Efficient Difference-in-Differences Estimation with Panel Data

Deng, Yuhao

University of Michigan

June 27, 2025

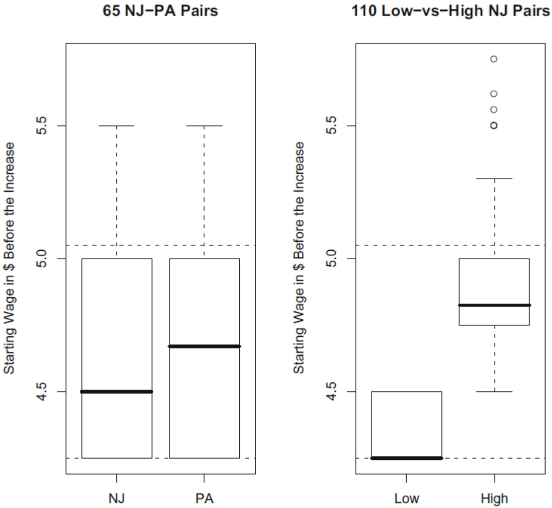
Outline

Minimum Wages and Employment

- *"The higher the minimum wage, the greater will be the number of covered workers who are discharged."* — George Stigler
- David Card and Alan Krueger's study
- New Jersey increased its state minimum wage from $4.25 to $5.05 per hour on April 1st, 1992
- Did the increase in the minimum wage in New Jersey reduce employment at fast-food restaurants?
- Treatment groups: (1) NJ vs PA, (2) low vs high in NJ

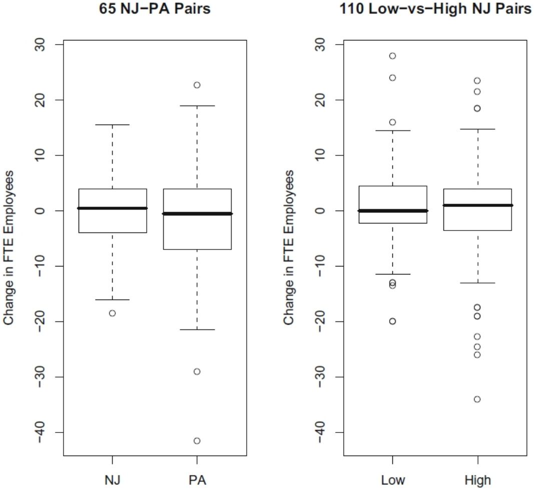
Minimum Wages and Employment

- Pre-treatment



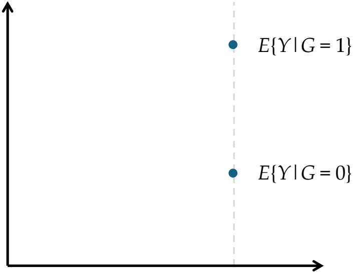
Minimum Wages and Employment

- Post-treatment



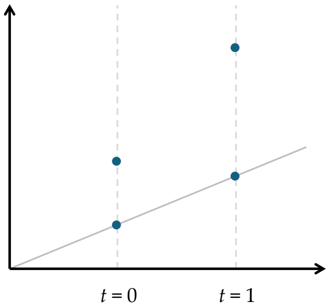
Difference-in-Differences

- Post-treatment period



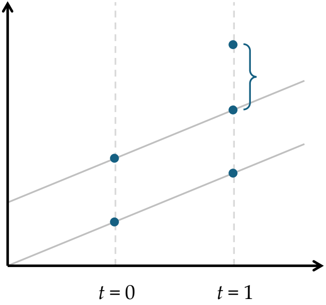
Difference-in-Differences

- Pre-treatment period



Difference-in-Differences

- A parallel trend



Formalization

- Group indicator $G \in \{0, 1\}$
- Period indicator $t \in \{0, 1\}$
- Potential outcome $Y_t(g)$, $g = 0, 1$, $t = 0, 1$
- Treatment indicator $D_t = Gt$
- Baseline covariates X

- Observed data $O = (X, G, Y_0, Y_1)$

Causal Estimand

- Average treatment effect on the treated (ATT)

$$\tau = E\{Y_1(1) - Y_1(0) \mid G = 1\}$$

- No anticipation: $Y_0(0) = Y_0(1)$
- Parallel trend:
  $E\{Y_1(0) - Y_0(0) \mid X, G = 1\} = E\{Y_1(0) - Y_0(0) \mid X, G = 0\}$
- Positivity: $P(G = 1) > c$, $P(G = 0 \mid X) > c$
- Consistency: $Y_t(G) = Y_t$

Structural Causal Model

- Unmeasured confounder U,

$$Y_t(g) = f(X, t, g) + U + \epsilon_t$$

- Difference in counterfactual outcomes under control between periods

$$Y_1(0) - Y_0(0) = f(X, 1, 0) - f(X, 0, 0) + \epsilon_1 - \epsilon_0$$

- Identical regardless of treatment assignment

Models

- Propensity score
$$\pi_g(x) = P(G = g \mid X = x)$$
- Outcome model
$$\mu_{g,t}(x) = E\{Y_t \mid G = g, X = x\}$$
- Increment
$$\delta_g(x) = E\{Y_1 - Y_0 \mid G = g, X = x\}$$

Identification

- ATT is identified by difference in differences,

$$
\begin{aligned}
\tau &= E\{Y_1(1) - Y_1(0) \mid G = 1\} \\
&= E(Y_1 - Y_0 \mid G = 1) - E\{E(Y_1 - Y_0 \mid X, G = 0) \mid G = 1\}
\end{aligned}
$$

- Outcome regression or weighting

$$
\begin{aligned}
\tau &= \frac{1}{P(G = 1)} \mathbb{P}\left[G\{\delta_1(X) - \delta_0(X)\}\right] \\
&= \frac{1}{P(G = 1)} \mathbb{P}\left[\left\{G - (1 - G)\frac{\pi_1(X)}{\pi_0(X)}\right\}(Y_1 - Y_0)\right]
\end{aligned}
$$

- Estimation efficiency?

Two-way Fixed Effects Model

- The simplest estimator by linear regression:

$$Y_t = \mu + \lambda G + \gamma t + \alpha D_t + \beta^\top X + u_t$$

- α is interpreted as ATT because

$$E(Y_1 - Y_0 \mid X, G) = \gamma + \alpha G$$

- Problems: model specification, efficiency

Regular and Asymptotically Linear Estimators

- We say $\hat{\theta}$ is a regular and asymptotic linear (RAL) estimator for θ, and φ is the influence function if

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(O_i) + o_p(1)$$

- There exists a unique influence function φ^{eff} such that for any φ,

$$\text{var}(\varphi) \geq \text{var}(\varphi^{eff})$$

- We call φ^{eff} the efficient influence function (EIF)

Efficient Influence Function

- EIF for τ:

$$\varphi^{eff} = \frac{1}{P(G=1)} \left\{ G - (1-G)\frac{\pi_1(X)}{\pi_0(X)} \right\} \{ Y_1 - Y_0 - \delta_0(X) - G\tau \}$$

- By solving the estimating equation $\mathbb{P}_n \varphi^{eff} = 0$, we obtain an estimator

$$\hat{\tau} = \frac{1}{\mathbb{P}_n(G)} \mathbb{P}_n \left\{ G - (1-G)\frac{\hat{\pi}_1(X)}{\hat{\pi}_0(X)} \right\} \{ Y_1 - Y_0 - \hat{\delta}_0(X) \}$$

- Asymptotic normality (under regularity conditions)

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, \mathrm{var}(\varphi^{eff}))$$

Asymptotic Properties

- Semiparametric efficiency: The asymptotic variance of $\hat{\tau}$ attains the semiparametric efficiency bound when all models are correctly specified
- Double robustness: The estimator $\hat{\tau}$ is consistent if either $\pi_g(x)$ or $\delta_0(x)$ is correctly specified
- Limitation: Unstable finite-sample performance

Targeted Minimum Loss Based Estimation

- Recall the EIF

$$\varphi^{eff} = \frac{1}{P(G=1)}\left\{G-(1-G)\frac{\pi_1(X)}{\pi_0(X)}\right\}\{Y_1 - Y_0 - \delta_0(X) - G\tau\}$$

- Targeted estimator as a substitution estimator

$$\tilde{\tau} = \frac{1}{\mathbb{P}_n(G)}\mathbb{P}_n[G\{\tilde{\delta}_1(X) - \tilde{\delta}_0(X)\}]$$

- To solve the EIF,

$$\mathbb{P}_n\left\{G-(1-G)\frac{\hat{\pi}_1(X)}{\hat{\pi}_0(X)}\right\}\{Y_1 - Y_0 - \tilde{\delta}_G(X)\} = 0$$

Targeted Minimum Loss Based Estimation

- Suppose we use OLS to model $\mu_{g,t}(x)$, we just need to add a "clever" covariate

$$\hat{H}_t(G, X) = (2t-1)\left\{G - \frac{\hat{\pi}_1(X)}{\hat{\pi}_0(X)}(1-G)\right\}$$

  in the model

$$Y_t = \mu_{G,t}(X) + \nu \hat{H}_t(G, X) + u_t$$

- The score function associated with ν solves

$$\mathbb{P}_n\left\{G - (1-G)\frac{\hat{\pi}_1(X)}{\hat{\pi}_0(X)}\right\}\{Y_1 - Y_0 - \tilde{\delta}_G(X)\} = 0$$

Link to Linear Models

- Consider the linear model

$$Y_{ti} = \mu + \lambda G_i + \gamma t + \alpha D_{ti} + \beta^\top X_i + \eta_1^\top G_i X_i + \eta_2^\top X_i t \\ + \eta_3^\top D_{ti} X_i + \nu \hat{H}_t(G_i, X_i) + u_{ti}$$

- The TMLE estimator is

$$\tilde{\tau} = \hat{\alpha} + \hat{\eta}_3^\top \sum_{i:G_i=1} \frac{X_i}{N_1} + \hat{\nu} \sum_{i:G_i=1} \frac{2/N_1}{\hat{\pi}_0(X_i)}$$

Asymptotic Properties

- The TMLE estimator has the same asymptotic properties as the estimating equation-based estimator
- Semiparametric efficiency
- Double robustness
- Probably better finite-sample performance

Simulation

- Data generated from a saturated model
- Methods: two-way fixed effects model (TWFE), saturated regression model (Satur), estimating equation based (DR), and TMLE

	TWFE	Satur	DR	TMLE
Saturated model, $n = 500$				
Bias	-0.235	-0.002	0.004	-0.002
SD	0.092	0.083	0.088	0.087
SE	0.086	0.072	0.087	0.083
CP	0.234	0.906	0.946	0.938
Saturated model, $n = 2000$				
Bias	-0.232	0.002	0.005	0.002
SD	0.046	0.041	0.044	0.042
SE	0.043	0.036	0.043	0.042
CP	0.001	0.914	0.943	0.945

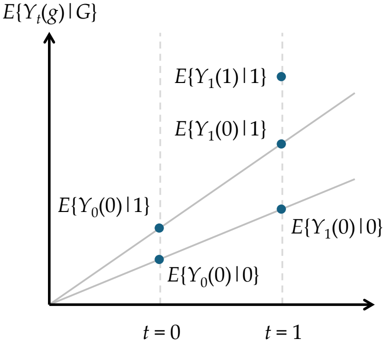
Simulation

- Skewed data; outcome regression model misspecified
- Methods: two-way fixed effects model (TWFE), saturated regression model (Satur), estimating equation based (DR), and TMLE

	TWFE	Satur	DR	TMLE
Misspecified model, $n = 500$				
Bias	-1.384	0.190	0.048	-0.001
SD	0.484	0.355	0.423	0.358
SE	0.435	0.357	0.413	0.352
CP	0.153	0.924	0.946	0.945
Misspecified model, $n = 2000$				
Bias	-1.412	0.162	0.012	-0.010
SD	0.240	0.180	0.210	0.178
SE	0.217	0.179	0.208	0.176
CP	0.000	0.855	0.949	0.944

Parallel Trend Assumption Revisited

- The parallel trend assumption may not hold for non-Gaussian outcomes
- Count data: rate difference
- Binary data: odds ratio



Transformed Parallel Trend

- Let $\mu^{d}_{g,t}(x) = E\{Y_t(d) \mid G = g, X = x\}$
- For a known transformation (link) function $h(\cdot)$,

$$h(\mu^{(0)}_{1,1}(X)) - h(\mu^{(0)}_{1,0}(X)) = h(\mu^{(0)}_{0,1}(X)) - h(\mu^{(0)}_{0,0}(X))$$

- $h(u) = u$: difference of means
- $h(u) = \log(u)$: ratio of means
- $h(u) = \log(u/(1-u))$: odds ratio for binary outcomes

Causal Estimand

- Conditional treatment effect

$$\tau(x) = h(\mu_{1,1}^{(1)}(x)) - h(\mu_{1,1}^{(0)}(x))$$

- Average treatment effect on the treated (ATT)

$$\tau = E\{h(\mu_{1,1}^{(1)}(X)) - h(\mu_{1,1}^{(0)}(X)) \mid G = 1\}$$

- $h(u) = u$: average difference in means
- $h(u) = \log(u)$: average ratio of means
- $h(u) = \log(u/(1-u))$: average odds ratio for binary outcomes

Identification

- Identification is achieved in a similar manner to conventional difference-in-differences
- A naive estimator based on regression

$$\hat{\tau} = \frac{1}{\mathbb{P}_n(G)}\mathbb{P}_n[G\{h(\hat{\mu}_{1,1}(X)) - h(\hat{\mu}_{1,0}(X)) - h(\hat{\mu}_{0,1}(X)) + h(\hat{\mu}_{0,0}(X))\}]$$

- How to improve efficiency and make inference?

Efficient Influence Function

- The EIF for τ is

$$
\begin{aligned}
\varphi^{eff} = & \frac{G}{P(G=1)} \sum_{t=0}^{1} (2t-1) \left\{ h'(\mu_{1,t}(X)) \{ Y_t - \mu_{1,t}(X) \} \right\} \\
& - \frac{1-G}{P(G=1)} \frac{\pi_1(X)}{\pi_0(X)} \sum_{t=0}^{1} (2t-1) \left\{ h'(\mu_{0,t}(X)) \{ Y_t - \mu_{0,t}(X) \} \right\} \\
& + \frac{G}{P(G=1)} \{ \tau(X) - \tau \}
\end{aligned}
$$

Efficient Estimation

- By solving the estimating equation $\mathbb{P}_n(\varphi^{eff}) = 0$, we obtain

$$\tilde{\tau} = \hat{\tau} + \frac{1}{\mathbb{P}_n(G)}\mathbb{P}_n\left[G\sum_{t=0}^{1}(2t-1)h'(\hat{\mu}_{1,t}(X))\{Y_t - \hat{\mu}_{1,t}(X)\}\right]$$
$$- \frac{1}{\mathbb{P}_n(G)}\mathbb{P}_n\left[(1-G)\frac{\hat{\pi}_1(X)}{\hat{\pi}_0(X)}\sum_{t=0}^{1}(2t-1)h'(\hat{\mu}_{0,t}(X))\{Y_t - \hat{\mu}_{0,t}(X)\}\right]$$

- Semiparametric efficiency (under regularity conditions)

$$\sqrt{n}(\tilde{\tau} - \tau) \xrightarrow{d} N(0, \mathrm{var}(\varphi^{eff}))$$

- No double robustness
- No simple form of TMLE

Estimation and Inference

- Fit the propensity score and the outcome regression model
- Calculate the naive regression estimator $\hat{\tau}$ and the semiparametric estimator $\tilde{\tau}$
- Plug the estimates into the EIF $\hat{\varphi}^{eff}$ and estimate the variance of $\tilde{\tau}$ by $\mathbb{P}_n\{\hat{\varphi}^{eff}\}^2/n$.

Family	Data support	Link	Interpretation
Gaussian	$(-\infty, +\infty)$	u	Average difference
Gaussian	$(0, +\infty)$	$\log(u)$	Average log ratio
Binomial	$\{0, 1\}$	$\log(u)$	Average log risk ratio
Binomial	$\{0, 1\}$	$\log(u/(1-u))$	Average log odds ratio
Quasibinomial	$(0, 1)$	$\log(u/(1-u))$	Average log odds ratio
Poisson	$\{0, 1, 2, \ldots\}$	$\log(u)$	Average log rate ratio
QuasiPoisson	$\{0, 1, 2, \ldots\}$	$\log(u)$	Average log rate ratio

Simulation: Binary Data

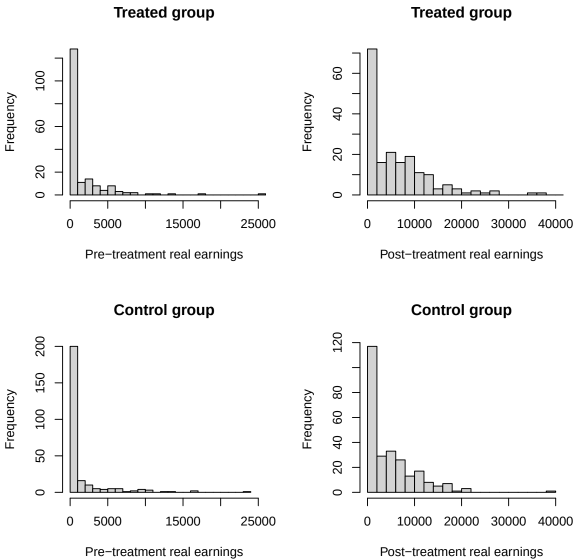
- Setting 1: correctly specified models
- Setting 2: outcome regression model misspecified (not consistent)

		Setting 1			Setting 2		
Size	Method	ΔG	Reg	Eff	ΔG	Reg	Eff
500	Bias	-0.067	-0.010	-0.010	-0.013	-0.154	-0.056
	SD	0.276	0.286	0.288	0.348	0.328	0.344
	SE			0.286			0.325
	CP			0.949			0.926
2000	Bias	-0.058	0.005	0.006	-0.053	-0.197	-0.103
	SD	0.136	0.139	0.139	0.172	0.164	0.170
	SE			0.142			0.160
	CP			0.956			0.890

Application to NSW Data

- The National Supported Work Demonstration (NSW) job training program
- 445 individuals with six baseline covariates (age, years of education, race, ethnicity, marital status, and possession of a degree)
- Treatment: guaranteed a job for 9–18 months (41%)
- Pre-treatment outcome: earnings in 1975
- Post-treatment outcome: earnings in 1978

Application to NSW Data



Application to NSW Data

- The data distribution is severely skewed (many zeros)
- Based on the estimate by TMLE, the job training program significantly increases real earnings

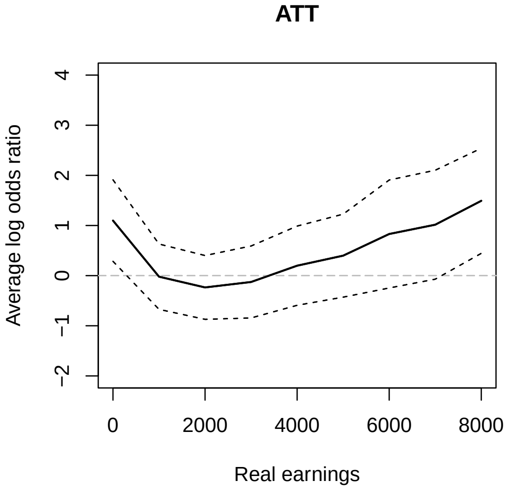
Method	Est	(SE)	P
TWFE	1529.2	(695.1)	0.028
Satur	1561.6	(714.6)	0.029
DR1	1562.6	(717.8)	0.029
DR2	1524.9	(725.9)	0.036
TMLE	1606.1	(728.0)	0.027

DR1 and DR2 use different outcome regression models.

Application to NSW Data

- We consider a binary outcome defined as $\tilde{Y}_t = I(Y_t > y)$
- Significant effect on increasing the employment (average log odds ratio 1.10, s.e. 0.42, $P = 0.008$)
- Significant effect on increasing the probability of having earnings greater than 8000 (average log odds ratio 1.49, s.e. 0.53, $P = 0.005$)

Application to NSW Data



Staggered Difference-in-Differences

- Multiple periods $t \in \{0, 1, \dots, T\}$
- Multiple groups $G \in \{1, \dots, T, \infty\}$
- Potential outcome $Y_t(g)$
- Group-time ATT

$$\tau_{g,t} = E\{Y_t(g) - Y_t(\infty) \mid G = g\}$$

- Aggregated ATT

$$\tau = \sum_{g,t} w_{g,t} \tau_{g,t}$$

Two-Way Fixed Effects Model

- Identification assumptions: parallel trend, no anticipation, positivity, consistency
- Linear model

$$Y_t = \lambda_t + \gamma_G + \alpha D_t + \beta^\top X + u_t$$

- Challenges in interpretation of α
- Negative weights

Aggregated ATT

- Define the ATT as

$$\tau = \frac{1}{\sum_{g=1}^{T} \sum_{t=g}^{T} P(G=g)} \sum_{g=1}^{T} \sum_{t=g}^{T} P(G=g)\tau_{g,t}$$

- Weighted by the probability of being treated

Why Not Efficient

- Identification based on the never-treated group

$$\tau_{g,t} = E(Y_t - Y_{g-1} \mid G = g) - E\{E(Y_t - Y_{g-1} \mid X, G = \infty) \mid G = g\}$$

- Identification based on the not-yet-treated group

$$\tau_{g,t} = E(Y_t - Y_{g-1} \mid G = g) - E\{E(Y_t - Y_{g-1} \mid X, G > t) \mid G = g\}$$

- It did not use all the information of untreated units

Doubly Robust AIPW Estimation

- A new identification formula:

$$\tau_{g,t} = E(Y_t - Y_{g-1} \mid G = g) - \sum_{k=g}^{t} E\{E(Y_k - Y_{k-1} \mid X, G > k) \mid G = g\}$$

- Estimation: augmented inverse probability weighting for $\tau_{g,t}$ and τ
- Double robustness; asymptotic normality
- Byproduct: ATT across groups τ_g, ATT across periods τ_t, ATT over length of exposure τ_{t-g}

Efficient Estimation

- Deriving the EIF needs considering the data generation mechanism
- Nonparametric structural causal model $\Delta Y_t = f(t, G, H_t, \epsilon_t)$
- Assume conditional parallel trend for $\Delta Y_t(\infty)$ given H_t
- Let $\sigma^2_{g,t}(H_t) = \mathrm{var}(\Delta Y_t \mid G = g, H_t)$

$$
\begin{aligned}
\varphi_{g,t} = {} & \frac{I(G=g)}{P(G=g)} \left\{ Y_t - Y_{g-1} - \sum_{k=g}^{t} \delta_k(H_k) - \tau_{g,t} \right\} \\
& - \frac{1}{P(G=g)} \sum_{k=g}^{t} I(G>k) \left[\sum_{l=k}^{T} \frac{\pi_l(H_k)}{\sigma^2_{l,k}(H_k)} \right]^{-1} \\
& \cdot \frac{\pi_g(H_k)}{\sigma^2_{G,k}(H_k)} \left\{ \Delta Y_k - \delta_k(H_k) \right\}
\end{aligned}
$$

- Simpler form under homoskedasticity

Simulation

- Homogeneous treatment effect
- Methods: two-way fixed effects model (TWFE), doubly robust (DR), estimating equation based (EIF), and TMLE

		Scenario 1: Homogeneous				
Size		TWFE	DRnt	DRny	EIF	TMLE
500	Bias	-0.024	0.021	0.012	0.002	0.002
	SD	0.086	0.298	0.231	0.123	0.123
	SE	0.125	0.243	0.200	0.125	0.125
	CP	0.992	0.912	0.928	0.966	0.967
2000	Bias	-0.026	-0.001	0.000	0.002	0.002
	SD	0.041	0.144	0.112	0.060	0.060
	SE	0.063	0.135	0.108	0.063	0.063
	CP	0.991	0.941	0.952	0.959	0.960

Simulation

- Heterogeneous treatment effects
- Methods: two-way fixed effects model (TWFE), doubly robust (DR), estimating equation based (EIF), and TMLE

		Scenario 2: Heterogeneous				
Size		TWFE	DRnt	DRny	EIF	TMLE
500	Bias	-0.474	0.249	0.241	-0.006	-0.006
	SD	0.086	0.298	0.231	0.123	0.123
	SE	0.126	0.243	0.200	0.126	0.126
	CP	0.003	0.709	0.681	0.972	0.969
2000	Bias	-0.468	0.236	0.238	0.004	0.004
	SD	0.042	0.144	0.112	0.060	0.060
	SE	0.063	0.135	0.108	0.063	0.063
	CP	0.000	0.503	0.352	0.960	0.962

Application to NCEE (Gaokao)

- Policy change: from ordered admission to parallel admission
- Data: 27 provinces, stem and non-stem, from 2007 to 2011
- Outcome: standardized justified envy (envy or not, number of envied students, distance of envy, number of unique blocks)

	EIF			TMLE		
Outcome	ATT	SE	*P*	ATT	SE	*P*
envy	-0.106	0.018	0.000	-0.106	0.018	0.000
nenvy	-0.054	0.006	0.000	-0.054	0.006	0.000
denvy_d	-0.036	0.004	0.000	-0.036	0.004	0.000
denvy_u	-0.223	0.036	0.000	-0.225	0.036	0.000

Acknowledgments

- Qinqing Liu, Xiang Peng, Tao Zhang (Soochow University)
- Haoyu Wei (University of California, San Diego)
- Le Kang (Nanjing University)