Generalized Imputation Estimators for Factor Models

Nicholas Brown and Kyle Butts

April 5, 2023

Introduction

We are interested in effects of an intervention.

Notation:

Observed outcomes have two potential states:

- Treated potential outcomes $y_{it}(1)$.
- Untreated potential outcomes $y_{it}(0)$.

The **treatment effect**, at time t for unit i is

$$\tau_{it} = y_{it}(1) - y_{it}(0), \tag{1}$$

where $y_{it}(0)$ is unobserved.

TWFE and Parallel Trends

In panel settings, researchers often assume that a parallel-trends type assumption holds for outcomes

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

with $\mathbb{E}\left[u_{it} \mid D_i = 1\right] = \mathbb{E}\left[u_{it} \mid D_i = 0\right] = 0$ for all t.

Under this assumption, treated and control units are (on average) on the same trend: λ_t

TWFE and Parallel Trends

In panel settings, researchers often assume that a parallel-trends type assumption holds for outcomes

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

with $\mathbb{E}\left[u_{it} \mid D_i = 1\right] = \mathbb{E}\left[u_{it} \mid D_i = 0\right] = 0$ for all t.

Under this assumption, treated and control units are (on average) on the same trend: λ_t

This is often undesirable, units select into treatment based on their economic trends all the time!

Example:

Walmart choses where to open new stores in the 90s.

- Interested on the labor market impacts

Example:

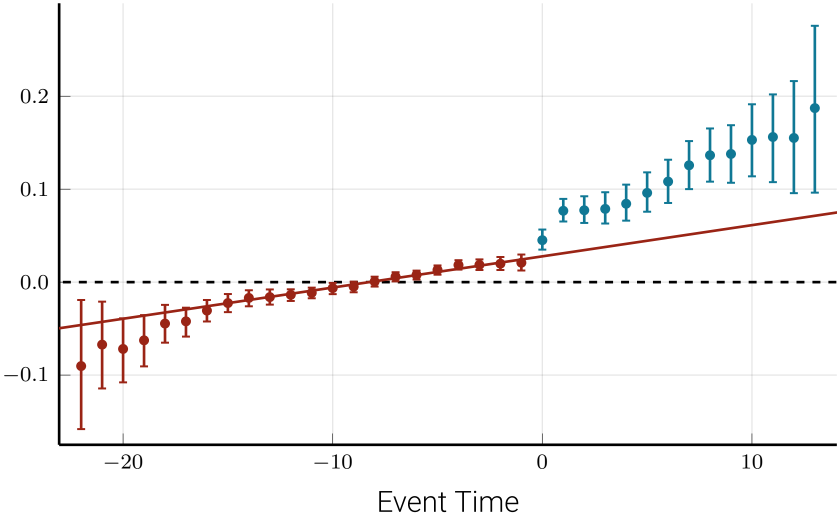
Walmart choses where to open new stores in the 90s.

- Interested on the labor market impacts

Untreated Model 1: $\text{employment}_{it}(0) = \text{macro}_t + \text{county}_i + u_{it}$.

- We would need to assume that treated counties and control counties are equally exposed to macroeconomic trends

Figure: TWFE Estimatied Effects of Walmart Entry on $\log$ Employment



Example:

Walmart choses where to open new stores in the 90s.

- Interested on the labor market impacts

Untreated Model 2: $\text{employment}_{it}(0) = \text{macro}_t * \text{county}_i + u_{it}$.

- Under this model, we can allow treated counties to have differential exposure to the macro shocks (e.g. manufacturing share)

Intuition of Factor Model

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

$$z_{it} = \gamma_i f_t$$

- f_t is the set of *'macroeconomic'* shocks (shifts) that all units experience
- γ_i is an individuals *exposure* to the shocks (shares)

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

Roadmap

Theory

Empirical Example

Model

N individuals observed for T times periods.

- Treatment begins **after** period T_0 (ignoring staggered treatment timing for this presentation).
- N_1 treated individuals, N_0 untreated individuals.

Model

N individuals observed for T times periods.

- Treatment begins **after** period T_0 (ignoring staggered treatment timing for this presentation).
- N_1 treated individuals, N_0 untreated individuals.

Untreated potential outcomes are given by a **factor model**:

$$y_{it}(0) = \mu_i + \lambda_t + \boldsymbol{f}_t' \boldsymbol{\gamma}_i + u_{it} \tag{2}$$

- $\boldsymbol{f}_t$: $p \times 1$ vector of unobserved common factors.
- $\boldsymbol{\gamma}_i$: $p \times 1$ vector of unobserved individual factor loadings.
- Nests the common TWFE model ($\boldsymbol{\gamma}_i = 0$).

Assumptions

Assumption: Arbitrary Treatment Effects

$$y_{it}(1) = y_{it}(0) + \tau_{it} \tag{3}$$

Assumptions

Assumption: Arbitrary Treatment Effects

$$y_{it}(1) = y_{it}(0) + \tau_{it} \tag{3}$$

Assumption: No Anticipation

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

- Treated units do not change their behavior before treatment.
- Can estimate and test for limited anticipation effects in our framework.

Assumptions

'Non-Parallel Trends'

Assumption: Selection into Treatment

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

with

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

- Relaxes parallel trends by allowing units to enter treatment based on exposure to macroeconomic shocks
- Does not let units enter into treatment based on unit specific shocks u_{it}

Selection into Treatment and Parallel Trends

In the two period case ($t = 1, 0$) consider the difference-in-differences estimand with parallel trends on the error term:

$$\mathbb{E}_i\left[y_{i1}(1) - y_{i0}(0) \mid D_i = 1\right] - \mathbb{E}_i\left[y_{i1}(1) - y_{i0}(0) \mid D_i = 0\right]$$

Selection into Treatment and Parallel Trends

In the two period case ($t = 1, 0$) consider the difference-in-differences estimand with parallel trends on the error term:

$$\begin{aligned} &\mathbb{E}_i\left[y_{i1}(1) - y_{i0}(0) \mid D_i = 1\right] - \mathbb{E}_i\left[y_{i1}(1) - y_{i0}(0) \mid D_i = 0\right] \\ &= \mathbb{E}_i\left[\tau_{i1} \mid D_i = 1\right] + \boldsymbol{f}_t'\big(\mathbb{E}_i\left[\boldsymbol{\gamma}_i \mid D_i = 1\right] - \mathbb{E}_i\left[\boldsymbol{\gamma}_i \mid D_i = 0\right]\big) \end{aligned}$$

This last term makes parallel trends not hold. That is, differential exposure to macroeconomic shocks violates parallel trends!

Identification under a factor model

There are many estimators for treatment effects under factor models:

1. Synthetic control (Abadie, 2021)
2. Matrix Completion (Athey et al., 2021)
3. Imputation Estimators (Gobillon and 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

ATT Identification

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

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

$$\mathbb{E}_i\left[y_{it}(0) \mid D_i = 1\right] = \lambda_t + \mathbb{E}_i\left[\mu_i \mid D_i = 1\right] + \boldsymbol{f}_t'\mathbb{E}_i\left[\boldsymbol{\gamma}_i \mid D_i = 1\right] \quad (4)$$

- **Insight:** Do not need to know each $\boldsymbol{\gamma}_i$ which would require large-T
  - $\rightarrow$ Only need to estimate $[\boldsymbol{\gamma}_i \mid D_i = 1]$.

ATT Identification

For now, ignore the additive fixed effects, so that

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

- Later, we remove the fixed effects with a within-transformation on y

ATT Identification

Let ‘pre’ denote the time periods before treatment $t \leq T_0$. If we observed the factors, $\boldsymbol{F}$, then for $t > T_0$,

$$\text{ATT}_t = \mathbb{E}_i \left[y_{it} - f_t(\boldsymbol{F}'_{\text{pre}} \boldsymbol{F}_{\text{pre}})^{-1} \boldsymbol{F}'_{\text{pre}} \boldsymbol{y}_{i,\text{pre}}) \mid D_i = 1 \right] \tag{5}$$

ATT Identification

Let 'pre' denote the time periods before treatment $t \leq T_0$. If we observed the factors, $\boldsymbol{F}$, then for $t > T_0$,

$$\text{ATT}_t = \mathbb{E}_i \left[y_{it} - f_t(\boldsymbol{F}'_{\text{pre}} \boldsymbol{F}_{\text{pre}})^{-1} \boldsymbol{F}'_{\text{pre}} \boldsymbol{y}_{i,\text{pre}}) \mid D_i = 1 \right] \tag{5}$$

What's happening here:

$$f_t \underbrace{\left(\boldsymbol{F}'_{\text{pre}} \boldsymbol{F}_{\text{pre}} \right)^{-1} \boldsymbol{F}'_{\text{pre}} \boldsymbol{y}_{i,\text{pre}} \right)}_{\to^p \, \mathbb{E}_i[\boldsymbol{\gamma}_i \,\mid\, D_i=1]} \to^p$$

ATT Identification

Let 'pre' denote the time periods before treatment $t \leq T_0$. If we observed the factors, $\boldsymbol{F}$, then for $t > T_0$,

$$\text{ATT}_t = \mathbb{E}_i \left[y_{it} - f_t (\boldsymbol{F}_{\text{pre}}' \boldsymbol{F}_{\text{pre}})^{-1} \boldsymbol{F}_{\text{pre}}' \boldsymbol{y}_{i,\text{pre}}) \mid D_i = 1 \right] \tag{5}$$

What's happening here:

$$f_t \underbrace{\left(\boldsymbol{F}_{\text{pre}}' \boldsymbol{F}_{\text{pre}} \right)^{-1} \boldsymbol{F}_{\text{pre}}' \boldsymbol{y}_{i,\text{pre}} \right)}_{\to^p \, \mathbb{E}_i[\boldsymbol{\gamma}_i \mid D_i = 1]} \to^p$$

This is a general imputation procedure that only requires $\sqrt{n}$-consistent estimation of the factors $\boldsymbol{F}$. This brings in a large literature on factor model estimation to causal-inference methods.

Removing additive effects

We consider the residuals after within-transforming

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

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

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:

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

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

- Our transformation removes (μ_i, λ_t) but preserves a common factor structure.

When is TWFE sufficient?

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

$$\mathbb{E}[\tilde{y}_{it} \mid D_i = 1] = \mathbb{E}[\tau_{it} \mid D_i = 1] = \tau_t \tag{6}$$

for $t > T_0$.

When is TWFE sufficient?

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

$$\mathbb{E}[\tilde{y}_{it} \mid D_i = 1] = \mathbb{E}[\tau_{it} \mid D_i = 1] = \tau_t \tag{6}$$

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 this condition.

Factor Identification

We consider instrumental-variables based identification of Ahn, Lee, and Schmidt (2013).

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

Factor Identification

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

1. Are correlated with the factor-loadings γ_i.
2. 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 would entail using time-invariant characteristics that we think are correlated with γ_i as instruments.

Roadmap

Theory

Empirical Example

Example

For example, consider our Walmart example. Since Walmart is likely targeting growing economies, we think that parallel trends would fail.

1. Plausibly, Walmart is not targetting a specific location based on local shocks, that is based on u_{it}.
2. More so, targetting places that are doing well due to national economic conditions, that is based on $\boldsymbol{f}_t \boldsymbol{\gamma}_i$.

Data

We construct a dataset following the description in Basker (2005).

- 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

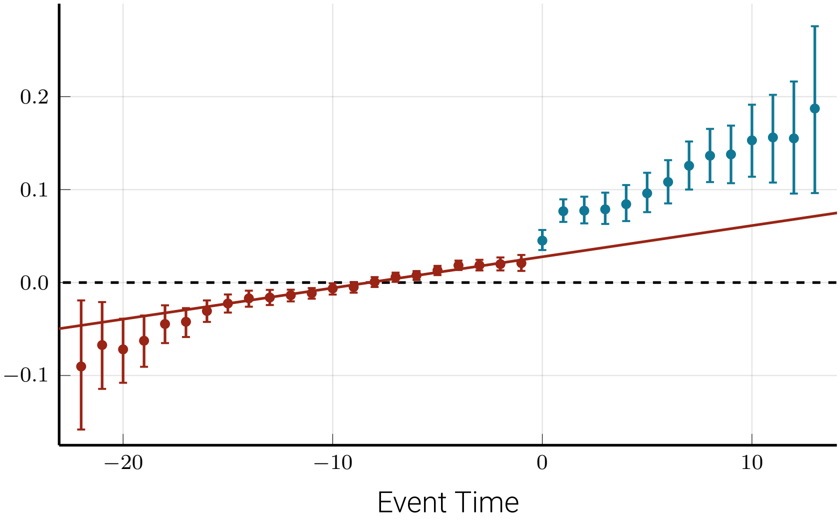
We construct a dataset following the description in Basker (2005).

- 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 Arcidiacono et al. (2020)

- 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.

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



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

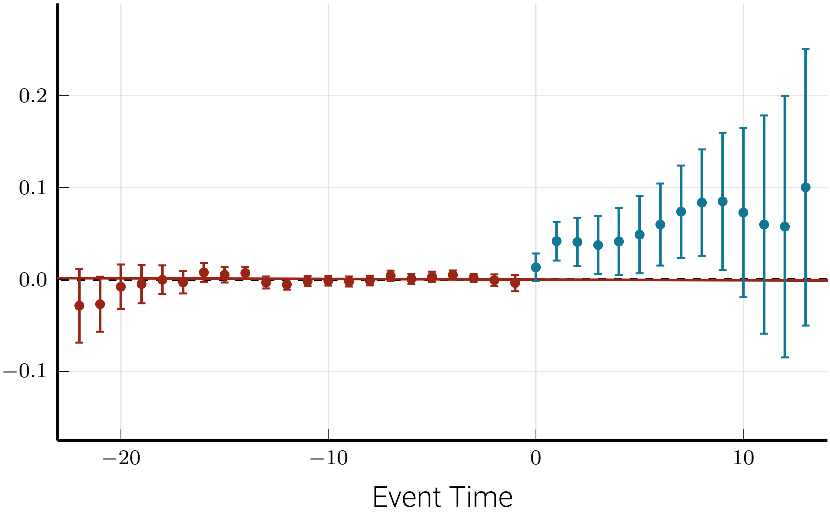
Turning to our factor model estimator, we use the following variables at their 1980 baseline values as instruments:

- 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)



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

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

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

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

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