Generalized Imputation Estimators for Factor Models

Nicholas Brown and Kyle Butts

February 3, 2023

Introduction

We are interested in effects of a treatment/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 is

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

Introduction

We aim to estimate average treatment effects on the treated, e.g. overall effect and event-study effects

- A particular form of **impulse response** with a binary treatment (Jordà, 2005)

Our goal is to 'impute' $y_{it}(0)$; that is, estimate what the treated units outcomes would have been without treatment.

TWFE and Parallel Trends

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

Under this model, we need a parallel trends assumption that the time shocks are experienced commonly between the control units and the treated units (θ_t is constant)

TWFE and Parallel Trends

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

Under this model, we need a parallel trends assumption that the time shocks are experienced commonly between the control units and the treated units (θ_t is constant)

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 on the same trend

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 on the same trend

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)

Empirical Macro

Large policy changes:

- Ramey (2016) gives examples of technological shocks, monetary policy shocks, and fiscal shocks
- e.g. Leblebicioğlu and Weinberger (2020) look at banking deregulation on the labor share using an event-study regression

Abnormal returns:

- Kothari and Warner (2007) review estimation difficulties
- Our method works in short-panels; helps with problems of structural breaks in long time-series
- Synthetic control fails even in reasonably long panels if there are highly autocorrelated errors

Empirical Macro

Marginal propensity to consume:

- Broda and Parker (2014) and Parker et al. (2013) estimate event-study analysis of 2008 stimulus payments.
- Estimation is complicated if payments are targeted based on economic status

Fiscal multipliers:

- Estimated via impulse response: Ilzetzki, Mendoza, and Végh (2013) and Brinca et al. (2016)
- Again, targeting of fiscal stimulus is non-random

Roadmap

Generalized Imputation Estimators

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**:

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

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

Assumptions

Arbitrary Treatment Effects:

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

Assumptions

Arbitrary Treatment Effects:

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

No Anticipation:

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

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

Assumptions

Selection into Treatment:

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

- **Mean Model:** $\mathbb{E}\left[y_{it}(0) \mid \mu_i, \boldsymbol{\gamma}_i, D_i\right] = \mu_i + \theta_t + \boldsymbol{f}_t' \mathbb{E}\left[\boldsymbol{\gamma}_i \mid D_i\right]$.
- Allows selection into treatment based on level and interactive fixed effects, but not idiosyncratic shocks

Selection into Treatment and Parallel Trends

In the two period case ($t = 1, 0$) consider the difference-in-differences estimand:

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

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

Selection into Treatment and Parallel Trends

In the two period case ($t = 1, 0$) consider the difference-in-differences estimand:

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

$$= \mathbb{E}\left[\tau_{it} \mid D_i = 1\right] + \boldsymbol{f}_t'\Big(\mathbb{E}\left[\boldsymbol{\gamma}_i \mid D_i = 1\right] - \mathbb{E}\left[\boldsymbol{\gamma}_i \mid D_i = 0\right]\Big)$$

Selection into Treatment and Parallel Trends

In the two period case ($t = 1, 0$) consider the difference-in-differences estimand:

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

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.

ATT Identification

Treated sample:

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

- **Insight:** Do not need to know $\boldsymbol{\gamma}_i$ which would require large-T

  $\rightarrow$ Only $\mathbb{E}\left[\boldsymbol{\gamma}_i \mid D_i = 1\right]$.

ATT Identification

Remove the additive fixed effects with a modified within-transformation:

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

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

$$\overline{y}_{0,t < T_0} = \frac{1}{N_0 T_0} \sum_{i=1}^{N} \sum_{t=1}^{T_0} (1 - D_i) y_{it}$$

- $\overline{y}_{0,t}$: never-treated cross-sectional averages.
- $\overline{y}_{i,t \leq T_0}$: pre-treated time averages.
- $\overline{y}_{0,t \leq T_0}$: overall never-treated pre-treated average.

Removing additive effects

We consider the residuals after within-transforming

$$\tilde{y}_{it} = y_{it} - \overline{y}_{0,t} - \overline{y}_{i,t\leq T_0} + \overline{y}_{0,t\leq T_0}$$

Removing additive effects

We consider the residuals after within-transforming

$$\tilde{y}_{it} = y_{it} - \overline{y}_{0,t} - \overline{y}_{i,t \leq T_0} + \overline{y}_{0,t \leq T_0}$$

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.

- **General result:** Our transformation removes (μ_i, θ_t) but preserves a common factor structure.

When is TWFE sufficient?

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.

$$\mathbb{E}\left[\tilde{y}_{it} \mid D_i = 1\right] = \mathbb{E}\left[\tau_{it} \mid D_i = 1\right] = \tau_t \tag{5}$$

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.

$$\mathbb{E}[\tilde{y}_{it} \mid D_i = 1] = \mathbb{E}[\tau_{it} \mid D_i = 1] = \tau_t \tag{5}$$

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.

ATT Identification

Okay, now to our cool result. If we know our factors, $\tilde{\boldsymbol{F}}$, Then for $t > T_0$,

$$\text{ATT}_t = \mathbb{E}\left[\tilde{y}_{it} - \tilde{\boldsymbol{f}}_t(\tilde{\boldsymbol{F}}'_{t\leq T_0}\tilde{\boldsymbol{F}}_{t\leq T_0})^{-1}\tilde{\boldsymbol{F}}'_{t\leq T_0}\boldsymbol{y}_{i,t\leq T_0}) \mid D_i = 1\right] \tag{6}$$

ATT Identification

Okay, now to our cool result. If we know our factors, $\tilde{\boldsymbol{F}}$, Then for $t > T_0$,

$$\text{ATT}_t = \mathbb{E}\left[\tilde{y}_{it} - \tilde{\boldsymbol{f}}_t(\tilde{\boldsymbol{F}}'_{t \leq T_0}\tilde{\boldsymbol{F}}_{t \leq T_0})^{-1}\tilde{\boldsymbol{F}}'_{t \leq T_0}\boldsymbol{y}_{i,t \leq T_0}) \mid D_i = 1\right] \tag{6}$$

What's happening here:

$$\tilde{\boldsymbol{f}}_t \underbrace{(\tilde{\boldsymbol{F}}'_{t \leq T_0}\tilde{\boldsymbol{F}}_{t \leq T_0})^{-1}\tilde{\boldsymbol{F}}'_{t \leq T_0}\boldsymbol{y}_{i,t \leq T_0})}_{\to^p \ \mathbb{E}[\boldsymbol{\gamma}_i \mid D_i = 1]}$$

ATT Identification

Okay, now to our cool result. If we know our factors, $\tilde{\boldsymbol{F}}$, Then for $t > T_0$,

$$\text{ATT}_t = \mathbb{E}\left[\tilde{y}_{it} - \tilde{\boldsymbol{f}}_t (\tilde{\boldsymbol{F}}'_{t \leq T_0} \tilde{\boldsymbol{F}}_{t \leq T_0})^{-1} \tilde{\boldsymbol{F}}'_{t \leq T_0} \boldsymbol{y}_{i,t \leq T_0}) \mid D_i = 1\right] \tag{6}$$

What's happening here:

$$\tilde{\boldsymbol{f}}_t \underbrace{(\tilde{\boldsymbol{F}}'_{t \leq T_0} \tilde{\boldsymbol{F}}_{t \leq T_0})^{-1} \tilde{\boldsymbol{F}}'_{t \leq T_0} \boldsymbol{y}_{i,t \leq T_0})}_{\to^p \ \mathbb{E}[\boldsymbol{\gamma}_i \mid D_i = 1]}$$

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

Factor Identification

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

- Allows fixed-T analysis.
- Provides moment conditions; inference is easy.

Factor Identification

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

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

Generalized Imputation Estimators

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).

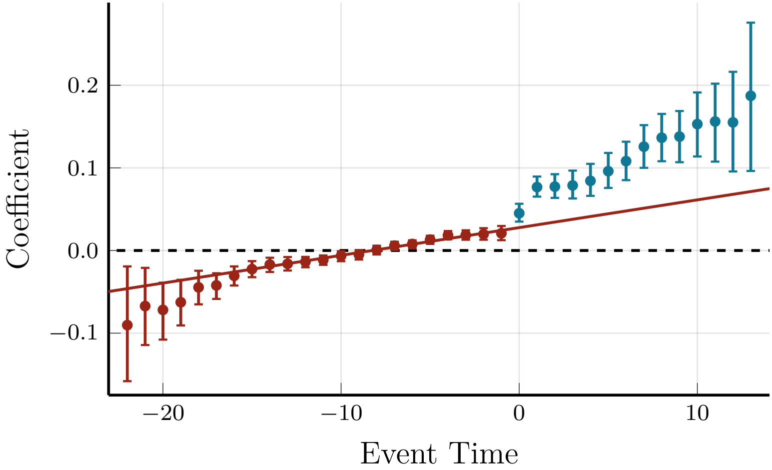
- In particular, we use the County Business Patterns dataset from 1964 and 1977-1999
- Subset to counties that (i) had more than 1500 employees overall in 1964 and (ii) had non-negative aggregate employment growth between 1964 and 1977

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.

TWFE Estimator

Figure: Effect of Walmart on County $\log$ Retail Employment



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:

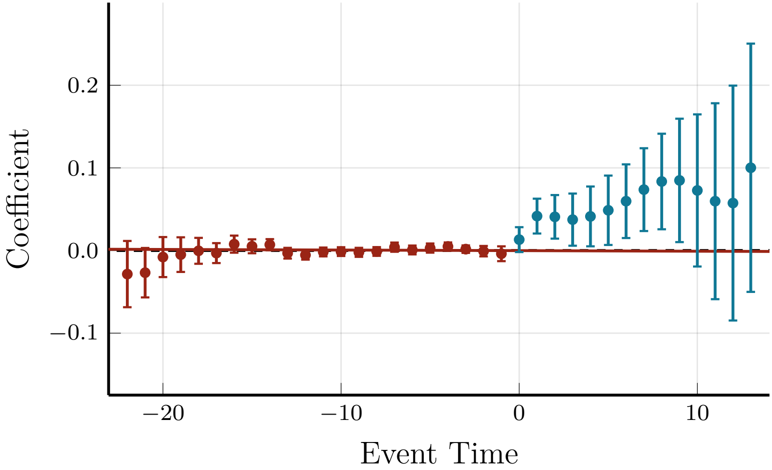
- share of population employed in manufacturing
- shares of population below and above the poverty line
- shares of population employed in the private-sector and by the government
- shares of population with high-school and college degrees

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

Generalized Imputation Estimator

Figure: Effect of Walmart on County log Retail Employment



Conclusion

- Present a fixed-T imputation procedure to identify treatment effects under a factor-model (nesting the TWFE model)
- Our estimator allows for valid identification of treatment effects when treatment is targeted based on a unit's economic trends.
- More work to be done on thinking through more explicit connections to VAR and local projection methods (Dube et al., 2022; Plagborg-Møller and Wolf, 2021)