Dynamic Treatment Effect Estimation with Interactive Fixed Effects and Short Panels

Kyle Butts and Nicholas Brown

September 27, 2023

https://kylebutts.com/files/JMP.pdf
https://kylebutts.com/files/JMP_slides.pdf

Roadmap

Motivation

General Identification Result

Example Estimators and Empirical Application

Motivation

Treatment is often targeted to places/units based on their economic trends:

- Place-based policies (**neumark2015place**)
  - $\rightarrow$ Target places with declining labor markets
- New apartment construction (**asquith2021local**; **pennington2021does**)
  - $\rightarrow$ Built in appreciating neighborhoods
- Walmart entry (**basker2005job**; **neumark2008effects**)
  - $\rightarrow$ Open stores in areas with growing retail spending

Standard difference-in-differences assumption of parallel trends is *implausible*

Motivation

In some settings, the causes of these trends are due to larger economic forces and not location-specific shocks:

- Place-based policies
  - $\rightarrow$ Decline of manufacturing hurting manufacturing hubs
- New apartment construction
  - $\rightarrow$ Changing preferences for walkable neighborhoods
- Walmart entry
  - $\rightarrow$ Growing employment increases disposable income

Units have differential exposure to these macroeconomic trends in ways that are correlated with treatment

Factor Model

This paper models differential trends using a **factor model** generalizes the two-way fixed effect model:

- A set of macroeconomic time shocks that are common across units
- Units vary in how affected they are by the shocks

This paper

We propose a **class of imputation-style treatment effect estimators** under a **factor model**:

- Our ‘imputation’ style estimator explicitly estimates untreated potential outcome, $y_{it}(0)$, in the post-treatment periods (similar to synthetic control)
- Treatment can be targeted based on a unit/location’s exposure to shocks (violates standard parallel trends)
- Our estimator is valid in small-T settings

Current approaches

Covariates in two-way fixed effect model

If you could observe 'exposure' to some macroeconomic trend, you could include it in a two-way fixed effect model:

$$y_{it}(0) = \mu_i + \lambda_t + X_i \beta_t + u_{it}$$

This allows X_i-specific trends, e.g. manufacturing share is given by X_i and each period's 'shocks' estimated by β_t

Current approaches

Covariates in two-way fixed effect model

If you could observe 'exposure' to some macroeconomic trend, you could include it in a two-way fixed effect model:

$$y_{it}(0) = \mu_i + \lambda_t + X_i \beta_t + u_{it}$$

This allows X_i-specific trends, e.g. manufacturing share is given by X_i and each period's 'shocks' estimated by β_t

Issues:

- You must *observe* the underlying 'exposure' variables
- Noisy measures of X_i only partially control for the problem [**kejriwal2021efficacy**] ▸ Simulation Evidence

Current approaches

Synthetic Control

The synthetic control estimator constructs a ‘control unit’ that has the same exposure to the macroeconomic trends.

- Synthetic control is consistent when $y_{it}(0)$ has a factor model structure if you have a sufficiently large number of pre-periods

Current approaches

Synthetic Control

The synthetic control estimator constructs a ‘control unit’ that has the same exposure to the macroeconomic trends.

- Synthetic control is consistent when $y_{it}(0)$ has a factor model structure if you have a sufficiently large number of pre-periods

Issues:

- In short-panels, you *over-fit* on noise and get bad estimates [**abadie2010synthetic**; **ferman2021synthetic**]
- Even if you have a large number of pre-periods, structural changes to the economy can make far-away pre-periods uninformative (e.g. the 2008 recession) [**abadie2021using**]

Contribution

There are many estimators for treatment effects under factor models:

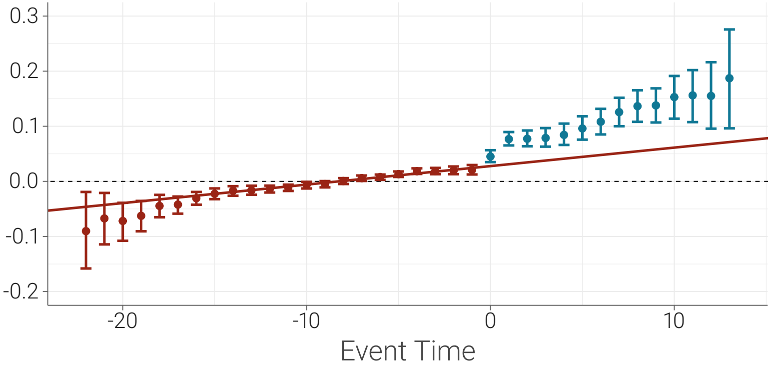
1. Synthetic control (**abadie2021using**)
2. Matrix Completion (**Athey˙et˙al˙2021**)
3. Imputation Estimators (**Gobillon˙Magnac˙2016**; **Xu˙2017**)

None of these are valid in short-T settings. Our paper introduces a general method that is valid in short-T settings.

- Unlocks a large econometric literature on factor model estimation and incorporates it into causal inference methods
  - → e.g. use some baseline covariates X_i as instruments. Need to be correlated with exposure to macroeconomic shocks

Preview of Application

Impact of new Walmart entry on $\log$ *retail employment. TWFE estimates*

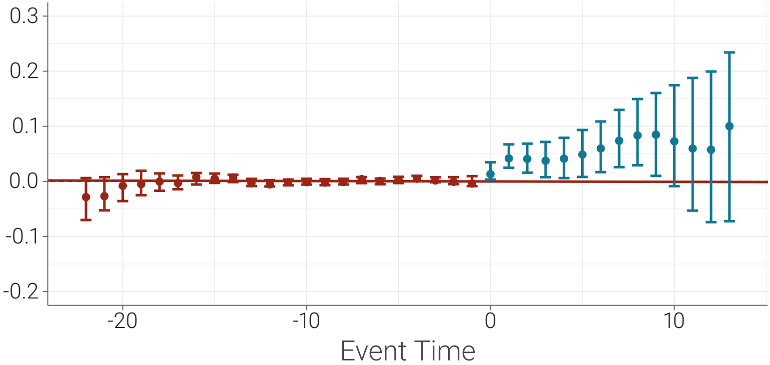


TWFE model expectedly is biased by non-parallel trends

(**neumark2008effects**; **basker2005job**)

Preview of Application

Factor model estimates



Factor model removed systematic trend in treated outcomes

Roadmap

Untreated Potential Outcomes

Factor Model

We observe a panel of observations denoted by unit $i \in \{1, \dots, N\}$ and by time period $t \in \{1, \dots, T\}$.

Untreated potential outcomes are given by a factor model:

$$y_{it}(0) = \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r} + u_{it} \tag{1}$$

- $f_{t,r}$ is the r-th factor (macroeconomic shock) at time t.
- $\gamma_{i,r}$ is unit i's factor loading (exposure) to the r-th factor.

Intuition of Factor Model

The intuition is very similar to that of a shift-share variable:

$$z_{it} = \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r}$$

- The $p \times 1$ vector $\boldsymbol{f}_t$ is the set of *'macroeconomic'* shocks (shifts) that all units experience
- $\boldsymbol{\gamma}_i$ is an individuals *exposure* (shares) to the shocks

The difference being that **we do not observe** the variables γ_i and f_t (like we don't observe fixed effects)

Two-way Fixed Effect vs. Factor Model

The factor model is a generalization of the TWFE model. If $\boldsymbol{f}_t = (\lambda_t, 1)'$ and $\boldsymbol{\gamma}_i = (1, \mu_i)'$, then (1) becomes:

$$y_{it}(0) = \lambda_t + \mu_i + u_{it}$$

Since TWFE is the work-horse model used by applied researchers, later we will explicitly add unit and time fixed-effects back in.

Treatment Effects

For now, assume there is a single treatment that turns on in some period $T_0 + 1$. Define D_i to be a dummy to denote which units receive treatment and d_{it} to equal 1 when treatment is active.

Treatment Effects

For now, assume there is a single treatment that turns on in some period $T_0 + 1$. Define D_i to be a dummy to denote which units receive treatment and d_{it} to equal 1 when treatment is active.

We are interested in event-study style average treatment effects. For each t, we define

$$\text{ATT}_t \equiv \mathbb{E}\left[y_{it}(1) - y_{it}(0) \mid D_i = 1\right],$$

where $y_{it}(0)$ is the (unobserved) untreated potential outcome.

Assumptions

'Non-Parallel Trends'

Assumption: Selection into Treatment

$$y_{it}(0) = \boldsymbol{f}_t' \boldsymbol{\gamma}_i + u_{it},$$

where for all t,

$$\mathbb{E}\left[u_{it} \mid \boldsymbol{\gamma}_i, D_i = 1\right] = \mathbb{E}\left[u_{it} \mid \boldsymbol{\gamma}_i, D_i = 0\right] = 0,$$

- Relaxes parallel trends by allowing units to enter treatment based on exposure to macroeconomic shocks
- Treatment can *not* be correlated with unit-time specific shocks u_{it}

Assumptions

Additional assumptions

Assumption: Arbitrary Treatment Effects

$$y_{it}(1) = y_{it}(0) + \tau_{it} \tag{2}$$

Assumption: No Anticipation

$$y_{it}(0) = y_{it} \text{ when } d_{it} = 0$$

ATT_t Identification

For a given t, the average outcome for the treated sample:

$$\begin{aligned} \text{ATT}_t &\equiv \mathbb{E}_i\left[y_{it}(1) \mid D_i = 1\right] - \mathbb{E}_i\left[y_{it}(0) \mid D_i = 1\right] \\ &= \mathbb{E}_i\left[y_{it}(1) \mid D_i = 1\right] - \boldsymbol{f}_t'\mathbb{E}_i\left[\boldsymbol{\gamma}_i \mid D_i = 1\right], \end{aligned}$$

where the equality comes from our selection into treatment assumption.

Insight: Estimating each $\boldsymbol{\gamma}_i$ requires large-T

- We only need to estimate $[\boldsymbol{\gamma}_i \mid D_i = 1]$ which is possible in small-T settings

ATT$_t$ Identification

Suppose we observed the $T \times p$ matrix of factors, $\boldsymbol{F}$. Let 'pre' denote the time periods before treatment $t \leq T_0$.

Then for $t > T_0$,

$$\mathbb{E}_i \left[y_{it} - \boldsymbol{f}_t' (\boldsymbol{F}_{\text{pre}}' \boldsymbol{F}_{\text{pre}})^{-1} \boldsymbol{F}_{\text{pre}}' \underbrace{\boldsymbol{y}_{i,\text{pre}}}_{\boldsymbol{F}_{\text{pre}} \boldsymbol{\gamma}_i + \boldsymbol{u}_{i,\text{pre}}} \mid D_i = 1 \right]$$

$$= \mathbb{E}_i \left[y_{it} - \boldsymbol{f}_t' \boldsymbol{\gamma}_i \mid D_i = 1 \right]$$

$$= \mathbb{E}_i \left[y_{it} - y_{it}(0) \mid D_i = 1 \right]$$

$$= ATT_t$$

ATT_t Identification

General Procedure

$$ATT_t = \mathbb{E}_i \left[y_{it} - \boldsymbol{f}_t' (\boldsymbol{F}_{\text{pre}}' \boldsymbol{F}_{\text{pre}})^{-1} \boldsymbol{F}_{\text{pre}}' \boldsymbol{y}_{i,\text{pre}} \mid D_i = 1 \right]$$

Consistency possible with $\sqrt{n}$-consistent estimation of the factors $\boldsymbol{F}$.

▸ Requirements for factor estimates, $\hat{\boldsymbol{F}}$

- This brings in a large literature on factor model estimation to causal-inference methods

  $\rightarrow$ Will illustrate multiple estimators of $\boldsymbol{F}$ in application.

ATT_t Identification

General Procedure

$$ATT_t = \mathbb{E}_i \left[y_{it} - \boldsymbol{f}_t' (\boldsymbol{F}_{\text{pre}}' \boldsymbol{F}_{\text{pre}})^{-1} \boldsymbol{F}_{\text{pre}}' \boldsymbol{y}_{i,\text{pre}} \mid D_i = 1 \right]$$

Consistency possible with $\sqrt{n}$-consistent estimation of the factors $\boldsymbol{F}$.

▸ Requirements for factor estimates, $\hat{\boldsymbol{F}}$

- This brings in a large literature on factor model estimation to causal-inference methods
  - $\rightarrow$ Will illustrate multiple estimators of $\boldsymbol{F}$ in application.
- Use only untreated observations, $d_{it} = 0$, for estimation of $\boldsymbol{F}$ to avoid bias.
- Staggered treatment 'imputes' $y_{it}(0)$ seperately for each treatment-timing group (changing pre)

Removing additive effects

Now, we extend our base model to include additive effects

$$y_{it} = \mu_i + \lambda_t + \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r} + u_{it}$$

We within-transform the outcome to remove the fixed effects:

$$\tilde{y}_{it} = y_{it} - \overline{y}_{0,t} - \overline{y}_{i,pre} + \overline{y}_{0,pre}$$

- $\overline{y}_{0,t}$: never-treated cross-sectional averages.
- $\overline{y}_{i,pre}$: pre-treated time averages.
- $\overline{y}_{0,pre}$: overall never-treated pre-treated average.

▸ Test for TWFE Model Sufficiency

Removing additive effects

$$\tilde{y}_{it} = y_{it} - \overline{y}_{0,t} - \overline{y}_{i,pre} + \overline{y}_{0,pre}$$

After performing our transformation, we have:

$$\mathbb{E}\left[\tilde{y}_{it} \mid D_i = 1\right] = \mathbb{E}\left[d_{it}\tau_{it} + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\gamma}}_i \mid D_i = 1\right]$$

where $\tilde{\boldsymbol{f}}_t$ are the pre-treatment demeaned factors and $\tilde{\boldsymbol{\gamma}}_i$ are the never-treated demeaned loadings.

- **Novel result:** Our transformation removes (μ_i, λ_t) but preserves a common factor structure $\implies$ our imputation argument holds on transformed outcomes.

Roadmap

Motivation

General Identification Result

Example Estimators and Empirical Application

Empirical Setting

We reevaluate the affect of Walmart openings on local labor markets.

Mixed results in empirical literature [**basker2005job**; **neumark2008effects**; **basker2007good**; **volpe2022economic**].

Empirical Setting

We reevaluate the affect of Walmart openings on local labor markets.

Mixed results in empirical literature [**basker2005job**; **neumark2008effects**; **basker2007good**; **volpe2022economic**].

Walmart opens stores based on local economic trajectories.

- Plausibly, Walmart is not targeting a specific location based on local shocks, i.e. based on u_{it}.
- Identification is based on assumption that Walmart picks places with growing retail sector due to national economic conditions, i.e. based on $\boldsymbol{f}_t \boldsymbol{\gamma}_i$.

Data

We construct a dataset following the description in **basker2005job**.

- In particular, we use the County Business Patterns dataset from 1964 and 1977-1999
- Subset to counties that (i) had more than 1500 employees overall in 1964 and (ii) had non-negative aggregate employment growth between 1964 and 1977

Data

We construct a dataset following the description in **basker2005job**.

- In particular, we use the County Business Patterns dataset from 1964 and 1977-1999
- Subset to counties that (i) had more than 1500 employees overall in 1964 and (ii) had non-negative aggregate employment growth between 1964 and 1977

We use a geocoded dataset of Walmart openings from **arcidiacono2020competitive**

- Treatment dummy is equal to one if the county has any Walmart in that year and our group variable denotes the year of entrance for the *first* Walmart in the county.

Initial Estimates

To show problems with selection, we estimate a TWFE imputation model [**Wooldridge˙2021**; **Borusyak˙Jaravel˙Spiess˙2021**]:

$$\log(y_{it}) = \mu_i + \lambda_t + \sum_{\ell=-22}^{13} \tau^\ell d_{it}^\ell + e_{it}, \tag{3}$$

- y_{it} include county-level retail employment and wholesale employment.
- d_{it}^ℓ are event-time dummies for being ℓ years away from when the initial Walmart opens in a county.

Figure: Effect of Walmart on County $\log$ Retail Employment (TWFE Estimate)

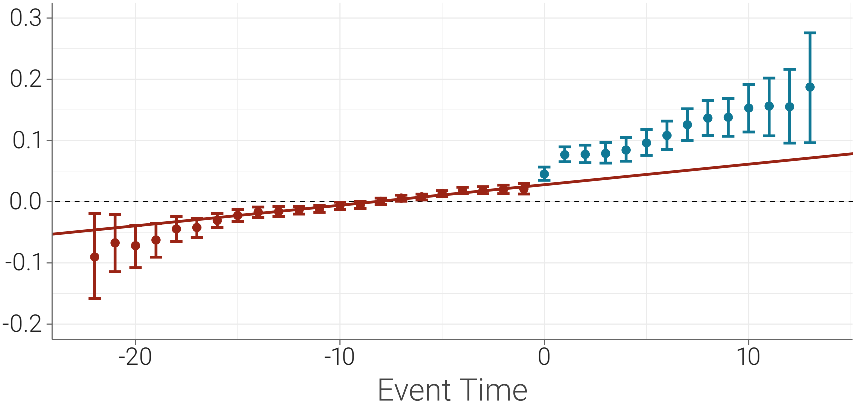
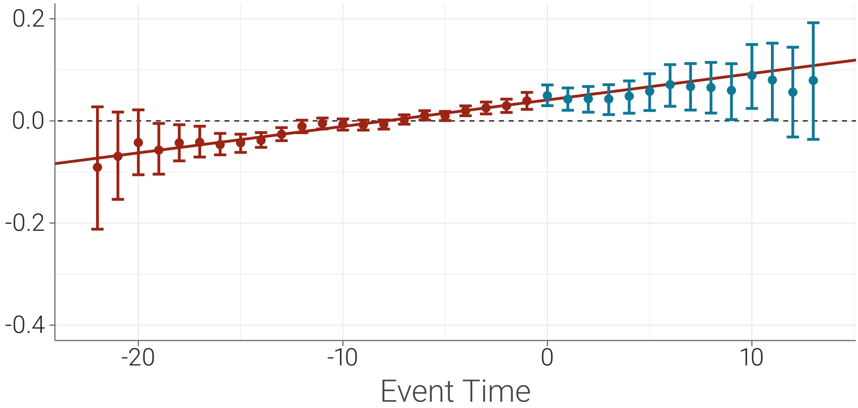


Figure: Effect of Walmart on County $\log$ Wholesale Employment (TWFE Estimate)



Factor Identification

Strategy 1: IV strategy

We consider instrumental-variables based identification strategy as proposed in **Ahn˙Lee˙Schmidt˙2013**.

- Allows fixed-T identification of $\boldsymbol{F}$.
- A GMM estimator $\implies$ inference is standard

▸ Quasi-long Differencing Details

Factor Identification

Strategy 1: IV strategy

Intuitively, we need a set of instruments that we think:

- (Relevancy) Are correlated with the factor-loadings γ_i.
- (Exclusion) Satisfy an exclusion restriction on u_{it}. We can't pick up on (i, t) shocks that are correlated with treatment

We think the best IV strategy entails using time-invariant characteristics X_i that we think are correlated with γ_i as instruments

Factor Model

Turning to our factor model estimator, we use the following variables at their 1980 baseline values as instruments:

- share of population employed in manufacturing
- shares of population below and above the poverty line
- shares of population employed in the private-sector and by the government
- shares of population with high-school and college degrees

Factor Model

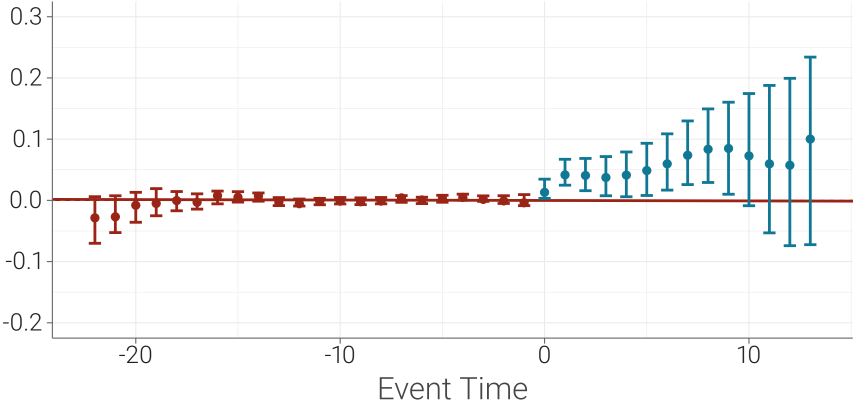
Turning to our factor model estimator, we use the following variables at their 1980 baseline values as instruments:

- share of population employed in manufacturing
- shares of population below and above the poverty line
- shares of population employed in the private-sector and by the government
- shares of population with high-school and college degrees

Think that these are predictive of the kinds of economic trends Walmart may be targeting

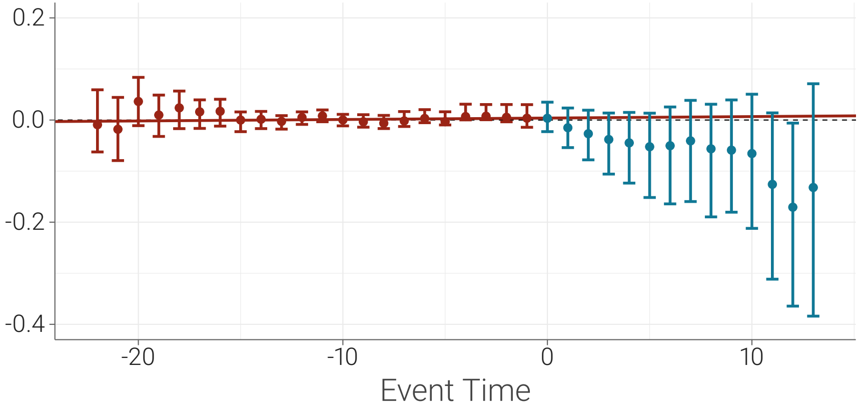
- Using baseline values helps us avoid picking up on concurrent shocks that are correlated with walmart opening

Figure: Effect of Walmart on County $\log$ Retail Employment (Factor Model)



Increase of retail employment of around 5%, consistent with estimates in **basker2005job**

Figure: Effect of Walmart on County $\log$ Wholesale Employment (Factor Model)



Consistent with estimates in **basker2005job**.

Alternative identification strategies

Strategy 2: Principal Components

An alternative identification strategy is a principal component decomposition of outcomes.

- This method requires no additional variables
- Requires either a large number of pre-periods [**Xu˙2017**] or error term u_{it} to be uncorrelated [**Imbens˙Kallus˙Mao˙2021**]

▸ Principal Components Details

Alternative identification strategies

Strategy 3: Common Correlated Effects

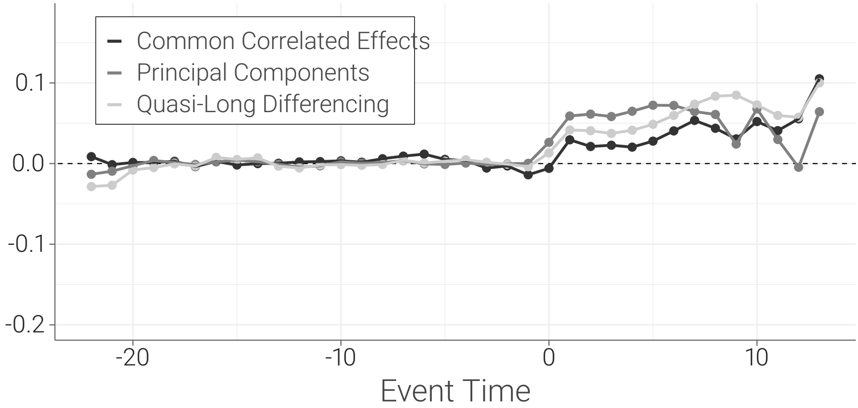
The common correlated effects estimate is based on the availability of a set of additional covariates $\boldsymbol{x}_{it}$ that are affected by the same factors as y_{it} [**Pesaran˙2006**]

- Cross-sectional averages of $\boldsymbol{x}_{it}$ across never-treated i become proxies $\hat{\boldsymbol{F}}_t$. Need $\geq p$ covariates.

In our Walmart setting, we use the log employment for the manufacturing, construction, agriculture, and healthcare 2-digit NAICS codes for $\boldsymbol{x}_{it}$

▸ Common Correlated Effects Details

Figure: Effect of Walmart on County $\log$ Retail Employment (Factor Model)



Consistent with estimates in **basker2005job**.

Conclusion

- Present a fixed-T imputation procedure to identify treatment effects under a factor-model
- Allows for differential trends between treated and control groups based on differential exposure to macroeconomic trends
- Proposed instrument-based identification of factors by using baseline characteristics that correlate with the factor-loadings

Removing additive effects

We consider the residuals after within-transforming

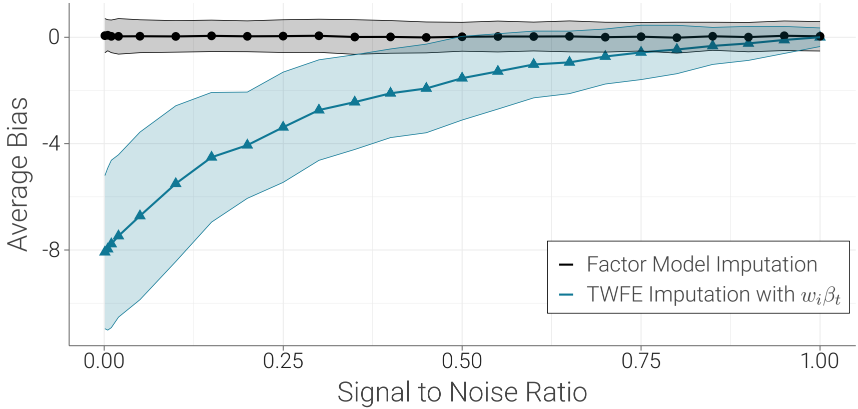
$$\tilde{y}_{it} = y_{it} - \overline{y}_{0,t} - \overline{y}_{i,pre} + \overline{y}_{0,pre},$$

$$\overline{y}_{i,pre} = \frac{1}{T_0} \sum_{t=1}^{T_0} y_{it}$$

$$\overline{y}_{0,t} = \frac{1}{N_0} \sum_{i=1}^{N} (1 - D_i) y_{it}$$

$$\overline{y}_{0,pre} = \frac{1}{T_0} \sum_{t=1}^{T_0} y_{0,t}$$

Figure: TWFE model with noisy proxy variable $w_i = X_i + v_i$



Back

Test for TWFE Model

If $\mathbb{E}\left[\boldsymbol{\gamma}_i \mid D_i\right] = \mathbb{E}\left[\boldsymbol{\gamma}_i\right]$, the ATTs are identified by the modified TWFE transformation.

$$\mathbb{E}\left[\tilde{y}_{it} \mid D_i = 1\right] = \mathbb{E}\left[\tau_{it} \mid D_i = 1\right] = \tau_t \tag{4}$$

for $t > T_0$.

Test for TWFE Model

If $\mathbb{E}[\boldsymbol{\gamma}_i \mid D_i] = \mathbb{E}[\boldsymbol{\gamma}_i]$, the ATTs are identified by the modified TWFE transformation.

$$\mathbb{E}[\tilde{y}_{it} \mid D_i = 1] = \mathbb{E}[\tau_{it} \mid D_i = 1] = \tau_t \tag{4}$$

for $t > T_0$.

- Says TWFE is sufficient even if there are factors, so long as exposure to these factors are the same between treated and control group.

In the paper, we provide tests for (4)

Back

Factor Identification

We cannot identify $\boldsymbol{F}$ or $\boldsymbol{\gamma}_i$ seperately from one another.

- Rotation problem means we can only identify $A\boldsymbol{F}$ for some matrix A.

Consider some estimator $\boldsymbol{F}(\theta)$ such that the true factor matrix $\boldsymbol{F} \in \text{col}(\boldsymbol{F}(\theta))$

- **Examples:** common correlated effects, principal components, quasi-differencing.

Then using $\boldsymbol{F}(\theta)$ in place of $\boldsymbol{F}$ still identifies ATT_t.

- $\boldsymbol{f}_t(\boldsymbol{F}_{\text{pre}}'\boldsymbol{F}_{\text{pre}})^{-1}\boldsymbol{F}_{\text{pre}}$ is invariant to rotating by any matrix A.

Back

Quasi-long Differencing Details

The quasi-long differencing method **Ahn˙Lee˙Schmidt˙2013** normalize the factors:

$$\boldsymbol{F}(\boldsymbol{\theta}) = \begin{pmatrix} -\boldsymbol{I}_p \\ \boldsymbol{\Theta} \end{pmatrix}$$

- Recall, this normalization does not impact imputation.

Quasi-differencing transformation: $\boldsymbol{H}(\boldsymbol{\theta}) = [\boldsymbol{\Theta}, \boldsymbol{I}_{(T-p)}]$. For all θ, we have

$$\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{F}(\boldsymbol{\theta}) = 0$$

▸ Back

Factor Identification

This transformation creates a set of moments:

$$\mathbb{E}\left[\boldsymbol{W}_i'\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{y}_i \mid D_i = 0\right]$$

- W_i is a $(T-p) \times w$ matrix of instruments.
- W_i must be exogenous after removing factors.
- $\hat{\boldsymbol{\theta}}$ is Fixed-T consistent.

Back

Principal Components Details

The principal component analysis takes the $T \times T$ matrix:

$$\mathbb{E}_i \left[\boldsymbol{y}_i \boldsymbol{y}_i' \mid D_i = 0 \right]$$

- Use the never-treated sample to estimate the covariance matrix.

The first p eigenvectors of the PC-decomposition will serve as the estimate of $\boldsymbol{F}$.

- Consistently spans the column space of $\boldsymbol{F}$ if $t \to \infty$ or if error term u_{it} is independent within i.

Common Correlated Effects Details

The common-correlated effects model assumes there are $K \geq p$ covariates that each take the form of:

$$x_{k,it} = \sum_{r=1}^{p} \xi_{i,r}^{k} f_{t,r} + \nu_{it}^{k},$$

where $f_{t,r}$ are the same factor shocks as the original outcome model.

The factor proxies $\boldsymbol{F}_t$ are formed as cross-sectional averages of x for the never-treated sample:

$$\hat{\boldsymbol{F}}_t' = \left(\mathbb{E}_i \left[x_{1,it} \mid D_i = 0\right], \ldots, \mathbb{E}_i \left[x_{K,it} \mid D_i = 0\right]\right)$$