Difference-in-differences with Economic Factors and the Case of Housing Returns*

Jiyuan Huang
University of Zurich
Swiss Finance Institute

Per Östberg
University of Zurich
Swiss Finance Institute

This version: November 12, 2023

First version: February 21, 2022

Abstract

This paper studies how to incorporate economic factors in difference-in-differences and document their empirical relevance. We show that even under random assignment directly adding factors with unit-specific loadings into the difference-in-differences estimation results in biased estimates. This bias, which we term the "bad time control problem" arises when the treatment effect covaries with the factor variation. Researchers often control for factor structures by using: (i) unit time trends, (ii) pre-treatment covariates interacted with a time trend and (iii) group-time dummies. We show that all these methods suffer from the bad time control problem and/or omitted factor bias. We propose two solutions to the bad time control problem. To evaluate the relevance of the factor structure we study US housing returns with bank deregulation. Proper control of macroeconomic factors significantly lowers the over-rejection rate of the bank deregulation index from 34% to 7%, and the estimated factor loadings differ systematically across different geographic regions. This results in substantially altered treatment effects.

Keywords: Difference-in-differences, Factor models, House prices.

JEL: C22, C54, G28, R30.

*We are grateful for financial support from the Swiss National Science Foundation (SNSF) through the grant "Trading and Financing during Market Stress #100018_172679". Additionally, we thank Christoph Basten, Nick Brown, Brant Callaway, Michel Habib, Mathias Hoffmann, Xiangying Huang, Philipp Lentner, Steven Ongena, Jon Roth, Olivier Scaillet, Ross Valkanov, Alexander Wagner, Michael Wolf, Jiri Woschitz and seminar participants at the KOF-ETH-UZH International Economic Policy Seminar, NHH Bergen, SFI meetings in Gerzensee, IAAE meetings in Oslo, Econometrics Society 2023 meetings in Barcelona, Beijing and Singapore, and Brown bag seminar at University of Zurich, for valuable comments.

Jiyuan Huang (Jiyuan.Huang@bf.uzh.ch) and Per Östberg (per.oestberg@bf.uzh.ch): Department of Banking and Finance, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland.

1 Introduction

Over the last decades with the adoption of quasi-experimental techniques identification in economics and finance has significantly improved. We have surveyed papers in the *American Economic Review* and 12% of the published papers in 2015 and 2016 use difference-in-differences for identification.[1,2] The majority (71%) of these papers use the two-way fixed effect (TWFE) estimator that only captures very restricted factor variation. The TWFE estimator has also been extensively employed when studying variables that have been documented to have a factor structure (e.g., stock returns and housing returns) and it has been shown theoretically that its omission leads to biased estimates.[3] This raises a number of questions that this paper addresses. First, if factors are observable how should they be included in the difference-in-differences framework? Second, applied researchers use techniques that partially control for the factor structure (e.g., unit time trends), but are these techniques sufficient and unbiased? Third, if we include a factor structure does this alter the conclusions of the difference-in-differences analysis?

An intuitive way to control for the factor structure is to augment the TWFE by allowing for unit-specific loadings (λ_i) interacted with the factor realizations (F_t) in the difference-in-differences estimation. We call this estimation technique the full-sample estimator. Despite its intuitive appeal, we show that the full-sample estimator in general leads to biased estimates of the average treatment effect on treated (ATT). Intuitively, if the *true* treatment effect is time-varying, but a non-dynamic estimator is used and the *true* treatment effect covaries with the factor realizations then the *estimated* factor loadings will capture some of the treatment effect. The end result is estimated treatment effects that capture factor variation and therefore biased estimates of the ATT, we term this bias the "bad time control problem." The bad time control problem only depends on the covariance between the treatment effect and the factor realizations implying that it exists even under random assignment. That is introducing factors, as is common in finance, might result in bias treatment effects in perfect experiments.

In fact, other variants of the full sample estimator are also commonly used. For example,

[1] A summary of our survey is attached at the end of the paper.

[2] de Chaisemartin, and D'Haultfoeuille (2020) report that 19% of papers published in the AER in years 2010 to 2012 use the two-way fixed effects estimator.

[3] Gobillon and Magnac (2016) show that in the presence of an omitted factor structure the two-way fixed effect estimator is inconsistent.

it is common to augment the TWFE estimator by introducing unit-specific time trends (we denote this UTT), which is in effect the full sample estimator where factor realizations are replaced by a time trend. In the context of divorce law, Wolfers (2006) credibly shows that introducing time trends may result in biased treatment effects in the presence of time variation in treatment effects. Additionally, he advocates for pre-treatment estimation of the time trends and documents that the estimates treatment effects are extremely sensitive to how time trends are controlled for. A related alternative is to interact pre-treatment covariates with a time-trend (we denote this CTT - covariates time-trend), this restricts the unit-specific loadings to a linear function of pre-treatment covariates. Since these two augmentations are restricted versions of the full-sample estimator, they are susceptible to the bad time control problem. In our survey, six (five) out of 21 DiD papers use the UTT (CTT) estimators. However, only 18 out of 217 specifications combine time-trends (unit and covariate) with a dynamic estimator.[4,5] That implies for the remaining specifications, if the true treatment effect covaries with trends the estimated ATT's are biased.[6]

One solution to the bad time control problem, which we call the pre-treatment estimator, is to only use pre-treatment variation to estimate the factor model and then subtract the factor variation when estimating the difference. Intuitively, with this estimation procedure factor loadings cannot be affected by time-variation in treatment.

Thus, there are two potential biases facing empirical researchers. On the one hand, in the absence of random assignment the researcher may face omitted factor bias (Gobillon and Magnac, 2016). And on the other hand introducing commonly used factor controls might lead to the bad time control problem. To evaluate the empirical relevance of the two biases we conduct placebo interventions and empirical replications using real estate returns.

We choose real estate returns for a number of reasons. First, there is extensive evidence

[4]For details see Appendices C and D. Dynamic treatment effects are most often used in event-strudy graphs.

[5]Bailey and Goodman-Bacon (2015) interact pre-treatment covariates with a time trend while estimating treatment effects for various event periods. Additionally, in Figures 3 and 4 in Currie, Davis, Greenstone and Walker (2015) there are treatment effects estimated per period while factor variation is saturated using time trends. Finally, columns 1-3 of Table 5 in Bøler, Moxnes and Ulltveit-Moe (2015) use dynamic treatment effects and a unit time-trend.

[6]Another commonly used augmentation that controls for factor variation is to introduce group-time dummies. Each group is assigned a dummy and then these dummies are interacted with dummies for time periods (which could either represent single or multiple time periods). Intuitively, it reduces the bias from omitting the factor structure by removing the between group factor variation. If the dummies aggregate over multiple periods (e.g., one dummy for each five year period) then the estimator suffers from the bad time control problem since it is possible that the dummy factor variation covaries with the treatment effect. See section 3.3 for more details.

that real estate returns exhibit a factor structure.[7] Second, interventions are often used where parallel trends may not hold. Specifically, we consider the deregulation index created by Rice and Strahan (2010) which is used in conjunction with real estate returns by Favara and Imbs (2015).[8]

Figure 1 reprints Figure B1 of Rice and Strahan (2010) and the picture indicates that the coasts deregulate faster than the midwest, suggesting that deregulation is not randomly assigned. Additionally, Kroszner and Strahan (1999) show that political economy is a determinant of banking deregulation. Third, due to data availability, researchers often resort to using low frequency yearly data which implies higher post-treatment factor volatility that is not controlled for by time fixed effects. Intuitively, the benchmark model is irrelevant when the post-treatment time tends to zero. Fourthly, the interventions studied often have many post-treatment periods.

In designing our placebo analysis we modify the deregulation index of Rice and Strahan. To create binary and sharp interventions (the index ranges from 0-4 and is staggered) we randomly sample an index value and all states with a greater or equal index value are considered treated. This maintains the spatial correlation in treatment while avoiding the problems associated with staggered and continuous treatment. In the absence of simulated treatment effects the TWFE estimator rejects 32% of the time. However, introducing state time-trends results in a rejection rate of 92% and controlling for economic factor using the full-sample estimator results in a rejection rate of 54%. In contrast, the pre-treatment estimator with optimally selected economic factors results in a rejection rate of 8%. Our placebo analysis highlights how important choice of estimation method is for this application.

To evaluate the empirical relevance of factors and factor estimation methods in difference-in-differences we revisit specifications of Favara and Imbs (2015) and Zevelev (2021). In the case of Favara and Imbs (2015), introducing optimally selected factors renders estimated

[7] For example, Cotter, Gabriel and Roll (2014) document the explanatory ability of macroeconomic factors in the cross-section of MSA (metropolitan statistical area) housing returns. Further, arbitrage pricing theory (APT) models with macroeconomic factors have been used in Chan et al. (1990), statistical factors (PCA) are employed by Titman and Warga (1986) while equity based factors such as the Fama-French factors, momentum and liquidity have studied in the real estate context by Peterson and Hsieh (1997), Hung and Glascock (2010) and Cannon and Cole (2010).

[8] The effect of interstate banking deregulation has been extensively studied, among other things it has been documented to lead to less pronounced business cycles (Morgan, Rime and Strahan, 2004) per capital growth in income and output (Jayaratne and Strahan, 1996), credit costs of borrowers (Rice and Strahan, 2010), lower Income inequality (Beck, Levine and Levkov, 2010) and reallocation across sectors (Acharya, Imbs and Sturgess, 2011).

treatment effects insignificant both when using the full and pre-treatment estimators. We also considered all possible factor combinations. For the full-sample estimator adding up to 3 factors implies that only 41.9% of estimated treatment effects remain statistically significant. For the pre-treatment estimator adding up to 3 factors implies that only 53.3% of the treatment effects are still significant. In addition, we show that the estimated state loadings differ systematically across U.S. regions.

Zevelev (2021) studies a Texas collateral reform. Given the importance of oil for Texas he uses the full-sample estimator with the oil price as a factor. In addition he introduces state time-trends. In the absence of factor controls the estimated treatment effect has the opposite sign. This highlights the importance of using factors especially when treatment controls are such diverse units as states. In most cases, introducing our economic factors reduces estimated treatment effect, suggesting that economic factors are important controls. One strength of Zevelev's paper is that he analyses the treatment effect only considering neighboring states and neghboring counties. In these cases, it seems as if adding factors has a lower impact on treatment effects. Intuitively, the improved internal validity might imply that the factor loadings of treated and control are more similar. Overall, our empirical replications highlight the importance of factor controls.

This paper makes a number of contributions. First, we introduce the bad time control problem that is a sibling of the bad control problem. Bad controls are controls that are affected by treatment *status* while the bad time control problem arises due to a correlation between the treatment effect *value* and controls (factors in our case). Second, we characterize the bias of the full-sample estimator and commonly used augmentations of TWFE estimators. Third, we show that for housing returns the TWFE estimator does not capture sufficient factor variation, especially when interventions are at the state level and the sample is as diverse as the entire United States.

The rest of the paper is organized as follows. Section 2 examines the related literature while section 3 describes our theoretical results and Section 4 presents our simulations. Our empirical evidence can be found in Section 5 and Section 6 concludes.

2 Related Literature

Our paper contributes to the broad effort examining empirical methods in economics in finance (Bertrand et al., 2004, Gormley and Matsa, 2014 Petersen, 2008 MacKinnon and Webb, 2017, Roberts and Whited, 2012).

Important work challenges economic and empirical validity of the assumptions made in the difference-in-differences framework. Karpoff and Wittry (2018) use institutional and legal context to challenge that antitakeover laws are exogenous. They show that inference for nine outcome variables in nine studies are significantly altered when legal and historical context is taken into account. Additionally, Berg et al. (2021) show that spatial interaction and competition is not consistent with the stable unit value assumption (SUTVA). Further, Boehmer et al. (2020) document the importance of spillovers in financial regulatory experiments. In a recent paper, Nyborg and Woschitz (2023) argue that difference-in-differences models could be biased when bond yields have a term structure (a form of a factor model).

Not only is the parallel trends challenged due to economic context, Roth (2022) highlights the econometric difficulty in testing for parallel trends. Roth and Sant'Anna (2023) derives a parallel trends test for non-linear functional forms. Rambachan and Roth (2023) relax the strict parallel trends assumption and offer novel inference methods in this scenario. Roth et al. (2022) provide an extensive review of recent work on difference-in-differences.

Given all of the above, like Gobillon and Magnac (2016) we depart from the parallel trends assumption. Among other things they show that, given an omitted factor structure the TWFE estimator is inconsistent (our Proposition 1 proves an analogous result). Econometrically, our paper builds on this result and shows that just including factors may result in the bad time control problem.

A recent strand of literature considers identification in an interactive fixed effect setting (i.e., there is a factor structure, but both factors and loadings are unobserved). In an elegant paper, Callaway and Karami (2022) show that GMM can estimate the ATT consistently under the assumption that there exists covariates with time-invariant effects. Brown and Butts (2023) develop an identification strategy that consistently estimates factor structure using control samples while we mainly focus on evaluating the performance of currently widely-used methods empirically.

Caetano et al. (2022) introduce covariates into the difference-in-difference framework and carefully determine under what conditions the ATT can be recovered. An additional insight of this paper is that using the TWFE with covariates has five potential bias terms. As such this paper carefully defines and significantly extends the bad control problem. In their setting potential covariates are affected by treatment *status* and exists even if treatment effects are homogeneous. In contrast our factors are by definition unaffected by treatment status, but the bad time control problem does not exist if treatment effects are time-invariant.

Another strand shows that even under parallel trends the TWFE estimator may assign wrong and even negative weights to treated observations. The intuition for this is that with heterogeneous and staggered treatment already treated units may become comparison units and resulting in very poor treatment effect estimates (de Chaisemartin and D'Haultfoeuille, 2020, Borosyak, Jaravel and Spiess, 2021, Goodman-Bacon, 2021, Sun and Abraham, 2021). The relevance of this issue and methodological choice in general for accounting and finance is highlighted by the impressive survey contained in Baker et al. (2022).

3 Combining Factors with the DiD

In this section we describe the implications for inference using difference-in-difference when the true data generating process has a linear factor structure. First, we characterize the bias when the factor is omitted. Second, we establish that the full-sample estimator suffers from the bad time control problem. Additionally, we characterize the degree to which commonly used techniques such as unit time trends, covariate time trends, and dummy factors suffer from the bad time control problem. Finally, we provide two solutions to researchers that want to control for time trends. As a first possible solution, we show that combining dynamic treatment effects with the full-sample estimator eliminates a bad time control problem. As a second alternative, we propose and prove that the two-step pre-treatment estimator results in unbiased estimation of both the treatment effect and factor loadings. The idea behind this estimator is analogous to the framework developed by Brown and Butts (2023).

We start by describing our setting and assumptions. The observed sample consists of the following set of variables $\{Y_{it}, F_t, D_i, P_t\}$ where i and t indicates unit and time, respectively. Y_{it} is the observed outcome. The potential outcome if not treated is indicated by $Y_{it}(0)$ while the potential outcome if treated is given by $Y_{it}(1)$. F_t is $r \times 1$ vector of *observed* time-specific factor realizations and λ_i is $r \times 1$ vector of *unobserved* individual-specific corresponding factor loadings, where r is the number of factors.[9,10] D_i is a unit indicator that takes the value of one when unit i belongs to the treated group. P_t is a post-treatment indicator that takes the value of one for all $t \geq T^*$ when treatment occurs at time T^*. We assume that there are at least two units ($N \geq 2$), one treated and one control. Additionally, we assume that the sample covers at least three time periods and at least two post treatment periods ($T \geq 3$, $T - T^* + 1 \geq 2$).[11]

Assumption 1 (Random sampling)**.** *An observed sample consists of* $\{Y_{it}, D_i, P_t, F_t\}_{i=1,t=1}^{N,\ T}$. *Observed samples are independently and identically distributed.*

This standard assumption ensures that we can examine the properties of estimators. However, importantly, it does not imply that within a particular sample our variables are independently distributed.

Assumption 2 (Conditional parallel trends)**.** *Conditional on the factor structure, there are parallel trends in potential outcomes.*

$$\begin{aligned}
&\mathbb{E}\left[Y_{it}(0) - Y_{is}(0) - Y_{jt}(0) + Y_{js}(0) | \, \lambda_i, \lambda_j, F_t, F_s, D_i = d_1, D_j = d_2, P_t = p_1, P_s = p_2\right] \\
=&\mathbb{E}\left[Y_{it}(0) - Y_{is}(0) - Y_{jt}(0) + Y_{js}(0) | \, \lambda_i, \lambda_j, F_t, F_s\right] \qquad\qquad \forall (d_1, d_2, p_1, p_2) \in \{0,1\}^4
\end{aligned}$$

This assumption implies that double-demeaned (in both unit and time dimensions) potential outcomes have the same expectation irrespective of treatment status when factor structure is controlled for. In essence, compared to the standard parallel trends assumption we add the requirement of conditioning on the factor structure.

Assumption 3 (Strict exogeneity)**.** *Potential outcomes are mean independent with factor loadings and factor realizations assigned to other units and periods.*

$$\mathbb{E}\left[Y_{it}(0) | \, \lambda_i, F_t\right] = \mathbb{E}\left[Y_{it}(0) | \, \boldsymbol{\lambda}, \boldsymbol{F}\right]$$

$$\mathbb{E}\left[Y_{it}(1) | \, \lambda_i, F_t\right] = \mathbb{E}\left[Y_{it}(1) | \, \boldsymbol{\lambda}, \boldsymbol{F}\right]$$

[9] For simplicity we consider a single factor, but all of our results can be generalized into a multi-factor setting.

[10] Note that a time fixed effects can be represented as a factor that all units have identical loadings to. Symmetrically, the unit fixed effect can be seen as time-invariant factor, but with unit specific loadings.

[11] We require at least two post-treatment periods since bad time control problem requires time-varying treatment effects.

Since we are introducing fixed effects we need to make the assumption of strict exogeneity. This assumption has two important implications. First, potential outcomes are uncorrelated with past and future factor realizations once current the current factor realization has been conditioned upon. Second, potential outcomes are uncorrelated with the loadings of all other units.

Assumption 4 (Bilinear factor structure)**.** *The expectation of the potential outcome if not treated is a bilinear function of* λ *and* F*.*

$$\mathbb{E}\left[Y_{it}(0) - Y_{is}(0) - Y_{jt}(0) + Y_{js}(0)\right] = (\lambda_i - \lambda_j)'(F_t - F_s)$$

This is assumption implies that the data generating process (DGP) contains a factor structure in addition to the fixed effects. Assuming that the factor structure is linear has the advantage that our DGP has the best possible chance to match the linear estimators that are frequently used in practice. Allowing for another functional form does not change the spirit of our conclusion, but will possibly introduce other bias terms.

Assumption 5 (Stable unit treatment value assumption, SUTVA)**.** *Potential outcomes are independent from the treatment status of other units and periods.*

$$Y_{it}(0), Y_{it}(1) \perp\!\!\!\perp \boldsymbol{D}_{-i}, \boldsymbol{P}_{-t}$$

We assume there are no spillovers of treatment on potential outcomes across units and time.

Assumption 6 (Stable loading and factor assumption)**.** *Factor loadings and factor realizations are independent of the treatment assignment of other units and periods.*

$$\lambda_i \perp\!\!\!\perp \boldsymbol{D}_{-i} \qquad F_t \perp\!\!\!\perp \boldsymbol{P}_{-t}$$

We assume that there are no spillovers of treatment on loadings and factor realizations.

Our setting differs to that of Caetano et al. (2022) in that we do not assume that covariates (factors in our case) are affected by treatment status, but we require treatment heterogeneity (in the time dimension). Further, our settings differs from papers examining inference in staggered difference-in-differences in that we have strict interventions (e.g., de Chaisemartin and D'Haultfoeuille, 2020 and Goodman-Bacon, 2021), but we do require the assumption that trends are only parallel once the factor structure has been conditioned upon.

In the next few sub-sections we consider whether a number of commonly used estimators provide unbiased estimates of the average treatment effect on treated (ATT). Define $\Delta_{it} = Y_{it}(1) - Y_{it}(0)$ as the *unobserved* true treatment effect for each unit and period. The goal of difference-in-differences estimators is to measure the average treatment effect on treated (ATT),

$$\alpha^{\text{ATT}} = \mathbb{E}[\Delta_{it}|D_i = 1, P_t = 1].$$

3.1 Two-Way Fixed Effect estimator

The classical Two-Way Fixed Effect (TWFE) estimator can be defined as,

$$(\hat{\alpha}^{\text{TWFE}}, \hat{\gamma}_i^{\text{TWFE}}, \hat{\eta}_t^{\text{TWFE}}) = \underset{\alpha,\gamma,\eta}{argmin}\Big\{\sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{i,t} - \gamma_i - \eta_t - \alpha D_i P_t)^2\Big\} \tag{1}$$

Proposition 1. *Given an omitted linear factor structure, the Two-Way Fixed Effect estimator estimates $\hat{\alpha}^{TWFE}$ as,*

$$\mathbb{E}[\hat{\alpha}^{TWFE}] = \alpha^{ATT} + (\mathbb{E}[\lambda_i|D_i = 1] - \mathbb{E}[\lambda_i|D_i = 0])(\mathbb{E}[F_t|P_t = 1] - \mathbb{E}[F_t|P_t = 0]). \tag{2}$$

Proof See Appendix A.2.

Thus, the TWFE estimator is in general biased. The bias term, $(\mathbb{E}[\lambda_i|D_i = 1]-\mathbb{E}[\lambda_i|D_i = 0])(\mathbb{E}[F_t|P_t = 1] - \mathbb{E}[F_t|P_t = 0])$, is non-zero if loadings of treated and control units are not equal and the pre to post factor realizations are not equal. To arrive at the bias expression, we double-demean the outcome variable (adjust for fixed effects) and proceed to calculate the omitted factor bias. This result is equivalent to Eq. (21) in Gobillon and Magnac (2016).[12]

Let us consider an example of how the omitted factor bias could affect the estimated treatment effect when using the TWFE estimator. Suppose we were to examine the return of target firms in mergers where the treatment event is the merger. Usually target firms are smaller and therefore also are expected to have lower market betas. If stock returns are determined by the market model, then the bias of the TWFE would be given by the

[12] They prove inconsistency of the difference-in-differences estimator *even* when the factor is deterministic. For our purposes, we are going to treat the factor as a random variable (similar to (Bai, 2009)) since this corresponds closer to the economic setting we are interested in.

difference in market betas between merging and non-merging firms multiplied by the return difference. Intuitively, the omitted factor bias would be increasing in time, as longer time-periods imply larger factor volatility.

Given that the TWFE estimator suffers from the omitted factor bias we now proceed to examine a number of estimators that at least partially control for factor structures.

3.2 Full-sample estimator

If the factors are observable, one potential solution would be to simply introduce the factors and estimate unit loadings jointly with the treatment effect. We refer to this method as the full-sample (FS) estimator,

$$(\hat{\alpha}^{\text{FS}}, \hat{\gamma}_i^{\text{FS}}, \hat{\eta}_t^{\text{FS}}, \hat{\lambda}_i^{\text{FS}}) = \underset{\alpha,\gamma,\eta,\lambda}{argmin} \Big\{ \sum_{t=1}^{T} \sum_{j=1}^{N} (Y_{i,t} - \gamma_i - \eta_t - \lambda_i F_t - \alpha D_i P_t)^2 \Big\} \tag{3}$$

where λ_i are unit specific loadings and F_t are observable factor realizations. In practice it is straightforward to estimate unit specific loadings, one interacts unit dummies with the time-series of factor realizations to allow for unit specific sensitivities to the factors.

Proposition 2. *When the factor realizations are exogenously determined, the FS estimator can be expressed as,*

$$\mathbb{E}[\hat{\alpha}^{FS}] = \alpha^{ATT} + w^{FS} Cov(F_t, \Delta_{it} | D_i P_t = 1). \tag{4}$$

where the bias (bad time control problem) is given by $w^{FS} Cov(F_t, \Delta_{it} | D_i P_t = 1)$. The term w^{FS} is,

$$\mathbb{E}\left[\frac{N_T T_P - 1}{N_T T_P} \cdot \frac{\overline{F}_{pre} - \overline{F}}{(1 - \overline{P})\sigma^2_{F,pre} + \overline{P}\sigma^2_{F,post}} \right]. \tag{5}$$

Proof See Appendix A.3.

Proposition 2 illustrates that only in partcular circumstances will the full-sample estimator uncover the true ATT. It will be unbiased if there is no difference in factor realizations between the pre and post treatment periods ($\mathbb{E}[F_t | P_t = 0] - \mathbb{E}[F_t | P_t = 1]$) or the factor realizations do not covary with the treatment effect ($\text{Cov}(F_t, \Delta_{it} | D_i P_t = 1) = 0$). The mechanism behind this result is that if the true ATT is time varying, but the full sample estimator assumes constant treatment effects and as a result estimated treatment effect may capture factor variation.

The covariance term in the bad time control is scaled by w^{FS}, which mainly depends on factor volatility. $(N_T T_P - 1)/(N_T T_P)$, is a degree of freedom adjustment where N_T is the number of treated units and T_P is the number of post-treatment periods. The numerator of the second part of the expression, $\overline{F}_{\text{pre}} - \overline{F}$, captures the difference in mean factor realizations between the pre-treatment period and the overall period, while the denominator is the probability weighted convex combination of pre and post factor volatility. This means that if we keep pre and post factor volatility constant, the absolute size of the full sample bias is increasing in total factor volatility. This suggests that empirical applications where treatment coincides with regime shifts in volatilty, such as the gloabal financial crisis are particularly vulnerable to bad time controls.

The paper by Zevelev (2021) discussed in section 5.4 uses the full sample estimator.

Corollary 1. *If the treatment effect is time invariant ($\forall t,\ \Delta_{it} = \Delta_i$) then the full-sample estimator is unbiased since $Cov(F_t, \Delta_i | D_i P_t = 1) = 0$*

An implication of corollary 1 is that the bias is not generated by lack of random assignment. The bias is driven by the covariance in the time dimension.

Corollary 2. *The bias of the full-sample estimator is independent of the unit loadings (λ_i), therefore the full sample estimator is biased even under random assignment.*

One implication of Corollary 2 is that the full-sample estimator may be more biased than the classical TWFE estimator. The intuition is that if the loadings of treated and control units are sufficiently close ($\mathbb{E}[\lambda_i | D_i = 1] \simeq \mathbb{E}[\lambda_i | D_i = 0]$) then the bias in the TWFE estimator will not be large while if the treatment effect significantly covaries with the factor then the full-sample may be more biased than the TWFE. The implication of this is that if the researcher has close to random assignment (so limited omitted factor bias) and includes omitted factors the result may be even more biased estimates of the ATT than if the factors were omitted.

It is important to consider what happens to the bad time control problem as the number of units increase and as the number of time periods increase. Intuitively, the number of units does not affect the size of the bad time control problem since it is driven by covariances in the time dimension. Asymptotically, the bad time control problem tends to zero as the ratio of treated time periods T_P to total time periods (T) (since $\mathbb{E}[P_t]$ can be expressed as T_P/T) tends to zero.[13]

In practice, researchers add unit-specific time trends, termed unit time trend (UTT) estimator, to control for time-varying heterogeneity. However, it is a special case of the full sample estimator and therefore also suffers from the bad time control problem. We consider the unit-specific linear time trend $F_t = t$ as an example, but clearly this can be generalized to polynomial time trends.

Corollary 3. *The unit time trend (UTT) estimator is unbiased if and only if the treatment effect is orthogonal to the linear time trend over the treated observations ($Cov\,(\,\Delta_{it}, t|\,D_iP_t = 1) = 0$).*

Proof See Appendix A.4.

Another common way to augment TWFE estimator is by adding covariates interacted with a time trend. To avoid the bad control problem (controls that are affected by treatment status), researchers often use pre-treatment covariates and interact them with a time trend (e.g., $X_{i0} \cdot t$), which by definition is unaffected by treatment status. We show that this augmentation may circumvent the bad control problem, but leads to a bad time control problem.

Proposition 3. *The covariate time trend estimator leads to a bad time control problem as long as $Cov\,(\,\Delta_{it}, X_{i0} \cdot t|\,D_iP_t = 1) \neq 0$*

Proof See Appendix A.5.

3.3 The Dummy factor estimator

A commonly used augmentation of the difference-in-differences that partially controls for factor variation is to introduce group-time dummies. Each group is assigned a dummy (e.g., firms in the same industry) and then these dummies are interacted with dummies for time periods (which could either represent single or multiple time periods). We call this control procedure the dummy factor method. 7 out of 21 DiD papers in our survey use group-time dummies.

[13] In the admittedly unrealistic setting with observable loadings and unobservable factors the full-sample estimator would still be biased, but the bias would depend on the covariance of the loadings with treatment heterogeneity in the unit dimension.

The dummy factor is defined by the granularity chosen along the unit and time dimension. We examine two limiting cases of the dummy factor. First, we consider the case where we define the group dummies R_i for $|R| < N$ groups interacted with T period dummies (i.e., we have the most amount of granularity in the time dimension and less than full granularity in the unit dimension). Second, we consider the case when we have N groups (i.e., full granularity in the unit dimension) interacted with S_t time periods, where $|S| < T$.

Proposition 4. *(i) The dummy factor with $R \times T$ dummies does not suffer from the full-sample bias. The weighting of observations is altered, implying that the ATT is generally not estimated. (ii) The dummy factor with $N \times S$ suffers from being a convex combination of the full-sample bias and the weighting error.*

Proof See Appendix A.6.

In conclusion, the dummy factor estimate is a convex combination of of TWFE estimates for each group. Omitted factor bias of the dummy factor estimator, compared to the TWFE estimator, is reduced because factor structure variation across groups is eliminated. When all groups have the same treated observation ratio, the dummy factor estimator degenerates into the TWFE estimator.

When loadings are balanced within each group, the omitted factor bias is zero, even though the dummy factor estimator is still biased because of the weighting issue. The weighting issue is not as severe as in staggered DiD, because the weighting is gruaranteed to be between 0 and 1. In other words, the estimated treatment effect term is a convex combination of the true treatment (unlike in de Chaismartin and D'Haultfoeuille, 2020) so it guarantees that the estimator is not negative if true treatment effects are all positive.

3.4 Pre-treatment estimator

Given a sufficiently long time-series another possible solution is to estimate factor loadings only using pre-treatment variation and then use the estimated loadings when estimating the ATT in the full sample. We refer to this two step procedure as the pre-treatment (PT) estimator. First loadings are estimated over pre-treatment periods,

$$(\hat{\lambda}_i^{\mathrm{PT}}) = \underset{\lambda}{argmin}\Big\{ \sum_{t=1}^{T}(1-P_t)\sum_{j=1}^{N}(Y_{i,t}-\gamma_i-\eta_t-\lambda_i F_t)^2\Big\} \tag{6}$$

and the estimated loadings ($\hat{\lambda}_i$) are then used in the full sample when estimating the ATT,

$$(\hat{\alpha}^{PT}, \hat{\gamma_i}^{\text{PT}}, \hat{\eta_t}^{\text{PT}}) = \underset{\alpha,\gamma,\eta}{argmin}\Big\{\sum_{t=1}^{T}\sum_{j=1}^{N}(Y_{i,t} - \gamma_i - \eta_t - \hat{\lambda}_i^{\text{PT}} F_t - \alpha D_i P_t)^2\Big\}. \tag{7}$$

The pre-treatment estimator avoids estimated loadings capture the treatment effect variation and therefore does not lead to biased estimation.

Proposition 5. *The pre-treatment estimator results in an unbiased estimate of the Average Treatment on Treated (ATT).*

Proof See Appendix A.7.

The pre-treatment estimator has a number of advantages over the GMM estimator with dynamic treatment effects proposed by Callaway and Karami (2022). First, recovering the ATT and standard error using the dynamic estimator requires additional calculations. Second, it is not completely clear how the dynamic estimator performs when treatment is staggered. The advantages of the dynamic estimator is that it does not require a long time-series to estimate loading and provides useful information about the treatment effect over time.

There are multiple drawbacks of using the pre-treatment estimator. Like all imputation estimators () it only uses part of the sammple implying that we lose power. Additionally, we know from finance applications () that in practice loadings are likely to be time-varying. Further, when the researcher does not know what the true factors are the pre-treatment estimator cannot be used.

4 Simulation Evidence: Different factor estimation methods

To illustrate our theoretical findings, we simulate data according to the following data generating process,

$$Y_{it} = \gamma_i + \eta_t + \lambda_i F_t + \Delta_{it} D_i \times P_t + \varepsilon_{it} \tag{8}$$

$$\gamma_i \sim N(0, \sigma_\gamma^2) \qquad \eta_t \sim N(0, \sigma_\eta^2) \qquad \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$$

γ_i are unit fixed effects, η_t are time fixed effects, $\lambda_i F_t$ represents the factor structure, and Δ_{it} is the heterogeneous and time-varying treatment effects. By construction the unit and time fixed effects are independent from our other key quantities. In order to allow for loading differences between treated and control we simulate loadings as follows,

$$\lambda_i = H_i + \mu(D_i - \mathbb{E}[D_i]) \quad H_i \sim N(0,1)$$

where the parameter μ allows us to shift the loading of treated units while maintaining a mean loading of zero. Our factor realizations are given by,

$$F_t = Q_t + \nu(t - \mathbb{E}[t]) \quad Q_t \sim N(0,1)$$

where the key parameter is ν that captures the time-trend of our factor. As our loadings, our factors are modeled to have mean zero. The last ingredient of our simulation is our treatment effects which are simulated as follows,

$$\Delta_{it} = ATT + \sigma_\Delta(\psi U_i + \phi V_t + (1 - \psi^2 - \phi^2)W_{it})$$
$$\text{corr}(U_i, P_i) = \rho_{\Delta,\lambda} \qquad corr(V_t, Q_t) = \rho_{\Delta,F} \qquad U_i, V_t, W_{it} \sim N(0,1)$$

where the true ATT is 1.0, U_i is a unit specific treatment effect, V_t is the time specific treatment effect and the term $(1 - \psi^2 - \phi^2)W_{it}$ ensures that the total variance is kept constant. Crucially, the parameter $\text{corr}(U_i, H_i) = \rho_{\Delta,\lambda}$ allows for a possible correlation between the loadings and the treatment effect and similarly $corr(V_t, Q_t) = \rho_{\Delta,F}$ allows for a correlation between the treatment effect and the factor realization.

For each sample we estimate three models: (i) the TWFE estimator defined in Eq. (1), (ii) the full-sample estimator defined in Eq. (3) and (iii) the pre-treatment estimator defined in Eq. (6). We expect the TWFE to suffer from omitted factor bias, and we expect the full sample estimator to perform worse than the pre-treatment estimator when $\rho_{\Delta,F}$ is different from zero. In this setting, we expect the pre-treatment to always be unbiased.

We perform 1000 iterations for each cell. In our simulation analysis, we set number of units $N = 1000$, number of treated unites $N_T = 0.3N = 300$, number of periods $T = 20$, number of post-treatment periods $T_P = 0.5T = 10$. We also set the baseline value of stardard deviation of unit fixed effect $\sigma_\gamma = 1$, standard deviation of time fixed effect $\sigma_\eta = 3$, standard deviation of error term $\sigma_\varepsilon = 3$, standard deviation of treatment effect $\sigma_\Delta = 2$, loadings difference between treatment group and control group $\mu = 0.4$, linear time trend $\nu = 1.5$, degree of time-varition in the treatment effect $\phi = 0.6$, degree of unit heterogeneity in the treatment effect $\psi = 0.6$, correlation between treatment effect and loadings $\rho_{\Delta,\lambda} = 0.6$, and the correlation between treatment effect and factors $\rho_{\Delta,F} = 0.6$.

Table 1: Homogeneous Treatment Effects

We set $\sigma_\Delta = 0$ (homogeneous treatment effects). The loading difference μ and factor trend ν are the parameters of interest. All other parameter are the same as in our baseline setting. The mean of two-way fixed effect estimators is displayed in the first line without any parentheses, the mean of full sample estimator is displayed in the second line with round parentheses, and the mean of pre-treatment estimators is displayed in the third line with square parentheses.

TWFE Estimator (FS Estimator) [PT Estimator]		Loading Difference μ					
		0	0.2	0.4	0.6	0.8	1
Factor Trend ν	0	0.9958	1.0075	0.9824	1.0070	1.0029	0.9969
		(0.9988)	(1.0044)	(0.9993)	(1.0036)	(1.0031)	(0.9941)
		[0.9986]	[1.0037]	[0.9998]	[1.0022]	[1.0041]	[0.9936]
	0.5	0.9947	1.0437	1.1183	1.1434	1.2331	1.2454
		(0.9977)	(0.9964)	(1.0033)	(0.9995)	(0.9976)	(1.0000)
		[0.9983]	[0.9967]	[1.0055]	[1.0001]	[0.9976]	[0.9999]
	1	1.0030	1.1072	1.1973	1.3525	1.4156	1.5557
		(1.0018)	(0.9975)	(1.0015)	(0.9982)	(0.9966)	(1.0000)
		[1.0020]	[0.9978]	[1.0012]	[0.9979]	[0.9964]	[0.9997]
	1.5	0.9998	1.1387	1.3233	1.4781	1.6147	1.7812
		(0.9991)	(1.0025)	(1.0032)	(0.9960)	(1.0007)	(0.9949)
		[0.9973]	[1.0038]	[1.0029]	[0.9975]	[1.0005]	[0.9954]
	2	1.0059	1.2167	1.4339	1.6442	1.8459	2.0769
		(0.9969)	(0.9960)	(1.0048)	(1.0059)	(0.9909)	(1.0000)
		[0.9971]	[0.9963]	[1.0042]	[1.0061]	[0.9916]	[1.0011]
	2.5	1.0051	1.2607	1.5119	1.7841	2.0723	2.3543
		(1.0005)	(0.9957)	(0.9975)	(0.9989)	(1.0010)	(1.0055)
		[1.0017]	[0.9981]	[0.9970]	[1.0000]	[0.9997]	[1.0011]

Table 1 is designed to illustrate Proposition 1, that is we assume the treatment effect is homogeneous and time in-variant. On the horizontal axis we change the loading difference of treated and control units. In the left most column there is no loading differential while on in the right most the loading differential is one corresponding to one standard deviation of the loading (λ). On the vertical axis we allow for an increasing factor trend, from none to 1.6 factor deviations. If there is no loading difference or no factor trend the TWFE is unbiased (there is only sampling error around the true ATT of 1). However, as we move diagonally introducing both a loading differential and factor trend the TWFE estimator gets significantly biased. In the bottom right cell it estimates an ATT of 2.35. The full sample estimator, however, is unbiased estimator when the treatment effect is homogeneous. The reason is straightforward. Homogeneous treatment effects imply zero correlation between treatment effect and factor realizations.

Table 2: Heterogeneous Treatment Effects

The loading difference μ and factor trend ν are the parameters of interest. All other parameter are the same as baseline setting. The mean of two-way fixed effect estimators is displayed in the first line without any parentheses, the mean of full sample estimator is displayed in the second line with round parentheses, and the mean of pre-treatment estimators is displayed in the third line with square parentheses.

TWFE Estimator (FS Estimator) [PT Estimator]		Loading Difference μ					
		0	0.2	0.4	0.6	0.8	1
Factor Trend ν	0	0.9810	1.0048	0.9698	1.0018	1.0122	0.9860
		(0.9895)	(1.0091)	(0.9948)	(0.9981)	(1.0156)	(0.9859)
		[0.9839]	[1.0009]	[0.9872]	[0.9971]	[1.0133]	[0.9827]
	0.5	1.0187	1.0356	1.1295	1.1334	1.2078	1.2415
		(0.9796)	(0.9509)	(0.9651)	(0.9459)	(0.9198)	(0.9476)
		[1.0222]	[0.9886]	[1.0167]	[0.9900]	[0.9724]	[0.9959]
	1	1.0046	1.1134	1.2058	1.3700	1.4069	1.5719
		(0.9160)	(0.9108)	(0.9176)	(0.9088)	(0.8905)	(0.9196)
		[1.0036]	[1.0040]	[1.0098]	[1.0154]	[0.9877]	[1.0160]
	1.5	1.0067	1.1228	1.3278	1.4740	1.6078	1.7869
		(0.8638)	(0.8522)	(0.8641)	(0.8494)	(0.8577)	(0.8593)
		[1.0042]	[0.9880]	[1.0074]	[0.9934]	[0.9936]	[1.0011]
	2	1.0125	1.2299	1.4341	1.6485	1.8406	2.0701
		(0.8126)	(0.8236)	(0.8177)	(0.8165)	(0.7948)	(0.8029)
		[1.0037]	[1.0095]	[1.0045]	[1.0104]	[0.9863]	[0.9943]
	2.5	1.0221	1.2675	1.4957	1.8047	2.0882	2.3723
		(0.7775)	(0.7756)	(0.7622)	(0.7974)	(0.7796)	(0.7882)
		[1.0187]	[1.0049]	[0.9808]	[1.0207]	[1.0156]	[1.0191]

Turning to Table 2, we now set the $\sigma_\Delta = 2$ so the treatment effect is heterogeneous and time-varying. The bias of the TWFE remains the same as with homogeneous treatment effect. Also, as expected, the pre-treatment estimator performs as in Table 1. However, treatment heterogeneity implies that the full sample estimator (within parentheses) becomes biased. The bias is only present when there is a factor trend (in the first row the estimated ATT of the full-sample estimator is 1). However as the factor trend increases and it reaches 2.5 the estimated ATT using the full sample estimator is 0.78. Finally, the estimated ATT using the full sample estimator is independent of the loading difference as we move along columns verifying corollary 2. Even when the treatment is randomly assigned ($\mu = 0$ in our case), the full sample estimator is still biased. On the contrary, as suggested by corollary 1, if the factor does not have a trend ($\nu = 0$ in our case), the full sample estimator remains unbiased no matter how unbalanced the treatment and control groups are.

In Table 3 we vary degree of treatment effect heterogeneity in the unit (columns) and

Table 3: Treatment Effect Heterogeneity Asymmetry

The treatment effect unit heterogeneity ψ and time heterogeneity ϕ are the parameters of interest. All other parameter are the same as baseline setting. The mean of two-way fixed effect estimators is displayed in the first line without any parentheses, the mean of full sample estimator is displayed in the second line with round parentheses, and the mean of pre-treatment estimators is displayed in the third line with square parentheses.

TWFE Estimator (**FS Estimator**) [**PT Estimator**]		TE Unit Heterogeneity ψ					
		0	0.2	0.4	0.6	0.8	1
TE Time Heterogeneity ϕ	0	1.3058	1.3183	1.3084	1.3133	1.3069	1.3071
		(1.0004)	(1.0002)	(0.9938)	(1.0007)	(1.0016)	(1.0018)
		[0.9995]	[0.9996]	[0.9921]	[1.0032]	[1.0006]	[1.0008]
	0.2	1.3364	1.3320	1.3163	1.3011	1.3406	
		(0.9508)	(0.9507)	(0.9590)	(0.9569)	(0.9551)	
		[1.0006]	[0.9971]	[1.0051]	[1.0011]	[1.0048]	
	0.4	1.2982	1.2910	1.3265	1.3445	1.3000	
		(0.9036)	(0.8956)	(0.9083)	(0.9050)	(0.9056)	
		[0.9932]	[0.9904]	[1.0018]	[1.0072]	[0.9950]	
	0.6	1.3183	1.3198	1.3011	1.3181	1.3062	
		(0.8652)	(0.8617)	(0.8456)	(0.8640)	(0.8610)	
		[1.0054]	[1.0029]	[0.9887]	[1.0008]	[0.9999]	
	0.8	1.3468	1.3248	1.2973	1.3057		
		(0.8226)	(0.8005)	(0.8023)	(0.8139)		
		[1.0224]	[0.9962]	[0.9802]	[0.9994]		
	1	1.3545					
		(0.7575)					
		[1.0132]					

time (rows) dimension. Regarding the full sample estimator (within parentheses), it performs the same irrespective of the amount of variance explained by unit heterogeneity (moving across columns), but performs worse the more variance is explained by time heterogeneity (moving between rows). Throughout this table we have kept the baseline assumption of a factor trend of $\nu = 1.5$ illustrating that the performance of the full sample estimator can either be degraded by the increasing factor trend (as in Table 2) or as in this table through changing the degree of time variation in the treatment effect.

Appendix B provides additional simulations. The first table in Appendix B is designed to verify that the necessary condition for the full sample bias is a correlation between factors and treatment effects. Along the rows we alter the correlation between the treatment effect and factors and along columns we alter the correlation between treatment effects and loadings. As expected, the performance of the full sample estimator does not change with the correlation of treatment effects with loadings, but substantially deteriorates once the treatment effects correlate with factor realizations.

Table Appendix B2 is designed to show that the full-sample bias is independent of the degree of random assignment and correlation between the treatment effect and loadings. Along both dimensions, the point estimate of the full sample estimator (within parentheses) is always around 0.85. This illustrates that the bad time control is independent from characteristics on the unit dimension.

5 Empirical Evidence

The goal of this section is to provide an empirical illustration of the importance of including factors in difference-in-difference models and evaluate the performance of the discussed estimators. We choose to focus on housing returns for several reasons. First, there is an extensive literature highlighting the importance of factors in real estate returns (see below). Second, the TWFE estimator is used frequently. Third, these studies often consider up to 10 years of data which means significant factor variation. Fourth, state interventions which cluster geographically are often considered, implying that parallel trends might not hold.

We identified two papers using data from the Federal Housing Finance Agency (FHFA) where all additional data needed is readily available. Table 4 of Favara and Imbs (2015) studies the impact of interstate bank branching deregulation on housing returns and Table B2 of Zevelev (2020) that studies the impact of allowing home equity loans on Texas house prices. We revisit these two results using the full-sample and pre-treatment estimators.

There are a significant amount of factors that have been shown to be relevant in housing returns. For example, models based on the arbitrage pricing theory (APT) with macroeconomic factors have been used in Chen et al. (1990), and Cotter et al. (2014), statistical factors (PCA) are employed by Titman and Warga (1986) while equity based factors such as the Fama-French factors, momentum and liquidity have studied in the real estate context by Peterson and Hsieh (1997), Hung and Glascock (2010) and Cannon and Cole (2010). Given the plethora of choice, we decided to pick the US aggregate economic factors used by Cotter, Gabriel and Roll (2014). The factors are: the loan-to-value ratio (LTV), mortgage-backed securities issuance (PrivMBS), payroll employment (Payems), equity markets (S&P500), industrial production (Indpro), PPI materials prices (PPIitm), personal Income (Income), consumer sentiment (Umcsent), building permits (Permit1), and the Federal Funds rate (Fedfunds).

Additionally, given that Zevelev (2021) uses the full-sample estimator with the oil price as a factor, we include the oil price in the possible set of factors.

5.1 Factor Selection

Bai and Ng (2002) note that when the factors are observable then the factor selection boils down to a model selection problem and the penalty term does not have to take into account the size of the cross-section. Additionally, the Bayesian Information Criterion (BIC) consistently estimates the number of factors unlike the Akaike Infomation Criterion (AIC) which selects too many factors. Therefore we use the BIC for model selection with penalty parameters based on the number of periods T.[14]

Table 4: Factor selection based on BIC

The factor in this table are selected based on BIC. We consider all possible factor combinations and select the specification with the minimum BIC. In Panel A we consider the full-sample estimator and in Panel B we consider the pre-treatment estimator. *unit*-level (*range*) indicates the data contains observations in *range* and minimun data unit is *unit*. For example, County-level (US) sample indicates the BIC value based on county-level data among all the united states.

Sample	Granularity	Optimal factor combination
Section A: Full sample estimator		
US	County-level	Fedfunds Indpro Payems Permit1 PPIitm S&P500
US	ZIP5-level	Fedfunds Indpro Payems Permit1 PPIitm Umcsent S&P500 Income OilPrice
Border States	ZIP5-level	Fedfunds Indpro Payems PPIitm S&P500 Income OilPrice
Border	ZIP5-level Border	Fedfunds Indpro Permit1 PPIitm Income OilPrice
Section B: Pre-treatment estimator		
US	County-level	Fedfunds Payems Income
US	ZIP5-level	Income Umcsent S&P500
Border State	ZIP5-level	Indpro Payems Permit1 PPIitm Income
Border	ZIP5-level	Indpro PPIitm

Table 4 presents the results of our model selection exercise. We perform the model selection at different levels of granularity since Favara and Imbs (2015) use county level data while Zevelev (2021) uses five digit zip code (Zip 5) data. Additionally, and importantly, Zevelev (2021) considers three different samples, a sample that contains all US states, a sample that only consider states which share a border with Texas and finally only counties that are close to Texas state borders.

[14]It is important to note that factor selection based on the BIC will not necessarily select the factors that have the largest impact on the treatment effect since it is based on maximizing the log-likelihood. It is possible that alternative factor combinations would affect the estimated treatment effects more.

Panel B uses the pre-treatment period for model selection. Thus, the first row of Panel A and B performs factor selection for the Favara and Imbs (2015) setting while the remaining rows of both panels perform model selection for the settings considered in Zevelev (2021). In general, the selected factors make intuitive sense. Interest rates (Fedfunds), industrial activity (Indpro) and production costs (PPIitm) are often selected.[15]

5.2 Placebo Interventions

In this section we use real estate return data combined with the deregulation index of Rice and Strahan (2010) to evaluate the performance of the different estimators under placebo interventions. An implication of using a misspecified benchmark model is some form of bias. In turn, the bias results in the over-rejection of the null hypothesis that the estimated treatment effect is equal to the true treatment effect. The TWFE estimator is likely to suffer from omitted factor bias and as a result this results in biased coefficients. In contrast, the full sample estimator, if well specified, reduces the omitted factor bias, but introduces the bad time control problem - another form of bias and hence a source of over-rejection. That is, the relative over-rejection rates of the two estimators are likely to vary according to the setting of the placebo interventions. Given enough data, time-invariant loadings, and that we have selected the "true" factors the pre-treatment estimator should not have a biased rejection rate.

To evaluate estimators in practice we do not only consider random interventions, but also interventions that are based on the deregulation index of Rice and Strahan (2010). Additionally, to highlight the weakness of the full-sample estimator we also consider placebo interventions where we introduce a dynamic treatment effect which decays over time. Finally, since we use the actual index (and therefore potentially true treatments), we consider the actual index but its influence on pre-index returns.

It is important to note that using two-way clustered standard errors does not deal with the over-rejection since the source of the bias that we are studying comes from the product of the unit and time dimension. Dealing with serial correlation in each of the dimensions separately does not deal with the product and therefore neither the omitted factor bias or

[15] We complement our BIC based model selection by examining which factors price the cross-section of real estate returns using Fama-MacBeth (1973) regressions. We consider multiple sample periods and different levels of granularity. Across specifications, we find evidence for Fedfunds, PPIitm and Income being priced in the cross-section.

the bad time control problem.

Specifically, we use county-level real estate return data from the Federal Housing Finance Agency (FHFA). The data spans the period from 1976 to 2020. In order to eliminate the impact from the potential true treatment effect, we only use data from 1976 to 1990 (the first act in Rice and Strahan (2010) is effective at Jan 1, 1994). We implement two different state-level placebo interventions. The first intervention is to randomly choose 25 states as treated group and the rest are control groups like in (Bertrand, Duflo, & Mullainathan, 2004). The second one is to first randomly pick a threshold between 1 to 4 and randomly pick a year between 1994 to 2005, and then define states with deregulation equal or higher than the threshold in that specific year as the treated states.[16]. In both interventions, all states are treated simultaneously (the intervention is non-staggered) and the treatment year is random but selected such that there are at least three years prior and post treatment. So the earliest treatment year is 1979 and the last possible treatment year is 1987. For each intervention, we only keep 7 years of data (three years pre- and post-treatment) to form placebo samples.

Besides two placebo interventions, we also implement two kinds of placebo treatments. In the first setting, we set the placebo treatment effect to a constant zero. In the second setting, we set the size of the placebo treatment effect into one standard deviation of the dependent variable at the treatment year and the placebo treatment effect decays to zero in three years. We examine the Wald test significance in which the null hypothesis is that the estimated treatment effect is equal to the placebo treatment effect. We set the significant level at 5% level and cluster standard errors at the state and year levels. In total, there are four specifications we investigate: random treated states - no treatment effects, deregulation states - no treatment effects, random treated states - decaying treatment effects, and deregulation treated states - decaying treatment effects.

We simulate 1,000 samples in each specification and estimate five different models. First, we use the plain TWFE estimator:

$$lnP_{i,t} - lnP_{i,t-1} = \gamma_i + \eta_t + \alpha D_i P_t + \varepsilon_{it},$$

where $lnP_{i,t} - lnP_{i,t-1}$ is the real estate return from year $t-1$ to t for county i, γ_i and η_t

[16] We remove cases where the number of treated state is smaller than 10 or greater than 40. (MacKinnon & Webb, 2017), show that standard errors are biased when the number of units within a cluster is smaller than 10.

are county and year fixed effects, respectively.

Second, we add a linear state time-trend to the TWFE estimator:

$$lnP_{i,t} - lnP_{i,t-1} = \gamma_i + \eta_t + \alpha D_i P_t + \lambda_i \times t + \varepsilon_{it},$$

where t is a linear time trend and λ_i is a state-specific loading.

Third, we use the full-sample estimator with economic factors:

$$lnP_{i,t} - lnP_{i,t-1} = \gamma_i + \eta_t + \alpha D_i P_t + \sum_{k=1}^{3} \lambda_{ki} F_{kt} + \varepsilon_{it},$$

where F_{kt} is the factor realization of factor k at time t and λ_i is the factor loading for state i. We use the three factors selected in 5.1.

Lastly, we use the pre-treatment estimator (described in Eq. 6 and 7) instead of full-sample estimator with state time-trend and optimally selected economic factors.

Table 5: Rejection rates of placebo Interventions

This table presents rejection rates estimated using TWFE estimator without controlling factors, full-sample and pre-treatment estimator using state-specific time trends and economic factors Fedfunds, PPIitm and Income. Rejection rates mean the percentage of significant Wald test results among 1,000 placebo samples in which the null hypothesis is the estimated treatment effect is equal to the placebo treatment effect. The standard errors calculated are clustered by state and year level. The significance level is set at 5%.

	Factor Control Strategy		
	No Factor	Time Trend	Economic Factors
Panel A: random treated states - no treatment effects			
Two-way fixed effect estimator	5.2%		
Full-sample estimator		5.9%	5.2%
Pre-treatment estimator		5.2%	4.6%
Panel B: deregulation states - no treatment effects			
Two-way fixed effect estimator	**34.1**%		
Full-sample estimator		30.3%	15.6%
Pre-treatment estimator		34.5%	35.3%
Panel C: random treated states - decaying treatment effects			
Two-way fixed effect estimator	5.2%		
Full-sample estimator		**90.0**%	**43.5**%
Pre-treatment estimator		5.2%	4.6%
Panel D: deregulation states - decaying treatment effects			
Two-way fixed effect estimator	34.1%		
Full-sample estimator		89.6%	38.6%
Pre-treatment estimator		**34.5**%	**7.0**%

We expect that the TWFE estimator will be unbiased and reject 5% samples when treatments are randomly assigned. If the TWFE estimator over-rejects, it suggests the treatment assignment is not random and there may exist missing factors. Moreover, we expect the full-sample estimator will overwhelmingly over-reject when there is a decaying treatment effect, compared to the case of no treatment effects. If factors are correctly selected, we expect the pre-treatment estimator to reject 5% samples even when treatment is NOT randomly assigned.

The results are presented in Table 5. In Panel A when treated states are randomly picked all of the estimators have a rejection rate of around 5%. However, in Panel C where we maintain random treatment, but introduce a decaying treatment effect the rejection rate of the full-sample estimator increases dramatically. The full-sample estimator with a time trend rejects 90% of the time while it rejects 43.5% of the time with economic factors.

In Panel B, we use interventions based on the deregulation index. In this case, the TWFE has a rejection rate of 34.1% suggesting that the parallel assumption may not hold in this setting.

Panel D considers interventions based on the deregulation index with decaying treatment effects. The rejection rate of the full-sample estimator is 89.8% and 38.6% when a time trend and economic factors are used, respectively. Using economic factors the pre-treatment estimator has a rejection rate of 7%. However, using state-level trends in conjunction with the pre-treatment estimator does not result in an improvement over the TWFE estimator.

Figure 1 presents histograms of the t values of our estimators. The top row considers randomly assigned interventions while the bottom row considers interventions based on the deregulation index. The left most column displays the t values of the TWFE estimator, the middle column considers the full-sample estimator while the right most column considers the pre-treatment estimator. All of the plots in this figure are based on the setting with decaying treatment effects and economic factors. Under random assignment, both the TWFE estimator and the pre-treatment estimator perform well. However, even under random assignment the full-sample estimator is biased. The performance of the TWFE and full-sample estimators deteriorates substantially when we use intervention based on the deregulation index. The distribution of the full-sample estimator is both skewed and flat while the distribution of the TWFE estimator is right skewed. The pre-treatment estimator is substantially better than the other two and appears close to the student´s t distribution,

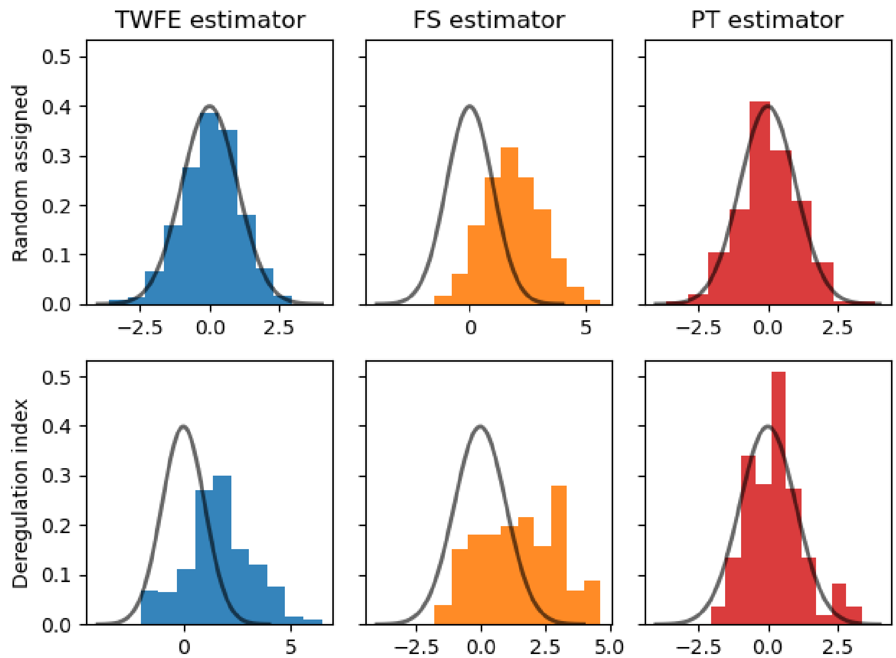

Figure 1: The distribution of t-values with decaying treatment effects

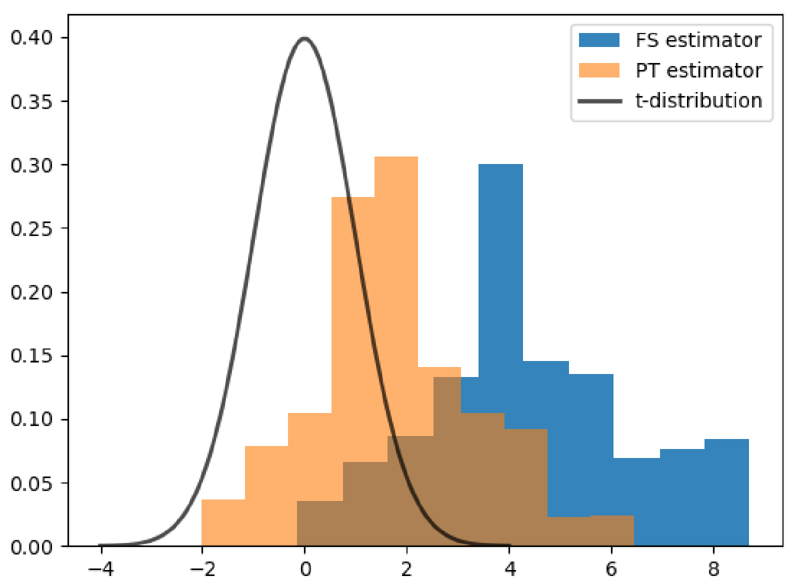

Figure 2: The distribution of t-values when treated states are determined by the deregulation index, state time trends are controlled and decaying treatment effects

but appears less smooth than under random assignment.

Figure 2 presents histograms of our t values when time trends are used instead of economic factors. The commonly used augmentation of the TWFE estimator appears to be more biased.

Taken together our placebo interventions suggest: (1) the choice of method matters substantially, (2) the bad time control problem can be substantial - almost doubling the rejection rate and (3) the pre-treatment estimator with economic factors performs close to as expected.

5.3 Favara and Imbs (2015)

For economists and policy makers it is important to understand the impact of local credit expansions on local asset prices. An increase in local house prices following a local credit expansion provides evidence that non-local assets are not perfect substitutes. Favara and Imbs (2015) use the state deregulation index introduced by Rice and Strahan (2010) to relate increases in local credit supply to local house prices.[17] Using a staggered difference-in-difference they find that an increase in the deregulation index results in an increase in local house prices by 1.2%.[18]

As with many quasi-natural experiments it is likely that deregulation is not randomly assigned. Indeed, Kroszner and Strahan (1999) study the causes of interstate banking deregulation and comment "We find that deregulation occurs earlier in states with fewer small banks, in states where small banks are financially weaker, and in states with more small, presumably bank-dependent, firms. Also, a larger insurance industry delays deregulation when banks may compete in the sale of insurance products. Interest group factors related to the relative strength of potential winners (large banks and small firms) and losers (small banks and the rival insurance firms) thus can explain the timing of branching deregulation across states." This suggests that treated and control units may have different loadings to factors.

We incorporate a factor structure into Eq. (2) of Favara and Imbs (2015). This implies

[17]The effect of interstate banking deregulation has been extensively studied, among other things it has been documented to lead to less pronounced business cycles (Morgan, Rime and Strahan, 2004) per capital growth in Income and output (Jayaratne and Strahan, 1996), credit costs of borrowers (Rice and Strahan, 2010), lower Income inequality (Beck, Levine and Levkov, 2010) and reallocation across sectors (Acharya, Imbs and Sturgess, 2011)

[18]Other papers that analyze house prices in a difference-in-difference setting includes Blickli (2018) and Di Maggio and Kermani (2017).

we estimate the following,

$$lnP_{c,t} - lnP_{c,t-1} = \beta_1 D_{s,t-1} + \beta_2 D_{s,t-1} \times \eta_c^s + \beta_3 \mathbf{X}_{c,t} + \alpha_c + \gamma_t + \sum_{k=1}^{K} \lambda_{c,k} \times F_{t,k} + \varepsilon_{c,t}$$

where $P_{c,t}$ denotes house price index, $D_{s,t-1}$ denotes deregulation index, η_c^s denotes housing supply (in)elasticity, $\mathbf{X}_{c,t}$ denotes county-level control variables, α_c and γ_t are county and year fixed effects respectively. Indexes c refers to counties, s to states, and t to years. We add factors based on the selection procedure described above where $\lambda_{c,k}$ refers to the loading to factor k of county c and $F_{t,k}$ is the factor realization of factor k at date t.

Table 6: Incorporating factors into Favara and Imbs (2015)

This table presents the original results of Favara and Imbs (2015) and our full sample estimators as well as pre-treatment estimators. Column (1) replicates column (3) of table 4 in Favara and Imbs (2015). Columns (2) and (3) add the factors with significant premia. Columns (4) and (5) introduce the factors selected in Panel A of Table 3. Columns (6) and (7) add the factors selected in Panel B of Table 3. The standard errors reported are clustered by state. ***, **, * represent significance at 10%, 5%, and 1% respectively.

Variable	Original (1)	FS Est. (2)	PT Est. (3)	FS Est. (4)	PT Est. (5)	FS Est. (6)	PT Est. (7)
Deregulation index	0.0122***	-0.0002	0.0009	-0.0034	-0.0029	0.0005	0.0017
	(0.002)	(0.006)	(0.006)	(0.013)	(0.014)	(0.006)	(0.006)
Deregulation index	-0.005***	-0.003	-0.004	-0.002	-0.024	-0.003	-0.003
× house supply elasticity	(0.000)	(0.002)	(0.0024)	(0.004)	(0.005)	(0.002)	(0.002)
County-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County & Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fedfunds × State dummies		✔	✔	✔	✔	✔	✔
Indpro × State dummies				✔	✔		
Payems × State dummies				✔	✔	✔	✔
Permit1 × State dummies				✔	✔		
PPIitm × State dummies		✔	✔	✔	✔		
Umcsent × State dummies							
SP500 × State dummies				✔	✔		
Income × State dummies		✔	✔			✔	✔

The results are presented in Table 6. In all specifications, the introduction of the factors renders the estimated treatment effect economically and statistically insignificant.

Since factor selection could be argued is somewhat arbitrary we performed the same analysis using all possible factor combinations while considering up to nine factors. In Table 7 we report the number of factors used and the number of significant treatment effects and the total of factor combinations. We present results separately for the full and pre-treatment estimators. Given the shorter time horizon there is a lower maximum number of factors for the pre-treatment estimators and PrivMBS is not available before

Table 7: Combinatorial Factor Selection

This table uses all possible factor combinations and record the number of significant point estimates of the treatment effect as well as the number of factors. In the full sample case, we use all 9 factors while in the pre-treatment case we exclude PrivMBS due to data availability.

Panel A: FS estimator									
# of factors	9	8	7	6	5	4	3	2	1
# significant	0/1	0/9	0/36	0/84	1/126	7/126	20/84	25/36	9/9
	[0%]	[0%]	[0%]	[0%]	[0.8%]	[5.6%]	[23.8%]	[69.4%]	[100%]
Panel B: PT estimator									
# of factors	3	2	1						
# significant	29/56	14/28	6/8						
	[51.8%]	[50%]	[75%]						

1994 implying that it cannot be used with the pre-treatment estimator. Using the full-sample estimator only 7 out of 126 combinations are significant when using four factors. For the pre-treatment estimator when we include 3 factors only 29 out of 56 combinations are statistically significant.

To examine whether there are systematic differences in loadings we display the deregulation index of Rice and Strahan (2010) taken from Favara and Imbs (2015) in Figure 3 and in Figure 4. we display our estimated state loadings. Examining the loadings visually it seems as if they cluster geographically.

5.4 Zevelev (2021)

Zevelev (2021) studies the effect of a constitutional amendment in Texas that legalized home equity loans. He finds that this increases Texas house prices by 4%. We introduce a factor structure into the Zevelev's equation (static DID) which implies we estimate,

$$y_{i,s,t} = \alpha_i + \theta_t + \beta_{DID} Texas_s \times Post_t + \Gamma X_{i,s,t} + \sum_{k=1}^{K} \lambda_{s,k} \times F_{t,k} + \varepsilon_{i,s,t}$$

where $y_{i,s,t}$ is log real house price index. Index i refers to five digit zip code, s refers to state and year t. α_i is zip code fixed effect while θ_t are time fixed effects. $Texas_s$ is a dummy variable that takes the value 1 if it is the state Texas and $Post_t$ takes the value 1 if it is in post-treatment preiod ($t \geq 1998$).

Interestingly Zevelev (2021) introduces a factor structure in some of his specifications. He controls for the interaction between the oil price and MSA dummies as well as a time-trend that is interacted with state dummies.

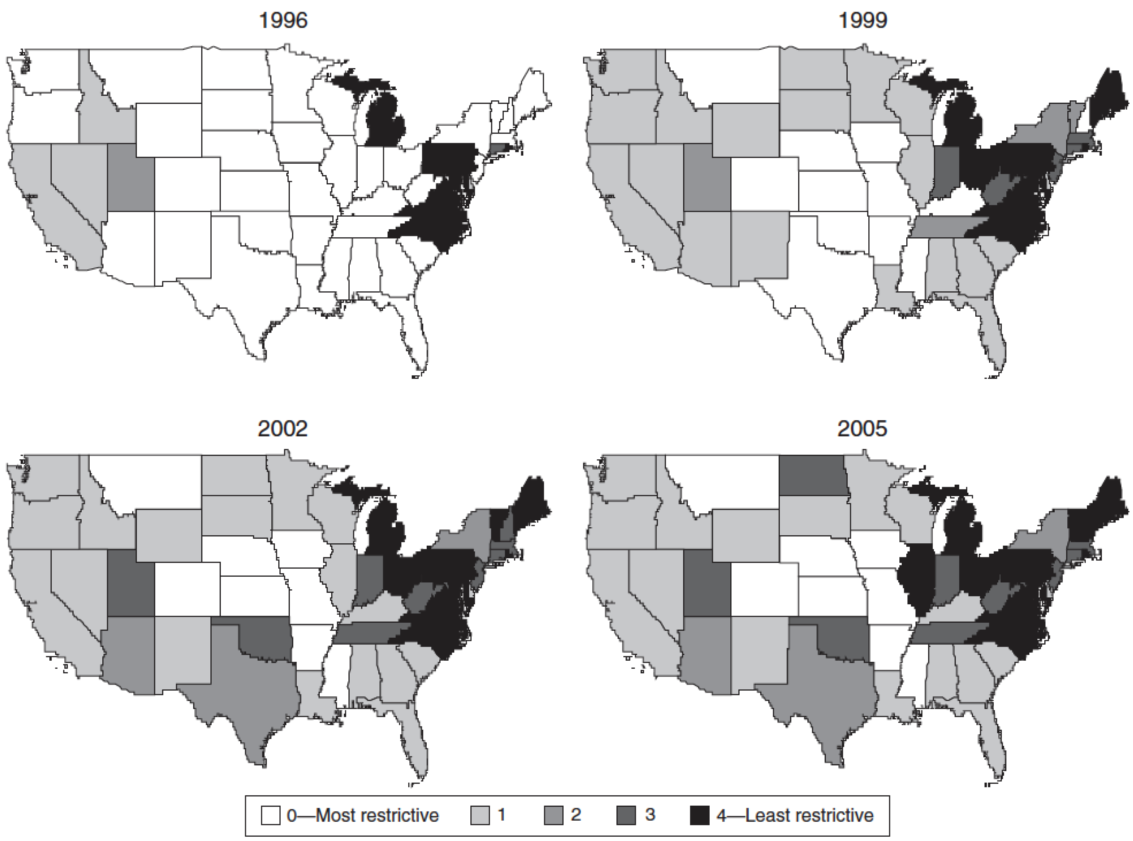

FIGURE B1. RICE-STRAHAN (2010) DEREGULATION INDEX BY STATE AND YEAR

Source: Rice and Strahan (2010).

Figure 3: Interstate Branching Deregulation Index

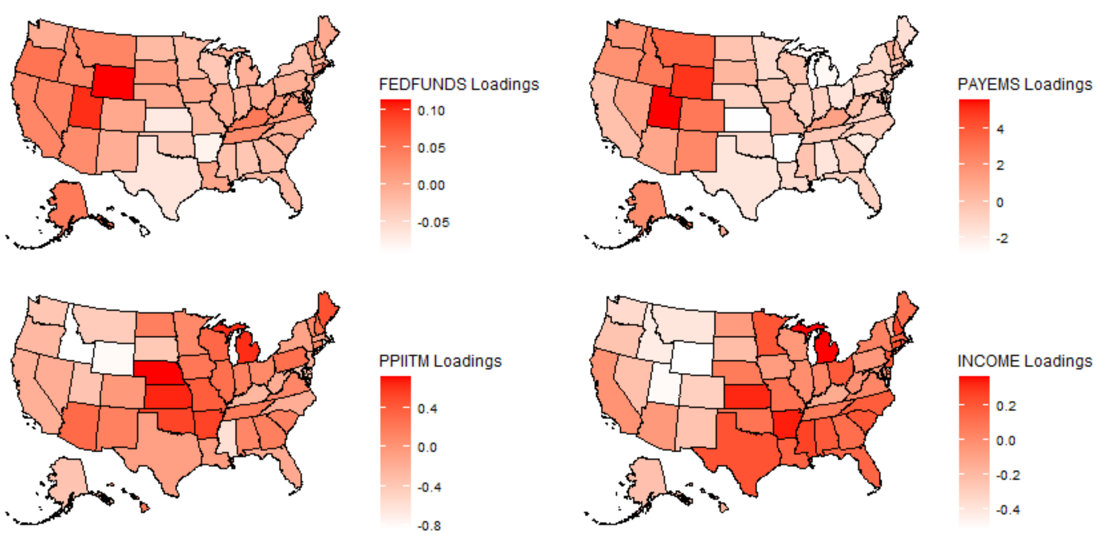

Figure 4: Estimated State Loadings

Table 8: Incorporating factors into Zevelev (2021)

This table presents the original results of Zevelev (2021) and our full sample estimators as well as pre-treatment estimators. Column (1), (5), (9) replicates column (1) to (3) of table B.2 in Zevelev (2021). Columns (2), (6), (10) remove factor controls from Zevelev (2021). Columns (3), (7), (11) introduce the factors selected in Panel A of Table 4. Columns (4), (8), (12) add the factors selected in Panel B of Table 4. The standard errors of full sample esitmators and pre-treatment estimators are clustered by zip-code level. ***, **, * represent significance at 10%, 5%, and 1% respectively.

Variable	Original Paper	Without Factor	FS Estimator	PT Estimator
Panel A: United States	(1)	(2)	(3)	(4)
TexasPost	0.0350***	-0.0387	-0.0123**	-0.0147***
	(0.0099)	(0.0040)	(0.0062)	(0.0040)
Zipcode & Year FE	✔	✔	✔	✔
State time trend	✔			
Oil× MSA dummies	✔		✔	
Fedfunds × State dummies			✔	✔
Indpro × State dummies			✔	
Payems × State dummies			✔	
Permit1 × State dummies			✔	
PPIitm × State dummies			✔	
Umcsent × State dummies			✔	✔
SP500 × State dummies			✔	✔
Income × State dummies			✔	
Panel B: Border States	(5)	(6)	(7)	(8)
TexasPost	0.0616***	0.0015	0.0372***	0.0390***
	(0.0221)	(0.0051)	(0.0041)	(0.0048)
Zipcode & Year FE	✔	✔	✔	✔
State time trend	✔			
Oil× MSA dummies	✔		✔	
Fedfunds × State dummies			✔	
Indpro × State dummies			✔	✔
Payems × State dummies			✔	✔
Permit1 × State dummies				✔
PPIitm × State dummies			✔	✔
Umcsent × State dummies				
SP500 × State dummies			✔	
Income × State dummies			✔	✔
Panel C: Border Counties	(9)	(10)	(11)	(12)
TexasPost	0.0476**	0.0024	0.0032	0.0450***
	(0.0151)	(0.0113)	(0.0095)	(0.0097)
Zipcode & Year FE	✔	✔	✔	✔
State time trend	✔			
Oil× MSA dummies	✔		✔	
Fedfunds × State dummies			✔	
Indpro × State dummies			✔	
Payems × State dummies				
Permit1 × State dummies			✔	
PPIitm × State dummies			✔	✔
Umcsent × State dummies				
SP500 × State dummies				
Income × State dummies			✔	✔

In Table 8 we introduce economic factors into Table B2 of Zevelev (2021). Panel A provides results when we consider the entire United States. We show that without any factor controls Zevelev's result changes sign and is statistically significant. This highlights the importance of including factor controls. Additionally, the point estimates of both the full and pre-treatment estimators are negative and significant.

In Panel B, we present replication results for border states. Without factor controls the treatment effect is rendered insignificant. Introducing optimally selected factors reduces the treatment effects from 0.0616 to 0.039 (pre) and 0.0372 (full), but the point estimates remain statistically significant.

Finally, in Panel C we replicate the results for border counties. Again, without factor controls the treatment effect is rendered insignificant. This is also the case for the full sample estimator. Interestingly, the point estimate of the pre treatment estimator is very close to what is found in the original paper. Although the introduction of factors provides mixed results, it is clear that they are essential for the estimated treatment effects.

6 Conclusion

For almost 30 years factor models were the standard methodology used to analyze housing returns. The advent of quasi-experimental techniques that offer improved identification has resulted in a shift in research methodology from factor models to difference-in-differences estimators. We show that it is far from obvious how to incorporate the factor model into the difference-in-differences framework. The TWFE estimator is generally biased when factors are omitted, but so is the full-sample estimator. The TWFE estimator is preferred when assignment is close to random while the full sample estimator is unbiased when treatment is time-invariant.

Researchers frequently augment the TWFE estimator to control for factor variation. We show that the resulting estimators often suffer from the bad time control problem. In our placebo analysis we find that the full sample estimator performs worse than the TWFE estimator suggesting that the bad time control problem is significant when studying housing returns. Further, we revisit the results of Favara and Imbs (2015) and Zevelev (2021) while incorporating relevant factors. In both cases we find that the factor model explains significant variation and should therefore be included. Additionally, depending on method and specification the estimated treatment effect may be significantly changed. Overall, this paper provides methods for incorporating factor models into difference-in-differences regressions while showing that it is also necessary when studying housing returns.

Future work should consider other dependent variables which have been shown to have a factor structure where difference-in-differences are often used. Given the importance of factors for interest rates (e.g., Litterman and Scheinkman, 1991), we suspect that in these applications it is particularly beneficial to augment the difference-in-differences analysis to control for factor variation.

Appendix A Proofs

A.1 Useful lemmas

Denote the sample mean of treatment effects on treated $\overline{\Delta}^{\text{ATT}}$ as

$$\overline{\Delta}^{\text{ATT}} = \frac{\sum_{i,t} D_i P_t \Delta_{it}}{\sum_{i,t} D_i P_t}$$

Lemma 1. *The expectation of the sample mean of treatment effects on treated is equal to the average treatment effect on treated.*

$$\mathbb{E}[\overline{\Delta}^{ATT}] = \alpha^{ATT}$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\overline{\Delta}^{\text{ATT}}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t} D_i P_t \Delta_{it}}{\sum_{i,t} D_i P_t}\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\left.\frac{\sum_{i,t} D_i P_t \Delta_{it}}{\sum_{i,t} D_i P_t}\right| \boldsymbol{D}, \boldsymbol{P}\right]\right] \\
=&\mathbb{E}\left[\frac{1}{\sum_{i,t} D_i P_t} \sum_{i,t} \mathbb{E}\left[D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right]\right]
\end{aligned}$$

Since $Y_{it}(0)$ and $Y_{it}(1)$ are independent from D_j and P_s when $j \neq i$ and $s \neq t$,

$$\begin{aligned}
&\mathbb{E}\left[D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\mathbb{E}\left[D_i P_t \Delta_{it} \middle| D_i, P_t\right] \\
=&D_i P_t\, \mathbb{E}\left[\Delta_{it} \middle| D_i, P_t\right] \\
=&\mathbb{1}_{\{D_i P_t = 1\}}\, \mathbb{E}\left[\Delta_{it} \middle| D_i P_t = 1\right]
\end{aligned}$$

In reason that $\mathbb{E}\left[\Delta_{it} \middle| D_i P_t = 1\right]$ is a constant,

$$\begin{aligned}
&\mathbb{E}\left[\overline{\Delta}^{\text{ATT}}\right] \\
=&\mathbb{E}\left[\frac{1}{\sum_{i,t} D_i P_t} \sum_{i,t} \mathbb{E}\left[D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right]\right] \\
=&\mathbb{E}\left[\Delta_{it} \middle| D_i P_t = 1\right] \mathbb{E}\left[\frac{\sum_{i,t} \mathbb{1}_{\{D_i P_t = 1\}}}{\sum_{i,t} D_i P_t}\right] \\
=&\alpha^{\text{ATT}}
\end{aligned}$$

□

Denote the sample mean of λ for the treated (control) group $\overline{\lambda}_D$ ($\overline{\lambda}_C$) as

$$\overline{\lambda}_D = \frac{\sum_i D_i \lambda_i}{\sum_i D_i} \qquad \overline{\lambda}_C = \frac{\sum_i (1-D_i)\lambda_i}{\sum_i 1 - D_i}$$

Denote the sample mean of F for the post-treatment (pre-treatment) period $\overline{F}_{\text{post}}$ ($\overline{F}_{\text{pre}}$) as

$$\overline{F}_{\text{post}} = \frac{\sum_t P_t F_t}{\sum_i P_t} \qquad \overline{F}_{\text{pre}} = \frac{\sum_i (1-P_t) F_t}{\sum_i 1 - P_t}$$

Using the method similar to lemma 1, it is not hard to verify that

$$\begin{aligned}
\mathbb{E}\left[\overline{\lambda}_D\right] &= \mathbb{E}\left[\lambda_i \middle| D_i = 1\right] \\
\mathbb{E}\left[\overline{\lambda}_C\right] &= \mathbb{E}\left[\lambda_i \middle| D_i = 0\right] \\
\mathbb{E}\left[\overline{F}_{\text{post}}\right] &= \mathbb{E}\left[F_t \middle| P_t = 1\right] \\
\mathbb{E}\left[\overline{F}_{\text{pre}}\right] &= \mathbb{E}\left[F_t \middle| P_t = 0\right]
\end{aligned}$$

Denote the sample covariance between factor realizations F_t and individual treatment effect Δ_{it} as

$$Q_{F,\Delta}^{\text{ATT}} = \frac{\sum_{i,t} D_i P_t F_t (\Delta_{it} - \overline{\Delta}^{\text{ATT}})}{\sum_{i,t} D_i P_t}$$

Lemma 2. *The expectation of the sample covariance between factor realizations and individual treatment effect is equal to the overall covariance adjusted by degree of freedom.*

$$\mathbb{E}\left[Q_{F,\Delta}^{ATT}\right] = \left(1 - \frac{1}{E\left[\sum_{it} D_i P_t\right]}\right) cov\left(F_t, \Delta_{it} \middle| D_i P_t = 1\right)$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[Q_{F,\Delta}^{\text{ATT}}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t} D_i P_t F_t (\Delta_{it} - \overline{\Delta}^{\text{ATT}})}{\sum_{i,t} D_i P_t}\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{i,t} D_i P_t F_t (\Delta_{it} - \overline{\Delta}^{\text{ATT}})}{\sum_{i,t} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}\right]\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t} \mathbb{E}\left[D_i P_t F_t (\Delta_{it} - \overline{\Delta}^{\text{ATT}}) \middle| \boldsymbol{D}, \boldsymbol{P}\right]}{\sum_{i,t} D_i P_t}\right]
\end{aligned}$$

Since both F_t and Δ_{it} are independent from D_j and P_s when $j \neq i$ and $s \neq t$,

$$
\begin{aligned}
&\mathbb{E}\left[D_i P_t F_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\mathbb{1}_{\{D_i P_t = 1\}} \mathbb{E}\left[F_t \Delta_{it} \middle| D_i = 1, P_t = 1\right] \\
=&\mathbb{1}_{\{D_i P_t = 1\}} \left(\operatorname{cov}\left(F_t, \Delta_{it} \middle| D_i = 1, P_t = 1\right) + \mathbb{E}\left[F_t \middle| P_t = 1\right] \mathbb{E}\left[\Delta_{it} \middle| D_i = 1, P_t = 1\right]\right)
\end{aligned}
$$

On the other hand, Δ_{it} is independent from F_s when $s \neq t$ and therefore $E[F_t \Delta_{js}] = E[F_t]E[\Delta_{js}] + \mathbb{1}_{\{t=s\}} \operatorname{cov}(F_t, \Delta_{jt})$. Thus, we obtain

$$
\begin{aligned}
&\sum_{i,t} \mathbb{E}\left[D_i P_t F_t \overline{\Delta}^{\text{ATT}} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\sum_{i,t} \mathbb{E}\left[D_i P_t F_t \frac{\sum_{j,s} D_j P_s \Delta_{js}}{\sum_{j,s} D_j P_s} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\frac{1}{N_D T_P} \mathbb{E}\left[\left(\sum_{(i,t):D_i P_t = 1} F_t\right)\left(\sum_{(j,s):D_j P_s = 1} \Delta_{js}\right)\right] \\
=&\frac{1}{N_D T_P}\left(\mathbb{E}\left[\sum_{(i,t):D_i P_t = 1} F_t\right]\right)\left(\mathbb{E}\left[\sum_{(j,s):D_j P_s = 1} \Delta_{js}\right]\right) \\
&+ \frac{1}{N_D T_P}\left(\mathbb{E}\left[\sum_{(i,t):D_i P_t = 1} \operatorname{cov}(F_t, \Delta_{js})\right]\right) \\
=&N_D T_P \, \mathbb{E}\left[F_t \middle| P_t = 1\right] \mathbb{E}\left[\Delta_{it} \middle| D_i = 1, P_t = 1\right] + \operatorname{cov}\left(F_t, \Delta_{it} \middle| D_i = 1, P_t = 1\right)
\end{aligned}
$$

Therefore, the expectation of sample covariance is

$$
\begin{aligned}
&\mathbb{E}\left[Q_{F,\Delta}^{\text{ATT}}\right] \\
=&\mathbb{E}\left[\frac{N_D T_P \, \mathbb{E}\left[F_t \middle| P_t = 1\right] \mathbb{E}\left[\Delta_{it} \middle| D_i = 1, P_t = 1\right] + N_D T_P \operatorname{cov}\left(F_t, \Delta_{it} \middle| D_i = 1, P_t = 1\right)}{N_D T_P}\right] \\
&- \mathbb{E}\left[\frac{N_D T_P \, \mathbb{E}\left[F_t \middle| P_t = 1\right] \mathbb{E}\left[\Delta_{it} \middle| D_i = 1, P_t = 1\right] + \operatorname{cov}\left(F_t, \Delta_{it} \middle| D_i = 1, P_t = 1\right)}{N_D T_P}\right] \\
=&\left(1 - \frac{1}{\mathbb{E}\left[\sum_{i,t} D_i P_t\right]}\right) \operatorname{cov}\left(F_t, \Delta_{it} \middle| D_i = 1, P_t = 1\right)
\end{aligned}
$$

□

A.2 Proof of Proposition 1

The traditional way to estimate treatment effect is the two-way fixed effect difference-in-difference estimator. The definition of two-way fixed effect is the following.

$$(\hat{\alpha}^{\text{TWFE}}, \hat{\gamma}_i^{\text{TWFE}}, \hat{\eta}_t^{\text{TWFE}}) = \underset{\alpha,\gamma,\eta}{argmin}\Big\{ \sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{it} - \gamma_i - \eta_t - \alpha D_i P_t)^2 \Big\} \tag{9}$$

Given observed $\{D_i\}$ and $\{P_t\}$, we can regress $D_i P_t$ on unit dummies, and time dummies. The residuals are defined as u_{it}^{TWFE}.

$$D_i P_t = \kappa_i^{\text{TWFE}} + \zeta_t^{\text{TWFE}} + u_{it}^{\text{TWFE}} \tag{10}$$

It can be verified that the residual u_{it}^{TWFE} is the two-way demeaned $D_i P_t$,

$$u_{it}^{\text{TWFE}} = (D_i - \overline{D})(P_t - \overline{P})$$
$$\text{where} \quad \overline{D} = \frac{1}{N}\sum_{i=1}^{N} D_i \quad \overline{P} = \frac{1}{T}\sum_{t=1}^{N} P_t$$

In reason that u_{it}^{TWFE} is the residual in regression (10), we can obtain

$$\forall t, \quad \sum_i u_{it}^{\text{TWFE}} = 0 \quad \Rightarrow \quad \sum i, t u_{it}^{\text{TWFE}} Y_{1t} = 0$$
$$\forall i, \quad \sum_t u_{it}^{\text{TWFE}} = 0 \quad \Rightarrow \quad \sum i, t u_{it}^{\text{TWFE}} Y_{i1} = 0$$
$$\forall (i,t) : D_i P_t = 1 \quad u_{it}^{\text{TWFE}} = (1 - \overline{D})(1 - \overline{P})$$

Then we can derive that

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} Y_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} Y_{it}(0) \middle| \boldsymbol{D}, \boldsymbol{P}\right] + \mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} (Y_{it}(0) - Y_{i1}(0) - Y_{1t}(0) + Y_{11}(0)) \middle| \boldsymbol{D}, \boldsymbol{P}\right] + \mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} (\lambda_i - \lambda_1)(F_t - F_1) \middle| \boldsymbol{D}, \boldsymbol{P}\right] + \mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} \lambda_i F_t \middle| \boldsymbol{D}, \boldsymbol{P}\right] + \mathbb{E}\left[\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}\right]
\end{aligned}$$

Following the Frisch-Waugh-Lovell theorem, the two-way fixed effect estimator can be written as

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\mathrm{TWFE}}\middle|\boldsymbol{D},\boldsymbol{P}\right]\\
=&\mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\mathrm{TWFE}}Y_{it}}{\sum_{i,t}u_{it}^{\mathrm{TWFE}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P}\right]\\
=&\mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\mathrm{TWFE}}D_iP_t\Delta_{it}}{\sum_{i,t}u_{it}^{\mathrm{TWFE}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P}\right]+\mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\mathrm{TWFE}}\lambda_iF_t}{\sum_{i,t}u_{it}^{\mathrm{TWFE}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P}\right]\\
=&\mathbb{E}\left[\frac{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{TWFE}}\Delta_{it}}{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{TWFE}}}\middle|\boldsymbol{D},\boldsymbol{P}\right]+\mathbb{E}\left[\frac{\sum_i\Big((D_i-\overline{D})\lambda_i\Big)\sum_t\Big((P_t-\overline{P})F_t\Big)}{\sum_{(i,t):D_iP_t=1}(1-\overline{D})(1-\overline{P})}\middle|\boldsymbol{D},\boldsymbol{P}\right]\\
=&\mathbb{E}\left[\frac{\sum_{(i,t):D_iP_t=1}\Delta_{it}}{\sum_{(i,t):D_iP_t=1}1}\middle|\boldsymbol{D},\boldsymbol{P}\right]+\mathbb{E}\left[\frac{\sum_i\Big((D_i-\overline{D})\lambda_i\Big)\sum_t\Big((P_t-\overline{P})F_t\Big)}{NT\overline{D}(1-\overline{D})\overline{P}(1-\overline{P})}\middle|\boldsymbol{D},\boldsymbol{P}\right]\\
=&\mathbb{E}\left[\overline{\Delta}^{\mathrm{ATT}}\middle|\boldsymbol{D},\boldsymbol{P}\right]+\mathbb{E}\left[(\overline{\lambda}_{\mathrm{D}}-\overline{\lambda}_{\mathrm{C}})(\overline{F}_{\mathrm{Post}}-\overline{F}_{\mathrm{Pre}})\middle|\boldsymbol{D},\boldsymbol{P}\right]
\end{aligned}
$$

According to lemma 1, we can get

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\mathrm{TWFE}}\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[\hat{\alpha}^{\mathrm{TWFE}}\middle|\boldsymbol{D},\boldsymbol{P}\right]\right]\\
=&\alpha^{\mathrm{ATT}}+(E[\lambda_i|D_i=1]-E[\lambda_i|D_i=0])(E[F_t|P_t=1]-E[F_t|P_t=0])
\end{aligned}
$$

A.3 Proof of Proposition 2

Two-way fixed effect estimator totally ignores the presence of the factor structure. A straight-forward idea is to add all the factors as control variables. To be explicit, full sample estimator is defined as

$$
(\hat{\alpha}^{\mathrm{FS}},\hat{\gamma}_i^{\mathrm{FS}},\hat{\eta}_t^{\mathrm{FS}},\hat{\lambda}_i^{\mathrm{FS}})=\underset{\alpha,\gamma,\eta,\lambda}{argmin}\Big\{\sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{it}-\gamma_i-\eta_t-\lambda_iF_t-\alpha D_iP_t)^2\Big\} \tag{11}
$$

Compared to TWFE estimator, a full sample estimator adds observed factor realizations as covariates. Suppose treatment group D_i, treatment time P_t and factor realizations F_t is given, we regress D_iP_t on unit dummies, time dummies and factor realizations, and define the residual as u_{it}^{FS}.

$$
D_iP_t=\kappa_i^{\mathrm{FS}}+\zeta_t^{\mathrm{FS}}+\xi_i^{\mathrm{FS}}F_t+u_{it}^{\mathrm{FS}} \tag{12}
$$

In reason that u^{FS} is the residual of regression (12), it satisfies the following equations.

$$\forall t, \quad \sum_i u_{it}^{\text{FS}} = 0$$
$$\forall i, \quad \sum_t u_{it}^{\text{FS}} = 0$$
$$\forall i, \quad \sum_t u_{it}^{\text{FS}} F_t = 0$$

Then it implies that

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} Y_{it} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} Y_{it}(0) \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right] + \mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} (Y_{it}(0) - Y_{i1}(0) - Y_{1t}(0) + Y_{11}(0)) \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right] + \mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} (\lambda_i - \lambda_1)(F_t - F_1) \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right] + \mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right] \\
=&\mathbb{E}\left[\sum_{i,t} u_{it}^{\text{FS}} D_i P_t \Delta_{it} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F}\right]
\end{aligned}$$

Due to the degree of freedom, there exists multiple solutions of regression (12). However, it does not change the values of residuals, which is what we are interested in. One of the possible solutions is

$$\begin{aligned}
&\xi_i^{\text{FS}} = 0 && \text{if } D_i = 0 \\
&\xi_i^{\text{FS}} = \overline{P}(1 - \overline{P}) \frac{\overline{F}_{\text{post}} - \overline{F}_{\text{pre}}}{\sigma_F^2} && \text{if } D_i = 1 \\
&\kappa_i^{\text{FS}} = 0 && \text{if } D_i = 0 \\
&\kappa_i^{\text{FS}} = \overline{P} - \xi_i^{\text{FS}} \overline{F} && \text{if } D_i = 1 \\
&\zeta_t^{\text{FS}} = -\overline{\kappa}^{\text{FS}} - \overline{\xi}^{\text{FS}} F_t && \text{if } P_t = 0 \\
&\zeta_t^{\text{FS}} = -\overline{\kappa}^{\text{FS}} - \overline{\xi}^{\text{FS}} F_t + \overline{D} && \text{if } P_t = 1
\end{aligned}$$

where

$$\sigma_F^2 = \frac{1}{T} \sum_{t=1}^{T} (F_t - \overline{F})^2 \qquad \overline{\kappa}^{\text{FS}} = \frac{1}{N} \sum_{i=1}^{N} \kappa_i^{\text{FS}} \qquad \overline{\xi}^{\text{FS}} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{\text{FS}}$$

In this case, we can estimate the value of a full sample estimator. Unfortunately, the full sample estimator is biased

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\mathrm{FS}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]\\
=&\mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\mathrm{FS}}Y_{it}}{\sum_{i,t}u_{it}^{\mathrm{FS}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]\\
=&\mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\mathrm{FS}}D_iP_t\Delta_{it}}{\sum_{i,t}u_{it}^{\mathrm{FS}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]\\
=&\mathbb{E}\left[\frac{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{FS}}\overline{\Delta}^{\mathrm{ATT}}}{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{FS}}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]+\mathbb{E}\left[\frac{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{FS}}(\Delta_{it}-\overline{\Delta}^{\mathrm{ATT}})}{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{FS}}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]\\
=&\mathbb{E}\left[\overline{\Delta}^{\mathrm{ATT}}\frac{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{FS}}}{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{FS}}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]+\mathbb{E}\left[\frac{\sum_{(i,t):D_iP_t=1}(1-\kappa_i^{\mathrm{FS}}-\zeta_t^{\mathrm{FS}}-\xi_i^{\mathrm{FS}}F_t)(\Delta_{it}-\overline{\Delta}^{\mathrm{ATT}})}{\sum_{(i,t):D_iP_t=1}u_{it}^{\mathrm{FS}}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]\\
=&\mathbb{E}\left[\overline{\Delta}^{\mathrm{ATT}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]+\frac{(\overline{F}_{\mathrm{pre}}-\overline{F})}{\sigma_F^2+(\overline{F}_{\mathrm{post}}-\overline{F})(\overline{F}_{\mathrm{pre}}-\overline{F})}\mathbb{E}\left[Q_{F,\Delta}^{\mathrm{ATT}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]
\end{aligned}
$$

We can not get the analytical solution of $\mathbb{E}\left[\hat{\alpha}^{\mathrm{FS}}\right]$ without adding assumption, because the variation of the factor σ_F^2 and the sample covariance $Q_{F,\Delta}^{\mathrm{ATT}}$ both depend on the factor realizations in a specific sample. If we assume the factor realizations F_t are exogenously determined (i.e. they are not random across samples), we can get a simplified expression.

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\mathrm{FS}}\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[\hat{\alpha}^{\mathrm{FS}}\middle|\boldsymbol{D},\boldsymbol{P}\right]\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[\overline{\Delta}^{\mathrm{ATT}}\middle|\boldsymbol{D},\boldsymbol{P}\right]\right]\\
&+\mathbb{E}\left[\frac{\overline{F}_{\mathrm{pre}}-\overline{F}}{\sigma_F^2+(\overline{F}_{\mathrm{post}}-\overline{F})(\overline{F}_{\mathrm{pre}}-\overline{F})}\mathbb{E}\left[Q_{F,\Delta}^{\mathrm{ATT}}\middle|\boldsymbol{D},\boldsymbol{P}\right]\right]\\
=&\alpha^{\mathrm{ATT}}+w^{FS}\mathrm{cov}\left(F_t,\Delta_{it}\middle|D_i=1,P_t=1\right)
\end{aligned}
$$

where

$$
w^{FS}=\mathbb{E}\left[\frac{N_TT_P-1}{N_TT_P}\cdot\frac{\overline{F}_{\mathrm{pre}}-\overline{F}}{\sigma_F^2+(\overline{F}_{\mathrm{post}}-\overline{F})(\overline{F}_{\mathrm{pre}}-\overline{F})}\right]
$$

A.4 Proof of Corollary 3

In practice, researchers add unit-specific time trend, termed unit time trend (UTT) estimator, to control for time-varying heterogeneity. However, it is a special case of the full sample estimator and have the same issue of “bad time control problem”.

We take unit-specific linear time trend $F_t = t$ as an example. If the time trend is polynomial, it does not make an essential difference. Since time trend is exogenously determined and non-random across different samples, we can plug $F_t = t$ into the estimation result of the full sample estimator and get the bias of unit time trend estimator. Suppose the treatment happens at time $T - T_P + 1$.

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\text{UTT}}\right] \\
=&\alpha^{\text{ATT}} + \mathbb{E}\left[\frac{N_T T_P - 1}{N_T T_P} \cdot \frac{\frac{-T_P}{2}}{\frac{T^2-1}{12} + \frac{T-T_P}{2}\frac{-T_P}{2}}\right] \text{cov}(\Delta_{it}, t | D_i P_t = 1) \\
=&\alpha^{\text{ATT}} + w^{\text{UTT}} \text{cov}(\Delta_{it}, t | D_i P_t = 1)
\end{aligned}
$$

where

$$
w^{\text{UTT}} = \mathbb{E}\left[\frac{N_T T_P - 1}{N_T T_P} \cdot \frac{-6T_P}{T^2 - 1 - 3T_P(T - T_P)}\right] < 0
$$

A.5 Proof of Proposition 3

In order to avoid bad control problem, researchers add pre-treatment covariate interacted with time trend termed covariate time trend (CTT) estimator. However, CTT has a similar bias as the full sample estimator, because of "bad time control problem". For simplicity, we assume the time trend is a linear time trend, but the argument is also valid for polynomial time trend. Let $Z_{it} = X_{i0} \cdot t$. The definition of covariate time trend estimator is:

$$
(\hat{\alpha}^{\text{CTT}}, \hat{\gamma}_i^{\text{CTT}}, \hat{\eta}_t^{\text{CTT}}, \hat{\beta}^{\text{CTT}}) = \underset{\alpha,\gamma,\eta,\beta}{argmin} \Big\{ \sum_{t=1}^{T} \sum_{i=1}^{N} (Y_{it} - \gamma_i - \eta_t - \beta Z_{it} - \alpha D_i P_t)^2 \Big\} \tag{13}
$$

We regress $D_i P_t$ on unit dummies, time dummies, and Z_{it}:

$$
D_i P_t = \kappa_i^{\text{CTT}} + \zeta_t^{\text{CTT}} + \xi^{\text{CTT}} Z_{it} + u_{it}^{\text{CTT}} \tag{14}
$$

The estimate of ξ is the coefficient of a covariant in the two-way fixed effect estimator.

$$
\hat{\xi}^{\text{CTT}} = \frac{\sum_{it}(D_i - \overline{D})(P_t - \overline{P})(Z_{it} - \overline{Z}_{i.} - \overline{Z}_{.t} + \overline{Z})}{\sum_{it}(Z_{it} - \overline{Z}_{i.} - \overline{Z}_{.t} + \overline{Z})^2}
$$

Thus, the residual of regression 14 can be written as:

$$
u_{it}^{\text{CTT}} = \left(D_i - \overline{D}\right)\left(P_t - \overline{P}\right) - \hat{\xi}^{\text{CTT}} Z_{it}
$$

where

$$
\overline{Z}_{i.} = \frac{\sum_{t=1}^{T} Z_{it}}{T} \qquad \overline{Z}_{.t} = \frac{\sum_{i=1}^{N} Z_{it}}{N} \qquad \overline{Z} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}}{NT} \qquad \overline{Z}^{\text{ATT}} = \frac{\sum_{i,t} D_i P_t Z_{it}}{\sum_{i,t} D_i P_t}
$$

Denote the sample covariance between covariate time trend Z_{it} and individual treatment effect Δ_{it} as

$$Q_{Z,\Delta}^{\text{ATT}} = \frac{\sum_{i,t} D_i P_t Z_{it}(\Delta_{it} - \overline{\Delta}^{\text{ATT}})}{\sum_{i,t} D_i P_t}$$

In the best possible setting, researchers correctly identify the factor structure (i.e. $\lambda_i = X_{i0}$ and $F_t = t$). However, the covariate time trend estimator is still subject to the bad time control problem:

$$\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\text{CTT}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right]\\
=&\mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{CTT}} Y_{it}}{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right]\\
=&\mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t \Delta_{it}}{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right] + \mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{CTT}} \lambda_i F_t}{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right]\\
=&\mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t \overline{\Delta}^{\text{ATT}}}{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right] + \mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t \left(\Delta_{it} - \overline{\Delta}^{\text{ATT}}\right)}{\sum_{i,t} u_{it}^{\text{CTT}} D_i P_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right]\\
=&\mathbb{E}\left[\overline{\Delta}^{\text{ATT}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right] + \frac{-\hat{\xi}^{\text{CTT}}}{(1-\overline{D})(1-\overline{P}) - \hat{\xi}^{\text{CTT}}\overline{Z}^{\text{ATT}}}\,\mathbb{E}\left[Q_{Z,\Delta}^{\text{CTT}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{Z}\right]
\end{aligned}$$

It shows that the covariate time trend estimator exists the bad time control problem in general. Besides, the covariate time trend estimator may suffer from the misspecification problem and generate additional bias terms if the pre-treatment covariate does not fully correlated with the factor loading or the factor realization is not a linear time trend.

A.6 Proof of Proposition 4

A.6.1 Unit group interacted with time

Suppose g_i indicates the group unit i belongs to and we define dummy factor estimator with unit group interacted with time as

$$(\hat{\alpha}^{\text{DF1}}, \hat{\gamma}_i^{\text{DF1}}, \hat{\omega}_{rt}^{\text{DF1}}) = \underset{\alpha,\gamma,\omega}{argmin}\Big\{\sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{it} - \gamma_i - \omega_{g_i,t} - \alpha D_i P_t)^2\Big\} \tag{15}$$

we regress $D_i P_t$ on unit dummies, and group dummies interacted with time dummies.

$$D_i P_t = \kappa_i^{\text{DF1}} + \theta_{rt}^{\text{DF1}} + u_{it}^{\text{DF1}} \tag{16}$$

Let R_g be the ratio of treated units overall all units in group g,

$$R_g = \frac{\sum_i D_i \mathbb{1}_{\{i\in g\}}}{\sum_i \mathbb{1}_{\{i\in g\}}}$$

Then, the residual of regression 16 is:

$$u_{it}^{\text{DF1}} = (D_i - R_{g_i})(P_t - \overline{P})$$

Denote $N_{D,g}$ is the number of treated observations in group g, $\overline{\lambda}_{D,g}$ is the mean of factor loadings λ of the treated observations in group g, $\overline{\lambda}_{C,g}$ is the mean of factor loadings λ of the non-treated observations in group r, and $\overline{\Delta}_g^{\text{ATT}}$ is the mean of the treatment effects Δ_{it} for the treated observations in group g.

Following the Frisch-Waugh-Lovell theorem, we can show that the dummy factor estimator is biased in two ways.

$$\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\text{DF1}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\text{DF1}}Y_{it}}{\sum_{i,t}u_{it}^{\text{DF1}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\text{DF1}}\Delta_{it}D_iP_t}{\sum_{i,t}u_{it}^{\text{DF1}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] + \mathbb{E}\left[\frac{\sum_{i,t}u_{it}^{\text{DF1}}\lambda_iF_t}{\sum_{i,t}u_{it}^{\text{DF1}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t}(1-R_{g_i})(1-\overline{P})D_iP_t\Delta_{it}}{\sum_{i,t}(1-R_{g_i})(1-\overline{P}))D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
&+\mathbb{E}\left[\frac{\left(\sum_i(D_i-R_{g_i})\lambda_i\right)\left(\sum_t(P_t-\overline{P})F_t\right)}{\sum_{i,t}u_{it}^{\text{DF1}}D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t}(1-R_{g_i})D_iP_t\Delta_{it}}{\sum_{i,t}(1-R_{g_i})D_iP_t}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
&+\mathbb{E}\left[\frac{\left(\sum_i(D_i-R_{g_i})\lambda_i\right)T\overline{P}(\overline{F}_{\text{post}}-\overline{F})}{\left(\sum_i(1-R_{g_i})D_i\right)(1-\overline{P})T\overline{P}}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
=&\mathbb{E}\left[\frac{\sum_g N_gR_g(1-R_g)\overline{\Delta}_g^{\text{ATT}}}{\sum_g N_gR_g(1-R_g)}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right] \\
&+\mathbb{E}\left[\frac{\sum_g N_gR_g(1-R_g)\left(\overline{\lambda}_{D,r}-\overline{\lambda}_{C,r}\right)\left(\overline{F}_{\text{post}}-\overline{F}_{\text{pre}}\right)}{\sum_g N_gR_g(1-R_g)}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{G}\right]
\end{aligned}$$

Given the assignment of groups, the expectation of a dummy factor estimator can be expressed as a convex combination of TWFE estimates for each group.

$$\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\text{DF1}}\middle|\boldsymbol{G}\right] \\
=&\sum_g\omega_g^{\text{DF1}}\,\mathbb{E}\left[\Delta_{it}\middle|\,g_i=g,D_i=1,P_t=1\right] \\
&+\sum_g\omega_g^{\text{DF1}}\left(\mathbb{E}\left[\lambda_i\middle|\,g_i=g,D_i=1\right]-\mathbb{E}\left[\lambda_i\middle|\,g_i=g,D_i=0\right]\right)\left(\mathbb{E}\left[F_t\middle|\,P_t=1\right]-\mathbb{E}\left[F_t\middle|\,P_t=0\right]\right)
\end{aligned}$$

where

$$\omega_g^{\mathrm{DF1}} = \frac{N_g R_g(1-R_g)}{\sum_g N_g R_g(1-R_g)}$$

We find that weighting problem can be rewritten as the true ATT plus a covariance term like bad time control problem. The weighting issue in Chaismartin and d'Haullfulle can be also rewrite in this way. Let $\overline{R}^{\mathrm{ATT}}$ be the sample mean of treated unit ratios for treatment observations and $Q_{G,\Delta}^{\mathrm{ATT}}$ be the sample covariance of group treatment ratio R_{g_i} and treatment effect Δ_{it} for the treated observations.

$$\overline{R}^{\mathrm{ATT}} = \frac{\sum_{i,t} D_i P_t R_{g_i}}{\sum_{i,t} D_i P_t}$$

$$Q_{G,\Delta}^{\mathrm{ATT}} = \frac{\sum_{i,t} D_i P_t \left(R_{g_i} - \overline{G}^{\mathrm{ATT}}\right)\left(\Delta_{it} - \overline{\Delta}^{\mathrm{ATT}}\right)}{\sum_{i,t} D_i P_t}$$

When loadings are balanced within each group ($\forall$group g, $\mathbb{E}\left[\lambda_i \middle| g_i = g, D_i = 1\right] - \mathbb{E}\left[\lambda_i \middle| g_i = g, D_i = 0\right]$), the omitted factor bias will be equal to zero. But then, the dummy factor estimator is still biased because of the weighting issue. The dummy factor estimator will be shown as

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{i,t}(1-R_{g_i})D_i P_t \Delta_{it}}{\sum_{i,t}(1-R_{g_i})D_i P_t}\middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{G}\right]\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\overline{\Delta}_{\mathrm{ATT}}\middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{G}\right]\right] - \mathbb{E}\left[\frac{1}{1-\overline{G}^{\mathrm{ATT}}}\,\mathbb{E}\left[Q_{G,\Delta}^{\mathrm{DF1}}\middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{G}\right]\right] \\
=&\alpha^{\mathrm{ATT}} - \frac{\mathrm{cov}\left(\Delta_{it}, R_{g_i}\middle| D_i P_t = 1\right)}{1 - \mathbb{E}\left[R_{g_i}\middle| D_i P_t = 1\right]}
\end{aligned}$$

Fortunately, the weighting issue is not extremely severe in the dummy factor estimator, because the weighting is grantee to be between 0 and 1 (unlike in Chaismartin and d'Haullfulle) and the estimator does not become negative if true treatment effects are all positive.

$$\begin{aligned}
&\alpha^{\mathrm{ATT}} - \frac{\mathrm{cov}\left(\Delta_{it}, R_{g_i}\middle| D_i P_t = 1\right)}{1 - \mathbb{E}\left[R_{g_i}\middle| D_i P_t = 1\right]} \\
=&\alpha^{\mathrm{ATT}} - \frac{\mathbb{E}\left[\Delta_{it} R_{g_i}\middle| D_i P_t = 1\right] - \mathbb{E}\left[\Delta_{it}\middle| D_i P_t = 1\right]\mathbb{E}\left[R_{g_i}\middle| D_i P_t = 1\right]}{1 - \mathbb{E}\left[R_{g_i}\middle| D_i P_t = 1\right]} \\
\geq&\alpha^{\mathrm{ATT}} - \frac{\mathbb{E}\left[\Delta_{it}\middle| D_i P_t = 1\right] - \mathbb{E}\left[\Delta_{it}\middle| D_i P_t = 1\right]\mathbb{E}\left[R_{g_i}\middle| D_i P_t = 1\right]}{1 - \mathbb{E}\left[R_{g_i}\middle| D_i P_t = 1\right]} \\
=&\alpha^{\mathrm{ATT}} - \alpha^{\mathrm{ATT}} \\
=&0
\end{aligned}$$

A.6.2 Time group interacted with unit

Suppose h_t indicates the time group that time t belongs to and we define dummy factor estimator with time group interacted with unit as

$$(\hat{\alpha}^{\text{DF2}}, \hat{\eta}_t^{\text{DF2}}, \hat{\omega}_{is}^{\text{DF2}}) = \underset{\alpha,\eta,\omega}{argmin} \Big\{ \sum_{t=1}^{T} \sum_{i=1}^{N} (Y_{it} - \eta_t - \omega_{i,h_t} - \alpha D_i P_t)^2 \Big\} \tag{17}$$

we regress $D_i P_t$ on time dummies, and time group dummies interacted with unit dummies.

$$D_i P_t = \zeta_t^{\text{DF2}} + \theta_{is_t}^{\text{DF2}} + u_{it}^{\text{DF2}}$$

Define K_h is the ratio of treated time periods overall all time periods in group h,

$$K_h = \frac{\sum_t P_t \mathbb{1}_{\{t \in h\}}}{\sum_i \mathbb{1}_{\{t \in h\}}}$$

Similarly, we can get

$$u_{it}^{\text{DF2}} = (D_i - \overline{D})(P_t - K_{h_t})$$

Following the Frisch-Waugh-Lovell theorem, we can show that the dummy factor estimator with time group interacted with unit.

$$\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{\text{DF2}} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{DF2}} Y_{it}}{\sum_{i,t} u_{it}^{\text{DF2}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{DF2}} \Delta_{it} D_i P_t}{\sum_{i,t} u_{it}^{\text{DF2}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] + \mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{DF2}} \lambda_i F_t}{\sum_{i,t} u_{it}^{\text{DF2}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t}(1-\overline{D})(1-K_{h_t}) D_i P_t \Delta_{it}}{\sum_{i,t}(1-\overline{D}) D_i P_t (1-K_{h_t})} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
&+ \mathbb{E}\left[\frac{\left(\sum_i (D_i - \overline{D})\lambda_i\right)\left(\sum_t (P_t - K_{h_t}) F_t\right)}{\sum_{i,t} u_{it}^{\text{DF2}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t}(1-K_{h_t}) D_i P_t \Delta_{it}}{\sum_{i,t}(1-K_{h_t}) D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
&+ \mathbb{E}\left[\frac{N\overline{D}(\overline{\lambda}_D - \overline{\lambda})\left(\sum_t (P_t - K_{h_t}) F_t\right)}{N\overline{D}(1-\overline{D})\left(\sum_t (P_t - K_{h_t}) P_t\right)} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
=&\mathbb{E}\left[\frac{\sum_{i,t}(1-K_{h_t}) D_i P_t \Delta_{it}}{\sum_{i,t}(1-K_{h_t}) D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right] \\
&+ \mathbb{E}\left[\frac{\sum_t (P_t - K_{h_t}) F_t}{\sum_t (P_t - K_{h_t}) P_t}\left(\overline{\lambda}_{\text{D}} - \overline{\lambda}_{\text{C}}\right) \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{K}\right]
\end{aligned}$$

Because time groups are continuous in time, most of the time groups are either fully before the treatment date or fully after the treatment date except for one group within which the treatment takes place. This suggests most of K_h are either 0 or 1. For all time t whose $K_{h_t} = 0$ or $K_{h_t} = 0$, it does contributes to neither the treatment effect part nor the bias part. It means that when using a time group dummy factor only keeps the data of one time group - the time group within which the treatment happens. However, for that time group, the dummy factor estimator with time group interacted with unit suffers from the bad time control problem like the one with unit group interacted with time.

A.7 Proof of Proposition 5

Given a sufficiently long time-series it is possible to estimate factor loadings only using pre-treatment variation and then use the estimated loadings when estimating the ATT in the full sample. We refer to this two step procedure as the pre-treatment (PT) estimator. First loadings are estimated over pre-treatment periods,

$$(\hat{\lambda}_i^{\mathrm{PT}}) = \underset{\lambda}{argmin}\Big\{\sum_{t=1}^{T}(1-P_t)\sum_{i=1}^{N}(Y_{it}-\gamma_i-\eta_t-\lambda_i F_t)^2\Big\}$$

and the estimated loadings $(\hat{\lambda}_i)$ are then used in the full sample when estimating the ATT,

$$(\hat{\alpha}^{PT}, \hat{\gamma_i}^{\mathrm{PT}}, \hat{\eta_t}^{\mathrm{PT}}) = \underset{\alpha,\gamma,\eta}{argmin}\Big\{\sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{it}-\gamma_i-\eta_t-\hat{\lambda}_i^{\mathrm{PT}}F_t-\alpha D_i P_t)^2\Big\}. \tag{18}$$

The pre-treatment estimator estimates the loadings using the sample of the pre-treatment period in order to avoid estimated loadings captures the treatment effect variation.

Due to collinearity, we, without loss of generality, assume the estimated loading of unit 1 is 0 and the average of time fixed effects is 0, i.e. $\hat{\lambda}_1^{PT} = 0$, $\sum_t \zeta_t^{PT,k} = 0$. Define $w_{it}^{PT,k}$ is the residual of $F_t \cdot \mathbb{1}_{i=k}$ on unit dummies, time dummies, and the rest of factors in the pre-treatment period.

$$F_t\mathbb{1}[i=k] = \kappa_i^{PT,k} + \zeta_t^{PT,k} + \sum_{j\neq k \wedge j\neq 1}\xi_j^{PT,k}F_t\mathbb{1}[i=j] + v_{it}^{PT,k}$$

we can verify that

$$\begin{aligned}
\xi_i^{PT,k} &= \xi_j^{PT,k} && \text{if} \quad i \neq k \wedge i \neq 1 \wedge j \neq k \wedge j \neq 1 \\
\phi_i^{PT,k} &= -\frac{\xi_i^{PT,k}}{T - T_P} \sum_{t:P_t=0} F_t && \text{if} \quad i \neq k \wedge i \neq 1 \\
v_{it}^{PT,k} &= 0 && \text{if} \quad i \neq k \wedge l \neq 1 \\
v_{kt}^{PT,k} &= -v_{1t}^{PT,k}
\end{aligned}$$

Based on the Frisch-Waugh-Lovell theorem, the loading estimate of unit k ($\hat{\lambda}_k^{PT}$) can be written as:

$$\begin{aligned}
\hat{\lambda}_k^{PT} =& \frac{\sum_{(i,t):P_t=0} v_{it}^{PT,k} Y_{it}}{\sum_{(i,t):P_t=0} v_{it}^{PT,k} F_t \mathbb{1}_{i=k}} \\
=& \frac{\sum_{(i,t):P_t=0} v_{it}^{PT,k} \lambda_i F_t + \sum_{(i,t):P_t=0} v_{it}^{PT,k} \varepsilon_{it}}{\sum_{(i,t):P_t=0} v_{it}^{PT,k} F_t \mathbb{1}_{i=k}} \\
=& \frac{\sum_{t:P_t=0} v_{1t}^{PT,k} \lambda_1 F_t + \sum_{t:P_t=0} v_{kt}^{PT,k} \lambda_k F_t}{\sum_{t:P_t=0} v_{kt}^{PT,k} F_t} + \frac{\sum_{(i,t):P_t=0} v_{it}^{PT,k} \varepsilon_{it}}{\sum_{(i,t):P_t=0} v_{it}^{PT,k} F_t \mathbb{1}_{i=k}} \\
=& \frac{-\lambda_1 \left(\sum_{t:P_t=0} v_{kt}^{PT,k} F_t \right) + \lambda_k \left(\sum_{t:P_t=0} v_{kt}^{PT,k} F_t \right)}{\sum_{t:P_t=0} v_{kt}^{PT,k} F_t} + \frac{\sum_{(i,t):P_t=0} v_{it}^{PT,k} \varepsilon_{it}}{\sum_{(i,t):P_t=0} v_{it}^{PT,k} F_t \mathbb{1}_{i=k}} \\
=& \lambda_k - \lambda_1 + \frac{\sum_{(i,t):P_t=0} v_{it}^{PT,k} \varepsilon_{it}}{\sum_{(i,t):P_t=0} v_{it}^{PT,k} F_t \mathbb{1}_{i=k}}
\end{aligned}$$

Then, we regress $Y_{it} - \hat{\lambda}_i^{PT} F_t$ on unit dummies, time dummies and treatment dummies and get the pre-treatment estimator.

$$\begin{aligned}
& \mathbb{E}\left[\hat{\alpha}^{PT} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F} \right] \\
=& \mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{TWFE}} (Y_{it} - \hat{\lambda}_i^{PT} F_t)}{\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F} \right] \\
=& \mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{TWFE}} \Delta_{it} D_i P_t}{\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F} \right] + \mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{TWFE}} (\lambda_i - \hat{\lambda}_i^{PT}) F_t}{\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F} \right] \\
=& \mathbb{E}\left[\frac{\sum_{i,t:} D_i P_t \Delta_{it}}{\sum_{i,t} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F} \right] + \mathbb{E}\left[\frac{\sum_{i,t} u_{it}^{\text{TWFE}} \lambda_1 F_t}{\sum_{i,t} u_{it}^{\text{TWFE}} D_i P_t} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F} \right] \\
=& \mathbb{E}\left[\overline{\Delta}^{\text{ATT}} \middle| \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{F} \right]
\end{aligned}$$

We can show that pre-treatment estimator is unbiased.

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}^{PT}\right] \\
=&\,\mathbb{E}\left[\mathbb{E}\left[\hat{\alpha}^{PT}\middle|\boldsymbol{D},\boldsymbol{P},\boldsymbol{F}\right]\right] \\
=&\,\alpha^{\text{ATT}}
\end{aligned}
$$

Appendix B Additional Simulations

Table B1: Correlation between Treatment effects and Factor Structure

This table presents how estimates change along with the correlation between loadings and treatment effects $\rho_{\Delta,\lambda}$ and correlation between factors and treatment effects $\rho_{\Delta,F}$. All other parameter are the same as baseline setting. The mean of two-way fixed effect estimators is displayed in the first line without any parentheses, the mean of full sample estimator is displayed in the second line with round parentheses, and the mean of pre-treatment estimators is displayed in the third line with square parentheses.

TWFE Estimator (FS Estimator) [PT Estimator]		Correlation between Loadings and TE $\rho_{\Delta,\lambda}$					
		0	0.2	0.4	0.6	0.8	1
Correlation between Factors and TE $\rho_{\Delta,F}$	0	1.3161	1.3252	1.3179	1.3424	1.3471	1.3005
		(1.0021)	(1.0016)	(1.0020)	(0.9998)	(1.0024)	(1.0035)
		[1.0098]	[1.0066]	[1.0067]	[1.0052]	[1.0024]	[1.0054]
	0.2	1.3315	1.3285	1.3212	1.3178	1.2856	1.3155
		(0.9485)	(0.9469)	(0.9492)	(0.9578)	(0.9318)	(0.9583)
		[0.9956]	[0.9936]	[0.9965]	[1.0005]	[0.9825]	[0.9979]
	0.4	1.2988	1.2862	1.3037	1.2961	1.3558	1.2758
		(0.9037)	(0.8900)	(0.8943)	(0.8886)	(0.9294)	(0.8857)
		[0.9938]	[0.9856]	[0.9914]	[0.9804]	[1.0273]	[0.9774]
	0.6	1.3206	1.3189	1.2996	1.2873	1.2742	1.3026
		(0.8673)	(0.8607)	(0.8487)	(0.8405)	(0.8621)	(0.8630)
		[1.0077]	[1.0019]	[0.9825]	[0.9809]	[0.9919]	[0.9969]
	0.8	1.3433	1.3232	1.3067	1.3387	1.3221	1.2988
		(0.8202)	(0.8002)	(0.8105)	(0.8251)	(0.8295)	(0.8122)
		[1.0189]	[0.9946]	[0.9966]	[1.0158]	[1.0121]	[0.9926]
	1	1.3476	1.3134	1.2950	1.2967	1.3757	1.3014
		(0.7554)	(0.7690)	(0.7658)	(0.7679)	(0.7731)	(0.7719)
		[1.0063]	[0.9971]	[0.9950]	[0.9910]	[1.0327]	[0.9987]

Table B2: Unit Dimension Irrelevance

This table presents how estimates change along with the loading difference μ and correlation between loadings and treatment effects $\rho_{\Delta,\lambda}$. All other parameter are the same as baseline setting. The mean of two-way fixed effect estimators is displayed in the first line without any parentheses, the mean of full sample estimator is displayed in the second line with round parentheses, and the mean of pre-treatment estimators is displayed in the third line with square parentheses.

TWFE Estimator **(FS Estimator)** **[PT Estimator]**		Loading Difference μ					
		0	0.2	0.4	0.6	0.8	1
Correlation between Loadings and TE $\rho_{\Delta,\lambda}$	0	0.9957	1.1476	1.3190	1.5148	1.6653	1.7614
		(0.8554)	(0.8637)	(0.8699)	(0.8566)	(0.8532)	(0.8751)
		[0.9923]	[0.9970]	[1.0079]	[1.0099]	[1.0023]	[1.0091]
	0.2	1.0069	1.1718	1.3267	1.4874	1.5940	1.7955
		(0.8571)	(0.8581)	(0.8621)	(0.8511)	(0.8386)	(0.8711)
		[1.0074]	[1.0020]	[1.0020]	[0.9951]	[0.9803]	[1.0023]
	0.4	1.0085	1.1402	1.3010	1.4801	1.6821	1.7797
		(0.8631)	(0.8458)	(0.8455)	(0.8574)	(0.8884)	(0.8419)
		[1.0001]	[0.9928]	[0.9886]	[0.9962]	[1.0328]	[0.9851]
	0.6	1.0001	1.1574	1.3002	1.4449	1.5691	1.8105
		(0.8668)	(0.8566)	(0.8493)	(0.8392)	(0.8648)	(0.8470)
		[1.0054]	[1.0002]	[0.9831]	[0.9809]	[0.9943]	[0.9932]
	0.8	1.0211	1.1664	1.3131	1.4811	1.6185	1.7735
		(0.8709)	(0.8520)	(0.8637)	(0.8689)	(0.8757)	(0.8617)
		[1.0215]	[0.9980]	[1.0031]	[1.0116]	[1.0111]	[0.9981]
	1	1.0044	1.1412	1.2972	1.4659	1.6616	1.7427
		(0.8558)	(0.8436)	(0.8635)	(0.8798)	(0.8657)	(0.8638)
		[1.0091]	[0.9803]	[0.9972]	[1.0052]	[1.0107]	[0.9935]

Appendix C Detailed literature review

We review 21 papers that use difference-in-difference or closely related methodology that we found in our literature review. For each paper, we use the following presentation:

Authors (year), Title.
DiD: Which tables use difference-in-differences methodology.
Estimator: What difference-in-difference estimator is used in the respective difference-in-differences tables.
Dimension: The unit dimension (e.g., plant), time frequency (e.g., monthly)) of the paper.
Factor Control: A description of the factor control that is used in the respective tables.
Description of variables and regressions: Brief description of the independent and dependent variables or mechanism being tested.
Heterogeneous: Whether the paper estimates heterogeneous treatment effects. We take a liberal classification here and describe also subsample analysis that acknowledges that the treatment effects are heterogeneous.
Dynamic: Whether a dynamic difference-in-differences estimator is used and if so which specifications with a factor structure use the dynamic estimator.
Staggered: Whether the difference-in-differences is staggered.

1. **Collard-Wexler and de Loecker (2015)**, Reallocation and Technology: Evidence from the US Steel Industry.
   DiD: Tables 5, 9.
   Estimator: Dummy Factor (Table 5, columns 2, 3, 4, Table 9, Columns 3, 4, 6, 7) TWFE (Table 9, Column 8).
   Dimensions: Plant $\times$ Year.
   Description of factor control: Dummy Factor: Year $\times$ Firm, Year $\times$ State, Firm $\times$ Year $\times$ State.
   Description of variables and regressions: Main independent variable is a dummy variable indicating whether a plant is vertically integrated interacted with a time dummy. The tables measure the (negative) technology premium associated with old technology.

Heterogeneous: No

Dynamic: No

Staggered: The implied DiD could be staggered since plants could theoretically be classified as vertically integrated and then change status at a later date.

2. **Cicala (2015)**, When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation. **DiD:** Tables 2, 3, 4, 5, 6, 7.

   Estimator: TWFE (Table 2, columns 1-6, Table 3, columns 2, 4-6, Table 4, columns 1-6, Table 6, columns 1-6, Table 7, columns 1-6) Dummy Factor (Table 3, column 3) Unit time trend (Online Appendix Table B.5).

   Dimensions: Facility × Month.

   Description of factor control: Dummy Factor (Table 3, column 3) – Facility × Year. Unit time trend (Online Appendix Table B.5) State-specific quadratic time trend.

   Description of variables and regressions: DiD that relates deregulation to the price paid for coal by power plants.

   Heterogeneous: Yes p.432 discusses the heterogeneity of treatment effects.

   Dynamic: Yes, Figure 5 presents dynamic treatment effects.

   Dynamic used with factor control: No

   Staggered: The deregulation is staggered.

3. **Currie, Davis, Greenstone and Walker (2015)**, Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings.

   DiD: Tables 2, 3, 4, 5, 6.

   Estimator: Dummy Factor (Table 2, columns 1-8, Table 4, columns 1-8, Table 6, columns 1-8), Covariates time trend (Table 2-6).

   Dimensions: Plant × Year.

   Description of factor control: Dummy Factor: Plant × Distance-bin, State × year, Plant × year, County × year. Covariates time trend: 1990 census tract characteristics interacted with quadratic time trends.

   Description of variables and regressions: Dependent variable pollution / birthweight and independent variable is plant openings and closings.

Heterogeneous: No

Dynamic: Yes, Figure 3 and 4.

Dynamic used with factor control: Yes, Figure 3 and 4 which use Dummy Factor and Covariates time trend. No, (Table 2, columns 1-8, Table 4, columns 1-8, Table 5, columns 1-5, Table 6, columns 1-8)

Staggered: Yes

4. **Favara and Imbs (2015)**, Credit Supply and the Price of Housing. Staggered.

   DiD: Table 2, Table 3, Table 4, Table 6.

   Estimator: TWFE (Table 2, 3, 4, 6)

   Dimensions: County × Year.

   Description of factor control: None

   Description of variables and regressions: Dependent variables loan outcomes and housing returns. Independent variable is state-wide banking deregulation index (developed by Rice and Strahan, 2010).

   Heterogeneous: Yes, allows for different treatment effects across counties depending on their house price elasticity.

   Dynamic: No

   Staggered: Yes

5. **Hackmann, Kolstad and Kowalski (2015)**, Adverse Selection and an Individual Mandate: When Theory Meets Practice.

   DiD: Table 2, 4.

   Estimator: DiD (Table 2,4)

   Dimensions: State × Year.

   Description of factor control: None

   Description of variables and regressions: Dependent variables are insurance coverage, log premiums or log average costs and independent variable is regulation change in Massachusetts that mandated insurance.

   Heterogeneous: No

Dynamic: No

Staggered: No

6. **Bailey and Goodman-Bacon (2015)**, The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans.

   DiD: Table 2- 5.

   Estimator: Dummy Factor (Table 2, Table 3, Table 4, Table 5) Covariate time trend (Table 2, column 2, column 3, Table 3, Table 4, Table 5)

   Dimensions: County × Year.

   Description of factor control: Dummy Factor: Urban × Year, State × Year.

   Covariate time trend: 1960 characteristics interacted with linear time trend: share of population: in urban area, in rural area, under 5 years of age, 65 or older, non-white, with 12 or more years of education, with less than 4 years of education, in households with Income less than $3000. In households with Incomes greater than $10000, total active MDs.

   Description of variables and regressions: Dependent variable average mortality rate and main independent variable is a dummy variables indicating the introduction of community health centres.

   Heterogeneous: Yes, Table 2, Panel A considers all ages while Panel B only considers people over 50 years. Table 3 stratifies treatment effects on mortality causes (e.g., heart disease). Table 4 stratifies treatment effects over 1960 characteristics and census regions. Table 5 stratifies results over household Income.

   Dynamic: Yes, Table 2, 3, and 4 consider dynamic treatment effects. Figures 5, 6 and 7 are dynamic.

   Dynamic used with factor control: Yes, Table 2, 3, and 4 consider dynamic treatment effects (treatment effects are estimated over different event time buckets). Figure 5, 6 and 7. No, Table 5.

   Staggered: Yes

7. **Burgess, Jedwab, Miguuel, and Morjaria Padró I Miquel (2015)**, The Value of Democracy: Evidence from Road Building in Kenya. **DiD:** Table 1, 2, 3, 5.

Estimator: TWFE (Table 1, column 1, Table 2, column 2) Covariate time trend (Table 1, column 2-5, Table 2, column 2 -5, Table 3, Table 5) Unit time trend (Table 1, column 5, Table 2, column 5, Table, 5 columns 3-4)

Dimensions: District × Year.

Description of factor control: Covariate time trend: Table 1, 2, 5: (Population, area, urbanization rate) × trend, (earnings, employment, cash crops) × trend, (Main highway, border, dist. Nairobi) × trend, District time trends. Table 3, Initial controls × trend. Unit time trend: (Table 1, column 5, Table 2, column 5, Table, 5 columns 3-4)

Description of variables and regressions: The dependent variable is the share of road expenditure normalized by population share. The independent variable is co-ethnicity of president.

Heterogeneous: No

Dynamic: No

Staggered: Yes

8. **Braguinsky, Ohyama, Okazaki, and Syverson (2015)**, Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry.

   DiD: Table 2, 3 and 6.

   Estimator: TWFE (Table 2, Table 3, Table 6, Columns 4-6)

   Dimensions: Plant × Year.

   Description of factor control: None

   Description of variables and regressions: The dependent variable is the economic performance. The main independent variable indicates whether the particular plant was acquired.

   Heterogeneous: Yes, Table 2, 3: treatment effects are stratified according to whether the acquisition is undertaken by a serial acquirer.

   Dynamic: No

   Staggered: Yes

9. **Pomeranz (2015)**, No Taxation without Information: Deterrence and Self-Enforcement

in the Value Added Tax

DiD: Table 4, 5, 6, 7

Estimator: TWFE (Table 4, 5, 6, 7)

Dimensions: Firm × Month.

Description of factor control: None

Description of variables and regressions: The dependent variable is the line item increase the main independent variable is letter from the tax office interacted with line item.

Heterogeneous: No

Dynamic: Yes, Figure 2 is dynamic, tracking treatment effects over time.

Dynamic used with factor control: No factor used.

Staggered: Yes

10. **de Janvry, Emerick, Gonzalez-Navarro and Sadoulet (2015)**, Delinking Land Rights from Land Use: Certification and Migration in Mexico

    DiD: Table 1, 4, 5, 6.

    Estimator: TWFE (Table 1, Columns 1,2,3,5,6 Table 4, Columns 1-2, Table 5, Table 6 Column 1) Dummy Factor (Table 1, Column 4, Table 4, Column 3, Table 6, Column 2)

    Dimensions: Household × Ejido × Year.

    Description of factor control: Dummy Factor: Table 1, Column 4 (State × Time), Table 4, Column 3 (High-yield × Time), Table 6, Column 2 (Progresa Treatment Locality × Time)

    Description of variables and regressions: The main dependent variable is in an indicator variable for whether households have a migrant and the main independent variable is whether a geographic area has been certified.

    Dynamic: No

    Staggered: Yes

11. **Yagan (2015)**, Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut

DiD: Table 2, 3, 4.

Estimator: DiD (Table 2, Column 1, 3, 4, 5, 7, 8, 10, 11, Table 3, Table 4, Columns 1-2) TWFE (Table 1, Column 3, 6, 9, 12, Table 4, Column 3, Column 6).

Dimensions: Firm × Year.

Description of factor control: None. The main dependent variables are firm investment, employ compensation and firm payout. The main independent variable is whether the firm is a C-Corp interacted with a time dummy indicating the 2003 tax cut.

Heterogeneous: No

Dynamic: Yes, Table 4, Columns 1-6 includes dummies for each of the treatment years.

Dynamic used with factor control: No factor used.

Staggered: No

12. **Lalive, Landais and Zweimüller (2015)**, Market Externalities of Large Unemployment Insurance Extension Programs

    DiD: Table 2, 3, 4.

    Estimator: TWFE (Table 2, Columns 1-2, Table 3, Table 4) Unit time trends (Table 2, Columns 3-6).

    Dimensions: Firm × Year.

    Description of factor control: Unit time trends: Region specific trends

    Description of variables and regressions: The main dependent variable is unemployment duration and the main independent variable indicated eligibility of the Regional Extension Benefit Program (REBP) which extended unemployment benefits for a large subset of Austrian workers.

    Heterogeneous: Yes, treatment effects are evaluated across employment and age.

    Dynamic: No

    Staggered: Effectively yes since there are two treatments.

13. **Muhlenbachs, Spiller, Timmins (2015)**, The Housing Market Impacts of Shale Gas Development

DiD: Table 2, 3, 4.

Estimator: Dummy Factor (Table 2, 3, 4) TWFE (Table 4)

Dimensions: Quarter × House

Description of factor control: Dummy Factor: Table 2 Panel A (County × year), Panel B (Census tract × year), Table 3 Panel B (County × year)

Description of variables and regressions: The main dependent variable is log sale prices of houses and the main independent variable is the number of wells at different distances from the property as well as whether the property is reliant on ground water.

Heterogeneous: Yes, both Table 3 and 4 considers subsamples in different panels.

Dynamic: No

Staggered: Yes, the number of wells are changing.

14. **Bøler, Moxnes and Ulltveit-Moe (2015)**, R&D, International Sourcing, and the Joint Impact on Firm Performance

DiD: Table 4, 5, 6, 7, 8, 9.

Estimator: TWFE (Table 4, Table 5, Columns 1-3, Table 7, Column 1, Table 8, Column 5) Unit time trends (Table 5, Columns 4-7, Table 6, Table 7, Column 2, Table 8, Columns 1-4, Table 9)

Dimensions: Firm × Year.

Description of factor control: Unit time trends, Table 5, Columns 4-7, Table 6, Table 7, Column 2, Table 8, Columns 1-4, Table 9

Description of variables and regressions: The paper considers as dependent variables R&D expenditure and number of imported products and the main independent variable captures whether the firm is eligible for tax credits.

Heterogeneous: Yes, Table 8 considers the origins of imported products.

Dynamic: Yes, Table 4, columns 1-3, Table 5, columns 1-3

Dynamic used with factor control: No (Table 4 no factor factor used, Table 5, columns 4-7, Table 6, Table 7, columns 2, Table 8, Columns 1-4, Table 9) Yes (Table 5, columns 1-3) .

Staggered: No

15. **Duggan, Garthwaite and Goyal (2016)**, The Market Impacts of Pharmaceutical Product Patents in Developing Countries: Evidence from India

    DiD: Table 4, 5, 6, 7, 8.

    Dimensions: Molecules × Quarter

    Estimator: TWFE (Table 4, Columns 2, 4, 6, 8, Table 6, Columns 2, 4, Table 7, Columns 2, 4, 6, 8) Unit time trends (Table 4, Columns 1, 3, 5, 7, Table 5, Table 6, Columns 1, 3, Table 7, Columns 1, 3, 5, 7, Table 8)

    Description of factor control: Unit time trends: $\lambda \times t \times I_{EverPatent}$, where t is a time indicator and $I_{EverPatent}$ is an indicator of whether the molecule ever has had a patent. (Table 4, Columns 1, 3, 5, 7, Table 5, Table 6, Columns 1, 3, Table 7, Columns 1, 3, 5, 7, Table 8)

    Description of variables and regressions: The paper examines the effect of molecule patents on prices and quantities sold.

    Heterogeneous: No.

    Dynamic: Yes, Figure 1, 2, 5, 6, 7, 8 and 9 are event studies.

    Dynamic used with factor control: No (Table 4, Columns 1, 3, 5, 7, Table 5, Table 6, Columns 1, 3, Table 7, columns 1, 3, 5, 7, Table 8).

    Staggered: Yes

16. **Jayaraman, Ray and De Véricourt (2016)**, Anatomy of a Contract Change

    DiD: Table 2, Columns 4, 5

    Estimator: DiD (Table 2, Columns 4, 5)

    Dimensions: Rice output in kg × day

    Description of factor control: None

    Description of variables and regressions: This paper studies the effect of a contract change on tea worker's productivity.

    Heterogeneous: No.

    Dynamic: Yes, Figure 9 is estimated with time-varying treatment effects (one for each of 17 weeks).

    Staggered: No

17. **Hoynes, Whitmore Schanzenbach and Almond (2016)**, Long-Run Impacts of Childhood Access to the Safety Net

    DiD: Table 2, 3, 4, 5, 6, 7, 8, 9.

    Estimator: Unit time trends & Covariate time trends (Table 2, 3, 4, 5, 6, 7, 8, 9)

    Dimensions: Individual × County × Birth year

    Description of factor control: Unit time trends & Covariate time trends (Table 2, 3, 4, 5, 6, 7, 8, 9), State specific cohort trend, county pre-treatment characteristics trend.

    Description of variables and regressions: This paper studies the effect of food stamps programs on long-run health outcomes.

    Heterogeneous: Yes, it is stratified across gender.

    Dynamic: No

    Staggered: Yes

18. **Pierce and Schott (2016)**, The Surprisingly Swift Decline of US Manufacturing Employment

    DiD: Table 1, 2, 3, 4, 5, 6, 7, 8, 9.

    Estimator: TWFE (Table 1, Table 2, Columns 1-4, 5, 6, Table 3, Columns 2-3, Table 7, Table 8, Table 9), Covariate time trend (Table 2, Column 5) Dummy factor (Table 3, Column 1, Table 4, Columns 1-4, Table 5, Columns 1-4, Table 6, Columns 1-4)

    Dimensions: Industry × year.

    Description of factor control: Covariate time trend (Table 2, Column 5): $ln(RDGP)_t \times ln(NP/Emp_{i,t})$, effectively a GDP factor interacted with a covariate. Dummy factor (Table 3, Column 1): Country × time, Country × industry, Industry × year (Table 4, 5, 6) Product × country, Country × time, Product × time.

    Description of variables and regressions: This paper studies the effect of tariff reduction on employment.

    Heterogeneous: No

    Dynamic: No.

    Staggered: No

19. **Muralidharan, Niehaus and Sukhtankar (2016)**, Building State Capacity: Evidence from Biometric Smartcards in India

    DiD: Table 2, 7.

    Estimator: TWFE (Table 2, Columns 5-8, Table 7, Columns 3-4)

    Dimensions: Household / Individual × mandal-district × week.

    Description of factor control: None.

    Description of variables and regressions: This paper studies the impact of "smart cards" on the functioning of the financial system.

    Heterogeneous: No

    Dynamic: No

    Staggered: Yes

20. **Sequiera, (2016)**, Corruption, Trade Costs, and Gains from Tariff Liberalization: Evidence from Southern Africa

    DiD: Table 5, 8, 9, 10, 11, 13, 15, 18, 19.

    Estimator: DiD (Table 5, Columns 4-6, Table 8, Table 9, Table 10, Table 11, Table 13, Table 18, Table 19, Panel B), ≈ TWFE (Table 15)

    Dimensions: Trade Gap × Year

    Description of factor control: None.

    Description of variables and regressions: This paper studies the effect of tariff changes on trade and bribery.

    Heterogeneous: No

    Dynamic: No

    Staggered: No

21. **Koudjis and Voth (2016)**, Leverage and Beliefs: Personal Experience and Risk-Taking in Margin Lending

    DiD: Table 5, 6, 7 (Panel B), 8, 9, 11, 12

    Estimator: ≈ TWFE (Table 5, Table 6, Table 7 Columns 3 & 6, Table 8, Table 9, Table 11, Table 12)

    Dimensions: Haircut × year

Description of factor control: None.

Description of variables and regressions: This paper examines of unexpected investor losses on the interest rates and haircuts charged.

Heterogeneous: Yes, treatment effects are stratified across exposure and whether it is a loans consortium.

Dynamic: No

Staggered: No

References

Acharya, V. V., Imbs, J., & Sturgess, J. (2011). Finance and efficiency: do bank branching regulations matter? *Review of Finance*, *15*(1), 135–172.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, *77*(4), 1229–1279.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Bailey, M. J., & Goodman-Bacon, A. (2015). The war on poverty's experiment in public medicine: Community health centers and the mortality of older americans. *American Economic Review*, *105*(3), 1067–1104.

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395.

Beck, T., Levine, R., & Levkov, A. (2010). Big bad banks? the winners and losers from bank deregulation in the united states. *The Journal of Finance*, *65*(5), 1637–1667.

Berg, T., Reisinger, M., & Streitz, D. (2021). Spillover effects in empirical corporate finance. *Journal of Financial Economics*, *142*(3), 1109–1127.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249–275.

Boehmer, E., Jones, C. M., & Zhang, X. (2020). Potential pilot problems: Treatment spillovers in financial regulatory experiments. *Journal of Financial Economics*, *135*(1), 68–87.

Bøler, E. A., Moxnes, A., & Ulltveit-Moe, K. H. (2015). R&d, international sourcing, and the joint impact on firm performance. *American Economic Review*, *105*(12), 3704–39.

Borusyak, K., Jaravel, X., & Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Brown, N., & Butts, K. (2023). Dynamic treatment effect estimation with interactive fixed effects and short panels.

Caetano, C., Callaway, B., Payne, S., & Rodrigues, H. S. (2022). Difference in differences with time-varying covariates. *arXiv preprint arXiv:2202.02903*.

Callaway, B., & Karami, S. (2022). Treatment effects in interactive fixed effects models with a small number of time periods. *Journal of Econometrics*.

Cannon, S. E., & Cole, R. A. (2011). Changes in reit liquidity 1988–2007: evidence from daily data. *The Journal of Real Estate Finance and Economics*, *43*(1), 258–280.

Chan, K. C., Hendershott, P. H., & Sanders, A. B. (1990). Risk and return on real estate: evidence from equity reits. *Real Estate Economics*, *18*(4), 431–452.

Cotter, J., Gabriel, S., & Roll, R. (2015). Can housing risk be diversified? a cautionary tale from the housing boom and bust. *The Review of Financial Studies*, *28*(3), 913–936.

Currie, J., Davis, L., Greenstone, M., & Walker, R. (2015). Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *American Economic Review*, *105*(2), 678–709.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–96.

Di Maggio, M., & Kermani, A. (2017). Credit-induced boom and bust. *The Review of Financial Studies*, *30*(11), 3711–3758.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, *6*(4), 379–406.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, *81*(3), 607–636.

Favara, G., & Imbs, J. (2015). Credit supply and the price of housing. *American Economic Review*, *105*(3), 958–92.

Gobillon, L., & Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, *98*(3), 535–551.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Gormley, T. A., & Matsa, D. A. (2014). Common errors: How to (and not to) control for unobserved heterogeneity. *The Review of Financial Studies*, *27*(2), 617–661.

Hung, K., & Glascock, J. (2010). Volatilities and momentum returns in real estate. *Journal of Real Estate Finance and Economics*, *41*, 26-149.

Jayaratne, J., & Strahan, P. E. (1996). The finance-growth nexus: Evidence from bank branch deregulation. *The Quarterly Journal of Economics*, *111*(3), 639–670.

Karpoff, J. M., & Wittry, M. D. (2018). Institutional and legal context in natural experiments: The case of state antitakeover laws. *The Journal of Finance*, *73*(2), 657–714.

Kroszner, R. S., & Strahan, P. E. (1999). What drives deregulation? economics and politics of the relaxation of bank branching restrictions. *The Quarterly Journal of Economics*, *114*(4), 1437–1467.

Litterman, R., & Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of fixed income*, *1*(1), 54–61.

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, *232*(2), 272–299.

MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, *32*(2), 233–254.

Morgan, D. P., Rime, B., & Strahan, P. E. (2004). Bank integration and state business cycles. *The Quarterly Journal of Economics*, *119*(4), 1555–1584.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.

Nyborg, K. G., & Woschitz, J. (2023). Robust difference-in-differences analysis when there is a term structure.

Petersen, M. A. (2008). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of financial studies*, *22*(1), 435–480.

Peterson, J. D., & Hsieh, C.-H. (1997). Do common risk factors in the returns on stocks and bonds explain returns on reits? *Real Estate Economics*, *25*(2), 321–345.

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, rdad018.

Rice, T., & Strahan, P. E. (2010). Does credit competition affect small-firm finance? *The Journal of Finance*, *65*(3), 861–889.

Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance1. In *Handbook of the economics of finance* (Vol. 2, pp. 493–572). Elsevier.

Rosenbaum, P. R., & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, *45*(2), 212–218.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–22.

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2022). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *arXiv preprint* *arXiv:2201.01194*.

Roth, J., & Sant'Anna, P. H. C. (2023). When is parallel trends sensitive to functional form? *Econometrica*, *91*(2), 737–747.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199.

Titman, S., & Warga, A. (1986). Risk and the performance of real estate investment trusts: A multiple index approach. *Real Estate Economics*, *14*(3), 414–431.

Zevelev, A. A. (2021). Does collateral value affect asset prices? evidence from a natural experiment in texas. *The Review of Financial Studies*, *34*(9), 4373–4411.