Causal Inference through the Method of Direct Estimation

Marc Ratkovic and Dustin Tingley Princeton University and Harvard University

Summary

Causal inference ⇔ ceteris paribus manipulation of treatment

Motivating Concerns

- Sensitivity to model misspecification
- Fail with high-dimensional covariates
- Limited to binary treatment, average effect
- Bespoke extensions

Method of Direct Estimation for counterfactual quantities

- Adaptive, nonparametrc regression
- Robust to monotonic transform of confounders
- Binary, categorical, continuous treatment
- IV, mediation arise naturally

Estimation Procedure

1. Construct tensor spline basis (*treatment* × *covariate* × *covariate*)
2. Marginal correlation screen (SIS, Fan and Lv)
3. Sparse Bayesian regression with endogenous tuning parameter (Ratkovic and Tingley 2017)
4. Predict desired counterfactual

Explaining Estimation Procedure

1. Model treatment heterogeneity as a nonparametric function of confounders
2. Millions of bases: screen to get to hundreds or thousands
3. Sparse model with endogenous tuning to achieve theoretical prediction bounds
4. Predict desired counterfactuals (treatment effect, IV, mediation)

Advantages

- Automated: no researcher-specified tuning parameters
- Binary, categorical, continuous treatment
- Robust to skewed covariates and interaction terms
- Stable algorithm
- Achieves minmax uncertainty bound
- Plug-in estimate for SEMs

Direct Extensions

- IV: Model heterogeneous first stage and second stage effects
- Mediation: Model heterogeneous direct and indirect effects

The Estimand

The Setup

- $Y_i(t)$: Potential outcome function; $t \in \{0,1\}$ or $t \in \Re$
- Confounders: X_i
- Conditional ignorability; SUTVA
- $[Y_i, T_i, X_i^\top]$: Observed

Estimands

- Binary effect: $E(Y_i(1) - Y_i(0)|X_i)$
- Dose effect: $\nabla_{T_i}(\delta) = \frac{1}{\delta}E(Y_i(T_i+\delta) - Y_i(T_i)|X_i)$
- Instantaneous effect: $\lim_{\delta\to 0}\nabla_{T_i}(\delta)$

The Observed Data Model

Outcome is fit to post-screened bases $R(T_i, X_i)$,

$$Y_i = R(T_i, X_i)^\top c + \epsilon_i; \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

Estimating effects

- $\widehat{E}(Y_i(1) - Y_i(0)|X_i) = (R(1, X_i) - R(0, X_i))^\top \widehat{c}$
- $\widehat{\nabla}_{T_i}(\delta) = \frac{1}{\delta}(R(T_i+\delta, X_i) - R(T_i, X_i))^\top \widehat{c}$
- Instantaneous effect: set $\delta = 10^{-5}$

Constructing Bases

- B-splines: 24 for treatment and each confounder
- Three-way interactions: $p = 10 \Rightarrow$ 1.45 million bases
- Screen: Maintain $100 \times (1 + n^{1/5}) \approx$ hundreds

The Sparse Bayesian Model

$$\begin{aligned}
Y_i|R(T_i,X_i), c, \sigma^2 &\sim \mathcal{N}(R(T_i,X_i)^\top c, \sigma^2)\\
c_k|\lambda, w_k, \sigma &\sim DE(\lambda w_k/\sigma)\\
\lambda^2|n, p, \rho &\sim \Gamma\left(n \times (\{log(n) + 2\log(p)) - p, \rho\right)\\
w_k|\gamma &\sim \text{generalizedGamma}(1, 1, \gamma)\\
\gamma &\sim \exp(1)
\end{aligned}$$

Properties

- Likelihood: Normal regression
- Tuning parameter $\widehat{\lambda}$ grows at Oracle Rate $\left(\sqrt{n\log(p)}\right)$
- Weights for $\widehat{c_k} \to 0$ grow as $\sqrt{\log(n)/n}$
- Weights for $\widehat{c_k} \nrightarrow 0$ grow as $1/n$
- $\widehat{\gamma}$: Global sparsity parameter; controls tail density of prior

Robustness 1: Model Specification

NSW experiment

- Treatment: Job training program
- Outcome: 1978 earnings; \$886=Experimental benchmark
- X: 1974/1975 income, no earnings in 1974/1975, age, race, . . .
- X reduced: X - no earnings in 1974/1975; no hs degree

	MDE	OLS Post LASSO
X	859	1423
X reduced	888	46
X, X^2	513	2306
X, X^2, interactions	338	8.22×10^{15}

Robustness 2: Algorithmic Stability

5 fits to same set of nonparametric bases (Truth = 0)

MDE	187	187	188	190	189
Horseshoe	1600	1873	997	1456	762
CV-LASSO	1719	1692	1755	1719	1755

Extension: IV Analysis

Instantaneous encouragement

$$\nabla_{Z_i}^{IV}(\delta) = \mathbb{E}(T_i(Z_i+\delta) - T_i(Z_i)|X_i)$$

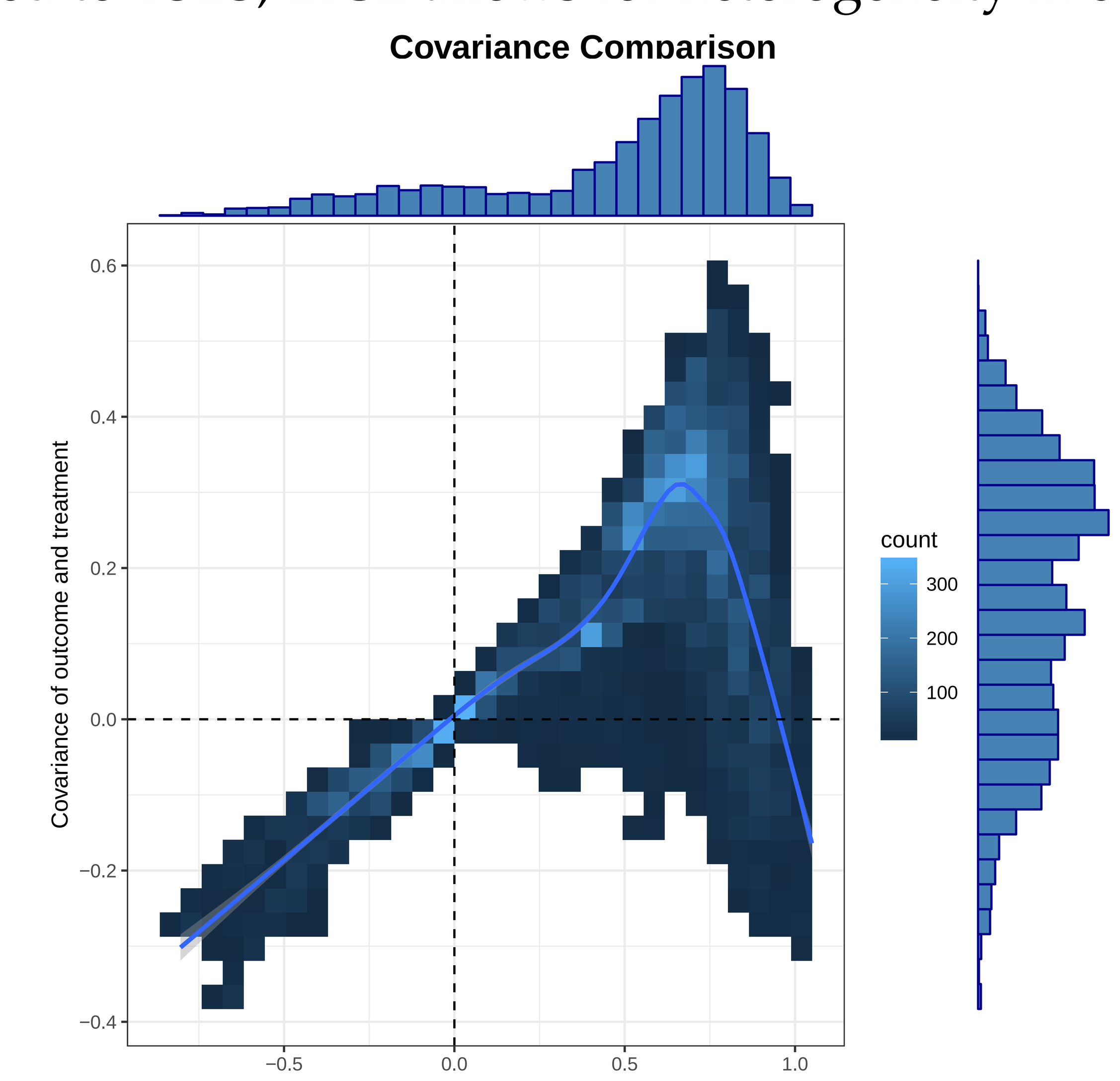
Instantaneous intent to treat on the treated

$$\nabla_{T_i}^{IV}(\delta) = \mathbb{E}\left\{Y_i\left(T_i(Z_i) + \nabla_{Z_i}^{IV}(\delta)\right) - Y_i\left(T_i(Z_i)\right)|X_i\right\}$$

Local **I**nstantaneous **C**ausal **E**ffect

$$\lim_{\delta\to 0} \nabla_{T_i}^{IV}(\delta)/\nabla_{Z_i}^{IV}(\delta)$$

Compared to TSLS, LICE allows for heterogeneity in each stage



Conclusion

- Automated, reliable causal effect estimation
- Structural equations (IV, Mediation)
- Future work: Repeated observations, limited d.v.'s