Dynamic Treatment Effect Estimation with Interactive Fixed Effects and Short Panels

Kyle Butts* and Nicholas Brown†

*University of Colorado Boulder †Queen's University

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

Roadmap

Motivation

General Identification Result

Empirical Application

Motivation

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

- New apartment construction (Asquith, Mast, and Reed, 2021; Pennington, 2021)
  - $\rightarrow$ Built in appreciating neighborhoods

Motivation

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

- New apartment construction (Asquith, Mast, and Reed, 2021; Pennington, 2021)
  - $\rightarrow$ Built in appreciating neighborhoods
- Walmart entry (Basker, 2005; Neumark, Zhang, and Ciccarella, 2008)
  - $\rightarrow$ Open stores in areas with growing retail spending

Motivation

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

- New apartment construction (Asquith, Mast, and Reed, 2021; Pennington, 2021)
  - $\rightarrow$ Built in appreciating neighborhoods
- Walmart entry (Basker, 2005; Neumark, Zhang, and Ciccarella, 2008)
  - $\rightarrow$ Open stores in areas with growing retail spending
- Place-based policies (Neumark and Simpson, 2015)
  - $\rightarrow$ Target places with declining labor markets

Motivation

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

- New apartment construction (Asquith, Mast, and Reed, 2021; Pennington, 2021)
  - $\rightarrow$ Built in appreciating neighborhoods
- Walmart entry (Basker, 2005; Neumark, Zhang, and Ciccarella, 2008)
  - $\rightarrow$ Open stores in areas with growing retail spending
- Place-based policies (Neumark and Simpson, 2015)
  - $\rightarrow$ Target places with declining labor markets

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:

- New apartment construction
  - $\rightarrow$ Changing preferences for walkable neighborhoods

Motivation

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

- New apartment construction
  - $\rightarrow$ Changing preferences for walkable neighborhoods
- Walmart entry
  - $\rightarrow$ Growing employment increases disposable income

Motivation

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

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

Modeling differential trends

This paper models differential trends using a **factor model** that is popular in the finance/macroeconomics literature:

- Each time period there is a set of *unobservable* macroeconomic shocks that are common across units
- Units vary based on their *unobservable* baseline characteristics in how impacted 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 the untreated potential outcome, $y_{it}(0)$, in the post-treatment periods
  - $\rightarrow$ Same strategy as synthetic control and imputation estimator in difference-in-differences context (Borusyak, Jaravel, and Spiess, Forthcoming)

This paper

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

- Our ‘imputation’ style estimator explicitly estimates the untreated potential outcome, $y_{it}(0)$, in the post-treatment periods
  - $\rightarrow$ Same strategy as synthetic control and imputation estimator in difference-in-differences context (Borusyak, Jaravel, and Spiess, Forthcoming)
- A factor model allows treatment can be targeted based on a unit/location’s exposure to common shocks
  - $\rightarrow$ Violates standard parallel trends

This paper

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

- Our ‘imputation’ style estimator explicitly estimates the untreated potential outcome, $y_{it}(0)$, in the post-treatment periods
  - $\rightarrow$ Same strategy as synthetic control and imputation estimator in difference-in-differences context (Borusyak, Jaravel, and Spiess, Forthcoming)
- A factor model allows treatment can be targeted based on a unit/location’s exposure to common shocks
  - $\rightarrow$ Violates standard parallel trends
- Our estimator is valid in small-T settings and under treatment effect heterogeneity

Current approaches

Standard factor model estimators

There are many factor model estimators that jointly estimate coefficients on treatment dummies and the factor model

Current approaches

Standard factor model estimators

There are many factor model estimators that jointly estimate coefficients on treatment dummies and the factor model

Issues:

- Similar to the modern difference-in-differences literature (Borusyak, Jaravel, and Spiess, Forthcoming; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021), standard estimators would face problems with negative weighting under treatment effect heterogeneity

This paper

Our treatment effect estimators allows any $\sqrt{N}$-consistent estimate of the macroeconomic shocks to be used to generate consistent treatment effect estimates

- Unlocks a large econometric literature on factor model estimation and incorporates it into causal inference methods
  - $\rightarrow$ e.g. principal components, quasi-long differencing, common-correlated effects, etc.

Current approaches

Two-way Fixed Effect Imputation Estimator

Borusyak, Jaravel, and Spiess (Forthcoming) and Gardner (2021) propose an 'imputation' estimator for two-way fixed effects model:

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

1. Estimate fixed effects using untreated/not-yet-treated observations ($d_{it} = 0$). Predict $y_{it}(0)$ out-of-sample for treated observations
2. Average $y_{it} - \hat{y}_{it}(0)$ for treated observations

Current approaches

Two-way Fixed Effect Imputation Estimator

Borusyak, Jaravel, and Spiess (Forthcoming) and Gardner (2021) propose an 'imputation' estimator for two-way fixed effects model:

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

1. Estimate fixed effects using untreated/not-yet-treated observations ($d_{it} = 0$). Predict $y_{it}(0)$ out-of-sample for treated observations
2. Average $y_{it} - \hat{y}_{it}(0)$ for treated observations

This paper extends this approach for factor models

Current approaches

Covariates in two-way fixed effect model

Extend model with a set of *observable* characteristics to allow for X_i-specific trends

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

- E.g. Demand shocks estimated by β_t for neighborhood characteristics X_i

▸ Simulation Evidence

Current approaches

Covariates in two-way fixed effect model

Extend model with a set of *observable* characteristics to allow for X_i-specific trends

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

- E.g. Demand shocks estimated by β_t for neighborhood characteristics X_i

Issues:

- The researcher might not be able to *observe* the underlying 'exposure' variables
- Noisy measures of true 'exposure' only partially control for the problem (Kejriwal, Li, and Totty, 2021)

Current approaches

Synthetic Control

The synthetic control estimator constructs a ‘control unit’ that has the same exposure to the macroeconomic trends (a form of imputation)

- 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 (a form of imputation)

- 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 (Abadie, Diamond, and Hainmueller, 2010; Ferman and Pinto, 2021)
- 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) (Abadie, 2021)

Current approaches

New synthetic-control style estimators

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.

Current approaches

Alternative short-T factor model estimators

Freyaldenhoven, Hansen, and Shapiro (2019) propose a method that uses some variable x_{it} that is affected by the same confounder that affects y_{it} but not affected by treatment.

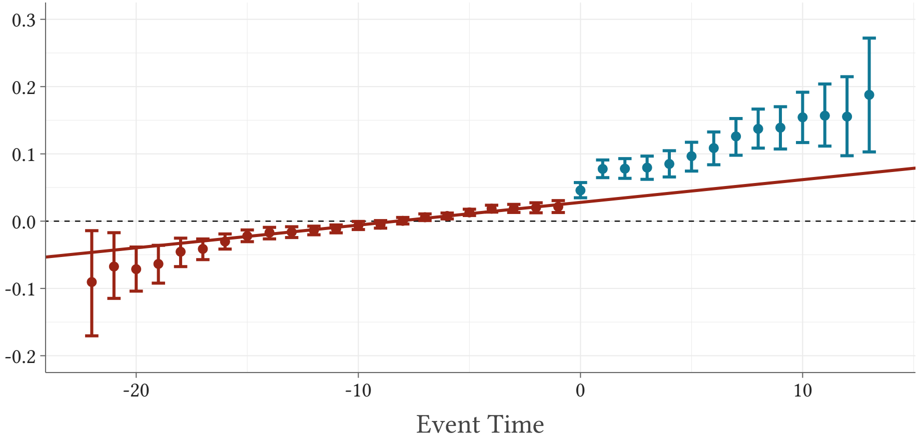
- This intuition is similar to a version of our estimator based on common-correlated effects (Brown, Butts, and Westerlund, 2023)
- The estimator is not heterogeneity-robust

Callaway and Karami (2023) propose a method for treatment effect estimation with a factor model

- The instrument they use would be valid for our quasi-long differencing estimator.

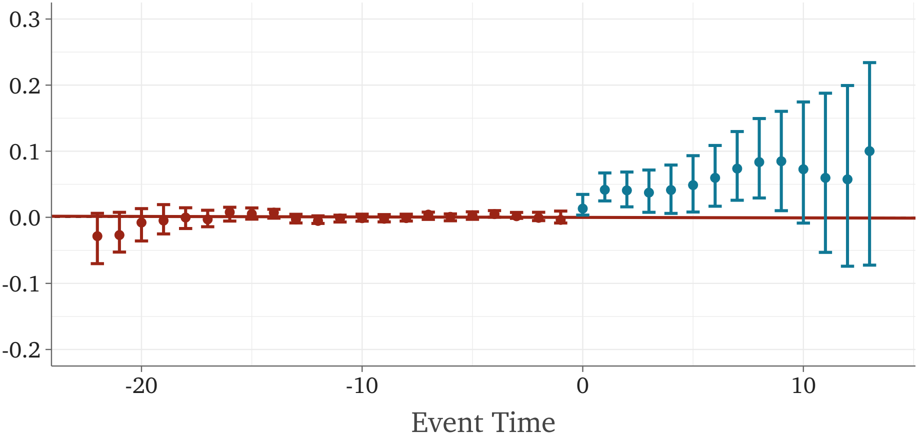
Preview of Application

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



Preview of Application

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



Roadmap

Motivation

General Identification Result

Empirical Application

Model

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

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.

▸ Selecting p ▸ Differences with Shift-Share IV

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 the TWFE model:

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

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.

- We assume $N_1 = \sum_i D_i$ and $N_0 = \sum_i D_i$ are non-vanishing as N grows.
- $T_0 \geq p$ for identification

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 assume $N_1 = \sum_i D_i$ and $N_0 = \sum_i D_i$ are non-vanishing as N grows.
- $T_0 \geq p$ for identification

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

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

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

Assumptions

Factor model

Assumption: Selection into Treatment

Untreated potential outcomes are given by

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

where $\mathbb{E}(u_{it} \mid \boldsymbol{\gamma}_i, D_i) = \mathbb{E}(u_{it} \mid \boldsymbol{\gamma}_i) = 0$ for all t.

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

Assumptions

Additional assumptions

Assumption: Arbitrary Treatment Effects

Treatment effects are left unrestricted (besides having finite moments):

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

Assumption: No Anticipation

Whenever $d_{it} = 0$, the observed $y_{it} = y_{it}(0)$.

ATT_t Identification

For a given t, our selection into treatment assumption implies:

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

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

- We only need to estimate $\mathbb{E}(\boldsymbol{\gamma}_i \mid D_i = 1)$ 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 rows corresponding to $t \leq T_0$. Then for $t > T_0$,

$$\begin{aligned}
&\mathbb{E}\left(y_{it} - \boldsymbol{f}_t'(\boldsymbol{F}_{\text{pre}}'\boldsymbol{F}_{\text{pre}})^{-1}\boldsymbol{F}_{\text{pre}}' \underbrace{\boldsymbol{y}_{i,\text{pre}}}_{\boldsymbol{F}_{\text{pre}}\boldsymbol{\gamma}_i + \boldsymbol{u}_{i,\text{pre}}} \mid D_i = 1 \right) \\
&= \mathbb{E}\big(y_{it} - \boldsymbol{f}_t'\boldsymbol{\gamma}_i \mid D_i = 1 \big) \\
&= \mathbb{E}(y_{it} - y_{it}(0) \mid D_i = 1) = ATT_t,
\end{aligned}$$

where the first equality comes from $\mathbb{E}(u_{it} \mid D_i = 1) = 0$.

ATT_t Identification

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

Technical Detail: There is a well-known identification issue that $\boldsymbol{f}_t$ and γ_i are only known up to rotation:

E.g. $\boldsymbol{f}_t'\gamma_i$ is the same multiply $\boldsymbol{f}_t$ by two and divide γ_i by two

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

ATT_t Identification

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

Technical Detail: There is a well-known identification issue that $\boldsymbol{f}_t$ and γ_i are only known up to rotation:

E.g. $\boldsymbol{f}_t'\gamma_i$ is the same multiply $\boldsymbol{f}_t$ by two and divide γ_i by two

- This is no problem since our procedure produces numerically identical results for any rotation of $\boldsymbol{F}$ (OLS intuition).
- Hence, we only care about the column span of $\boldsymbol{F}$.

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

Estimation of $\boldsymbol{F}$

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

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

- This brings in a large literature on factor model estimation to causal-inference methods
  - $\rightarrow$ Will illustrate multiple estimators of $\boldsymbol{F}$ in application.

Estimation of $\boldsymbol{F}$

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

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

- 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, $D_i = 0$, for estimation of $\boldsymbol{F}$ to avoid bias.
- Staggered treatment 'imputes' $y_{it}(0)$ separately for each treatment-timing group (changing pre)

Removing additive effects

Now, we extend our base model to include additive effects

$$y_{it} = \mu_i + \lambda_t + \boldsymbol{f}_t' * \boldsymbol{\gamma}_i + u_{it}$$

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

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

▸ Within-transform details ▸ 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:

$$\mathbb{E}(\tilde{y}_{it} \mid D_i = 1) = \mathbb{E}\Big(d_{it}\tau_{it} + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\gamma}}_i \mid D_i = 1\Big)$$

where $\tilde{\boldsymbol{f}}_t$ are the pre-treatment demeaned factors and $\tilde{\gamma}_i$ 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

Empirical Application

Empirical Setting

We reevaluate the affect of Walmart openings on local labor markets. Mixed results in empirical literature (Basker, 2005, 2007; Neumark, Zhang, and Ciccarella, 2008).

Empirical Setting

We reevaluate the affect of Walmart openings on local labor markets. Mixed results in empirical literature (Basker, 2005, 2007; Neumark, Zhang, and Ciccarella, 2008).

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 $\boldsymbol{f}_t' \gamma_i$.
  - $\rightarrow$ Intuitively this is reasonable given Walmart's keg and spoke approach

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.

Initial Estimates

To show problems with selection, we estimate a TWFE imputation model (Borusyak, Jaravel, and Spiess, Forthcoming; Wooldridge, 2021):

$$\log(y_{it}) = \mu_i + \lambda_t + \sum_{\ell=-22}^{13} \tau^\ell d_{it}^\ell + e_{it}, \tag{2}$$

- 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)

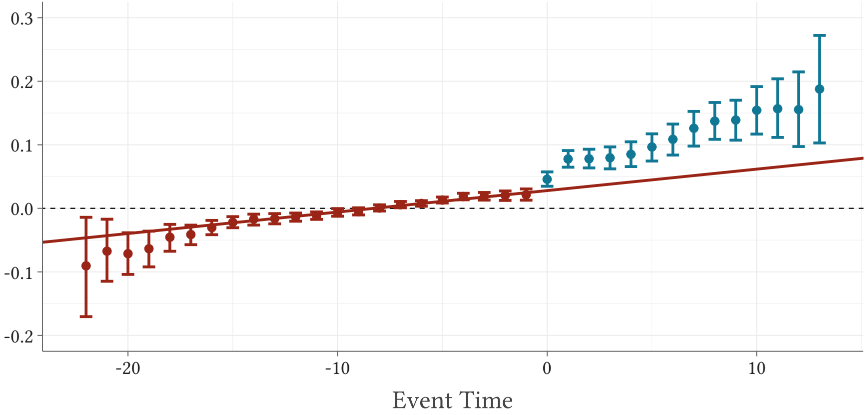
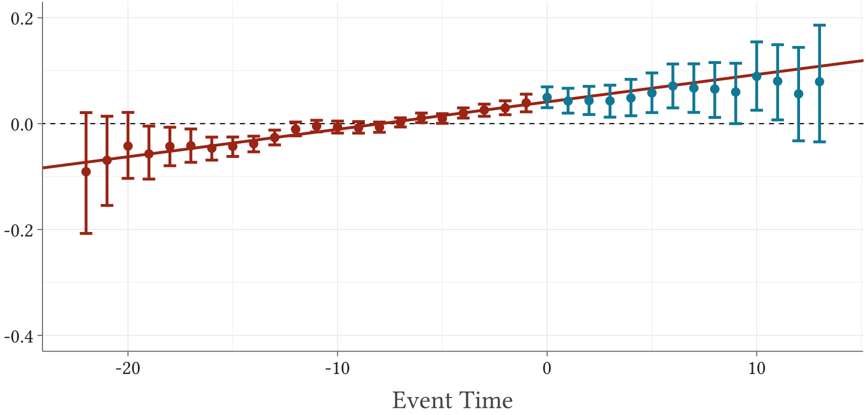


Figure: Effect of Walmart on County $\log$ Wholesale Employment (TWFE Estimate)



Factor Identification

Strategy 1: IV strategy

We consider instrumental-variables based identification strategy as proposed in Ahn, Lee, and Schmidt (2013).

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

▸ 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

▸ Quasi-long Differencing Details

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

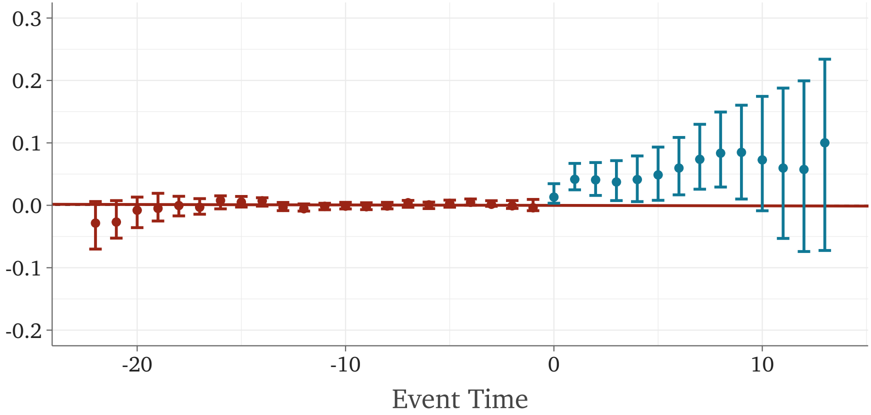
Turning to our factor model estimator, we use the following variables at their 1980 baseline values as instruments:

- 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 $\approx$ 5%, consistent with Basker (2005) and Stapp (2014)

Figure: Including instruments as $X_i\beta_t$ in TWFE Model

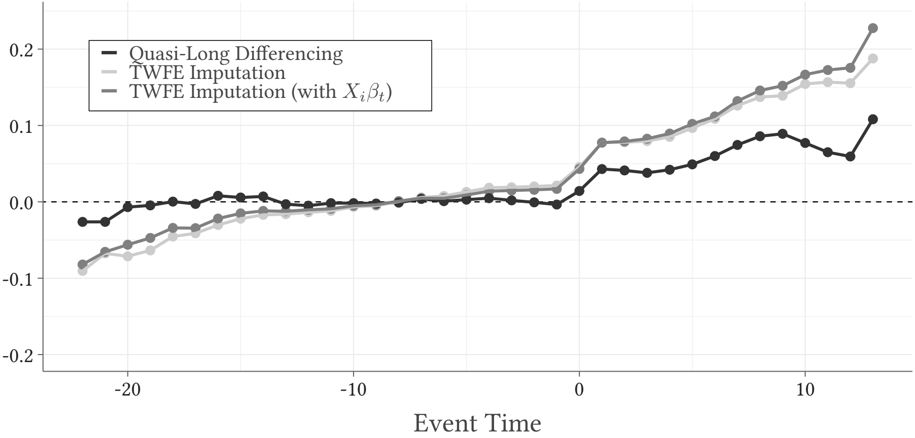


Figure: Effect of Walmart on County $\log$ Retail Employment (Factor Model)

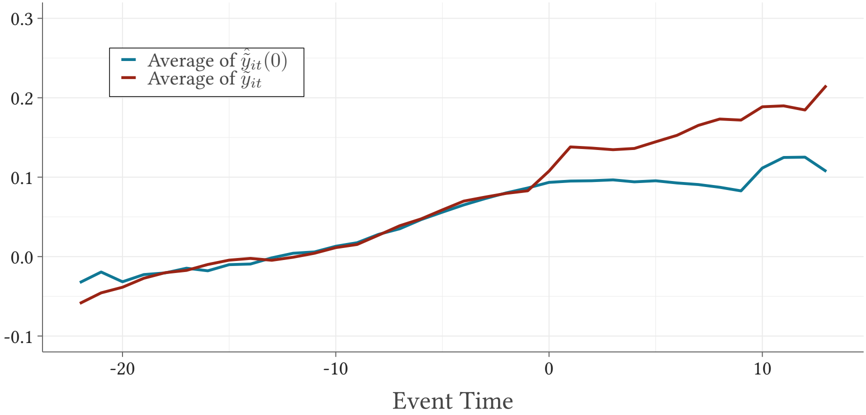
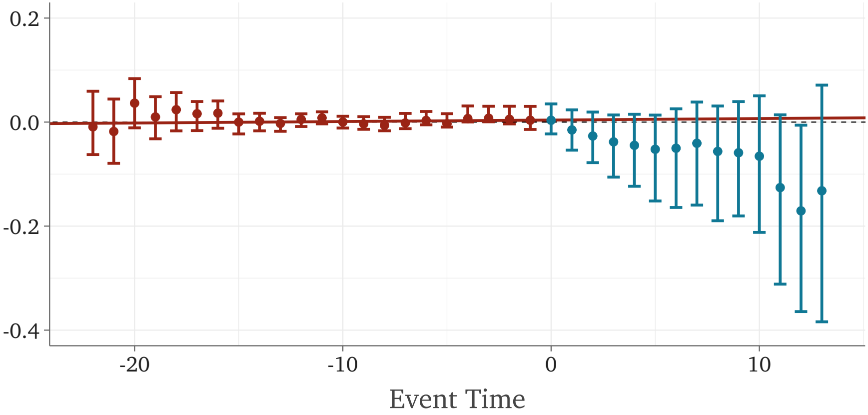


Figure: Effect of Walmart on County $\log$ Wholesale Employment (Factor Model)



Very noisy, but consistent with estimates in Basker (2005).

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 not be autocorrelated (Imbens, Kallus, and Mao, 2021)

▸ Principal Components Details

Alternative identification strategies

Strategy 3: Common Correlated Effects

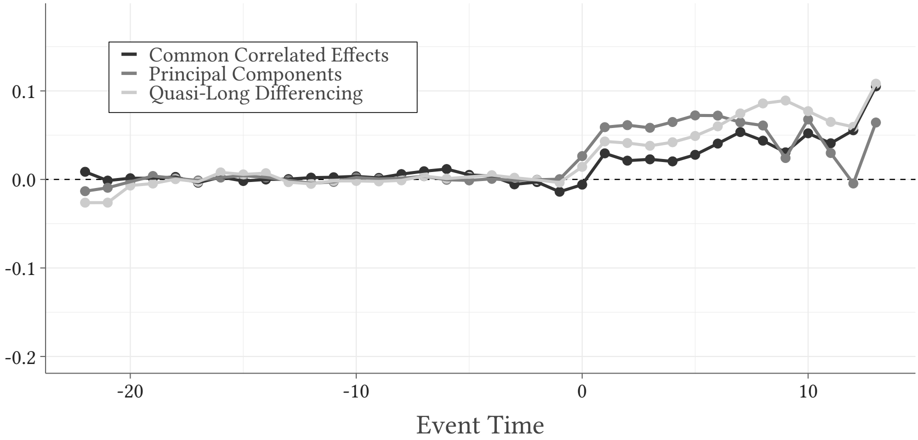
The common correlated effects estimate is based on the availability of a set of additional covariates $\boldsymbol{x}_{it}$ that are affected by the same factors as y_{it} (Freyaldenhoven, Hansen, and Shapiro, 2019; Pesaran, 2006)

- Cross-sectional averages of $\boldsymbol{x}_{it}$ across never-treated i become proxies $\hat{\boldsymbol{F}}_t$. Need $\geq p$ covariates Brown, Butts, and Westerlund (2023).

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: Generalized Procedure allows many factor estimators



Conclusion

Provide general identification results for ATTs under linear factor models.

- Generalizes the two-way fixed effect model and is estimable in short-T settings
- Can use multiple estimators for the factor space.
  - $\rightarrow$ Brown et al. (2023) for CCE.

Implement quasi-long-differencing estimator

- Find results on Walmart's affect on local labor markets similar to Basker (2005).

References I

Abadie, Alberto (2021). "Using synthetic controls: Feasibility, data requirements, and methodological aspects". *Journal of Economic Literature* 59.2, pp. 391–425. DOI: 10.1257/jel.20191450.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". *Journal of the American statistical Association* 105.490, pp. 493–505.

Ahn, Seung C, Young H Lee, and Peter Schmidt (2013). "Panel data models with multiple time-varying individual effects". *Journal of econometrics* 174.1, pp. 1–14. DOI: 10.1016/j.jeconom.2012.12.002.

Arcidiacono, Peter et al. (2020). "The competitive effects of entry: Evidence from supercenter expansion". *American Economic Journal: Applied Economics* 12.3, pp. 175–206. DOI: 10.2139/ssrn.3045492.

Asquith, Brian J, Evan Mast, and Davin Reed (2021). "Local effects of large new apartment buildings in low-income areas". *Review of Economics and Statistics*, pp. 1–46.

References II

Athey, Susan et al. (2021). "Matrix completion methods for causal panel data models". *Journal of the American Statistical Association* 116.536, pp. 1716–1730. DOI: 10.1080/01621459.2021.1891924.

Basker, Emek (Feb. 2005). "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion". *Review of Economics and Statistics* 87.1, pp. 174–183. DOI: 10.1162/0034653053327568.

— **(2007).** "When good instruments go bad: A reply to Neumark, Zhang, and Ciccarella". *Zhang, and Ciccarella (April 2007)*. DOI: 10.2139/ssrn.980988.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (Forthcoming). "Revisiting Event Study Designs: Robust and Efficient Estimation". Review of Economic Studies. DOI: 10.47004/wp.cem.2022.1122.

Brown, Nicholas, Kyle Butts, and Joakim Westerlund (2023). *Simple Difference-in-Differences Estimation in Fixed-T Panels*. eprint: arXiv:2301.11358.

References III

Callaway, Brantly and Sonia Karami (Mar. 2023). “Treatment effects in interactive fixed effects models with a small number of time periods”. *Journal of Econometrics* 233.1, pp. 184–208. DOI: 10.1016/j.jeconom.2022.02.001.

Callaway, Brantly and Pedro HC Sant’Anna (2021). “Difference-in-differences with multiple time periods”. *Journal of Econometrics* 225.2, pp. 200–230. DOI: 10.1016/j.jeconom.2020.12.001.

Ferman, Bruno and Cristine Pinto (2021). “Synthetic controls with imperfect pretreatment fit”. *Quantitative Economics* 12.4, pp. 1197–1221.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro (Sept. 2019). “Pre-Event Trends in the Panel Event-Study Design”. *American Economic Review* 109.9, pp. 3307–3338. DOI: 10.1257/aer.20180609.

Gardner, John (2021). *Two-Stage Difference-in-Differences*. eprint: arXiv:2207.05943.

References IV

Gobillon, Laurent and Thierry Magnac (July 2016). "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls". en. *Review of Economics and Statistics* 98.3, pp. 535–551. DOI: 10.1162/REST_a_00537.

Goodman-Bacon, Andrew (Dec. 2021). "Difference-in-differences with variation in treatment timing". *Journal of Econometrics* 225.2, pp. 254–277. DOI: 10.1016/j.jeconom.2021.03.014.

Imbens, Guido, Nathan Kallus, and Xiaojie Mao (2021). *Controlling for Unmeasured Confounding in Panel Data Using Minimal Bridge Functions: From Two-Way Fixed Effects to Factor Models*. eprint: arXiv:2108.03849.

Kejriwal, Mohitosh, Xiaoxiao Li, and Evan Totty (2021). *The Efficacy of Ability Proxies for Estimating the Returns to Schooling: A Factor Model-Based Evaluation*. DOI: 10.2139/ssrn.3843260.

References V

Neumark, David and Helen Simpson (2015). "Place-based policies". *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 1197–1287.

Neumark, David, Junfu Zhang, and Stephen Ciccarella (Mar. 2008). "The effects of Wal-Mart on local labor markets". *Journal of Urban Economics* 63.2, pp. 405–430. DOI: 10.1016/j.jue.2007.07.004.

Pennington, Kate (2021). *Does building new housing cause displacement?: the supply and demand effects of construction in San Francisco*. DOI: 10.2139/ssrn.3867764.

Pesaran, M Hashem (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure". *Econometrica* 74.4, pp. 967–1012.

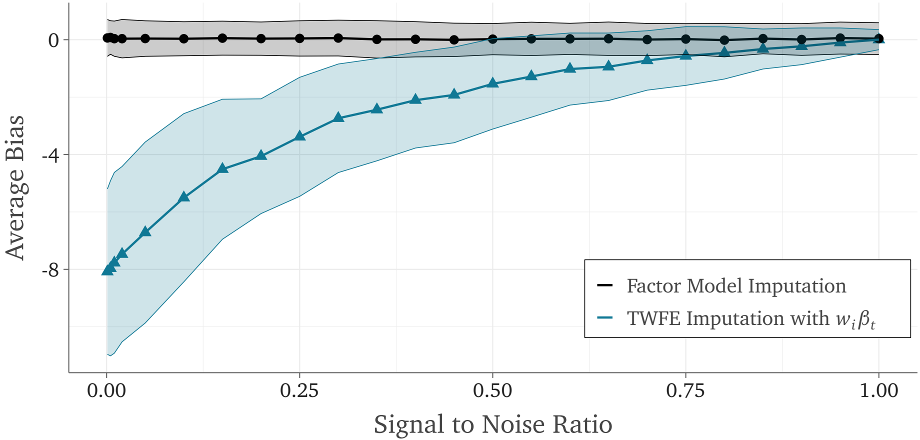
Stapp, Jacob (2014). "The Walmart Effect: Labor Market Implications in Rural and Urban Counties". *SS-AAEA Journal of Agricultural Economics* 2014.318-2016-9525.

References VI

Wooldridge, Jeffrey M (2021). *Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators*.

Xu, Yiqing (2017). "Generalized synthetic control method: Causal inference with interactive fixed effects models". *Political Analysis* 25.1, pp. 57–76. DOI: `10.1017/pan.2016.2`.

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



Back

Intuition of Factor Model

The intuition is very similar to that of the construction 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
- γ_i is an individuals *exposure* (shares) to the shocks

The difference being that **we do not observe** the variables γ_i and $\boldsymbol{f}_t$ (like we don't observe fixed effects)

Removing additive effects

We consider the residuals after within-transforming

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

where

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

Back

Test for TWFE Model

The ATTs are identified by the modified TWFE transformation if

$$\mathbb{E}(\boldsymbol{\gamma}_i \mid D_i) = \mathbb{E}[\boldsymbol{\gamma}_i] \tag{3}$$

For $t > T_0$

$$\mathbb{E}(\tilde{y}_{it} \mid D_i = 1) = \mathbb{E}(\tau_{it} \mid D_i = 1) = \tau_t$$

- 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 (3) under the quasi-long differencing identification strategy

Back

Factor Identification

We cannot identify $\boldsymbol{F}$ or γ_i separately 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.

Quasi-long Differencing Details

The quasi-long differencing method Ahn, Lee, and 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

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

Factor Identification

This transformation creates a set of moments:

$$\mathbb{E}\big(\boldsymbol{W}_i'\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{y}_i \mid D_i = 0\big)$$

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

Selection of p

The paper assumes that the number of factors, p, is known. Methods for selecting p depend on the factor estimator

- For the quasi-long differencing estimator, Ahn, Lee, and Schmidt (2013) provide a procedure for asymptotically determining p (given valid instruments):
  - $\rightarrow$ Start with $p = 0$ and estimate model. Calculate a J-statistic for the GMM model fit. If you reject null, then increase p.
  - $\rightarrow$ Continue this until you fail to reject the null. This asymptotically selects the correct p

- For principal-components, Xu (2017), discuss a similar selection procedure

- For the common-correlated effects estimator, Brown, Butts, and Westerlund (2023) show that you only need the number of covariates to be larger than p.

Principal Components Details

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

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

- 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:

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