \widehat{\Sigma}_K\right\|_F \left\|\widehat{\Sigma}^{-1}\right\|_F \left\|\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\left(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X)\right)\widehat{w}^{(k)}(X)\phi(X)\right]\right\|_2 \\
&= O_p\left(\left\|\widehat{\Sigma} - \widehat{\Sigma}_K\right\|_F\right) \qquad \text{(By the boundedness conditions in Assumption 5.4)}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\widehat{\Sigma} - \widehat{\Sigma}_K &= \widehat{\mathbb{E}}_{n_E}\left[w(X)^2\phi(X)\phi(X)^T\right] - \frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k\left[\widehat{w}^{(k)}(X)^2\phi(X)\phi(X)^T\right] \\
&= \frac{1}{K}\sum_{k=1}^{K}(\widehat{\mathbb{E}}_k - \mathbb{E})\left[\left(w(X)^2 - \widehat{w}^{(k)}(X)^2\right)\phi(X)\phi(X)^T\right] \\
&\quad + \mathbb{E}\left[\left(w(X)^2 - \widehat{w}^{(k)}(X)^2\right)\phi(X)\phi(X)^T\right] \\
\Rightarrow \left\|\widehat{\Sigma} - \widehat{\Sigma}_K\right\|_F &\leq \frac{1}{K}\sum_{k=1}^{K}\left\|(\widehat{\mathbb{E}}_k - \mathbb{E})\left[\left(w(X)^2 - \widehat{w}^{(k)}(X)^2\right)\phi(X)\phi(X)^T\right]\right\|_F
\end{aligned}$$

$$
\begin{aligned}
&+ \left\| \mathbb{E}\left[\left(w(X)^2 - \widehat{w}^{(k)}(X)^2 \right) \phi(X)\phi(X)^T \right] \right\|_F \\
&\leq \frac{1}{K} \sum_{k=1}^{K} \sum_{i,j=1}^{d} \Big| \underbrace{(\widehat{\mathbb{E}}_k - \mathbb{E})\left[\left(w(X)^2 - \widehat{w}^{(k)}(X)^2 \right) \phi(X)_i \phi(X)_j \right]}_{:=\delta_k} \Big| \\
&\quad + \left\| w - \widehat{w}^k \right\|_{L_2} \mathbb{E}\left[\left(w(X) + \widehat{w}^{(k)}(X) \right)^2 \left\| \phi(X)\phi(X)^T \right\|_F^2 \right]^{1/2} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Holder's inequality)}
\end{aligned}
$$

By our boundedness assumptions, the second term yields an $O_p\left(\|w - \widehat{w}^k\|_{L_2}\right) = O_p(r_\gamma(n_E) + r_{\pi_Z}(n_E))$ term in the expression for $O_p\left(\|\widehat{\Sigma} - \widehat{\Sigma}_K\|_F\right)$. To analyze the first term, let E_k represent the samples in the k^{th} fold of the E dataset. Then, $\delta_k \mid E_k$ has mean 0 since $\widehat{w}^{(k)}$ is independent from E_k due to the K-fold sample splitting. Then, we can apply Chebyshev's inequality to obtain

$$
\delta_k \mid E_k = O_p\left(n_E^{-1/2}\, \mathbb{E}\left[\left(w(X)^2 - \widehat{w}^{(k)}(X)^2 \right)^2 \phi(X)_i^2 \phi(X)_j^2 \middle| E_k \right]^{1/2} \right) = o_p(1/\sqrt{n_E})
$$

from the consistency assumptions for $\widehat{\gamma}^{(k)}, \widehat{\pi}_Z^{(k)}$ which translate into a consistency assumption for $\widehat{w}^{(k)}$. By the bounded convergence theorem, this implies that δ_k is also $o_p(1/\sqrt{n_E})$. Putting everything together, we obtain

$$
\|\lambda_{1,1}\|_2 = O_p(r_\gamma(n_E) + r_{\pi_Z}(n_E)) + o_p(1/\sqrt{n_E}).
$$

We now tackle $\lambda_{1,2}$:

$$
\begin{aligned}
&\lambda_{1,2} \\
&= \widehat{\Sigma}^{-1} \frac{1}{K} \sum_{k=1}^{K} \Big(\mathbb{E}[(\widehat{Y}^{(k)} - \widehat{w}^{(k)}(X)\widehat{\tau}^O(X))\widehat{w}^{(k)}(X)\phi(X)] \\
&\qquad\qquad - \mathbb{E}[(\widetilde{Y} - w(X)\tau^O(X))w(X)\phi(X)] \Big) \\
&\|\lambda_{1,2}\|_2 \\
&\leq \|\widehat{\Sigma}^{-1}\|_F \frac{1}{K} \sum_{k=1}^{K} \\
&\quad \cdot \sum_{i=1}^{d} \left| \mathbb{E}\left[\left(\widehat{w}^{(k)}(X)\widehat{Y}^{(k)} - w(X)\widetilde{Y} - \widehat{w}^{(k)}(X)^2\widehat{\tau}^O(X) + w(X)^2\tau^O(X) \right) \phi(X)_i \right] \right|
\end{aligned}
$$

$$\leq \|\widehat{\Sigma}^{-1}\|_F \frac{1}{K} \cdot \sum_{k=1}^{K} \sum_{i=1}^{d} \left\| \mathbb{E}\left[\widehat{w}^{(k)}(X)\widehat{Y}^{(k)} - w(X)\widetilde{Y} - \widehat{w}^{(k)}(X)^2\widehat{\tau}^O(X) + w(X)^2\tau^O(X) \middle| X \right] \right\|_{L_2} \|\phi(X)_i\|_{L_2}$$

Since the terms $|\phi(X)_i|$ are bounded by Assumption 5.4, and since we have $\widehat{\Sigma}^{-1} \xrightarrow{P} \Sigma^{-1}$ by the continuous mapping theorem, it suffices to study the term $\mathbb{E}\left[\widehat{w}^{(k)}(X)\widehat{Y}^{(k)} - w(X)\widetilde{Y} - \widehat{w}^{(k)}(X)^2\widehat{\tau}^O(X) + w(X)^2\tau^O(X) \middle| X \right]$:

$$\begin{aligned}
&\left\| \mathbb{E}\left[\widehat{w}^{(k)}(X)\widehat{Y}^{(k)} - w(X)\widetilde{Y} - \widehat{w}^{(k)}(X)^2\widehat{\tau}^O(X) + w(X)^2\tau^O(X) \middle| X \right] \right\|_{L_2} \\
&\leq \|\mathbb{E}[Y \mid Z = 1, X]\pi_Z(X)\{\widehat{w}^{(k)}(X)(1 - \widehat{\pi}_Z^{(k)}(x)) - w(X)(1 - \pi_Z(X))\}\|_{L_2} \\
&\quad + \|\mathbb{E}[Y \mid Z = 0, X](1 - \pi_Z(X))\{\widehat{w}^{(k)}(X)\widehat{\pi}_Z^{(k)}(x) - w(X)\pi_Z(X)\}\|_{L_2} \\
&\quad + \|\widehat{w}^{(k)}(X)^2\widehat{\tau}^O(X) - w(X)^2\tau^O(X)\|_{L_2} \\
&\lesssim \|\widehat{w}^{(k)} - w\|_{L_2} + \|\widehat{\gamma}^{(k)} - \gamma\|_{L_2} + \|\widehat{\pi}_Z^{(k)} - \pi_Z\|_{L_2} + \|\widehat{\tau}^O - \tau^O\|_{L_2} \\
&\qquad\qquad \text{(Boundedness assumptions)} \\
&\lesssim \|\widehat{\gamma}^{(k)} - \gamma\|_{L_2} + \|\widehat{\pi}_Z^{(k)} - \pi_Z\|_{L_2} + \|\widehat{\tau}^O - \tau^O\|_{L_2} \qquad \text{(Definition of } w(X)\text{)} \\
&\leq r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O)
\end{aligned}$$

where $\lesssim$ absorbs constants. Thus, $\|\lambda_{1,2}\|_2$ is $O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O)\right)$. Lastly, we note that $\lambda_{1,3}$ is the empirical process equivalent of $\lambda_{1,2}$ and thus, by leveraging sample splitting through arguments similar those used for the $\lambda_{1,1}$ term, we have that $\|\lambda_{1,3}\|_2$ is $o_p(1/\sqrt{n_E})$. Putting all $\lambda_{1,i}$ terms together, we have that λ_1 is $O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O)\right) + o_p(\sqrt{n_E})$. Recall that λ_2 is $O_p(1/\sqrt{n_E})$ and $\lambda_3 = 0$, we obtain the desired result:

$$\left\|\widehat{\theta} - \theta\right\|_2 = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O) + 1/\sqrt{n_E}\right).$$

Given that $\|\widehat{\tau} - \tau\|_{L_2} = \|(\theta - \widehat{\theta})^T\phi(X) + (\tau^O(X) - \widehat{\tau}^O(X))\|_{L_2}$, we further have

$$\left\|\widehat{\tau} - \tau\right\|_{L_2} = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O) + 1/\sqrt{n_E}\right)$$

by using the derived $\widehat{\theta}$ rates, the Cauchy-Schwartz inequality and the boundedness of $\|\phi(X)\|_2$ assumption. Our proof is now complete.

D.2.4 Proof of Theorem 5.8

We first study the convergence rate of $\widehat{\tau}^O$ using the conditions of Theorem 5.8. Assume that h^O and $\phi(x)$ solve the following joint optimization problem:

$$\widehat{h}^O, \widehat{\phi} = \arg\min_{h^O \in \mathbb{R}^d, \phi \in \Phi} \sum_{i=1}^{n_O} \left(\left(\frac{Y^O A^O}{\widehat{\pi}_A(X)} - \frac{Y^O(1-A^O)}{1-\widehat{\pi}_A(X)} \right) - (h^O)^T \phi(X^O) \right)^2$$

Then, $\widehat{\tau}^O(x) = (\widehat{h}^O)^T \widehat{\phi}(x)$. Thus, we write:

$$\begin{aligned}
\left\| \tau^O - \widehat{\tau}^O \right\|_{L_2} &\leq \left\| (h^O)^T \phi(X) - (\widehat{h}^O)^T \widehat{\phi}(X) \right\|_{L_2} \\
&\leq \left\| (h^O)^T \phi(X) - (\widehat{h}^O)^T \phi(X) \right\|_{L_2} + \left\| (\widehat{h}^O)^T (\phi(X) - \widehat{\phi}(X)) \right\|_{L_2} \\
&\lesssim \left\| h^O - \widehat{h}^O \right\|_2 + r_\phi(n_O) \qquad \text{(Boundedness assumptions)}
\end{aligned}$$

We further expand the first term:

$$\begin{aligned}
\left\| h^O - \widehat{h}^O \right\|_2 &= \left\| \mathbb{E}[\phi(X)\phi(X)]^{-1} \mathbb{E}[\widetilde{Y}\phi(X)] - \widehat{\mathbb{E}}_{n_O}[\widehat{\phi}(X)\widehat{\phi}(X)]^{-1} \widehat{\mathbb{E}}_{n_O}[\widetilde{Y}\widehat{\phi}(X)] \right\|_2 \\
&\qquad \left(\tilde{Y} := \tfrac{Y^O A^O}{\widehat{\pi}_A(X)} - \tfrac{Y^O(1-A^O)}{1-\widehat{\pi}_A(X)} \right) \\
&\leq \left\| \mathbb{E}[\phi(X)\phi(X)]^{-1} \mathbb{E}[\widetilde{Y}\phi(X)] - \mathbb{E}[\widehat{\phi}(X)\widehat{\phi}(X)]^{-1} \mathbb{E}[\widetilde{Y}\widehat{\phi}(X)] \right\|_2 \\
&\quad + \left\| \mathbb{E}[\widehat{\phi}(X)\widehat{\phi}(X)]^{-1} \mathbb{E}[\widetilde{Y}\widehat{\phi}(X)] - \widehat{\mathbb{E}}_{n_O}[\widehat{\phi}(X)\widehat{\phi}(X)]^{-1} \widehat{\mathbb{E}}_{n_O}[\widetilde{Y}\widehat{\phi}(X)] \right\|_2 \\
&= O_p(r_\phi(n_O) + 1/\sqrt{n_O})
\end{aligned}$$

Thus, $\left\| \tau^O - \widehat{\tau}^O \right\|_{L_2}$ is $O_p(r_\phi(n_O) + 1/\sqrt{n_O})$. Next, we build upon the insights provided by the Proof of Theorem 5.5. We note that we can apply the same analysis as in the Proof of Theorem 5.5 by using $\widehat{\phi}$ instead of ϕ and everything goes through except the λ_3 term which is not 0 since ν depends on ϕ and not $\widehat{\phi}$. Thus, the convergence rate of $\|\widehat{\nu} - \nu\|_2$ will be $O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_{\tau^O}(n_O) + 1/\sqrt{n_E}\right) = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_\phi(n_O) + 1/\sqrt{n_E} + 1/\sqrt{n_O}\right)$ plus a term that depends on the deviation between $\widehat{\phi}$ and ϕ. This term is given by:

$$\lambda_3 = \left\| \mathbb{E}\left[w(X)^2 \widehat{\phi}(X) \widehat{\phi}(X)^T \right]^{-1} \right.$$

$$\cdot \mathbb{E}\left[\left(YZ(1-\pi_Z(X)) - Y(1-Z)\pi_Z(X) - w(X)\tau^O(X)\right) w(X)\widehat{\phi}(X)\right] - \nu\Big\|_2$$

$$= \left\|\mathbb{E}\left[w(X)^2\widehat{\phi}(X)\widehat{\phi}(X)^T\middle|\gamma(X) \neq 0\right]^{-1} \mathbb{E}\left[w(X)^2\widehat{\phi}(X)\phi(X)^T\nu\middle|\gamma(X) \neq 0\right] - \nu\right\|_2$$

(Lemma 5.3)

$$= \left\|\mathbb{E}\left[w(X)^2\widehat{\phi}(X)\widehat{\phi}(X)^T\middle|\gamma(X) \neq 0\right]^{-1} \mathbb{E}\left[w(X)^2\widehat{\phi}(X)(\phi(X) - \widehat{\phi}(X))^T\nu\middle|\gamma(X) \neq 0\right]\right\|_2$$

$$= O_p(r_\phi(n_O))$$

This term then gets absorbed into $O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_\phi(n_O) + 1/\sqrt{n_E} + 1/\sqrt{n_O}\right)$. Thus, we obtain the desired results:

$$\|\widehat{\nu} - \nu\|_2 = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_\phi(n_O) + 1/\sqrt{n_E} + 1/\sqrt{n_O}\right),$$

and

$$\|\widehat{\tau} - \tau\|_{L_2} = O_p\left(r_\gamma(n_E) + r_{\pi_Z}(n_E) + r_\phi(n_O) + 1/\sqrt{n_E} + 1/\sqrt{n_O}\right).$$

D.3 Additional Experimental Details

D.3.1 Simulation Studies

Implementation Details The results for the parametric extension from Section 5.5.1 were generated on a consumer laptop equipped with a 13th Gen Intel Core i7 CPU. The execution took approximately 1.5 minutes using 20 concurrent workers. In contrast, the representation learning outcomes were derived using an NVIDIA Tesla T4 GPU on Google Colab [Google, 2024]. The execution took roughly 1.5 hours, with half the time spent on Algorithm 5.2 and the other half on learning $\widehat{\tau}^E(x)$ over 100 iterations.

The Random Forest (RF) models employed in Algorithm 5.1 make use of the `RandomForestRegressor` and `RandomForestClassifier` algorithms from the `scikit-learn` [Pedregosa et al., 2011] Python library. For the feed-

Table D.1: Hyperparameters for the synthetic-data experiments in Chapter 5..

Method	Model(s)	Algorithm	Hyperparameter	Value
Algorithm 5.1	Compliance	Random Forest (`scikit-learn`)	max_depth	3
			min_samples_leaf	50
Algorithm 5.1	Outcomes	Random Forest (`scikit-learn`)	max_depth	5
			min_samples_leaf	5
Algorithm 5.2	Representation CATE Compliance	Neural Network (`PyTorch`)	activation	ELU
			hidden units	2
			network depth	5
			weight_decay	0.02
			optimizer	Adam
			learning rate	0.01
			batch size	2000
			epochs	1000

forward neural networks within the representation learning component, we utilize the `nn` module from the `PyTorch` package [Paszke et al., 2019]. Details regarding the hyperparameters for these models are provided in Table D.1.

We configured the parameters for the Random Forest (RF) models based on the theoretical guidance outlined in Probst et al. [2019]. For the neural networks, we implemented early stopping using a validation dataset that constituted 20% of the total generated datasets.

Result for High-Dimensional DGP We perform additional experiments to highlight the effectiveness of our method in higher-dimensional settings. To this aim, we modify the DGP in Section 5.5.1 to include d features $X^d \in \mathbb{R}^d$, with both baselines and bias depending on all features as follows:

$$Y = 1 + A + X_1 + 2A\beta^T X + 0.5X_1^2 + 0.75AX_1^2 + U + 0.5\epsilon_Y$$
$$U \mid X, A \sim N\left(\gamma^T X\,(A - 0.5)\,, 0.75\right)$$

where the coefficients $\beta, \gamma \in [-1, 1]^d$ are set at random at the beginning of the simulation. In this setting, the bias function is given by $b(x) = -\gamma^T x$. We leave all other settings and parameters (including $n_O = n_E = 5,000$) unchanged and

Table D.2: MSE ± standard deviation for the Chapter 5 estimators in a high-dimensional synthetic data-generating process.

	$\widehat{\tau}^O(x)$	$\widehat{\tau}^E(x)$	$\widehat{\tau}(x)$
$d = 5$	1.40 ± 0.09	3.97 ± 1.21	0.40 ± 0.07
$d = 10$	3.25 ± 0.15	7.70 ± 1.54	1.25 ± 0.20
$d = 20$	9.32 ± 0.51	19.2 ± 2.58	4.05 ± 0.68
$d = 50$	37.1 ± 0.94	43.2 ± 2.89	9.39 ± 1.64

perform parametric extrapolation using Algorithm 5.1.

In Table D.2, we report the mean squared error (MSE) and standard deviation (SD) of predictions on a fixed sample of 1,000 points drawn from the same distribution as X, over 100 iterations and for various dimensions ($d \in 5, 10, 20, 50$). The high MSE of the IV estimator $\widehat{\tau}^E(x)$ reflects the challenges of estimating compliance in high-dimensional settings. Likewise, the observational data estimator $\widehat{\tau}^O(x)$ shows clear bias. In contrast, the combined data estimator $\widehat{\tau}(x)$ from Algorithm 5.1 significantly outperforms both, demonstrating improved accuracy in this high-dimensional context.

D.3.2 Impact of 401(k) Participation on Financial Wealth

Implementation Details The dataset from Chernozhukov and Hansen [2004] is comprised 9,915 observations with 9 covariates: age, income, education, family size, marital status, two-earner household status, defined benefit pension status, IRA participation, and home ownership indicators. We describe the features of the 401(k) dataset in Table D.3.

Given the heavy-tailed distribution of net worth measures, we perform a pre-processing step to remove outliers. Specifically, we eliminate the top and bottom 2.5% of observations, effectively narrowing the range of potential outcomes from $[-0.5\times10^6, 1.5\times10^6]$ to $[-1.4\times10^4, 1.34\times10^5]$. This adjustment leaves us with 9,419 observations, which are then evenly distributed between the ob-

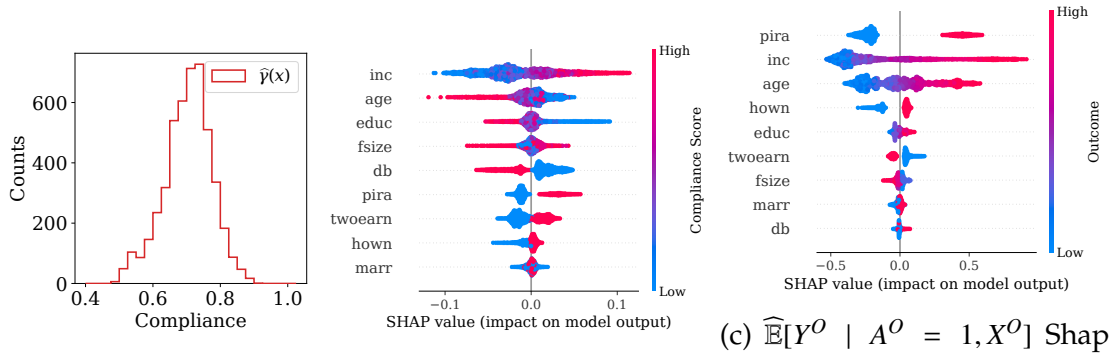
Table D.3: Description of the covariates in the 401(k) dataset used in the Chapter 5 application.

Name	Description	Type
age	age	continuous covariate
inc	income	continuous covariate
educ	years of completed education	continuous covariate
fsize	family size	continuous covariate
marr	marital status	binary covariate
two_earn	whether dual-earning household	binary covariate
db	defined benefit pension status	binary covariate
pira	IRA participation	binary covariate
hown	home ownership	binary covariate
e401	401(k) eligibility	binary instrument
p401	401(k) participation	binary treatment
net_tfa	net financial assets	continuous outcome

servational and experimental datasets. We find that this procedure improves the stability of regression and classification algorithms across different random data splits.

This dataset has previously been analyzed using Random Forest algorithms in [Chernozhukov et al., 2018a]. Consistent with this work, we employ the same models (`RandomForestRegressor` and `RandomForestClassifier` from `scikit-learn`) and use identical hyperparameters (n_estimators = 100, max_depth = 6, max_features = 3, min_samples_leaf = 10) for various regression and classification tasks outlined in Algorithm 5.1. For the second stage of Algorithm 5.1, we use a `Lasso` regressor from `scikit-learn` with a penalty of $\alpha = 0.07$ selected via 5-fold cross-validation.

In Figure D.1, we display several characteristics of the 401(k) dataset derived from the first stages of Algorithm 5.1. In particular, we illustrate the spread in compliance scores in IV dataset, as well as the impact of important features on the predictions of the compliance and outcome models, respectively. As noted



(a) $\widehat{\gamma}(x)$ Histogram (b) $\widehat{\gamma}(x)$ Shap Plot (c) $\widehat{\mathbb{E}}[Y^O \mid A^O = 1, X^O]$ Shap Plot

Figure D.1: First-stage diagnostics for the 401(k) experiment in Chapter 5. (D.1a): Distribution of estimated compliance scores for $x \in X^E$. (D.1b): Shapley plot [Lundberg and Lee, 2017] for the compliance model in the IV dataset with features arranged in decreasing order by feature importance. (D.1c): Shapley plot for the estimated outcome model $\widehat{\mathbb{E}}[Y^O \mid A^O = 1, X^O]$ in the observational dataset with features arranged in decreasing order by feature importance.

Table D.4: MSE ± standard deviation across different 401(k) data splits in the Chapter 5 application. Age: 40, Income: $30,000, Single.

Educ	$\widehat{\tau}^O$ (in 1,000$)	$\widehat{\tau}^E$ (in 1,000$)	$\widehat{\tau}$ (in 1,000$)
8	11.9 ± 2.18	10.0 ± 2.23	9.83 ± 2.22
10	11.8 ± 2.17	10.2 ± 2.42	9.99 ± 2.18
12	11.8 ± 2.22	9.88 ± 2.36	10.2 ± 2.20

in the main text, the compliance scores are relatively large and range between 0.49 and 0.90 (mean=0.70). Furthermore, the primary features influencing the compliance score model include income, age, and education. In contrast, the features impacting the outcome model $\widehat{\mathbb{E}}[Y^O \mid A^O = 1, X^O]$ are IRA participation, income, and age, with education having a significantly lesser effect. This motivated us to investigate how education influences the derived CATEs.

Quantifying Uncertainty Across Data Splits We quantify our claims for the 401(k) study by repeating the experiment across 100 different (O, E) splits of the original data. We calculate the means and standard deviations of the treatment

Table D.5: MSE ± standard deviation across different 401(k) data splits in the Chapter 5 application. Age: 40, Income: $30,000, Married.

Educ	$\widehat{\tau}^O$ (in 1,000$)	$\widehat{\tau}^E$ (in 1,000$)	$\widehat{\tau}$ (in 1,000$)
8	11.3 ± 2.40	9.49 ± 2.23	9.54 ± 2.50
10	11.3 ± 2.40	9.63 ± 2.40	9.59 ± 2.38
12	11.2 ± 2.41	9.93 ± 2.39	9.65 ± 2.29

effects by years of education for the two examples described in the paper, with the results shown in Table D.4 and Table D.5. We note that the original trend (biased observational estimates, accurate extrapolation to the no-compliance region) is largely preserved, and our method demonstrates the ability to interpolate well on average in the artificially introduced non-compliance region. However, the uncertainty, as reflected by the large standard deviations, is substantial enough that the results are not statistically significant, which limits the strength of the conclusions we can draw from this experiment (unfortunately!). This is most likely due to the prevalence of outliers, as net worth follows a heavy-tailed distribution, and RF regressors tend to overfit to these extreme values.

D.4 Limitations and Societal Impacts of Our Work

Our methodology hinges on several key assumptions, and violations can significantly affect the accuracy and reliability of our estimates. First, the standard IV assumptions (Assumption 5.1) must hold. If the instrument directly affects the outcome, is correlated with unobserved confounders, or is weak across all strata of covariates, our estimates may be biased and unreliable. Some of these issues can be mitigated in experimental settings where the instrument is fully randomized. Additionally, the unconfounded compliance assumption requires that compliance is independent of potential outcomes given the covariates. Vi-

olations here can also lead to biased estimates if unrecorded explanatory variables affect both outcomes and compliance. Lastly, our method relies on realizability assumptions regarding the bias function. If these assumptions do not hold, our estimates might be biased.

The societal impacts of our method stem from potential inaccuracies in treatment effect estimates and their subsequent use. Inaccurate treatment effect estimates could lead to a range of adverse outcomes, from a diminished user experience on online platforms to less effective healthcare recommendations, economic and public policies. Furthermore, while accurate estimates can provide substantial benefits, they must be used responsibly to avoid unintended consequences such as privacy concerns or potential biases in decision-making. It is thus crucial to apply these methods with careful consideration of ethical implications and societal impacts.

APPENDIX E

APPENDIX FOR CHAPTER 6

E.1 Extended Literature Review

We contextualize our work by surveying six strands of prior research. We group these into two categories: (1) *core* threads that directly motivate and inform our methodology, and (2) *auxiliary* threads that provide important theoretical and practical foundations but are not specific to our design.

E.1.1 Core Related Work

Identification and Estimation via Instrumental Variables Instrumental variable (IV) methods are widely used to estimate causal effects in the presence of endogenous treatment selection due to unmeasured confounding. Under classical IV assumptions—exclusion, independence, and relevance—these methods typically identify the local average treatment effect (LATE) [Angrist et al., 1996, Abadie et al., 2002, Abadie, 2003, Cheng et al., 2009, Ogburn et al., 2015], which pertains to compliers: units whose treatment responds to the instrument. However, compliers represent an unknown and potentially unrepresentative subpopulation, limiting the policy relevance of LATE.

To target the population average treatment effect (ATE), IV methods have traditionally relied on linear structural equation models (SEMs), in which the ATE corresponds to a regression coefficient under correct model specification [Goldberger, 1972]. Two-stage least squares (2SLS) is the canonical estimator in this setting, but its consistency and interpretability depend on strong linearity assumptions [Wooldridge, 2010, Clarke and Windmeijer, 2012]. More recent SEM-based IV approaches focus on estimating conditional effects, including

kernel IV [Singh et al., 2019], DeepIV [Hartford et al., 2017], and other moment-based estimators [Bennett et al., 2019, Syrgkanis et al., 2019].

We instead build on the framework of Wang and Tchetgen Tchetgen [2018], who introduce an alternative identification strategy based on an unconfounded compliance assumption. This allows point identification of the ATE while separating identification assumptions from estimation model assumptions. Their approach avoids reliance on parametric SEMs for ATE estimation and instead yields an efficient semiparametric estimator with an efficient influence function (EIF) that confers the estimator a multiple-robust property, *i.e.*the estimator remains consistent if one or several nuisance components are misspecified. This structure makes the estimators well-suited to nonparametric plug-in estimation using modern machine learning tools. This insight has been extended to develop multiply-robust CATE estimators that leverage binary instruments to adjust for unobserved confounding [Frauen and Feuerriegel, 2023], and to debias confounded observational data by incorporating (potentially weak or imperfect) instruments [Oprescu and Kallus, 2025]. Other recent work extends the framework of Wang and Tchetgen Tchetgen [2018] to nonparametric identification of ATEs under related assumptions [Neopane et al., 2025a].

We adopt the framework of Wang and Tchetgen Tchetgen [2018] as the foundation for our estimator, with the goal of efficiently and robustly estimating the ATE under adaptive, sequential data collection. Specifically, we derive the semiparametric efficiency bound for ATE estimation under arbitrary, covariate-dependent instrument-assignment policies and identify the optimal adaptive policy that minimizes this bound. We then introduce AMRIV, an adaptive, multiply robust estimator that attains the bound and enables valid inference in sequential experiments.

Adaptive Experimental Design for Treatment Effect Estimation A substantial literature studies adaptive algorithms for estimating the average treatment effect (ATE) efficiently and with minimal variance when the treatment itself can be directly assigned. This line of work was initiated by Hahn et al. [2011], who proposed a two-stage explore-then-commit design that asymptotically achieves the semiparametric efficiency bound, echoing early bandit-style adaptive allocation schemes. Fully sequential designs soon followed: Kato et al. [2020] introduced the Adaptive Augmented Inverse Propensity Weighting (A2IPW) estimator, which achieves variance-optimal Neyman allocation; Kato et al. [2021] extended this to settings with estimated policies and nuisance functions and demonstrated multiply robust consistency; and Cook et al. [2024] added principled policy truncation and developed the first anytime-valid confidence sequences for adaptive ATE estimation.

Parallel progress has come from an online-learning perspective. Recent methods such as Clip-OGD and its optimistic variants attain sublinear or logarithmic "Neyman regret" for the ATE [Dai et al., 2023, Neopane et al., 2025b,c]; low-switching policies achieve finite-sample optimality bounds [Li et al., 2024]; and newer designs jointly optimize over covariate and treatment dimensions [Kato et al., 2024]. Off-policy estimators with adaptively collected data have also obtained sharp error rates and regret guarantees [Lee and Ma, 2024]. Together, these advances form a mature toolkit for adaptive experimentation that combines adaptive nuisance learning, cross-fitting, policy truncation, regret-minimizing allocation, and time-uniform inference.

Our work builds directly on this literature but extends it to the less explored setting where *only an instrument can be assigned* and treatment uptake is endogenous. We integrate core components from the direct-treatment literature—

influence-function–based estimation, adaptive policy learning, cross-fitting, policy truncation, and sequential inference—to construct a unified estimator that retains semiparametric efficiency, multiply robust consistency, and time-uniform inference under noncompliance. Specifically, we are the first to: (i) derive the semiparametric efficiency bound for ATE estimation when only an instrument can be adaptively assigned; (ii) identify the variance-optimal allocation rule that balances outcome noise and compliance variability; and (iii) develop AMRIV, a multiply robust estimator that attains this bound and supports anytime-valid inference.

Adaptive Experimentation with Instrumental Variables Recent work has begun to explore adaptive experimentation in settings where treatments cannot be directly assigned, requiring the use of instrumental variables (IVs) to estimate causal effects under unobserved confounding. Broadly, these efforts fall into two categories: methods aimed at improving predictive accuracy in the presence of confounding, and approaches focused on adaptive design and data collection or regret minimization using bandit-style feedback.

The first group focuses on improving estimation efficiency in indirect experiments. Gupta et al. [2021] propose an adaptive framework for selecting among multiple data sources to efficiently estimate causal functionals such as the ATE. Their method, Online Moment Selection (OMS), chooses which source to query at each step based on moment conditions implied by a causal graph. While they address efficient data acquisition under structural constraints, their setting assumes passive data collection and differs from our focus on adaptive experimental design with noncompliance and endogenous treatment. Ailer et al. [2024] study sequential indirect experiment design in instrumental variable settings, focusing on partial identification of nonlinear treatment effect queries.

Rather than aiming for point estimation, their method adaptively tightens upper and lower bounds on a functional $Q[f]$ of the treatment effect by selecting experiments that reduce the gap between these bounds. In contrast, our work targets point identification and estimation of the ATE, and provides semiparametric efficiency and robustness guarantees under adaptive instrument assignment. Most closely related to our setting, Chandak et al. [2024] study adaptive instrument selection to improve sample efficiency in indirect experiments. They propose a general influence-function–based optimization procedure for selecting instruments that minimize the mean squared error of nonparametric IV estimators, such as DeepIV [Hartford et al., 2017]. However, their objective is variance reduction for prediction, not inference for causal estimands like the ATE. Their analysis is estimator-specific and does not characterize semiparametric efficiency bounds or multiply robust inference, which are central to our approach.

The second line of work focuses on regret minimization in settings with instrumental feedback. Zhao et al. [2024] use randomized instruments within a linear structural equation model to enable pure exploration for policy learning under unobserved confounding. Their focus is on identifying the best treatment arm using bandit-style algorithms with finite-sample confidence intervals and near-optimal sample complexity guarantees. Unlike our work, which targets semiparametric inference for the ATE, their goal is policy optimization rather than estimation. Della Vecchia and Basu [2025] study online instrumental variable regression with bandit feedback and propose regret bounds under endogeneity. Their focus is on prediction in stochastic settings with instrumental bandit structure, rather than causal effect estimation or statistical inference.

Our Contribution To our knowledge, we present the first estimator that is both semiparametrically efficient and multiply robust for ATE estimation under a binary instrument with adaptive assignment. We characterize the efficiency bound, derive the optimal allocation policy that balances outcome and compliance variance, and develop the AMRIV estimator that asymptotically attains this bound. Our results generalize prior work on ATE estimation to settings with endogenous treatments, establish asymptotic normality, and construct time-uniform asymptotic confidence sequences. Empirical studies confirm our method's efficiency, robustness, and practical viability.

E.1.2 Auxiliary Context

Semiparametric Efficiency and Influence-Function–Based Methods Our estimator builds on a long line of semiparametric inference techniques, particularly those using influence functions to achieve efficiency and robustness in the presence of nuisance components [Bickel et al., 1993, Robins et al., 1994]. Recent work has adapted these methods to flexible machine learning settings by incorporating sample-splitting and cross-fitting [Chernozhukov et al., 2018a, Kennedy, 2023d, Frauen and Feuerriegel, 2023]. In adaptive experiments, such techniques have been shown to yield efficient estimators without requiring Donsker conditions [Kato et al., 2020, 2021, Cook et al., 2024]. We extend these tools to a setting with endogenous treatment and adaptive assignment via a binary instrument.

Multiply Robust Estimation Multiply robust estimators remain consistent if any one of multiple nuisance components is correctly specified. In the IV context, this structure has been exploited to enhance robustness of ATE and CATE estimators [Wang and Tchetgen Tchetgen, 2018, Frauen and Feuerriegel, 2023].

We extend these ideas to adaptive settings, showing that AMRIV retains consistency even when some nuisance functions are misspecified, as long as at least one of $\delta(X)$ or $\delta^A(X)$ is consistently estimated.

Confidence Sequences and Anytime-Valid Inference Confidence sequences (CSs) provide coverage guarantees that hold uniformly over time, making them well-suited to adaptive experiments with interim monitoring or early stopping. Recent work has developed CSs for influence-function–based estimators using martingale techniques and empirical Bernstein bounds [Howard et al., 2021, Waudby-Smith et al., 2024b, Cook et al., 2024]. We build on this to construct asymptotically valid confidence sequences for our adaptive IV estimator, accounting for sequential dependence and cross-fitted nuisances.

E.2 Notation

Table E.1: Notation used in Chapter 6.

X	Observed covariates (feature vector) in $\mathbb{R}^m$.
Z	Binary instrument / encouragement, $Z \in \{0, 1\}$ (experimenter-assigned).
A	Binary treatment, $A \in \{0, 1\}$, determined endogenously by the unit.
Y	Real-valued observed outcome.
$Y(a)$, $Y(a, z)$	Potential outcome under treatment a (and instrument z when shown).
$A(z)$	Potential treatment taken under instrument z.
U	Unobserved confounder(s).
P	True (observable) distribution of (X, Z, A, Y).
τ	Population average treatment effect (ATE): $\tau := \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$.
$\mu^Y(z, x)$	$\mathbb{E}[Y \mid Z = z, X = x]$, instrument-conditional outcome regression.
$\mu^A(z, x)$	$\mathbb{E}[A \mid Z = z, X = x]$, instrument-conditional treatment regression.
$\delta^Y(x)$, $\delta^A(x)$	Instrument-induced shifts in Y and A, respectively. $\delta^Y(x) = \mu^Y(1, x) - \mu^Y(0, x)$, $\delta^A(x) = \mu^A(1, x) - \mu^A(0, x)$:
$\delta(x)$	Conditional average treatment effect (CATE) at x: $\delta(x) := \delta^Y(x)/\delta^A(x)$.
$\pi_t(x \mid \mathcal{H}_{t-1})$	Instrument assignment policy at time t: $\Pr(Z_t = 1 \mid X_t = x, \mathcal{H}_{t-1})$.
$\pi^*(x)$	Efficiency-optimal instrument assignment (Eq. (6.4)).
$\mathcal{H}_{t-1}$	History before round t: $\{(X_i, Z_i, A_i, Y_i)\}_{i=1}^{t-1}$.
T	Total number of rounds / samples.
T_0	Burn-in rounds (initial exploration).
k_t	Truncation parameter at time t (ensures $\pi_t \in [1/k_t, 1 - 1/k_t]$).
$\sigma^2(z, x)$	Residual variance: $\mathrm{Var}(Y - A\delta(X) \mid Z = z, X = x)$.
$V_{\mathrm{eff}}(\pi)$	Semiparametric efficiency bound under policy π (Eq. (6.3)).
$\phi(\cdot; \pi, \eta)$	Recentered efficient influence function (EIF) used in AMRIV (Eq. (6.2)).
$\widehat{f}$, $\widetilde{f}$	$\widehat{f}$: estimate of f; $\widetilde{f}$: plug-in or candidate function.
$\widehat{\tau}_T^{\mathrm{AMRIV}}$	AMRIV estimator after T rounds (Alg. 6.1).
$\widehat{\mathbb{E}}$	Empirical expectation (sample average).
$\|g\|_2$	L_2 norm: $\|g\|_2 := \mathbb{E}[g(X)^2]^{1/2}$.
$o_p(1)$, $O_p(\cdot)$	Standard probabilistic asymptotic order notation.
ε	Small positive constant (e.g., variance-floor in Eq. (6.6)).
$\mathcal{H}_{t-1}^{(j)}$	Temporal cross-fitting fold (e.g., $j \in \{0, 1\}$) used for sequential cross-fitting.

Table E.2: Representative nuisance estimators for AMRIV, together with convergence rates and suitable applications.

Estimator	Convergence rate	Best suited for
k-NN / kernel smoother	$O\big(n^{-\beta/(2\beta+d)}\big)$ for Hölder-β smooth functions Nickl and Pötscher [2007]	Low-dimensional, smooth problems
Random forest	$\tilde{O}\big(n^{-\beta/(2\beta+d)}\big)$ for Hölder-β smooth functions Scornet et al. [2015], Wager and Athey [2018b]	Moderate-dimensional, tabular data
Fully connected neural nets (ReLU)	$O\big(\sqrt{WL\log W}\, n^{-1/2}\big)$ for width W and depth L Yarotsky [2017]	High-dimensional or structured inputs
Neural nets (1-Lipschitz, bounded weights)	$O\Big(\sqrt{\prod_{l=1}^{L} M_l}\, n^{-1/4}\Big)$, where M_l bounds the Frobenius norm of layer l's weights Golowich et al. [2018]	High-dimensional or structured inputs

E.3 Practical Implementation of Nuisance and Variance Estimators

This section complements Remark 6.6 by outlining practical choices for nuisance and variance estimators, including nonnegativity constraints and online-update considerations.

Nuisance estimators Because AMRIV uses sequential cross-fitting, we require only standard nonparametric convergence rates to establish asymptotic normality (Theorems 6.8–6.9). Consequently, the analyst may flexibly choose estimators according to data structure and sample size. For representative options, see Table E.2.

The best choice depends on sample size, covariate dimension, and smoothness. In adaptive experiments with directly assigned treatments, k-NN and ran-

dom forests are standard baselines [Kato et al., 2020, Cook et al., 2024].

Online or streaming implementations Theoretical results assume that nuisance functions are re-estimated at each update, but full data storage is not required. Practical alternatives include:

1. **Online learners:** gradient-descent (SGD) updates, online random forests [Lakshminarayanan et al., 2014], or GLMs can update parameters incrementally using new data.
2. **Block sample-splitting:** reuse recent batches to update nuisance fits and discard older data while maintaining cross-fitting validity [Jacob, 2020].
3. **Sufficient-statistic updates:** for parametric or binned regressors, sufficient statistics such as running sums of (X, Z, A, Y) per cell suffice.

Non-negative variance estimation. Equation (6.6) enforces non-negativity via a small floor parameter ε. As an alternative, one may use a *self-normalized kernel variance estimator* that is fully nonparametric and guarantees $\widehat{\sigma}_t^2(z, x) \geq 0$ by construction.

For each (z, x), define

$$\widehat{\sigma}_t^2(z, x) = \frac{\sum_{s \leq t} K_h(\|X_s - x\|)\, \mathbb{I}\{Z_s = z\}\, (R_s - \widehat{\mu}_{t-1}(z, x))^2}{\sum_{s \leq t} K_h(\|X_s - x\|)\, \mathbb{I}\{Z_s = z\}},$$
$$\widehat{\mu}_{t-1}(z, x) = \frac{\sum_{s \leq t} K_h(\|X_s - x\|)\, \mathbb{I}\{Z_s = z\}\, R_s}{\sum_{s \leq t} K_h(\|X_s - x\|)\, \mathbb{I}\{Z_s = z\}},$$

where $R_s = Y_s - A_s \widehat{\delta}_{t-1}(X_s)$ and $K_h(\cdot)$ is a kernel with bandwidth h.

This estimator satisfies $\widehat{\sigma}_t^2(z, x) \geq 0$ by construction and is consistent for $\mathrm{Var}(Y - A\delta(X) \mid Z = z, X = x)$ under standard kernel conditions. **Caveat:** updating kernel weights online can be $O(t^2)$ in general; for large-scale data, the plug-in estimator in Eq. (6.6) is more computationally efficient.

Both the plug-in estimator with the floor ε and the kernel-based version

above preserve all asymptotic guarantees; the choice between them is primarily a trade-off between computational efficiency and strict non-negativity.

E.4 Asymptotic Confidence Sequences

The fixed–time intervals in Section 6.6 guarantee $(1-\alpha)$ coverage *solely* at a pre-specified sample size T. In practice, however, analysts often *peek* at interim results and may stop the study early once a decision rule is met [Ramdas et al., 2023], behavior that invalidates fixed-time intervals. To remain valid under such data-dependent stopping one needs a **confidence sequence (CS)**—a collection of intervals

$$[L_t,\ U_t]_{t\geq 1}$$

satisfying the time-uniform guarantee

$$P\left(\forall t \in \mathbb{N}^+ : \tau \in [L_t, U_t]\right) \ \geq \ 1-\alpha.$$

Constructing *non-asymptotic*, anytime-valid CSs can be difficult when the target estimand contains estimated nuisance functions. Fortunately, for AMRIV the nuisance-induced remainder is $o_p(t^{-1/2})$ under Theorem 6.8 assumptions, so an *asymptotic* CS—valid after a finite, burn-in phase— remains both tractable and practically useful.

Definition E.1 (Asymptotic time-uniform coverage [Dalal et al., 2024, Def. 2.1 & 2.3])**.** A sequence of random intervals $\widetilde{C}_t = [\widetilde{L}_t, \widetilde{U}_t]_{t\geq 1}$ is an asymptotic time uniform $(1-\alpha)$ confidence sequence (AsympCS) for a parameter τ if the following two conditions hold.

(i) **Asymptotic confidence sequence:** there exists an exact (potentially unknown) $(1-\alpha)$ confidence sequence $C_t^\star = [L_t^\star, U_t^\star]_{t\geq 1}$ such that $\widetilde{L}_t/L_t^\star \to 1$ and $\widetilde{U}_t/U_t^\star \to 1$ almost surely.

(ii) **Asymptotic time-uniform coverage:**

$$\lim_{T_0\to\infty} P\Big(\forall t \geq T_0 : \tau \in \widetilde{C}_t\Big) \geq 1-\alpha.$$

Definition E.1 can be read as follows: if one waits to "peek" until the sample size is sufficiently large ($T \geq T_0$ for some burn-in T_0), the band then covers the true parameter at *every* later time with probability approaching $1-\alpha$. In practice, the rare coverage failures occur almost exclusively during this short initial window; once past it, the intervals tighten rapidly and deliver appreciable power gains over fully non-asymptotic sequences [Waudby-Smith et al., 2024a, Cook et al., 2024].

Building on the methodologies of Waudby-Smith et al. [2024b] and Cook et al. [2024], we now present the corresponding asymptotic confidence-sequence (AsympCS) results for our estimator:

Theorem E.2 (AsympCS for AMRIV)**.** *Suppose Assumptions 5.2, 6.1 and 6.7 hold and there exists a non-adaptive policy $\pi(X) \in [\epsilon, 1-\epsilon]$ for some $\epsilon > 0$ such that the nuisances estimates $\widehat{\eta}_t$ and the adaptive assignment policy $\pi_t(X \mid \mathcal{H}_{t-1})$ are L_2-consistent relative to the truncation schedule, i.e. $k_t\|\widehat{f}_{t-1} - f\|_2 = o_p(1)$ and $k_t\|\pi_t - \pi\|_2 = o_p(1)$ for $f \in \{\mu^Y(0,\cdot), \mu^A(0,\cdot), \delta(\cdot), \delta^A(\cdot)\}$. Furthermore, assume $\|\hat{\delta}_{t-1} - \delta\|_2 = o_{a.s.}(1)$ and $\|\hat{\delta}_{t-1} - \delta\|_2\|\widehat{\delta}^A_{t-1} - \delta^A\|_2 = o_{a.s.}\Big(\sqrt{\frac{\log t}{t}}\Big)$. Let*

$$\widehat{V}_T := \frac{1}{T}\sum_{t=1}^{T}\Big(\phi(X_t, Z_t, A_t, Y_t;\, \pi_t, \widehat{\eta}_t) - \widehat{\tau}^{AMRIV}_T\Big)^2,$$

be the estimated variance of $\{\phi(X_t, Z_t, A_t, Y_t;\, \pi_t, \widehat{\eta}_t)\}$, and fix a user-specified $\rho > 0$. Then, for all $T > 0$, the interval

$$\tilde{C}^{\text{AsympCS}}_T := \left(\widehat{\tau}^{AMRIV}_T \pm \sqrt{\frac{2(T\widehat{V}_T\rho^2+1)}{T^2\rho^2}\,\log\left(\frac{\sqrt{T\widehat{V}_T\rho^2+1}}{\alpha}\right)}\right)$$

forms an asymptotic $(1-\alpha)$ *confidence sequence (as in Definition E.1) for* τ*. Furthermore, width of* $\tilde{C}_T^{\text{AsympCS}}$ *is (approximately) minimized at*

$$\rho^\star \;=\; \sqrt{\frac{-2\log\alpha \;+\; \log(-2\log\alpha+1)}{T}}.$$

Remark E.3 (Difference in convergence rates for fixed-time and anytime-valid inference.). The conditions under which the AsympCS in Theorem E.2 is valid are largely the same as those imposed for fixed time inference in Theorem 6.8. The main difference lies in the convergence conditions for the nuisance functions. While $\|\hat{\delta}_{t-1}-\delta\|_2$ and $\|\hat{\delta}_{t-1}-\delta\|_2\|\hat{\delta}^A_{t-1}-\delta^A\|_2$ are assumed to converge in probability for fixed time inference, confidence sequences require convergence almost surely, as noted by Waudby-Smith et al. [2024a]. However, the conditions for valid inference are not necessarily stricter than fixed-time inference, as the product error term is allowed to converge a slower $\sqrt{\log t/t}$ rate rather than at rate $t^{-1/2}$

Proof of Theorem E.2. For the proof of Theorem E.2, we rely on an existing result for asymptotic confidence sequences [Waudby-Smith et al., 2024a]. For completeness, we provide this result in Lemma E.4 below.

Lemma E.4 (Corollary 3.4 of Waudby-Smith et al. [2024a])**.** *Suppose* $\hat{\theta}_t$ *is an asymptotically linear estimator of* θ *with influence function* ϕ *that satisfies*

$$\hat{\theta}_t-\theta=\frac{1}{t}\sum_{i=1}^{t}\phi(X_i,Z_i,A_i,Y_i;\pi_i,\eta)+o_{a.s.}\left(\sqrt{\frac{\log t}{t}}\right). \tag{E.1}$$

Furthermore, suppose that $\mathrm{Var}(\phi)<\infty$*. Then,* $\left(\hat{\theta}_t\pm\sqrt{\frac{2(t\rho^2+1)}{t^2\rho^2}\log\left(\frac{\sqrt{t\rho^2+1}}{\alpha}\right)}\right)$ *forms a valid* $(1-\alpha)$*-AsympCS (as in Definition E.1) for* θ*.*

Using Lemma E.4, we only need to show that (i) the residual error of our estimator is of a smaller order than $\sqrt{\log t/t}$ almost surely and (ii) the variance of the limiting influence function ϕ is bounded.

Verifying Residual Error Our proof of the residual error bound follows similar steps to the proof of Theorem 6.8. We first rewrite the difference between our estimate $\hat{\tau}_t^{AMRIV}$ and τ as

$$\widehat{\tau}_t^{AMRIV} - \tau = \frac{1}{t}\sum_{i=1}^{t} \phi(X_i, A_i, Z_i, Y_i; \pi_i, \eta) - \frac{1}{t}\sum_{i=1}^{t} m_t, \tag{E.2}$$

where $m_i = \phi(X_i, A_i, Z_i, Y_i; \pi_i, \hat{\eta}_i) - \phi(X_i, A_i, Z_i, Y_i; \pi_t, \eta)$. We repeat the same arguments as the proof of Theorem 6.8, which shows that the cumulative residual error (i.e. the sum of m_t) vanish at $o_p(1/\sqrt{t})$ rates. Replacing assumptions $\|\hat{\delta}_{t-1} - \delta\|_2 = o_p(1)$ with $\|\hat{\delta}_{t-1} - \delta\|_2 = o_{a.s.}(1)$ and $\|\hat{\delta}_{t-1} - \delta\|_2 \|\hat{\delta}^A_{t-1} - \delta^A\|_2 = o_p(t^{-1/2})$ with $\|\hat{\delta}_{t-1} - \delta\|_2 \|\hat{\delta}^A_{t-1} - \delta^A\|_2 = o_{a.s.}\left(\sqrt{\frac{\log t}{t}}\right)$, we obtain $\sum_{i=1}^{t} m_t = o_{a.s.}(\sqrt{t \log t})$. Normalizing by t, we obtain the desired result.

Finite Variance of ϕ The finite variance of limiting influence function ϕ is immediate from Assumption 6.7 and the condition that $\pi(X) \in [\epsilon, 1-\epsilon]$ for some $\epsilon > 0$. Under these assumptions, for any tuple (X, Z, A, Y), ϕ is bounded almost surely as a function of some constant C,

$$|\phi(X, Z, A, Y; \pi, \eta)| = \frac{2}{\epsilon}\left(3C^2 + C^3\right) + C.$$

Using this bound, we now upper bound the variance as follows:

$$\text{Var}\,(\phi) = \mathbb{E}[\phi^2] - \mathbb{E}[\phi]^2 \leq \mathbb{E}[\phi^2] \leq \left(\frac{2}{\epsilon}\left(3C^2 + C^3\right) + C\right)^2.$$

Because the constant C is finite, the variance must also be finite, which completes our proof. By Lemma E.4, the confidence sequence in Theorem E.2 is a valid $(1-\alpha)$-AsympCS for τ. The proof for the approximate choice of ρ^* that minimizes the relative width of the confidence sequence at time T is provided in Waudby-Smith et al. [2024a, Appendix B.2]. □

E.5 Proof of Theorem 6.4 and Corollary 6.5

In semiparametric theory, the efficiency bound is determined by the variance of the EIF characterized in Eq. (6.2):

$$\begin{aligned} V_{\text{eff}}(\pi) &= \mathbb{E}[(\phi(X,Z,A,Y;\eta)-\tau)^2] \\ &= \mathbb{E}\Big[\mathbb{E}[(\phi(X,Z,A,Y;\eta)-\tau)^2 \mid X]\Big] \qquad \text{(Law of iterated expectations)} \end{aligned}$$

where

$$\begin{aligned} &\phi(X,Z,A,Y;\eta)-\tau \\ &= \underbrace{\frac{2Z-1}{Z\pi(X)+(1-Z)(1-\pi(X))}\frac{1}{\delta^A(X)}\Big[Y-A\delta(X)-\mu^Y(0,X)+\mu^A(0,X)\delta(X)\Big]}_{\Lambda_1} +\delta(X)-\tau \end{aligned}$$

First, we show that $\mathbb{E}[\Lambda_1 \mid X]=0$:

$$\begin{aligned} \mathbb{E}[\Lambda_1 \mid X] &= \pi(X)\mathbb{E}[\Lambda_1 \mid Z=1,X] + (1-\pi(X))\mathbb{E}[\Lambda_1 \mid Z=0,X] \\ &= \frac{\pi(X)}{\pi(X)}\frac{1}{\delta^A(X)}\Big(\mu^Y(1,X)-\mu^A(1,X)\delta(X)-\mu^Y(0,X)+\mu^A(0,X)\delta(X)\Big) \\ &\quad + \frac{1-\pi(X)}{1-\pi(X)}\frac{1}{\delta^A(X)}\underbrace{\Big(\mu^Y(0,X)-\mu^A(0,X)\delta(X)-\mu^Y(0,X)+\mu^A(0,X)\delta(X)\Big)}_{=0} \\ &\qquad\qquad \text{(Using the } \mu^A,\mu^Y \text{ definitions)} \\ &= \frac{1}{\delta^A(X)}(\mu^Y(1,X)-\mu^A(1,X)\delta(X)-\mu^Y(0,X)+\mu^A(0,X)\delta(X)) \\ &= 0 \end{aligned}$$

where in the last line we used the fact that $\delta(X)=\frac{\mu^Y(1,X)-\mu^Y(0,X)}{\mu^A(1,X)-\mu^A(0,X)}$ which implies the identity $\mu^Y(1,X)-\mu^A(1,X)\delta(X)=\mu^Y(0,X)-\mu^A(0,X)\delta(X)$.

Thus, we can expand $V_{\text{eff}}(\pi)$ as:

$$\begin{aligned} V_{\text{eff}}(\pi) &= \mathbb{E}\Big[\mathbb{E}[(\Lambda_1+\delta(X)-\tau)^2 \mid X]\Big] \\ &= \mathbb{E}\Big[\mathbb{E}[\Lambda_1^2 \mid X]+(\delta(X)-\tau)^2\Big] - 2\mathbb{E}\left[(\delta(X)-\tau)\mathbb{E}[\Lambda_1 \mid X]\right] \\ &= \mathbb{E}\Big[\mathbb{E}[\Lambda_1^2 \mid X]+(\delta(X)-\tau)^2\Big] \qquad \text{(Using } \mathbb{E}[\Lambda_1 \mid X]\text{=0)} \end{aligned}$$

It now remains to expand $\mathbb{E}[\Lambda_1^2 \mid X]$:

$$\begin{aligned}
\mathbb{E}[\Lambda_1^2 \mid X] &= \pi(X)\mathbb{E}[\Lambda_1^2 \mid Z=1,X] + (1-\pi(X))\mathbb{E}[\Lambda_1^2 \mid Z=0,X] \\
&= \frac{\pi(X)}{\pi(X)^2}\frac{1}{\delta^A(X)^2}\mathbb{E}\left[(Y - A\delta(X) - \mu^Y(0,X) + \mu^A(0,X)\delta(X))^2 \mid Z=1,X\right] \\
&\quad + \frac{1-\pi(X)}{(1-\pi(X))^2}\frac{1}{\delta^A(X)^2}\mathbb{E}\left[(Y - A\delta(X) - \mu^Y(0,X) + \mu^A(0,X)\delta(X))^2 \mid Z=0,X\right] \\
&= \frac{1}{\delta^A(X)^2}\Bigg(\frac{1}{\pi(X)}\operatorname{Var}(Y - A\delta(X) \mid Z=1,X) \\
&\qquad\qquad + \frac{1}{1-\pi(X)}\operatorname{Var}(Y - A\delta(X) \mid Z=0,X)\Bigg) \\
&= \frac{1}{\delta^A(X)^2}\left(\frac{\sigma^2(1,X)}{\pi(X)} + \frac{\sigma^2(0,X)}{1-\pi(X)}\right)
\end{aligned}$$

where we used the fact that $\mathbb{E}[Y - A\delta(X) \mid Z=0,X] = \mu^Y(0,X) - \mu^A(0,X)\delta(X)$ and $\mathbb{E}[Y - A\delta(X) \mid Z=1,X] = \mu^Y(1,X) - \mu^A(1,X)\delta(X) = \mu^Y(0,X) - \mu^A(0,X)\delta(X)$. Putting everything together, we obtain the result of Theorem 6.4:

$$V_{\text{eff}}(\pi) := \mathbb{E}\left[\frac{1}{\delta^A(X)^2}\left(\frac{\sigma^2(1,X)}{\pi(X)} + \frac{\sigma^2(0,X)}{1-\pi(X)}\right) + (\delta(X) - \tau)^2\right].$$

Then, the optimal policy $\pi^*(X)$ is given by:

$$\begin{aligned}
\pi^* &= \arg\min_\pi V_{\text{eff}}(\pi) \\
&= \arg\min_\pi \mathbb{E}\left[\frac{1}{\delta^A(X)^2}\left(\frac{\sigma^2(1,X)}{\pi(X)} + \frac{\sigma^2(0,X)}{1-\pi(X)}\right)\right] \\
&= \arg\min_\pi \mathbb{E}\left[\frac{1}{\delta^A(X)^2}\mathbb{E}\left[\left(\frac{\sigma^2(1,X)}{\pi(X)} + \frac{\sigma^2(0,X)}{1-\pi(X)}\right)\Big| X\right]\right] \\
\Rightarrow \pi^*(X) &= \arg\min_p \left(\frac{\sigma^2(1,X)}{p} + \frac{\sigma^2(0,X)}{1-p}\right)
\end{aligned}$$

The minimum is obtained when the derivative of the argument w.r.t. p is 0, *i.e.* $\frac{\sigma^2(0,X)}{(1-p)^2} - \frac{\sigma^2(1,X)}{p^2} = 0$. By solving for p, we obtain $p = \frac{\sqrt{\sigma^2(1,X)}}{\sqrt{\sigma^2(1,X)} + \sqrt{\sigma^2(0,X)}}$. Thus, we obtain the result of Corollary 6.5:

$$\pi^*(X) = \frac{\sqrt{\sigma^2(1,X)}}{\sqrt{\sigma^2(1,X)} + \sqrt{\sigma^2(0,X)}}.$$

E.6 Proof of Theorem 6.8

E.6.1 Preliminaries

Our asymptotic argument relies on a martingale central limit theorem under a Lindeberg-type condition. We use a streamlined version of the MDS central limit theorem originally due to Dvoretzky [1972], as presented in Zhang et al. [2021, Theorem 2].

Theorem E.5 (Martingale CLT, adapted from [Zhang et al., 2021, Thm. 2])**.** *Let* $\{(z_t, \mathcal{H}_t)\}_{t=1}^T$ *be a real-valued sequence where* $\bar{z}_T = \frac{1}{T}\sum_{t=1}^T z_t$ *such that:*

1. *(Martingale difference sequence)* $\{z_t\}_{t=1}^T$ *is a martingale difference sequence; that is,* $\mathbb{E}[z_t \mid \mathcal{H}_{t-1}] = 0$ *for every* $t \in [1, T]$.
2. *(Conditional variance convergence) There exists a constant* $\sigma^2 > 0$ *such that*
$$\frac{1}{T}\sum_{t=1}^T \mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}] \xrightarrow{p} \sigma^2,$$
3. *(Lindeberg condition) For every* $\epsilon > 0$,
$$\frac{1}{T}\sum_{t=1}^T \mathbb{E}\left[z_t^2 \mathbb{I}\{|z_t| > \epsilon\sqrt{T}\} \mid \mathcal{H}_{t-1}\right] \xrightarrow{p} 0.$$

Then,

$$\sqrt{T}\,\bar{z}_T \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

To begin our proof, we define $\psi_t := \phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) - \tau$, where ϕ is given in Eq. (6.9). Letting $\eta = \{\mu^Y(0, X), \mu^A(0, X), \delta^A(X), \delta(X)\}$ denote the true nuisance values, we decompose ψ_t as

$$\psi_t = \underbrace{\phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta) - \tau}_{z_t} + \underbrace{\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) - \phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta)}_{m_t},$$

such that

$$\sqrt{T}(\widehat{\tau}_T^{\text{AMRIV}} - \tau) = \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^T z_t\right) + \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^T m_t\right).$$

Then, the proof of Theorem 6.8 proceeds in three steps. We first verify that $\{z_t\}_{t=1}^T$ forms a martingale difference sequence. Then, we show that $\{z_t\}_{t=1}^T$ satisfies conditions (2)-(3) of Theorem E.5 with $\sigma^2 = V_{\text{eff}}(\pi)$. Finally, we will show that $\sqrt{T}(\frac{1}{T}\sum_{t=1}^T m_t) = o_p(1)$, thus concluding that

$$\sqrt{T}(\widehat{\tau}_T^{\text{AMRIV}} - \tau) \xrightarrow{d} \mathcal{N}(0, V_{\text{eff}}(\pi)).$$

E.6.2 MDS structure of z_t

We now show that $\{z_t\} = \{\phi_t(X_t, Z_t, A_t, Y_t; \pi_t, \eta) - \tau\}$ forms an MDS, *i.e.*$\mathbb{E}[z_t \mid \mathcal{H}_{t-1}] = 0$:

$$\begin{aligned}
&\mathbb{E}[z_t \mid \mathcal{H}_{t-1}] \\
&= \mathbb{E}\Big[\frac{2Z_t - 1}{Z_t\pi_t(X_t \mid \mathcal{H}_{t-1}) + (1 - Z_t)(1 - \pi_t(X_t \mid \mathcal{H}_{t-1}))}\frac{1}{\delta^A(X_t)} \\
&\qquad \cdot \Big[Y_t - A_t\delta(X_t) - \mu^Y(0, X_t) + \mu^A(0, X_t)\delta(X_t)\Big] + \delta(X_t)\Big|\mathcal{H}_{t-1}\Big] - \tau \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\frac{2Z_t - 1}{Z_t\pi_t(X_t \mid \mathcal{H}_{t-1}) + (1 - Z_t)(1 - \pi_t(X_t \mid \mathcal{H}_{t-1}))}\frac{1}{\delta^A(X_t)} \\
&\qquad \cdot \Big[Y_t - A_t\delta(X_t) - \mu^Y(0, X_t) + \mu^A(0, X_t)\delta(X_t)\Big]\Big|X_t, \mathcal{H}_{t-1}\Big]\Big|\mathcal{H}_{t-1}\Big] + \tau - \tau \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{From Eq. (6.1)} \\
&= \mathbb{E}\Big[\frac{1}{\delta^A(X_t)}\Big[\mu^Y(1, X_t) - \mu^A(1, X_t)\delta(X_t) - \mu^Y(0, X_t) + \mu^A(0, X_t)\delta(X_t)\Big] \\
&\qquad - \frac{1}{\delta^A(X_t)}\Big[\mu^Y(0, X_t) - \mu^A(0, X_t)\delta(X_t) - \mu^Y(0, X_t) + \mu^A(0, X_t)\delta(X_t)\Big]\Big|\mathcal{H}_{t-1}\Big] \\
&= 0
\end{aligned}$$

where we used the identity $\mu^Y(1, X_t) - \mu^A(1, X_t)\delta(X_t) = \mu^Y(0, X_t) - \mu^A(0, X_t)\delta(X_t)$ which follows from the definition of $\delta(X_t)$. Thus, $\{z_t\}$ is an MDS, owing to the fact that π_t is constructed from historical data only.

E.6.3 z_t satisfies conditions (2)–(3) of Theorem E.5

For condition (2), we first show that $\mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}] - V_{\text{eff}}(\pi) \xrightarrow{p} 0$:

$$\mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}] - V_{\text{eff}}(\pi)$$

$$= \text{Var}(z_t \mid \mathcal{H}_{t-1}) - \mathbb{E}\left[\frac{1}{\delta^A(X_t)^2}\left(\frac{\sigma^2(1,X_t)}{\pi(X_t)} + \frac{\sigma^2(0,X_t)}{1-\pi(X_t)}\right) + (\delta(X_t)-\tau)^2\right]$$

$$(\mathbb{E}[z_t \mid \mathcal{H}_{t-1}] = 0)$$

$$= \text{Var}(\phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta) \mid \mathcal{H}_{t-1})$$
$$- \mathbb{E}\left[\frac{1}{\delta^A(X_t)^2}\left(\frac{\sigma^2(1,X_t)}{\pi(X_t)} + \frac{\sigma^2(0,X_t)}{1-\pi(X_t)}\right) + (\delta(X_t)-\tau)^2\right]$$
$$= \mathbb{E}\left[\frac{1}{\delta^A(X_t)^2}\left(\frac{\sigma^2(1,X_t)}{\pi_t(X_t \mid \mathcal{H}_{t-1})} + \frac{\sigma^2(0,X_t)}{1-\pi_t(X_t \mid \mathcal{H}_{t-1})}\right) + (\delta(X_t)-\tau)^2 \middle| \mathcal{H}_{t-1}\right]$$

(η is the oracle nuisance set)

$$- \mathbb{E}\left[\frac{1}{\delta^A(X_t)^2}\left(\frac{\sigma^2(1,X_t)}{\pi(X_t)} + \frac{\sigma^2(0,X_t)}{1-\pi(X_t)}\right) + (\delta(X_t)-\tau)^2\right]$$
$$= \mathbb{E}\left[\frac{\sigma^2(1,X_t)}{\delta^A(X_t)^2}\left(\frac{\pi(X_t) - \pi_t(X_t \mid \mathcal{H}_{t-1})}{\pi_t(X_t \mid \mathcal{H}_{t-1})\pi(X_t)}\right)\middle| \mathcal{H}_{t-1}\right]$$
$$+ \mathbb{E}\left[\frac{\sigma^2(0,X_t)}{\delta^A(X_t)^2}\left(\frac{\pi_t(X_t \mid \mathcal{H}_{t-1}) - \pi(X_t)}{(1-\pi_t(X_t \mid \mathcal{H}_{t-1}))(1-\pi(X_t))}\right)\middle| \mathcal{H}_{t-1}\right]$$
$$\leq \frac{36C^2\epsilon_{\delta^A}^2 k_t}{\epsilon} |\mathbb{E}[\pi(X_t) - \pi_t(X_t \mid \mathcal{H}_{t-1})]| \lesssim k_t \|\pi_t - \pi\|_2 = o_p(1)$$

where in the last line we use the following boundedness conditions: (i) $|Y| \leq C$ from Assumption 6.7 and thus $|\delta(X)| \leq 2C$ and $\sigma^2(z, X_t) = \mathbb{E}[(Y - A\delta(X_t))^2 \mid Z = z, X_t] - \mathbb{E}[Y - A\delta(X_t) \mid Z = z, X_t]^2 \leq 18C^2$, (ii) $|\delta^A(X)|^{-1} \leq \epsilon_{\delta^A}$ for some $\epsilon_{\delta^A} > 0$ implicit in the (conditional) relevance in Assumption 6.1, (iii) $\pi(X_t), 1 - \pi(X_t) > \epsilon$ from Theorem 6.8 statement, (iv) $\pi(X_t), 1 - \pi(X_t) \geq 1/k_t$ by construction, and (v) the L_1 norm is bounded by the L_2 norm. Thus, setting $\sigma^2 := V_{\text{eff}}$, we have that each term converges in probability to σ^2, *i.e.*$\mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}] \xrightarrow{p} \sigma^2$, where σ^2 is finite by Assumption 6.1 and Assumption 6.7.

To complete condition (2), we now show that

$$\left|\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}] - \sigma^2\right| \xrightarrow{p} 0.$$

Let $a_t := \mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}]$ and $a := \sigma^2$. We have just established that $a_t \xrightarrow{p} a$, and under our boundedness assumptions, $\sup_t \mathbb{E}[a_t] < \infty$, so the sequence $\{a_t\}$ is uniformly

integrable. By the L^1 convergence theorem (e.g., Loève [1977]), this implies that $a_t \to a$ in L^1, i.e., $\mathbb{E}\left[\left|\mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}] - \sigma^2\right|\right] \to 0$. Therefore, by Cesàro averaging and Markov's inequality, we obtain $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[z_t^2 \mid \mathcal{H}_{t-1}] \xrightarrow{p} \sigma^2$, completing the verification of condition (2) in Theorem E.5.

Now we verify that z_t satisfies condition (3), the Lindeberg condition. We follow the same steps as in Cook et al. [2024]. Let $b_t := z_t^2 \cdot \mathbb{I}\{|z_t| > \delta\sqrt{T}\}$. Then $b_t = z_t^2$ with probability $\Pr(|z_t| > \delta\sqrt{T})$, and $b_t = 0$ otherwise. By Chebyshev's inequality,

$$\Pr(|z_t| > \delta\sqrt{T}) \le \frac{\mathrm{Var}(z_t)}{\delta^2 T}.$$

Since $\mathrm{Var}(z_t) = \mathbb{E}[z_t^2] < \infty$, it follows that $\lim_{T\to\infty}\frac{\mathrm{Var}(z_t)}{\delta^2 T} = 0$, which implies $b_t \xrightarrow{p} 0$ and hence $b_t \xrightarrow{d} 0$. Moreover, note that $|b_t| \le z_t^2$ and $\mathbb{E}[z_t^2] < \infty$. By the dominated convergence theorem,

$$\lim_{T\to\infty}\mathbb{E}[b_t] = \mathbb{E}\left[\lim_{T\to\infty} b_t\right] = 0.$$

Therefore,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[z_t^2 \cdot \mathbb{I}\left\{|z_t| > \delta\sqrt{T}\right\} \mid \mathcal{H}_{t-1}\right] \xrightarrow{p} 0,$$

verifying the Lindeberg-type condition required for the martingale CLT.

E.6.4 $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} m_t\right)$ is $o_p(1)$

We first decompose $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} m_t\right)$ as:

$$\begin{aligned}
&\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} m_t\right) \\
&= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}(\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) - \phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta))\right) \\
&= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}(\mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}] - \mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta) \mid \mathcal{H}_{t-1}])\right) \quad (\Delta^A)
\end{aligned}$$

$$+ \sqrt{T}\Big(\frac{1}{T}\sum_{t=1}^{T}\Big\{(\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) - \phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta))$$

$$- \mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) - \phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta) \mid \mathcal{H}_{t-1}]\Big\}\Big) \quad (\Delta^B)$$

where Δ^A is an asymptotic bias term due to nuisance estimation and Δ^B is the empirical process term. We bound these independently. Let $\Delta_t^A = \mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}] - \mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta) \mid \mathcal{H}_{t-1}]$. Then:

$$\begin{aligned}
&\Delta_t^A \\
&= \mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}] - \mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta) \mid \mathcal{H}_{t-1}] \\
&= \mathbb{E}[\mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid X_t, \mathcal{H}_{t-1}] \mid \mathcal{H}_{t-1}] - \mathbb{E}[\delta(X_t) \mid \mathcal{H}_{t-1}] \\
&= \mathbb{E}[\mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid Z_t = 1, X_t, \mathcal{H}_{t-1}]\pi_t(X_t \mid \mathcal{H}_{t-1}) \mid \mathcal{H}_{t-1}] \\
&\quad + \mathbb{E}[\mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid Z_t = 0, X_t, \mathcal{H}_{t-1}](1 - \pi_t(X_t \mid \mathcal{H}_{t-1})) \mid \mathcal{H}_{t-1}] \\
&\quad - \mathbb{E}[\delta(X_t) \mid \mathcal{H}_{t-1}] \\
&= \mathbb{E}\Bigg[\frac{1}{\widehat{\delta}_{t-1}^A(X_t)}(\mu^Y(1, X_t) - \mu^A(1, X_t)\widehat{\delta}_{t-1}(X_t) + \widehat{\mu}_{t-1}^Y(0, X_t) - \widehat{\mu}_{t-1}^A(0, X_t)\widehat{\delta}_{t-1}(X_t)) \\
&\qquad - \frac{1}{\widehat{\delta}_{t-1}^A(X_t)}(\mu^Y(0, X_t) - \mu^A(0, X_t)\widehat{\delta}_{t-1}(X_t) + \widehat{\mu}_{t-1}^Y(0, X_t) - \widehat{\mu}_{t-1}^A(0, X_t)\widehat{\delta}_{t-1}(X_t)) \\
&\qquad + \widehat{\delta}_{t-1}(X_t) - \delta(X_t)\Bigg|\mathcal{H}_{t-1}\Bigg] \\
&= \mathbb{E}\left[\frac{1}{\widehat{\delta}_{t-1}^A(X_t)}(\delta^Y(X_t) - \delta^A(X_t)\widehat{\delta}_{t-1}(X_t)) + \widehat{\delta}_{t-1}(X_t) - \delta(X_t)\middle|\mathcal{H}_{t-1}\right] \\
&= \mathbb{E}\left[\frac{\delta^A(X_t)}{\widehat{\delta}_{t-1}^A(X_t)}(\delta(X_t) - \widehat{\delta}_{t-1}(X_t)) + \widehat{\delta}_{t-1}(X_t) - \delta(X_t)\middle|\mathcal{H}_{t-1}\right] \\
&= \mathbb{E}\left[\frac{1}{\widehat{\delta}_{t-1}^A(X_t)}(\delta^A(X_t) - \widehat{\delta}_{t-1}^A(X_t))(\delta(X_t) - \widehat{\delta}_{t-1}(X_t))\middle|\mathcal{H}_{t-1}\right] \\
&\leq C\|\widehat{\delta}_{t-1} - \delta\|_2\|\widehat{\delta}_{t-1}^A - \delta^A\|_2 \qquad \text{(Assumption 6.7 and Cauchy-Schwarz)} \\
&= o_p(t^{-1/2})
\end{aligned}$$

By an argument similar to that of the previous section, $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\Delta_t^A\right) = o_p(1)$.

We now focus on the empirical process term Δ^B. We will show that $\mathbb{E}[\Delta^B] = 0$

and $\text{Var}(\Delta^B) = o_p(1)$ and then apply Chebyshev's inequality to reach the desired conclusion. We now turn to the empirical process term Δ^B. Our goal is to show that $\mathbb{E}[\Delta^B] = 0$ and $\text{Var}(\Delta^B) = o_p(1)$, which together imply that Δ^B is $o_p(1)$ by Chebyshev's inequality.

Let $\phi_t(\widehat{\eta}_t) := \phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t)$ and $\phi_t(\eta) := \phi(X_t, Z_t, A_t, Y_t; \pi_t, \eta)$. We tackle the mean:

$$\begin{aligned}\mathbb{E}[\Delta^B] &= \mathbb{E}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\phi_t(\widehat{\eta}_t) - \phi_t(\eta) - \mathbb{E}[\phi_t(\widehat{\eta}_t) - \phi_t(\eta) \mid \mathcal{H}_{t-1}]\right)\right] \\ &= \frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left(\mathbb{E}[\phi_t(\widehat{\eta}_t) - \phi_t(\eta)] - \mathbb{E}[\phi_t(\widehat{\eta}_t) - \phi_t(\eta)]\right) = 0. \quad \text{(Iterated expectations)}\end{aligned}$$

Let us now bound $\text{Var}(\Delta^B)$. Since the summands in Δ^B are conditionally mean-zero and adapted to the filtration $\mathcal{H}_{t-1}$, the cross-terms vanish by martingale difference independence. Thus:

$$\begin{aligned}\text{Var}(\Delta^B) &= \text{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\phi_t(\widehat{\eta}_t) - \phi_t(\eta) - \mathbb{E}[\phi_t(\widehat{\eta}_t) - \phi_t(\eta) \mid \mathcal{H}_{t-1}]\right)\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\text{Var}\left(\phi_t(\widehat{\eta}_t) - \phi_t(\eta) - \mathbb{E}[\phi_t(\widehat{\eta}_t) - \phi_t(\eta) \mid \mathcal{H}_{t-1}]\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\text{Var}\left(\phi_t(\widehat{\eta}_t) - \phi_t(\eta) \mid \mathcal{H}_{t-1}\right)\right] \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}[(\phi_t(\widehat{\eta}_t) - \phi_t(\eta))^2 \mid \mathcal{H}_{t-1}]\right]\end{aligned}$$

where we used the inequality $\text{Var}(X - \mathbb{E}[X \mid \mathcal{F}]) \leq \mathbb{E}[X^2]$ for any square-integrable X. We now stochastically bound $\mathbb{E}[(\phi_t(\widehat{\eta}_t) - \phi_t(\eta))^2 \mid \mathcal{H}_{t-1}]$. First, we note:

$$\begin{aligned}&\phi_t(\widehat{\eta}_t) - \phi_t(\eta) \\ &= \frac{2Z-1}{Z\pi_t(X_t) + (1-Z)(1-\pi_t(X_t))}\cdot \\ &\quad\left\{Y_t\frac{\delta^A(X_t) - \widehat{\delta}^A_{t-1}(X_t)}{\delta^A(X_t)\widehat{\delta}^A_{t-1}(X_t)} - A_t\left(\frac{\widehat{\delta}_{t-1}(X_t)}{\widehat{\delta}^A_{t-1}(X_t)} - \frac{\delta(X_t)}{\delta^A(X_t)}\right) + \left(\frac{\widehat{\mu}^Y_{t-1}(0, X_t)}{\widehat{\delta}^A_{t-1}(X_t)} - \frac{\mu^Y(0, X_t)}{\delta^A(X_t)}\right)\right.\end{aligned}$$

$$+\left(\frac{\widehat{\mu}^A_{t-1}(0,X_t)\widehat{\delta}_{t-1}(0,X_t)}{\widehat{\delta}^A_{t-1}(X_t)}-\frac{\mu^A(0,X_t)\delta(0,X_t)}{\delta^A(X_t)}\right)+(\widehat{\delta}_{t-1}(X_t)-\delta(X_t))\Big\}.$$

Then:

$$\begin{aligned}
&\mathbb{E}[(\phi_t(\widehat{\eta}_t)-\phi_t(\eta))^2 \mid \mathcal{H}_{t-1}] = \mathbb{E}\left[\mathbb{E}\left[(\phi_t(\widehat{\eta}_t)-\phi_t(\eta))^2 \mid X_t, \mathcal{H}_{t-1}\right]\right] \\
&\leq 2\mathbb{E}\left[\frac{1}{\pi_t(X_t)}\mathbb{E}[Y_t^2 \mid Z_t=1, X_t]\frac{(\delta^A(X_t)-\widehat{\delta}^A_{t-1}(X_t))^2}{(\delta^A(X_t)\widehat{\delta}^A_{t-1}(X_t))^2}\middle|\mathcal{H}_{t-1}\right] && \text{(a)} \\
&\quad + 2\mathbb{E}\left[\frac{1}{\pi_t(X_t)}\mathbb{E}[A_t^2 \mid Z_t=1, X_t]\left(\frac{\widehat{\delta}_{t-1}(X_t)}{\widehat{\delta}^A_{t-1}(X_t)}-\frac{\delta(X_t)}{\delta^A(X_t)}\right)^2\middle|\mathcal{H}_{t-1}\right] && \text{(b)} \\
&\quad + 2\mathbb{E}\left[\frac{1}{\pi_t(X_t)}\left(\frac{\widehat{\mu}^Y_{t-1}(0,X_t)}{\widehat{\delta}^A_{t-1}(X_t)}-\frac{\mu^Y(0,X_t)}{\delta^A(X_t)}\right)^2\middle|\mathcal{H}_{t-1}\right] && \text{(c)} \\
&\quad + 2\mathbb{E}\left[\frac{1}{\pi_t(X_t)}\left(\frac{\widehat{\mu}^A_{t-1}(0,X_t)\widehat{\delta}_{t-1}(0,X_t)}{\widehat{\delta}^A_{t-1}(X_t)}-\frac{\mu^A_{t-1}(0,X_t)\delta_{t-1}(0,X_t)}{\delta^A(X_t)}\right)^2\middle|\mathcal{H}_{t-1}\right] && \text{(d)} \\
&\quad + 2\mathbb{E}\left[(\widehat{\delta}_{t-1}(X_t)-\delta(X_t))^2\middle|\mathcal{H}_{t-1}\right] && \text{(e)} \\
&\quad - 2\mathbb{E}\left[\frac{1}{1-\pi_t(X_t)}\mathbb{E}[Y_t^2 \mid Z_t=0, X_t]\frac{(\delta^A(X_t)-\widehat{\delta}^A_{t-1}(X_t))^2}{(\delta^A(X_t)\widehat{\delta}^A_{t-1}(X_t))^2}\middle|\mathcal{H}_{t-1}\right] && \text{(a)} \\
&\quad - 2\mathbb{E}\left[\frac{1}{1-\pi_t(X_t)}\mathbb{E}[A_t^2 \mid Z_t=1, X_t]\left(\frac{\widehat{\delta}_{t-1}(X_t)}{\widehat{\delta}^A_{t-1}(X_t)}-\frac{\delta(X_t)}{\delta^A(X_t)}\right)^2\middle|\mathcal{H}_{t-1}\right] && \text{(b)} \\
&\quad - 2\mathbb{E}\left[\frac{1}{1-\pi_t(X_t)}\left(\frac{\widehat{\mu}^Y_{t-1}(0,X_t)}{\widehat{\delta}^A_{t-1}(X_t)}-\frac{\mu^Y(0,X_t)}{\delta^A(X_t)}\right)^2\middle|\mathcal{H}_{t-1}\right] && \text{(c)} \\
&\quad - 2\mathbb{E}\left[\frac{1}{1-\pi_t(X_t)}\left(\frac{\widehat{\mu}^A_{t-1}(0,X_t)\widehat{\delta}_{t-1}(0,X_t)}{\widehat{\delta}^A_{t-1}(X_t)}-\frac{\mu^A_{t-1}(0,X_t)\delta_{t-1}(0,X_t)}{\delta^A(X_t)}\right)^2\middle|\mathcal{H}_{t-1}\right] && \text{(d)} \\
&\quad - 2\mathbb{E}\left[(\widehat{\delta}_{t-1}(X_t)-\delta(X_t))^2\middle|\mathcal{H}_{t-1}\right] && \text{(e)}
\end{aligned}$$

Bounding all terms using Assumption 6.1 and Assumption 6.7, we have:

$$\begin{aligned}
&\mathbb{E}[(\phi_t(\widehat{\eta}_t)-\phi_t(\eta))^2 \mid \mathcal{H}_{t-1}] \\
&\leq 4k_tC^4\epsilon^2_{\delta^A}\|\widehat{\delta}^A_{t-1}-\delta^A\|_2^2 && \text{(a)} \\
&\quad + 8k_tC^2\epsilon^2_{\delta^A}(\|\widehat{\delta}_{t-1}-\delta\|_2^2 + 4C^2\|\widehat{\delta}^A_{t-1}-\delta^A\|_2^2) && \text{(b)} \\
&\quad + 8k_tC^2\epsilon^2_{\delta^A}(\|\widehat{\mu}^Y_{t-1}(0,\cdot)-\mu^Y(0,\cdot)\|_2^2 + C^2\|\widehat{\delta}^A_{t-1}-\delta^A\|_2^2) && \text{(c)} \\
&\quad + 8k_tC^2\epsilon^2_{\delta^A}(4C^4\|\widehat{\mu}^A_{t-1}(0,\cdot)-\mu^A(0,\cdot)\|_2^2 + 4C^4\|\widehat{\delta}^A_{t-1}-\delta^A\|_2^2 + C^2\|\widehat{\delta}_{t-1}-\delta\|_2^2) && \text{(d)}
\end{aligned}$$

$$+ \|\widehat{\delta}_{t-1}(X_t) - \delta(X_t)\|_2^2 \tag{e}$$
$$= o_p(1)$$

where the last line follows because terms (a)–(e) are $o_p(1)$ by the premise of Theorem 6.8. Since each term $\mathbb{E}\Big[(\phi_t(\widehat{\eta}_t) - \phi_t(\eta))^2 \mid \mathcal{H}_{t-1}\Big]$ is nonnegative, uniformly bounded, and $o_p(1)$, Cesàro averaging implies that

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[(\phi_t(\widehat{\eta}_t) - \phi_t(\eta))^2 \mid \mathcal{H}_{t-1}\Big] = o_p(1).$$

Thus, $\mathrm{Var}(\Delta^B) = o_p(1)$, and Chebyshev's inequality gives

$$\Pr\Big(|\Delta^B| \geq \varepsilon\Big) \leq \frac{\mathrm{Var}(\Delta^B)}{\varepsilon^2}, \qquad \forall \varepsilon > 0.$$

Therefore, $\Delta^B = o_p(1)$, as desired. Putting everything together, we conclude that

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} m_t\right) = o_p(1),$$

and the conclusion of Theorem 6.8 follows.

E.7 Proof of Theorem 6.9 and Corollary 6.10

Letting $\phi_t(\pi, \eta) := \phi(X_t, Z_t, A_t, Y_t; \pi, \eta)$ for any π, η and , we decompose $\widehat{\tau}_T^{\mathrm{AMRIV}} - \tau$ as follows:

$$\begin{aligned}
\widehat{\tau}_T^{\mathrm{AMRIV}} - \tau &= \frac{1}{T}\sum_{t=1}^{T} \phi_t(\pi_t, \widehat{\eta}_t) - \tau \\
&= \underbrace{\frac{1}{T}\sum_{t=1}^{T} \left(\phi_t(\pi_t, \widehat{\eta}_t) - \mathbb{E}[\phi_t(\pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}]\right)}_{\Delta^A} + \underbrace{\frac{1}{T}\sum_{t=1}^{T} \left(\mathbb{E}[\phi_t(\pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}] - \tau\right)}_{\Delta^B}
\end{aligned}$$

We will now show that Δ^A is $O_p(T^{-1/2})$ via a similar argument as in Appendix E.6 and Δ^B is $O_p\left(\|\widehat{\delta}_T^A - \delta^A\|_2 \|\widehat{\delta}_T - \delta\|_2\right)$.

Write $\Delta_t^A := \phi_t(\pi_t, \widehat{\eta}_t) - \mathbb{E}[\phi_t(\pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}]$ and note that Δ_t^A is an MDS by construction, *i.e.*$\mathbb{E}[\Delta_t^A \mid \mathcal{H}_{t-1}] = 0$. Let $\widetilde{V}(\pi) := \mathrm{Var}_\pi(\phi_t(\pi, \tilde{\eta}))$ where Var_π indicates

the variance over data where $Z \sim \text{Bern}(\pi(X_t))$. Thus, it suffices to show that $\mathbb{E}[(\Delta_t^A)^2 \mid \mathcal{H}_{t-1}] \xrightarrow{p} \sigma^2$, where $\sigma^2 = \widetilde{V}(\pi)$. Then, the result follows by tracing the rest of the proof in Appendix E.6.

Write $\Lambda_t := \phi_t(\pi_t, \widehat{\eta}_t) - \phi_t(\pi_t, \widetilde{\eta})$ and note

$$\text{Var}(\phi_t(\pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}) = \text{Var}(\phi_t(\pi_t, \tilde{\eta}) \mid \mathcal{H}_{t-1}) + \text{Var}(\Lambda_t \mid \mathcal{H}_{t-1}) + 2\ \text{Cov}(\phi_t(\tilde{\eta}), \Delta_t \mid \mathcal{H}_{t-1})$$

and thus

$$\begin{aligned}
&\left|\text{Var}(\phi_t(\pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}) - \text{Var}(\phi_t(\pi_t, \tilde{\eta}) \mid \mathcal{H}_{t-1})\right| \\
&\leq \text{Var}(\Lambda_t \mid \mathcal{H}_{t-1}) + 2\sqrt{\text{Var}(\Lambda_t \mid \mathcal{H}_{t-1})\,\text{Var}(\phi_t(\pi_t, \widetilde{\eta}) \mid \mathcal{H}_{t-1})}
\end{aligned}$$

Since, $\text{Var}(\phi_t(\pi_t, \widetilde{\eta}) \mid \mathcal{H}_{t-1})$ is bounded by Assumption 6.7, we just need to show that $\text{Var}(\Lambda_t \mid \mathcal{H}_{t-1}) = o_p(1)$:

$$\begin{aligned}
&\text{Var}(\phi_t(\pi_t, \widetilde{\eta}) \mid \mathcal{H}_{t-1}) \leq \mathbb{E}\left[(\phi_t(\pi_t, \widehat{\eta}_t) - \phi_t(\pi_t, \widetilde{\eta}))^2 \mid \mathcal{H}_{t-1}\right] \\
&\leq \tilde{C} k_t(\|\widehat{\delta}_{t-1} - \tilde{\delta}\|_2^2 + \|\widehat{\mu}_{t-1}^Y(0,\cdot) - \tilde{\mu}^Y(0,\cdot)\|_2^2 + \|\widehat{\mu}_{t-1}^A(0,\cdot) - \tilde{\mu}^A(0,\cdot)\|_2^2 + \|\widehat{\delta}_{t-1}^A - \tilde{\delta}^A\|_2^2) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Parallelogram law)} \\
&= o_p(1) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Theorem assumptions)}
\end{aligned}$$

where $\tilde{C}$ encompasses the constants ϵ and C from Assumption 6.7 and the theorem's premise. Thus, since $\text{Var}(\phi_t(\pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}) \xrightarrow{p} \text{Var}(\phi_t(\pi_t, \widetilde{\eta}) \mid \mathcal{H}_{t-1}) \xrightarrow{p} \widetilde{V}_\pi$, we can use Theorem E.5 from Appendix E.6 and retrace the same arguments to obtain $\Delta^A = O_p(T^{-1/2})$ due to the Martingale CLT.

Now, we study Δ^B:

$$\begin{aligned}
&\Delta_t^B \\
&= \mathbb{E}[\phi_t(\pi_t, \widehat{\eta}_t) \mid \mathcal{H}_{t-1}] - \mathbb{E}[\delta(X)] \\
&= \mathbb{E}[\mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid X_t, \mathcal{H}_{t-1}] \mid \mathcal{H}_{t-1}] - \mathbb{E}[\delta(X_t) \mid \mathcal{H}_{t-1}] \\
&= \mathbb{E}[\mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid Z_t = 1, X_t, \mathcal{H}_{t-1}]\pi_t(X_t \mid \mathcal{H}_{t-1}) \mid \mathcal{H}_{t-1}]
\end{aligned}$$

$$
\begin{aligned}
&+ \mathbb{E}[\mathbb{E}[\phi(X_t, Z_t, A_t, Y_t; \pi_t, \widehat{\eta}_t) \mid Z_t = 0, X_t, \mathcal{H}_{t-1}](1 - \pi_t(X_t \mid \mathcal{H}_{t-1})) \mid \mathcal{H}_{t-1}] \\
&- \mathbb{E}[\delta(X_t) \mid \mathcal{H}_{t-1}] \\
&= \mathbb{E}\Bigg[\frac{1}{\widehat{\delta}^A_{t-1}(X_t)}(\mu^Y(1, X_t) - \mu^A(1, X_t)\widehat{\delta}_{t-1}(X_t) + \widehat{\mu}^Y_{t-1}(0, X_t) - \widehat{\mu}^A_{t-1}(0, X_t)\widehat{\delta}_{t-1}(X_t)) \\
&\qquad - \frac{1}{\widehat{\delta}^A_{t-1}(X_t)}(\mu^Y(0, X_t) - \mu^A(0, X_t)\widehat{\delta}_{t-1}(X_t) + \widehat{\mu}^Y_{t-1}(0, X_t) - \widehat{\mu}^A_{t-1}(0, X_t)\widehat{\delta}_{t-1}(X_t)) \\
&\qquad + \widehat{\delta}_{t-1}(X_t) - \delta_{t-1}(X_t)\Big|\mathcal{H}_{t-1}\Bigg] \\
&= \mathbb{E}\left[\frac{1}{\widehat{\delta}^A_{t-1}(X_t)}(\delta^Y(X_t) - \delta^A(X_t)\widehat{\delta}_{t-1}(X_t)) + \widehat{\delta}_{t-1}(X_t) - \delta_{t-1}(X_t)\Big|\mathcal{H}_{t-1}\right] \\
&= \mathbb{E}\left[\frac{\delta^A(X_t)}{\widehat{\delta}^A_{t-1}(X_t)}(\delta(X_t) - \widehat{\delta}_{t-1}(X_t)) + \widehat{\delta}_{t-1}(X_t) - \delta_{t-1}(X_t)\Big|\mathcal{H}_{t-1}\right] \\
&= \mathbb{E}\left[\frac{1}{\widehat{\delta}^A_{t-1}(X_t)}(\delta^A(X_t) - \widehat{\delta}^A_{t-1}(X_t))(\delta(X_t) - \widehat{\delta}_{t-1}(X_t))\Big|\mathcal{H}_{t-1}\right] \\
&\leq C\|\widehat{\delta}_{t-1} - \delta\|_2\|\widehat{\delta}^A_{t-1} - \delta^A\|_2 \qquad \text{(Assumption 6.7 and Cauchy-Schwarz)}
\end{aligned}
$$

Thus, Δ^B_t is $O_p(\|\widehat{\delta}_T - \delta\|_2\|\widehat{\delta}^A_T - \delta^A\|_2)$, as desired. Corollary 6.10 follows immediately by noting that if either $\widehat{\delta}_{t-1}$ or $\widehat{\delta}^A_{t-1}$ are consistent(*i.e.*$o_p(1)$), then Δ^B is also consistent and $|\tau^{\text{AMRIV}}_T - \tau| = o_p(1)$.

E.8 Experimental Details

This appendix provides additional details for the simulation experiments described in Section 6.7, including exact hyperparameters, model components, and execution setup. All experiments were run on a Perlmutter compute node with 256 CPU cores at the National Energy Research Scientific Computing Center (NERSC) [National Energy Research Scientific Computing Center, 2025] and required approximately 40–50 minutes per configuration. Random Forests were implemented using `scikit-learn` [Pedregosa et al., 2011], and parallelization was handled via `joblib`. Full code for generating data, running experiments,

and reproducing all figures is available at `https://github.com/CausalML/Adaptive-IV`, with instructions in the `README.md`.

Each estimator was evaluated on 1000 independent synthetic trials. Simulations were run over $T = 2000$ rounds with a $T_0 = 200$ burn-in period, and nuisance estimators were updated in mini-batches of 200. For all adaptive methods, we applied the truncated optimal allocation policy from Eq. (6.7), with a truncation schedule $k_t = 2/0.999^t$. Oracle methods used ground-truth nuisance functions, while misspecified estimators were constructed by replacing $\mu^Y(1, X)$ with a constant regressor fit to the average oracle value.

Unless otherwise stated, outcome and residual variance functions were modeled via Random Forests with 100 trees, maximum depth 5, and minimum leaf size 5. The compliance model $\mu^A(1, X)$ was learned with a shallower forest (depth 3, minimum leaf size 30), and $\mu^A(0, X)$ was zero by construction due to one-sided noncompliance. For the A2IPW estimator, we followed Kato et al. [2020] and estimated outcome means and second moments using random forests (depth 5, leaf size 100) and used a Neyman-style allocation based on observed outcomes. All figures report results averaged over replicates, with confidence intervals based on empirical standard errors.

E.8.1 Simulation Studies with Synthetic Data

We generate the data sequentially for each time $t \in [1, T + T_0]$ using the following one-sided noncompliance setup:

$$X_t \sim \text{Unif}(0, 2)^d, \quad Z_t \sim \text{Bern}(\pi_t(X_t \mid \mathcal{H}_{t-1})) \qquad \text{(Covariates \& Instrument)}$$

$$\mu^A(0, X_t) = 0, \quad \delta^A(X_t) = \mu^A(1, X_t) = \sigma(2X_t[1]) \qquad \text{(Compliance Scores)}$$

$$C_t \sim \text{Bern}(\delta^A(X_t)), \quad A_t = C_t \cdot Z_t \qquad \text{(Treatment Assignment)}$$

$$U_t = u(1 - C_t) \qquad \text{(Unobserved Confounder)}$$

$$Y_t = f(A_t, X_t) + U_t + \epsilon_{A_t}, \quad \epsilon_{A_t} \sim \text{Unif}[-g(A_t, X_t), g(A_t, X_t)] \quad \text{(Outcome Function)}$$

where T_0 is the burn-in period, C_t is the (unknown) compliance indicator, σ is the logistic sigmoid, Unif and Bern are the uniform and the Bernoulli distributions, respectively. We utilize the following instantiations for d, u, f, g:

$$d = 5$$

$$u = -2.0$$

$$f(A, X) = 1 + A + X[1] + 2a\,(X^\top \beta) + 0.75a\,X[1]^2$$

$$g(A, X) = \sqrt{3 \cdot (v_1 \cdot A + (v_0 \cdot X[1] + v_1) \cdot (1 - A))}, \quad v_0 = 4.0, v_1 = 0.25$$

where $X[1]$ denotes the first coordinate of the covariate vector $X \in \mathbb{R}^d$ and $\beta \in \mathbb{R}^d, \beta \sim \text{Unif}[-1, 1]^d$ is a parameter vector that is fixed over the 1000 simulations (we used a seed of 1 for reproducibility purposes.). $g(A, X)$ was chosen such that $\text{Var}(\epsilon_A \mid X) = v_1 \cdot A + (v_0 \cdot X[1] + v_1) \cdot (A - 1)$ where v_0, v_1 are constants.

E.8.2 Simulation Studies with Semi-Synthetic Data

To complement our synthetic evaluation, we conduct additional experiments using a semi-synthetic setting derived from a real-world dataset collected by TripAdvisor. The original data-generating process (DGP) was introduced by Syrgkanis et al. [2019] and is publicly available on GitHub. In the original A/B test, users were randomly assigned to one of two groups: group A (instrument $Z = 1$) was offered a simplified membership sign-up experience, while group B ($Z = 0$) saw the default interface. This encouragement increased the likelihood of signing up for a membership (treatment A), though actual uptake remained endogenous due to user-specific factors.

The covariates $X \in \mathbb{R}^{10}$ capture rich pre-treatment user behavior and demographics. These include: prior platform revenue, visit frequencies to dif-

ferent TripAdvisor sections (hotels, restaurants, experiences, flights, and vacation rentals) over a 28-day pre-experimental window, engagement through free channels (e.g., email), locale information, and operating system type. The binary treatment A indicates whether the user became a member during the experiment, while the outcome Y records the total number of days the user visited TripAdvisor during the study period.

We preserve the original covariate structure and instrument assignment mechanism, but modify the outcome model to introduce heteroskedasticity by adding log-normal noise whose variance depends on treatment status. This choice reflects the heavy-tailed nature of usage metrics in online platforms [Barabasi, 2005, Krishnan et al., 2018], and results in a more realistic and challenging estimation task. The full data-generating process is provided below ("*" indicates same as original).

TripAdvisor Data-Generating Process

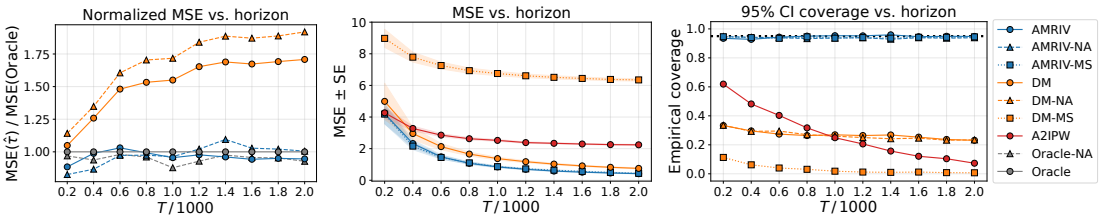
We simulate tuples $(X, A(0), A(1), Y(0), Y(1))$ via:

$$
\begin{aligned}
&X \sim \text{TripAdvisor pre-treatment covariates}, && (^*)\\
&\nu \sim \text{Unif}[-5, 5], && (\text{latent user heterogeneity}^*)\\
&A(1) \sim \text{Bernoulli}\left(0.8 \cdot \sigma(0.4X_1 + \nu)\right), \quad A(0) \sim \text{Bernoulli}(0.006), && (\text{compliance}^*)\\
&\varepsilon_1 \sim \text{LogNormal}(0, \sigma_1), \quad \varepsilon_0 \sim \text{LogNormal}(0, \sigma_0), && (\textbf{new:}\ \text{heavy-tailed errors})\\
&Y(1) = f(X) + 2\nu + 5 \cdot \mathbb{I}[X[1] > 0] + \varepsilon_1, && (\text{potential outcome for } A = 1^*)\\
&Y(0) = 2\nu + 5 \cdot \mathbb{I}[X[1] > 0] + \varepsilon_0, && (\text{potential outcome for } A = 0^*)
\end{aligned}
$$

where $\sigma_0 = 1.5$ and $\sigma_1 = 0.25$, and the structural CATE function is defined as:

$$f(X) = 0.8 + 0.5 \cdot \phi(X_1) - 3.0X[7],$$

with $\phi(X_1) := 5 \cdot \mathbb{I}[X[1] > 5] + 10 \cdot \mathbb{I}[X_1 > 15] + 5 \cdot \mathbb{I}[X_1 > 20]$.



(a) Efficiency (b) Consistency (c) Coverage

Figure E.1: Performance of AMRIV and baseline estimators on the TripAdvisor semi-synthetic experiment. **(a)** Efficiency: Normalized MSE versus an oracle benchmark. **(b)** Consistency: MSE ± standard error. **(c)** Coverage: Empirical coverage of nominal 95% confidence intervals.

We illustrate our results on Figure E.1. To maintain readability in the plots, we define the misspecified outcome model $\widehat{\mu}^Y(1, X)$ as the oracle $\mu^Y(1, X)$ plus a constant shift.

Adaptivity As shown in panel (a), adaptive allocation improves the efficiency of both the DM and AMRIV estimators (particularly at larger T), with AMRIV again approaching the oracle benchmark. Interestingly, AMRIV, AMRIV-NA, and Oracle-NA slightly outperform the fully adaptive Oracle in some regimes—likely due to extreme compliance scores in this DGP ($\delta^A(X) \to 0$), which inflate the asymptotic variance when using oracle denominators. In contrast, estimators with learned denominators can perform better in finite samples [Su et al., 2023, Kato et al., 2021]. This also helps explain the narrower variance gap between AMRIV and AMRIV-NA relative to the synthetic setting, as the variance is dominated by low-compliance regions rather than outcome variance between instrument arms.

Consistency Panel (b) confirms that AMRIV, AMRIV-NA, and DM all converge to the true τ, with AMRIV variants consistently achieving lower error. As expected, A2IPW fails to converge due to uncorrected confounding, while DM-MS diverges due to misspecification of $\delta(X)$. In contrast, AMRIV-MS remains

consistent—further validating the multiply-robust guarantee from Theorem 6.9.

Coverage Panel (c) shows that AMRIV, AMRIV-NA, and AMRIV-MS maintain valid 95% confidence interval coverage, consistent with the asymptotic normality result in Theorem 6.8. All other estimators—including DM-MS and A2IPW—under-cover severely as T increases, reflecting bias under misspecification or confounding.

E.9 Limitations and Broader Impacts

Limitations

While AMRIV is grounded in semiparametric theory and achieves strong empirical performance, there are several limitations we highlight. First, our method relies on standard IV identification assumptions (Assumption 6.1) and the unconfounded compliance assumption (Assumption 6.2), which—while weaker than ignorability—are still untestable and may be violated in practice. In particular, the exclusion restriction and the unconfounded compliance assumption may not hold even in observational settings where the instrument is randomized. Second, AMRIV assumes access to flexible, sequentially consistent nuisance estimators, which may be difficult to train or tune in low-data regimes or in the presence of heavy-tailed outcomes. Third, our analysis focuses on a binary instrument and binary treatment; extending the framework to multi-valued or continuous instruments remains an open challenge.

Broader Impacts

This work contributes to the growing intersection of causal inference and adaptive experimentation, enabling more data-efficient and statistically principled estimation in settings with noncompliance. Potential applications include

health interventions and online recommendation systems, where experimenters can encourage behavior but not enforce it. AMRIV allows experimenters to make better use of limited resources while supporting robust inference under endogenous treatment selection under unobserved confounding. However, we caution that the validity of conclusions drawn from AMRIV hinges on the identification assumptions and data quality. In high-stakes settings, particularly those involving marginalized or vulnerable populations, improper use or misinterpretation could lead to harmful decisions. We strongly recommend pairing AMRIV with domain expertise, sensitivity analysis, and uncertainty quantification to ensure responsible deployment and interpretation.

APPENDIX F

APPENDIX FOR CHAPTER 7

F.1 Extended Literature Review

Classical Spatiotemporal Causal Inference Early spatiotemporal causal inference methods–including spatial econometrics [Anselin, 2013], difference-in-differences [Keele and Titiunik, 2015], and synthetic controls [Ben-Michael et al., 2022]–provide useful frameworks for estimating treatment effects across regions but rely on strong assumptions such as parallel trends or stable treatment assignment. These approaches struggle with interference, nonlinear dependencies, and time-varying confounders, limiting their applicability in complex settings. More recent approaches for spatiotemporal causal inference handle time-varying confounding through inverse propensity weighting (IPW), typically by extending marginal structural models to the spatial or spatiotemporal domain. For instance, Papadogeorgou et al. [2022] and Zhou et al. [2024] employ IPW-style adjustments to estimate regional average treatment effects across space and time. However, these approaches cannot accommodate interference unless strong assumptions are made–e.g., defining a user-specified exposure mapping or restricting attention to hyper–local interactions (see also [Wang, 2021, Christiansen et al., 2022, Papadogeorgou et al., 2022, Zhang and Ning, 2023]). Such simplifications may be ill-suited for real-world systems with rich spatial dependencies. Moreover, even recent advances in this space remain limited; as noted by Zhou et al. [2024], the literature on spatiotemporal causal inference remains sparse, especially in settings with feedback loops or time-varying confounding.

Machine Learning for Spatiotemporal Modeling Spatiotemporal predictive modeling has seen rapid progress with the rise of deep learning. Convolu-

Table F.1: Comparison of prior neural G-computation methods and GST-UNet for spatiotemporal causal inference.

Aspect	**Prior Neural G-Computation**	**GST-UNet (ours)**
Data structure	Many independent temporal trajectories (e.g., patient sequences), with no inter-unit interactions such as spatial dependence.	A single spatiotemporal chain in which outcomes, covariates, and treatments evolve jointly across a lattice, with strong spatial coupling and interference.
Encoder	RNN- or Transformer-based encoder over time only.	A ConvLSTM–U-Net encoder that aggregates neighboring covariates and treatments before the G-heads, capturing interference and spatial confounding.
Training	Standard end-to-end training for i.i.d. trajectories; stability comes from large datasets rather than curriculum learning or spatial priors.	Curriculum-stabilized multi-head training for accurate pseudo-outcome generation under limited samples.
Theory	Classical G-formula under i.i.d. trajectories, with no single-chain guarantees.	Identification (Theorem 7.3) and consistency (Theorem F.2) under representation-based time-invariance for a single chain.

tional and recurrent neural networks are widely used for forecasting spatially indexed time series (e.g., weather or traffic) [Shi et al., 2015, Zhang et al., 2017], while graph-based methods (e.g., Graph WaveNet [Wu et al., 2019], Diffusion Convolutional RNN [Li et al., 2018]) capture non-Euclidean spatial dependencies. Vision transformer variants, including Video Swin Transformers [Liu et al., 2022] and TimeSformer [Bertasius et al., 2021], extend attention-based models to spatiotemporal video data. These architectures can learn complex non-local interactions over space and time. However, such models are typically optimized for prediction tasks and do not include causal adjustments. Without mechanisms like propensity modeling or G-computation, they remain ill-equipped

to estimate counterfactual outcomes or adjust for time-varying confounding. Some recent work integrates spatial representations for causal inference–e.g., Tec et al. [2023] incorporate non-local confounders using a UNet-based model–but these methods do not explicitly model dependencies over time or adjust for time-varying confounders.

Time-Series Causal Inference In the longitudinal domain, time-series causal inference has developed tools for handling temporal confounding using models such as marginal structural models [Robins et al., 2000], IPW-style estimation [Lim, 2018], and iterative G-computation [Robins and Hernan, 2008]. Recent ML-based extensions include recurrent networks [Bica et al., 2020b, Li et al., 2021, Seedat et al., 2022], Transformers [Melnychuk et al., 2022, Hess et al., 2024] and meta-learners [Frauen et al., 2025]. However, all these methods assume access to independent time series–e.g., across units or patients–which allows for pooling across trajectories. These methods do not consider spatial dependencies, interference, or scenarios with a single observed spatiotemporal realization. As such, while they may handle time-varying confounding, they do not generalize to our setting. Table F.1 summarizes the key methodological differences between GST-UNet and prior neural G-computation frameworks.

Neural-Based Spatiotemporal Causal Inference There has been limited work on neural models that explicitly address spatiotemporal causal inference. Tec et al. [2023] use a U-Net backbone to learn spatial representations for causal inference in air pollution studies but do not address time-varying confounding or feedback loops. Ali et al. [2024] present a U-Net–based architecture for predicting direct and indirect effects in climate contexts, but primarily focus on forecasting rather than causal identification. While these works highlight growing interest in neural approaches to causal inference in spatiotemporal do-

mains, none incorporate an iterative adjustment procedure like G-computation that handles time-varying confounders, leaving identification in these settings largely unaddressed.

Our Contribution The GST-UNet bridges these gaps by combining flexible spatiotemporal neural architectures with a theoretically grounded iterative G-computation framework. This allows valid estimation of potential outcomes in the presence of interference, spatial confounding, and time-varying confounding–without requiring practitioners to specify structural models or exposure mappings. To our knowledge, this is the first end-to-end framework to implement G-computation for causal inference over a single spatiotemporal trajectory. We integrate spatiotemporal processing via U-Nets and ConvLSTMs with a principled multi-head neural causal module, and we design a curriculum-based training strategy to stabilize learning of recursive pseudo-outcomes. Together, these components yield a ready-to-use tool for practitioners, with consistent identification guarantees and robust empirical performance. By abstracting away the modeling choices typically required in structural spatiotemporal methods, GST-UNet makes spatiotemporal causal estimation more accessible, interpretable, and reliable for real-world applications.

F.2 Proof of Theorem 7.3

We aim to show that under Assumption 7.1 and Assumption 7.2, the CAPOs in Eq. (7.1) can be identified recursively from a single time series via a sequence of conditional expectations.

Step 1: Recursive Decomposition for the Intractable Expectation We first demonstrate the recursive decomposition of the intractable expectation in the CAPO definition (Eq. (7.1)). While this expectation is theoretically well-defined,

it cannot be directly estimated in practice due to the limited availability of data. Specifically, we only observe a single time series, meaning we have just one sample of the history at time $t + \tau$ for each t. Nevertheless, as we will show, we can convert these expectations into expectations over prefix-based segments that allow us to estimate these quantities from the data.

Starting from $\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}]$, we have:

$$
\begin{aligned}
&\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}] \\
&= \mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_t = \mathbf{a}_t] \\
&\qquad\qquad \text{(Sequential ignorability and positivity (Assumption 7.1))} \\
&= \mathbb{E}[\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}^{\mathbf{a}}_{1:t+1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_t = \mathbf{a}_t] \qquad \text{(Law of total probability)} \\
&= \mathbb{E}[\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}^{\mathbf{a}}_{1:t+1}, \mathbf{A}_{t+1} = \mathbf{a}_{t+1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_t = \mathbf{a}_t] \\
&\qquad\qquad \text{(Sequential ignorability and positivity)} \\
&= \mathbb{E}\Big[\mathbb{E}[\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}^{\mathbf{a}}_{1:t+2}] \,\Big|\, \mathbf{H}^{\mathbf{a}}_{1:t+1}, \mathbf{A}_{t+1} = \mathbf{a}_{t+1}] \,\Big|\, \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_t = \mathbf{a}_t\Big] \\
&\qquad\qquad \text{(Law of total probability)} \\
&= \mathbb{E}\Big[\mathbb{E}[\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}^{\mathbf{a}}_{1:t+2}, \mathbf{A}_{t+2} = \mathbf{a}_{t+2}] \,\Big|\, \mathbf{H}^{\mathbf{a}}_{1:t+1}, \mathbf{A}_{t+1} = \mathbf{a}_{t+1}] \,\Big|\, \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_t = \mathbf{a}_t\Big] \\
&\qquad\qquad \text{(Sequential ignorability and positivity)} \\
&\dots \\
&= \mathbb{E}\Big[\dots \mathbb{E}[\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{A}_{t+\tau-1} = \mathbf{a}_{t+\tau-1}] \\
&\qquad\qquad\qquad \Big| \mathbf{H}^{\mathbf{a}}_{1:t+\tau-2}, \mathbf{A}_{t+\tau-2} = \mathbf{a}_{t+\tau-2}] \\
&\qquad\qquad\qquad \Big| \dots \\
&\qquad\qquad\qquad \Big| \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_t = \mathbf{a}_t\Big] \\
&\qquad\qquad \text{(Sequential ignorability and positivity)} \\
&= \mathbb{E}\Big[\dots \mathbb{E}[\mathbb{E}[\mathbf{Y}_{t+\tau} \mid \mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{A}_{t+\tau-1} = \mathbf{a}_{t+\tau-1}] \\
&\qquad\qquad\qquad \Big| \mathbf{H}^{\mathbf{a}}_{1:t+\tau-2}, \mathbf{A}_{t+\tau-2} = \mathbf{a}_{t+\tau-2}]
\end{aligned}
$$

$$\Big| \ \ldots$$

$$\Big| \ \mathbf{H}_{1:t} = \mathbf{h}_{1:t}, \mathbf{A}_t = \mathbf{a}_t \Big] \qquad \text{(Consistency)}$$

Thus, if we had multiple spatiotemporal time-series samples, we could directly estimate this nested expression from data, since the right-hand side depends solely on observed quantities, ensuring identifiability.

Step 2: From Intractable to Prefix-based Expectations We now show how to estimate the nested expectations using the prefix data. First, by Assumption 7.2, we can rewrite the inner-most expectation as

$$\begin{aligned}\mathbb{E}[\mathbf{Y}_{t+\tau} \mid \mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{A}_{t+\tau-1} = \mathbf{a}_{t+\tau-1}] &= \mathbb{E}_{\mathbf{P}}[\mathbf{Y}_{t+\tau} \mid \phi(\mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1})] \\ &= Q_\tau(\mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1}). \qquad \text{(Definition of } Q_\tau\text{)}\end{aligned}$$

By using Assumption 7.1, we can write this expectation over the prefix data which we have many samples of. Now consider the next nested expectation:

$$\begin{aligned}&\mathbb{E}[Q_\tau(\mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1}) \mid \mathbf{H}^{\mathbf{a}}_{1:t+\tau-2} = \mathbf{h}^{\mathbf{a}}_{1:t+\tau-2}, \mathbf{A}_{t+\tau-2} = \mathbf{a}_{t+\tau-2}] \\ &= \int Q_\tau(\mathbf{h}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1}) p(x_{t+\tau-1}, y_{t+\tau-1} \mid \mathbf{h}^{\mathbf{a}}_{t+\tau-2}, \mathbf{a}_{t+\tau-2}) d(x_{t+\tau-1}, y_{t+\tau-1}) \\ &= \int_P Q_\tau(\mathbf{h}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1}) p(x_{t+\tau-1}, y_{t+\tau-1} \mid \phi(\mathbf{h}^{\mathbf{a}}_{t+\tau-2}, \mathbf{a}_{t+\tau-2})) d(x_{t+\tau-1}, y_{t+\tau-1}) \\ &\qquad\qquad \text{(Assumption 7.2)} \\ &= \mathbb{E}_P[Q_\tau(\mathbf{H}^{\mathbf{a}}_{1:t+\tau-1}, \mathbf{a}_{t+\tau-1}) \mid \phi(\mathbf{H}^{\mathbf{a}}_{1:t+\tau-2}, \mathbf{A}_{t+\tau-2}) = \phi(\mathbf{h}^{\mathbf{a}}_{1:t+\tau-2}, \mathbf{a}_{t+\tau-2})] \\ &= Q_{\tau-1}(\mathbf{h}^{\mathbf{a}}_{1:t+\tau-2}, \mathbf{a}_{t+\tau-2})\end{aligned}$$

Tracing this argument recursively through the nested expectation in Step 1, we obtain:

$$\mathbb{E}[\mathbf{Y}_{t+\tau}[\mathbf{a}_{t:t+\tau-1}] \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}] = Q_1(\mathbf{h}_{1:t}, \mathbf{a}_t),$$

as desired. Thus, Q_1 – which can be estimated from the prefix data – recovers the CAPOs, under our assumptions, even from a single chain.

F.3 Consistency of the Iterative G-Computation Estimator

In this section, we state the conditions under which the iterative G-computation procedure in Section 7.4.2 yields a consistent estimator, and show that our implementation of the Q_k estimators satisfies these conditions.

Notation We denote the L_2 norm of a function f as $\|f\|_2 := \mathbb{E}_P[f(X)^2]^{1/2}$, where the expectation is over the probability distribution P. The notation $\widehat{f_n}$ represents the estimated value of a parameter or function learned on n data points, where f is the true value. For a sequence of random variables $\{Z_n\}_{n\geq 1}$ we write $Z_n = o_p(1)$ if $\Pr(|Z_n| > \varepsilon) \to 0$ for every $\varepsilon > 0$, *i.e.*$Z_n \xrightarrow{p} 0$.

To begin, we introduce the following stochastic equicontinuity condition from [Van Der Vaart et al., 1996]:

Definition F.1 (Stochastic equicontinuity [Van Der Vaart et al., 1996, Def. 1.5.7])**.** Let $(\mathcal{Z}, d)$ be a semi-metric space and $\{\widehat{f_n}\}_{n\geq 1} \subset \ell^\infty(\mathcal{Z})$ a sequence of random functions. It is *uniformly stochastically equi-continuous* if, for every $\epsilon > 0, \eta > 0$, there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\Big(\sup_{d(z,z')\leq\delta} \big|\widehat{f_n}(z) - \widehat{f_n}(z')\big| > \epsilon\Big) < \eta.$$

Stochastic equicontinuity ensures that, with high probability, each estimator changes only slightly when its input is perturbed by a small amount. It is strictly weaker than global Lipschitz continuity – any family that is Lipschitz on a bounded domain with constants bounded in probability automatically satisfies Definition F.1. We impose this condition in Theorem F.2 so that the $o_p(1)$ error in the learned embedding propagates to only $o_p(1)$ errors in the G-heads, making the recursive estimator consistent.

The following theorem restates Theorem 7.4 from the main text in full detail and provides its proof.

Theorem F.2 (Consistency under Uniform Stochastic Equicontinuity)**.** *Suppose the conditions of Theorem 7.3 hold, and let $\widehat{\phi}$ be a learned embedding. Define $\boldsymbol{Z}_k := (\boldsymbol{H}_{1:t+k}, \boldsymbol{A}_{t+k})$, and recursively define the learned estimators $\widehat{Q}_k(\mathbf{Z}_{k-1}; \widehat{\phi}) := \widehat{\mathbb{E}}_{\mathbf{P}}[\widehat{Q}_{k+1}(\boldsymbol{Z}_k; \widehat{\phi}) \mid \widehat{\phi}(\boldsymbol{Z}_k)]$ for $k = 1, \ldots, \tau$, with terminal condition $\widehat{Q}_{\tau+1}(\boldsymbol{Z}_\tau; \widehat{\phi}) = Y^{t+\tau}$. Assume that $\{\widehat{Q}_k\}_{k=1}^{\tau}$ are obtained via the iterative G-computation algorithm. If:*

(i) $\|\widehat{\phi} - \phi\|_2 = o_p(1)$*;*

(ii) $\|\widehat{Q}_k(\mathbf{Z}_{k-1}; \phi) - Q_k(\mathbf{Z}_{k-1}; \phi))\|_2 = o_p(1)$ *for all* k*;*

(iii) for every k *the random maps* $z \mapsto \widehat{Q}_k(h, a; z)$ *are stochastically equicontinuous on* $\operatorname{Im} \phi$ *(Definition F.1), and* $Q_k(\cdot)$ *is uniformly continuous there,*

then

$$\left\| \widehat{Q}_1(\mathbf{Z}_0; \widehat{\phi}) - Q_1(\mathbf{Z}_0; \phi) \right\|_2 = o_p(1).$$

Thus the recursive G-computation estimator is (probabilistically) consistent.

Proof. We proceed by reverse induction on k, starting from $k = \tau$ and working backward to $k = 1$. For each k, we aim to show:

$$\|\widehat{Q}_k(\mathbf{Z}_{k-1}; \widehat{\phi}) - Q_k(\mathbf{Z}_{k-1}; \phi)\|_2 = o_p(1).$$

Base case ($k = \tau$). By definition, $\widehat{Q}_{\tau+1}(\mathbf{Z}_\tau; \widehat{\phi}) = Y^{t+\tau}$, which is observed. Thus,

$$\widehat{Q}_\tau(\mathbf{Z}_{\tau-1}; \widehat{\phi}) = \widehat{\mathbb{E}}_{\mathbf{P}}[Y^{t+\tau} \mid \widehat{\phi}(\mathbf{Z}_\tau)] \quad \text{and} \quad Q_\tau(\mathbf{Z}_{\tau-1}; \phi) = \mathbb{E}[Y^{t+\tau} \mid \phi(\mathbf{Z}_\tau)].$$

We decompose the difference:

$$\left\| \widehat{Q}_\tau(\mathbf{Z}_{\tau-1}; \widehat{\phi}) - Q_\tau(\mathbf{Z}_{\tau-1}; \phi) \right\|_2 \leq \underbrace{\left\| \widehat{Q}_\tau(\mathbf{Z}_{\tau-1}; \widehat{\phi}) - \widehat{Q}_\tau(\mathbf{Z}_{\tau-1}; \phi) \right\|_2}_{\Lambda_1} + \underbrace{\left\| \widehat{Q}_\tau(\mathbf{Z}_{\tau-1}; \phi) - Q_\tau(\mathbf{Z}_{\tau-1}; \phi) \right\|_2}_{\Lambda_2}.$$

Term Λ_2 is $o_p(1)$ by assumption (ii). Term Λ_1 converges to zero in probability due to (i) $\|\widehat{\phi} - \phi\|_2 = o_p(1)$ and (iii) stochastic equicontinuity of $\widehat{Q}_\tau$. Therefore,

$$\left\|\widehat{Q}_\tau(\mathbf{Z}_{\tau-1}; \widehat{\phi}) - Q_\tau(\mathbf{Z}_{\tau-1}; \phi)\right\|_2 = o_p(1).$$

Inductive step. Suppose for some $k + 1 \leq \tau$ that

$$\left\|\widehat{Q}_{k+1}(\mathbf{Z}_k; \widehat{\phi}) - Q_{k+1}(\mathbf{Z}_k; \phi)\right\|_2 = o_p(1).$$

We now consider

$$\widehat{Q}_k(\mathbf{Z}_{k-1}; \widehat{\phi}) = \widehat{\mathbb{E}}_{\mathbf{P}}[\widehat{Q}_{k+1}(\mathbf{Z}_k; \widehat{\phi}) \mid \widehat{\phi}(\mathbf{Z}_k)], \quad Q_k(\mathbf{Z}_{k-1}; \phi) = \mathbb{E}[Q_{k+1}(\mathbf{Z}_k; \phi) \mid \phi(\mathbf{Z}_k)].$$

Again, decompose:

$$\begin{aligned}\left\|\widehat{Q}_k(\mathbf{Z}_{k-1}; \widehat{\phi}) - Q_k(\mathbf{Z}_{k-1}; \phi)\right\|_2 \leq &\left\|\widehat{Q}_k(\mathbf{Z}_{k-1}; \widehat{\phi}) - \widehat{Q}_k(\mathbf{Z}_{k-1}; \phi)\right\|_2 \\ &+ \left\|\widehat{Q}_k(\mathbf{Z}_{k-1}; \phi) - Q_k(\mathbf{Z}_{k-1}; \phi)\right\|_2.\end{aligned}$$

The second term is $o_p(1)$ by assumption (ii). The first term is also $o_p(1)$ because $\widehat{\phi} \to \phi$ in L_2 and the stochastic equicontinuity of $\widehat{Q}_k$ ensures that perturbations in ϕ yield small changes in predictions uniformly over $\operatorname{Im} \phi$. Thus,

$$\|\widehat{Q}_k(\mathbf{Z}_{k-1}; \widehat{\phi}) - Q_k(\mathbf{Z}_{k-1}; \phi)\|_2 = o_p(1).$$

By induction, the result holds for all $k = \tau, \tau - 1, \ldots, 1$, and in particular:

$$\|\widehat{Q}_1(\mathbf{Z}_0; \widehat{\phi}) - Q_1(\mathbf{Z}_0; \phi)\|_2 = o_p(1).$$

Thus, the proof is now complete. □

Example F.3 (Feed-forward or Convolutional Heads)**.**
Suppose each G-computation head $Q_k(\cdot; z)$ is implemented as a depth-d neural network given by

$$\Psi(z) = W_d \sigma_{d-1}(\cdots \sigma_1(W_1 z)),$$

where the activations σ_ℓ are Lipschitz continuous (e.g., ReLU, Leaky ReLU, SoftPlus, Tanh, Sigmoid, or ArcTan). If each layer weight satisfies a spectral norm bound $\|W_\ell\|_2 \leq \rho_\ell < \infty$, then Ψ is globally Lipschitz on $\mathbb{R}^h$ with constant $L = \prod_\ell \rho_\ell$, and thus uniformly continuous on any compact subset. This implies the stochastic equicontinuity condition in Definition F.1.

In practice, norm control can be enforced via weight decay, spectral normalization, or weight clipping during training. Similarly, the encoder output $\widehat{\phi}(H, A)$ can be bounded—e.g., through normalization or clipping—so its image lies in a compact subset of $\mathbb{R}^h$. Together, these ensure the continuity and equicontinuity conditions required by Theorem F.2.

The same argument applies to convolutional networks, since 2-D convolutions are linear operators whose induced matrix representations also admit spectral norm bounds controlled via spectral normalization.

F.4 Experimental Details

In this appendix, we provide further information on the simulation experiments (Section 7.6.1) and the real-world wildfire application (Section 7.6.2), including exact parameter settings, model architecture and execution details, hyperparameter selection strategies, and validation procedures. All code for generating, preprocessing, and analyzing both the synthetic and real-world datasets—and for training and evaluating GST-UNet—is available at `https://github.com/moprescu/GSTUNet`, with step-by-step replication instructions in the repository's `README.md`.

For both applications, GST-UNet employs a U-Net backbone with a single ConvLSTM layer (hidden dimension 32) and a contracting-expanding path of channel sizes $16 \to 32 \to 64 \to 128 \to 256$. The G-computation heads are im-

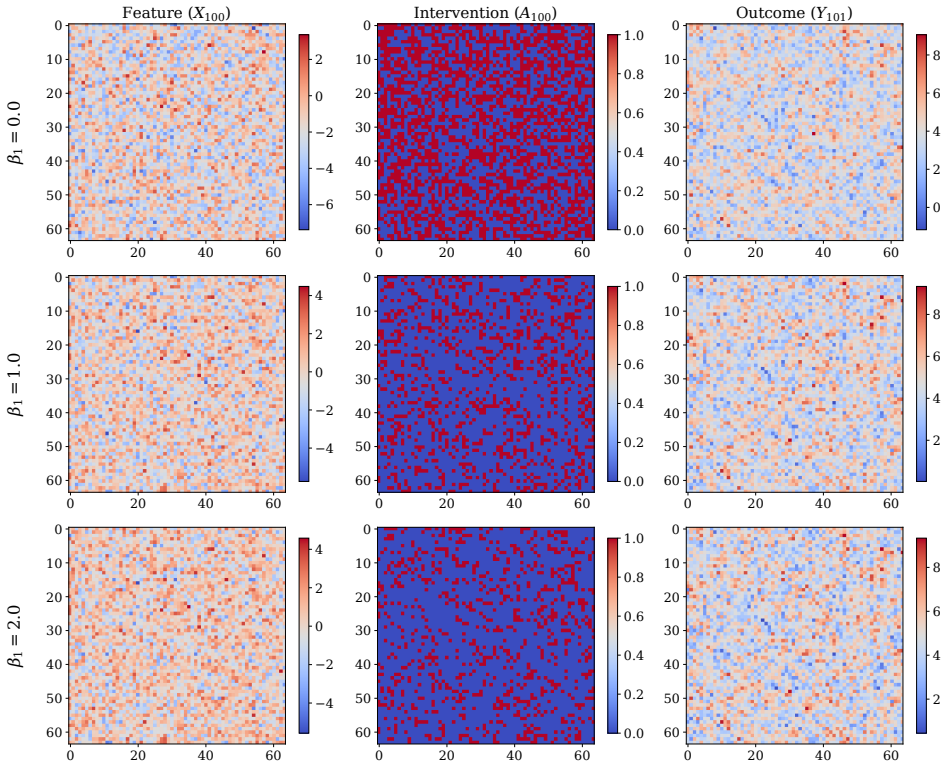


Figure F.1: Samples from the GST-UNet synthetic data-generating process at time $t = 100$, showing the covariate field $\mathbf{X}_{100}$, treatment field $\mathbf{A}_{100}$, and next-step outcome field $\mathbf{Y}_{101}$ for varying strengths of time-varying confounding $\beta_1 \in \{0.0, 1.0, 2.0\}$.

plemented as shallow feed-forward neural networks that operate on the U-Net feature maps at each grid cell for G-computation. In practice, to ensure stable ConvLSTM training and reduce computational overhead, we truncate the input history to a fixed length. All neural networks are implemented via the `nn` module in `PyTorch` [Paszke et al., 2019]. Experiments were conducted on an NVIDIA A100 (Ampere) GPU using the Perlmutter system at the National Energy Research Scientific Computing Center (NERSC). The synthetic experiments required roughly 55 minutes per hyperparameter set, while the wildfire experiment completed in about 5 minutes.

F.4.1 Synthetic Experiments

Data Simulation Process. For our primary simulation experiments, we generate $T = 200$ time steps on a 64×64 grid. The simulation parameters in the generating equations

$$\mathbf{X}_t = \alpha_0 + \alpha_1 \mathbf{X}_{t-1} + \alpha_2 \mathbf{A}_{t-1} + \alpha_3 (K_X * \mathbf{X}_{t-1}) + \epsilon_X,$$

$$\mathbf{A}_t \sim \text{Bern}\Big(\sigma(\beta_1(\beta_0 + \tfrac{1}{L}\sum_{l=0}^{L-1} K_A * \mathbf{X}_{t-l}))\Big),$$

$$\mathbf{Y}_t = \gamma_0 + \gamma_1 (K_{YA} * \mathbf{A}_{t-1}) + \gamma_2 \tfrac{1}{L}\sum_{l=1}^{L} (K_{YX} * \mathbf{X}_{t-l}) + \gamma_3 \mathbf{Y}_{t-1} + \epsilon_Y,$$

are given by:

- $\mathbf{X}_t$:

$$\alpha_0 = 0.5,\ \alpha_1 = 0.5,\ \alpha_2 = -2.0,\ \alpha_3 = 0.2,\ K_X = \begin{pmatrix} 0 & 0.45 & 0 \\ 0.15 & 0 & 0.35 \\ 0 & 0.05 & 0 \end{pmatrix}.$$

  where K_X influences how $\mathbf{X}$ diffuses across neighboring cells, with an asymmetry due to advection.

- $\mathbf{A}_t$:

$$\beta_0 = -1.0,\ \beta_1 \in \{0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0\},\ K_A = \frac{1}{16}\begin{pmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 8.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{pmatrix}.$$

- $\mathbf{Y}_t$:

$$\gamma_0 = 2.0,\ \gamma_1 = 1.5,\ \gamma_2 = 0.5,\ \gamma_3 = 0.5$$

$$K_{YX} = \frac{1}{16}\begin{pmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 8.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{pmatrix},\ K_{YA} = \frac{1}{16}\begin{pmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 8.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{pmatrix}.$$

Table F.2: Hyperparameters and search ranges used in GST-UNet, with the best validation values shown in bold.

Hyperparameter	**Model(s)**	**Value Range**
Batch size	All models	$\{2, \mathbf{4}, 8\}$
Learning rate	All models	$\{10^{-4}, \mathbf{5 \times 10^{-4}}, 10^{-3}\}$
Scheduler patience	All models	$\{3, \mathbf{5}, 10\}$
Early stopping patience	All models	$\{5, \mathbf{10}\}$
Curriculum period	GST-UNet	$\{1, \mathbf{3}, 5, 7\}$
Curriculum learning rate	GST-UNet	$\{10^{-4}, \mathbf{5 \times 10^{-4}}, 10^{-3}\}$
UNet output dim d_h	GST-UNet	$\{8, \mathbf{16}, 32\}$
G-head hidden size	GST-UNet	$\{\mathbf{8}, 16\}$
G-head layers	GST-UNet	$\{\mathbf{1}, 2, 3\}$

Table F.3: Ablation on spatial kernel size for GST-UNet at horizon $\tau = 5$. Removing neighbor aggregation (1×1 kernel) degrades performance, confirming the need to model spatial spill-overs.

Kernel	β_1=0.0	0.5	1.0	1.5	2.0
3×3	**0.33 ± 0.004**	**0.35 ± 0.004**	**0.40 ± 0.005**	**0.44 ± 0.004**	**0.40 ± 0.005**
1×1	0.53 ± 0.004	0.55 ± 0.005	0.54 ± 0.005	0.60 ± 0.007	0.64 ± 0.006

We use $L = 5$ temporal lags for $\mathbf{X}$ and $\mathbf{Y}$, a seed of 42 for reproducibility. The parameter values were chosen such that the simulation remains stable (*i.e.*, the process does not diverge). See Figure F.1 for representative $t = 100$ snapshots of X_{100}, A_{100}, and Y_{101} under varying β_1.

For each β_1, we first generate a factual dataset of length $T = 200$ (*i.e.*, $\{(\mathbf{X}_t, \mathbf{A}_t, \mathbf{Y}_t)\}_{t=1}^{200}$). We then create $n_{\text{test}} = 50$ test histories of length $l_H = 10$. For each test history, we simulate 100 trajectories under a randomly chosen (yet fixed over the test data) counterfactual intervention of length $\tau = 10$, and average the outcomes at each step to approximate the true CAPOs. This procedure yields a final test set of shape $n_{\text{test}} \times (l_H + \tau + 1) \times 64 \times 64$, *i.e.*, $50 \times 21 \times 64 \times 64$.

Neural Architectures The **GST-UNet** comprises a single ConvLSTM layer (hidden dimension 32), followed by a U-Net with channel sizes $16 \to 32 \to 64 \to$

$128 \rightarrow 256$. Its G-computation heads are shallow feed-forward networks operating on the final U-Net feature maps at each grid cell; both the U-Net's output dimension (d_h) and the G-head architecture (number of layers, hidden size) are treated as hyperparameters. The **UNet+** baseline uses the same ConvLSTM+U-Net backbone as GST-UNet but outputs a single channel ($d_h = 1$), omitting any G-computation. For direct comparison, we also implement **STCINet** [Ali et al., 2024] with an identical ConvLSTM+U-Net backbone, and retaining their original Latent Factor Model (LFM) details.

IPWUNet Baseline We adapt the Inverse Propensity Weighting (IPW) estimator from [Zhou et al., 2024] to the spatiotemporal setting. Given estimated propensities $\hat{\pi}(\mathbf{a}_l \mid \mathbf{H}_{1:l})$, the estimator is defined as:

$$\hat{Y}^{\text{IPW}}_{t+\tau} = \left(\prod_{l=t}^{t+\tau} \frac{\mathbb{I}[\mathbf{A}_l = \mathbf{a}_l]}{\hat{\pi}(\mathbf{a}_l \mid \mathbf{H}_{1:l})} \right), \quad \text{CAPO} = \mathbb{E}[\hat{Y}^{\text{IPW}}_{t+\tau} \mid \mathbf{H}_{1:t} = \mathbf{h}_{1:t}].$$

We implement the IPWUNet baseline by reusing the UNet+ architecture (U-Net + ConvLSTM + Attention) for both propensity estimation and outcome prediction. Specifically, we first train the propensity model with a binary cross-entropy loss to estimate $\hat{\pi}(\mathbf{A}_t \mid \mathbf{H}_t)$ at each time t. We then freeze this model and use the estimated weights to train a second instance of the same architecture with a weighted MSE loss, where pseudo-outcomes are reweighted by the estimated inverse propensities along the counterfactual treatment path. While this allows partial adjustment for time-varying confounding, the method does not correct for spatial interference and is sensitive to small propensity values, which can lead to high variance.

Training Details We randomly initialize all model parameters (GST-UNet and baselines) with Xavier uniform weights [Glorot and Bengio, 2010]. We use the Adam optimizer [Kingma and Ba, 2015] with an initial learning rate, halving

Table F.4: Effect of increasing trajectory length T on RMSE in the GST-UNet synthetic experiment for confounding strength $\beta_1 = 2.0$. GST-UNet improves as more trajectory data are observed, while the baselines remain biased.

Model	**T=100**	**T=200**	**T=400**	**T=600**	**T=800**
UNet+	0.78	0.81	0.82	0.95	0.87
STCINet	0.80	0.90	1.04	1.02	0.91
GST-UNet	**0.69**	**0.40**	**0.32**	**0.32**	**0.36**

it whenever the validation loss plateaus for a specified scheduler patience. To mitigate overfitting, we adopt early stopping when the validation loss fails to improve for a specified early stopping patience epochs. Validation uses 40 of the 190 training prefixes, and the total training is capped at 100 epochs. We tune the following hyperparameters: *(i)* batch size, learning rate, scheduler patience, and early stopping patience (common to all models); *(ii)* for GST-UNet, the curriculum period and learning rate for curriculum phases, the U-Net output dimension d_h, and the number and width of hidden layers in the feed-forward G-heads. Table F.2 lists the hyperparameter ranges considered, with the values yielding the best validation performance in **bold**.

Evaluation Procedure We evaluate each model by averaging the root mean square error (RMSE) of the estimated CAPOs against ground truth across 50 test trajectories. Table 7.1 in the main text reports RMSE ± standard deviation for horizon lengths $\tau \in \{5,\ 10\}$ and $\beta_1 \in \{0, 0.5, 1.0, 1.5, 2.0\}$.

Effect of Varying T We ran additional simulations varying the trajectory length $T \in \{100, 200, 400, 600, 800\}$ and $\beta_1 = 2.0$ while keeping the grid size fixed ($d_x = d_y = 64$). Results are shown in Table F.4. GST-UNet consistently improves with more data, while the baselines remain biased—even as T increases. This highlights the importance of adjusting for time-varying confounding: without it, there is a persistent asymptotic bias.

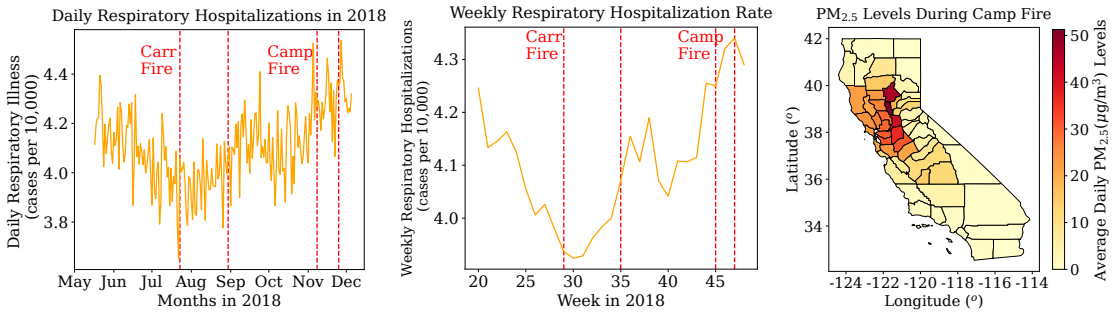


Figure F.2: Wildfire application data summary. **(Left)** Daily respiratory hospitalizations incidence (cases per 10,000). **(Center)** Weekly aggregated respiratory hospitalizations incidence. **(Right)** Average daily $PM_{2.5}$ during the Camp Fire.

Effect of Neighbor Aggregation To evaluate the importance of spatial spillover modeling, we ablate the convolutional kernel used in the ConvLSTM encoder. Table F.3 compares GST-UNet with a standard 3×3 kernel against a variant that removes neighbor aggregation by using a 1×1 kernel. Across all levels of confounding strength (β_1), performance deteriorates markedly when neighbor information is excluded, with RMSE increasing by 30–40%. This confirms that explicitly aggregating information from nearby locations is essential for capturing spatial interference and achieving unbiased counterfactual estimates.

F.4.2 Wildfire Application

Data Preprocessing and Interpolation We analyze daily, county-level data from Letellier et al. [2025] that include $PM_{2.5}$ (particulate matter $< 2.5\,\mu m$), hospitalization counts for respiratory and cardiovascular conditions, and weather variables (temperature, precipitation, humidity, radiation, wind), plus population estimates from the California Department of Finance. Our study period spans weeks 20–48 (May 18–December 2, 2018), covering both the Carr and Camp fires. As illustrated in Figure F.2, daily and weekly aggregated respiratory illness rates rise around these events, while $PM_{2.5}$ levels also surge during

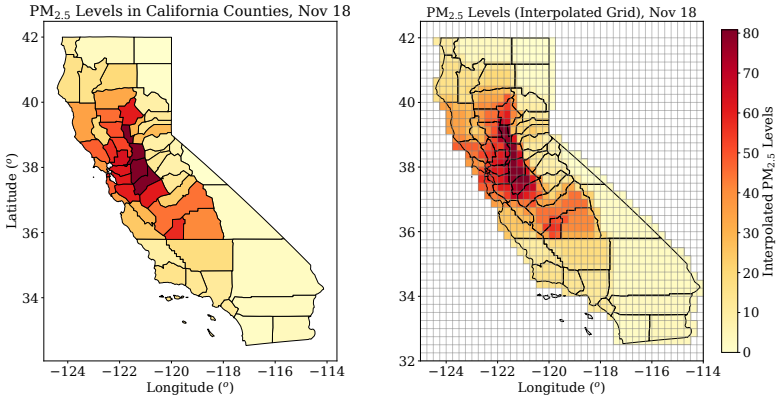


Figure F.3: Example of county-level (*left*) vs. grid-interpolated (*right*) $PM_{2.5}$ levels on November 18 (during the Camp Fire). The interpolation converts county measurements into the 40×44 spatiotemporal lattice used by GST-UNet.

the Camp Fire.

To align with our spatiotemporal framework, we use `geopandas` [Jordahl et al., 2020] to interpolate county-level covariates, $PM_{2.5}$, and hospitalizations onto a latitude–longitude grid from 32°N to 42°N latitude and -125° to -114° longitude, at a resolution of 0.25°. Each grid cell's values are an area-weighted average of the counties it intersects, yielding a 40×44 spatial lattice. We mask out non-California cells by setting them to zero, thus obtaining a consistent dataset for further analysis. Figure F.3 illustrates how the raw county-level data compare to the interpolated grid for $PM_{2.5}$ on November 18.

Model Training and Validation We train GST-UNet, UNet+, STCINet, and IP-WUNet with a prediction horizon of $\tau = 10$ days. All models use a shared set of hyperparameters: batch size = 4, learning rate = 5×10^{-4}, scheduler patience = 5, early stopping patience = 10, and a curriculum period = 5 (with curriculum learning rate = 5×10^{-4}). For GST-UNet, we additionally set the U-Net output dimension to $d_h = 16$, the G-head hidden layer size to 8, and the number of G-head layers to 1. We optimize a mean squared error (MSE) loss with two adjustments: (1) we mask grid cells outside California to exclude them from the loss compu-

Table F.5: Estimated county-level increases in respiratory emergency department visits attributable to the wildfire event, with 95% bootstrap confidence intervals. Population is reported in units of 10,000; counties marked with * have smaller populations and therefore greater uncertainty.

County	Mean	2.5%	97.5%	Population ($\times 10^4$)	Interval Width / Pop.
Tehama	37	-126	158	6.4	44.4
Butte	168	30	325	23.0	12.8
Glenn*	-52	-262	39	2.8	107.6
Colusa*	13	-158	107	2.1	124.0
Sutter	-18	-170	70	9.6	24.9
Napa	81	-41	192	13.9	16.8
Lake	103	-66	203	6.4	41.8
Solano	38	-79	173	44.6	5.6
Sacramento	202	-107	484	153.9	3.8

tation, and (2) we apply cell-specific weights proportional to the number of grid cells per county to avoid bias toward geographically larger counties. Hyperparameter tuning and validation are performed using data from the first 50 days of the wildfire season. Using the selected configuration, we generate counterfactual predictions for the Camp Fire peak period (November 8–17) by iteratively applying each trained model with increasing history lengths. We note that counties with populations below 20,000–30,000 can yield unreliable incidence rate estimates (baseline daily rates of approximately 4 cases per 10,000 individuals); in Figure 7.3, we denote these high-uncertainty counties with hatched markings.

Bootstrap Confidence Intervals We compute 95% bootstrap confidence intervals for all models using $n = 40$ bootstrap samples, balancing statistical rigor with computational load. Counties with populations below 20,000–30,000 tend to yield unstable incidence rate estimates, driven by low baseline daily counts (approximately 4 cases per 10,000), and are excluded from the analysis. These counties are indicated with hatching in Figure 7.3, a choice further supported by the bootstrap results. In Table F.5, we report bootstrap intervals for the coun-

ties closest to the Camp Fire. Glenn and Colusa exhibit disproportionately wide intervals–reflecting the uncertainty introduced by their small population sizes–and this further justifies their exclusion from the final analysis.

F.5 Limitations and Broader Impacts

Limitations While GST-UNet demonstrates strong empirical performance and theoretical grounding, several limitations should be acknowledged. First, our method relies on standard causal identification assumptions, including no unobserved confounding (Assumption 7.1), which is inherently untestable and may not hold in all real-world settings. Second, our framework assumes the existence of a time-invariant representation of the spatiotemporal process (Assumption 7.2)–a useful but idealized condition that may be violated in domains with highly non-stationary or regime-shifting dynamics. Finally, GST-UNet is designed for gridded spatiotemporal data and assumes a regular spatial lattice; while this is common in environmental and health applications, adapting the framework to irregular spatial structures (e.g., graphs or administrative boundaries) is an important direction for future work.

Broader Impacts This work advances machine learning by introducing a spatiotemporal causal inference framework for estimating treatment effects in complex real-world settings. The GST-UNet has broad applications in public health, environmental science, and social policy, where understanding interventions supports evidence-based decisions. For example, it can inform pollution control, wildfire response, or health resource allocation. However, like all observational methods, GST-UNet depends on the quality and completeness of the data, as well as the assumptions stated in this work. We caution against uncritical use in high-stakes settings, as violations of model assumptions or data biases

can lead to misleading conclusions. We encourage responsible deployment—especially in contexts affecting vulnerable populations—and recommend pairing our method with domain expertise, sensitivity analysis, and uncertainty quantification.

APPENDIX G

APPENDIX FOR CHAPTER 8

G.1 Extended Literature Review

The **Spatial Deconfounder** draws on three strands of prior work: (i) spatial causal inference under interference and spatially structured confounding, (ii) deconfounding methods for ATE estimation with unobserved confounders, and (iii) deep learning for spatial and latent structure modeling. We detail each in the sections that follow.

G.1.1 Spatial Causal Inference Under Interference and Spatially Structured Confounding

Classical Spatial Causal Inference Most estimators of direct and spillover effects assume that bias can be removed by conditioning on *observed* covariates (together with a specified exposure mapping or interference structure). Design-based work—grounded in exposure mappings, partial-interference designs, and randomization inference—derives estimators or hypothesis tests under known neighborhood or network structure [e.g., Hudgens and Halloran, 2008, Sobel, 2006, Aronow and Samii, 2017, Forastiere et al., 2021, Tchetgen Tchetgen et al., 2021]. Model-based strategies then adjust for that structure while still relying on measured covariates or correct functional form: spatial autoregressive and two-stage least-squares estimators for spatial-lag/lagged-error models [Anselin, 1988], and spline/GAM or restricted spatial regression approaches that treat residual spatial trend as a nuisance to improve precision and approximate balance [e.g., Hanks et al., 2015]. Deep graph/convolutional architectures can pool information across nearby units to improve prediction or imputation, but by themselves do not furnish identification without additional

causal assumptions [Kipf and Welling, 2017]. Domain-specific simulators (e.g., wildfire spread or atmospheric transport) encode spatial dependence through process-based physics and are often used as inputs to causal analyses, yet they typically still condition on observed drivers or require design-identifying assumptions [e.g. Larsen et al., 2022, Zigler et al., 2025]. All of the above *presume exchangeability given observed covariates (or a valid design)*; if important spatial determinants of treatment and outcome are unmeasured, residual confounding bias can remain.

Spatial Confounding and Bias-Adjustment Methods. A growing literature tackles *unmeasured* spatial confounding directly. One family augments outcome models with latent spatial random effects (e.g., BYM/ICAR or GMRF priors) to soak up smooth hidden structure; this can reduce bias when the confounder is well captured by the basis, but may leave bias or distort fixed effects under misspecification [Rue and Held, 2005, Hodges and Reich, 2010]. Restricted spatial regression and related orthogonalization schemes constrain the latent field away from covariates to mitigate bias [Hanks et al., 2015]. Building on this idea, Dupont et al. [2022] (SPATIAL+) explicitly orthogonalizes spatial structure in the covariates from the outcome trend to purge bias from unmeasured *spatial* confounding. Propensity-score strategies that incorporate spatial proximity—such as distance-adjusted propensity score matching—aim to proxy smooth unmeasured confounders via geography [Papadogeorgou et al., 2019a]. Instrumental-variable designs exploit exogenous spatial shocks (e.g., wind direction, policy boundaries, thermal inversions) to identify causal effects despite hidden confounding, but require strong relevance/exclusion conditions that are difficult to validate under interference [e.g., Angrist et al., 1996, Imbens and Rubin, 2015, Deryugina et al., 2019]. Finally, Bayesian frameworks that jointly model interfer-

ence and latent spatial fields (e.g., Papadogeorgou and Samanta, 2023) achieve identification under specified priors and structural assumptions. In short, existing approaches either (i) assume smoothly varying latent fields or valid instruments or (ii) rely on strong parametric priors. None exploit interference patterns themselves as a *signal* for nonparametrically recovering the hidden confounder, nor do they aim to explicitly reconstruct the unobserved confounding process—a gap our Spatial Deconfounder addresses.

G.1.2 Deconfounding Methods for ATE Estimation with Unobserved Confounders

When confounders are unmeasured, point identification of causal effects generally fails. One approach is to derive bounds through sensitivity analysis [e.g., VanderWeele et al., 2015, Dorn et al., 2025b, Oprescu et al., 2023, Frauen et al., 2023], trading identifiability for robustness. Another is the *deconfounder* framework, which fits a factor model to multiple causes in order to infer a substitute for the latent confounder, thereby restoring point identification [Wang and Blei, 2019, Bica et al., 2020a, Hatt and Feuerriegel, 2024]. This stream is closest in spirit to our work: like us, it leverages multiplicity of treatments as a proxy for hidden structure. However, existing deconfounder methods require datasets with many simultaneous treatments (e.g., recommender systems, panel data) and assume no interference. Our approach resolves both limitations: interference itself naturally generates multiple-cause treatment vectors, enabling latent field recovery even with a single treatment type.

G.1.3 Deep Learning for Spatial and Latent Structure Modeling

Deep Learning for Spatial Modeling Modern deep architectures capture rich spatial structure but, on their own, remain predictive rather than identifying. U-

Nets and encoder–decoder variants model multi-scale patterns on grids [Ronneberger et al., 2015b, Oktay et al., 2018]; graph neural networks extend to irregular domains [Kipf and Welling, 2017, Hamilton et al., 2017, Veličković et al., 2018]; and patch-wise transformers model long-range dependencies on images and geospatial rasters [Dosovitskiy et al., 2021, Liu et al., 2021]. Spatiotemporal extensions (e.g., ConvLSTM and graph/vision transformers) further capture dynamics [Shi et al., 2015]. These tools provide flexible representations but require additional causal structure for identification.

Deep Latent Variable Models Finally, conditional variational autoencoders (C-VAEs) and related deep generative models are widely used for representation learning with latent factors [Kingma and Welling, 2013, Sohn et al., 2015]. Beyond C-VAEs, the broader family of latent-variable models includes variational autoencoders with structured priors [Rezende et al., 2014, Maaløe et al., 2016], disentangled representation learning [Higgins et al., 2017], normalizing flows [Rezende and Mohamed, 2015], and diffusion-based generative models [Ho et al., 2020, Kingma et al., 2021], all of which offer flexible ways to recover hidden structure from high-dimensional data. While these methods are not causal in themselves, they provide natural tools for reconstructing latent processes from observed multi-cause data. In our framework, a C-VAE combined with a spatial prior enables smooth, nonparametric recovery of a substitute confounder from local treatment vectors, which is then used for causal identification. Other architectures (e.g., diffusion models or flow-based methods) could, in principle, be substituted, but the key contribution lies in adapting deep latent-factor reconstruction to the spatial interference setting, where treatments on neighboring units jointly reveal the latent field.

G.1.4 Causal Generative Models

Recent work has proposed using expressive generative models as parameterizations of structural causal models. One stream of work uses autoregressive flows to obtain identifiable SCMs given a causal ordering [e.g., Javaloy et al., 2023, Khemakhem et al., 2021]. Others combine diffusion- or GAN-based models with structural equations to model complex, high-dimensional counterfactuals [Sanchez and Tsaftaris, 2022, Kocaoglu et al., 2018]. However, all of the methods assume unconfoundedness and are thus orthogonal to our Spatial Deconfounder. A different stream of literature combines causal inference and generative modeling under hidden confounding [e.g., Xia et al., 2021, Almodóvar et al., 2025]. Similar to our work, the recently proposed DeCaFlow [Almodóvar et al., 2025] extends this line by learning confounded SCMs with causal normalizing flows and variational inference based on the deconfounder framework. However, these works are restricted to specific variables types, e.g., continuous treatments, and do not apply to the spatial setting. Building upon proxy variables, follow-up work on the deconfounder clarifies identifiability conditions in multi-cause settings [Wang and Blei, 2021]. Similarly, this work assumes multiple treatments in an independent setting and does not apply to spatial causal inference tasks.

G.1.5 Deep Identifiable Models and Network Deconfounding

A complementary line of work focuses on identifiability in deep latent variable models. Sparse deep generative models establish identifiability of VAEs under sparsity constraints Moran et al. [2022], while Intact-VAE [Wu and Fukumizu, 2021] and β-Intact-VAE [Wu and Fukumizu, 2022] provide identifiable generative models for causal inference under unobserved confounding, IVs, proxies,

and networked confounding. Applications to medical data show how identifiable VAEs can recover meaningful latent prognostic factors [Ma et al., 2023]. These methods are typically designed for i.i.d. or network-structured observations and often rely on known adjacency structure, e.g., using neighbor information to help identify latent confounders in network deconfounding tasks. Our Spatial Deconfounder differs by targeting a specific spatial setting with localized grid-interference. More importantly, we note that our Spatial Deconfounder is not limited to the use of a C-VAE. The framework is model-agnostic and can be combined with other generative factor models. In contrast to these identifiable deep models, our focus is on a spatial–interference design: we show that interference-generated multi-cause vectors $(A_s, A_{\mathcal{N}_s})$, together with a spatial prior on Z, are sufficient to identify both direct and spillover effects without specifying a parametric latent-field model.

Our Work Our contribution lies at the intersection of spatial causal inference, methods for deconfounding under unobserved confounding, and modern deep latent-variable modeling. Existing approaches to spatial interference either assume that all relevant confounders are observed, or else mitigate bias through strong structural assumptions and priors—for example, by imposing smooth latent fields, leveraging restrictive IV conditions, or specifying parametric Bayesian models. In parallel, the "deconfounder" framework demonstrates that multiplicity of causes can be exploited to infer substitutes for unobserved confounders, thereby restoring point identification; however, these methods are designed for i.i.d. settings with many simultaneous treatments (e.g., recommender systems, panels), and do not naturally extend to spatial domains where interference and locality are intrinsic.

The *Spatial Deconfounder* closes this gap. We treat interference itself as the

source of multi-cause information: treatment vectors on a unit and its neighbors contain precisely the dependence needed to reveal the hidden confounding field. By training a C-VAE with a spatial prior, we nonparametrically reconstruct a smooth latent confounder from these local treatment vectors. This substitute confounder can then be used to adjust for bias, enabling identification and estimation of both direct and spillover effects. Crucially, our method achieves this without committing to a fully specified latent-field model or relying on IV-style exclusion restrictions, thereby combining the flexibility of nonparametric deconfounding with the structural realities of spatial interference.

G.2 Proofs and Additional Results

We first provide background by stating supporting definitions and lemmas. Then we prove our main theorem on the identifiability of the treatment effects.

G.2.1 Supporting Lemmas and definitions

Definition G.1 (Ignorability)**.** The grid treatment $(a_s, \mathbf{a}_{\mathcal{N}_s})$ is *ignorable* given Z_s,$\mathbf{X}_s$,$\mathbf{X}_{\mathcal{N}_s}$, if for all $s = 1, \ldots, n$ and for all $(a, \mathbf{a}_{\mathcal{N}}) \in \mathcal{A}^{|\mathcal{S}|}$

$$(A_s, \mathbf{A}_{\mathcal{N}_s}) \perp\!\!\!\perp Y_s(a, \mathbf{a}_{\mathcal{N}}) \mid Z_s, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}. \tag{G.1}$$

Definition G.2 (Factor models)**.** A factor model of the assigned spatial treatments is a latent-variable model

$$p_\phi(z_{1:|\mathcal{S}|}, \boldsymbol{x}_{1:|\mathcal{S}|}, \boldsymbol{x}_{\mathcal{N}_{1:|\mathcal{S}|}}, a_{1:|\mathcal{S}|}, \mathbf{a}_{\mathcal{N}_{1:|\mathcal{S}|}}) \tag{G.2}$$

$$= p(z_{1:|\mathcal{S}|}, \boldsymbol{x}_{1:|\mathcal{S}|}, \boldsymbol{x}_{\mathcal{N}_{1:|\mathcal{S}|}}) \prod_{s=1}^{|\mathcal{S}|} p_\phi(a_s \mid z_s, \boldsymbol{x}_s, \boldsymbol{x}_{\mathcal{N}_s}) \prod_{k \in \mathcal{N}_s} p_\phi(a_k \mid z_s, \boldsymbol{x}_s, \boldsymbol{x}_{\mathcal{N}_s}) \tag{G.3}$$

rendering the assigned treatments conditionally independent.

Lemma G.3. *For the relation between the substitute confounder and factor models, it holds under weak regularity conditions*

1. *Assume that the true distributions of the treatments* $p(a_{1:|\mathcal{S}|}, \mathbf{a}_{\mathcal{N}_{1:|\mathcal{S}|}})$ *can be represented by a factor model employing the substitute confounder* Z*, i.e.,* $p_\phi(z_{1:|\mathcal{S}|}, \boldsymbol{x}_{1:|\mathcal{S}|}, \boldsymbol{x}_{\mathcal{N}_{1:|\mathcal{S}|}}, a_{1:|\mathcal{S}|}, \mathbf{a}_{\mathcal{N}_{1:|\mathcal{S}|}})$*. With the assumption of latent field sufficiency (see Assumption 8.5), the assigned treatments* $(a, \mathbf{a}_\mathcal{N})$ *are ignorable given* Z_s*,* $\mathbf{X}_s$*, and* $\mathbf{X}_{\mathcal{N}_s}$*, i.e.,*

$$(A_s, \mathbf{A}_{\mathcal{N}_s}) \perp\!\!\!\perp Y_s(a, \mathbf{a}_\mathcal{N}) \mid Z_s, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}. \tag{G.4}$$

2. *A factor model that represents the distribution of the assigned treatments always exists.*

Proof. The statement follows from Proposition 5 in Wang and Blei [2019]. □

G.2.2 Proof of the Main Theorem

Theorem 8.7 (Causal identifiability)**.** *Suppose Assumptions 8.2–8.6 hold. Let* Z_s *be a piecewise constant function of the assigned neighborhood exposure and covariates* $(a, \mathbf{a}_\mathcal{N}, \boldsymbol{x}, \boldsymbol{x}_\mathcal{N})$ *and let the outcome be a separable function of the observed and unobserved variables:*

$$\begin{aligned}
&\mathbb{E}_Y[Y_s(a, \mathbf{a}_\mathcal{N}) \mid \mathbf{X}_s = \boldsymbol{x}, \mathbf{X}_{\mathcal{N}_s} = \boldsymbol{x}_\mathcal{N}, Z_s = z] \\
&\qquad = f_1(a, \mathbf{a}_\mathcal{N}, \boldsymbol{x}, \boldsymbol{x}_\mathcal{N}) + f_2(z), \qquad (8.14) \\
&\mathbb{E}_Y[Y_s \mid A_s = a, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_\mathcal{N}, \mathbf{X}_s = \boldsymbol{x}, \mathbf{X}_{\mathcal{N}_s} = \boldsymbol{x}_\mathcal{N}, Z_s = z] \\
&\qquad = f_3(a, \mathbf{a}_\mathcal{N}, \boldsymbol{x}, \boldsymbol{x}_\mathcal{N}) + f_4(z), \qquad (8.15)
\end{aligned}$$

for continuously differentiable functions f_1, f_2, f_3, f_4*. Consequently, the direct and spillover effects are identifiable as*

$$\begin{aligned}
\tau_{\text{dir}} = \mathbb{E}_{\mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z}\Big[&\mathbb{E}_Y[Y_s \mid A_s = 1, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s] \\
&- \mathbb{E}_Y[Y_s \mid A_s = 0, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]\Big], \qquad (8.16)
\end{aligned}$$

$$\tau_{\text{spill}} = \mathbb{E}_{\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s},Z}\Big[\mathbb{E}_Y[Y_s \mid a, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}^{(1)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s] - \mathbb{E}_Y[Y_s \mid a, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}^{(0)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]\Big]. \tag{8.17}$$

Proof. First, observe that by the power-property and the separability of the outcome, we have

$$\mathbb{E}_Y[Y_s(a, \mathbf{a}_{\mathcal{N}})] = \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}},Z}[\mathbb{E}_Y[Y_s(a, \mathbf{a}_{\mathcal{N}}) \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]] \tag{G.5}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(a, \mathbf{a}_{\mathcal{N}}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] + \mathbb{E}_Z[f_2(Z_s)]. \tag{G.6}$$

For the direct and indirect effects τ_{dir} and τ_{ind} follows

$$\tau_{dir} = \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(A_s = 1, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] - \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(A_s = 0, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] \tag{G.7}$$

$$= \int_{C(1,0)} \nabla_\nu \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(\nu, \mathbf{a}_{\mathcal{N}}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] d\nu, \quad \nu \in \mathbb{R} \tag{G.8}$$

and

$$\tau_{ind} = \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}^{(1)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] - \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}^{(0)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] \tag{G.9}$$

$$= \int_{C(a^{(1)}_{\mathcal{N}_s}, a^{(0)}_{\mathcal{N}_s}))} \nabla_\kappa \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(a_s, \mathbf{A}_{\mathcal{N}_s} = \kappa, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] d\kappa, \quad \kappa \in \mathbb{R}^{|\mathcal{S}|-1}. \tag{G.10}$$

We thus need to find an expression for the gradient to rewrite the integral in terms of observable quantities.

To do so, we first consider the conditional expected outcome. By Assumption 8.6 there exists a function g such that $Z = g(a, \mathbf{a}_{\mathcal{N}}, \mathbf{X}, \mathbf{X}_{\mathcal{N}})$. Therefore, it holds

$$\mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}},Z}[\mathbb{E}_Y[Y_s \mid A_s = a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_{\mathcal{N}_s}, Z_s]] \tag{G.11}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[\mathbb{E}_Y[Y_s \mid A_s = a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s = g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] \tag{G.12}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[\mathbb{E}_Y[Y_s(a_s, \mathbf{a}_{\mathcal{N}_s}) \mid A_s = a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s = g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})]], \tag{G.13}$$

where the latter equality follows from Assumption 8.2.

As $Y_s(a_s, \mathbf{a}_{\mathcal{N}_s}) \perp\!\!\!\perp A_s, \mathbf{A}_{\mathcal{N}_s} \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s$ (by Lemma G.3) and the outcomes are assumed to be separable, it follows

$$\mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}},Z}[\mathbb{E}_Y[Y_s \mid A_s = a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]] \tag{G.14}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[\mathbb{E}_Y[Y_s(a_s, \mathbf{a}_{\mathcal{N}_s}) \mid \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s = g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})]] \tag{G.15}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] + \mathbb{E}_Z[f_2(g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}))]. \tag{G.16}$$

Recall that by the definition of the conditional expected outcome, we have

$$\mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}},Z}[\mathbb{E}_Y[Y_s \mid A_s = a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]] = \tag{G.17}$$

$$\mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_3(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] + \mathbb{E}_Z[f_4(g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}))]. \tag{G.18}$$

Now, we are ready to consider the gradients in G.9. Observe that for the gradients of the conditional outcome, it holds

$$\nabla_{a_s}\mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}},Z}[\mathbb{E}_Y[Y_s \mid a_s, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s]] \tag{G.19}$$

$$= \nabla_{a_s}\mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_1(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] + \nabla_{a_s}\mathbb{E}_Z[f_2(g(a_s, \mathbf{a}_{\mathcal{N}_s}))] \tag{G.20}$$

$$= \nabla_{a_s}\mathbb{E}_{\mathbf{X},\mathbf{X}_{\mathcal{N}}}[f_3(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s})] + \nabla_{a_s}\mathbb{E}_Z[f_4(g(a_s, \mathbf{a}_{\mathcal{N}_s}))] \tag{G.21}$$

with a similar expression for $\nabla_{\mathbf{a}_{\mathcal{N}_s}}$. Note that, up to a set of Lebesgue measure zero, the gradients of f_2 and f_4 disappear, i.e.,

$$\nabla_{a_s}\mathbb{E}_Z[f_2(g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}))] = \nabla_{g(a_s,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})} f_2 \nabla_{a_s} g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}) = 0 \tag{G.22}$$

and

$$\nabla_{a_s}\mathbb{E}_Z[f_4(g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}))] = \nabla_{g(a_s,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})} f_4 \nabla_{a_s} g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}) = 0 \tag{G.23}$$

as

$$\nabla_{a_s} g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}) = 0.$$

Similarly,

$$\nabla_{\mathbf{a}_{\mathcal{N}_s}}\mathbb{E}_Z[f_2(g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}))] = \nabla_{\mathbf{a}_{\mathcal{N}_s}}\mathbb{E}_Z[f_4(g(a_s, \mathbf{a}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}))] = 0.$$

Overall, we receive

$$\nabla_{a_s}\mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_1(a_s,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] = \nabla_{a_s}\mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(a_s,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] \tag{G.24}$$

and

$$\nabla_{\mathbf{a}_{\mathcal{N}_s}}\mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_1(a_s,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] = \nabla_{\mathbf{a}_{\mathcal{N}_s}}\mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(a_s,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})]. \tag{G.25}$$

Finally, we can identify the direct treatment τ_{dir} effect as

$$\tau_{dir} = \int_{C(1,0)} \nabla_\nu \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_1(\nu,\mathbf{a}_{\mathcal{N}},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})]d\nu, \quad \nu\in\mathbb{R} \tag{G.26}$$

$$= \int_{C(1,0)} \nabla_\nu \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(\nu,\mathbf{a}_{\mathcal{N}},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})]d\nu, \quad \nu\in\mathbb{R} \tag{G.27}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(A_s=1,\mathbf{a}_{\mathcal{N}},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] - \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(A_s=0,\mathbf{a}_{\mathcal{N}},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] \tag{G.28}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(A_s=1,\mathbf{a}_{\mathcal{N}},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] + \mathbb{E}_Z[f_4(Z_s)] \tag{G.29}$$

$$- \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(A_s=0,\mathbf{a}_{\mathcal{N}},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] - \mathbb{E}_Z[f_4(Z_s)] \tag{G.30}$$

$$= \mathbb{E}_{Z,\mathbf{X},\mathbf{X}_{\mathcal{N}}}\Big[\mathbb{E}_Y[Y_s \mid a_s{=}1,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s},Z_s] - \mathbb{E}_Y[Y_s \mid a_s{=}0,\mathbf{a}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s},Z_s]\Big] \tag{G.31}$$

and similarly the indirect treatment effect τ_{ind} as

$$\tau_{ind} = \int_{C(a^{(1)}_{\mathcal{N}_s},a^{(0)}_{\mathcal{N}_s})} \nabla_\kappa \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_1(a_s,\mathbf{A}_{\mathcal{N}_s}=\kappa,\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})]d\kappa \tag{G.32}$$

$$= \int_{C(a^{(1)}_{\mathcal{N}_s},a^{(0)}_{\mathcal{N}_s})} \nabla_\kappa \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(a_s,\mathbf{A}_{\mathcal{N}_s}=\kappa,\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})]d\kappa \tag{G.33}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(a_s,\mathbf{a}^{(1)}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] - \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(a_s,a^{(0)}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] \tag{G.34}$$

$$= \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(a_s,\mathbf{a}^{(1)}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] + \mathbb{E}_Z[f_4(Z_s)] \tag{G.35}$$

$$- \mathbb{E}_{\mathbf{X},\mathbf{X}_N}[f_3(a_s,\mathbf{a}^{(0)}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s})] - \mathbb{E}_Z[f_4(Z_s)] \tag{G.36}$$

$$= \mathbb{E}_{Z,\mathbf{X},\mathbf{X}_{\mathcal{N}}}\Big[\mathbb{E}_Y[Y_s \mid a_s,\mathbf{a}^{(1)}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s},Z_s] - \mathbb{E}_Y[Y_s \mid a_s,\mathbf{a}^{(0)}_{\mathcal{N}_s},\mathbf{X}_s,\mathbf{X}_{\mathcal{N}_s},Z_s]\Big] \tag{G.37}$$

Overall, we proved that the substitute confounder generated by our spatial deconfounder renders the treatment effects identifiable. □

G.2.3 Sensitivity to Proxy Error

Proposition G.4 (Sensitivity of identified effects to proxy error)**.** *Let*

$$m(a, \mathbf{a}_{\mathcal{N}}, x, x_{\mathcal{N}}, z) := \mathbb{E}[Y_s \mid A_s = a, \mathbf{A}_{\mathcal{N}_s} = \mathbf{a}_{\mathcal{N}}, \mathbf{X}_s = x, \mathbf{X}_{\mathcal{N}_s} = x_{\mathcal{N}}, Z_s^\star = z]$$

denote the (oracle) outcome regression indexed by the population substitute confounder $Z_s^\star$. Assume that m is L-Lipschitz in z, uniformly over $(a, \mathbf{a}_{\mathcal{N}}, x, x_{\mathcal{N}})$:

$$\left|m(a, \mathbf{a}_{\mathcal{N}}, x, x_{\mathcal{N}}, z) - m(a, \mathbf{a}_{\mathcal{N}}, x, x_{\mathcal{N}}, z')\right| \le L\|z - z'\| \qquad \forall z, z'.$$

Define the identified direct and spillover effect functionals (cf. Eqs. (8.3) *and* (8.4)*) evaluated at a generic proxy W_s by*

$$\tau_{\text{dir}}(W) := \mathbb{E}\Big[m(1, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, W_s) - m(0, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, W_s)\Big],$$
$$\tau_{\text{spill}}(W) := \mathbb{E}\Big[m(a, \mathbf{a}^{(1)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, W_s) - m(a, \mathbf{a}^{(0)}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, W_s)\Big].$$

Then, for any proxy $\hat{Z}_s$,

$$\left|\tau_{\text{dir}}(\hat{Z}) - \tau_{\text{dir}}(Z^\star)\right| \le 2L\,\mathbb{E}\Big[\|\hat{Z}_s - Z_s^\star\|\Big], \qquad \left|\tau_{\text{spill}}(\hat{Z}) - \tau_{\text{spill}}(Z^\star)\right| \le 2L\,\mathbb{E}\Big[\|\hat{Z}_s - Z_s^\star\|\Big].$$

In particular, if $\mathbb{E}\|\hat{Z}_s - Z_s^\star\| \to 0$, then $\tau_{\text{dir}}(\hat{Z}) \to \tau_{\text{dir}}(Z^\star)$ and $\tau_{\text{spill}}(\hat{Z}) \to \tau_{\text{spill}}(Z^\star)$.

Proof. For the direct effect,

$$\begin{aligned}
&\left|\tau_{\text{dir}}(\hat{Z}) - \tau_{\text{dir}}(Z^\star)\right| \\
&= \Big|\mathbb{E}\Big[m(1, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s) - m(0, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s) \\
&\qquad\qquad - m(1, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s^\star) + m(0, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s^\star)\Big]\Big| \\
&\le \mathbb{E}\Big[\left|m(1, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s) - m(1, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s^\star)\right|\Big] \\
&\quad + \mathbb{E}\Big[\left|m(0, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, \hat{Z}_s) - m(0, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}, Z_s^\star)\right|\Big] \\
&\le L\,\mathbb{E}\Big[\|\hat{Z}_s - Z_s^\star\|\Big] + L\,\mathbb{E}\Big[\|\hat{Z}_s - Z_s^\star\|\Big] = 2L\,\mathbb{E}\Big[\|\hat{Z}_s - Z_s^\star\|\Big],
\end{aligned}$$

where the first inequality is the triangle inequality and the second uses the Lipschitz condition. The spillover bound is identical, replacing $(1, \mathbf{A}_{\mathcal{N}_s})$ and $(0, \mathbf{A}_{\mathcal{N}_s})$ by $(a, \mathbf{a}^{(1)}_{\mathcal{N}_s})$ and $(a, \mathbf{a}^{(0)}_{\mathcal{N}_s})$. □

G.3 Implementation Details

This section provides implementation details for our experimental setup. We cover four aspects:

1. **Semi-synthetic data generation:** construction of counterfactual outcomes under interference and spatial confounding using the `SpaCE` benchmark framework, with hidden confounders simulated by masking key covariates.
2. **Predictive model:** how the outcome model f is estimated with ensembles of machine-learning models, including convolutional networks for spatial structure.
3. **Software and hyperparameters:** the AutoML framework used for training and tuning, along with default settings.
4. **Benchmarks:** implementation details for baseline methods.

Semi-Synthetic Outcomes Recall from Section 8.6 that we construct counterfactual outcomes via

$$\hat{Y}_s = f(A_s, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s) + R_s \quad \text{or} \quad \hat{Y}_s = f(A_s, \mathbf{A}_{\mathcal{N}_s}, \mathbf{X}_s, \mathbf{X}_{\mathcal{N}_s}) + R_s,$$

where f is a predictive model learned from real-world environmental data and R_s are exogenous, spatially correlated residuals with the same distribution as the endogenous residuals.

Predictive Model with Interference We estimate the function f using ensembles of machine-learning models, with ensemble weights determined by predictive accuracy on held-out validation data. Following Tec et al. [2024] and the benchmarking guidelines of Curth et al. [2021], this avoids bias toward causal estimators tied to a single model class. To capture spatial structure, we include ResNet-18 [He et al., 2016] as one of the base learners. Training and hyperparameter tuning are automated with the `AutoGluon` Python package

[Erickson et al., 2020], which performs model selection, hyperparameter search, and overfitting control with minimal human intervention. Default settings for `AutoGluon` are summarized in Table G.1.

Table G.1: Hyperparameters used in AutoML for the Spatial Deconfounder experiments.

Parameter	Value
`package`	`AutoGluon v1.4.0`
`fit.presets`	`good_quality`
`fit.tuning_data`	custom split using Algorithm G.1
`fit.use_bag_holdout`	true
`fit.time_limit`	null
`feature_importance.time_limit`	900
`AG_AUTOMM.optim.max_epochs`	10
`AG_AUTOMM.model.timm_image.checkpoint_name`	`resnet18`

Spatially-Aware Train-Validation Split. We implement a *spatially-aware* train-validation data split [Roberts et al., 2017] that takes interference into account to avoid overfitting due to spatial correlations. We only consider nodes with complete neighborhoods for training and validation. This spatial splitting strategy identifies a limited number of validation nodes and applies breadth-first search to exclude their adjacent neighbors from the training dataset. For this study, we define each grid cell to have edges connecting it to its 8 surrounding cells. This algorithm is described in Algorithm G.1.

Synthetic Residual Generation Following the approach established in Tec et al. [2024], we generate synthetic residuals using a Gaussian Markov Random Field (GMRF) from a spatial graph. Specifically, we sample the synthetic residuals according to: $\boldsymbol{R} \sim_{\text{iid}} \text{MultivariateNormal}(\mathbf{0}, \hat{\lambda}(\mathbb{D} - \hat{\rho}\mathbb{A}\mathbb{D})^{-1})$, where $\mathbb{A}$ represents the spatial graph's adjacency matrix, $\mathbb{D}$ denotes a diagonal matrix containing the

Algorithm G.1 Spatially-aware validation split selection with radius and complete neighborhoods

Input: Graph as a map of neighbors $s \mapsto \mathbb{N}_s$, where $\mathbb{N}_s \subset \mathbb{S}$ is the set of neighbors of s

Params: Fraction α of seed validation points (default $\alpha = 0.02$); number of BFS levels L to include in the validation set (default $L = 1$); buffer size B indicating the number of BFS levels to leave outside training and validation (default $B = 1$); radius r_m of the model used when determining the split (default $r_m = 1$)

Output: Training nodes $\mathbb{T} \subset \mathbb{S}$ and validation nodes $\mathbb{V} \subset \mathbb{S}$

 # Helper function to check whether a node has a complete r-hop neighborhood

function HASCOMPLETENEIGHBORHOOD(s, r)
    expected_count $\leftarrow (2r + 1)^2$ ▷ for a square grid
    actual_neighbors $\leftarrow$ GetNeighborsWithinRadius(s, r)
    return |actual_neighbors| = expected_count
end function

 # Filter to nodes with complete neighborhoods

$\mathbb{S}_{\text{valid}} \leftarrow \{ s \in \mathbb{S} : \text{HASCOMPLETENEIGHBORHOOD}(s, r_m) \}$

 # Initialize validation set with seed nodes from valid nodes only

$\mathbb{V} \leftarrow \text{SampleWithoutReplacement}(\mathbb{S}_{\text{valid}}, \alpha)$

 # Expand validation set with neighbors

for $\ell \in \{0, \ldots, L - 1\}$ **do**
    tmp $\leftarrow \mathbb{V}$
    for $s \in$ tmp **do**
        $\mathbb{V} \leftarrow \mathbb{V} \cup \mathbb{N}_s$
    end for
end for

 # Compute buffer

$\mathbb{B} \leftarrow \mathbb{V}$
for $b \in \{0, \ldots, B - 1 + r_m\}$ **do**
    tmp $\leftarrow \mathbb{B}$
    for $s \in$ tmp **do**
        $\mathbb{B} \leftarrow \mathbb{B} \cup \mathbb{N}_s$
    end for
end for

 # Exclude buffer from training set

$\mathbb{T} \leftarrow \mathbb{S}_{\text{valid}} \setminus \mathbb{B}$
return $\mathbb{T}, \mathbb{V}$

degree (number of neighbors) for each spatial location, $\hat{\rho}$ parameterizes the spatial dependence between observations and their neighbors (estimated from the true residuals obtained from f), and $\hat{\lambda}$ is calibrated to preserve the exact variance of the observed residuals. See Tec et al. [2024] for additional details.

Model	Iterations	Tuning Metric	Value
C-VAE-SPATIAL+	100	weight_decay_C-VAE	loguniform between 1e-4 and 1e-3
		beta_max (β) ($r_d = 1, PM_{2.5}$)	loguniform between 1e-8 and 10
		beta_max (β) ($r_d = 1, SO_4$)	loguniform between 1e-5 and 10
		beta_max (β) ($r_d = 2$)	loguniform between 1e-5 and 1e-4
		lam_t	loguniform between 1e-5 and 1.0
		lam_y	loguniform between 1e-5 and 1.0
C-VAE-UNET	60	weight_decay_C-VAE	loguniform between 1e-4 and 1e-3
		beta_max (β)	loguniform between 1e-3 and 1
		weight_decay_head	loguniform between 1e-4 and 1e-3
		unet_base_chan	16 or 32
DAPSM	N/A	propensity_score_penalty_value	choose from [0.001, 0.01, 0.1, 1.0]
		propensity_score_penalty_type	l1 or l2
		spatial_weight	uniform between 0.0 and 1.0
GCNN	N/A	hidden_dim	16 or 32
		hidden_layers	1 or 2
		weight_decay	loguniform between 1e-6 and 1e-1
		lr	1e-3 or 3e-4
		epochs	1000 or 2500
		dropout	loguniform between 1e-3 to 0.5
SPATIAL+	2,500	lam_t	loguniform between 1e-5 and 1.0
		lam_y	loguniform between 1e-5 and 1.0
SPATIAL	2,500	lam	loguniform between 1e-5 and 1.0
UNET	50	unet_base_chan	choose from [8, 16, 32]

Table G.2: Hyperparameter configurations evaluated on the validation set for each Spatial Deconfounder model, with the number of Ray Tune trials reported for each model.

Benchmark Training and Hyperparameter Tuning To ensure a fair comparison, we use the RAY TUNE [Liaw et al., 2018] framework for hyperparameter tuning. For all but DAPSM, the tuning metric is implemented as mean-squared error

(MSE) from a validation set obtained with the spatially-aware splitting method in Algorithm G.1. We use this splitting algorithm for computing the tuning metric since random splitting would result in extreme overfitting [Roberts et al., 2017]. For DAPSM we use the covariate balance criterion following Papadogeorgou et al. [2019a]. After selecting the best hyperparameters, the method is retrained on the full data. Table G.2 summarizes our hyperparameter search space for different baseline models. For C-VAE models with radius R evaluated on a dataset of radius r_d, training and validation are restricted to nodes with radius $r_m = \max(r_d, \mathrm{R})$. Each C-VAE model also specifies a latent confounder dimension $d_Z \in \{1, 2, 4, 8, 16, 32\}$. The licenses of the data sources used for training are summarized in the supplement of Tec et al. [2024], which allow sharing and reuse for non-commercial purposes.

G.4 Further Experimental Results

Our full experimental results are available for local confounding and spatial confounding at Table G.3 and Table G.4, respectively. There is a general pattern that C-VAE models tend to outperform benchmarks in estimating direct effects. In particular, C-VAE are the only local confounding methods that can also estimate spillover effects. In spatial confounding datasets with $r_d = 1$, deconfounders tend to have better direct effect and spillover estimation than UNET. We also plot the latent space of the C-VAE in Figure G.1 to show that it recovers the large-scale spatial structure of the true confounder.

Table G.3: Performance of the Spatial Deconfounder and baselines under *local confounding*. Results averaged over 10 runs with 95% confidence intervals. r_d: neighborhood radius in data generation; R: neighborhood radius used by the deconfounder. Lower values for ATE and SPILL indicate less bias; p indicates the predictive-check p-value, with values near 0.5 indicating good model fit.

Setup	Confounder	Method	DIR	SPILL	p
$PM_{2.5} \rightarrow m$ ($r_d = 1$)	ρ_{pop}	C-VAE-SPATIAL+ (R=0)	0.15 ± 0.11	n/a	0.36 ± 0.07
		C-VAE-SPATIAL+ (R=1)	0.05 ± 0.02	0.34 ± 0.08	0.35 ± 0.09
		C-VAE-SPATIAL+ (R=2)	0.07 ± 0.02	0.52 ± 0.08	0.35 ± 0.03
		DAPSM	0.25 ± 0.01	n/a	n/a
		GCNN	0.36 ± 0.03	n/a	n/a
		S2SLS-LAG1	**0.03 ± 0.00**	n/a	n/a
		SPATIAL+	0.13 ± 0.04	n/a	n/a
		SPATIAL	0.10 ± 0.07	n/a	n/a
	q_{summer}	C-VAE-SPATIAL+ (R=0)	0.15 ± 0.07	n/a	0.38 ± 0.08
		C-VAE-SPATIAL+ (R=1)	**0.04 ± 0.01**	**0.42 ± 0.08**	0.37 ± 0.07
		C-VAE-SPATIAL+ (R=2)	0.04 ± 0.01	0.44 ± 0.09	0.36 ± 0.04
		DAPSM	0.30 ± 0.03	n/a	n/a
		GCNN	0.41 ± 0.03	n/a	n/a
		S2SLS-LAG1	0.20 ± 0.00	n/a	n/a
		SPATIAL+	0.13 ± 0.04	n/a	n/a
		SPATIAL	0.10 ± 0.07	n/a	n/a
$PM_{2.5} \rightarrow m$ ($r_d = 2$)	ρ_{pop}	C-VAE-SPATIAL+ (R=0)	0.11 ± 0.02	n/a	0.35 ± 0.03
		C-VAE-SPATIAL+ (R=1)	0.05 ± 0.02	0.15 ± 0.05	0.34 ± 0.04
		C-VAE-SPATIAL+ (R=2)	**0.04 ± 0.03**	0.24 ± 0.06	0.35 ± 0.04
		DAPSM	0.16 ± 0.01	n/a	n/a
		GCNN	0.18 ± 0.03	n/a	n/a
		S2SLS-LAG1	0.07 ± 0.00	n/a	n/a
		SPATIAL+	0.10 ± 0.02	n/a	n/a
		SPATIAL	0.17 ± 0.03	n/a	n/a

	q_{summer}	C-VAE-SPATIAL+ (R=0)	0.13 ± 0.05	n/a	0.36 ± 0.04
		C-VAE-SPATIAL+ (R=1)	**0.04 ± 0.02**	**0.11 ± 0.05**	0.36 ± 0.04
		C-VAE-SPATIAL+ (R=2)	0.07 ± 0.02	0.19 ± 0.06	0.36 ± 0.04
		DAPSM	0.20 ± 0.01	n/a	n/a
		GCNN	0.16 ± 0.05	n/a	n/a
		S2SLS-LAG1	0.09 ± 0.00	n/a	n/a
		SPATIAL+	0.11 ± 0.02	n/a	n/a
		SPATIAL	0.17 ± 0.03	n/a	n/a
$SO_4 \rightarrow PM_{2.5}$ ($r_d = 1$)	NH_4	C-VAE-SPATIAL+ (R=0)	0.22 ± 0.04	n/a	0.40 ± 0.05
		C-VAE-SPATIAL+ (R=1)	**0.07 ± 0.03**	**0.64 ± 0.10**	0.38 ± 0.04
		C-VAE-SPATIAL+ (R=2)	**0.07 ± 0.03**	0.16 ± 0.06	0.39 ± 0.06
		DAPSM	1.44 ± 0.00	n/a	n/a
		GCNN	0.52 ± 0.16	n/a	n/a
		S2SLS-LAG1	0.09 ± 0.00	n/a	n/a
		SPATIAL+	0.11 ± 0.03	n/a	n/a
		SPATIAL	0.08 ± 0.02	n/a	n/a
	OC	C-VAE-SPATIAL+ (R=0)	0.07 ± 0.03	n/a	0.41 ± 0.02
		C-VAE-SPATIAL+ (R=1)	0.08 ± 0.03	**0.69 ± 0.10**	0.41 ± 0.03
		C-VAE-SPATIAL+ (R=2)	0.11 ± 0.04	0.90 ± 0.12	0.44 ± 0.02
		DAPSM	1.45 ± 0.00	n/a	n/a
		GCNN	0.77 ± 0.22	n/a	n/a
		S2SLS-LAG1	**0.00 ± 0.00**	n/a	n/a
		SPATIAL+	0.11 ± 0.03	n/a	n/a
		SPATIAL	0.08 ± 0.02	n/a	n/a
$SO_4 \rightarrow PM_{2.5}$ ($r_d = 2$)	NH_4	C-VAE-SPATIAL+ (R=0)	**0.07 ± 0.04**	n/a	0.48 ± 0.06
		C-VAE-SPATIAL+ (R=1)	0.08 ± 0.03	0.13 ± 0.05	0.44 ± 0.03
		C-VAE-SPATIAL+ (R=2)	0.12 ± 0.04	**0.09 ± 0.04**	0.43 ± 0.03
		DAPSM	1.23 ± 0.00	n/a	n/a
		GCNN	0.26 ± 0.09	n/a	n/a
		S2SLS-LAG1	0.10 ± 0.00	n/a	n/a
		SPATIAL+	0.13 ± 0.07	n/a	n/a
		SPATIAL	0.29 ± 0.01	n/a	n/a
	OC	C-VAE-SPATIAL+ (R=0)	0.10 ± 0.07	n/a	0.43 ± 0.04

C-VAE-SPATIAL+ (R=1)	**0.06 ± 0.03**	**0.18 ± 0.09**	0.43 ± 0.03
C-VAE-SPATIAL+ (R=2)	0.12 ± 0.06	0.35 ± 08	0.43 ± 0.04
DAPSM	1.24 ± 0.01	n/a	n/a
GCNN	0.30 ± 0.10	n/a	n/a
S2SLS-LAG1	0.21 ± 0.00	n/a	n/a
SPATIAL+	0.13 ± 0.07	n/a	n/a
SPATIAL	0.29 ± 0.01	n/a	n/a

Table G.4: Performance of the Spatial Deconfounder and baselines under *spatial confounding*. Results averaged over 10 runs with 95% confidence intervals. r_d: neighborhood radius in data generation; R: neighborhood radius used by the deconfounder. Lower values for ATE and SPILL indicate less bias. p indicates the predictive-check p-value, with values near 0.5 indicating good model fit.

Environment	Confounder	Method	DIR	SPILL	p
$PM_{2.5} \rightarrow m$ ($r_d = 1$)	ρ_{pop}	C-VAE-UNET (R=0)	0.11 ± 0.04	n/a	0.34 ± 0.04
		C-VAE-UNET (R=1)	0.05 ± 0.01	0.22 ± 0.06	0.34 ± 0.03
		C-VAE-UNET (R=2)	**0.04 ± 0.02**	**0.12 ± 0.06**	0.36 ± 0.06
		DAPSM	0.20 ± 0.01	n/a	n/a
		GCNN	0.17 ± 0.06	n/a	n/a
		S2SLS-LAG1	0.05 ± 0.00	n/a	n/a
		SPATIAL+	0.27 ± 0.18	n/a	n/a
		SPATIAL	0.06 ± 0.06	n/a	n/a
		UNET	0.06 ± 0.01	0.17 ± 0.04	n/a
	q_{summer}	C-VAE-UNET (R=0)	**0.04 ± 0.02**	n/a	0.35 ± 0.02
		C-VAE-UNET (R=1)	0.06 ± 0.02	0.13 ± 0.07	0.33 ± 0.02
		C-VAE-UNET (R=2)	**0.04 ± 0.02**	**0.10 ± 0.05**	0.36 ± 0.05
		DAPSM	0.28 ± 0.04	n/a	n/a
		GCNN	0.23 ± 0.03	n/a	n/a
		S2SLS-LAG1	0.16 ± 0.00	n/a	n/a

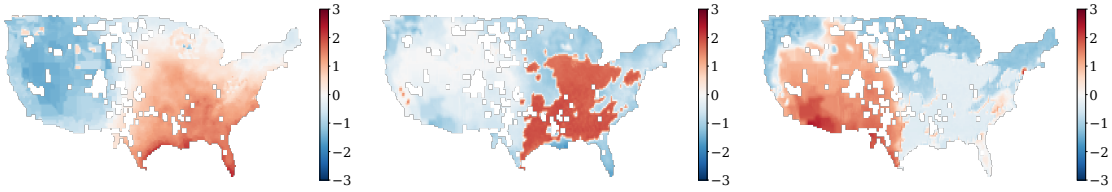
		SPATIAL+	0.27 ± 0.18	n/a	n/a
		SPATIAL	0.07 ± 0.06	n/a	n/a
		UNET	**0.04 ± 0.01**	**0.10 ± 0.05**	n/a
$PM_{2.5} \rightarrow m\ (r_d = 2)$	ρ_{pop}	C-VAE-UNET (R=0)	0.09 ± 0.03	n/a	0.32 ± 0.04
		C-VAE-UNET (R=1)	0.15 ± 0.01	**0.09 ± 0.03**	0.31 ± 0.04
		C-VAE-UNET (R=2)	0.15 ± 0.01	0.13 ± 0.05	0.29 ± 0.06
		DAPSM	0.15 ± 0.02	n/a	n/a
		GCNN	0.15 ± 0.04	n/a	n/a
		S2SLS-LAG1	0.06 ± 0.00	n/a	n/a
		SPATIAL+	0.08 ± 0.04	n/a	n/a
		SPATIAL	**0.05 ± 0.02**	n/a	n/a
		UNET	0.15 ± 0.01	0.15 ± 0.03	n/a
	q_{summer}	C-VAE-UNET (R=0)	**0.05 ± 0.01**	n/a	0.30 ± 0.05
		C-VAE-UNET (R=1)	0.14 ± 0.01	0.07 ± 0.03	0.30 ± 0.05
		C-VAE-UNET (R=2)	0.15 ± 0.01	**0.06 ± 0.03**	0.33 ± 0.04
		DAPSM	0.21 ± 0.01	n/a	n/a
		GCNN	0.23 ± 0.03	n/a	n/a
		S2SLS-LAG1	0.10 ± 0.00	n/a	n/a
		SPATIAL+	0.07 ± 0.03	n/a	n/a
		SPATIAL	**0.05 ± 0.02**	n/a	n/a
		UNET	0.15 ± 0.00	0.08 ± 0.04	n/a
$SO_4 \rightarrow PM_{2.5}\ (r_d = 1)$	NH_4	C-VAE-UNET (R=0)	0.18 ± 0.03	n/a	0.44 ± 0.03
		C-VAE-UNET (R=1)	0.05 ± 0.02	0.22 ± 0.03	0.45 ± 0.03
		C-VAE-UNET (R=2)	**0.04 ± 0.02**	0.37 ± 0.06	0.43 ± 0.03
		DAPSM	1.56 ± 0.00	n/a	n/a
		GCNN	0.55 ± 0.09	n/a	n/a
		S2SLS-LAG1	0.22 ± 0.00	n/a	n/a
		SPATIAL+	0.06 ± 0.05	n/a	n/a
		SPATIAL	**0.04 ± 0.01**	n/a	n/a
		UNET	**0.04 ± 0.01**	**0.19 ± 0.04**	n/a
	OC	C-VAE-UNET (R=0)	**0.04 ± 0.02**	n/a	0.46 ± 0.02
		C-VAE-UNET (R=1)	0.06 ± 0.02	0.09 ± 0.04	0.44 ± 0.03
		C-VAE-UNET (R=2)	0.06 ± 0.02	0.18 ± 0.06	0.45 ± 0.03

		DAPSM	1.57 ± 0.00	n/a	n/a
		GCNN	0.42 ± 0.15	n/a	n/a
		S2SLS-LAG1	0.13 ± 0.00	n/a	n/a
		SPATIAL+	0.06 ± 0.05	n/a	n/a
		SPATIAL	**0.04 ± 0.01**	n/a	n/a
		UNET	0.07 ± 0.02	**0.05 ± 0.02**	n/a
$SO_4 \rightarrow PM_{2.5}$ ($r_d = 2$)	NH_4	C-VAE-UNET (R=0)	**0.04 ± 0.02**	n/a	0.43 ± 0.04
		C-VAE-UNET (R=1)	0.13 ± 0.02	**0.05 ± 0.02**	0.45 ± 0.03
		C-VAE-UNET (R=2)	0.15 ± 0.01	0.07 ± 0.03	0.45 ± 0.03
		DAPSM	1.47 ± 0.00	n/a	n/a
		GCNN	0.66 ± 0.21	n/a	n/a
		S2SLS-LAG1	0.16 ± 0.00	n/a	n/a
		SPATIAL+	0.06 ± 0.02	n/a	n/a
		SPATIAL	0.06 ± 0.05	n/a	n/a
		UNET	0.15 ± 0.01	0.11 ± 0.04	n/a
	OC	C-VAE-UNET (R=0)	**0.04 ± 0.02**	n/a	0.43 ± 0.02
		C-VAE-UNET (R=1)	0.12 ± 0.02	**0.06 ± 0.03**	0.43 ± 0.03
		C-VAE-UNET (R=2)	0.13 ± 0.03	0.07 ± 0.02	0.44 ± 0.03
		DAPSM	1.49 ± 0.01	n/a	n/a
		GCNN	0.67 ± 0.12	n/a	n/a
		S2SLS-LAG1	0.09 ± 0.00	n/a	n/a
		SPATIAL+	0.05 ± 0.02	n/a	n/a
		SPATIAL	0.06 ± 0.05	n/a	n/a
		UNET	0.15 ± 0.01	0.08 ± 0.04	n/a

G.5 Additional Robustness Tests

G.5.1 Treatment Sparsity

The results in Table G.5 examine our method under sparse treatment conditions with 30% and 10% of grid cells receiving treatment. Despite similar performance under moderate treatment sparsity (30%), C-VAE-SPATIAL+ consid-



(a) Summer humidity $U(s)$ (b) PC1 of latent space Z_s (c) PC2 of latent space Z_s

Figure G.1: Reconstructed latent confounder in the Spatial Deconfounder experiment. The first principal component of the learned latent representation captures treatment variation, while the second principal component recovers large-scale structure of the true unobserved spatial confounder.

erably outperforms SPATIAL+ when sparsity is extreme (10%), underscoring the value of our framework for direct effect estimation in highly sparse conditions. In addition, the predictive p-value is lower as treatment sparsity increases, showing worse model calibration in sparse settings.

Table G.5: Performance of the Spatial Deconfounder and baselines under **sparse** *local confounding*. Results averaged over 10 runs with 95% confidence intervals. r_d: neighborhood radius in data generation; R: neighborhood radius used by the deconfounder. Lower values for ATE and SPILL indicate less bias. p indicates the predictive p-value, with values near 0.5 indicating good model fit. Percentage in environment denotes the fraction of observations receiving treatment.

Setup	Confounder	Method	DIR	SPILL	p
$SO_4 \rightarrow PM_{2.5}$ ($r_d = 1$) (10%)	NH_4	C-VAE-SPATIAL+ (R=0)	0.07 ± 0.04	n/a	0.28 ± 0.02
		C-VAE-SPATIAL+ (R=1)	0.19 ± 0.08	**0.80 ± 0.23**	0.28 ± 0.01
		C-VAE-SPATIAL+ (R=2)	0.14 ± 0.08	1.20 ± 0.12	0.29 ± 0.01
		DAPSM	**0.02 ± 0.00**	n/a	n/a
		GCNN	0.42 ± 0.07	n/a	n/a
		S2SLS-LAG1	0.04 ± 0.00	n/a	n/a
		SPATIAL+	0.68 ± 0.21	n/a	n/a
		SPATIAL	0.17 ± 0.11	n/a	n/a

	OC	C-VAE-SPATIAL+ (R=0)	**0.05 ± 0.03**	n/a	0.27 ± 0.01
		C-VAE-SPATIAL+ (R=1)	0.14 ± 0.05	**0.71 ± 0.17**	0.29 ± 0.02
		C-VAE-SPATIAL+ (R=2)	0.08 ± 0.03	1.09 ± 0.18	0.30 ± 0.02
		DAPSM	**0.05 ± 0.02**	n/a	n/a
		GCNN	0.69 ± 0.20	n/a	n/a
		S2SLS-LAG1	0.26 ± 0.00	n/a	n/a
		SPATIAL+	0.55 ± 0.19	n/a	n/a
		SPATIAL	0.17 ± 0.11	n/a	n/a
$SO_4 \rightarrow PM_{2.5}$ ($r_d = 1$) (30%)	NH_4	C-VAE-SPATIAL+ (R=0)	0.14 ± 0.03	n/a	0.33 ± 0.02
		C-VAE-SPATIAL+ (R=1)	0.18 ± 0.06	0.42 ± 0.11	0.35 ± 0.03
		C-VAE-SPATIAL+ (R=2)	0.12 ± 0.07	**0.25 ± 0.11**	0.34 ± 0.02
		DAPSM	1.00 ± 0.00	n/a	n/a
		GCNN	0.34 ± 0.12	n/a	n/a
		S2SLS-LAG1	**0.03 ± 0.00**	n/a	n/a
		SPATIAL+	0.12 ± 0.05	n/a	n/a
		SPATIAL	0.16 ± 0.03	n/a	n/a
	OC	C-VAE-SPATIAL+ (R=0)	0.13 ± 0.03	n/a	0.31 ± 0.03
		C-VAE-SPATIAL+ (R=1)	0.15 ± 0.06	0.35 ± 0.09	0.35 ± 0.02
		C-VAE-SPATIAL+ (R=2)	0.11 ± 0.05	**0.27 ± 0.10**	0.36 ± 0.03
		DAPSM	1.00 ± 0.00	n/a	n/a
		GCNN	0.35 ± 0.14	n/a	n/a
		S2SLS-LAG1	**0.07 ± 0.00**	n/a	n/a
		SPATIAL+	0.12 ± 0.05	n/a	n/a
		SPATIAL	0.15 ± 0.03	n/a	n/a

G.5.2 Performance Under Single-Cause Confounders

We evaluate our method under violation of Assumption 8.5 by introducing a localized single-cause unobserved confounder named SC. We select $C = \{c_1, \ldots, c_n\}$ as cluster centers, drawn uniformly from the set of spatial sites, where $n = \lceil s|\mathcal{S}| \rceil$ and s denotes the sparsity. Each cluster center is assigned a peak in-

tensity $\alpha_c \sim U(0.5, 1.0)$. for any site s, the resulting single-cause confounder is

$$SC_s = \max_{c \in C} \alpha_c \exp\left(-\frac{d(s,c)}{2}\right)$$

where $d(s,c)$ is the shortest distance path between s and c. We then inject SC into both the treatment and outcome by adding $0.8 \times \text{std}(X) \times SC$ to each variable where X denotes the respective treatment or outcome variable. The treatments are binarized by applying a threshold.

Table G.6 presents the performance of our methods when Assumption 8.5 is violated. When the unobserved confounder exhibits greater localization (10%), C-VAE-SPATIAL+ shows larger bias in the direct effect estimate compared to SPATIAL+. However, with a moderately sparse unobserved confounder, C-VAE-SPATIAL+ achives comparable performance to SPATIAL+.

Table G.6: Performance of the Spatial Deconfounder and baselines under *local confounding* with **single-cause unobserved confounder** SC. Results averaged over 10 runs with 95% confidence intervals. r_d: neighborhood radius in data generation; R: neighborhood radius used by the deconfounder. Lower values for ATE and SPILL indicate less bias. p indicates the predictive p-value, with values near 0.5 indicating good model fit. Percentage in environment denotes the fraction of observations receiving treatment.

Setup	Confounder	Method	DIR	SPILL	p
$PM_{2.5} \rightarrow m$ ($r_d = 1$) (10%)	SC	C-VAE-SPATIAL+ (R=0)	0.11 ± 0.08	n/a	0.40 ± 0.02
		C-VAE-SPATIAL+ (R=1)	0.11 ± 0.06	**0.44 ± 0.14**	0.40 ± 0.02
		C-VAE-SPATIAL+ (R=2)	0.08 ± 0.02	0.62 ± 0.07	0.40 ± 0.03
		DAPSM	0.52 ± 0.01	n/a	n/a
		GCNN	0.13 ± 0.03	n/a	n/a
		S2SLS-LAG1	0.20 ± 0.00	n/a	n/a

		SPATIAL+	**0.04 ± 0.01**	n/a	n/a
		SPATIAL	0.06 ± 0.07	n/a	n/a
$PM_{2.5} \rightarrow m\ (r_d = 1)$ (30%)	SC	C-VAE-SPATIAL+ (R=0)	**0.07 ± 0.02**	n/a	0.38 ± 0.02
		C-VAE-SPATIAL+ (R=1)	0.08 ± 0.02	**0.26 ± 0.07**	0.39 ± 0.03
		C-VAE-SPATIAL+ (R=2)	0.10 ± 0.04	1.14 ± 1.37	0.42 ± 0.05
		DAPSM	0.58 ± 0.00	n/a	n/a
		GCNN	0.16 ± 0.05	n/a	n/a
		S2SLS-LAG1	0.23 ± 0.00	n/a	n/a
		SPATIAL+	0.09 ± 0.01	n/a	n/a
		SPATIAL	0.08 ± 0.02	n/a	n/a

APPENDIX H

APPENDIX FOR CHAPTER 9

H.1 Notation

Table H.1: Notation used in Chapter 9.

Notation	Meaning
$\mathcal{G}$	Network with N nodes
$\mathcal{N}, \mathcal{E}$	Node set and edge set of $\mathcal{G}$
$\mathcal{N}_i,\ \mathcal{N}_{-i}$	Neighborhood of unit i and its complement in the network
n_i	Degree of node i, i.e., number of neighbors of i
$T_i,\ T_{\mathcal{N}_i}$	Binary treatment of unit i and treatment vector of its neighbors
$\mathbf{X}$	Covariates, taking values in $\mathcal{X}$
Y	Outcome, taking values in $\mathcal{Y}$
g	Analyst-specified exposure mapping, $g : [0,1]^n \to \mathcal{Z}$
g^*	True exposure mapping
Z	Neighborhood exposure summary, $Z = g(T_{\mathcal{N}})$
$Y(t,z)$	Potential outcome under individual treatment t and neighborhood exposure z
$\psi(t,z)$	Average potential outcome (APO)
$\mu(t,z,\mathbf{x})$	Conditional average potential outcome (CAPO)
$Q^{\pm}(t,z,\mathbf{x})$	Upper and lower conditional quantiles induced by the sensitivity bounds $b^{\pm}(z,\mathbf{x})$
$b^{\pm}(z,\mathbf{x})$	Upper and lower bounds on the exposure-propensity shift due to misspecification
$\gamma_u^{\pm}(t,z,\mathbf{x})$	Upper tail conditional moment, $\mathbb{E}[(Y - Q^{\pm})_+ \mid t,z,\mathbf{x}]$
$\gamma_l^{\pm}(t,z,\mathbf{x})$	Lower tail conditional moment, $\mathbb{E}[(Q^{\pm} - Y)_+ \mid t,z,\mathbf{x}]$
$\pi^t(\mathbf{x})$	Unit-level propensity score
$\pi^g(z \mid \mathbf{x})$	Exposure propensity induced by the mapping g
η	Collection of nuisance functions
$\phi_{t,z}^{+}(S;\eta),\ \phi_{t,z}^{-}(S;\eta)$	Upper and lower orthogonal pseudo-outcomes
$\widehat{\mu}^{\pm}(t,z,\mathbf{x})$	Estimated upper and lower CAPO bounds
$\widehat{\psi}^{\pm}(t,z)$	Estimated upper and lower APO bounds

H.2 Extended theory

H.2.1 Summary of Bounds

Potential outcomes	$\mu^+(t,z,\mathbf{x}) = Q^+(t,z,\mathbf{x}) + \frac{1}{b^-(z,\mathbf{x})}\gamma_u^+(t,z,\mathbf{X}) - \frac{1}{b^+(z,\mathbf{x})}\gamma_u^-(t,z,\mathbf{X})$ $\mu^-(t,z,\mathbf{x}) = Q^-(t,z,\mathbf{x}) + \frac{1}{b^+(z,\mathbf{x})}\gamma_u^-(t,z,\mathbf{X}) - \frac{1}{b^-(z,\mathbf{x})}\gamma_l^-(t,z,\mathbf{X})$
Pseudo-outcomes (discrete)	$\phi_{t,z}^+(S;\widehat{\eta}) = \widehat{Q}^+(t,z,\mathbf{X}) + \frac{\widehat{\gamma}_u^+(t,z,\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\widehat{\gamma}_l^+(t,z,\mathbf{X})}{b^+(z,\mathbf{X})}$ $+\frac{\mathbf{1}_{[T=t]}\,\mathbf{1}_{[Z=z]}}{\widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(Z\|\mathbf{X})}\left[\frac{(Y-\widehat{Q}^+(t,Z,\mathbf{X}))_+-\widehat{\gamma}_u^+(t,Z,\mathbf{X})}{b^-(Z,\mathbf{X})} - \frac{(\widehat{Q}^+(t,Z,\mathbf{X})-Y)_+-\widehat{\gamma}_l^+(t,Z,\mathbf{X})}{b^+(Z,\mathbf{X})}\right]$ $\phi_{t,z}^-(S;\widehat{\eta}) = \widehat{Q}^-(t,z,\mathbf{X}) + \frac{\widehat{\gamma}_u^-(t,z,\mathbf{X})}{b^+(z,\mathbf{X})} - \frac{\widehat{\gamma}_l^-(t,z,\mathbf{X})}{b^-(z,\mathbf{X})}$ $+\frac{\mathbf{1}_{[T=t]}\,\mathbf{1}_{[Z=z]}}{\widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(Z\|\mathbf{X})}\left[\frac{(Y-\widehat{Q}^-(t,Z,\mathbf{X}))_+-\widehat{\gamma}_u^-(t,Z,\mathbf{X})}{b^+(Z,\mathbf{X})} - \frac{(\widehat{Q}^-(t,Z,\mathbf{X})-Y)_+-\widehat{\gamma}_l^-(t,Z,\mathbf{X})}{b^-(Z,\mathbf{X})}\right]$
Pseudo-outcomes (continuous)	$\phi_{t,z}^+(S;\widehat{\eta}) = \widehat{Q}^+(t,z,\mathbf{X}) + \frac{\widehat{\gamma}_u^+(t,z,\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\widehat{\gamma}_l^+(t,z,\mathbf{X})}{b^+(z,\mathbf{X})}$ $+\frac{\mathbf{1}_{[T=t]}\,K_h(Z-z)}{\widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(Z\|\mathbf{X})}\left[\frac{(Y-\widehat{Q}^+(t,Z,\mathbf{X}))_+-\widehat{\gamma}_u^+(t,Z,\mathbf{X})}{b^-(Z,\mathbf{X})} - \frac{(\widehat{Q}^+(t,Z,\mathbf{X})-Y)_+-\widehat{\gamma}_l^+(t,Z,\mathbf{X})}{b^+(Z,\mathbf{X})}\right]$ $\phi_{t,z}^-(S;\widehat{\eta}) = \widehat{Q}^-(t,z,\mathbf{X}) + \frac{\widehat{\gamma}_u^-(t,z,\mathbf{X})}{b^+(z,\mathbf{X})} - \frac{\widehat{\gamma}_l^-(t,z,\mathbf{X})}{b^-(z,\mathbf{X})}$ $+\frac{\mathbf{1}_{[T=t]}\,K_h(Z-z)}{\widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(Z\|\mathbf{X})}\left[\frac{(Y-\widehat{Q}^-(t,Z,\mathbf{X}))_+-\widehat{\gamma}_u^-(t,Z,\mathbf{X})}{b^+(Z,\mathbf{X})} - \frac{(\widehat{Q}^-(t,Z,\mathbf{X})-Y)_+-\widehat{\gamma}_l^-(t,Z,\mathbf{X})}{b^-(Z,\mathbf{X})}\right]$

Table H.2: Summary of the partial-identification bounds developed in Chapter 9 for causal inference under misspecified exposure mappings.

H.2.2 Continuous Neighborhood Exposure

This section gives the continuous-Z analogues of Theorem 9.14 and Corollary 9.17, as well as the corresponding *sharpness* and *validity* guarantees for the estimated bounds (complementary to the the discrete-Z results in the main text).

When Z is continuous, point evaluation at $Z = z$ is non-regular. Following the standard approach in orthogonal learning for continuous exposures, we therefore target a *kernel-localized* (bandwidth-indexed) version of the bound functional. Under smoothness in z, these localized targets converge to the original (pointwise) bounds as $h \downarrow 0$, at the usual bias–variance tradeoff governed by (n, h).

Assumption H.1 (Kernel localization)**.** Let Z be continuous and let $K_h(u) = \frac{1}{h}K(u/h)$, where K is bounded, integrates to 1, and $\int K(u)^2\,du < \infty$. Let $h = h_n \downarrow 0$ with $nh_n \to \infty$.

Kernel-localized targets Fix (t, z) and let $h > 0$. For continuous Z, define the localized selection weight

$$\kappa_{t,z,h}(S) := \frac{\mathbf{1}_{[T=t]}\, K_h(Z - z)}{\pi^t(\mathbf{X})\, \pi^g(Z \mid \mathbf{X})}, \tag{H.1}$$

as in the continuous-Z modification of the proof of Theorem 9.11. Define the kernel-localized pseudo-outcome $\phi^+_{t,z,h}(S; \widehat{\eta})$ as Eq. (9.35) with $\omega_{z,h}(Z) = K_h(Z - z)$.

When $\widehat{\eta} = \eta$, define the associated bandwidth-indexed functionals by

$$\mu^+_h(t, z, \mathbf{x}) := \mathbb{E}\Big[\phi^+_{t,z,h}(S; \eta) \mid \mathbf{X} = \mathbf{x}\Big], \qquad \psi^+_h(t, z) := \mathbb{E}\Big[\phi^+_{t,z,h}(S; \eta)\Big]. \tag{H.2}$$

Under standard smoothness in z, $\mu^+_h(t, z, \mathbf{x}) \to \mu^+(t, z, \mathbf{x})$ and $\psi^+_h(t, z) \to \psi^+(t, z)$ as $h \downarrow 0$ (see Remark 9.12).

Relative to the discrete-Z case, kernel localization inflates the second-order remainder by a factor $h^{-1/2}$ (reflecting $\int K_h^2 = O(1/h)$). This propagates to the final-stage CAPO rate and yields the usual $\sqrt{nh}$ scaling for the (smoothed) APO.

Theorem H.2 (Second-order nuisance error (continuous Z))**.** *Assume Z is continuous and Assumptions 9.13 and H.1 hold. Let $\widehat{\eta} = (\widehat{\pi}^t, \widehat{\pi}^g, \widehat{Q}^+, \widehat{\gamma}^+_u, \widehat{\gamma}^+_l)$ be the cross-fitted nuisances used in $\phi^+_{t,z,h}(S; \widehat{\eta})$ (Eq.* (9.35) *with $\omega_{z,h}(Z) = K_h(Z - z)$).*

Define nuisance error rates (in L_2 norms over the appropriate arguments) by

$$r_{n,\pi} := \|\widehat{\pi}^t - \pi^t\|_2 + \|\widehat{\pi}^g - \pi^g\|_2, \qquad r_{n,Q} := \|\widehat{Q}^+ - Q^+\|_2, \tag{H.3}$$

$$r_{n,\gamma} := \|\widehat{\gamma}^+_u - \gamma_u(\widehat{Q}^+; \cdot)\|_2 + \|\widehat{\gamma}^+_l - \gamma_l(\widehat{Q}^+; \cdot)\|_2, \tag{H.4}$$

where the norms are taken over the random variables that the corresponding nuisance is evaluated on (e.g., $(Z, \mathbf{X})$ for $\pi^g(Z \mid \mathbf{X})$, $Q^+(t, Z, \mathbf{X})$, and $\gamma^\pm(t, Z, \mathbf{X})$).

Then, the conditional bias induced by nuisance estimation satisfies

$$\left\| \mathbb{E}\left[\phi^+_{t,z,h}(S; \widehat{\eta}) - \phi^+_{t,z,h}(S; \eta) \mid \mathbf{X} \right] \right\|_2 = O_p\left(\frac{r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q}}{\sqrt{h}} \right). \tag{H.5}$$

Corollary H.3 (Quasi-oracle rates and inference (continuous Z))**.** *Assume the conditions of Theorem H.2 and that the second-stage regression learner $\widehat{\mathbb{E}}_n[\cdot \mid \mathbf{X} = \mathbf{x}]$ satisfies Assumption 9.15 with rate δ_n when regressing $\phi^+_{t,z,h}(S; \eta)$ on $\mathbf{X}$.*

Then:

CAPO rates: The CAPO upper-bound estimator satisfies

$$\|\widehat{\mu}^+_h(t, z, \cdot) - \mu^+_h(t, z, \cdot)\|_2 = O_p\left(\delta_n + \frac{r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q}}{\sqrt{h}} \right). \tag{H.6}$$

APO rates: The APO upper-bound estimator $\widehat{\psi}^+(t, z) = \mathbb{E}_n[\widehat{\phi}^+_{t,z,h}]$ satisfies

$$|\widehat{\psi}^+_h(t, z) - \psi^+_h(t, z)| = O_p\left(\frac{1}{\sqrt{nh}} + \frac{r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q}}{\sqrt{h}} \right). \tag{H.7}$$

$\sqrt{nh}$-CLT (central limit theorem) for the (smoothed) APO. If $r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q} = o_p(n^{-1/2})$, then

$$\sqrt{nh}\left(\widehat{\psi}^+_h(t, z) - \psi^+_h(t, z) \right) \rightsquigarrow \mathcal{N}(0,\ V^+_h(t, z)), \tag{H.8}$$

where one valid asymptotic variance target is $V^+_h(t, z) := \mathrm{Var}(\sqrt{h}\, \phi^+_{t,z,h}(S; \eta))$.

Finally, if the smoothing bias satisfies $|\psi^+_h(t, z) - \psi^+(t, z)| = o((nh)^{-1/2})$ (e.g., via undersmoothing under z-smoothness), then the CLT holds with $\psi^+(t, z)$ in place of $\psi^+_h(t, z)$.

Sharpness and validity of the estimated bounds The previous results control the second-order remainder and deliver quasi-oracle rates for the localized targets (μ^+_h, ψ^+_h). We now record the two complementary guarantees from the

main text in their continuous-Z versions: (i) consistency for the *sharp* identified bounds, and (ii) *validity* of the resulting intervals under potentially misspecified cutoffs. As before, the statements hold for both endpoints (+/−); we write them for the upper endpoint for brevity, with the lower endpoint following analogously by sign-swapping in the pseudo-outcome.

Proposition H.4 (Consistency for sharp bounds (continuous Z))**.** *Assume the conditions of Corollary H.3 and consider the corresponding lower-bound estimator* $\widehat{\mu_h^-}(t, z, \cdot)$ *constructed from the lower-bound pseudo-outcome (defined analogously to Eq.* (9.35)*). Suppose* $\delta_n = o_p(1)$ *and*

$$\frac{r_{n,\pi}\, r_{n,\gamma} + r_{n,Q}^2}{\sqrt{h}} = o_p(1). \tag{H.9}$$

Then,

$$\|\widehat{\mu_h^\pm}(t, z, \cdot) - \mu_h^\pm(t, z, \cdot)\|_2 = o_p(1), \qquad |\widehat{\psi_h^\pm}(t, z) - \psi_h^\pm(t, z)| = o_p(1). \tag{H.10}$$

Consequently, the estimated CAPO and APO intervals converge to the sharp *kernel-localized identified intervals for the bandwidth-indexed targets.*

Moreover, if the smoothing bias vanishes at the appropriate rate (e.g., $|\psi_h^\pm(t, z) - \psi^\pm(t, z)| = o((nh)^{-1/2})$*), then the estimated intervals are asymptotically sharp for the original pointwise bounds as* $h \downarrow 0$*.*

Corollary H.5 (Asymptotic validity under misspecified cutoffs (continuous Z))**.** *Fix measurable cutoffs* $\overline{Q}^\pm(t, z, \mathbf{x})$ *(not necessarily equal to the sharp cut-offs) and let* $\overline{\mu}_h^\pm(t, z, \mathbf{x}; \overline{Q}^\pm)$ *and* $\overline{\psi}_h^\pm(t, z; \overline{Q}^\pm)$ *denote the resulting (possibly non-sharp) kernel-localized bound functionals induced by these cutoffs (i.e., the targets obtained by replacing* $Q^\pm$ *in the pseudo-outcomes and taking the conditional/unconditional expectations as in Eq.* (H.2)*). Then, the induced intervals*

$$[\overline{\mu_h^-}(t, z, \mathbf{x}; \overline{Q}^-),\ \overline{\mu}_h^+(t, z, \mathbf{x}; \overline{Q}^+)] \quad \textit{and} \quad [\overline{\psi_h^-}(t, z; \overline{Q}^-),\ \overline{\psi}_h^+(t, z; \overline{Q}^+)] \tag{H.11}$$

are (not necessarily sharp) valid CAPO and APO intervals for the kernel-localized targets. Moreover, if $\widehat{Q}^{\pm} \to \overline{Q}^{\pm}$ *in* L_2 *and either*

(i) $(\widehat{\pi}^l, \widehat{\pi}^g)$ *is consistent*, or

(ii) the corresponding tail-moment regressions $(\widehat{\gamma}_u^{\pm}, \widehat{\gamma}_l^{\pm})$ *are consistent for the targets induced by* $\overline{Q}^{\pm}$,

then the estimated endpoints converge to the induced (conservative) targets and the resulting (C)APO intervals remain asymptotically valid, though potentially conservative. If $\overline{Q}^{\pm}$ *equals the sharp cut-offs, then the induced bounds coincide with the sharp bounds, and the intervals are asymptotically sharp as well.*

Conclusion For continuous neighborhood exposure, our estimation and theory proceed exactly as in the discrete-Z case, except that (i) the indicator $\mathbf{1}_{[Z=z]}$ in the selection weight is replaced by kernel localization $K_h(Z-z)$ and (ii) the conditional pmf $\pi^g(z \mid \mathbf{X})$ is replaced by the conditional density $\pi^g(Z \mid \mathbf{X})$. This replacement yields an effective sample size nh around z, which inflates the second-order remainder by a factor $h^{-1/2}$ and leads to $\sqrt{nh}$ scaling for APO inference. Under smoothness in z, the bandwidth-indexed targets $(\mu_h^{\pm}, \psi_h^{\pm})$ converge to the pointwise bounds $(\mu^{\pm}, \psi^{\pm})$ as $h \downarrow 0$, yielding the usual bias-variance tradeoff in (n, h). The proofs in Appendix G.2 show that all continuous-Z results follow from the discrete-Z proofs by replacing $\mathbf{1}_{[Z=z]}$ by $K_h(Z-z)$ and tracking $\int K_h^2 = O(1/h)$.

H.3 Main Text Proofs

H.3.1 Auxiliary theory

Our bounds employ a sensitivity method proposed in Frauen et al. [2023]. However, the original contribution proposes bounds in the presence of unobserved confounding, whereas we are targeting a different setting. Below, we present

Theorem 1 in Frauen et al. [2023] adapted to our setting.

Theorem H.6. *Let $b^-(z, \mathbf{x}) \leq b^+(z, \mathbf{x})$ with $b^-(z, \mathbf{x}) \in (0, 1]$ and $b^+(z, \mathbf{x}) \in [1, \infty)$, such that for all $z, \mathbf{x}$*

$$b^-(z, \mathbf{x}) \leq \frac{p(g^*(t_{\mathcal{N}}) = z \mid \mathbf{x})}{p(g(t_{\mathcal{N}}) = z \mid \mathbf{x})} \leq b^+(z, \mathbf{x}) \tag{H.12}$$

and define $\alpha^{\pm}(z, \mathbf{x}) := \frac{(1-b^{\mp}(z,\mathbf{x}))b^{\pm}(z,\mathbf{x})}{b^{\pm}(z,\mathbf{x})-b^{\mp}(z,\mathbf{x})}$. Furthermore, let $F_Y(y) := F_Y(y \mid t, z, \mathbf{x})$ denote the conditional cumulative distribution function (CDF) of Y. For $Y \in \mathbb{R}$ continuous, we define

$$p^+(y \mid t, z, \mathbf{x}) = \begin{cases} \frac{1}{b^+(z,\mathbf{x})} p(y \mid t, z, \mathbf{x}), \text{ if } F(y) \leq \alpha^+(z, \mathbf{x}), \\ \frac{1}{b^-(z,\mathbf{x})} p(y \mid t, z, \mathbf{x}), \text{ if } F(y) > \alpha^+(z, \mathbf{x}), \end{cases} \tag{H.13}$$

and for $Y \in \mathbb{R}$ discrete, we define the probability mass function

$$P^+(y \mid t, z, \mathbf{x}) = \begin{cases} \frac{1}{b^+(z,\mathbf{x})} P(y \mid t, z, \mathbf{x}), & \text{if } F(y) < \alpha^+(z, \mathbf{x}), \\ \frac{1}{b^-(z,\mathbf{x})} P(y \mid t, z, \mathbf{x}), & \text{if } F(y-1) > \alpha^+(z, \mathbf{x}), \\ \frac{1}{b^+(z, \mathbf{x})}(\alpha^+(z, \mathbf{x}) - F(y-1)) & \\ \quad + \frac{1}{b^-(z, \mathbf{x})}(F(y) - \alpha^+(z, \mathbf{x})), & \text{otherwise.} \end{cases} \tag{H.14}$$

The lower bound $p^-(y \mid t, z, \mathbf{x})$ is defined through exchanging the signs in α and b. Let $F^{\pm}(y)$ denote the conditional CDF with regard to $p^{\pm}(y \mid t, z, \mathbf{x})$. Then, for all $y \in \mathcal{Y}$

$$F^+(y) \leq \inf_{\tilde{P} \in \mathcal{M}} F_{\tilde{P}}(y), F^-(y) \geq \inf_{\tilde{P} \in \mathcal{M}} F_{\tilde{P}}(y), \tag{H.15}$$

i.e., the bounds are valid, and

$$F^+(y) = \inf_{\tilde{P} \in \mathcal{M}} F_{\tilde{P}}(y), F^-(y) = \inf_{\tilde{P} \in \mathcal{M}} F_{\tilde{P}}(y), \tag{H.16}$$

i.e., the bounds are sharp, if Z is continuous or if Z is discrete and $\frac{1}{b^+(z,\mathbf{x})} \geq \pi^g(z \mid \mathbf{x})$.

H.3.2 Proof of Theorem 9.9

Theorem 9.9. *[Sharp Bounds] Let $Q^{\pm}(t, z, \mathbf{x})$ be defined as in Eq.* (9.21)*, and let $(u)_+ = \max\{u, 0\}$. Then the sharp CAPO upper and lower bounds are*

$$\begin{aligned}\mu^{\pm}(t, z, \mathbf{x}) = Q^{\pm}(t, z, \mathbf{x}) &+ \frac{1}{b^{\mp}(z, \mathbf{x})}\mathbb{E}\left[(Y - Q^{\pm}(t, z, \mathbf{x}))_+ \mid t, z, \mathbf{x}\right] \\ &- \frac{1}{b^{\pm}(z, \mathbf{x})}\mathbb{E}\left[(Q^{\pm}(t, z, \mathbf{x}) - Y)_+ \mid t, z, \mathbf{x}\right].\end{aligned} \tag{9.22}$$

Proof. Throughout the proof we focus on the upper bound for continuous outcomes. The other cases follow analogously. Recall the definition of

$$Q^{\pm}(t, z, \mathbf{x}) := \inf\left\{y \mid F_Y(y \mid t, z, \mathbf{x}) \geq \frac{(1 - b^{\mp}(z, \mathbf{x}))b^{\pm}(z, \mathbf{x})}{b^{\pm}(z, \mathbf{x}) - b^{\mp}(z, \mathbf{x})}\right\}, \tag{H.17}$$

when $b^-(z, \mathbf{x}) < 1 < b^+(z, \mathbf{x})$, and $Q^{\pm}(t, z, \mathbf{x}) = Q(t, z, \mathbf{x}) := \inf\{y \mid F_Y(y \mid t, z, \mathbf{x}) \geq \frac{1}{2}\}$ otherwise.

By applying Theorem H.6, the sharp upper and lower bounds on the conditional potential outcome $\mu(t, z, \mathbf{x})$ are given by

$$\mu^{\pm}(t, z, \mathbf{x}) = \frac{1}{b^{\pm}(z, \mathbf{x})}\int_{-\infty}^{Q^{\pm}(z,\mathbf{x})} y \, \mathrm{d}\mu + \frac{1}{b^{\mp}(z, \mathbf{x})}\int_{Q^{\pm}(z,\mathbf{x})}^{\infty} y \, \mathrm{d}\mu \tag{H.18}$$

$$= \frac{1}{b^{\pm}(z, \mathbf{x})} \cdot \alpha^{\pm}\mathrm{LCTE}^{\pm}_{\alpha}(t, z, \mathbf{x}) + \frac{1}{b^{\mp}(z, \mathbf{x})} \cdot (1 - \alpha^{\pm})\mathrm{CVaR}^{\pm}_{\alpha}(t, z, \mathbf{x}) \tag{H.19}$$

where we define $\alpha^{\pm} := \frac{(1-b^{\mp}(z,\mathbf{x}))b^{\pm}(z,\mathbf{x})}{b^{\pm}(z,\mathbf{x})-b^{\mp}(z,\mathbf{x})}$. Here, the $\mathrm{CVaR}^{\pm}$ denotes the *conditional value at risk* at level $\alpha^{\pm}$ with corresponding quantiles $Q^+(t, z, \mathbf{x})/Q^-(t, z, \mathbf{x})$ defined as

$$\mathrm{CVaR}^{+}_{\alpha}(t, z, \mathbf{x}) := \min_{q \in \mathbb{R}}\Big\{q + \frac{1}{1 - \alpha^+}\mathbb{E}[\,(Y - q)_+ \mid t, z, \mathbf{x}]\Big\} \tag{H.20}$$

$$= Q^+(t, z, \mathbf{x}) + \frac{b^- - b^+}{(1 - b^+)b^-}\mathbb{E}[(Y - Q^+(t, z, \mathbf{x}))_+ \mid t, z, \mathbf{x}], \tag{H.21}$$

$$\mathrm{CVaR}^{-}_{\alpha}(t, z, \mathbf{x}) := \min_{q \in \mathbb{R}}\Big\{q + \frac{1}{1 - \alpha^-}\mathbb{E}[\,(Y - q)_+ \mid t, z, \mathbf{x}]\Big\} \tag{H.22}$$

$$= Q^-(t, z, \mathbf{x}) + \frac{b^+ - b^-}{(1 - b^-)b^+}\mathbb{E}[(Y - Q^-(t, z, \mathbf{x}))_+ \mid t, z, \mathbf{x}] \tag{H.23}$$

where $(u)_+ = \max\{u, 0\}$, and $\text{LCTE}^{\pm}$ the *lower conditional tail expectation* at level $\alpha^{\pm}$ with corresponding quantiles $Q^+(t,z,\mathbf{x}))/Q^-(t,z,\mathbf{x}))$ defined as

$$\begin{aligned}\text{LCTE}^+_\alpha(t,z,\mathbf{x}) :=& \sup_{q\in\mathbb{R}}\Big\{ q \;-\; \frac{1}{\alpha^+}\,\mathbb{E}[(\,q - Y\,)_+ \mid t,z,\mathbf{x}]\Big\} \\ =& Q^+(t,z,\mathbf{x})) - \frac{b^+(z,\mathbf{x}) - b^-(z,\mathbf{x})}{(1-b^-(z,\mathbf{x}))b^+(z,\mathbf{x})}\mathbb{E}[(Q^+(t,z,\mathbf{x})) - Y)_+ \mid t,z,\mathbf{x}],\end{aligned} \tag{H.24}$$

$$\begin{aligned}\text{LCTE}^-_\alpha(t,z,\mathbf{x}) :=& \sup_{q\in\mathbb{R}}\Big\{ q \;-\; \frac{1}{\alpha^-}\,\mathbb{E}[(\,q - Y\,)_+ \mid t,z,\mathbf{x}]\Big\} \\ =& Q^-(t,z,\mathbf{x})) - \frac{b^-(z,\mathbf{x}) - b^+(z,\mathbf{x})}{(1-b^+(z,\mathbf{x}))b^-(z,\mathbf{x})}\mathbb{E}[(Q^-(t,z,\mathbf{x})) - Y)_+ \mid t,z,\mathbf{x}].\end{aligned} \tag{H.25}$$

With these reformulations of CVaR and LCTE then follows the desired result

$$\mu^{\pm}(t,z,\mathbf{x}) = Q^{\pm}(t,z,\mathbf{x}) + \frac{1}{b^{\mp}(z,\mathbf{x})}\mathbb{E}[(Y - Q^{\pm}(t,z,\mathbf{x}))_+ \mid t,z,\mathbf{x}] \tag{H.26}$$

$$- \frac{1}{b^{\pm}(z,\mathbf{x})}\mathbb{E}[(Q^{\pm}(t,z,\mathbf{x}) - Y)_+ \mid t,z,\mathbf{x}]. \tag{H.27}$$

□

H.3.3 Proof of Theorem 9.11

Theorem 9.11. *Let $S = (\mathbf{X}, Y, T, Z)$ and fix (t,z). Define*

$$\omega_{z,h}(Z) := \begin{cases} \mathbf{1}_{[Z=z]}, & \text{if } Z \text{ is discrete,} \\ K_h(Z-z), & \text{if } Z \text{ is continuous,}\end{cases}$$

and let $\pi^g(Z \mid \mathbf{X})$ denote the conditional pmf when Z is discrete and the conditional density when Z is continuous. Let

$$\widehat{\eta} = (\widehat{\pi}^t, \widehat{\pi}^g, \widehat{Q}^+, \widehat{\gamma}^+_u, \widehat{\gamma}^+_l) \tag{9.34}$$

be estimates of the nuisance functions. Then an orthogonal pseudo-outcome for the CAPO upper bound $\mu^+(t,z,\mathbf{x})$ is

$$\phi^+_{t,z}(S;\widehat{\eta}) = \widehat{Q}^+(t,z,\mathbf{X}) + \frac{\widehat{\gamma}^+_u(t,z,\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\widehat{\gamma}^+_l(t,z,\mathbf{X})}{b^+(z,\mathbf{X})} \tag{9.35}$$

$$+\frac{\mathbf{1}_{[T=t]}\omega_{z,h}(Z)}{\widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(Z\mid\mathbf{X})}\Bigg[\frac{(Y-\widehat{Q}^+(t,Z,\mathbf{X}))_+-\widehat{\gamma}_u^+(t,Z,\mathbf{X})}{b^-(Z,\mathbf{X})} - \frac{(\widehat{Q}^+(t,Z,\mathbf{X})-Y)_+-\widehat{\gamma}_l^+(t,Z,\mathbf{X})}{b^+(Z,\mathbf{X})}\Bigg].$$

Moreover, when $\widehat{\eta}=\eta$, the pseudo-outcome is unbiased for its target bound functional.

Proof. We begin with the case where Z is discrete (including binary), so $\omega_{z,h}(Z)=\mathbf{1}_{[Z=z]}$. We discuss the continuous-Z modification at the end.

Fix (t,z) and abbreviate

$$p(t,z\mid\mathbf{X}):=\pi^t(\mathbf{X})\,\pi^g(z\mid\mathbf{X}),\qquad \alpha:=\alpha^+(z,\mathbf{X}),\qquad Q:=Q^+(t,z,\mathbf{X}). \tag{H.28}$$

Define the (unnormalized) conditional tail moments

$$\gamma_u:=\gamma_u^+(t,z,\mathbf{X}):=\mathbb{E}[(Y-Q)_+\mid T=t,Z=z,\mathbf{X}]\,, \tag{H.29}$$

$$\gamma_l:=\gamma_l^+(t,z,\mathbf{X}):=\mathbb{E}[(Q-Y)_+\mid T=t,Z=z,\mathbf{X}]\,. \tag{H.30}$$

The sharp upper bound can be written as

$$\mu^+(t,z,\mathbf{X})=Q+\frac{\gamma_u}{b^-(z,\mathbf{X})}-\frac{\gamma_l}{b^+(z,\mathbf{X})}, \tag{H.31}$$

which matches the first line of Eq. (9.35) when $\widehat{\eta}=\eta$.

Step 1: Reparameterization of μ^+ as a convex combination of CVaR/LCTE functionals Define the upper-tail and lower-tail pseudo-outcomes at level α (see, e.g., Dorn et al. [2025a], Oprescu et al. [2023])

$$H_u(y,q):=q+\frac{1}{1-\alpha}(y-q)_+,\qquad H_l(y,q):=q-\frac{1}{\alpha}(q-y)_+. \tag{H.32}$$

Their conditional expectations at the true quantile Q are the conditional upper CVaR and lower conditional tail expectation (LCTE), respectively:

$$\theta_u(\mathbf{X}):=\mathbb{E}[H_u(Y,Q)\mid T=t,Z=z,\mathbf{X}]=Q+\frac{1}{1-\alpha}\gamma_u \tag{H.33}$$

$$\theta_l(\mathbf{X}) := \mathbb{E}[H_l(Y, Q) \mid T = t, Z = z, \mathbf{X}] = Q - \frac{1}{\alpha}\gamma_l. \tag{H.34}$$

Now set the weights

$$w_u(\mathbf{X}) := \frac{1-\alpha}{b^-(z, \mathbf{X})}, \qquad w_l(\mathbf{X}) := \frac{\alpha}{b^+(z, \mathbf{X})}. \tag{H.35}$$

By the definition of $\alpha^+(z, \mathbf{X})$, one has

$$w_u(\mathbf{X}) + w_l(\mathbf{X}) = \frac{1-\alpha}{b^-(z, \mathbf{X})} + \frac{\alpha}{b^+(z, \mathbf{X})} = 1. \tag{H.36}$$

Therefore,

$$w_u(\mathbf{X})\theta_u(\mathbf{X}) + w_l(\mathbf{X})\theta_l(\mathbf{X}) = (w_u + w_l)Q + \frac{w_u}{1-\alpha}\gamma_u - \frac{w_l}{\alpha}\gamma_l \tag{H.37}$$

$$= Q + \frac{1}{b^-}\gamma_u - \frac{1}{b^+}\gamma_l \tag{H.38}$$

$$= \mu^+(t, z, \mathbf{X}), \tag{H.39}$$

so μ^+ is a (convex) linear combination of the two tail functionals.

Step 2: The recentered efficient influence function for μ^+ Our orthogonal pseudo-outcome is the recentered efficient influence function (REIF) of $\mu^+(t, z, \mathbf{X})$. Since $w_u(\mathbf{X}), w_l(\mathbf{X})$ are known functions of $(b^\pm, \alpha)$ (hence fixed with respect to the data-generating distribution), linearity of REIFs implies

$$\phi^+_{t,z}(S\,; \eta) := \mathrm{REIF}(\mu^+(t, z, \mathbf{X})) = w_u(\mathbf{X})\,\phi_u(S\,; \eta) + w_l(\mathbf{X})\,\phi_l(S\,; \eta), \tag{H.40}$$

where $\phi_u(S\,; \eta) := \mathrm{REIF}(\theta_u(\mathbf{X}))$ and $\phi_l(S\,; \eta) := \mathrm{REIF}(\theta_l(\mathbf{X}))$.

Define the selection weight

$$\kappa_{t,z}(S) := \frac{\mathbf{1}_{[T=t]}\,\mathbf{1}_{[Z=z]}}{\pi^t(\mathbf{X})\,\pi^g(z \mid \mathbf{X})}. \tag{H.41}$$

By the known REIFs for conditional CVaR/LCTE functionals (e.g., Dorn et al. [2025a], Oprescu et al. [2023]),

$$\phi_u(S\,; \eta) = \theta_u(\mathbf{X}) + \kappa_{t,z}(S)(H_u(Y, Q) - \theta_u(\mathbf{X})), \quad \phi_l(S\,; \eta) = \theta_l(\mathbf{X}) + \kappa_{t,z}(S)(H_l(Y, Q) - \theta_l(\mathbf{X})). \tag{H.42}$$

Moreover, these REIFs are *orthogonal with respect to Q*: the cutoff Q is characterized as the optimizer of the corresponding tail objective (equivalently, the Rockafellar–Uryasev CVaR variational form), so the envelope/first-order condition yields $\partial_q \mathbb{E}[H_u(Y,q) \mid t,z,\mathbf{X}]|_{q=Q} = 0$ and $\partial_q \mathbb{E}[H_l(Y,q) \mid t,z,\mathbf{X}]|_{q=Q} = 0$ (see Dorn et al. [2025a], Oprescu et al. [2023]).

Finally, substituting (H.42) into (H.40), using $\theta_u(\mathbf{X}) = Q+\gamma_u/(1-\alpha)$ and $\theta_l(\mathbf{X}) = Q - \gamma_l/\alpha$, and simplifying with $w_u/(1-\alpha) = 1/b^-$ and $w_l/\alpha = 1/b^+$ yields exactly Eq. (9.35).

Step 3: Unbiasedness and orthogonality

Orthogonality (Neyman-orthogonality) follows because $\phi^+_{t,z}$ is a linear combination of orthogonal REIFs for θ_u and θ_l (linearity preserves orthogonality), and because θ_u, θ_l themselves are orthogonal both to the selection nuisance (π^t, π^g) and to the regression nuisances via the standard conditional-mean EIF from Eq. ((H.42)). Orthogonality with respect to Q is guaranteed by the envelope/first-order condition (FOC) argument above.

Unbiasedness follows by iterated expectations: conditional on $\mathbf{X}$,

$$\mathbb{E}\left[\frac{\mathbf{1}_{[T=t]}\mathbf{1}_{[Z=z]}}{p(t,z \mid \mathbf{X})}\left\{\frac{(Y-Q)_+ - \gamma_u}{b^-} - \frac{(Q-Y)_+ - \gamma_l}{b^+}\right\} \,\middle|\, \mathbf{X}\right] \tag{H.43}$$

$$= \mathbb{E}\left[\frac{(Y-Q)_+ - \gamma_u}{b^-} - \frac{(Q-Y)_+ - \gamma_l}{b^+} \,\middle|\, T=t, Z=z, \mathbf{X}\right] \tag{H.44}$$

$$= 0, \tag{H.45}$$

so $\mathbb{E}[\phi^+_{t,z}(S;\eta) \mid \mathbf{X}] = \mu^+(t,z,\mathbf{X})$. This completes the proof for discrete Z.

Continuous Z When Z is continuous, evaluation at $Z = z$ is not pathwise differentiable. We instead use kernel localization: replace $\mathbf{1}_{[Z=z]}$ in $\kappa_{t,z}$ by $\omega_{z,h}(Z) = K_h(Z-z)$ and replace the pmf $\pi^g(z \mid \mathbf{X})$ by the conditional density $\pi^g(Z \mid \mathbf{X})$ to

define the localized weight

$$\kappa_{t,z,h}(S) := \frac{\mathbf{1}_{[T=t]}\, K_h(Z-z)}{\pi^t(\mathbf{X})\,\pi^g(Z \mid \mathbf{X})}. \tag{H.46}$$

Then Eq. (H.42) and the linearity relation from Eq. (H.40) hold verbatim with $\kappa_{t,z}$ replaced by $\kappa_{t,z,h}$, yielding the localized pseudo-outcome in Eq. (9.35). The same iterated-expectations argument gives $\mathbb{E}[\phi^+_{t,z,h}(S;\eta) \mid \mathbf{X}] = \mu^+_h(t,z,\mathbf{X})$, and under standard smoothness in z, $\mu^+_h(t,z,\mathbf{X}) \to \mu^+(t,z,\mathbf{X})$ as $h \downarrow 0$. □

H.3.4 Proof of Theorem 9.14

Theorem 9.14. *Assume Z is discrete and Assumption 9.13 holds. Let $\widehat{\eta} = (\widehat{\pi}^t, \widehat{\pi}^g, \widehat{Q}^+, \widehat{\gamma}^+_u, \widehat{\gamma}^+_l)$ be the cross-fitted nuisances used in $\phi^+_{t,z}(S;\widehat{\eta})$ from Theorem 9.11. Define $r_{n,\pi} := \|\widehat{\pi}^t - \pi^t\|_2 + \|\widehat{\pi}^g - \pi^g\|_2$, $r_{n,Q} := \|\widehat{Q}^+ - Q^+\|_2$, and $r_{n,\gamma} := \|\widehat{\gamma}^+_u - \gamma_u(\widehat{Q}^+;\cdot)\|_2 + \|\widehat{\gamma}^+_l - \gamma_l(\widehat{Q}^+;\cdot)\|_2$, where*

$$\gamma_u(\widehat{Q}^+;\mathbf{X}) := \mathbb{E}[(Y - \widehat{Q}^+(\mathbf{X}))_+ \mid T=t, Z=z, \mathbf{X}], \tag{9.40}$$

$$\gamma_l(\widehat{Q}^+;\mathbf{X}) := \mathbb{E}[(\widehat{Q}^+(\mathbf{X}) - Y)_+ \mid T=t, Z=z, \mathbf{X}]. \tag{9.41}$$

Then

$$\left\| \mathbb{E}\left[\phi^+_{t,z}(S;\widehat{\eta}) - \phi^+_{t,z}(S;\eta) \mid \mathbf{X}\right] \right\|_2 = O_p\Big(r_{n,\pi} r_{n,\gamma} + r^2_{n,Q}\Big). \tag{9.42}$$

Proof. We prove the statement for discrete Z. Throughout, fix (t,z) and suppress (t,z) in the notation whenever clear. Because we use K-fold cross-fitting, for any observation in a held-out fold the nuisance estimates $\widehat{\eta} = (\widehat{\pi}^t, \widehat{\pi}^g, \widehat{Q}^+, \widehat{\gamma}^+_u, \widehat{\gamma}^+_l)$ are functions of the training folds only; hence, when taking expectations over the held-out fold, we may treat $\widehat{\eta}$ as fixed (formally, condition on the training sample).

Let $A := \mathbf{1}_{[T=t]}\mathbf{1}_{[Z=z]}$ and write the true and estimated joint propensities as

$$\pi(\mathbf{X}) := \pi^t(\mathbf{X})\pi^g(z \mid \mathbf{X}), \qquad \widehat{\pi}(\mathbf{X}) := \widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(z \mid \mathbf{X}). \tag{H.47}$$

Also denote the (population) conditional means at an arbitrary cutoff $\widehat{Q}$:

$$\gamma_u(\widehat{Q};\mathbf{X}) := \mathbb{E}\Big[(Y-\widehat{Q}(\mathbf{X}))_+ \mid T=t, Z=z, \mathbf{X}\Big], \tag{H.48}$$

$$\gamma_l(\widehat{Q};\mathbf{X}) := \mathbb{E}\Big[(\widehat{Q}(\mathbf{X})-Y)_+ \mid T=t, Z=z, \mathbf{X}\Big] \tag{H.49}$$

with $\gamma_u(Q^+;\mathbf{X}) = \gamma_u(\mathbf{X})$ and $\gamma_l(Q^+;\mathbf{X}) = \gamma_l(\mathbf{X})$.

Step 1: Conditional expectation of the estimated pseudo-outcome For discrete Z, the pseudo-outcome simplifies (since A forces $Z = z$ inside the square bracket) to

$$\phi(S;\widehat{\eta}) = \widehat{Q}(\mathbf{X}) + \frac{\widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})} \tag{H.50}$$

$$+ \frac{A}{\widehat{\pi}(\mathbf{X})}\left[\frac{(Y-\widehat{Q}(\mathbf{X}))_+ - \widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{(\widehat{Q}(\mathbf{X})-Y)_+ - \widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})}\right]. \tag{H.51}$$

Taking conditional expectations given $\mathbf{X}$ and using $\mathbb{E}[A \mid \mathbf{X}] = \pi(\mathbf{X})$ yields

$$\mathbb{E}[\phi(S;\widehat{\eta}) \mid \mathbf{X}] = \widehat{Q}(\mathbf{X}) + \frac{\widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})} \tag{H.52}$$

$$+ \frac{\pi(\mathbf{X})}{\widehat{\pi}(\mathbf{X})}\left[\frac{\gamma_u(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\gamma_l(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})}\right] \tag{H.53}$$

$$= \underbrace{\widehat{Q}(\mathbf{X}) + \frac{\gamma_u(\widehat{Q}^+;\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\gamma_l(\widehat{Q}^+;\mathbf{X})}{b^+(z,\mathbf{X})}}_{=:\,\mu_{\widehat{Q}^+}(\mathbf{X})} \tag{H.54}$$

$$+ \left(\frac{\pi(\mathbf{X})}{\widehat{\pi}(\mathbf{X})} - 1\right)\left[\frac{\gamma_u(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\gamma_l(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})}\right] \tag{H.55}$$

Moreover, by Theorem 9.11 (applied with true nuisances), we yield

$$\mathbb{E}[\phi(S;\eta) \mid \mathbf{X}] = \mu^+(\mathbf{X}) := Q(\mathbf{X}) + \frac{\gamma_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\gamma_l(\mathbf{X})}{b^+(z,\mathbf{X})}. \tag{H.56}$$

Thus, we arrive at

$$\mathbb{E}[\phi(S;\widehat{\eta}) - \phi(S;\eta) \mid \mathbf{X}] = \mu_{\widehat{Q}^+}(\mathbf{X}) - \mu^+(\mathbf{X}) \tag{H.57}$$

$$+ \left(\frac{\pi(\mathbf{X})}{\widehat{\pi}(\mathbf{X})} - 1\right)\left[\frac{\gamma_u(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\gamma_l(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})}\right], \tag{H.58}$$

and

$$\left\|\mathbb{E}[\phi(S;\widehat{\eta}) - \phi(S;\eta) \mid \mathbf{X}]\right\|_2 \leq \underbrace{\left\|\mu_{\widehat{Q}}(\mathbf{X}) - \mu^+(\mathbf{X})\right\|_2}_{\text{cutoff-induced error}} \tag{H.59}$$

$$+ \underbrace{\left\|\frac{\pi(\mathbf{X})}{\widehat{\pi}(\mathbf{X})} - 1\right\|_2 \left\|\frac{\gamma_u(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\gamma_l(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})}\right\|_2}_{\text{propensity} \times \text{regression product term}}$$

(H.60)

where the last inequality is due to the triangle inequality and Cauchy–Schwarz inequality.

Step 2: Bounding the product term by $O_p(r_{n,\pi} r_{n,\gamma})$ By Assumption 9.13, $\widehat{\pi}(\mathbf{X}) \geq \varepsilon$ a.s., hence

$$\left\|\frac{\pi(\mathbf{X})}{\widehat{\pi}(\mathbf{X})} - 1\right\|_2 = \left\|\frac{\pi(\mathbf{X}) - \widehat{\pi}(\mathbf{X})}{\widehat{\pi}(\mathbf{X})}\right\|_2 \leq \varepsilon^{-1} \|\widehat{\pi} - \pi\|_2. \tag{H.61}$$

Since $\pi = \pi^t \pi^g$ and $\widehat{\pi} = \widehat{\pi}^t \widehat{\pi}^g$,

$$\widehat{\pi} - \pi = (\widehat{\pi}^t - \pi^t)\widehat{\pi}^g + \pi^t(\widehat{\pi}^g - \pi^g), \tag{H.62}$$

so by the triangle inequality and $0 \leq \pi^t, \widehat{\pi}^g \leq 1$ (discrete Z),

$$\|\widehat{\pi} - \pi\|_2 \leq \|\widehat{\pi}^t - \pi^t\|_2 + \|\widehat{\pi}^g - \pi^g\|_2. \tag{H.63}$$

Therefore

$$\left\|\frac{\pi}{\widehat{\pi}} - 1\right\|_2 \leq \varepsilon^{-1}\Big(\|\widehat{\pi}^t - \pi^t\|_2 + \|\widehat{\pi}^g - \pi^g\|_2\Big) \tag{H.64}$$

and it remains to bound the second factor. Since $b^-(z,\mathbf{X})$ is bounded away from 0, we have

$$\left\|\frac{\gamma_u(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_u(\mathbf{X})}{b^-(z,\mathbf{X})} - \frac{\gamma_l(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_l(\mathbf{X})}{b^+(z,\mathbf{X})}\right\|_2 \leq \varepsilon^{-1}\Big(\|\widehat{\gamma}_u - \gamma_u(\widehat{Q}^+;\cdot)\|_2 + \|\widehat{\gamma}_l - \gamma_l(\widehat{Q}^+;\cdot)\|_2\Big), \tag{H.65}$$

by the triangle inequality. Combining with the previous inequality yields

$$\underbrace{\left\|\frac{\pi}{\widehat{\pi}} - 1\right\|_2}_{=O_p(r_{n,\pi})} \cdot \underbrace{\left\|\frac{\gamma_u(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_u(\mathbf{X})}{b^-} - \frac{\gamma_l(\widehat{Q}^+;\mathbf{X}) - \widehat{\gamma}_l(\mathbf{X})}{b^+}\right\|_2}_{=O_p(r_{n,\gamma})} = O_p(r_{n,\pi} r_{n,\gamma}), \tag{H.66}$$

where the second $O_p(r_{n,\gamma})$ is by definition of $r_{n,\gamma}$ (as the L_2 rate for estimating the conditional tail means at the cutoff used in the pseudo-outcome).

Step 3: Bounding the cutoff-induced term by $O_p(r_{n,Q}^2)$ We now need to control the term $\mu_{\widehat{Q}^+}(\mathbf{X}) - \mu^+(\mathbf{X})$. Fix (t, z) and $\mathbf{x}$, and define the scalar function

$$\mathcal{L}_{\mathbf{x}}(q) := q + \frac{1}{b^-(z,\mathbf{x})}\,\mathbb{E}[(Y-q)_+ \mid T=t, Z=z, \mathbf{X}=\mathbf{x}] \tag{H.67}$$

$$- \frac{1}{b^+(z,\mathbf{x})}\,\mathbb{E}[(q-Y)_+ \mid T=t, Z=z, \mathbf{X}=\mathbf{x}]\,. \tag{H.68}$$

By construction,

$$\mu_{\widehat{Q}^+}(\mathbf{x}) = \mathcal{L}_{\mathbf{x}}(\widehat{Q}^+(t,z,\mathbf{x})), \qquad \mu^+(\mathbf{x}) = \mathcal{L}_{\mathbf{x}}(Q^+(t,z,\mathbf{x})), \tag{H.69}$$

where $Q^+(t,z,\mathbf{x})$ is the optimal cutoff from Theorem 9.9.

Assume (as is standard for quantile/CVaR-style expansions) the conditional CDF $F_{Y|t,z,\mathbf{x}}$ is differentiable in a neighborhood of $Q^+(t,z,\mathbf{x})$ with density $f_{Y|t,z,\mathbf{x}}$ bounded by $\bar{f} < \infty$. Then, $\mathcal{L}_{\mathbf{x}}$ is differentiable and

$$\mathcal{L}'_{\mathbf{x}}(q) = 1 - \frac{1 - F_{Y|t,z,\mathbf{x}}(q)}{b^-(z,\mathbf{x})} - \frac{F_{Y|t,z,\mathbf{x}}(q)}{b^+(z,\mathbf{x})}. \tag{H.70}$$

Moreover, $\mathcal{L}'_{\mathbf{x}}$ is Lipschitz with

$$|\mathcal{L}''_{\mathbf{x}}(q)| = \left|\left(\frac{1}{b^-(z,\mathbf{x})} - \frac{1}{b^+(z,\mathbf{x})}\right) f_{Y|t,z,\mathbf{x}}(q)\right| \le \bar{f}\left(\frac{1}{b^-(z,\mathbf{x})} + \frac{1}{b^+(z,\mathbf{x})}\right) \le L, \tag{H.71}$$

for a finite constant L (uniform in $\mathbf{x}$ by Assumption 9.13).

By optimality of $Q^+(t,z,\mathbf{x})$ for $\mathcal{L}_{\mathbf{x}}$, we have $\mathcal{L}'_{\mathbf{x}}(Q^+(t,z,\mathbf{x})) = 0$. Therefore, by the fundamental theorem of calculus,

$$\left|\mathcal{L}_{\mathbf{x}}(\widehat{Q}^+(t,z,\mathbf{x})) - \mathcal{L}_{\mathbf{x}}(Q^+(t,z,\mathbf{x}))\right| = \left|\int_{Q^+(t,z,\mathbf{x})}^{\widehat{Q}^+(t,z,\mathbf{x})} (\mathcal{L}'_{\mathbf{x}}(u) - \mathcal{L}'_{\mathbf{x}}(Q^+(t,z,\mathbf{x})))\,\mathrm{d}u\right|$$

$$\leq \int_{Q^+(t,z,\mathbf{x})}^{\widehat{Q}^+(t,z,\mathbf{x})} L\,|u - Q^+(t,z,\mathbf{x})|\,\mathrm{d}u$$
$$\leq \frac{L}{2}\,|\widehat{Q}^+(t,z,\mathbf{x}) - Q^+(t,z,\mathbf{x})|^2.$$

Taking $L_2(P_{\mathbf{X}})$ norms yields

$$\|\mu_{\widehat{Q}^+} - \mu^+\|_2 = O_p\Big(\|\widehat{Q}^+ - Q^+\|_2^2\Big) = O_p(r_{n,Q}^2). \tag{H.72}$$

Conclusion. Combining Step 2 and Step 3 above in the decomposition from Eq. (H.58) yields

$$\Big\|\mathbb{E}\Big[\phi_{t,z}^+(S;\widehat{\eta}) - \phi_{t,z}^+(S;\eta) \mid \mathbf{X}\Big]\Big\|_2 = O_p\Big(r_{n,\pi}\, r_{n,\gamma} + r_{n,Q}^2\Big), \tag{H.73}$$

which is exactly Eq. (9.42). □

H.3.5 Proof of Corollary 9.17

Corollary 9.17. *Suppose Assumptions 9.13 and 9.15 hold, and let $r_{n,\pi}, r_{n,\gamma}, r_{n,Q}$ be as in Theorem 9.14. For the CAPO upper-bound estimator,*

$$\|\widehat{\mu}^+(t,z,\cdot) - \mu^+(t,z,\cdot)\|_2 = O_p\Big(\delta_n + r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2\Big). \tag{9.44}$$

In particular, if $r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2 = o_p(\delta_n)$, then $\|\widehat{\mu}^+(t,z,\cdot) - \mu^+(t,z,\cdot)\|_2 = O_p(\delta_n)$.
For the APO upper-bound estimator $\widehat{\psi}^+(t,z) = \mathbb{E}_n[\widehat{\phi}_{t,z}^+]$,

$$\Big|\widehat{\psi}^+(t,z) - \psi^+(t,z)\Big| = O_p\Big(n^{-1/2} + r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2\Big). \tag{9.45}$$

If moreover $r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2 = o_p(n^{-1/2})$, then

$$\sqrt{n}\left(\widehat{\psi}^+(t,z) - \psi^+(t,z)\right) \rightsquigarrow \mathcal{N}(0, V^+(t,z)), \tag{9.46}$$

where $V^+(t,z) := \mathrm{Var}(\phi_{t,z}^+(S;\eta))$.

Proof. We prove the CAPO and APO statements for the upper bound; the lower-bound case follows by the same argument with the sign-swapped pseudo-outcome.

CAPO bound rate. Let $m^+_{t,z}(\mathbf{x}) := \mathbb{E}[\widehat{\phi}^+_{t,z} \mid \mathbf{X} = \mathbf{x}]$ denote the conditional mean of the (cross-fitted) pseudo-outcome. By Assumption 9.15,

$$\|\widehat{\mu}^+(t,z,\cdot) - m^+_{t,z}(\cdot)\|_2 = O_p(\delta_n). \tag{H.74}$$

By the triangle inequality,

$$\|\widehat{\mu}^+(t,z,\cdot) - \mu^+(t,z,\cdot)\|_2 \le \|\widehat{\mu}^+(t,z,\cdot) - m^+_{t,z}(\cdot)\|_2 + \|m^+_{t,z}(\cdot) - \mu^+(t,z,\cdot)\|_2. \tag{H.75}$$

The second term is precisely the conditional bias induced by nuisance estimation. Applying Theorem 9.14 yields

$$\|m^+_{t,z}(\cdot) - \mu^+(t,z,\cdot)\|_2 = O_p(r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q}). \tag{H.76}$$

Combining the two bounds gives the stated CAPO rate.

APO rate and asymptotic normality. Recall $\widehat{\psi}^+(t,z) = \mathbb{E}_n[\widehat{\phi}^+_{t,z}]$. Decompose

$$\widehat{\psi}^+(t,z) - \psi^+(t,z) = (\mathbb{E}_n - \mathbb{E})[\phi^+_{t,z}(S;\eta)] + \mathbb{E}\Big[\phi^+_{t,z}(S;\widehat{\eta}) - \phi^+_{t,z}(S;\eta)\Big] + R_n, \tag{H.77}$$

where $R_n := (\mathbb{E}_n - \mathbb{E})[\phi^+_{t,z}(S;\widehat{\eta}) - \phi^+_{t,z}(S;\eta)]$ is an empirical-process term. Under Assumption 9.13(iii), $\phi^+_{t,z}(S;\cdot)$ is uniformly bounded, and with cross-fitting $R_n = o_p(n^{-1/2})$ by standard arguments (conditioning on training folds and applying Hoeffding/Bernstein inequalities).

The first term is $O_p(n^{-1/2})$ by the CLT. The second term is bounded by Theorem 9.14 (after integrating over $\mathbf{X}$), yielding

$$\Big|\mathbb{E}\Big[\phi^+_{t,z}(S;\widehat{\eta}) - \phi^+_{t,z}(S;\eta)\Big]\Big| = O_p(r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q}). \tag{H.78}$$

This proves Eq. (9.46). If additionally $r_{n,\pi} r_{n,\gamma} + r^2_{n,Q} = o_p(n^{-1/2})$, then the nuisance-induced bias term and R_n are $o_p(n^{-1/2})$, hence

$$\sqrt{n}(\widehat{\psi}^+(t,z) - \psi^+(t,z)) = \sqrt{n}(\mathbb{E}_n - \mathbb{E})[\phi^+_{t,z}(S;\eta)] + o_p(1) \rightsquigarrow \mathcal{N}(0, V^+(t,z)), \tag{H.79}$$

with $V^+(t,z) = \mathrm{Var}(\phi^+_{t,z}(S;\eta))$. □

H.3.6 Proof of Proposition 9.18

Proposition 9.18. *Assume the conditions of Corollary 9.17 hold. Suppose* $\delta_n = o_p(1)$ *and* $r_{n,Q} = o_p(1)$*, and in addition either* $r_{n,\pi} = o_p(1)$ *or* $r_{n,\gamma} = o_p(1)$*. Then*

$$\|\widehat{\mu}^{\pm}(t,z,\cdot) - \mu^{\pm}(t,z,\cdot)\|_2 = o_p(1) \tag{9.47}$$

and

$$|\widehat{\psi}^{\pm}(t,z) - \psi^{\pm}(t,z)| = o_p(1). \tag{9.48}$$

Consequently, the estimated CAPO and APO intervals converge to the sharp identified intervals.

Proof. We show the claim for the CAPO upper bound; the other bounds (CAPO lower, APO upper/lower) follow similarly.

By Corollary 9.17,

$$\|\widehat{\mu}^{+}(t,z,\cdot) - \mu^{+}(t,z,\cdot)\|_2 = O_p\Big(\delta_n + r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2\Big). \tag{H.80}$$

Under the proposition assumptions, $\delta_n = o_p(1)$ and $r_{n,Q} = o_p(1)$. Moreover, if either $r_{n,\pi} = o_p(1)$ or $r_{n,\gamma} = o_p(1)$, then $r_{n,\pi} r_{n,\gamma} = o_p(1)$. Therefore, the right-hand side is $o_p(1)$, implying $\|\widehat{\mu}^{+}(t,z,\cdot) - \mu^{+}(t,z,\cdot)\|_2 = o_p(1)$.

For APOs, the corresponding statement follows from the APO rate in Corollary 9.17 and the same convergence of the remainder. Finally, repeating the same argument for the lower bound (using its analogous pseudo-outcome) establishes convergence of both endpoints and, hence, convergence of the estimated intervals to the sharp identified intervals. □

H.3.7 Proof of Corollary 9.19

Corollary 9.19. *Assume the conditions of Corollary 9.17 hold. Let $\overline{Q}^{\pm}(t, z, \mathbf{x})$ be any measurable cutoff and define the induced bounds*

$$\overline{\mu}^{\pm}(t, z, \mathbf{x}; \overline{Q}^{\pm}) = \overline{Q}^{\pm}(t, z, \mathbf{x}) + \frac{1}{b^{\mp}(z, \mathbf{x})}\mathbb{E}\left[(Y - \overline{Q}^{\pm}(t, z, \mathbf{x}))_+ \mid t, z, \mathbf{x}\right] \tag{9.49}$$
$$- \frac{1}{b^{\pm}(z, \mathbf{x})}\mathbb{E}\left[(\overline{Q}^{\pm}(t, z, \mathbf{x}) - Y)_+ \mid t, z, \mathbf{x}\right],$$

and analogously define $\overline{\psi}^{\pm}(t, z) := \mathbb{E}[\overline{\mu}^{\pm}(t, z, \mathbf{X})]$. Then

$$[\overline{\mu}^{-}(t, z, \mathbf{x}), \overline{\mu}^{+}(t, z, \mathbf{x})] \tag{9.50}$$

is a valid, though not necessarily sharp, CAPO interval, and likewise

$$[\overline{\psi}^{-}(t, z), \overline{\psi}^{+}(t, z)] \tag{9.51}$$

is a valid APO interval. Moreover, if $\widehat{Q}^{\pm} \to \overline{Q}^{\pm}$ in L_2 and either (i) $(\widehat{\pi}^t, \widehat{\pi}^s)$ is consistent, or (ii) $(\widehat{\gamma}_u^{\pm}, \widehat{\gamma}_l^{\pm})$ is consistent for the tail-moment targets induced by $\overline{Q}^{\pm}$, then the resulting estimated CAPO and APO intervals converge to

$$[\overline{\mu}^{-}, \overline{\mu}^{+}] \quad \textit{and} \quad [\overline{\psi}^{-}, \overline{\psi}^{+}], \tag{9.52}$$

respectively, and are asymptotically valid, though potentially conservative. If $\overline{Q}^{\pm} = Q^{\pm}$, then the bounds coincide with the sharp bounds.

Proof. We prove the CAPO claim; the APO claim follows by taking expectations over $\mathbf{X}$.

Step 1: Any cutoff induces a valid (conservative) interval. Fix $(t, z, \mathbf{x})$ and define for any scalar cutoff q the upper and lower tail objectives

$$\mathcal{L}_{\mathbf{x}}^{+}(q) := q + \frac{1}{b^{-}(z, \mathbf{x})}\mathbb{E}[(Y - q)_+ \mid T = t, Z = z, \mathbf{X} = \mathbf{x}] \tag{H.81}$$
$$- \frac{1}{b^{+}(z, \mathbf{x})}\mathbb{E}[(q - Y)_+ \mid T = t, Z = z, \mathbf{X} = \mathbf{x}], \tag{H.82}$$

$$\mathcal{L}_{\mathbf{x}}^{-}(q) := q + \frac{1}{b^{+}(z,\mathbf{x})}\mathbb{E}[(Y-q)_{+} \mid T=t, Z=z, \mathbf{X}=\mathbf{x}] \tag{H.83}$$

$$- \frac{1}{b^{-}(z,\mathbf{x})}\mathbb{E}[(q-Y)_{+} \mid T=t, Z=z, \mathbf{X}=\mathbf{x}]\,. \tag{H.84}$$

By Theorem 9.9 (equivalently, the standard Rockafellar–Uryasev variational form),

$$\mu^{+}(t,z,\mathbf{x}) = \inf_{q} \mathcal{L}_{\mathbf{x}}^{+}(q), \qquad \mu^{-}(t,z,\mathbf{x}) = \sup_{q} \mathcal{L}_{\mathbf{x}}^{-}(q), \tag{H.85}$$

with optimizers $q = Q^{+}(t,z,\mathbf{x})$ and $q = Q^{-}(t,z,\mathbf{x})$. Hence, for any measurable $\overline{Q}^{+}(t,z,\mathbf{x})$ and $\overline{Q}^{-}(t,z,\mathbf{x})$,

$$\overline{\mu}^{+}(t,z,\mathbf{x};\overline{Q}^{+}) := \mathcal{L}_{\mathbf{x}}^{+}(\overline{Q}^{+}(t,z,\mathbf{x})) \;\geq\; \mu^{+}(t,z,\mathbf{x}) \tag{H.86}$$

$$\overline{\mu}^{-}(t,z,\mathbf{x};\overline{Q}^{-}) := \mathcal{L}_{\mathbf{x}}^{-}(\overline{Q}^{-}(t,z,\mathbf{x})) \;\leq\; \mu^{-}(t,z,\mathbf{x}), \tag{H.87}$$

so $[\overline{\mu}^{-}(t,z,\mathbf{x}), \overline{\mu}^{+}(t,z,\mathbf{x})]$ contains the sharp CAPO interval and is therefore valid.

Step 2: Convergence to the induced bounds. Fix measurable cutoffs $\overline{Q}^{\pm}$ and define the induced hinge-mean targets

$$\overline{\gamma}_{u}^{\pm}(t,z,\mathbf{x}) := \mathbb{E}\Big[(Y - \overline{Q}^{\pm}(t,z,\mathbf{x}))_{+} \mid T=t, Z=z, \mathbf{X}=\mathbf{x}\Big], \tag{H.88}$$

$$\overline{\gamma}_{l}^{\pm}(t,z,\mathbf{x}) := \mathbb{E}\Big[(\overline{Q}^{\pm}(t,z,\mathbf{x}) - Y)_{+} \mid T=t, Z=z, \mathbf{X}=\mathbf{x}\Big]. \tag{H.89}$$

Let $\overline{\eta}^{\pm} := (\pi^{t}, \pi^{g}, \overline{Q}^{\pm}, \overline{\gamma}_{u}^{\pm}, \overline{\gamma}_{l}^{\pm})$. By Theorem 9.11, the corresponding pseudo-outcome is conditionally unbiased: $\mathbb{E}[\phi_{t,z}^{\pm}(S;\overline{\eta}^{\pm}) \mid \mathbf{X}] = \overline{\mu}^{\pm}(t,z,\mathbf{X};\overline{Q}^{\pm})$.

Now consider the estimated pseudo-outcome $\phi_{t,z}^{\pm}(S;\widehat{\eta}^{\pm})$ and write $\widehat{Q} = \widehat{Q}^{\pm}$, $\overline{Q} = \overline{Q}^{\pm}$ for brevity. The same conditional-expectation algebra as in the proof of Theorem 9.14 yields the decomposition

$$\mathbb{E}[\phi_{t,z}^{\pm}(S;\widehat{\eta}^{\pm}) \mid \mathbf{X}] = \mu_{\widehat{Q}}^{\pm}(\mathbf{X}) + \left(\frac{\pi(\mathbf{X})}{\widehat{\pi}(\mathbf{X})} - 1\right)\Delta_{\widehat{Q}}^{\pm}(\mathbf{X}), \tag{H.90}$$

where $\mu_{\widehat{Q}}^{\pm}(\mathbf{X})$ is the induced bound functional evaluated at $\widehat{Q}$ (i.e., $\mathcal{L}_{\mathbf{X}}^{\pm}(\widehat{Q})$) and $\Delta_{\widehat{Q}}^{\pm}(\mathbf{X})$ collects the conditional-mean regression errors at cutoff $\widehat{Q}$.

First, since $(u)_+$ is 1-Lipschitz, for each $\mathbf{X}$,

$$|\gamma_u(\widehat{Q};\mathbf{X}) - \gamma_u(\overline{Q};\mathbf{X})| \le |\widehat{Q}(\mathbf{X}) - \overline{Q}(\mathbf{X})| \tag{H.91}$$

$$|\gamma_l(\widehat{Q};\mathbf{X}) - \gamma_l(\overline{Q};\mathbf{X})| \le |\widehat{Q}(\mathbf{X}) - \overline{Q}(\mathbf{X})|. \tag{H.92}$$

Using $b^\pm$ bounded away from 0, this implies $\|\mu^\pm_{\widehat{Q}} - \overline{\mu}^\pm(\cdot;\overline{Q})\|_2 \lesssim \|\widehat{Q} - \overline{Q}\|_2 = o_p(1)$ whenever $\widehat{Q} \to \overline{Q}$ in L_2.

Second, for the product term, Assumption 9.13 implies $\|\pi/\widehat{\pi} - 1\|_2$ is bounded, and, if $(\widehat{\pi}^t, \widehat{\pi}^g)$ is consistent, then $\|\pi/\widehat{\pi} - 1\|_2 = o_p(1)$. Moreover,

$$\|\gamma_u(\widehat{Q};\cdot) - \widehat{\gamma}^\pm_u\|_2 \le \|\overline{\gamma}^\pm_u - \widehat{\gamma}^\pm_u\|_2 + \|\gamma_u(\widehat{Q};\cdot) - \gamma_u(\overline{Q};\cdot)\|_2 \le \|\overline{\gamma}^\pm_u - \widehat{\gamma}^\pm_u\|_2 + \|\widehat{Q} - \overline{Q}\|_2, \tag{H.93}$$

and similarly for the lower hinge mean. Hence, if $(\widehat{\gamma}^\pm_u, \widehat{\gamma}^\pm_l)$ is consistent for the induced targets $(\overline{\gamma}^\pm_u, \overline{\gamma}^\pm_l)$ and $\widehat{Q} \to \overline{Q}$, then $\|\Delta^\pm_{\widehat{Q}}\|_2 = o_p(1)$, so the product term is $o_p(1)$ even if $(\widehat{\pi}^t, \widehat{\pi}^g)$ is misspecified (but bounded away from 0).

Combining the two parts gives

$$\|\mathbb{E}[\phi^\pm_{t,z}(S;\widehat{\eta}^\pm) \mid \mathbf{X}] - \overline{\mu}^\pm(t,z,\mathbf{X};\overline{Q}^\pm)\|_2 = o_p(1). \tag{H.94}$$

Under Assumption 9.15 with $\delta_n = o_p(1)$, the final-stage regression therefore yields $\|\widehat{\mu}^\pm(t,z,\cdot) - \overline{\mu}^\pm(t,z,\cdot;\overline{Q}^\pm)\|_2 = o_p(1)$, and the sample-average estimator gives $\widehat{\psi}^\pm(t,z) \to \overline{\psi}^\pm(t,z)$. Thus the estimated (C)APO intervals converge to the induced (conservative) intervals and are asymptotically valid. If $\overline{Q}^\pm = Q^\pm$, then $\overline{\mu}^\pm = \mu^\pm$ and the limits coincide with the sharp bounds. □

H.4 Appendix Proofs

H.4.1 Proof of Theorem H.2

Theorem H.2 (Second-order nuisance error (continuous Z))**.** *Assume Z is continuous and Assumptions 9.13 and H.1 hold. Let $\widehat{\eta} = (\widehat{\pi}^t, \widehat{\pi}^g, \widehat{Q}^+, \widehat{\gamma}^+_u, \widehat{\gamma}^+_l)$ be the cross-fitted*

nuisances used in $\phi^{+}_{t,z,h}(S;\widehat{\eta})$ (Eq. (9.35) with $\omega_{z,h}(Z) = K_h(Z-z)$).

Define nuisance error rates (in L_2 norms over the appropriate arguments) by

$$r_{n,\pi} := \|\widehat{\pi}^t - \pi^t\|_2 + \|\widehat{\pi}^g - \pi^g\|_2, \qquad r_{n,Q} := \|\widehat{Q}^{+} - Q^{+}\|_2, \tag{H.3}$$

$$r_{n,\gamma} := \|\widehat{\gamma}^{+}_u - \gamma_u(\widehat{Q}^{+};\cdot)\|_2 + \|\widehat{\gamma}^{+}_l - \gamma_l(\widehat{Q}^{+};\cdot)\|_2, \tag{H.4}$$

where the norms are taken over the random variables that the corresponding nuisance is evaluated on (e.g., $(Z,\mathbf{X})$ for $\pi^g(Z \mid \mathbf{X})$, $Q^{+}(t,Z,\mathbf{X})$, and $\gamma^{\pm}(t,Z,\mathbf{X})$).

Then, the conditional bias induced by nuisance estimation satisfies

$$\Big\| \mathbb{E}\Big[\phi^{+}_{t,z,h}(S;\widehat{\eta}) - \phi^{+}_{t,z,h}(S;\eta) \mid \mathbf{X}\Big] \Big\|_2 = O_p\!\left(\frac{r_{n,\pi}\, r_{n,\gamma} + r_{n,Q}^2}{\sqrt{h}} \right). \tag{H.5}$$

Proof. We mirror the proof of Theorem 9.14 and highlight only the changes required for continuous Z. Fix (t,z) and suppress (t,z) in the notation. As before, by cross-fitting, we may condition on the training folds and treat $\widehat{\eta}$ as fixed when taking expectations over the held-out fold.

Key modification For continuous Z, define

$$A_h := \mathbf{1}_{[T=t]} K_h(Z-z), \qquad \pi(Z,\mathbf{X}) := \pi^t(\mathbf{X})\,\pi^g(Z \mid \mathbf{X}), \qquad \widehat{\pi}(Z,\mathbf{X}) := \widehat{\pi}^t(\mathbf{X})\widehat{\pi}^g(Z \mid \mathbf{X}), \tag{H.95}$$

so that the (true) localized selection weight is

$$\kappa_{t,z,h}(S) = \frac{A_h}{\pi(Z,\mathbf{X})}. \tag{H.96}$$

The discrete-Z algebra carries through with A replaced by A_h and $\pi(\mathbf{X})$ replaced by $\pi(Z,\mathbf{X})$. The only substantive difference is that L_2-norms of kernel-weighted terms pick up a factor $h^{-1/2}$ via $\int K_h(u)^2\,du = O(1/h)$.

A useful kernel moment bound Under Assumptions 9.13 and H.1, there exists a constant $C < \infty$ such that, for any square-integrable measurable function $G(S)$,

$$\left\|\mathbb{E}[\kappa_{t,z,h}(S)\, G(S) \mid \mathbf{X}]\right\|_2 \leq \frac{C}{\sqrt{h}} \|G(S)\|_2. \tag{H.97}$$

Indeed, by conditional Cauchy–Schwarz,

$$\left(\mathbb{E}[\kappa_{t,z,h} G \mid \mathbf{X}]\right)^2 \leq \mathbb{E}[\kappa_{t,z,h}^2 \mid \mathbf{X}]\, \mathbb{E}[G^2 \mid \mathbf{X}], \tag{H.98}$$

and $\mathbb{E}[\kappa_{t,z,h}^2 \mid \mathbf{X}]$ is of order $1/h$ because K_h^2 integrates to $O(1/h)$ and $\pi^t(\mathbf{X}), \pi^g(\cdot \mid \mathbf{X})$ are bounded away from 0 (overlap).

Step 1: Conditional expectation decomposition Write $\phi_h(S; \cdot)$ for Eq. (9.35) with $\omega_{z,h}(Z) = K_h(Z - z)$. As in the discrete proof, take conditional expectations given $\mathbf{X}$ and use iterated expectations to replace the in-sample hinge terms by their corresponding conditional-mean targets (evaluated at the cutoff used in the pseudo-outcome). This yields a decomposition of the form

$$\mathbb{E}[\phi_h(S; \widehat{\eta}) - \phi_h(S; \eta) \mid \mathbf{X}] = \underbrace{\left(\mu_{h,\widehat{Q}^+}(\mathbf{X}) - \mu_h^+(\mathbf{X})\right)}_{\text{cutoff-induced error}} + \underbrace{\mathbb{E}\left[\left(\frac{\pi(Z,\mathbf{X})}{\widehat{\pi}(Z,\mathbf{X})} - 1\right) \kappa_{t,z,h}(S)\, \Delta_{\widehat{Q}^+}(S) \;\middle|\; \mathbf{X}\right]}_{\text{propensity} \times \text{regression product term}}, \tag{H.99}$$

where $\mu_{h,\widehat{Q}^+}(\mathbf{X})$ denotes the bound functional induced by the cutoff $\widehat{Q}^+$ (holding the remaining targets at their population values for that cutoff), and $\Delta_{\widehat{Q}^+}(S)$ collects the hinge-mean regression discrepancies at cutoff $\widehat{Q}^+$ (the continuous-Z analogue of the bracketed term in Eq. (H.58) of the discrete proof).

Step 2: Bounding the product term By overlap, $\widehat{\pi}(Z, \mathbf{X})$ is bounded away from 0; hence

$$\left\|\frac{\pi(Z,\mathbf{X})}{\widehat{\pi}(Z,\mathbf{X})} - 1\right\|_2 \lesssim \|\widehat{\pi}^t - \pi^t\|_2 + \|\widehat{\pi}^g - \pi^g\|_2 = O_p(r_{n,\pi}), \tag{H.100}$$

where norms are taken over the arguments on which the nuisances are evaluated (here $(Z, \mathbf{X})$ for π^g). Moreover, by definition of $r_{n,\gamma}$, $\|\Delta_{\widehat{Q}^+}(S)\|_2 = O_p(r_{n,\gamma})$.

Applying Eq. (H.97) with $G(S) := \left(\frac{\pi}{\widehat{\pi}} - 1\right)\Delta_{\widehat{Q}^+}(S)$ gives

$$\left\|\mathbb{E}\left[\left(\frac{\pi}{\widehat{\pi}} - 1\right)\kappa_{t,z,h}\Delta_{\widehat{Q}^+} \mid \mathbf{X}\right]\right\|_2 \leq \frac{C}{\sqrt{h}}\left\|\left(\frac{\pi}{\widehat{\pi}} - 1\right)\Delta_{\widehat{Q}^+}\right\|_2 = O_p\left(\frac{r_{n,\pi}\, r_{n,\gamma}}{\sqrt{h}}\right). \tag{H.101}$$

Step 3: Bounding the cutoff-induced term The discrete proof bounds the cutoff-induced term using the envelope/FOC property of the cutoff and a second-order Taylor expansion, yielding a quadratic dependence on $\widehat{Q}^+ - Q^+$. The same argument applies here pointwise in the arguments of the cutoff (the cutoff remains an optimizer of the same tail objective, now for the localized target), so

$$|\mu_{h,\widehat{Q}^+}(\mathbf{X}) - \mu_h^+(\mathbf{X})| \lesssim \mathbb{E}\left[\kappa_{t,z,h}(S)\left|\widehat{Q}^+(t,Z,\mathbf{X}) - Q^+(t,Z,\mathbf{X})\right|^2 \;\middle|\; \mathbf{X}\right]. \tag{H.102}$$

Applying Eq. (H.97) with $G(S) := \left|\widehat{Q}^+(t,Z,\mathbf{X}) - Q^+(t,Z,\mathbf{X})\right|^2$ and using the same bounded-moment simplification as in the discrete proof gives

$$\|\mu_{h,\widehat{Q}^+} - \mu_h^+\|_2 = O_p\left(\frac{r_{n,Q}^2}{\sqrt{h}}\right). \tag{H.103}$$

Conclusion. Combining Steps 2–3 in Eq. (H.99) yields

$$\left\|\mathbb{E}[\phi_h(S;\widehat{\eta}) - \phi_h(S;\eta) \mid \mathbf{X}]\right\|_2 = O_p\left(\frac{r_{n,\pi}\, r_{n,\gamma} + r_{n,Q}^2}{\sqrt{h}}\right), \tag{H.104}$$

which is exactly the claim. □

H.4.2 Proof of Corollary H.3

Corollary H.3 (Quasi-oracle rates and inference (continuous Z))**.** *Assume the conditions of Theorem H.2 and that the second-stage regression learner $\widehat{\mathbb{E}}_n[\cdot \mid \mathbf{X} = \mathbf{x}]$ satisfies Assumption 9.15 with rate δ_n when regressing $\phi_{t,z,h}^+(S;\eta)$ on $\mathbf{X}$.*

Then:

<u>CAPO rates:</u> The CAPO upper-bound estimator satisfies

$$\|\widehat{\mu}_h^+(t,z,\cdot) - \mu_h^+(t,z,\cdot)\|_2 = O_p\left(\delta_n + \frac{r_{n,\pi}\, r_{n,\gamma} + r_{n,Q}^2}{\sqrt{h}}\right). \tag{H.6}$$

APO rates: The APO upper-bound estimator $\widehat{\psi}^+(t,z) = \mathbb{E}_n[\widehat{\phi}^+_{t,z,h}]$ satisfies

$$|\widehat{\psi}^+_h(t,z) - \psi^+_h(t,z)| = O_p\left(\frac{1}{\sqrt{nh}} + \frac{r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q}}{\sqrt{h}}\right). \tag{H.7}$$

$\sqrt{nh}$-CLT (central limit theorem) for the (smoothed) APO. If $r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q} = o_p(n^{-1/2})$, then

$$\sqrt{nh}\left(\widehat{\psi}^+_h(t,z) - \psi^+_h(t,z)\right) \rightsquigarrow \mathcal{N}(0,\ V^+_h(t,z))\,, \tag{H.8}$$

where one valid asymptotic variance target is $V^+_h(t,z) := \mathrm{Var}(\sqrt{h}\,\phi^+_{t,z,h}(S;\eta))$.

Finally, if the smoothing bias satisfies $|\psi^+_h(t,z) - \psi^+(t,z)| = o((nh)^{-1/2})$ (e.g., via under-smoothing under z-smoothness), then the CLT holds with $\psi^+(t,z)$ in place of $\psi^+_h(t,z)$.

Proof. We follow the proof of Corollary 9.17, replacing Theorem 9.14 by Theorem H.2 and tracking the kernel-induced scaling.

CAPO rate Let $m^+_{t,z,h}(\mathbf{x}) := \mathbb{E}[\widehat{\phi}^+_{t,z,h} \mid \mathbf{X} = \mathbf{x}]$ denote the conditional mean of the (cross-fitted) localized pseudo-outcome. By Assumption 9.15,

$$\|\widehat{\mu}^+_h(t,z,\cdot) - m^+_{t,z,h}(\cdot)\|_2 = O_p(\delta_n). \tag{H.105}$$

By the triangle inequality,

$$\|\widehat{\mu}^+_h(t,z,\cdot) - \mu^+_h(t,z,\cdot)\|_2 \le \|\widehat{\mu}^+_h(t,z,\cdot) - m^+_{t,z,h}(\cdot)\|_2 + \|m^+_{t,z,h}(\cdot) - \mu^+_h(t,z,\cdot)\|_2. \tag{H.106}$$

The second term is exactly the conditional nuisance-induced bias controlled by Theorem H.2, giving the stated CAPO rate.

APO rate and $\sqrt{nh}$ asymptotic normality Recall $\widehat{\psi}^+_h(t,z) = \mathbb{E}_n[\widehat{\phi}^+_{t,z,h}]$. Decompose

$$\widehat{\psi}^+_h(t,z) - \psi^+_h(t,z) = (\mathbb{E}_n - \mathbb{E})[\phi^+_{t,z,h}(S;\eta)] + \mathbb{E}\Big[\phi^+_{t,z,h}(S;\widehat{\eta}) - \phi^+_{t,z,h}(S;\eta)\Big] + R_{n,h}, \tag{H.107}$$

where $R_{n,h} := (\mathbb{E}_n - \mathbb{E})[\phi^+_{t,z,h}(S;\widehat{\eta}) - \phi^+_{t,z,h}(S;\eta)]$.

Under Assumption H.1 and overlap, $\mathrm{Var}(\phi^+_{t,z,h}(S;\eta)) = O(1/h)$, so $(\mathbb{E}_n - \mathbb{E})[\phi^+_{t,z,h}(S;\eta)] = O_p((nh)^{-1/2})$ by the CLT. With cross-fitting and the same conditioning argument as in the discrete proof, $R_{n,h} = o_p((nh)^{-1/2})$.

The bias term is controlled by Theorem H.2 after integrating over $\mathbf{X}$, yielding the stated APO rate. If additionally $r_{n,\pi} r_{n,\gamma} + r^2_{n,Q} = o_p(n^{-1/2})$, then the bias term and $R_{n,h}$ are $o_p((nh)^{-1/2})$, implying

$$\sqrt{nh}(\widehat{\psi}^+_h(t,z) - \psi^+_h(t,z)) = \sqrt{nh}(\mathbb{E}_n - \mathbb{E})[\phi^+_{t,z,h}(S;\eta)] + o_p(1) \rightsquigarrow \mathcal{N}(0, V^+_h(t,z)), \tag{H.108}$$

with $V^+_h(t,z) = \mathrm{Var}(\sqrt{h}\,\phi^+_{t,z,h}(S;\eta))$. The final undersmoothing statement follows by adding/subtracting $\psi^+(t,z)$ and using the assumed bias condition. □

H.4.3 Proof of Proposition H.4

Proposition H.4 (Consistency for sharp bounds (continuous Z))**.** *Assume the conditions of Corollary H.3 and consider the corresponding lower-bound estimator $\widehat{\mu}^-_h(t,z,\cdot)$ constructed from the lower-bound pseudo-outcome (defined analogously to Eq.* (9.35)*). Suppose $\delta_n = o_p(1)$ and*

$$\frac{r_{n,\pi}\, r_{n,\gamma} + r^2_{n,Q}}{\sqrt{h}} = o_p(1). \tag{H.9}$$

Then,

$$\|\widehat{\mu}^\pm_h(t,z,\cdot) - \mu^\pm_h(t,z,\cdot)\|_2 = o_p(1), \qquad |\widehat{\psi}^\pm_h(t,z) - \psi^\pm_h(t,z)| = o_p(1). \tag{H.10}$$

Consequently, the estimated CAPO and APO intervals converge to the sharp *kernel-localized identified intervals for the bandwidth-indexed targets.*

Moreover, if the smoothing bias vanishes at the appropriate rate (e.g., $|\psi^\pm_h(t,z) - \psi^\pm(t,z)| = o((nh)^{-1/2})$), then the estimated intervals are asymptotically sharp for the original pointwise bounds as $h \downarrow 0$.

Proof. The argument is identical to the proof of Proposition 9.18, replacing Corollary 9.17 by Corollary H.3. Under the stated assumptions, $\delta_n = o_p(1)$ and $\frac{r_{n,\pi} r_{n,\gamma} + r_{n,Q}^2}{\sqrt{h}} = o_p(1)$, hence both CAPO endpoints converge in L_2 to the sharp kernel-localized endpoints $\mu_h^{\pm}(t, z, \cdot)$. The APO convergence follows from the APO rate in Corollary H.3. Finally, if the smoothing bias vanishes at the stated rate, the same conclusion holds for the pointwise (unsmoothed) bounds. □

H.4.4 Proof of Corollary H.5

Corollary H.5 (Asymptotic validity under misspecified cutoffs (continuous Z))**.** *Fix measurable cutoffs $\overline{Q}^{\pm}(t, z, \mathbf{x})$ (not necessarily equal to the sharp cut-offs) and let $\overline{\mu}_h^{\pm}(t, z, \mathbf{x}; \overline{Q}^{\pm})$ and $\overline{\psi}_h^{\pm}(t, z; \overline{Q}^{\pm})$ denote the resulting (possibly non-sharp) kernel-localized bound functionals induced by these cutoffs (i.e., the targets obtained by replacing $Q^{\pm}$ in the pseudo-outcomes and taking the conditional/unconditional expectations as in Eq.* (H.2)*). Then, the induced intervals*

$$[\overline{\mu}_h^{-}(t, z, \mathbf{x}; \overline{Q}^{-}),\ \overline{\mu}_h^{+}(t, z, \mathbf{x}; \overline{Q}^{+})] \quad \textit{and} \quad [\overline{\psi}_h^{-}(t, z; \overline{Q}^{-}),\ \overline{\psi}_h^{+}(t, z; \overline{Q}^{+})] \tag{H.11}$$

are (not necessarily sharp) valid CAPO and APO intervals for the kernel-localized targets. Moreover, if $\widehat{Q}^{\pm} \to \overline{Q}^{\pm}$ in L_2 and either

(i) $(\widehat{\pi}^t, \widehat{\pi}^g)$ *is consistent*, or

(ii) the corresponding tail-moment regressions $(\widehat{\gamma}_u^{\pm}, \widehat{\gamma}_l^{\pm})$ are consistent for the targets induced by $\overline{Q}^{\pm}$,

then the estimated endpoints converge to the induced (conservative) targets and the resulting (C)APO intervals remain asymptotically valid, though potentially conservative. If $\overline{Q}^{\pm}$ equals the sharp cut-offs, then the induced bounds coincide with the sharp bounds, and the intervals are asymptotically sharp as well.

Proof. We adapt the proof of Corollary 9.19 and indicate only the continuous-Z differences.

Step 1: Any cutoffs induce a valid (conservative) localized interval Fix $(t, z, \mathbf{x})$. For each exposure level u, the discrete-Z proof shows (via the same Rockafellar-Uryasev tail objectives) that evaluating the tail objective at an arbitrary cutoff $\overline{Q}^{\pm}(t, u, \mathbf{x})$ yields conservative endpoints $\overline{\mu}^{\pm}(t, u, \mathbf{x}; \overline{Q}^{\pm})$ that contain the sharp pointwise endpoints $\mu^{\pm}(t, u, \mathbf{x})$.

Kernel localization preserves this ordering because $K_h(\cdot) \geq 0$ and integrates to 1. Indeed, the continuous-Z localized targets are obtained by the same conditional-expectation construction as in Eq. (H.2), and the weight $\kappa_{t,z,h}(S)$ transports pointwise statements in u into their localized analogues around z. Therefore,

$$\overline{\mu}_h^{-}(t, z, \mathbf{x}; \overline{Q}^{-}) \;\leq\; \mu_h^{-}(t, z, \mathbf{x}) \;\leq\; \mu_h^{+}(t, z, \mathbf{x}) \;\leq\; \overline{\mu}_h^{+}(t, z, \mathbf{x}; \overline{Q}^{+}), \tag{H.109}$$

such that the CAPO interval is valid (though not necessarily sharp). The APO claim follows by taking expectations over $\mathbf{X}$.

Step 2: Convergence to the induced (conservative) localized bounds The convergence argument follows Step 2 of the discrete-Z proof, with the single change that kernel-weighted terms are controlled using the bound in Eq. (H.97) (hence the extra factor $h^{-1/2}$ in intermediate inequalities). Under $\widehat{Q}^{\pm} \to \overline{Q}^{\pm}$ in L_2 and either (i) $(\widehat{\pi}^t, \widehat{\pi}^g)$ consistent or (ii) $(\widehat{\gamma}_u^{\pm}, \widehat{\gamma}_l^{\pm})$ consistent for the targets induced by $\overline{Q}^{\pm}$, the same decomposition yields

$$\left\| \mathbb{E}\left[\phi_{t,z,h}^{\pm}(S; \widehat{\eta}^{\pm}) \mid \mathbf{X} \right] - \overline{\mu}_h^{\pm}(t, z, \mathbf{X}; \overline{Q}^{\pm}) \right\|_2 = o_p(1). \tag{H.110}$$

With Assumption 9.15 and $\delta_n = o_p(1)$, the second-stage regression therefore implies $\|\widehat{\mu}_h^{\pm}(t, z, \cdot) - \overline{\mu}_h^{\pm}(t, z, \cdot; \overline{Q}^{\pm})\|_2 = o_p(1)$, and the sample-average estimator yields $\widehat{\psi}_h^{\pm}(t, z) \to \overline{\psi}_h^{\pm}(t, z; \overline{Q}^{\pm})$. Thus the estimated intervals converge to the induced (conservative) localized intervals and remain asymptotically valid. If $\overline{Q}^{\pm} = Q^{\pm}$, these limits coincide with the sharp localized bounds. □

BIBLIOGRAPHY

Alberto Abadie. Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, 113(2):231–263, 2003.

Alberto Abadie, Joshua Angrist, and Guido Imbens. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117, 2002.

Alberto Abadie, Jiaying Gu, and Shu Shen. Instrumental variable estimation with first-stage heterogeneity. *Journal of Econometrics*, 240(2):105425, 2024.

Alekh Agarwal, Yuda Song, Wen Sun, Kaiwen Wang, Mengdi Wang, and Xuezhou Zhang. Provable benefits of representational transfer in reinforcement learning. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2114–2187. PMLR, 2023.

Alan Agresti. *Foundations of linear and generalized linear models*. John Wiley & Sons, 2015.

Amir Ahmadi-Javid. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123, 2012.

Chunrong Ai and Xiaohong Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71 (6):1795–1843, 2003.

Elisabeth Ailer, Niclas Dern, Jason S Hartford, and Niki Kilbertus. Targeted sequential indirect experiment design. *Advances in Neural Information Processing Systems*, 37:122029–122053, 2024.

Sahara Ali, Omar Faruque, and Jianwu Wang. Estimating direct and indirect causal effects of spatiotemporal interventions in presence of spatial interference. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 213–230. Springer, 2024.

Alejandro Almodóvar, Adrián Javaloy, Juan Parras, Santiago Zazo, and Isabel Valera. Decaflow: A deconfounding causal generative model. In *Advances in Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=fwsRLgEg8M`.

Heresh Amini, Mahdieh Danesh-Yazdi, Qian Di, Weeberb Requia, Yaguang Wei, Yara Abu-Awad, Liuhua Shi, Meredith Franklin, Choong-Min Kang, Jack Wolfson, Peter James, Rima Habre, Qiao Zhu, Joshua Apte, Zorana Andersen, Itai kloog, Francesca Dominici, Petros Koutrakis, and Joel Schwartz. Hyperlocal super-learned PM2.5 components across the contiguous us, 2022.

Philip Amortila, Dylan J Foster, Nan Jiang, Ayush Sekhari, and Tengyang Xie. Harnessing density ratios for online reinforcement learning. *arXiv preprint arXiv:2401.09681*, 2024.

Isaiah Andrews, James H Stock, and Liyang Sun. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11:727–753, 2019.

Marcus Ang, Jie Sun, and Qiang Yao. On the dual representation of coherent risk measures. *Annals of Operations Research*, 262:29–46, 2018.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

Luc Anselin. *Spatial Econometrics: Methods and Models*, volume 4. Springer Science & Business Media, 1988.

Luc Anselin. *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media, 2013.

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912—-1947, 2017.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

Onur Atan, James Jordon, and Mihaela Van der Schaar. Deep-treat: Learning optimal personalized treatments from observational data using neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.

Susan Athey, Raj Chetty, and Guido Imbens. Combining experimental and observational data to estimate treatment effects on long term outcomes. *arXiv preprint arXiv:2006.09676*, 2020.

Jean-Yves Audibert and Alexandre B. Tsybakov. Fast learning rates for plug-in classifiers. *The Annals of Statistics*, 35(2):608–633, 2007.

Alex Ayoub, Kaiwen Wang, Vincent Liu, Samuel Robertson, James McInerney, Dawen Liang, Nathan Kallus, and Csaba Szepesvári. Switching the loss reduces the cost in batch reinforcement learning. *International Conference of Machine Learning*, 2024.

Kishan Panaganti Badrinath and Dileep Kalathil. Robust reinforcement learning using least squares policy iteration with provable performance guarantees. In *International Conference on Machine Learning*, pages 511–520. PMLR, 2021.

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Albert-Laszlo Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

Keith Battocchi, Eleanor Dillon, Maggie Hei, Greg Lewis, Paul Oka, Miruna Oprescu, and Vasilis Syrgkanis. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/microsoft/EconML, 2019. Version 0.13.

Alexandre Belloni, Victor Chernozhukov, Iván Fernández-Val, and Christian Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298, 2017a.

Alexandre Belloni, Victor Chernozhukov, Ivan Fernandez-Val, and Christian Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298, 2017b.

Eli Ben-Michael, Avi Feller, and Jesse Rothstein. Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):351–381, 2022.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

Andrew Bennett and Nathan Kallus. The variational method of moments. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):810–841, 2023.

Andrew Bennett, Nathan Kallus, and Tobias Schnabel. Deep generalized method of moments for instrumental variable analysis. *Advances in neural information processing systems*, 32, 2019.

Andrew Bennett, Nathan Kallus, Miruna Oprescu, Wen Sun, and Kaiwen Wang. Efficient and sharp off-policy evaluation in robust markov decision processes. *Advances in Neural Information Processing Systems*, 37:112962–113000, 2025.

Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 813–824. PMLR, 18–24 Jul 2021.

Pallab K Bhattacharya and Ashis K Gangopadhyay. Kernel and nearest-neighbor estimation of a conditional quantile. *The Annals of Statistics*, pages 1400–1415, 1990.

Ioana Bica, Ahmed Alaa, and Mihaela Van Der Schaar. Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In *International conference on machine learning*, pages 884–895. PMLR, 2020a.

Ioana Bica, Ahmed M. Alaa, James Jordon, and Mihaela van der Schaar. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In *International Conference on Learning Representations*, 2020b.

Peter J Bickel, Chris AJ Klaassen, Peter J Bickel, Ya'acov Ritov, J Klaassen, Jon A Wellner, and YA'Acov Ritov. *Efficient and adaptive estimation for semiparametric models*, volume 4. Johns Hopkins University Press Baltimore, 1993.

Marianne P Bitler, Jonah B Gelbach, and Hilary W Hoynes. What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, 96(4):988–1012, 2006.

Iavor Bojinov and Neil Shephard. Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*, 2019.

Matteo Bonvini, Edward Kennedy, Valerie Ventura, and Larry Wasserman. Sensitivity analysis for marginal structural models. *arXiv preprint arXiv:2210.04681*, 2022.

Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

Haïm Brezis and Petru Mironescu. Where sobolev interacts with gagliardo–nirenberg. *Journal of functional analysis*, 277(8):2839–2864, 2019.

David Bruns-Smith and Angela Zhou. Robust fitted-q-evaluation and iteration under sequentially exogenous unobserved confounders. *arXiv preprint arXiv:2302.00662*, 2023.

David A Bruns-Smith. Model-free and model-based policy evaluation when causality is uncertain. In *International Conference on Machine Learning*, pages 1116–1126. PMLR, 2021.

Rich Caruana. Multitask learning. *Machine learning*, 28:41–75, 1997.

Wayne E Cascio. Wildland fire smoke and human health. *Science of the total environment*, 624:586–595, 2018.

Centers for Disease Control and Prevention. Behavioral risk factor surveillance system (BRFSS), 2010.

Domagoj Cevid, Loris Michel, Jeffrey Näf, Peter Bühlmann, and Nicolai Meinshausen. Distributional random forests: Heterogeneity adjustment and multivariate distributional regression. *Journal of Machine Learning Research*, 23(333): 1–79, 2022.

Gary Chamberlain. Efficiency bounds for semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 567–596, 1992.

Yash Chandak, Shiv Shankar, Vasilis Syrgkanis, and Emma Brunskill. Adaptive instrument design for indirect experiments. In *International Conference on Learning Representations*, 2024.

Jonathan Chang, Kaiwen Wang, Nathan Kallus, and Wen Sun. Learning bellman complete representations for offline policy evaluation. In *International Conference on Machine Learning*, pages 2938–2971. PMLR, 2022.

Weilin Chen, Ruichu Cai, Zeqin Yang, Jie Qiao, Yuguang Yan, Zijian Li, and Zhifeng Hao. Doubly robust causal effect estimation under networked interference via targeted learning. In *International Conference on Machine Learning (ICML)*, 2024a.

Weilin Chen, Ruichu Cai, Zeqin Yang, Jie Qiao, Yuguang Yan, Zijian Li, and Zhifeng Hao. Doubly robust causal effect estimation under networked interference via targeted learning. In *Proceedings of the 41st International Conference on Machine Learning*, 2024b.

Xiaohong Chen and Demian Pouzo. Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, 152(1):46–60, 2009.

Zonghao Chen, Ruocheng Guo, Jean-François Ton, and Yang Liu. Conformal counterfactual inference under hidden confounding. In *Conference on Knowledge Discovery and Data Mining (KDD)*, 2024c.

David Cheng and Tianxi Cai. Adaptive combination of randomized and observational data. *arXiv preprint arXiv:2111.15012*, 2021.

Jing Cheng, Dylan S Small, Zhiqiang Tan, and Thomas R Ten Have. Efficient nonparametric estimation of causal effects in randomized trials with noncompliance. *Biometrika*, 96(1):19–36, 2009.

Lu Cheng, Ruocheng Guo, Raha Moraffah, Paras Sheth, K. Selcuk Candan, and Huan Liu. Evaluation Methods and Measures for Causal Learning Algorithms . *IEEE Transactions on Artificial Intelligence*, 3(06):924–943, 2022.

Victor Chernozhukov and Christian Hansen. The effects of 401 (k) participa-

tion on the wealth distribution: an instrumental quantile regression analysis. *Review of Economics and statistics*, 86(3):735–751, 2004.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018a.

Victor Chernozhukov, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research, 2018b.

Victor Chernozhukov, Ivan Fernandez-Val, and Martin Weidner. Network and panel quantile effects via distribution regression. *Journal of Econometrics*, 2020.

Victor Chernozhukov, Carlos Cinelli, Whitney Newey, Amit Sharma, and Vasilis Syrgkanis. Long story short: Omitted variable bias in causal machine learning. Technical report, National Bureau of Economic Research, 2022a.

Victor Chernozhukov, Whitney Newey, Rahul Singh, and Vasilis Syrgkanis. Automatic debiased machine learning for dynamic treatment effects and general nested functionals. *arXiv preprint arXiv:2203.13887*, 2022b.

Yinlam Chow, Aviv Tamar, Shie Mannor, and Marco Pavone. Risk-sensitive and robust decision-making: a cvar optimization approach. *Advances in neural information processing systems*, 28, 2015.

Rune Christiansen, Matthias Baumann, Tobias Kuemmerle, Miguel D Mahecha, and Jonas Peters. Toward causal inference for spatio-temporal data: Conflict

and forest loss in Colombia. *Journal of the American Statistical Association*, 117 (538):591–601, 2022.

Paul S Clarke and Frank Windmeijer. Instrumental variable estimators for binary outcomes. *Journal of the American Statistical Association*, 107(500):1638–1652, 2012.

Stephanie E Cleland, Marc L Serre, Ana G Rappold, and J Jason West. Estimating the acute health impacts of fire-originated PM2.5 exposure during the 2017 California wildfires: Sensitivity to choices of inputs. *Geohealth*, 5(7): e2021GH000414, 2021.

Bénédicte Colnet, Julie Josse, Gaël Varoquaux, and Erwan Scornet. Causal effect on a target population: a sensitivity analysis to handle missing covariates. *Journal of Causal Inference*, 10(1):372–414, 2022.

Thomas Cook, Alan Mishler, and Aaditya Ramdas. Semiparametric efficient inference in adaptive experiments. In *Causal Learning and Reasoning*, pages 1033–1064. PMLR, 2024.

Jerome Cornfield, William Haenszel, E Cuyler Hammond, Abraham M Lilienfeld, Michael B Shimkin, and Ernst L Wynder. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer institute*, 22(1):173–203, 1959.

Amanda Coston, Edward Kennedy, and Alexandra Chouldechova. Counterfactual predictions under runtime confounding. *Advances in neural information processing systems*, 33:4150–4162, 2020.

Stephen Coussens and Jann Spiess. Improving inference from simple instruments through compliance estimation. *arXiv preprint arXiv:2108.03726*, 2021.

Richard K Crump, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik. Nonparametric tests for treatment effect heterogeneity. *The Review of Economics and Statistics*, 90(3):389–405, 2008.

Alicia Curth, Ahmed M Alaa, and Mihaela van der Schaar. Estimating structural target functions using machine learning and influence functions. *arXiv preprint arXiv:2008.06461*, 2020.

Alicia Curth, David Svensson, Jim Weatherall, and Mihaela van der Schaar. Really doing great at estimating CATE? a critical look at ML benchmarking practices in treatment effect estimation. In *Advances Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Will Dabney, Georg Ostrovski, David Silver, and Rémi Munos. Implicit quantile networks for distributional reinforcement learning. In *International conference on machine learning*, pages 1096–1105. PMLR, 2018a.

Will Dabney, Mark Rowland, Marc Bellemare, and Rémi Munos. Distributional reinforcement learning with quantile regression. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018b.

Jessica Dai, Paula Gradu, and Christopher Harshaw. Clip-ogd: An experimental design for adaptive neyman allocation in sequential experiments. *Advances in Neural Information Processing Systems*, 36:32235–32269, 2023.

Abhinandan Dalal, Patrick Blöbaum, Shiva Kasiviswanathan, and Aaditya Ramdas. Anytime-valid inference for double/debiased machine learning of causal parameters. *arXiv preprint arXiv:2408.09598*, 2024.

Stephanie DeFlorio-Barker, James Crooks, Jeanette Reyes, and Ana G Rappold. Cardiopulmonary effects of fine particulate matter exposure among older

adults, during wildfire and non-wildfire periods, in the United States 2008–2010. *Environmental health perspectives*, 127(3):037006, 2019.

Riccardo Della Vecchia and Debabrota Basu. Stochastic online instrumental variable regression: Regrets for endogeneity and bandit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 16190–16198, 2025.

Tatyana Deryugina, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif. The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219, 2019.

Qian Di, Heresh Amini, Liuhua Shi, Itai Kloog, Rachel Silvern, James Kelly, M. Benjamin Sabath, Christine Choirat, Petros Koutrakis, Alexei Lyapustin, Yujie Wang, Loretta J. Mickley, and Joel Schwartz. An ensemble-based model of pm2.5 concentration across the contiguous united states with high spatiotemporal resolution. *Environment International*, 130:104909, 2019.

Nishanth Dikkala, Greg Lewis, Lester Mackey, and Vasilis Syrgkanis. Minimax estimation of conditional moment models. *Advances in Neural Information Processing Systems*, 33:12248–12262, 2020.

Jacob Dorn and Kevin Guo. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *Journal of the American Statistical Association*, 0(0):1–13, 2022. doi: 10.1080/01621459.2022.2069572.

Jacob Dorn and Kevin Guo. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *Journal of the American Statistical Association*, 118(544):2645–2657, 2023.

Jacob Dorn, Kevin Guo, and Nathan Kallus. Doubly-valid/doubly-sharp sensitivity analysis for causal inference with unmeasured confounding. *Journal of the American Statistical Association*, 120(549):331–342, 2025a.

Jacob Dorn, Kevin Guo, and Nathan Kallus. Doubly-valid/doubly-sharp sensitivity analysis for causal inference with unmeasured confounding. *Journal of the American Statistical Association*, 120(549):331–342, 2025b.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Yihan Du, Siwei Wang, and Longbo Huang. Provably efficient risk-sensitive reinforcement learning: Iterated cvar and worst path. In *The Eleventh International Conference on Learning Representations*, 2022.

Yaqi Duan, Chi Jin, and Zhiyuan Li. Risk bounds and rademacher complexity in batch reinforcement learning. In *International Conference on Machine Learning*, pages 2892–2902. PMLR, 2021.

Emiko Dupont, Simon N Wood, and Nicole H Augustin. Spatial+: A novel approach to spatial confounding. *Biometrics*, 78(4):1279–1290, 2022.

Aryeh Dvoretzky. Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, volume 6, pages 513–536. University of California Press, 1972.

Anouar El Ghouch and Marc G Genton. Local polynomial quantile regression with parametric features. *Journal of the American Statistical Association*, 104 (488):1416–1429, 2009.

Kevin Elie-Dit-Cosaque and Véronique Maume-Deschamps. Random forest estimation of conditional distribution functions and conditional quantiles. *Electronic Journal of Statistics*, 16(2):6553–6583, 2022.

Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505*, 2020.

Theodoros Evgeniou and Massimiliano Pontil. Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 109–117, 2004.

Max H Farrell, Tengyuan Liang, and Sanjog Misra. Deep neural networks for estimation and inference. *Econometrica*, 89(1):181–213, 2021.

Sergio Firpo. Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276, 2007.

Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918, 2021.

Laura Forastiere, Edoardo M. Airoldi, and Fabrizia Mealli. Estimands and identification of average causal effects on networks. *Journal of the American Statistical Association*, 117(541):607–618, 2022.

Dylan J Foster and Vasilis Syrgkanis. Orthogonal statistical learning. *The Annals of Statistics*, 51(3):879–908, 2023a.

Dylan J Foster and Vasilis Syrgkanis. Orthogonal statistical learning. *The Annals of Statistics*, 51(3):879–908, 2023b.

Dennis Frauen and Stefan Feuerriegel. Estimating individual treatment effects under unobserved confounding using binary instruments. In *International Conference on Learning Representations*, 2023.

Dennis Frauen, Valentyn Melnychuk, and Stefan Feuerriegel. Sharp bounds for generalized causal sensitivity analysis. *Advances in Neural Information Processing Systems*, 2023.

Dennis Frauen, Valentyn Melnychuk, and Stefan Feuerriegel. Sharp bounds for generalized causal sensitivity analysis. *Advances in Neural Information Processing Systems*, 36, 2024.

Dennis Frauen, Konstantin Hess, and Stefan Feuerriegel. Model-agnostic meta-learners for estimating heterogeneous treatment effects over time. In *International Conference on Learning Representations*, 2025.

Qiyang Ge, Xuelin Huang, Shenying Fang, Shicheng Guo, Yuanyuan Liu, Wei Lin, and Momiao Xiong. Conditional generative adversarial networks for individualized treatment effect estimation and treatment selection. *Frontiers in genetics*, page 1578, 2020.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

Arthur S Goldberger. Structural equation methods in the social sciences. *Econometrica: Journal of the Econometric Society*, pages 979–1001, 1972.

Noah Golowich, Alexander Rakhlin, and Ohad Shamir. Size-independent sample complexity of neural networks. In *Conference On Learning Theory*, pages 297–299. PMLR, 2018.

Google. Google colab. `https://colab.research.google.com/`, 2024. Accessed: April 2024.

Vineet Goyal and Julien Grand-Clement. Robust markov decision processes: Beyond rectangularity. *Mathematics of Operations Research*, 48(1):203–226, 2023.

DRAFT GUIDANCE. Adaptive designs for clinical trials of drugs and biologics. *Center for Biologics Evaluation and Research (CBER)*, 2018.

Shantanu Gupta, Zachary Lipton, and David Childers. Efficient online estimation of causal effects by deciding what to observe. *Advances in Neural Information Processing Systems*, 34:20995–21007, 2021.

Jinyong Hahn, Keisuke Hirano, and Dean Karlan. Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29(1): 96–108, 2011.

P Richard Hahn, Jared S Murray, and Carlos M Carvalho. Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis*, 15(3):965–1056, 2020.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 2017.

Ephraim M Hanks, Erin M Schliep, Mevin B Hooten, and Jennifer A Hoeting. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4):243–254, 2015.

Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep iv: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, pages 1414–1423. PMLR, 2017.

Tobias Hatt and Stefan Feuerriegel. Sequential deconfounding for causal inference with unobserved confounders. In *Causal Learning and Reasoning*, pages 934–956. PMLR, 2024.

Tobias Hatt, Jeroen Berrevoets, Alicia Curth, Stefan Feuerriegel, and Mihaela van der Schaar. Combining observational and randomized data for estimating heterogeneous treatment effects. *arXiv preprint arXiv:2202.12891*, 2022.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

James J Heckman, Jeffrey Smith, and Nancy Clements. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4):487–535, 1997.

Konstantin Hess, Dennis Frauen, Valentyn Melnychuk, and Stefan Feuerriegel. Igc-net for conditional average potential outcome estimation over time. In *The Fourteenth International Conference on Learning Representations*, 2024.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. Beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017.

Jennifer L Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

Oliver Hines, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 76(3):292–304, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Kate Ho and Adam Rosen. Partial identification in applied research: Benefits and challenges. In Bo Honore, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, editors, *Advances in Economics and Econometrics: Eleventh World Congress (Econometrics Society Monographs)*, volume II, pages 307–359. Cambridge University Press, Cambridge, 2017.

James S Hodges and Brian J Reich. Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325–334, 2010.

Maike Hohberg, Peter Pütz, and Thomas Kneib. Treatment effects beyond the mean using distributional regression: Methods and guidance. *PloS one*, 15(2): e0226514, 2020.

Ronald A Howard and James E Matheson. Risk-sensitive markov decision processes. *Management science*, 18(7):356–369, 1972.

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 18:1–29, 2021.

Jesse Y. Hsu and Dylan S. Small. Calibrating sensitivity analyses to observed covariates in observational studies. *Biometrics*, 69(4):803–811, 2013. ISSN 0006341X, 15410420.

Aiwei Huang, Madhurima Chandra, and Laura Malkhasyan. Weak instrumental variables: Limitations of traditional 2sls and exploring alternative instrumental variable estimators. *arXiv preprint arXiv:2104.12370*, 2021.

Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

Hidehiko Ichimura and Whitney K Newey. The influence function of semiparametric estimators. *Quantitative Economics*, 13(1):29–61, 2022.

Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

Guido Imbens, Nathan Kallus, and Xiaojie Mao. Controlling for unmeasured confounding in panel data using minimal bridge functions: From two-way fixed effects to factor models. *arXiv preprint arXiv:2108.03849*, 2021.

Guido Imbens, Nathan Kallus, Xiaojie Mao, and Yuhao Wang. Long-term causal inference under persistent confounding via data combination. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(2):362–388, 2025.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Garud N Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

Daniel Jacob. Cross-fitting and averaging for machine learning estimation of heterogeneous treatment effects. *arXiv preprint arXiv:2007.02852*, 2020.

Adrián Javaloy, Pablo Sánchez-Martín, and Isabel Valera. Causal normalizing flows: from theory to practice. *Advances in Neural Information Processing Systems*, 2023.

Andrew Jesson, Sören Mindermann, Yarin Gal, and Uri Shalit. Quantifying ignorance in individual-level causal-effect estimates under hidden confounding. In *International Conference on Machine Learning*, pages 4829–4838. PMLR, 2021.

Nan Jiang and Lihong Li. Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 652–661. PMLR, 2016.

Fredrik Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In *International conference on machine learning*, pages 3020–3029. PMLR, 2016.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, Max Albert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe, and François Leblanc. geopandas/geopandas: v0.8.1. Zenodo, July 2020. URL `https://doi.org/10.5281/zenodo.3946761`.

Nathan Kallus. What's the harm? sharp bounds on the fraction negatively affected by treatment. *Advances in Neural Information Processing Systems*, 35: 15996–16009, 2022.

Nathan Kallus. Treatment effect risk: Bounds and inference. *Management Science*, 69(8):4579–4590, 2023.

Nathan Kallus and Xiaojie Mao. Debiased inference on identified linear functionals of underidentified nuisances via penalized minimax estimation. *arXiv preprint arXiv:2208.08291*, 2022.

Nathan Kallus and Miruna Oprescu. Robust and agnostic learning of conditional distributional treatment effects. In *International Conference on Artificial Intelligence and Statistics*, pages 6037–6060. PMLR, 2023a.

Nathan Kallus and Miruna Oprescu. Robust and agnostic learning of conditional distributional treatment effects. In *International Conference on Artificial Intelligence and Statistics*, pages 6037–6060. PMLR, 2023b.

Nathan Kallus and Masatoshi Uehara. Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *The Journal of Machine Learning Research*, 21(1):6742–6804, 2020.

Nathan Kallus and Masatoshi Uehara. Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning. *Operations Research*, 70(6):3282–3302, 2022.

Nathan Kallus and Angela Zhou. Confounding-robust policy improvement. *Advances in neural information processing systems*, 31, 2018.

Nathan Kallus and Angela Zhou. Confounding-robust policy evaluation in

infinite-horizon reinforcement learning. *Advances in neural information processing systems*, 33:22293–22304, 2020.

Nathan Kallus, Aahlad Manas Puli, and Uri Shalit. Removing hidden confounding by experimental grounding. *Advances in neural information processing systems*, 31, 2018.

Nathan Kallus, Xiaojie Mao, and Angela Zhou. Interval estimation of individual-level causal effects under unobserved confounding. In *Proceedings of the 22nd International Conference on Aritificial Intelligence and Statistics (AISTATS) 2019*, volume 89, 2019.

Nathan Kallus, Xiaojie Mao, Kaiwen Wang, and Zhengyuan Zhou. Doubly robust distributionally robust off-policy evaluation and learning. In *International Conference on Machine Learning*, pages 10598–10632. PMLR, 2022.

Nathan Kallus, Xiaojie Mao, and Masatoshi Uehara. Localized debiased machine learning: Efficient inference on quantile treatment effects and beyond. *Journal of Machine Learning Research*, 25(16):1–59, 2024.

Hyunseung Kang, Anru Zhang, T Tony Cai, and Dylan S Small. Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *Journal of the American statistical Association*, 111 (513):132–144, 2016.

Masahiro Kato, Takuya Ishihara, Junya Honda, and Yusuke Narita. Efficient adaptive experimental design for average treatment effect estimation. *arXiv preprint arXiv:2002.05308*, 2020.

Masahiro Kato, Kenichiro McAlinn, and Shota Yasui. The adaptive doubly ro-

bust estimator and a paradox concerning logging policy. *Advances in neural information processing systems*, 34:1351–1364, 2021.

Masahiro Kato, Akihiro Oga, Wataru Komatsubara, and Ryo Inokuchi. Active adaptive experimental design for treatment effect estimation with covariate choice. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 23291–23323. PMLR, 2024.

Luke J Keele and Rocio Titiunik. Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155, 2015.

Edward H Kennedy. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656, 2019.

Edward H. Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008 – 3049, 2023a. doi: 10.1214/23-EJS2157.

Edward H Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008–3049, 2023b.

Edward H Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008–3049, 2023c.

Edward H Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008–3049, 2023d.

Edward H Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *Handbook of statistical methods for precision medicine*, pages 207–236, 2024.

Edward H. Kennedy, Sivaraman Balakrishnan, and Max G'Sell. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008–2030, 2020. doi: 10.1214/19-AOS1874.

David M Kent, Peter M Rothwell, John PA Ioannidis, Doug G Altman, and Rodney A Hayward. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials*, 11(1):1–11, 2010.

Ilyes Khemakhem, Ricardo Monti, Robert Leech, and Aapo Hyvarinen. Causal autoregressive flows. In *International conference on artificial intelligence and statistics*, 2021.

Khashayar Khosravi, Greg Lewis, and Vasilis Syrgkanis. Non-parametric inference adaptive to intrinsic dimension. In Bernhard Schölkopf, Caroline Uhler, and Kun Zhang, editors, *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177 of *Proceedings of Machine Learning Research*, pages 373–389. PMLR, 11–13 Apr 2022. URL `https://proceedings.mlr.press/v177/khosravi22a.html`.

Ayush Khot, Miruna Oprescu, Maresa Schröder, Ai Kagawa, and Xihaier Luo. Spatial deconfounder: Interference-aware deconfounding for spatial causal inference. *arXiv preprint arXiv:2510.08762*, 2025.

Amirhossein Kiani, Chris Wang, and Angela Xu. Sepsis world model: A mimic-based openai gym" world model" simulator for sepsis treatment. *arXiv preprint arXiv:1912.07127*, 2019.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems*, 2021.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *International Conference on Learning Representations*, 2013.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL `https://openreview.net/forum?id=SJU4ayYgl`.

Michael C Knaus. Double machine learning-based programme evaluation under unconfoundedness. *The Econometrics Journal*, 25(3):602–627, 2022.

Murat Kocaoglu, Christopher Snyder, Alexandros G. Dimakis, and Sriram Vishwanath. Causalgan: Learning causal implicit generative models with adversarial training. In *International Conference on Learning Representations*, 2018.

Roger Koenker. *Quantile regression*, volume 38. Cambridge university press, 2005.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

J Kolter. The fixed points of off-policy td. *Advances in Neural Information Processing Systems*, 24, 2011.

Adit Krishnan, Ashish Sharma, and Hari Sundaram. Insights from the long-tail: Learning latent representations of online user behavior in the presence of skew and sparsity. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 297–306, 2018.

Navdeep Kumar, Kfir Levy, Kaixin Wang, and Shie Mannor. Efficient policy iteration for robust markov decision processes via regularization. *arXiv preprint arXiv:2205.14327*, 2022.

Navdeep Kumar, Esther Derman, Matthieu Geist, Kfir Yehuda Levy, and Shie Mannor. Policy gradient for rectangular robust markov decision processes. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL `https://openreview.net/forum?id=NLpXRrjpa6`.

Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

Milan Kuzmanovic, Tobias Hatt, and Stefan Feuerriegel. Deconfounding temporal autoencoder: estimating treatment effects over time using noisy proxies. In *Machine Learning for Health*, pages 143–155. PMLR, 2021.

Mark J Laan and James M Robins. *Unified methods for censored longitudinal data and causality*. Springer, 2003.

Balaji Lakshminarayanan, Daniel M Roy, and Yee Whye Teh. Mondrian forests: Efficient online random forests. *Advances in neural information processing systems*, 27, 2014.

Alexandra Larsen, Shu Yang, Brian J Reich, and Ana G Rappold. A spatial causal analysis of wildland fire-contributed pm2. 5 using numerical model output. *The annals of applied statistics*, 16(4):2714, 2022.

Jeonghwan Lee and Cong Ma. Off-policy estimation with adaptively collected data: the power of online learning. *Advances in Neural Information Processing Systems*, 37:133908–133947, 2024.

Giovanni Leoni. *A first course in Sobolev spaces*. American Mathematical Soc., 2017.

Liu Leqi and Edward H Kennedy. Median optimal treatment regimes. *arXiv preprint arXiv:2103.01802*, 2021.

Noemie Letellier, Maren Hale, Kasem U Salim, Yiqun Ma, Francois Rerolle, Lara Schwarz, and Tarik Benmarhnia. Applying a two-stage generalized synthetic control approach to quantify the heterogeneous health effects of extreme weather events: A 2018 large wildfire in California event as a case study. *Environmental Epidemiology*, 9(1):e362, 2025.

Greg Lewis and Vasilis Syrgkanis. Double/debiased machine learning for dynamic treatment effects. In *NeurIPS*, pages 22695–22707, 2021.

Jiachun Li, David Simchi-Levi, and Yunxiao Zhao. Optimal adaptive experimental design for estimating treatment effect. *arXiv preprint arXiv:2410.05552*, 2024.

Rui Li, Stephanie Hu, Mingyu Lu, Yuria Utsumi, Prithwish Chakraborty, Daby M Sow, Piyush Madan, Jun Li, Mohamed Ghalwash, Zach Shahn, et al. G-net: a recurrent network approach to g-computation for counterfactual prediction under a dynamic treatment regime. In *Machine Learning for Health*, pages 282–299. PMLR, 2021.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018.

Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E. Gonzalez,

and Ion Stoica. Tune: A research platform for distributed model selection and training, 2018.

Bryan Lim. Forecasting treatment responses over time using recurrent marginal structural networks. *Advances in neural information processing systems*, 31, 2018.

Yiqi Lin, Frank Windmeijer, Xinyuan Song, and Qingliang Fan. On the instrumental variable estimation with many weak and invalid instruments. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae025, 2024.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *IEEE/CVF International Conference on Computer Vision*, 2021.

Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. Video swin transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3202–3211, 2022.

Michel Loève. *Elementary probability theory*. Springer, 1977.

Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. *Advances in neural information processing systems*, 30, 2017.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Wenao Ma, Cheng Chen, Jill Abrigo, Calvin Hoi-Kwan Mak, Yuqi Gong, Nga Yan Chan, Chu Han, Zaiyi Liu, and Qi Dou. Treatment outcome prediction for intracerebral hemorrhage via generative prognostic model with

imaging and tabular data. In *International conference on medical image computing and computer-assisted intervention*, 2023.

Lars Maaløe, Casper Kaae Sønderby, Søren Kaae Sønderby, and Ole Winther. Auxiliary deep generative models. In *International Conference on Machine Learning*, 2016.

Shie Mannor, Ofir Mebel, and Huan Xu. Robust mdps with k-rectangular uncertainty. *Mathematics of Operations Research*, 41(4):1484–1509, 2016.

Nicolai Meinshausen and Greg Ridgeway. Quantile regression forests. *Journal of Machine Learning Research*, 7(6), 2006.

Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Causal transformer for estimating counterfactual outcomes. In *International Conference on Machine Learning*, pages 15293–15329. PMLR, 2022.

Volodymyr Mnih. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

Gemma Elyse Moran, Dhanya Sridhar, Yixin Wang, and David Blei. Identifiable deep generative models via sparse decoding. *Transactions on Machine Learning Research*, 2022.

Rémi Munos and Csaba Szepesvári. Finite-time bounds for fitted value iteration. *Journal of Machine Learning Research*, 9(5), 2008.

Hongseok Namkoong, Ramtin Keramati, Steve Yadlowsky, and Emma Brunskill. Off-policy policy evaluation for sequential decisions under unobserved confounding. *Advances in Neural Information Processing Systems*, 33:18819–18831, 2020.

National Energy Research Scientific Computing Center. NERSC. https://nersc.gov, 2025. Accessed: 2025-05-14.

Ojash Neopane, Aaditya Ramdas, and Aarti Singh. Optimistic algorithms for adaptive estimation of the average treatment effect. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 45895–45910. PMLR, 2025a.

Ojash Neopane, Aaditya Ramdas, and Aarti Singh. Logarithmic neyman regret for adaptive estimation of the average treatment effect. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 4303–4311. PMLR, 2025b.

Ojash Neopane, Aaditya Ramdas, and Aarti Singh. Optimistic algorithms for adaptive estimation of the average treatment effect. *arXiv preprint arXiv:2502.04673*, 2025c.

Jerzy Neyman. Optimal asymptotic tests of composite hypotheses. *Probability and statsitics*, pages 213–234, 1959.

Jerzy Neyman. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. In *Breakthroughs in statistics: Methodology and distribution*, pages 123–150. Springer, 1992.

Richard Nickl and Benedikt M Pötscher. Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov-and Sobolev-type. *Journal of Theoretical Probability*, 20(2):177–199, 2007.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.

Arnab Nilim and Laurent El Ghaoui. Robust control of markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

Elizabeth L Ogburn, Andrea Rotnitzky, and James M Robins. Doubly robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77(2):373–396, 2015.

Elizabeth L. Ogburn, Rohit Bhattacharya, Fredrik Sävje, Ilya Shpitser, Rachel Stringham, Lingjiao Li, and Maya Mathur. Causal inference in the presence of interference. *Annual Review of Statistics and Its Application*, 11:341–367, 2024.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. In *Proceedings of the 1st Conference on Medical Imaging with Deep Learning*, 2018.

Tomasz Olma. Nonparametric estimation of truncated conditional expectation functions. *arXiv preprint arXiv:2109.06150*, 2021.

Miruna Oprescu and Nathan Kallus. Estimating heterogeneous treatment effects by combining weak instruments and observational data. *Advances in Neural Information Processing Systems*, 37:118777–118806, 2025. URL `https://proceedings.neurips.cc/paper_files/paper/2024/file/d738aefead8500f5aed667f0a7ca7b7c-Paper-Conference.pdf`.

Miruna Oprescu, Vasilis Syrgkanis, and Zhiwei Steven Wu. Orthogonal random forest for causal inference. In *International Conference on Machine Learning*, pages 4932–4941. PMLR, 2019.

Miruna Oprescu, Jacob Dorn, Marah Ghoummaid, Andrew Jesson, Nathan Kallus, and Uri Shalit. B-learner: Quasi-oracle bounds on heterogeneous causal effects under hidden confounding. In *International Conference on Machine Learning*, pages 26599–26618. PMLR, 2023.

Miruna Oprescu, Andrew Bennett, and Nathan Kallus. Low-rank mdps with continuous action spaces. In *International Conference on Artificial Intelligence and Statistics*, pages 4069–4077. PMLR, 2024.

Miruna Oprescu, Brian M Cho, and Nathan Kallus. Efficient adaptive experimentation with noncompliance. *Advances in Neural Information Processing Systems*, 2025. To appear.

Miruna Oprescu, David K Park, Xihaier Luo, Shinjae Yoo, and Nathan Kallus. Gst-unet: A neural framework for spatiotemporal causal inference with time-varying confounding. *Advances in Neural Information Processing Systems*, 2025. To appear.

Kishan Panaganti, Zaiyan Xu, Dileep Kalathil, and Mohammad Ghavamzadeh. Robust reinforcement learning using offline data. *Advances in neural information processing systems*, 35:32211–32224, 2022.

Georgia Papadogeorgou and Srijata Samanta. Spatial causal inference in the presence of unmeasured confounding and interference. *arXiv preprint arXiv:2303.08218*, 2023.

Georgia Papadogeorgou, Christine Choirat, and Corwin M Zigler. Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. *Biostatistics*, 20(2):256–272, 2019a.

Georgia Papadogeorgou, Fabrizia Mealli, and Corwin M Zigler. Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics*, 75(3):778–787, 2019b.

Georgia Papadogeorgou, Kosuke Imai, Jason Lyall, and Fan Li. Causal inference with spatio-temporal data: estimating the effects of airstrikes on insurgent violence in iraq. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):1969–1999, 2022.

Junhyung Park, Uri Shalit, Bernhard Schölkopf, and Krikamol Muandet. Conditional distributional treatment effect with kernel conditional mean embeddings and u-statistic regression. In *International Conference on Machine Learning*, pages 8401–8412. PMLR, 2021.

Samir Passi and Solon Barocas. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 39–48, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. NeurIPS, 2019.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

James M Poterba, Steven F Venti, and David A Wise. 401 (k) plans and tax-deferred saving. *Studies in the Economics of Aging*, pages 105–142, 1994.

Philipp Probst, Marvin N Wright, and Anne-Laure Boulesteix. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3):e1301, 2019.

Wei Qian and Yuhong Yang. Kernel estimation and model combination in a bandit problem with covariates. *Journal of Machine Learning Research*, 17(149): 1–37, 2016.

Jeffrey S Racine and Kevin Li. Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach. *Journal of Econometrics*, 201(1): 72–94, 2017.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4): 576–601, 2023.

Colleen E Reid, Michael Brauer, Fay H Johnston, Michael Jerrett, John R Balmes, and Catherine T Elliott. Critical review of health impacts of wildfire smoke exposure. *Environmental health perspectives*, 124(9):1334–1343, 2016a.

Colleen E Reid, Michael Jerrett, Ira B Tager, Maya L Petersen, Jennifer K Mann, and John R Balmes. Differential respiratory health effects from the 2008 northern California wildfires: A spatiotemporal approach. *Environmental research*, 150:227–235, 2016b.

Patrik Reizinger, Luigi Gresele, Jack Brady, Julius Von Kügelgen, Dominik Zietlow, Bernhard Schölkopf, Georg Martius, Wieland Brendel, and Michel

Besserve. Embrace the gap: Vaes perform independent mechanism analysis. *Advances in Neural Information Processing Systems*, 2022.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, 2015.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning*, 2014.

David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017.

James Robins and Miguel Hernan. Estimation of the causal effects of time-varying exposures. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*, pages 553–599, 2008.

James M Robins, Steven D Mark, and Whitney K Newey. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, pages 479–495, 1992.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology, 2000.

James M Robins, Lingling Li, Rajarshi Mukherjee, Eric Tchetgen Tchetgen, and Aad van der Vaart. Minimax estimation of a functional on a structured high-dimensional model. *The Annals of Statistics*, 45(5):1951–1987, 2017.

R Tyrrell Rockafellar. *Conjugate duality and optimization*. SIAM, 1974.

R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015a.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015b.

Paul R Rosenbaum. *Observational Studies*. Springer, 2002.

Paul R Rosenbaum and Donald B Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218, 1983.

Paul R Rosenbaum, P Rosenbaum, and Briskman. *Design of observational studies*, volume 10. Springer, 2010.

Evan TR Rosenman, Guillaume Basse, Art B Owen, and Mike Baiocchi. Combining observational and experimental datasets using shrinkage estimators. *Biometrics*, 79(4):2961–2973, 2023.

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58, 1978.

Donald B Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, pages 1151–1172, 1984.

Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.

Andrzej Ruszczyński and Alexander Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31(3):433–452, 2006.

Pedro Sanchez and Sotirios A. Tsaftaris. Diffusion causal models for counterfactual estimation. In *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177 of *Proceedings of Machine Learning Research*, pages 647–668. PMLR, 2022. URL `https://proceedings.mlr.press/v177/sanchez22a.html`.

Anton Schick. On asymptotically efficient estimation in semiparametric models. *The Annals of Statistics*, pages 1139–1151, 1986.

Johannes Schmidt-Hieber. Nonparametric regression using deep neural networks with relu activation function. *The Annals of Statistics*, 48(4):1875–1897, August 2020. doi: 10.1214/19-AOS1875.

Maresa Schröder, Miruna Oprescu, Stefan Feuerriegel, and Nathan Kallus. Causal inference on networks under misspecified exposure mappings: A partial identification framework. *arXiv preprint arXiv:2602.03459*, 2026.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Erwan Scornet, Gérard Biau, and Jean-Philippe Vert. Consistency of random forests. *The Annals of Statistics*, 43(4):1716–1741, August 2015. doi: 10.1214/15-AOS1321.

Nabeel Seedat, Fergus Imrie, Alexis Bellot, Zhaozhi Qian, and Mihaela van der Schaar. Continuous-time modeling of counterfactual outcomes using neural controlled differential equations. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*. PMLR, 2022.

Vira Semenova and Victor Chernozhukov. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24(2):264–289, 2021.

Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, pages 3076–3085. PMLR, 2017.

Claudia Shi, David Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, 32, 2019.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.

Rahul Singh, Maneesh Sahani, and Arthur Gretton. Kernel instrumental variable regression. *Advances in Neural Information Processing Systems*, 32, 2019.

Michael E Sobel. What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407, 2006.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. *Advances in Neural Information Processing Systems*, 2015.

Chao Song, Yaode Wang, Xiu Yang, Yili Yang, Zhangying Tang, Xiuli Wang, and Jay Pan. Spatial and temporal impacts of socioeconomic and environmental factors on healthcare resources: a county-level bayesian local spatiotemporal regression modeling study of hospital beds in southwest china. *International Journal of Environmental Research and Public Health*, 17(16):5890, 2020.

Charles J Stone. Consistent nonparametric regression. *The annals of statistics*, pages 595–620, 1977.

Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053, 1982.

Fangzhou Su, Wenlong Mou, Peng Ding, and Martin J Wainwright. When is the estimated propensity score better? high-dimensional analysis and bias correction. *arXiv preprint arXiv:2303.17102*, 2023.

Vasilis Syrgkanis, Victor Lei, Miruna Oprescu, Maggie Hei, Keith Battocchi, and

Greg Lewis. Machine learning estimation of heterogeneous treatment effects with instruments. *Advances in Neural Information Processing Systems*, 32, 2019.

Ichiro Takeuchi, Quoc V Le, Timothy D Sears, Alexander J Smola, and Chris Williams. Nonparametric quantile estimation. *Journal of machine learning research*, 7, 2006.

Zhiqiang Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.

Eric J Tchetgen Tchetgen, Isabel R Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network. *Journal of the American Statistical Association*, 116(534):833–844, 2021.

Mauricio Tec, James G Scott, and Corwin M Zigler. Weather2vec: Representation learning for causal inference with non-local confounding in air pollution and climate studies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14504–14513, 2023.

Mauricio Tec, Ana Trisovic, Michelle Audirac, Sophie Woodward, Jie Hu, Naeem Khoshnevis, and Francesca Dominici. SpaCE: The spatial confounding environment. In *International Conference on Representation Learning*, 2024.

Anastasios A Tsiatis. *Semiparametric theory and missing data*, volume 4. Springer, 2006.

John Tsitsiklis and Benjamin Van Roy. Analysis of temporal-diffference learning with function approximation. *Advances in neural information processing systems*, 9, 1996.

Masatoshi Uehara, Masaaki Imaizumi, Nan Jiang, Nathan Kallus, Wen Sun, and Tengyang Xie. Finite sample analysis of minimax offline reinforcement

learning: Completeness, fast rates and first-order efficiency. *arXiv preprint arXiv:2102.02981*, 2021.

U.S. Census Bureau. 2010 census, 2010.

Mark J Van der Laan and James M Robins. *Unified methods for censored longitudinal data and causality*, volume 5. Springer, 2003.

Mark J van der Laan, Sherri Rose, Wenjing Zheng, and Mark J van der Laan. Cross-validated targeted minimum-loss-based estimation. *Targeted learning: causal inference for observational and experimental data*, pages 459–474, 2011.

A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Aad W Van Der Vaart, Jon A Wellner, Aad W van der Vaart, and Jon A Wellner. *Weak convergence*. Springer, 1996.

Tyler J VanderWeele, Eric J Tchetgen Tchetgen, and M Elizabeth Halloran. Interference and sensitivity analysis. *Statistical science: A Review Journal of the Institute of Mathematical Statistics*, 29(4):687, 2015.

Stijn Vansteelandt and Marshall Joffe. Structural nested models and g-estimation: the partially realized promise. *Statistical Science*, 29(4):707–731, 2014.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. In *International conference on learning representations*, 2018.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018a.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018b.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press, 2019.

Jie Wang, Rui Gao, and Hongyuan Zha. Reliable off-policy evaluation for reinforcement learning. *Operations Research*, 72(2):699–716, 2024a.

Kaiwen Wang, Nathan Kallus, and Wen Sun. Near-minimax-optimal risk-sensitive reinforcement learning with cvar. In *International Conference on Machine Learning*, pages 35864–35907. PMLR, 2023a.

Kaiwen Wang, Kevin Zhou, Runzhe Wu, Nathan Kallus, and Wen Sun. The benefits of being distributional: Small-loss bounds for reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2023b.

Kaiwen Wang, Dawen Liang, Nathan Kallus, and Wen Sun. A reductions approach to risk-sensitive reinforcement learning with optimized certainty equivalents. *arXiv preprint arXiv:2403.06323*, 2024b.

Kaiwen Wang, Owen Oertell, Alekh Agarwal, Nathan Kallus, and Wen Sun. More benefits of being distributional: Second-order bounds for reinforcement learning. *International Conference of Machine Learning*, 2024c.

Kaiwen Wang, Nathan Kallus, and Wen Sun. The central role of the loss function in reinforcement learning. *Statistical Science*, 40(4):597–622, 2025.

Linbo Wang and Eric Tchetgen Tchetgen. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(3):531–550, 2018.

Ye Wang. Causal inference under temporal and spatial interference. *arXiv e-prints*, pages arXiv–2106, 2021.

Yixin Wang and David Blei. A proxy variable view of shared confounding. In *International Conference on Machine Learning*, 2021.

Yixin Wang and David M Blei. The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528):1574–1596, 2019.

Yue Wang and Shaofeng Zou. Online robust reinforcement learning with model uncertainty. *Advances in Neural Information Processing Systems*, 34:7193–7206, 2021.

Larry Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.

Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H Kennedy, and Aaditya Ramdas. Time-uniform central limit theory and asymptotic confidence sequences. *The Annals of Statistics*, 52(6):2613–2640, 2024a.

Ian Waudby-Smith, Lili Wu, Aaditya Ramdas, Nikos Karampatziakis, and Paul Mineiro. Anytime-valid off-policy inference for contextual bandits. *ACM/IMS Journal of Data Science*, 1(3):1–42, 2024b.

Wolfram Wiesemann, Daniel Kuhn, and Berç Rustem. Robust markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013.

Wikipedia. Camp Fire (2018) — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Camp%20Fire%20(2018)&oldid=1271689743, 2025. [Online; accessed 29-January-2025].

Jeffrey M Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2nd edition, 2010.

Pengzhou Wu and Kenji Fukumizu. Towards principled causal effect estimation by deep identifiable models. *arXiv preprint arXiv:2109.15062*, 2021. URL https://arxiv.org/abs/2109.15062.

Pengzhou Abel Wu and Kenji Fukumizu. β-intact-VAE: Identifying and estimating causal effects under limited overlap. In *International Conference on Learning Representations*, 2022.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pages 1907–1913, 2019.

Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference. *Advances in Neural Information Processing Systems*, 2021.

Liyuan Xu, Yutian Chen, Siddarth Srinivasan, Nando de Freitas, Arnaud Doucet, and Arthur Gretton. Learning deep features in instrumental variable regression. In *International Conference on Learning Representations*, 2021.

Wenhao Xu, Xuefeng Gao, and Xuedong He. Regret bounds for Markov decision processes with recursive optimized certainty equivalents. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato,

and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 38400–38427. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/xu23d.html.

Steve Yadlowsky, Hongseok Namkoong, Sanjay Basu, John Duchi, and Lu Tian. Bounds on the conditional and average treatment effect with unobserved confounding factors. *The Annals of Statistics*, 50(5):2587–2615, 2022.

Shu Yang and Peng Ding. Combining multiple observational data sources to estimate causal effects. *Journal of the American Statistical Association*, 2019.

Yuhong Yang and Dan Zhu. Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *The Annals of Statistics*, 30(1):100–121, 2002.

Dmitry Yarotsky. Error bounds for approximations with deep relu networks. *Neural Networks*, 94:103–114, 2017.

Mingzhang Yin, Claudia Shi, Yixin Wang, and David M Blei. Conformal sensitivity analysis for individual treatment effects. *Journal of the American Statistical Association*, pages 1–14, 2022.

Keming Yu and MC1614628 Jones. Local linear quantile regression. *Journal of the American statistical Association*, 93(441):228–237, 1998.

Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Kelly Zhang, Lucas Janson, and Susan Murphy. Statistical inference with m-estimators on adaptively collected data. *Advances in neural information processing systems*, 34:7460–7471, 2021.

Wenjia Zhang and Kexin Ning. Spatiotemporal heterogeneities in the causal effects of mobility intervention policies during the covid-19 outbreak: A spatially interrupted time-series (sits) analysis. *Annals of the American Association of Geographers*, 113(5):1112–1134, 2023.

Yao Zhao, Kwang-Sung Jun, Tanner Fiez, and Lalit Jain. Adaptive experimentation when you can't experiment. *Advances in neural information processing systems*, 37:121928–121991, 2024.

Lingxiao Zhou, Kosuke Imai, Jason Lyall, and Georgia Papadogeorgou. Estimating heterogeneous treatment effects for spatio-temporal causal inference: How economic assistance moderates the effects of airstrikes on insurgent violence. *arXiv preprint arXiv:2412.15128*, 2024.

Tianhui Zhou, William E Carson IV, and David Carlson. Estimating potential outcome distributions with collaborating causal networks. *Transactions on machine learning research*, 2022:https–openreview, 2022.

Corwin Zigler, Vera Liu, Fabrizia Mealli, and Laura Forastiere. Bipartite interference and air pollution transport: estimating health effects of power plant interventions. *Biostatistics*, 26(1):kxae051, 2025.