Machine Learning Estimation of Heterogeneous Treatment Effects with Instruments

Vasilis Syrgkanis[1] Victor Lei[2] Miruna Oprescu[1] Maggie Hei[1] Keith Battocchi[1] Greg Lewis[1]

[1]Microsoft Research, New England (ALICE) [2]TripAdvisor

github.com/microsoft/EconML/tree/master/prototypes/dml_iv

Summary

- We estimate heterogeneous causal effects in the presence of unobserved confounders with the aid of a valid instrument
- We build robust methods that leverage arbitrary machine learning models to account for heterogeneous treatment effects and compliance
- We partnered with TripAdvisor to apply these methods to an A/B test with an intent-to-treat structure
- This work resolves an open question in the literature[1]

TripAdvisor's Questions

- What is the *causal effect* of becoming a member on TripAdvisor on downstream activity on the webpage?
- How does that effect vary with observable characteristics of the user?
- Useful for understanding the quality of membership offering, improvements, and targeting

Standard approach: Let's run an A/B test
Problem: We cannot enforce the treatment

Membership is an action that requires user engagement!

Formal Model

- T → treatment, *e.g. user membership*
- Y → outcome, *e.g. webpage activity*
- X → features that capture heterogeneity, *e.g. user features*
- Z → instrumental variable: a variable that affects the treatment T but does not affect the outcome Y other than through the treatment *e.g. assignment in A/B test*

Structural equations:

$$Y = \theta(X)\cdot T + f_0(X) + e \qquad (1)$$

$$T = g_0(X,Z) + \eta \qquad (2)$$

where $\mathbb{E}[e \mid X,Z] = 0$.

Limitations of Typical IV Methods

- Cannot estimate a complex $\theta(X)$
- Do not account for compliance heterogeneity
- Do not leverage Machine Learning estimation techniques

Example: Let $\tilde Y = Y - \mathbb{E}[Y|X], \tilde T = T - \mathbb{E}[T|X], \tilde Z = Z - \mathbb{E}[Z|X], \beta_0(X) = \mathbb{E}[\tilde T \tilde Z|X]$

(2) proposes the following robust estimate of the average treatment effect $\hat\theta$:

$$\mathbb{E}[\theta(X)] = \hat\theta = \frac{\mathbb{E}[\tilde Y \tilde Z]}{\mathbb{E}[\tilde T \tilde Z]} \text{ (ATE)}$$

Since $\hat\theta\,\mathbb{E}[\tilde T \tilde Z] = \mathbb{E}[\theta(X)\tilde T \tilde Z]$ in the limit, then:

$$\hat\theta = \frac{\mathbb{E}[\theta(X)\beta_0(X)]}{\mathbb{E}[\beta_0(X)]}$$

which is **consistent only if $\theta(X)$ is constant** (no heterogeneity) **or $\theta(X)$ and $\beta_0(X)$ are independent** (uniform compliance)!

DRIV Algorithm

Doubly Robust Instrumental Variable (DRIV) Treatment Effect Estimation

1. Consider the **compliance score**[3]:

$$\Delta(X) = (2Z-1)\frac{\mathbb{P}(T=1|Z=1,X) - \mathbb{P}(T=1|Z=0,X)}{2} \qquad (3)$$

Define the following residuals:

$$\tilde Y = Y - \mathbb{E}[Y|X], \quad \tilde T = T - \mathbb{E}[T|X], \quad \tilde Z = Z - \mathbb{E}[Z|X] \qquad (4)$$

2. Estimate **preliminary $\hat\theta(X)$**:

$$\hat\theta = \underset{\theta(\cdot)}{\operatorname{argmin}}\, \mathbb{E}\left[\left(\tilde Y - \theta(X)\cdot\Delta(X)\right)^2\right] \qquad (5)$$

3. Estimate **robust final $\theta(X)$**:

$$\min_{\theta(\cdot)} \mathbb{E}\left[\left(\hat\theta(X) + \frac{\tilde Y - \hat\theta(X)\cdot\tilde T}{\Delta(X)} - \theta(X)\right)^2\right] \qquad (6)$$

* More generally, when T and Z are arbitrary and as in (1)-(2), $\theta(X)$ is given by:

$$\min_{\theta(\cdot)} \mathbb{E}\left[\left(\hat\theta(X) + \frac{(\tilde Y - \hat\theta(X)\cdot\tilde T)\tilde Z}{\hat\beta(X)} - \theta(X)\right)^2\right]$$

where $\beta_0(X) = \mathbb{E}[\tilde T \tilde Z|X]$ and $\tilde Z/\,\hat\beta(X)$ is known as the compliance score[3].

Reduction to a Set of Regression/ Classification Steps

In the above algorithm, the steps marked with ★,★ are classifications and regression tasks, respectively. Benefits of this approach:

- **Statistical and computational benefits** of modern ML approaches (forests, regularized linear models, SVM, DNNs etc.)
- **Cross-validation** for model selection and hyperparameter tuning
- **Interpretability** of estimated models (SHAP, Lime, Influence functions)

Properties of the DRIV method

- Loss function for final estimate satisfies **Neyman orthogonality**[4]
- Mean-Squared-Error of final $\theta(X)$ **robust to errors in auxiliary Classifications/ Regressions**
- Approach extends **beyond recommendation A/B tests, to linear-in-treatment IV setting**
- When (6) supports CI construction, Neyman orthogonality typically **preserves the validity of the intervals**

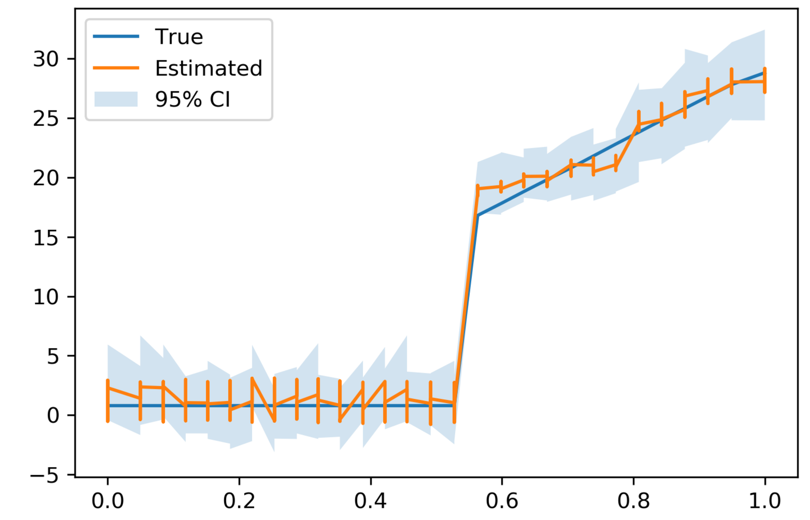


Fig. 1: Example of DRIV estimates and confidence intervals for synthetic data with a step treatment effect.
The *x* and *y* axes represent the feature of heterogeneity and the treatment effect, respectively.
The shaded region depicts the 95% confidence intervals which attain ~95% coverage in Monte Carlo experiments.

TripAdvisor's Findings

Ran a 4M user A/B test with half receiving a new, easier sign-up process.

- Easier sign-up process incentivizes membership
- Outcome was the number of visits in the next 14 days

Table 1: ATE Estimates (Normalized)

Nuisance	Method	ATE Est [95% CI]	Nuisance	Method	ATE Est [95% CI]
LM	DMLATEIV	0.117 [-0.051, 0.285]	GB	DMLATEIV	0.127 [-0.031, 0.285]
LM	DRIV	0.113 [-0.052, 0.279]	GB	DRIV	0.125 [-0.061, 0.311]

High Level Take-Aways

- Large heterogeneity based on which pages were recently visited
- Large heterogeneity based on platform of access (e.g. iPhone, Linux etc.)
- Results enable better targeting of the right user populations and improvements of membership offering for user segments with small effects

Fig 2: Random Forest Heterogeneity with Linear Model Residualization

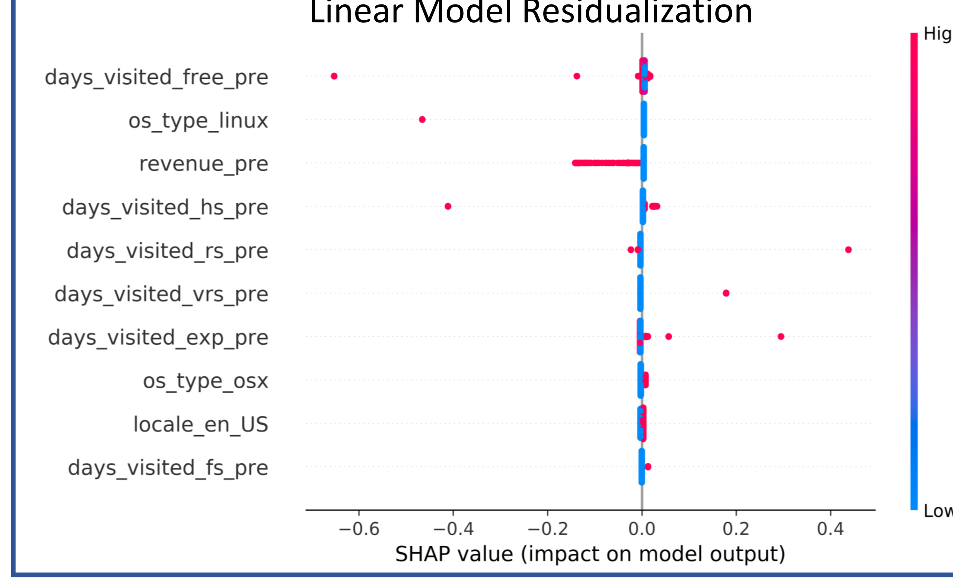
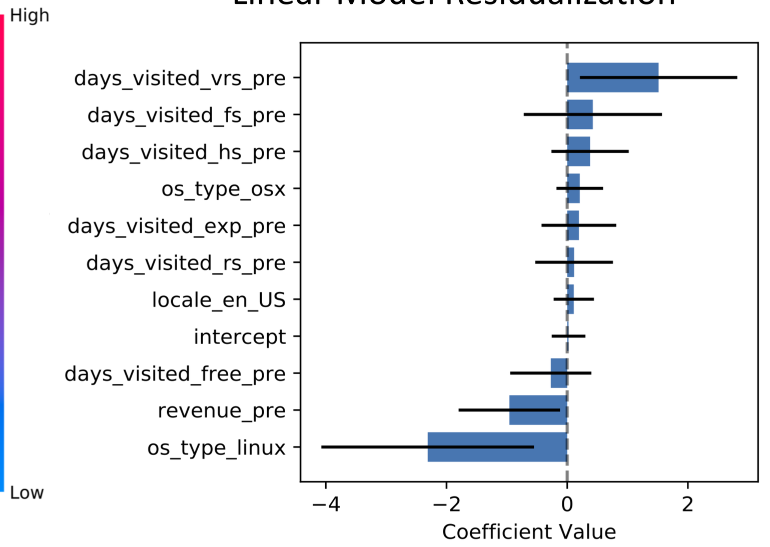


Fig 3: Linear Model Heterogeneity with Linear Model Residualization



Try It Out!

- Code: github.com/microsoft/EconML/tree/master/prototypes/dml_iv

```
# Define the DRIV algorithm and nuisance functions
dr_cate = IntentToTreatDRIV(model_y_x = RandomForestRegressor(),
                            model_t_xz = RandomForestClassifier(),
                            prel_model_effect = RandomForestRegressor(),
                            final_model_effect = Linear Regression())
# Fit estimator and calculate treatment effects
dr_cate.fit(Y, T, X, Z)
te_pred = dr_cate.effect(X_test)
```

Fig. 4: Snapshot of the Python code for applying the DRIV method

- EconML python library for ML Estimation of Heterogeneous Treatment Effects
  - github.com/microsoft/EconML
  - pip install econml

Selected References

1. Nie, X., & Wager, S. *Quasi-oracle estimation of heterogeneous treatment effects*. arXiv preprint, 2017.
2. Chernozhukov, V., Chetverikov, D., et al. *Double/debiased machine learning for treatment and structural parameters*. The Econometrics Journal, 2018.
3. Abadie, A. *Semiparametric instrumental variable estimation of treatment response models*. Journal of Econometrics, 2003.
4. Chernozhukov, V., Nekipelov, D., et al. *Plug-in regularized estimation of high-dimensional parameters in nonlinear semiparametric models*. arXiv preprint, 2018.