Causal inference for large dimensional non-stationary panels with two-way endogenous treatment

Junting Duan, Markus Pelger and Ruoxuan Xiong

Stanford University and Emory University

Motivation

Problem: Large dimensional panel data (both N and T large) with missing entries is prevalent

Goal: Impute missing values and provide entrywise inference for imputations

Casual inference: Unobserved counterfactual outcomes can be modeled as missing values

Challenges:

- Latent dependency structure of panel generally unknown
- Non-stationary data in time and cross-section
- Complex and potentially endogenous missing patterns

Need for new methods:

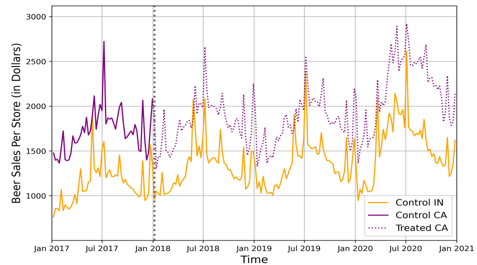
⇒ Existing imputation methods with inferential theory require stationary panels or restrictive assumptions on missingness

⇒ Violated by many empirical applications

A Motivating Example: Liberalization of Marijuana

The effect of legalization of recreational marijuana on per-capita beer sales (Li and Sonnier (2023))

- Economic question: Is marijuana a substitute or complement for alcohol?
- Data: 208 weekly observations between 2017-2020 for 45 states in the U.S.
- Treatment: 6 treated states legalize marijuana at different time points + 39 control states



Challenges:

1. Outcome data is non-stationary with common trends
2. Endogenous treatment pattern with confounders in both cross-sectional and temporal dimensions
3. Complex latent interactive confounders (factors)

Model Setup

Our solution: Approximate factor model with k latent factors and two-way non-stationary fixed effects for control panel Y with missing observations: N units over T time periods

$$Y_{it} = \underbrace{\mu + \alpha_i + \xi_t + \Lambda_i^\top F_t}_{\text{common component}} + \underbrace{\epsilon_{it}}_{\text{error}}$$

α_i, ξ_t: **two-way fixed effects**

- Allow for arbitrary non-stationary time trends in ξ_t
- Allow for confounders in both cross-sectional and time series dimensions
- Generally more efficient than subsuming fixed effects by latent factors

$\Lambda_i^\top F_t$: k **latent factors**

- Precise estimation by explaining most variations in the outcome variable
- Data-driven approach to learn latent interactive confounders

$\Rightarrow$ **Two special cases**:

1. When we remove $\Lambda_i^\top F_t$, degenerate to a difference-in-difference (DID) framework
2. When we remove α_i, ξ_t (include them in factors), degenerate to pure factor model

Model Setup: Observation Pattern

Observation matrix $W = [W_{it}]$ for panel Y: $W_{it} = \begin{cases} 1 & Y_{it} \text{ is observed} \\ 0 & Y_{it} \text{ is missing} \end{cases}$

1. W can have any pattern with sufficiently many observations for each unit and time:

(a) Simultaneous adoption (b) Staggered adoption (c) Treatment switch on/off

2. W can have complicated dependency on the following components:
   - Observed characteristics O
   - Unit fixed effects α and latent factor loadings Λ
   - Time fixed effects $\{\xi_s\}_{s=1}^{t}$ up to time t
3. W can have arbitrary time series dependency: $W_{it} \not\perp W_{is} | I$, e.g. $I = \{O, \alpha, \Lambda, \xi\}$

$\Rightarrow$ Most general assumptions on the missing pattern in this literature ▸ assumptions

Challenges

Challenge: How can we estimate the fixed effects?

- Complicated observation pattern and its dependency on the fixed effects and factor model

$\Rightarrow$ Least square estimation is biased:

$$(\tilde{\mu}, \tilde{\alpha}, \tilde{\xi}, \tilde{\Lambda}, \tilde{F}) = \arg \min_{\mu,\alpha,\xi,\Lambda,F} \sum_{i,t} W_{it}(Y_{it} - \mu - \alpha_i - \xi_t - \Lambda_i^\top F_t)^2$$

This paper: Propose a novel method, Within-Transform-PCA (wi-PCA), to consistently estimate the common components and make inference on the imputed values

Contribution

Methodology:

- Most general assumptions on the missing pattern in this literature:
  Missingness can depend on both the two-way fixed effects and latent factor model
- A novel estimator for two-way fixed effects and factor model by carefully weighting all the temporal and cross-sectional observations
- Entrywise inferential theory for estimator and feasible approach to construct confidence intervals

Empirics:

Effect of legalization of recreational marijuana on per-capita beer sales

$\Rightarrow$ Superior performance compared to special cases of difference-in-differences and PCA methods

- More accurate imputation out-of-sample
- Only fixed-effects (DID) have omitted variable bias
- PCA methods are unstable and have large variance
- Omitting fixed effects or factors leads to excessive treatment effects with spurious significance

Related Literature (Incomplete and Partial List)

Large dimensional factor modeling

- Full observations with inferential theory: Bai and Ng 2002, Bai 2003, Fan, Liao and Mincheva 2013, Pelger and Xiong 2021a+b
- Partial observations: Stock and Watson 2002, Jin, Miao and Su 2021, Bai and Ng 2021a, Cahan, Bai and Ng 2022, Xiong and Pelger 2022, Duan, Pelger and Xiong 2023, Ng and Scanlan 2024

Causal inference in panels

- One treated unit: Abadie, Diamond and Hainmueller 2010, 2015
- Block pattern: Xu 2017, Arkhangelsky, Athey, Hirshberg, Imbens and Wager 2021
- Staggered adoption: Athey and Imbens 2022
- General pattern: De Chaisemartin and D'Haultfoeuille 2020, Athey, Bayati, Doudchenko, Imbens and Khosravi 2021, Arkhangelsky and Imbens 2022

Estimator and Inferential Theory

Our Method: wi-PCA

Estimator: **wi-PCA = within transform + PCA**

$$Y_{it} = \underbrace{\mu + \alpha_i + \xi_t + \Lambda_i^\top F_t}_{\text{common component}} + \underbrace{\epsilon_{it}}_{\text{error}}$$

Step 1: (within-transform)

Estimate the grand mean μ and two-way fixed effects ξ and α

Step 2: (PCA)

Estimate the latent factor structure $\Lambda_i^\top F_t$ and common component from within-transformed panel

Step 1 – Estimate the Two-Way Fixed Effects

Step 1: Estimate grand mean μ and two-way fixed effects ξ and α as weighted averages of observed Y

$$\tilde{\mu} = \sum_{i=1}^{N} \sum_{t=1}^{T} M_{it}^{\mu} Y_{it}$$

$$\tilde{\xi}_t = \sum_{i=1}^{N} M_{it}^{\xi} Y_{it} - \tilde{\mu}$$

$$\tilde{\alpha}_i = \sum_{t=1}^{T} M_{it}^{\alpha} (Y_{it} - \tilde{\xi}_t) - \tilde{\mu}$$

- Weight characterized by M_{it}^{μ}, M_{it}^{ξ} and M_{it}^{α}

$\Rightarrow$ Different weights for different cases

$\Rightarrow$ Note: $\tilde{\xi}_t$ and $\tilde{\alpha}_i$ are not symmetric!

Reason: Observation pattern arbitrarily dependent in time dimension, but conditionally independent in cross-sectional dimension

Intuition for Constructing Weights

All entries in Y are observed

- Simple within estimator: $M_{it}^{\mu} = 1/(NT)$, $M_{it}^{\xi} = 1/N$ and $M_{it}^{\alpha} = 1/T$

Missing entries in Y

- Weights have to be constructed to consistently estimate $\mu + \alpha_i + \beta_t$:

$$\tilde{\mu} + \tilde{\alpha}_i + \tilde{\xi}_t - (\mu + \alpha_i + \xi_t)$$

$$= \sum_{s=1}^{T} M_{is}^{\alpha} \left(\Lambda_i^\top F_s + \epsilon_{is} \right) - \sum_{s=1}^{T} M_{is}^{\alpha} \sum_{j=1}^{N} M_{js}^{\xi} \left(\alpha_j + \Lambda_j^\top F_s + \epsilon_{js} \right) + \sum_{j=1}^{N} M_{jt}^{\xi} \left(\alpha_j + \Lambda_j^\top F_t + \epsilon_{jt} \right)$$

- M_{it}^{μ}: does not affect error; without loss of generality, set $M_{it}^{\mu} = \frac{W_{it}}{\sum_{s=1}^{T} \sum_{j=1}^{N} W_{js}}$
- M_{it}^{α}: set $M_{it}^{\alpha} = \frac{W_{it}}{\sum_{s=1}^{T} W_{is}}$, so that the first error term is $o_p(1)$
- M_{it}^{ξ}: selection is main challenge

  $\Rightarrow$ Different weights for three different cases of observation patterns

Step 1 – Estimate the Two-Way Fixed Effects

Case 1: Known observation probability $p_{it} = P(W_{it} = 1)$ (design-based settings)

- Adjust the observed Y_{it} by the inverse observation probability p_{it}

⇒ Correct for the bias when observations are not missing uniformly at random

Construct the weights as (Hajek estimator)

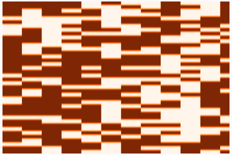
$$M_{it}^{\xi} = \left(\sum_{j=1}^{N} \frac{W_{jt}}{p_{jt}} \right)^{-1} \frac{W_{it}}{p_{it}}$$

Step 1 – Estimate the Two-Way Fixed Effects

Case 2: Unknown observation probability p_{it} with short-term missingness with factor structure (observational study)

We can estimate the probability p_{it} if

- The observation probability p_{it} can be factorized into a one-factor model as $p_{it} = u_i v_t$
- No long-term dependency of time series missingness
  Example: Treatment assignments switch on and off



$\Rightarrow$ Replace p_{jt} by $\bar{W}_{j,\cdot}$ (up to a scaling constant) in M^{ξ}

$$M^{\xi}_{it} = \left(\sum_{j=1}^{N} \frac{W_{jt}}{\bar{W}_{j,\cdot}} \right)^{-1} \frac{W_{it}}{\bar{W}_{i,\cdot}}$$

Step 1 – Estimate the Two-Way Fixed Effects

Case 3: Unknown observation probability p_{it} with monotone missingness (observational study)

Time series observation patterns are monotone i.e. $W_{i1} \geq W_{i2} \geq \cdots \geq W_{iT}$ holds for all i

Examples:

(a) Simultaneous adoption

(b) Staggered adoption

⇒ Use fully observed control units to estimate the time fixed effects

$$M_{it}^{\xi} = \begin{cases} N_c^{-1} & \text{if unit } i \text{ is observed for all times } t = 1, \cdots, T \\ 0 & \text{otherwise} \end{cases}$$

⇒ Three cases altogether cover the important examples of observation patterns

⇒ Consistent estimation even if observation patterns depend on both α_i and ξ_t

Step 2 – Estimate the Factor Structure and Common Component

Step 2: Estimate the factor structure $\Lambda_i^\top F_t$

- Within-transform Y to $\dot{Y}_{it} = Y_{it} - \tilde{\mu} - \tilde{\xi}_t - \tilde{\alpha}_i$
- Estimate the second moment matrix $\Sigma = \dot{Y}\dot{Y}^\top / T$ as

$$\tilde{\Sigma}_{ij} = \frac{1}{|Q_{ij}|} \sum_{t \in Q_{ij}} \dot{Y}_{it} \dot{Y}_{jt}$$

  where $Q_{ij} = \{t : W_{it} = W_{jt} = 1\}$ (Xiong and Pelger (2022))
- Estimate loadings as $\sqrt{N}$ times the eigenvectors of the k largest eigenvalues of $\tilde{\Sigma}/N$
- Estimate latent factors by regressing the observed $\dot{Y}$ on $\tilde{\Lambda}$

$$\tilde{F}_t = \left(\frac{1}{N} \sum_{i=1}^{N} W_{it} \tilde{\Lambda}_i \tilde{\Lambda}_i^\top \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} W_{it} \tilde{\Lambda}_i \dot{Y}_{it} \right)$$

$\Rightarrow$ Estimate common components/impute missing entries as $\tilde{C}_{it} = \tilde{\mu} + \tilde{\xi}_t + \tilde{\alpha}_i + \tilde{\Lambda}_i^\top \tilde{F}_t$

Asymptotic Results

Theorem: Consistency

Let $\delta_{N,T} = \min(N, T)$. Under general observation pattern and data-generating model ▸ assumptions.
Assume that $N, T \to \infty$ and one of the following cases holds:

1. The observation probability p_{it} is known
2. The observation probability p_{it} is unknown and satisfies $p_{it} = u_i v_t$. The time series observation pattern only has short-term conditional dependency
3. The observation probability p_{it} is unknown. The time series observation patterns is monotone, and in addition, $\mathbb{E}[\| N_{co}^{-1} \sum_{i \in \mathcal{N}_{co}} \Lambda_i \|^4] \leq M/\delta_{N,T}^2$

Then our proposed estimator with corresponding weights in $\tilde{\mu}$, $\tilde{\xi}_t$ and $\tilde{\alpha}_i$ is consistent for the common components of Y

$$\sqrt{\delta_{N,T}}\left((\tilde{\mu} + \tilde{\alpha}_i + \tilde{\xi}_t) - (\mu + \alpha_i + \xi_t)\right) = O_p(1)$$
$$\sqrt{\delta_{N,T}}\left(\tilde{C}_{it} - C_{it}\right) = O_p(1)$$

Asymptotic Results

Theorem: Asymptotic normality

Let $\delta_{N,T} = \min(N, T)$. Under general observation pattern and data-generating model ▸ assumptions.
Assume that $N, T \to \infty$ and one of the above cases holds. Under additional assumptions, our proposed estimator with proper weights is asymptotically normal

$$\sqrt{\delta_{N,T}} \cdot \sigma_{C,it}^{-1} \left(\tilde{C}_{it} - C_{it} \right) \xrightarrow{d} \mathcal{N}(0, 1)$$

where $\sigma_{C,it}^2 = \lim_{N,T\to\infty} \frac{\delta_{N,T}}{N} \sigma_{C,it,1}^2 + \frac{\delta_{N,T}}{T} \sigma_{C,it,2}^2$ with some $\sigma_{C,it,1}^2$ and $\sigma_{C,it,2}^2$

First-stage estimation error of fixed effects carries over to the second-stage estimation of factors

$\Rightarrow$ The form of asymptotic variance $\sigma_{C,it}^2$ is very complicated

$\Rightarrow$ We propose practical feasible estimator of asymptotic variance

Application to Causal Inference

Definition and Estimation of Treatment Effects

Treatment effect

- Individual treatment effect of unit i at time t

$$\tau_{it} = Y_{it}^{(\mathrm{tr})} - Y_{it}^{(\mathrm{ct})},$$

where $Y_{it}^{(\mathrm{tr})}$ and $Y_{it}^{(\mathrm{ct})}$ denote the treated and control outcomes of unit i at time t

- Our focus: Average treatment effect on a treated unit i over treated time periods (ATT)

$$\tau_i = \frac{1}{T_{i,tr}} \sum_{t \in \mathcal{T}_{i,tr}} \left(Y_{it}^{(\mathrm{tr})} - Y_{it}^{(\mathrm{ct})} \right) = \frac{1}{T_{i,tr}} \sum_{t \in \mathcal{T}_{i,tr}} \tau_{it},$$

where $\mathcal{T}_{i,tr}$ and $T_{i,tr}$ denote the set and number of treated time periods of unit i

Estimation of treatment effect

- Feasible estimator of τ_i imputes all values in $Y^{(\mathrm{ct})}$ with $\tilde{C}_{it}$ estimated with wi-PCA:

$$\hat{\tau}_i = \frac{1}{T_{i,tr}} \sum_{t \in \mathcal{T}_{i,tr}} \left(Y_{it}^{(\mathrm{tr})} - \tilde{C}_{it} \right)$$

⇒ Analogous estimation of average treatment effect over units or over both units and time

Feasible Variance Estimator for ATT

Example: Construct confidence intervals for the average treatment effect on the treated (ATT)

$$\left[\hat{\tau}_i - z_{1-\alpha/2}\sqrt{V_i}, \hat{\tau}_i + z_{1-\alpha/2}\sqrt{V_i}\,\right]$$

We propose to estimate variance V_i by resampling bootstrap

$\Rightarrow$ Simple to implement; good performance in large panels; accommodate general missing patterns

$\Rightarrow$ Estimating only the variance and leveraging the theoretical normal distribution have a better coverage than estimating the complete distribution ▸ simulations

Three-step procedure in resampling bootstrap

1. Construct the bootstrap sample
   - Sample $N-1$ units from all units besides i-th unit with replacement
   - Bootstrapped version of i: Estimated components of i plus a draw of $(\epsilon_{i1}, \cdots, \epsilon_{iT})$
2. Estimate τ_i on the bootstrap sample
3. Estimate V_i from the sample variance of bootstrapped estimates of τ_i

Empirical Results

Empirical Application – Legalization of Marijuana

The effect of legalization of recreational marijuana on per-capita beer sales (Li and Sonnier (2023))

Data: 208 weekly observations between 2017-2020 for 45 contiguous states in U.S.

- Weekly beer sales revenue from the Nielsen retail scanner data set from the Kilts Center for Marketing at Chicago Booth (factor analysis of the data in Guha and Ng (2019))
- Yearly state population from the U.S. Census Bureau

Treatment: 6 treated states with staggered adoption + 39 control states

Compare the estimation of our estimator, wi-PCA, with three methods

- TWFE: Two-way fixed effects estimator (special case without latent factor structure)
- PCA: PCA estimator in Xiong and Pelger (2022) (special case without fixed effects)
- Block-PCA: PCA estimator that estimates factors only from fully observed blocks in Xu (2017)

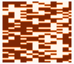
Extensive ▸ simulations demonstrate the superior performance of wi-PCA compared with three methods

Empirical Application – Legalization of Marijuana

Compare estimation accuracy of different estimators via synthetic treatments

- Uniformly random treatment
- Endogenous treatment assignment

Data: Control panel that consists of 208 weekly observations for 39 control states

		wi-PCA (FE+factor model)			PCA (factor model only)			Block-PCA (factor model only)			TWFE (FE only)
		$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	
	RMSE_Y	2.21	**1.93**	2.01	2.53	2.64	2.95	-	-	-	2.54
	$\lvert \mathrm{Bias}_{\mathrm{ATT}} \rvert$	0.16	**0.14**	0.16	0.22	0.59	0.98	-	-	-	0.22
	$\mathrm{RMSE}_{\mathrm{ATT}}$	1.24	**1.11**	1.18	1.68	1.76	2.03	-	-	-	1.57
	RMSE_Y	2.42	2.09	**1.94**	2.37	4.61	5.45	2.31	2.25	2.13	2.45
	$\lvert \mathrm{Bias}_{\mathrm{ATT}} \rvert$	**0.26**	0.27	0.26	0.83	3.76	4.19	0.45	0.32	0.30	0.42
	$\mathrm{RMSE}_{\mathrm{ATT}}$	**0.99**	1.03	1.06	1.67	4.08	4.52	1.57	1.13	1.18	1.53
	RMSE_Y	4.30	**3.92**	4.10	5.00	6.77	7.46	4.88	7.16	5.93	4.60
	$\lvert \mathrm{Bias}_{\mathrm{ATT}} \rvert$	0.75	**0.47**	0.57	1.45	4.47	5.25	0.90	1.26	1.04	0.59
	$\mathrm{RMSE}_{\mathrm{ATT}}$	2.89	**2.77**	2.91	4.12	5.43	6.06	3.88	5.11	4.21	3.50

$\Rightarrow$ Our estimator is robust to the number of latent factors (extends to more factors extensions)

$\Rightarrow$ Our estimator provides the most precise estimation for different cases

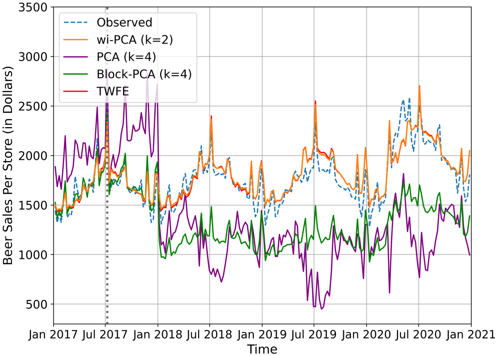
Empirical Application – Legalization of Marijuana

Estimation of the average treatment effects (ATT) for the 6 actually treated states

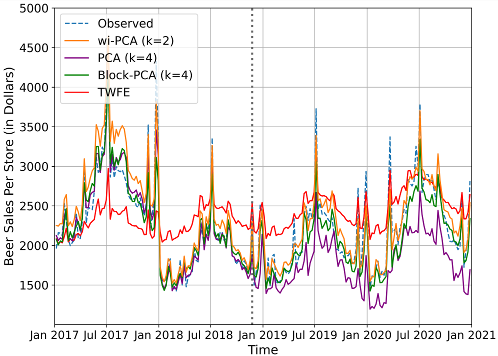
Treat time	k	California Jan. 1, 2018 ATT	SE	Michigan Dec. 1, 2019 ATT	SE	Illinois Jan. 1, 2020 ATT	SE	Massachusetts Nov. 20, 2018 ATT	SE	Nevada Jul. 1, 2017 ATT	SE	Maine Oct. 9, 2020 ATT	SE
wi-PCA (FE+factor model)	1	0.52	3.78	0.43	1.17	0.32	1.21	-0.48	1.11	-0.47	3.18	-1.15	1.11
	2	0.57	3.64	0.44	1.17	0.31	1.22	-0.37	1.07	-0.47	3.02	-0.95	1.18
	3	0.55	3.68	0.52	1.12	0.09	1.12	-0.42	1.28	-0.48	3.05	-0.96	1.47
	4	0.54	3.69	1.00	1.14	0.08	1.08	-0.86	1.54	-0.47	3.02	-0.16	0.60
PCA (factor model only)	1	2.96	3.85	3.19	1.64	1.20	1.64	-1.29	2.36	3.02	3.10	-1.32	1.34
	2	7.44	3.61	2.74	1.15	2.27	1.21	3.63	1.49	5.35	3.01	-1.06	1.11
	3	7.65	3.62	6.56	2.44	5.39	2.37	4.18	2.35	5.49	3.04	-0.36	1.13
	4	8.01	3.60	6.66	2.31	5.55	2.27	4.28	2.36	4.94	3.03	0.32	1.08
Block-PCA (factor model only)	1	-0.30	4.26	1.65	1.52	-0.05	1.53	-4.41	2.50	-1.24	3.44	-1.32	1.19
	2	4.25	6.65	0.38	1.46	0.33	1.31	1.16	1.41	7.79	8.32	-1.18	1.09
	3	0.18	6.14	0.32	1.41	0.36	1.29	1.36	1.74	2.84	6.37	-0.91	1.31
	4	1.69	6.92	0.30	1.44	0.33	1.41	1.26	1.83	4.80	6.68	0.31	1.15
TWFE (FE only)		0.56	3.73	1.73	1.64	0.02	1.62	-2.94	2.29	-0.47	3.04	-1.28	1.19

- Different estimators give substantially different point estimates of ATT
- Standard errors of the benchmarks are generally much larger than with wi-PCA
- wi-PCA is extremely stable for different number of latent factors (e.g., MA and NV)

Empirical Application – Legalization of Marijuana



(a) Nevada (NV)



(b) Massachusetts (MA)

- **NV**: PCA and block-PCA find a treatment effect due to a bad model fit
- **MA**: TWFE suffers from an omitted variable bias and neglects relevant time-series variation
- Note: Same degrees of freedom for wi-PCA, PCA, and block-PCA

Conclusion

wi-PCA: Novel method for causal inference in large panels

- Explicitly combines non-stationary two-way fixed effects and latent factor model and allows for endogenous treatment that depends on both of them
- Entrywise inferential theory for imputed values
- Feasible variance estimators for average treatment effects

⇒ Easy to implement and broadly applicable

wi-PCA: Superior simulation and empirical performance

- wi-PCA has highest out-of-sample accuracy compared to DID and PCA-only-methods
- PCA-only-methods: extremely unstable for different number of factors and large variance
- DID: omitted variable bias

⇒ More credible economic conclusions with wi-PCA

Appendix

Assumption: Observation Pattern of Control Panel Y

1. W is independent of F and ϵ (but can depend on α, ξ and Λ)
2. Overlap: the conditional observation probability $\mathbb{P}(W_{it} = 1 | I_{it}) \geq \eta > 0$
3. Sufficiently many observations: $\frac{1}{N}\sum_{i=1}^{N} W_{it} \geq q > 0$ and $\frac{1}{T}\sum_{t=1}^{T} W_{it} \geq q > 0$ for any i, t
4. Conditional independence of cross-sectional missingness: $W_{it} \perp\!\!\!\perp W_{js} | I_{it} \cup I_{js}$ for any $i \neq j$ and t, s

We present the assumptions of a simplified factor model with two-way fixed effects which captures the main insight of the general model

Assumption S1: Simplified Factor Model with Fixed Effects

There exists a constant $M < \infty$ such that

1. Fixed effects: $\alpha_i \overset{i.i.d.}{\sim} (0, \sigma_\alpha^2)$ and $\mathbb{E}[\alpha_i^4 | I_i] \leq M$. Furthermore, $\sum_{i=1}^N \alpha_i = 0$.
2. Factors: $F_t \overset{i.i.d.}{\sim} (0, \Sigma_F)$ and $\mathbb{E}\|F_t\|^8 \leq M$ for any t. Furthermore, $\sum_{t=1}^T F_t = 0$.
3. Loadings: $\Lambda_i \overset{i.i.d.}{\sim} (0, \Sigma_\Lambda)$ and $\mathbb{E}[\|\Lambda_i\|^8 | I_i] \leq M$ for any i. Furthermore, $\sum_{i=1}^N \Lambda_i = 0$, and for any t, $N^{-1} \sum_{i=1}^N W_{it} \Lambda_i \Lambda_i^\top \xrightarrow{p} \Sigma_{\Lambda,t}$ with some positive definite matrix $\Sigma_{\Lambda,t}$.
4. Idiosyncratic errors: $\epsilon_{it} \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$ and $\mathbb{E}[\epsilon_{it}^8] \leq M$.
5. Independence: α, ξ, F, Λ and ϵ are mutually independent.

Simulation Design

The data-generating process is

$$Y_{it} = \mu + \alpha_i + \xi_t + \Lambda_i F_t + \epsilon_{it},$$

where $\mu = 1$, $\alpha_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $F_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\Lambda_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0,4)$, stationary and non-stationary time fixed effects ξ_t

- Stationary time fixed effects: $\xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- Non-stationary time fixed effects: $\xi_t = 0.05t + \mathcal{N}(0,1)$ for any t (include a time trend)

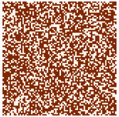
We compare the relative MSE ($\mathcal{S}$: set of either observed, missing or all entries in Y)

$$\text{relative MSE}_{\mathcal{S}} = \frac{\sum_{(i,t)\in\mathcal{S}} (\tilde{C}_{it} - C_{it})^2}{\sum_{(i,t)\in\mathcal{S}} C_{it}^2}$$

Three different observation patterns

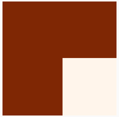
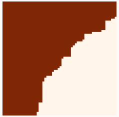
- Missing-at-random
- Simultaneous treatment adoption
- Staggered treatment adoption

Simulation Results

			wi-PCA (FE+factor model)		PCA (factor model only)			Block-PCA (factor model only)			TWFE (FE only)
	ξ_t	$\mathcal{S}$	known	unknown	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	
		obs	0.055	0.055	0.360	0.147	0.088	-	-	-	0.270
	S	**miss**	**0.056**	**0.056**	0.377	0.157	0.093	-	-	-	0.283
		all	0.055	0.055	0.364	0.149	0.089	-	-	-	0.273
		obs	0.036	0.036	0.292	0.145	0.059	-	-	-	0.177
	N	**miss**	**0.037**	**0.037**	0.305	0.157	0.063	-	-	-	0.186
		all	0.036	0.036	0.294	0.148	0.059	-	-	-	0.179

- wi-PCA is more efficient than PCA with $k = 3$
- PCA with $k < 3$ suffers from an omitted variable bias
- TWFE suffers from an an omitted variable bias

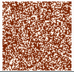
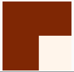
Simulation Results

	ξ_t	$\mathcal{S}$	wi-PCA (FE+factor model) known	wi-PCA (FE+factor model) unknown	PCA (factor model only) $k=1$	PCA $k=2$	PCA $k=3$	Block-PCA (factor model only) $k=1$	Block-PCA $k=2$	Block-PCA $k=3$	TWFE (FE only)
	S	obs	0.064	0.069	0.369	0.156	0.104	0.372	0.161	0.110	0.283
		miss	0.108	**0.105**	0.414	0.210	0.165	0.412	0.204	0.152	0.314
		all	0.077	0.079	0.383	0.172	0.122	0.384	0.174	0.122	0.292
	N	obs	0.048	0.051	0.376	0.209	0.091	0.359	0.186	0.080	0.207
		miss	0.055	**0.054**	0.594	0.573	0.527	0.476	0.322	0.154	0.162
		all	0.051	0.052	0.460	0.349	0.259	0.404	0.238	0.108	0.190
	S	obs	0.054	0.050	0.362	0.132	0.088	0.363	0.131	0.086	0.283
		miss	0.083	**0.079**	0.431	0.325	0.213	0.440	0.310	0.176	0.206
		all	0.058	0.055	0.371	0.159	0.106	0.373	0.157	0.100	0.271
	N	obs	0.056	0.045	0.301	0.178	0.078	0.287	0.126	0.083	0.216
		miss	0.097	**0.081**	0.911	0.759	0.602	1.135	1.065	0.494	0.175
		all	0.068	0.055	0.487	0.353	0.237	0.541	0.408	0.207	0.204

- wi-PCA is more efficient than block-PCA by using all the data
- PCA and block-PCA can suffer from the correlation between time trend and missing pattern
- TWFE suffers from an an omitted variable bias

Simulation: Comparison of Confidence Intervals by Different Methods

- Bootstrapped variance with normal distribution has a better coverage than bootstrapped distribution

		Bootstrapped Variance with Normal Distribution			Bootstrapped Distribution		
	ξ_t	95%	90%	80%	95%	90%	80%
	S	95.0%	89.8%	80.0%	93.4%	88.6%	78.8%
	N	95.0%	89.8%	80.0%	93.4%	88.6%	78.8%
	S	94.3%	89.2%	79.5%	92.4%	87.5%	80.4%
	N	94.3%	89.2%	79.5%	92.4%	87.5%	80.4%
	S	94.1%	89.2%	78.1%	92.6%	87.2%	76.4%
	N	94.1%	89.2%	78.1%	92.6%	87.2%	76.4%

Empirical Study: Synthetic Treatment Assignment with More Factors

Results for the uniformly random treatment pattern

		$RMSE_Y$	$\lvert Bias_{ATT} \rvert$	$RMSE_{ATT}$
wi-PCA (FE+factor model)	k=1	2.207	0.156	1.236
	k=2	1.933	0.139	1.110
	k=3	2.012	0.158	1.179
	k=4	2.033	0.167	1.206
	k=5	2.026	0.173	1.214
	k=6	2.051	0.177	1.258
	k=7	2.061	0.177	1.291
	k=8	2.067	0.177	1.303
PCA (factor model only)	k=1	2.525	0.216	1.677
	k=2	2.642	0.586	1.755
	k=3	2.952	0.979	2.032
	k=4	3.302	1.357	2.368
	k=5	3.598	1.637	2.615
	k=6	3.764	1.806	2.782
	k=7	3.883	1.929	2.908
	k=8	3.978	2.022	3.002
Block-PCA (factor model only)	k=1	-	-	-
	k=2	-	-	-
	k=3	-	-	-
	k=4	-	-	-
	k=5	-	-	-
	k=6	-	-	-
	k=7	-	-	-
	k=8	-	-	-
TWFE (FE only)		2.576	0.218	1.604

Empirical Study: Synthetic Treatment Assignment with More Factors

Results for the simultaneous treatment adoption pattern

		$RMSE_Y$	$\lvert Bias_{ATT} \rvert$	$RMSE_{ATT}$
wi-PCA (FE+factor model)	k=1	2.424	0.255	0.995
	k=2	2.090	0.275	1.029
	k=3	1.942	0.262	1.056
	k=4	1.909	0.284	1.084
	k=5	1.955	0.299	1.138
	k=6	2.006	0.299	1.166
	k=7	2.025	0.301	1.177
	k=8	2.039	0.304	1.184
PCA (factor model only)	k=1	2.374	0.825	1.672
	k=2	4.605	3.760	4.076
	k=3	5.450	4.190	4.523
	k=4	5.123	4.131	4.459
	k=5	5.038	4.138	4.467
	k=6	5.009	4.147	4.478
	k=7	5.005	4.152	4.483
	k=8	5.012	4.151	4.482
Block-PCA (factor model only)	k=1	2.307	0.450	1.568
	k=2	2.246	0.318	1.128
	k=3	2.125	0.304	1.176
	k=4	2.085	0.284	1.218
	k=5	2.037	0.289	1.228
	k=6	1.968	0.282	1.209
	k=7	1.940	0.272	1.208
	k=8	1.924	0.286	1.204
TWFE (FE only)		2.445	0.415	1.528

Empirical Study: Synthetic Treatment Assignment with More Factors

Results for the staggered treatment adoption pattern

		$RMSE_Y$	$\|Bias_{ATT}\|$	$RMSE_{ATT}$
wi-PCA (FE+factor model)	k=1	4.299	0.751	2.893
	k=2	3.923	0.475	2.770
	k=3	4.097	0.574	2.911
	k=4	4.185	0.642	2.985
	k=5	4.241	0.692	3.068
	k=6	4.270	0.733	3.102
	k=7	4.289	0.735	3.114
	k=8	4.278	0.726	3.096
PCA (factor model only)	k=1	5.004	1.445	4.119
	k=2	6.766	4.472	5.430
	k=3	7.463	5.246	6.063
	k=4	7.715	5.437	6.242
	k=5	7.794	5.558	6.369
	k=6	7.900	5.647	6.467
	k=7	7.985	5.719	6.539
	k=8	8.016	5.764	6.580
Block-PCA (factor model only)	k=1	4.877	0.902	3.884
	k=2	7.163	1.256	5.114
	k=3	5.934	1.035	4.212
	k=4	6.237	1.015	4.456
	k=5	6.243	1.009	4.434
	k=6	6.359	1.008	4.521
	k=7	6.617	1.039	4.731
	k=8	6.725	1.034	4.793
TWFE (FE only)		4.603	0.593	3.498