Synthetic Instrumental Variables

in Diff-in-Diff designs with unmeasured confounding

Ahmet Gulek (MIT) and Jaume Vives-i-Bastida (MIT)

August 28, 2024

Overview

- Motivation
- Empirical example
- Synthetic IV estimator
    - Bias, consistency and inference
- Simulations
- Empirical checks

Motivation

- **IV-DiD Design**: (Y, R, Z) + pre-treatment period
- **Common problem**: worried about DiD due to unmeasured confounding (PT may not hold).
- **Solution**: find an IV
  - Example: Shift-share IV methods compare groups or regions more- and less-exposed to a treatment.
- **But**: What to do if the instrument is also correlated with unobserved confounders? (e.g. reduced-form has pre-trends)
- **This paper**: a *synthetic IV estimator*
  - Combines IV and SCs to address unmeasured confounding
  - Biased, but consistent under assumptions when TSLS is not
  - Bias depends on pre-treatment fit and instrument strength
  - Can be combined with other estimators to become 'doubly-robust'
  - Exhibits good properties in empirical applications and simulations

An Example

- **Question**: What is the impact of immigrants on natives' labor market outcomes?
- Setting: Syrian Refugees in Turkey
- Syrian Civil war started in March 2011
- By 2017, 6 million Syrians had sought shelter outside of Syria
    - Primarily in Turkey (3.5 million)
- Time span studied 2004–2016.
- **Outcome** Y: employment rate of natives (at region-year level)
- **Treatment** R: Refugee/native ratio
- **Instrument** Z: Weighted-distance from the border
- **Confounder** U: region economic trends

Time-series of the number of refugees in Turkey

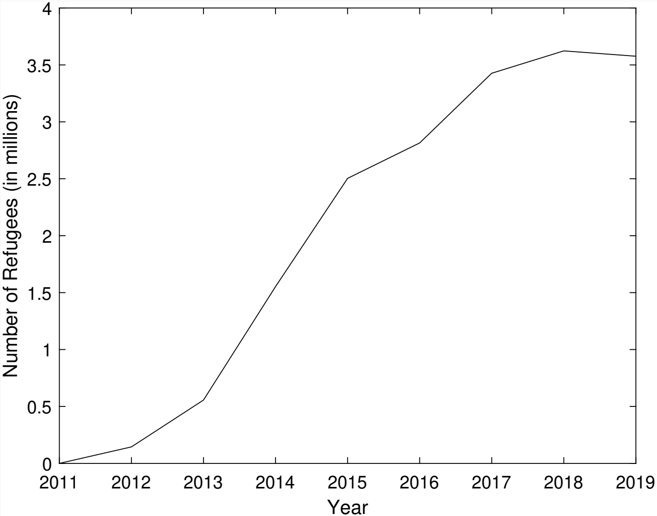


Figure 1: Number of refugees in Turkey

Distribution of Syrian refugees in Turkey

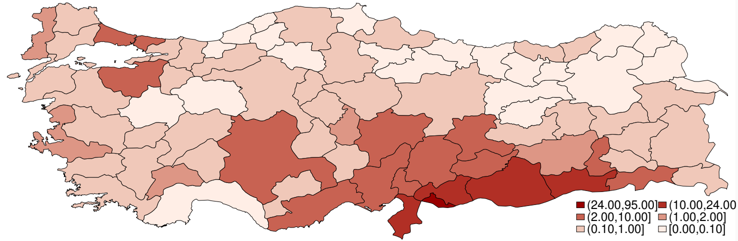


Figure 2: Number of refugees per 100 natives in 2015

Instrument construction

First-stage

$$R_{jt} = \sum_{k \neq 2010} \theta_k(\mathbb{1}\{t = k\} \times Z_j) + f_j + f_t + \eta_{jt}$$

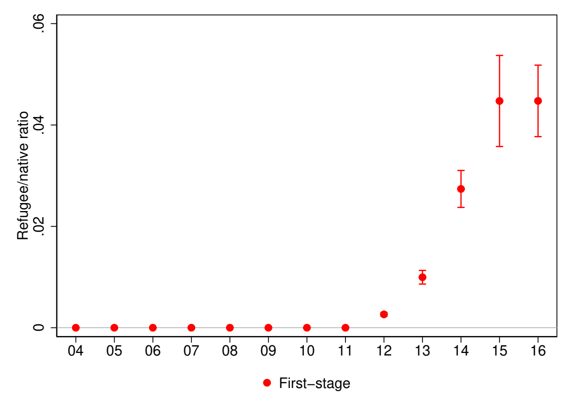


Figure 3: First-stage

Reduced-form: Employment rate

- Reduced form exhibits pre-trends (confounder U matters)

$$Y_{jt} = \sum_{k \neq 2010} \beta_k (\mathbb{1}\{t = k\} \times Z_j) + f_j + f_t + \epsilon_{jt}$$

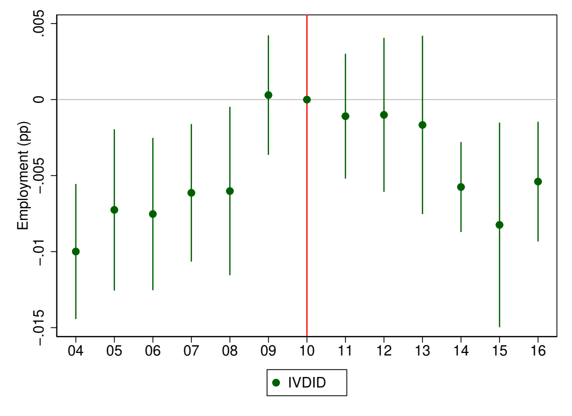


Figure 4: RF wage-employment rate

Common problem in shift-share designs

- Example from Autor, Dorn and Hanson (2013).

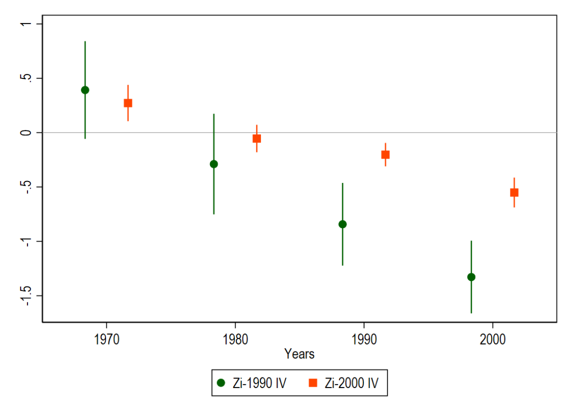


Figure 5: Percent manufacturing employment, treatment in 1990

Potential solutions

Common solutions in empirical papers

- Adding **two-way FEs** ($U_{it} = \alpha_i + \delta_t$), region FEs ($U_{it} = \alpha_g \times \delta_t$)
- Estimating a **linear trend** ($U_{it} = \alpha_i \times t$).
- Using observed **covariates** ($U_{it} = g(X_{it})$).
- But, in many economic examples agents make decisions on time varying unobservables (e.g. Llull 2017).

Current papers addressing related concerns:

1. **Boryusak and Hull 2023**: identification and counterfactual inference in shift-share designs when instruments mechanically depends on covariates.
2. **Arkhangelsky and Korovkin 2023**: IV with unobserved confounding when exogeneous variation is given by aggregate time-series shocks
3. **Danieli et al. 2024**: negative controls for IV. Show IV-DiD unbiased under PT ($U_{it} = \alpha_i + \delta_t$).

In this paper: address unmeasured confounding ($U_{it} = \mu_i' F_t$) in the framework of IV-DiD using SC.

General idea

- What to do when Z_{it} is correlated with $U_{it} = \mu_i' F_t$?
- If we could control for U_{it} we would be done
- **Idea**: in a pre-period we create a proxy for U_{it} then use IV on the "debiased" data

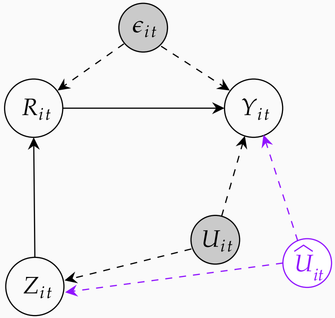


Figure 6: Triangular system with proxy

SC program: for each $j \in \{1, \dots, J\}$ we find the synthetic control weights $\hat{w}_j$ by solving the following program for $t \in \{1, \dots, T_0\}$. For each j,

$$\hat{w}_j^{SC} \in \text{argmin}_{w \in \mathcal{W}} \| Y_j^{T_0} - Y_{-j}^{T_0 \prime} w \|^2,$$

Today: $\mathcal{W} = \Delta^{J-1}$.

However, our results follow for any estimator in

$$\mathcal{W} = \{ w \in \mathbb{R}^J \mid \|w\|_1 \leq C \}.$$

- Nests the standard SC program with simplex constraint: $C = 1$.
- Allows for a variety of SC procedures (e.g. with l_1 penalties).
- While our results are valid for $\mathcal{W}$ estimators, in practice simplex regularization performs well.
- If R_{it} exists in pre-period we compute the weights on the residual of Y on R or match R.

The Synthetic IV

Step 2: Given $\{\tilde{Y}_{it}, \tilde{R}_{it}, \tilde{Z}_{it}\}_{t>T_0}^T$, we estimate the first stage and reduced form by OLS

$$\tilde{\pi} \in \text{argmin}_{\pi}(\tilde{Y} - \tilde{Z}\pi)'(\tilde{Y} - \tilde{Z}\pi),$$
$$\tilde{\beta} \in \text{argmin}_{\beta}(\tilde{R} - \tilde{Z}\beta)'(\tilde{R} - \tilde{Z}\beta),$$

where $\tilde{Z}$ includes an intercept.

Then, the estimated average marginal effect is given by:

$$\tilde{\theta}^{TSLS} = \left(\sum_{it} \tilde{Z}_{it}\tilde{R}_{it}\right)^{-1} \sum_{it} \tilde{Z}_{it}\tilde{Y}_{it}.$$

- Also possible to not debias Z, $\tilde{\theta}_{YR}^{TSLS}$, or to only debias Z, $\tilde{\theta}_{Z}^{TSLS}$.
- But, debiasing both will help in finite samples.
- Intuition from TSLS with covariates differs because we are estimating P_U.

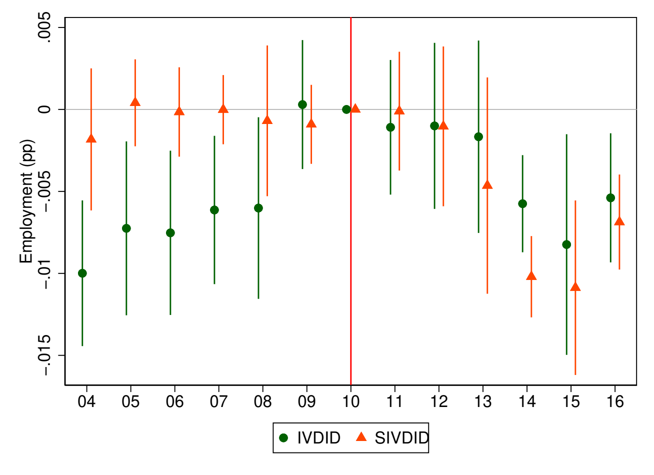
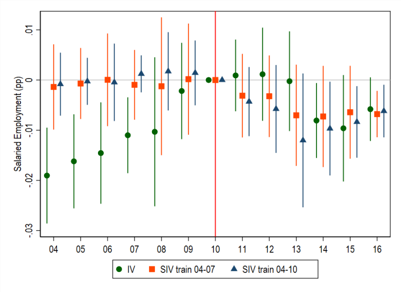
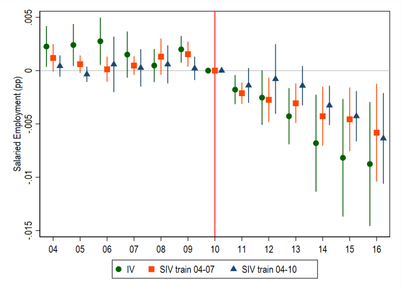


Figure 7: RF wage-employment rate

Reduced-form re-visited II



(a) Men: salaried employment



(b) Women: formal salaried employment

Figure 8: Additional examples of IV vs SIV

Treatment effects

- Using SIV can matter in practice

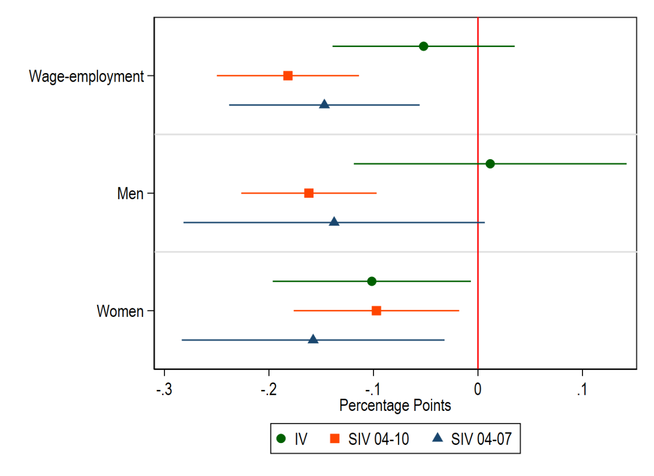


Figure 9: RF wage-employment rate

Setting

Linear Triangular System: J units, $T = T_0 + T_1$ time periods

$$Y_{it} = \theta R_{it} + \mu_i' F_t + \epsilon_{it},$$
$$R_{it} = \gamma Z_{it} + A_{it} + \eta_{it},$$

where $Z_{it} = 0$ for $t < T_0$.

Assumptions

- Partial independence condition: $\epsilon_{it}, \eta_{it} \perp Z_{it}$
- Bounded primitives: μ_i, F_t, A_{it}, Z_{it} are bounded
- Strong instrument and enough residual variation
$$\frac{1}{JT} Z'(I - P_U)Z \xrightarrow{p} Q > 0$$
- Errors: ϵ_{it} and η_{it} are $i.i.d$ mean zero, with bounded variance, covariance and fourth moments
- Stable rank: $\frac{\|Y^{T_0}\|_F^2}{\|Y^{T_0}\|_2^2} \leq \bar{r}$ and $\lambda_1(Y^{T_0}) \leq \bar{\sigma}_1$ (signal to noise assumption)

more detail

The Synthetic TSLS identification

As in FE models,

$$
\begin{aligned}
\tilde{Y}_{it} &= Y_{it} - \hat{Y}_{it}^{SC} \\
&= \theta R_{it} + \mu_i' F_t + \epsilon_{it} - \sum_{j \neq i} \hat{w}_{ij}^{SC} Y_{jt} \\
&= \theta \tilde{R}_{it} + (\mu_i - \sum_{j \neq i} \hat{w}_{ij}^{SC} \mu_j)' F_t + \epsilon_{it} - \sum_{j \neq i} \hat{w}_{ij}^{SC} \epsilon_{jt}.
\end{aligned}
$$

The **synthetic** TSLS is given by

$$
\begin{aligned}
\tilde{\theta}^{TSLS} &= \left(\sum_{it} \tilde{Z}_{it} \tilde{R}_{it} \right)^{-1} \sum_{it} \tilde{Z}_{it} \tilde{Y}_{it} \\
&= \theta + \left(\sum_{it} \tilde{Z}_{it} \tilde{R}_{it} \right)^{-1} \sum_{it} \tilde{Z}_{it} \left(\mu_i - \sum_{j \neq i} \hat{w}_{ij}^{SC} \mu_j \right)' F_t \\
&\quad + \left(\sum_{it} \tilde{Z}_{it} \tilde{R}_{it} \right)^{-1} \sum_{it} \tilde{Z}_{it} \left(\epsilon_{it} - \sum_{j \neq i} \hat{w}_{ij}^{SC} \epsilon_{jt} \right).
\end{aligned}
$$

Bound on factor term

In general, it need not be $o_p(1)$, the expected fit in the pre-treatment period plays a key role.

Theorem (Bound - MAD)

Under assumptions, for $t > T_0$ the following bound holds for all J, T and T_0

$$\frac{1}{JT_1}\left|\mathbb{E}\left[\sum_{it}\tilde{Z}_{it}\tilde{\mu}_i' F_t\right]\right| \leq \left(\frac{\bar{F}^2 k}{\xi}\right) c_z \left(2\sqrt{\frac{J}{T_0}}\sigma_\epsilon + \mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}_{jt}|\right]\right.$$
$$\left. + \mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{R}_{jt}|\right]\right)$$

Bound on factor term

Theorem (Bound)

Under assumptions, for $t > T_0$ the following bound holds for all J, T_1 and T_0

$$\frac{1}{JT_1}\left|\mathbb{E}\left[\sum_{it}\tilde{Z}_{it}\tilde{\mu}_i' F_t\right]\right| \leq \left(\frac{\bar{F}^2 k}{\underline{\xi}}\right) c_z \left(2\sqrt{\frac{J}{T_0}}\sigma_\epsilon + \bar{r}\bar{\sigma}_1\left[\frac{1}{\sqrt{T_0 J}} + \sqrt{\frac{J}{T_0}}\right]\right)$$

Furthermore, as $JT_1 \to \infty$ and $\bar{r}\bar{\sigma}_1\sqrt{\frac{J}{T_0}} \to 0$,

$$\frac{1}{JT_1}\sum_{it}\tilde{Z}_{it}\tilde{\mu}_i' F_t \xrightarrow{p} 0.$$

Consistency result

Theorem (Consistency)

Under assumptions, as $JT_1 \to \infty$ *and* $\bar{r}\bar{\sigma}_1\sqrt{\frac{J}{T_0}} \to 0$,

$$\tilde{\theta}^{TSLS} - \theta \xrightarrow{p} 0,$$

$$\tilde{\theta}_Z^{TSLS} - \theta \xrightarrow{p} 0,$$

$$\tilde{\theta}_{YR}^{TSLS} - \theta \xrightarrow{p} 0$$

Finite sample bias increases with

- **Noise**: σ_ϵ vs. $\mu_i' F_t$, worse fit
- **Short pre-periods**: $\sqrt{\frac{J}{T_0}}$, worse over-fitting
- **Correlation with confounder**: first stage gets worse

$$\frac{1}{JT_1}\sum_{it} \tilde{Z}_{it}\tilde{R}_{it} \xrightarrow{p} Q,$$

for $\tilde{\theta}^{TSLS}$, $\tilde{\theta}_{YR}^{TSLS}$ and $\tilde{\theta}_Z^{TSLS}$

Asymptotic normality

Theorem (Asymptotic normality)

Under assumptions, for $t > T_0$*, if* $\sqrt{\frac{T_1}{T_0}}(1+J)\bar{r}\bar{\sigma}_1 \to 0$ *as* $\sqrt{\frac{J}{T_1}} \to 0$*, then*

$$\frac{\sqrt{JT_1}(\tilde{\theta}^{SIV} - \theta)}{v_{JT}} \xrightarrow{d} N(0,1)$$

where $v_{JT_1} = (\mathbb{E}\tilde{Z}'\tilde{R})^{-2}\left(\mathbb{E}\left[\sum_{it} var(\tilde{Z}_{it}\tilde{\epsilon}_{it} \mid Z, w)\right]\right)$ *and*

$$\sum_{it} var(\tilde{Z}_{it}\tilde{\epsilon}_{it} \mid Z, w) = \sigma^2 \sum_{it} \Delta_{it}(Z, w)$$

- We require large number of pre and post periods $\frac{J}{T_1} \to 0$
- Condition on weight sequence such that $\|w^i\|_1 \le C$, then we do not need $J/T_1 \to 0$
- Alternatively, we also explore randomization inference

Additional estimators

- The instrument can remove some of the noise!

Suppose we can decompose $Z_{it} = Z_i \lambda_t$

Projected synthetic estimator

1. "de-noise" by projecting: $Y_z = Z(Z'Z)^{-1}Z'Y$, where Z is $J \times 1$
2. Use the de-noised outcomes to get the SC

$$w_j^P \in \text{argmin}_{w \in \Delta^{J-1}} \| Y_{z,j}^{T_0} - Y_{z,-j}^{T_0}{}' w \|^2$$

3. Compute $\tilde{\theta}_P^{TSLS}$ using the projected debiased data $\tilde{Y}^P$, $\tilde{R}^P$, and Z.

Bound comparison

- Baseline synthetic estimator biased when σ_ϵ high.
- Projected estimator less biased when σ_ϵ, but otherwise could fit unobserved *worse*.

SIV

$$\frac{1}{JT_1}\left|\mathbb{E}\left[\sum_{it}\tilde{Z}_{it}\tilde{\mu}_i' F_t\right]\right| \leq \left(\frac{\bar{F}^2 k}{\xi}\right) c_z \left(2\sqrt{\frac{J}{T_0}}\sigma_\epsilon + \mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}_{jt}|\right]\right)$$

Projected SIV

$$\frac{1}{JT_1}\left|\mathbb{E}\left[\sum_{it}\tilde{Z}_{it}\tilde{\mu}_i' F_t\right]\right| \leq \left(\frac{\bar{F}^2 k}{\xi}\right) c_z \left(2\sqrt{\frac{1}{T_0}}\sigma_\epsilon + \mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}^P_{jt}|\right]\right)$$

Ensemble estimator for $\alpha \in [0,1]$

$$\tilde{\theta}^E(\alpha) = \alpha\tilde{\theta} + (1-\alpha)\tilde{\theta}^P,$$

Simulation design

$$
\begin{aligned}
Y_{it} &= \beta R_{it} + \mu_i' f_t + \epsilon_{it}, \\
R_{it} &= (\gamma Z_{it} + \eta_{it}) * \mathbb{1}(t \geq T_0) \\
Z_{it} &= Z_i' g_t * \mathbb{1}(t \geq T_0)
\end{aligned}
$$

Time series structure:

$$
\begin{aligned}
f_t &= \kappa_f f_{t-1} + u_{ft}, \\
g_t &= \kappa_g g_{t-1} + u_{gt},
\end{aligned}
$$

Error structure:

$$
\begin{pmatrix} u_{ft} \\ g_{ft} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_f^2 & \rho_g \sigma_f \sigma_g \\ \rho_g \sigma_f \sigma_g & \sigma_g^2 \end{bmatrix}\right)
$$

$$
\begin{pmatrix} Z_i \\ \mu_i \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_z^2 & \rho_z \sigma_z \sigma_\mu \\ \rho_z \sigma_z \sigma_\mu & \sigma_\mu^2 \end{bmatrix}\right),
$$

$$
\begin{pmatrix} \epsilon_{it} \\ \eta_{it} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_\epsilon^2 & \rho \sigma_\epsilon \sigma_\lambda \\ \rho \sigma_\epsilon \sigma_\lambda & \sigma_\lambda^2 \end{bmatrix}\right).
$$

Calibrated simulation

Simulation based on Syrian example:

- $T = 16$, $T_0 = 10$, $J = 26$
- Target $\theta = -0.16$
- Noise level: $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.035$
- Instrument strength: match F-stat of 153; $\sigma_Z^2 = 0.54$, $\gamma = 3.16$
- Signal level: $\sigma_\mu^2 = 0.25$, $k = 2$ (from PCA)

Simulations designs for 10000 draws, we vary:

1. ρs to change degree of confounding
2. σ_ϵ to change noise level

Baseline

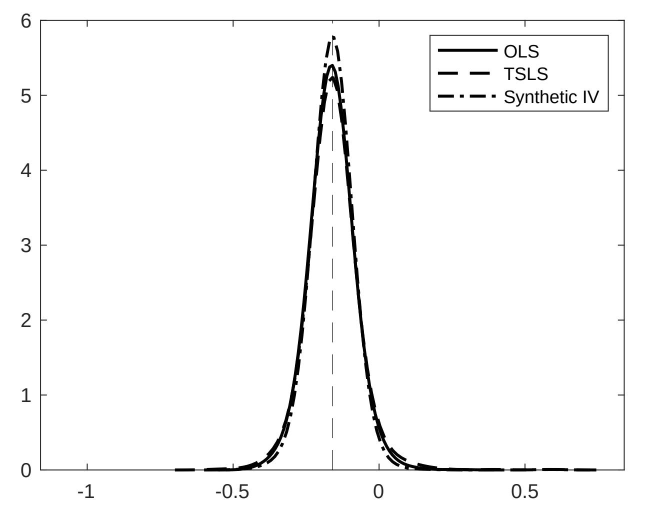


Figure 10: $\rho = \rho_z = \rho_g = 0$, 95% coverage is 0.96

Endogeneity

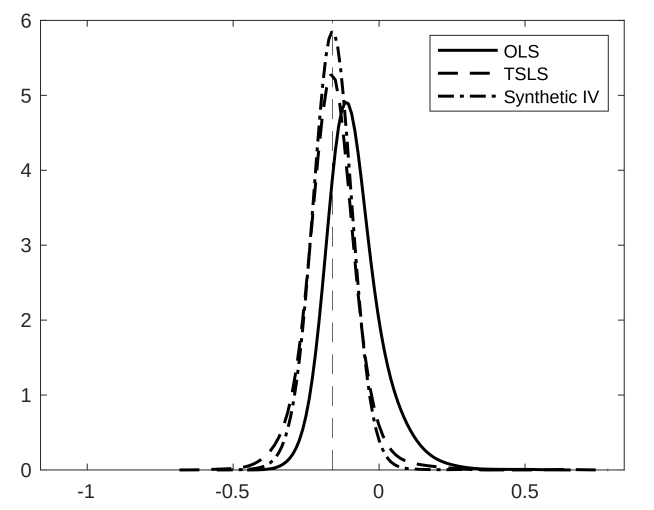


Figure 11: $\rho = 0.5, \rho_z = \rho_g = 0$, 95% coverage is 0.96

Endogeneity + OVB

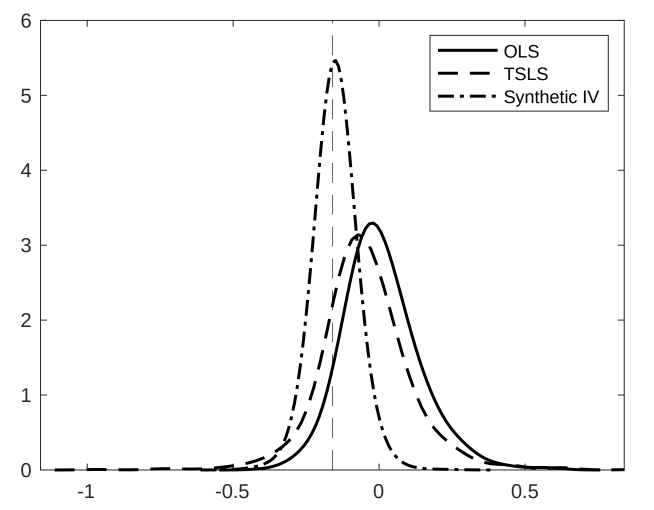


Figure 12: $\rho = \rho_z = \rho_g = 0.5$, 95% coverage is 0.94

Increasing the correlation with U

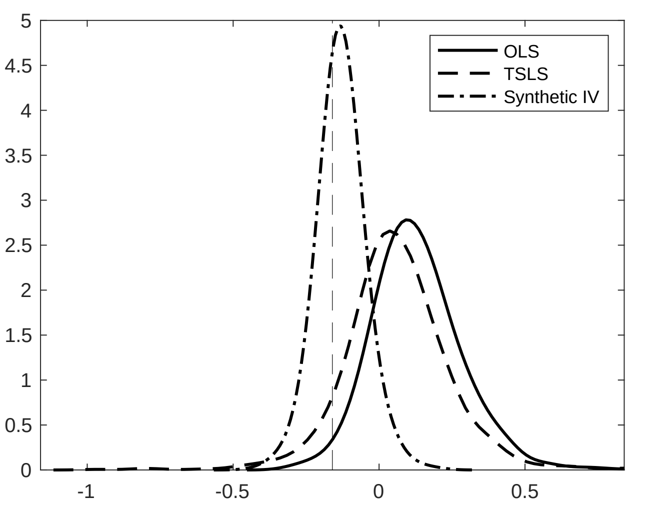


Figure 13: $\rho = \rho_z = \rho_g = 0.7$, 95% coverage is 0.89

Increasing the correlation with U

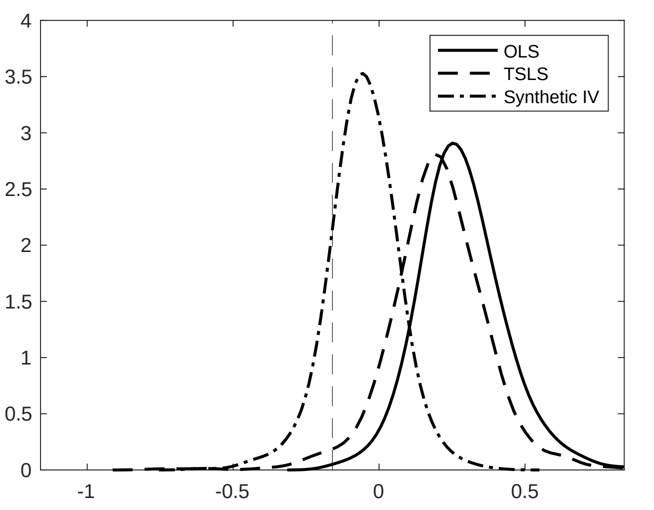
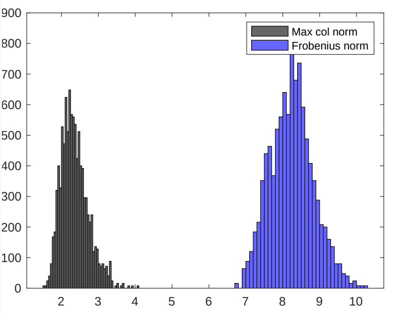


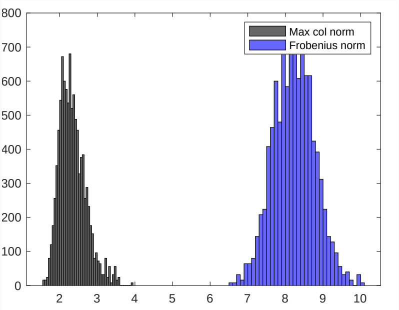
Figure 14: $\rho = \rho_z = \rho_g = 0.9$, 95% coverage is 0.70

Why is coverage good?

- Density check: no unit receives all weight



(a) $\rho = \rho_z = \rho_g = 0.5$



(b) $\rho = \rho_z = \rho_g = 0.9$

Increasing the noise

Model	Mean β	Var	$\lvert$Bias$\rvert$	MSE
$\rho = \rho_z = \rho_g = 0.5$				
OLS (twfe)	0.029	0.019	0.180	0.055
TSLS (twfe)	-0.047	0.025	0.112	0.038
SIV	-0.145	0.006	0.014	0.006
projected SIV	-0.188	0.015	0.028	0.016
SIV + projected	-0.151	0.006	0.008	0.006
SIV Z not debiased	-0.126	0.007	0.033	0.008

Table 1: $\sigma_\epsilon = 2\sigma_{Syria}$

Increasing the noise

Model	Mean β	Var	$|$Bias$|$	MSE
$\rho = \rho_z = \rho_g = 0.5$				
OLS (twfe)	0.069	0.026	0.229	0.078
TSLS (twfe)	-0.047	0.028	0.112	0.041
SIV	-0.134	0.012	0.025	0.013
projected SIV	-0.183	0.023	0.023	0.024
SIV + projected	-0.144	0.012	0.015	0.012
SIV Z not debiased	-0.113	0.012	0.046	0.014

Table 2: $\sigma_\epsilon = 4\sigma_{Syria}$

Increasing the noise

Model	Mean β	Var	$\lvert$Bias$\rvert$	MSE
$\rho = \rho_z = \rho_g = 0.5$				
OLS (twfe)	0.203	0.067	0.363	0.199
TSLS (twfe)	-0.048	0.046	0.111	0.058
SIV	-0.101	0.045	0.058	0.048
projected SIV	-0.152	0.070	0.007	0.071
SIV + projected	-0.119	0.044	0.041	0.046
SIV Z not debiased	-0.081	0.037	0.078	0.043

Table 3: $\sigma_\epsilon = 8\sigma_{Syria}$

Recap

- Good finite sample Bias + MSE

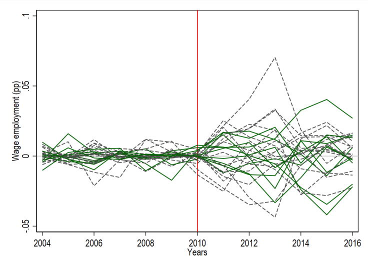
but the method is sensitive to

1. High correlation with the unobserved con-founder
2. High noise-signal ratio
3. **Ensemble estimator**: SIV + projected estimator may offer "double-robustness"

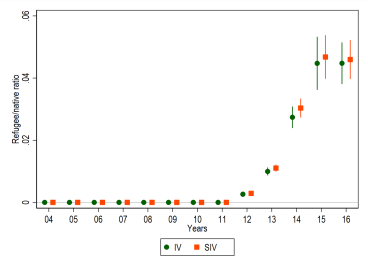
Diagnostic checks:

1. Asses good pre-treatment fit (MAD plots)
2. Back testing the start of the treatment
3. Check the instrument strength
4. Check density of weights

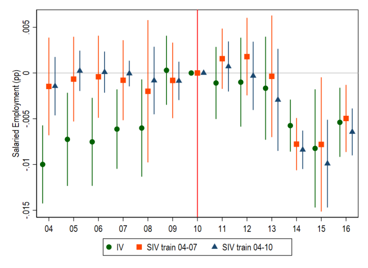
Empirical checks I



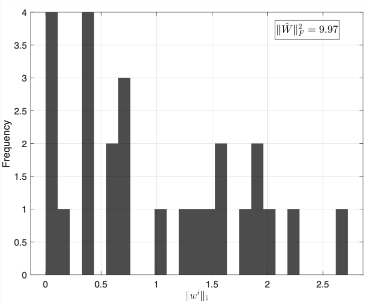
(a) Debiased Data



(b) Debiased First-Stage



(a) Debiased Reduced-Form



(b) Weight density

ADH revisited

- Decreased effect for 1999-2000

Table 4: China shock effect

	1990–2000	2000-2007	1990–2007
	(1)	(2)	(3)
2SLS	-0.888	-0.718	-0.746
	(0.181)	(0.064)	(0.068)
SIV	-0.588	-0.726	-0.703
	(0.198)	(0.070)	(0.067)

Notes: The first row replicates columns 1–3 of Table 2 in ADH 2013. The second row presents the estimates using the SIV. The SC weights are estimated using the manufacturing growth rates in 1970 and 1980.

Ranking example I

- Use promotion contracts (Z) as an IV for producer ranking (R) in digital platform to study the effect on sales (Y)
- A/B test in which ranks were randomized available!

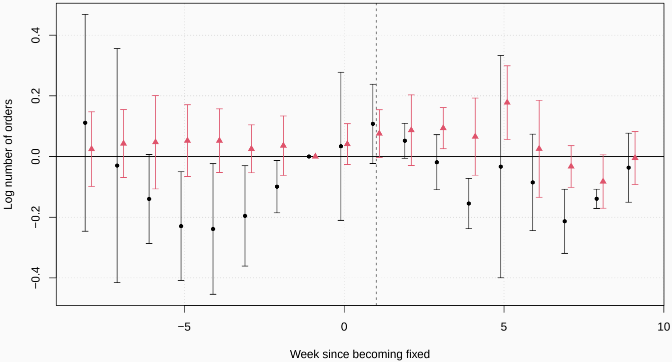


Figure 18: Reduced form event study of promotion on log sales.

Ranking example II

- IV biased upwards as expected from positive OVB
- SIV recovers the A/B test

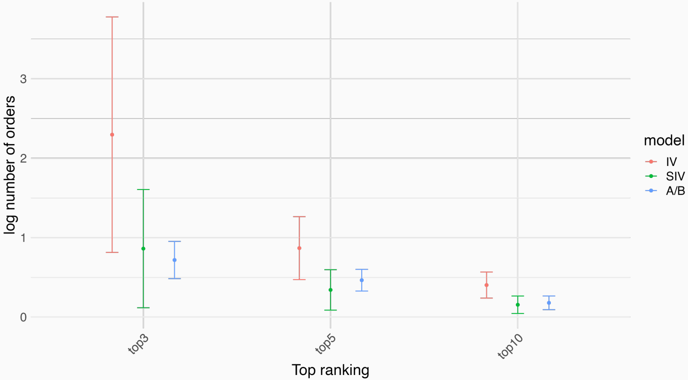
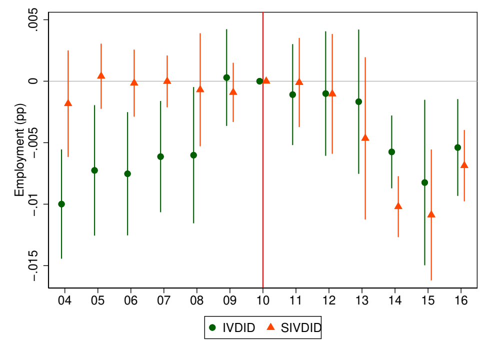


Figure 19: Treatment effects of being in top ranks on sales.

Conclusion

- Un-measured confounding can bias DiD and IVDiD estimates
- The *synthetic* IV offers one solution based on SC
- We derive conditions for consistency and asymptotic normality
- Show the applicability of the method empirically and through simulations



New draft coming soon!

Additional slides

Regression Set-up

- Let $R_{j,t}$ denote the refugee/native ratio at province-year level
- Refugee location choice is endogenous: use travel distance as an instrument:

$$Z_{j,t} = \underbrace{\bar{H}_t}_{\text{shift}} \times \underbrace{Z_j}_{\text{share}}$$
$$Z_j = \sum_{s=1}^{13} \lambda_s \frac{1}{d_{j,s}} \tag{1}$$

where

- $\bar{H}_t$ is the number of refugees in Turkey in year t.
- $d_{j,s}$ is the travel distance between Turkish region p and Syrian governorate s
- λ_s is the weight given to Syrian governorate s
- today: $\lambda_s = \pi_s$, population share of Syrian governorate s

return

Setting

- Two unobserved (grey) confounders.
- Goal: estimate causal effect of R_{it} on Y_{it}.

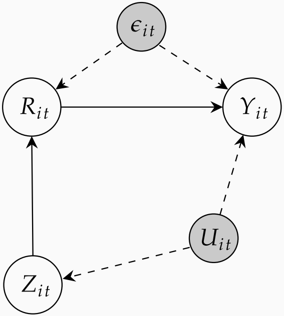


Figure 20: Triangular system with unknown confounders.

OLS

- OLS of Y on R valid if $R_{it} \perp \epsilon_{it}, U_{it}$ (R exogenous).

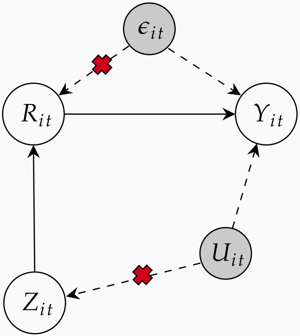


Figure 21: Example of independence relationships for OLS.

- IV requires that $Z_{it} \perp \epsilon_{it}, U_{it}$.

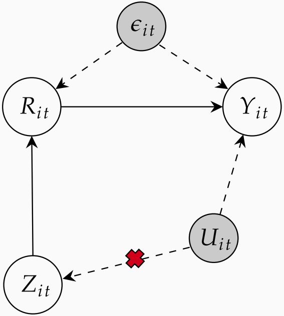


Figure 22: Example of independence relationships for IV.

Model Assumptions I

Assumption (A1)

Independence condition: $\epsilon_{it}, \eta_{it} \perp Z_{it}$

Assumption (A2)

- $|F_{lt}| \leq \bar{F}$.
- $F^{'}_{T_0} F_{T_0}$ *has minimum eigenvalue* ξ *such that* $\xi / T_0 > 0$.
- $\mu_i \in \mathcal{M}$ *with* $diam(\mathcal{M}) = \sup\{\|t - s\| : \text{for } t, s \in \mathcal{M}\} \leq c_\mu$.
- $Z_{it} \in \mathcal{Z}$ *such that* $diam(\mathcal{Z}) = \sup\{\|t - s\| : \text{for } t, s \in \mathcal{Z}\} \leq c_z$.
- *The instrument* Z_{it} *and the unobserved factor structure satisfy*

$$\frac{1}{JT} \sum_{it} Z_{it}^2 \xrightarrow{p} Q_Z > 0,$$

as $JT \rightarrow \infty$ *and*

$$\frac{1}{JT} Z'(I - P_U) Z \xrightarrow{p} Q > 0,$$

for projection matrix U.

Model Assumptions II

Assumption (A3)

With probability one,

$$\frac{\|Y^{T_0\prime}\|_F^2}{\|Y^{T_0\prime}\|_2^2} \leq \bar{r},$$

and the largest singular value satisfies $\sigma_1(Y^{T_0}) \leq \bar{\sigma}_1$*, where* $\bar{r}$ *and* $\bar{\sigma}_1$ *may depend on* J *and* T_0*.*

Theorem (Bias bound)

Under A1-A3, for $t > T_0$ *the following bound holds for all* J, T *and* T_0

$$\frac{1}{JT}\left|\mathbb{E}\left[\sum_{it} \tilde{Z}_{it}\tilde{\mu}_i' F_t\right]\right| \leq \left(\frac{\bar{F}^2 k}{\xi}\right) c_z \left(2\sqrt{\frac{J}{T_0}}\sigma_\epsilon + \bar{r}\bar{\sigma}_1\left[\frac{1}{\sqrt{T_0 J}} + \sqrt{\frac{J}{T_0}}\right]\right)$$

Furthermore, as $JT \to \infty$ *and* $\bar{r}\bar{\sigma}_1\sqrt{\frac{J}{T_0}} \to 0$,

$$\frac{1}{JT}\sum_{it} \tilde{Z}_{it}\tilde{\mu}_i' F_t \xrightarrow{p} 0.$$

Two additional estimators

- The instrument can remove some of the noise!

Suppose we can decompose $Z_{it} = Z_i \lambda_t$

Projected synthetic estimator

1. "de-noise" by projecting: $Y_z = Z(Z'Z)^{-1}Z'Y$, where Z is $J \times 1$
2. Use the de-noised outcomes to get the SC

$$w_j^P \in \text{argmin}_{w \in \Delta^{J-1}} \| Y_j^{T_0} - Y_{z,-j}^{T_0}{}' w \|^2$$

Aggregated synthetic estimator

1. Let $Q_i = \sum_{t<T_0} Z_{it} Y_{it}$
2. Match the aggregated values

$$w_j^{Agg} \in \text{argmin}_{w \in \Delta^{J-1}} \| Q_i - Q_{-i}' w \|^2$$

Ensemble estimator

- Baseline synthetic estimator biased when σ_ϵ high.
- Projected/Agg estimators less biased when σ_ϵ, but otherwise could fit unobserved *worse*.

Ensemble estimator for $\alpha \in [0, 1]$

$$\tilde{\theta}^E(\alpha) = \alpha\tilde{\theta} + (1 - \alpha)\tilde{\theta}^P,$$

- α hyper-parameter can be chosen to minimise *MSE* in a validation period.

return

Event study simulation

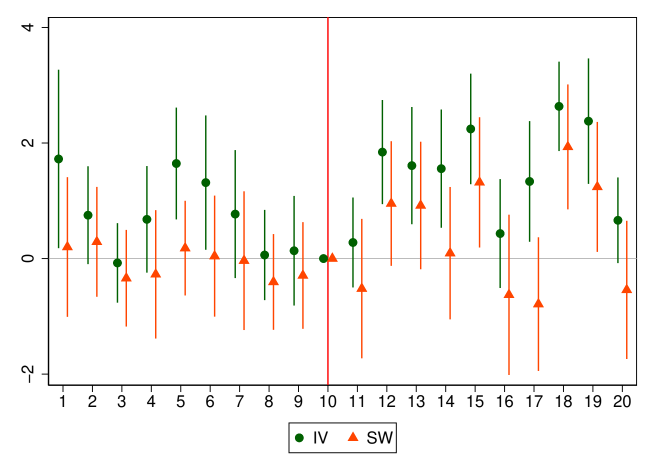


Figure 23: Example event study simulation.

Empirical correlation

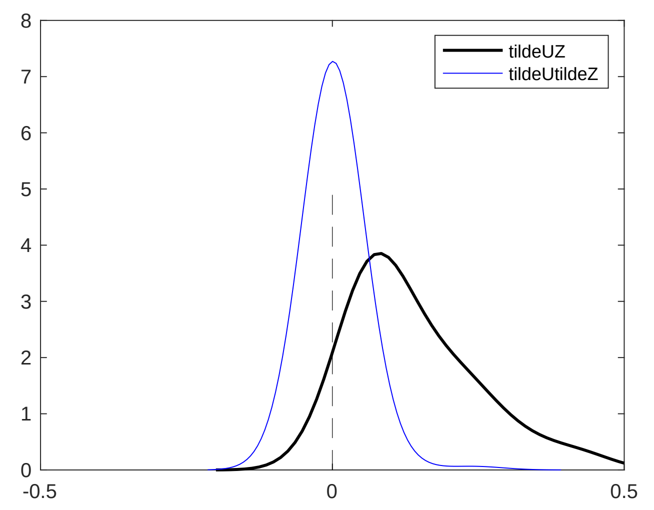


Figure 24: Factor correlation for $\rho = 0.7$.

Debiasing the first stage?

We want to compare

- $\tilde{\beta} = (\sum_{it} \tilde{R}_{it}\tilde{R}_{it})^{-1} \sum_{it} \tilde{R}_{it}\tilde{Z}_{it}$,
- $\tilde{\beta}_Z = (\sum_{it} \tilde{R}_{it}\tilde{R}_{it})^{-1} \sum_{it} \tilde{R}_{it}Z_{it}$.

Lemma

Consider an instrument of the form:

$$Z_{it} = A_{it} + \alpha\mu_i F_t,$$

where $A_{it} \sim_{iid} N(0, \sigma_A^2)$ *and* $\mu_i' F_t \perp A_{it}$. *Suppose that* $\tilde{\mu}_i' F_t = o_p(1)$, *then*

$$\frac{\tilde{\beta}}{\tilde{\beta}_Z} = \frac{\sum_{it} \tilde{R}_{it}\tilde{Z}_{it}}{\sum_{it} \tilde{R}_{it}Z_{it}} \xrightarrow{p} \xi \geq 1.$$

- So de-biasing the instrument could matter asymptotically.

More details

Randomization Inference

- Asymptotic results require large T. Often we have short panels.
- Randomization inference common practice in SC studies
    - e.g. Abadie et al. (2010), Firpo and Possebom (2018)...
- Challenge is that we have a continuous treatment and an instrument.

Two approaches:

1. **Split conformal inference**: define a blank period and perform permutation tests comparing the distribution of realized $\tilde{\theta}_t$ with blank period $\tilde{\theta}_t^b$.
    1.1 Chernozhukov et al. (2021), Abadie and Zhao (2022).
2. **Randomization inference**: permute instrument-treatment pairs across units, compare permuted estimates $\tilde{\theta}_b$ with realized $\tilde{\theta}$.
    - Imbens and Rosenbaum (2003).

Split conformal inference

1. Split T_0 period into a *training* period and a *blank* period (starting at $T_b < T_0$).
2. Compute SC weights in the training period and define debiased quantities.
3. Run reduced form event regression to get estimates $\{\tilde{\theta}_{T^b}, \ldots, \tilde{\theta}_{T_0}, \ldots, \tilde{\theta}_T\}$.
4. Generate a permutation $(T - T_0) \times 1$ vector $e_\pi = (\tilde{\theta}_{\pi(1)}, \ldots, \tilde{\theta}_{\pi(T-T_0)})$.
5. Compute test statistic: $S(e) = 1/(T - T_0)\|e\|_q$, q=1.
6. Permutation p-value:

$$\hat{p} = \frac{1}{\Pi} \sum_{\pi \in \Pi} 1(S(\tilde{\theta}_\pi) \geq S(\tilde{\theta}_{t > T_0}))$$

- Requires exchangeability across time of ϵ_{it}, F_t.
- Requires blank time periods.

Randomization inference

1. In the pre-period compute the SC weights and generate the debiased quantities.
2. Define the set of permutations of the J units: $\mathcal{P}(J)$.
3. For a given permutation $\pi \in \mathcal{P}(J)$, compute

$$\tilde{\theta}_\pi = \left(\sum_{it} \tilde{Z}_{\pi(i)t} \tilde{R}_{\pi(i)t} \right)^{-1} \sum_{it} \tilde{Z}_{\pi(i)t} \tilde{Y}_{it},$$

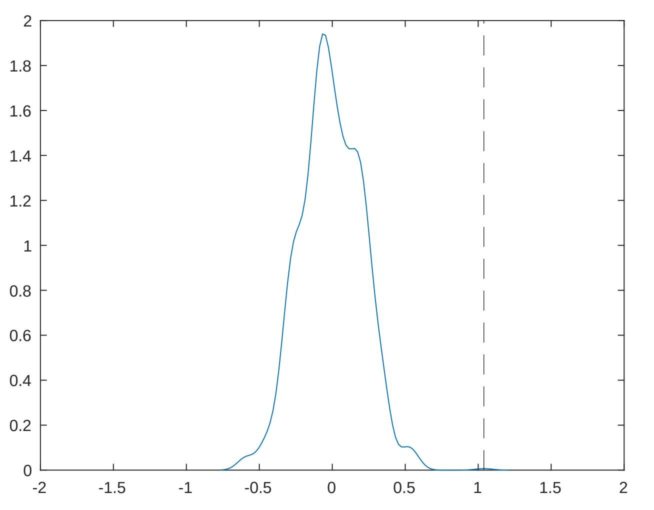
   where we permute the individuals for Z and R but not Y.
4. p-value:

$$\hat{p} = \frac{1}{\mathcal{P}(J)} \sum_{\pi \in \mathcal{P}(J)} P(\tilde{\theta}_\pi \geq \tilde{\theta}))$$

- Requires exchangeability across units of ϵ_{it}, μ_i.
- Approximation given the number of permutations.

Randomization inference

- Permutation distribution example for simulation design with moderate correlation.



Back-testing the RF

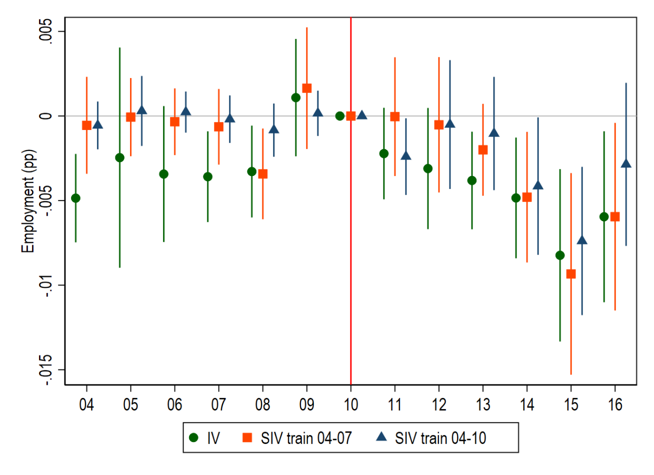


Figure 25: RF wage-employment rate

return

Endogeneity

Model	Mean β	Var	$\|$Bias$\|$	MSE
$\rho = 0.5, \rho_z = \rho_g = 0$				
OLS (twfe)	1.245	0.012	0.245	0.072
TSLS (twfe)	0.994	0.017	0.006	0.017
SIV	0.995	0.011	0.005	0.011
projected SIV	0.975	0.043	0.025	0.044
SIV + projected	0.994	0.010	0.006	0.010
SIV Z not debiased	0.992	0.011	0.008	0.010

Table 5: $\mathrm{T}_0 = 20, T = 30, J = 20, \sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Endogeneity + OVB

Model	Mean β	Var	$\lvert$Bias$\rvert$	MSE
$\rho = \rho_z = \rho_g = 0.5$				
OLS (twfe)	1.379	0.019	0.379	0.162
TSLS (twfe)	1.260	0.069	0.260	0.136
SIV	1.030	0.013	0.030	0.014
projected SIV	0.901	0.046	0.099	0.056
SIV + projected	1.008	0.013	0.008	0.013
SIV Z not debiased	1.079	0.017	0.079	0.023

Table 6: $T_0 = 20, T = 30, J = 20, \sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Increasing OVB

Model	Mean β	Var	$\vert$Bias$\vert$	MSE
$\rho = 0.5, \rho_z = \rho_g = 0.7$				
OLS (twfe)	1.501	0.021	0.501	0.272
TSLS (twfe)	1.505	0.084	0.505	0.339
SIV	1.080	0.024	0.080	0.031
projected SIV	0.935	0.062	0.065	0.066
SIV + projected	1.038	0.025	0.038	0.027
SIV Z not debiased	1.200	0.028	0.200	0.068

Table 7: $T_0 = 20, T = 30, J = 20, \sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Increasing OVB

Model	Mean β	Var	$\|$Bias$\|$	MSE
$\rho = 0.5, \rho_z = \rho_g = 0.9$				
OLS (twfe)	1.661	0.018	0.661	0.455
TSLS (twfe)	1.826	0.064	0.826	0.747
SIV	1.268	0.054	0.268	0.125
projected SIV	1.142	0.087	0.142	0.107
SIV + projected	1.207	0.056	0.207	0.099
SIV Z not debiased	1.519	0.050	0.519	0.328

Table 8: $T_0 = 20, T = 30, J = 20, \sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Increasing OVB

Model	Mean β	Var	$\|$Bias$\|$	MSE
$\rho = 0.5, \rho_g = \rho_z = 1$				
OLS (twfe)	1.745	0.010	0.745	0.566
TSLS (twfe)	2.006	0.034	1.006	1.045
SIV	1.879	3.338	0.879	4.107
projected SIV	2.141	1.213	1.141	2.514
SIV + projected	2.033	1.446	1.033	2.511
SIV Z not debiased	2.024	0.025	1.024	1.082

Table 9: $T_0 = 20$, $T = 30$, $J = 20, \sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Increasing the noise

Model	Mean β	Var	$\|$Bias$\|$	MSE
$\rho = \rho_g = \rho_z = 0.5, \sigma_\epsilon = 2$				
OLS (twfe)	1.625	0.037	0.625	0.428
TSLS (twfe)	1.252	0.089	0.252	0.152
SIV	1.076	0.049	0.076	0.054
projected SIV	0.849	0.131	0.151	0.154
SIV + projected	1.014	0.049	0.014	0.049
SIV Z not debaised	1.130	0.053	0.130	0.067

Table 10: $\mathrm{T}_0 = 20$, $T = 30, J = 20, \sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Increasing J/T_0

- Accentuates other problems, but is less important in low noise settings.

Model	Mean β	Var	|Bias|	MSE
$\rho = \rho_g = \rho_z = 0.5$, $T_0 = 10, J = 20$				
OLS (twfe)	1.365	0.018	0.365	0.151
TSLS (twfe)	1.222	0.058	0.222	0.107
SIV	1.023	0.018	0.023	0.019
projected SIV	0.929	0.053	0.071	0.058
SIV + projected	1.004	0.016	0.004	0.016
SIV Z not debiased	1.076	0.017	0.076	0.022

Table 11: T=20, $\sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Increasing J/T_0

Model	Mean β	Var	$\lvert$Bias$\rvert$	MSE
$\rho = \rho_g = \rho_z = 0.5$, $T_0 = 10$, $J = 40$				
OLS (twfe)	1.377	0.013	0.377	0.155
TSLS (twfe)	1.256	0.041	0.256	0.107
SIV	1.028	0.009	0.028	0.010
projected SIV	0.949	0.029	0.051	0.032
SIV + projected	1.014	0.008	0.014	0.008
SIV Z not debiased	1.061	0.009	0.061	0.014

Table 12: T=20, $\sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$

Coverage

- Good coverage in good settings
- Under coverage in higher bias settings

Coverage $\alpha = 0.05$			
	T=30	T=40	T=50
$\rho = \rho_g = \rho_z = 0.0$	0.981	0.962	0.952
$\rho = \rho_g = \rho_z = 0.3$	0.976	0.944	0.96
$\rho = \rho_g = \rho_z = 0.5$	0.960	0.945	0.923
$\rho = \rho_g = \rho_z = 0.7$	0.904	0.808	0.792

Table 13: $T_0 = 20, J = 20, \sigma_\epsilon = 0.5, \sigma_z = 1, \sigma_{other} = 0.5, \kappa = 0.5$