Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

The PCDID Approach: Difference-in-Differences when Trends are Potentially Unparallel and Stochastic

Marc K. Chan & Simon S. Kwok

The PCDID Approach: Difference-in-Differences when Trends are Potentially Unparallel and Stochastic

Marc K. Chan[*,1], Simon S. Kwok[2]

[1]University of Melbourne

[2]The University of Sydney

We would like to thank Manuel Arellano, Mario Fiorini, John Ham, Rosa Matzkin, Tim Moore, Morten Nielsen, Tim Robinson, Chris Skeels, Yongcheol Shin, Katrien Stevens, Chris Taber, seminar participants at Adelaide, ANU, Deakin, HKU, Melbourne, Monash, NYUAD, Oxford, Sydney, UTS, and participants of the Econometrics Colloquium (UQ), IAAE, AASLE and IWEEE conferences for their valuable comments. The paper has greatly benefited from suggestions by the Editor, Associate Editor and the five referees. Special thanks to James Ziliak who provided the key years of welfare caseload data to us.

Chan: Department of Economics, Faculty of Business and Economics, University of Melbourne, Parkville VIC 3010, Australia. Email: marc.chan@unimelb.edu.au.

Kwok: School of Economics, The University of Sydney, NSW 2006 Australia; Email: simon.kwok@sydney.edu.au.

[*]marc.chan@unimelb.edu.au

Abstract

We develop a class of regression-based estimators, called Principal Components Differ-ence-in-Differences estimators (PCDID), for treatment effect estimation. Analogous to a control function approach, PCDID uses factor proxies constructed from control units to control for unobserved trends, assuming that the unobservables follow an interactive effects structure. We clarify the conditions under which the estimands in this regression-based approach represent useful causal parameters of interest. We establish consistency and asymptotic normality results of PCDID estimators under minimal assumptions on the specification of time trends. The PCDID approach is illustrated in an empirical exercise that examines the effects of welfare waiver programs on welfare caseloads in the US.

Keywords: principal components difference-in-differences, interactive fixed effects, factor-augmented regressions, treatment effects, parallel trends

1 Introduction

The difference-in-differences (DID) method is a workhorse for policy evaluation in empirical economics and other disciplines. Its key underlying assumption is that trends are parallel among the control and treated units. In this paper, we develop a class of regression-based estimators, called Principal Components Difference-in-Differences Estimators (PCDID), that can be applied to scenarios in which trends are potentially unparallel and stochastic among control and treated units. Unlike existing approaches such as synthetic control, unconfoundedness and matrix completion estimators, PCDID uses factor-augmented regressions to estimate specific parameters related to treatment effects. Specifically, PCDID does the following: (1) use a data-driven method (principal component analysis (PCA)) on data from control units to form factor proxies that capture the endogeneity arising from unparallel trends; (2) among treated unit(s), run regressions using the factor proxies as extra covariates. Our method is analogous to the control function approach in the microeconometric literature, in the sense that the factor proxies play the same role as control functions.

Our main theoretical findings are as follows. First, we clarify the conditions under which the estimands in this regression-based approach represent useful causal parameters of interest. This is useful because the recent literature has shown that, under treatment effect heterogeneity, a standard DID two-way fixed effect (2wfe) regression may yield an estimate that no longer represents a useful causal parameter of interest, but instead a weighted average of treatment effects where some of the weights can be negative (e.g., Borusyak and Jaravel (2017), Abraham and Sun (2018), Athey and Imbens (2018), de Chaisemartin and D' Haultfœuille (2018), Goodman-Bacon (2018)). Because our factor-augmented regression extends the 2wfe regression by incorporating unparallel trends, it is important to show that it targets meaningful causal parameters under reasonable assumptions. Second, based on the various ways the factor-augmented regression can be carried out, we consider three different PCDID estimators (basic, mean-group, pooled), and establish consistency and asymptotic normality results for each estimator with respect to its target causal parameter. Their differences in identification conditions and rates of convergence are clarified. Importantly, we show that asymptotic normality holds under minimal assumptions on the specification of trends, e.g., it encompasses nonstationary trends. Thus standard inference (e.g., t-tests with standard critical values) is valid and does not depend on the unknown nature of the time trends. Third, based on the PCDID approach, we develop a test of parallel trends under the functional form specification of the interactive effects model. We establish consistency and asymptotic normality results of the test statistic and, as in PCDID estimators, standard inference is valid and does not depend on the unknown nature of the time trends. We discuss how the test is related to the parallel trend assumptions in the fully nonparametric DID framework of Callaway and Sant'Anna (2018). We then consider Monte Carlo simulations and an empirical illustration that examines the effects of welfare waiver programs on welfare caseloads in the US. Extra results such as micro/group-level data are found in Online Appendix.

The PCDID approach is related to the large literature on factor-augmented regression models (e.g., Stock and Watson (2002), Bernanke et al. (2005), Bai and Ng (2006)), which forecast one or several time series using a large number of predictor series. We extend this literature to treatment effect estimation. Our stepwise implementation is also related to common correlated effects (CCE) estimators pioneered by Pesaran (2006). The key difference is that PCDID specifically exploits the DID data structure when it constructs factor proxies.[1] Other estimators in interactive effects models, such as Bai (2009) or Moon and Weidner (2015), are typically based on the assumption of homogeneous parameters, and they do not exploit the specific data structure as in our approach. Bai (2009) has recently been used or adapted for treatment effect estimation, e.g., Kim and Oka (2014), Gobillon and Magnac (2016). Its asymptotic properties under treatment effect heterogeneity and its robustness to different factor dynamics have not been fully studied.

The PCDID approach is related to the rapidly growing literature on synthetic control (SC), unconfoundedness and matrix completion estimators (e.g., Abadie et al. (2010), Hsiao et al. (2012), Xu (2017), Ferman and Pinto (2018), Ben-Michael et al. (2018), Athey et al. (2018), Chernozhukov et al. (2018), Arkhangelsky et al. (2019)). In this literature, both the control panel and pre-intervention data of treated units (an “L-shaped” matrix, as in Athey et al. (2018)) are used to impute the counterfactual outcomes of treated units in post-intervention periods; the treatment effect is then estimated as the difference between these counterfactual outcomes and observed outcomes. Athey et al. (2018) show that the original SC method (Abadie et al. (2010)), vertical regression (SC methods) and horizontal regression (unconfoundedness methods) all belong to the same class of matrix completion methods based on matrix factorization (i.e., low-rank matrices) with different restrictions/regularizations. Gobillon and Magnac (2016) and the generalized SC estimator (GSC) of Xu (2017) also consider this step-wise approach, by first applying Bai’s estimator on the control panel and pre-intervention data of treated units to form counterfactual outcomes of treated units in post-intervention periods.

Unlike the above approaches, PCDID exploits the data structure in the form of two rectangular-shaped matrices. It first constructs control functions (factor proxies) from the control panel exclusively. Then, it runs a factor-augmented regression using all (pre- and post-intervention) data of treated units, with the control functions as additional covariates. By using the entire sample period for estimation, PCDID is advantageous in terms of reducing finite sample bias and improving asymptotic efficiency. This is especially true when policy intervention is potentially correlated with the unobserved factors, a case where the post-intervention counterfactual outcome is tricky to estimate. For example, in our application where waiver programs were introduced when welfare caseloads started to reverse trends, PCDID tends to yield more credible estimates than GSC. The regression-based nature of PCDID estimators, however, necessitates assumptions on treatment effect dynamics to ensure convergence to well-defined causal parameters of interest. While we also propose extensions of PCDID regressions to capture such dynamics, we recommend researchers to use other approaches if they suspect that the dynamics are complicated.

We note that all our results are derived under the specification assumption of interactive effects, i.e., the unobservables can be factorized as described in Athey et al. (2018). Moreover, as in this literature we require some never-treated units and we focus on binary treatment for analytical tractability (see our pooled estimator for discussion). Interactive effects impose a structure on inferring how the unparallel trends behave, based on observed variations in outcomes across units and time. They are less general than many other structures for the unobservables. In fact, Callaway and Sant'Anna (2018) show that, in a fully nonparametric DID setting, the parallel trend assumption is untestable but a stronger, augmented version of this assumption is testable. Interactive effects are more general than two-way fixed effects but it comes with costs; namely, large-T asymptotics are typically required, and the asymptotic distributions and valid inference procedures are non-trivial under weak assumptions on time trends.[2] Whether interactive effects are reasonable and important depend on the specific application. For instance, in the welfare caseload example, each state's pre-existing welfare program differs in generosity and structure and the potential recipient populations are different. This implies that variations in trends that represent macroeconomic conditions or national sentiment on welfare policy (e.g., changes in welfare stigma) will have heterogeneous effects on welfare participation in different states, resulting in more elastic responses in some states than others.

The paper is organized as follows. Section 2 describes the basic framework. Section 3 discusses the full model. Section 4 discusses the PCDID estimators and the parallel trend test. Section 5 reports results from Monte Carlo simulations. Sections 6 discusses the empirical analysis. Section 7 concludes. Supplementary materials are included in an Online Appendix.

2 Basic Framework

We first illustrate the basic framework using potential outcomes. There are *N* units and *T* periods. Let $y_{it}(1)$ and $y_{it}(0)$ be the potential outcomes of unit *i* in period *t* with and without treatment, respectively. The treatment effect (TE) is denoted by $\Delta_{it} := y_{it}(1) - y_{it}(0)$. Let D_{it} = 1 if unit *i* receives treatment in period *t*, D_{it} = 0 otherwise. The observed outcome is $y_{it} = D_{it} y_{it}(1) + (1 - D_{it}) y_{it}(0)$. Denote the set of never-treated units as *C* (called "control units"; there are N_C such units), and the rest of units as *E* (called "treated units"; there are N_E such units). Suppose once a unit receives treatment, it remains so thereafter, i.e., D_{it} = 1 implies $D_{i,t+k} = 1$ for all *k* > 0. Let a treated unit $i \in E$ have T_{0i} pre-intervention periods and T_{1i} post-intervention periods, i.e., its first treated period is $T_{0i} + 1$. Note that T_{0i} can be the same or differ across $i \in E$. We can then express D_{it} as a product of two indicator functions: $D_{it} = 1_{\{i \in E\}} 1_{\{t > T_{0i}\}}$. The observed outcome can be rewritten as $y_{it} = \Delta_{it} 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + y_{it}(0)$.

We now impose a key functional form specification on potential outcome: $y_{it}(0) = \varsigma_i + \mu_i' f_t + \tilde{\epsilon}_{it}$, where ς_i, μ_i, f_t, and $\tilde{\epsilon}_{it}$ are all individually unobserved. The term $\mu_i' f_t$ is known as a factor structure or interactive effects, which contains an $\ell \times 1$ vector of factor loadings μ_i and time-varying factors f_t. The observed outcome is:

$$y_{it} = \Delta_{it} 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + \varsigma_i + \mu_i' f_t + \tilde{\epsilon}_{it}, \qquad (1)$$

The factor structure is used in various literatures such as the SC and matrix completion literatures (e.g., Xu (2017), Ferman and Pinto (2018), Ben-Michael et al. (2018), Athey et al. (2018), Chernozhukov et al. (2018), Arkhangelsky et al. (2019)). In the latter literatures, $\mu_i' f_t$ is seen as a low-rank (less complex) matrix which, together with idiosyncratic noise $\tilde{\epsilon}_{it}$, generates $y_{it}(0)$.

Equation (1) reduces to the two-way fixed effects (2wfe) model when factor loadings are homogeneous across units: $\mu_i = \mu_0$ for all i. The factor structure is then $\mu_0' f_t$, which can be re-expressed as a scalar time fixed effect τ_t, yielding $y_{it} = \Delta_{it} 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + \varsigma_i + \tau_t + \tilde{\epsilon}_{it}$. In this case, regardless of the specification of f_t, the trend τ_t can be eliminated in a 2wfe regression (when $\Delta_{it} \equiv \Delta_0$). Recent work on this model have also shown that, when Δ_{it} is heterogeneous, a standard 2wfe regression yields an estimate that may no longer represent a useful causal parameter of interest, but instead a weighted average of treatment effects where some of the weights can be negative.

We assume factor loadings are heterogeneous, which permits trends to be potentially unparallel across units, albeit parsimoniously via the factor structure. The benchmark is what we call "weak parallel trend (PTW)" under the functional form specification in (1):

$$\textbf{PTW}: \qquad E(\mu_i \mid i \in E) = E(\mu_i \mid i \in C), \qquad (2)$$

where the average factor loadings are the same between control and treated units. Section 4.4 provides a formal analysis of PTW when covariates are present, and discusses its relationship with the parallel trend assumption in a fully nonparametric framework (Callaway and Sant'Anna (2018)).[3]

The factor structure is part of the error term in a 2wfe regression. For illustration, consider the simplest possible setting (homogeneous TE and intervention date, single factor, etc.) where we can convert it into a nuisance term $\left(\bar{\mu}_E - \bar{\mu}_C\right)\left(\bar{f}_{post} - \bar{f}_{pre}\right)$, which is the difference in average loading between control and treated units multiplied by the difference in average factor between pre- and post-intervention periods. The consistency of 2wfe is determined by the limit of this term being zero, which depends crucially on whether PTW holds and the factor specification. The possibility of stochastic trends complicates the analysis further. For example, if f_t is a random walk, then $\left(\bar{f}_{post} - \bar{f}_{pre}\right)$ is divergent as T increases, implying that even the slightest deviation from PTW can result in a divergent nuisance term that is detrimental to 2wfe regression. A key feature of PCDID is that it controls for the factor structure under minimal assumptions on factor specification.

A related issue is efficiency. While a direct comparison between 2wfe and PCDID is generally not possible, we expect PCDID to be more efficient because it controls for the factor structure (see also Section 5). There is a special case in which 2wfe is more efficient than PCDID – when factor loadings are homogeneous ($\mu_i = \mu_0 \forall i$). In Online Appendix, we show this result formally via the Hausman equality, based on the insight that the nuisance term is identically zero.

Unlike synthetic-control type estimators, PCDID estimators are regression-based. It is therefore important to show that the PCDID regression coefficients converge to meaningful causal parameters of interest. Our key results show that the key causal parameters identified and estimated are ITET and ATET, defined as

$$\textbf{ITET}: \quad \bar{\Delta}_i := E(\Delta_{it} \mid t > T_{0i}) \text{ for fixed i} \in \text{E}, \qquad (3)$$

$$\textbf{ATET}: \quad \bar{\bar{\Delta}} := E(\bar{\Delta}_i \mid i \in E). \qquad (4)$$

The basic building block is the ITET $\bar{\Delta}_i$, which is the treatment effect of a unit $i \in E$ averaged over post-intervention periods. The ATET $\bar{\bar{\Delta}}$ is the average of the ITET across units $i \in E$. Except in ITET analysis where a treated unit *i* is taken as given (and hence μ_i is non-random), we consider all quantities indexed by *i* as random across units. Under weak regularity conditions, the basic PCDID estimator converges to $\bar{\Delta}_i$ when $N_C, T \to \infty$ (Theorem 1), whereas the simple mean-group PCDID estimator (PCDID-MG) converges to $\bar{\bar{\Delta}}$ when $N_C, N_E, T \to \infty$ (Theorem 2). Moreover, the t-statistic from these estimators converges to a standard normal distribution and does not depend on the specification of f_t.

There are many alternative definitions of causal parameters, some of which are more natural under certain settings. For example, Callaway and Sant' Anna (2018) consider DID with multiple time periods in a fully nonparametric setting. They introduce *group-time average treatment effects* (in our notations, $ATT(g,t) := E(\Delta_{it} \mid T_{0i}+1=g)$), which is identified under the assumption that the potential outcome $y_{it}(0)$ satisfies a parallel trend assumption for all $t \geq g$ given observed covariates (note that this assumption is violated in our context because the factor structure generates potentially unparallel trends). They show how this definition can be used to examine general dynamic treatment effects and aggregated into various summary measures of causal effects in the related literature. Reflecting the nature of our estimation method, our causal parameters consider aggregation over the time dimension first (ITET) before the cross-sectional dimension (ATET), an approach that has been considered by some studies, e.g., Goodman-Bacon (2018).

Although some of the modeling assumptions that we introduce are rather technical, the key intuition that underlies our method is analogous to the control function approach in the microeconometric literature. We use data from control units ($i \in C$) to form $\ell \geq 1$ control functions that capture the unobserved factor structure, $\mu_i' f_t$, which creates the endogeneity problem as illustrated earlier. Then, in a factor-augmented regression for $i \in E$, the endogeneity is corrected by including the control functions as extra covariates in the regression.

3 The Model

We consider an extension of equation (1), assuming the potential outcome $y_{it}(0)$ is a linear function of covariates:

$$y_{it} = \Delta_{it} 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + \beta_i' x_{it} + \varsigma_i + \mu_i' f_t + \tilde{\epsilon}_{it}, \qquad (5)$$

where the $k \times 1$ vector x_{it} stores the time-varying covariates with unit-specific parameters β_i. The TE is decomposed as

$$\Delta_{it} = \bar{\Delta}_i + \tilde{\Delta}_{it}, \quad (6)$$

where $\bar{\Delta}_i := E(\Delta_{it} \mid t > T_{0i})$ is the ITET of unit *i*, a key estimand in this paper. The term $\tilde{\Delta}_{it}$ represents the deviation of Δ_{it} from the ITET, and it is the demeaned, time-varying idiosyncratic component of Δ_{it} for $t > T_{0i}$; by construction, $E(\tilde{\Delta}_{it} \mid t > T_{0i}) = 0$ for each $i \in E$. Substituting $\Delta_{it} = \bar{\Delta}_i + \tilde{\Delta}_{it}$ into (5) yields a reduced-form model, which is the main model that we examine:

$$y_{it} = \bar{\Delta}_i 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + \beta_i' x_{it} + \varsigma_i + \mu_i' f_t + \epsilon_{it}, \qquad (7)$$

where $\epsilon_{it} := \tilde{\Delta}_{it} 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + \tilde{\epsilon}_{it}$ is the idiosyncratic error in the reduced-form model. The composite term $\tilde{\Delta}_{it} 1_{\{i \in E\}} 1_{\{t > T_{0i}\}}$ equals $\tilde{\Delta}_{it}$ when $i \in E$ and $t > T_{0i}$, and it equals zero otherwise. It does not contain $\tilde{\Delta}_{it}$ for any $i \in C$ or $t \le T_{0i}$. The decomposition in equation (6) implies that the composite term has zero mean. To identify ITET and ATET, the errors ϵ_{it} need to satisfy some assumptions. The first key assumption is:

Assumption E (predeterminedness, treatment and intervention statuses):

(i) $E(\epsilon_{it} \mid \bar{\Delta}_i, 1_{\{i \in E\}}, 1_{\{t > T_{0i}\}}, \beta_i, \varsigma_i, \mu_i, f_t, x_{it}) = 0$ for each *i* and *t*.

(ii) $0 < E(1_{\{i \in E\}}) < 1$ for each *i*.

(iii) for each $i \in E$, $T_{1i} / T \xrightarrow{p} \kappa_i$ as $T, T_{1i} \to \infty$, where $0 < \kappa_i < 1$.

Assumption E(i) exerts orthogonality between the idiosyncratic errors ϵ_{it} $(= \tilde{\epsilon}_{it} + \tilde{\Delta}_{it}$) and the conditioning components.[4] Importantly, it imposes restrictions on treatment effect dynamics. Given each $i \in E$, the idiosyncratic TE component $\tilde{\Delta}_{it}$ is orthogonal to f_t and x_{it} over $t > T_{0i}$, i.e., TE dynamics is temporally uncorrelated with factors and covariates. In addition, **E(i)** rules out long-run time trends in TE dynamics. Section 4.5 shows how these restrictions can be relaxed by modifying the PCDID regressions.

Note that **E(i)** does not preclude the conditioning components from being correlated with one another. For example, loadings may differ between control and treated units ($1_{\{i\in E\}}$, μ_i); factors may differ between pre- and post-intervention periods ($1_{\{t>T_{0i}\}}$, f_t); the ITET, policy intervention date and loadings may be correlated ($\bar{\Delta}_i, T_{0i}$, μ_i). Similarly, the time-varying covariates x_{it} may be correlated (but not collinear) with all the other conditioning components.[5] To the extent that correlation is not equivalent to causation, we emphasize that x_{it} *cannot* be affected (causally) by the treatment. This assumption is implicit in our model specification that the untreated potential outcome $y_{it}(0)$ is a linear function of x_{it}. Relaxation of this assumption is out of our scope and it is left for future research.

Assumption E(ii) is similar to Gobillon and Magnac (2016). It assumes the presence of control (i.e., never-treated) units, ruling out more extreme forms of staggered adoption in which all units receive treatment eventually; the literature has shown that, in such cases, the DID-2wfe estimator may represent uninterpretable weighted averages of treatment effects (Section 2). **Assumption E(iii)** assumes the proportions of pre- and post-intervention periods do not vanish to zero in the limit.

Next, we discuss the assumptions on factors and factor loadings. Denote $F := [f_1, \ldots, f_T]'$ and $X_i := \begin{bmatrix} 1 & ,..., & 1 \\ x_{i1} & ,..., & x_{iT} \end{bmatrix}'$, and let $G_i := (F, X_i)$. Denote $\mu_C := [\mu_{N_E+1}, \ldots, \mu_N]'$. As a convention, the norm of a matrix *A* is given by $\|A\| = [trace(A'A)]^{1/2}$, and "$a.s.$" stands for "almost surely".

Assumption F (factors and covariates): Let $\Gamma := diag(T^{r_1}, \ldots, T^{r_\ell}, T^{0.5}, \ldots, T^{0.5})$ where $r_1, \ldots, r_\ell \geq 0.5$. For each *i*, the following conditions are satisfied:

(i) For all *T*, $E\left\|\Gamma^{-1}G_i'G_i\Gamma^{-1}\right\|^2 \leq c$ for some constant *c* > 0.

(ii) $\text{plim}_{T\to\infty}\Gamma^{-1}G_i'G_i\Gamma^{-1}$ is positive definite $a.s.$.

Assumption FL$_C$ (factor loadings of control units): The following conditions are satisfied:

(i) For all $i \in C, E\|\mu_i\|^2 \leq c$ for some constant $c > 0$.

(ii) $\text{plim}_{N_C \to \infty} \frac{1}{N_C} \mu_C' \mu_C$ is positive definite.

Assumption F(i) accommodates a wide range of factor dynamics (deterministic and random, stationary and nonstationary) with possibly heterogeneous normalization orders. The factors may contain a unit root (e.g., Bai (2004)), display long-range dependence (e.g., Ergemen and Velasco (2017), Ergemen (2019)), or exhibit structural breaks (e.g., Chen et al. (2014)). **Assumption F(ii)** ensures sufficient variability in factors and covariates and assumes linear independence, ruling out multicollinearity. **Assumption FL$_C$(i)** is a standard moment condition on the factor loadings of control units. **Assumption FL$_C$(ii)** ensures sufficient variability and assumes linear independence in factor loadings among control units as the number of control units grow large. Both **Assumptions F** and **FL$_C$(ii)** are crucial for asymptotic identification of the linear factor space (of dimension ℓ). Given the assumptions, the model given by (5)-(7) rules out the possibility that some treated units are exposed to factors that do not affect control unit outcomes. The factors and factor loadings may be deterministic or random – in the former case, the probability limits in the assumptions reduce to deterministic limits.

We want to highlight that identifying ITET and ATET requires identification of $\mu_i' f_t$ but not μ_i and f_t separately. This is evident from (5). Although μ_i and f_t are only separately identified up to a rotation matrix, we are not interested in, nor required to, identify and estimate them separately. Because $\mu_i' f_t = \mu_i' R R^{-1} f_t$ for all invertible rotation matrix R, it follows that both ITET and ATET are uniquely determined regardless of the choice of R.

4 Estimation Procedure and Key Results

4.1 Construction of Factor Proxies

The PCDID approach uses the following procedure to construct factor proxies, which serve as control functions in factor-augmented regression:

1. For each $i \in C$, perform a linear projection of y_{it} on x_{it} using data from $t = 1,\ldots,T$. Obtain the residuals $\hat{u}_{it}$ and form the $N_C \times T$ residual matrix $\hat{u}_C$.

2. Apply PCA on the $N_C \times N_C$ sample covariance matrix $S := \hat{u}_C \hat{u}_C' / T$. With a pre-specified p, construct the $T \times p$ factor proxy matrix $\hat{F} := \frac{\hat{u}_C' W}{N_C}$, where W contains p columns of eigenvectors associated with the p largest eigenvalues of S.

The factor proxies (columns of $\hat{F}$) can be viewed as weighted averages of the residuals from step 1 over the control panel. The sets of weights are determined by PCA such that the p linear combinations of the residuals jointly explain the most time-variation of control unit outcomes after partialling out the covariates. The residuals are, by construction of projection, orthogonal to the covariates. As long as the covariates are not linear combinations of the factors in model (7) (as implied by Assumption F(ii)), the residuals will preserve the rank and other information of the factor space. Note that the covariates and factors may be correlated. For instance, in Online Appendix, we consider a model in which the covariates are a linear function of factors plus idiosyncratic errors: $X_i = F\Pi_i + V_i$, which is a special case studied by Pesaran (2006) and Bai (2009). The PCDID procedure remains valid for this example.

The total number of factors ℓ and the factor dynamics (as characterized by the normalization orders $r_1,\ldots,r_\ell$) are unknown to users. To extract factors of different dynamical complexity, we implement the above procedure in a recursive and adaptive manner (see Online Appendix). It uses the growth ratio (GR) test of Ahn and Horenstein (2013) to determine the unknown number of factors associated with each level of dynamical complexity. We also implement conservative versions of the GR test, motivated by Moon and Weidner (2015)'s result that overestimation of ℓ is less problematic than underestimation.[6] The finite-sample performance of the procedure is evaluated in simulations (Section 5.1).

4.2 ITET Estimation

We now discuss the intuition behind the identification and estimation of ITET $\bar{\Delta}_i := E(\Delta_{it} \mid t > T_{0i})$, which is our causal parameter of interest for a given treated unit $i \in E$. All quantities indexed by $i \in E$ (e.g., μ_i) is viewed as given. The factor proxies $\hat{F}$ span the same linear space as the true factors F in the limit as the control panel grows in both the time and cross-sectional dimensions (see lemma A4(d)-(g) in Online Appendix). This large-sample result justifies the use of $\hat{F}$ as an approximation for F, and motivates the following factor-augmented time series regression (*aka PCDID regression*) for treated unit $i \in E$ using data from $t = 1,..,T$:

$$y_{it} = b_{0i} + \delta_i 1_{\{t > T_{0i}\}} + a_i' \hat{f}_t + b_{1i}' x_{it} + e_{it}, \quad (8)$$

where the $p \times 1$ vector $\hat{f}_t$ is the transpose of the t^{th} row of $\hat{F}$.

We now present the asymptotic theory of estimating ITET $\bar{\Delta}_i$. We rewrite the PCDID regression (8) in vector form. Let $1_{post,i}$ be the $T \times 1$ vector consisting of T_{0i} zeros followed by T_{1i} ones. For treated unit $i \in E$,

$$y_i = \delta_i 1_{post,i} + \hat{F} a_i + X_i b_i + e_i, \quad (9)$$

where $y_i = [y_{i1}, \ldots, y_{iT}]'$, $e_i = [e_{i1}, \ldots, e_{iT}]'$, $b_i = [b_{0i}, b_{1i}']'$, and $\hat{F} := \hat{u}_C' W / N_C$, the $T \times \ell$ factor proxy matrix. The PCDID estimator of $\bar{\Delta}_i$ is $\hat{\delta}_i$, the least squares estimator of δ_i:

$$\hat{\delta}_i = (1_{post,i}' M_{[\hat{F}, X_i]} 1_{post,i})^{-1} 1_{post,i}' M_{[\hat{F}, X_i]} y_i, \qquad (10)$$

where $M_A = I - A(A'A)^{-1} A'$ for a given matrix A.

In the following, we list and discuss the key assumptions (**Assumptions AI*$_i$*** and **ES**) for the asymptotic analysis of PCDID estimation. Define $\tilde{G}_i = \{1_{post,i}, G_i\} = \{1_{post,i}, F, X_i\}$.

Assumption AI*$_i$* (asymptotic identification, PCDID estimator): For each $i \in E$,

(i) $\rho_i := \mathrm{plim}_{T\to\infty} \dfrac{1}{T} 1_{post,i}^{\prime} M_{G_i} 1_{post,i}$ exists and is strictly positive $a.s.$.

(ii) $\xi_i^2 := \mathrm{plim}_{T\to\infty} E\left(\left\| \dfrac{1}{\sqrt{T}} 1_{post,i}^{\prime} M_{G_i} \epsilon_i \right\|^2 \middle| \tilde{G}_i \right)$ exists and is strictly positive $a.s.$.

Assumption ES (strict exogeneity in time): For each *i*, $E(\epsilon_{it} \mid \tilde{G}_i) = 0$ $a.s.$ for all *t*.

Assumption AI*$_i$*(i) is key to asymptotic identification of $\bar{\Delta}_i$. It rules out multi-collinearity of the intervention status $1_{post,i}$ with X_i and F in the limit. Multi-collinearity occurs when X_i or F is a step function that jumps at $T_{0i}+1$ but constant otherwise, which is rare in practice. Gobillon and Magnac (2016) imposes an analogous condition under finite *T* without covariates. **Assumption AI*$_i$*(ii)** ensures that $\hat{\delta}_i$ has a non-degenerate distribution after normalization. It rules out unbounded TE dynamics (see Section 4.5). **Assumption ES** strengthens **Assumption E(i)** to strict exogeneity of ϵ_{it} $(=\tilde{\epsilon}_{it} + \tilde{\Delta}_{it})$ on the time series of *F*, X_i and $1_{post,i}$. This is crucial for the conditioning argument that leads to our asymptotic normality result under nonstationary factors. Given each $i \in E$, it assumes that the time-varying TE component $\tilde{\Delta}_{it}$ is orthogonal to contemporaneous, leads and lags of f_t and x_{it} over $t > T_{0i}$.

More technical assumptions (**Assumptions IE, M, D*$_i$***) are detailed in Online Appendix due to space limitation. These are regularity conditions that are relatively standard in the interactive effects and time series literatures. We briefly summarize them below and highlight the key properties and restrictions. **Assumption IE** accommodates heteroskedasticity and weak dependence (e.g., cross-sectional and serial correlations) in ϵ_{it}, but rules out factor structure in TE dynamics. It assumes the higher-order moments of ϵ_{it} are bounded, which again rules out unbounded TE dynamics. **Assumption M** exerts control on the dependence among various model components (idiosyncratic errors, factors, factor loadings, covariates and their cross-products). For analytical convenience, covariates and factors are assumed to be orthogonal (**Assumption MX(iii)**), although this is not necessary for the PCDID approach to deliver valid results (see Section 4.1). **Assumption D$_i$** governs the dynamical properties of regressors, factors and idiosyncratic errors by allowing them to be mixing and heteroskedastic over time. This enables us to apply the functional central limit theorem of DeJong and Davidson (2000).

Theorem 1 states the consistency and asymptotic normality results related to $\hat{\delta}_i$:

Theorem 1. Suppose Assumptions E, F, FL$_C$, AI$_i$, IE and M hold. Then, as $T, N_C \to \infty$ jointly and $\frac{\sqrt{T}}{N_C} \to 0$, we have for each $i \in E$:

(a) $\hat{\delta}_i \xrightarrow{p} \bar{\Delta}_i$.

(b) $\sqrt{T}\sigma_{Ti}^{-1}(\hat{\delta}_i - \bar{\Delta}_i) \xrightarrow{d} N(0,1)$ if additionally Assumptions ES and D$_i$ hold, where $\sigma_{Ti}^2 := \frac{\xi_{Ti}^2}{R_{Ti}^2}$, $\xi_{Ti}^2 := \frac{1}{T} 1_{post,i}' M_{G_i} \Omega_{ii} M_{G_i} 1_{post,i}$, $\Omega_{ii} = Var(\epsilon_i \mid \tilde{G}_i)$ and $R_{Ti} := \frac{1}{T} 1_{post,i}' M_{G_i} 1_{post,i}$.

The consistency and asymptotic normality result is robust to a wide range of factor dynamics, including nonstationary processes. This may seem surprising given that the DID estimator is potentially inconsistent when the factors contain a unit root (see Section 2). The intuition behind the result is that the outcomes and factors form a cointegrating relationship in our model. Since the time dummy $1_{post,i}$ is bounded, its coefficient δ_i can be estimated consistently by least squares method at the $\sqrt{T}$-rate regardless of the factor dynamics.

The factor proxies contain not only *F* but also idiosyncratic errors of control units. The “measurement error” thus introduced creates an endogeneity issue when the idiosyncratic errors of treated and control units are correlated. The asymptotic condition $\sqrt{T} / N_C \to 0$ is necessary to remove the asymptotic bias due to the measurement error.

An important implication of Theorem 1(b) is that the limiting distribution of the *studentized statistic* remains to be *N*(0, 1) under a wide range of factors (stationary and nonstationary) allowed by **Assumption F**.[7] Inference can be carried out using the studentized form of the PCDID estimate and standard critical values. To compute the *t*-statistic, the population standard deviation σ_{Ti} may be replaced by a suitable sample analog that reflects the error dependence structure, e.g., the Newey-West HAC standard error is nonparametric and is widely used in the time series literature. An alternative approach is bootstrapping the *t*-statistic.[8] More details about inference procedures are given in Section 5.2 and Online Appendix.

4.3 ATET Estimation

We now discuss the identification and estimation of ATET $\bar{\bar{\Delta}} := E(\bar{\Delta}_i \mid i \in E)$, the key estimand in this section.[9] This estimand is built on the ITET $\bar{\Delta}_i$ (see Section 2), now viewed as random across $i \in E$ (similarly for μ_i, etc.). We break down the treatment effect Δ_{it} using decomposition (6) and the definition of ATET, obtaining

$$\Delta_{it} = \bar{\Delta}_i + \tilde{\Delta}_{it} = \bar{\bar{\Delta}} + \upsilon_i + \tilde{\Delta}_{it}, \quad (11)$$

where $\upsilon_i := \bar{\Delta}_i - \bar{\bar{\Delta}}$ is the unit-specific deviation of the ITET from the ATET. By construction, $E(\upsilon_i \mid i \in E) = 0$. Our principal case is *heterogeneous ITET over treated units*, i.e., υ_i varies over $i \in E$. The Online Appendix discusses the case of homogeneous ITET over treated units, i.e., $\upsilon_i = 0$ for all $i \in E$, which is between our principal case and full homogeneity of Δ_{it} (i.e., $\Delta_{it} = \Delta_0$).

4.3.1 Simple Mean-group Estimator

The *simple mean-group estimator* (PCDID-MG) is defined as the simple average of ITET estimates:

$$\hat{\delta}^{mg} = \frac{1}{N_E}\sum_{i\in E}\hat{\delta}_i. \qquad (12)$$

The assumptions (listed in Appendix) are similar to those imposed for ITET estimation. We highlight the key differences below. To account for multiple treated units, **Assumptions FL** and **AI**$_{mg}$ involve relatively standard regularity conditions on treated units. For example, the key asymptotic identification **Assumption AI**$_{mg}$**(i)** rules out multi-collinearity of $1_{post,i}$ with X_i and *F for each* $i \in E$ (so that $\hat{\delta}_i$ is well defined), and **AI**$_{mg}$**(ii)** assumes that TE dynamics are bounded for each $i \in E$. The same principle applies to the technical assumption **MM** (Online Appendix). To derive consistency and asymptotic normality, **Assumption RT**$_{mg}$ imposes a set of weak regularity conditions that place control on the cross-sectional variation and higher-order moments of the term υ_i ($:= \bar{\Delta}_i - \bar{\bar{\Delta}}$). No parametric assumptions are involved. Importantly, none of the assumptions preclude $\bar{\Delta}_i$ from being cross-sectionally correlated with $1_{\{i\in E\}}, T_{0i}$, β_i, ς_i, μ_i, and x_{it}.

The asymptotic result for PCDID-MG is presented below. Note that we no longer need **Assumptions ES** and **D**$_i$ to derive asymptotic normality.

Theorem 2 (**simple mean-group estimator**). Suppose υ_i varies over $i \in E$ and satisfies Assumption RT$_{mg}$, and Assumptions E, F, FL, AI$_{mg}$, IE and MM hold. As $T, N_E, N_C \to \infty$ jointly and $\frac{T}{N_C} \to 0$, we have the following results:

(a) $\hat{\delta}^{mg} \xrightarrow{p} \bar{\bar{\Delta}}$.

(b) $\sqrt{N_E}\,\tilde{\varsigma}_{N_E}^{-1}(\hat{\delta}^{mg} - \bar{\bar{\Delta}}) \xrightarrow{d} N(0,1)$, where $\bar{\tilde{\varsigma}}_{N_E}^{2} := Var\left(\frac{1}{\sqrt{N_E}}\sum_{i\in E}\upsilon_i\right)$.

A number of remarks are in order. First, PCDID-MG is $\sqrt{N_E}$ -consistent under heterogeneous ITET. This is because υ_i dominates the idiosyncratic errors ϵ_{it} in the limit. Second, compared to ITET estimation, a stronger asymptotic condition $T / N_C \to 0$ is required to remove the asymptotic bias due to the estimation error of factor proxies. This is necessary when idiosyncratic errors in the treated and control panels are correlated. Third, as in Theorem 1, the limiting distribution of the studentized PCDID-MG statistic remains to be $N(0, 1)$ under a wide range of factors (stationary and nonstationary), which greatly facilitates inference procedures. This justifies the nonparametric variance estimator as in Pesaran (2006): $Var(\hat{\delta}^{mg}) = \frac{1}{N_E(N_E-1)}\sum_{i\in E}(\hat{\delta}_i - \hat{\delta}^{mg})^2$, which provides a convenient way to form the t-statistic. We show its consistency when υ_i are uncorrelated across $i \in E$. More details regarding inference can be found in Online Appendix.

Theorem 3. Suppose the assumptions in Theorem 2 holds, and that $Cov(\upsilon_i, \upsilon_j) = 0$ for $i \neq j$. Then $N_E Var(\hat{\delta}^{mg}) - \bar{\varsigma}^2_{N_E} \xrightarrow{p} 0$ as $T, N_E, N_C \to \infty$ jointly and $\frac{T}{N_C} \to 0$.

In Online Appendix, we show that the asymptotic efficiency of the PCDID-MG estimator depends on the extent of treatment effect heterogeneity in the model. We consider the case of homogeneous ITET, which bridges the interactive effects estimator of Bai (2009) where the analysis is confined to the case of homogeneous slope parameters (corresponding to full TE homogeneity ($\Delta_{it} = \Delta_0$) in our context) and stationary factors. In this case, the idiosyncratic errors are no longer dominated in the limit, necessitating a different approach. Invoking extra assumptions and the joint central limit theorem of Phillips and Moon (1999), we show that the PCDID-MG estimator achieves a faster $\sqrt{N_E T}$ rate of convergence, a result comparable to Theorem 2 in Bai (2009).

4.3.2 Pooled Estimator: An Evaluation

A more "familiar" approach involves a factor-augmented *panel* regression using data from the treated panel: for $i \in E$ and $t = 1,..T$,

$$y_{it} = b_{0i} + \delta 1_{\{t>T_{0i}\}} + a_i' \hat{f}_t + b_{1i}' x_{it} + e_{it}. \quad (13)$$

The *pooled estimator* is the least squares estimator of δ given by

$$\hat{\delta}^{pl} = \left(\sum_{i \in E} 1_{post,i}' M_{[\hat{F},X_i]} 1_{post,i} \right)^{-1} \sum_{i \in E} 1_{post,i}' M_{[\hat{F},X_i]} y_i. \qquad (14)$$

which is analogous to the widely popular least square estimator in 2wfe regression. Thus it is useful to derive the theoretical properties of the pooled estimator (see Online Appendix). For the principal case of heterogeneous ITET, we have two main findings. First, the pooled estimator requires extra assumptions to identify ATET unless, say, when $T_{0i} = T_0$ (homogeneous intervention date) or T_{0i} is randomly assigned as in an experiment. This is because $\hat{\delta}^{pl}$ generally identifies a weighted function of ITETs; a similar condition can be found in Goodman-Bacon (2018) for DID models. Second, and somewhat surprisingly, PCDID-MG is asymptotically more efficient than the pooled estimator (provided that identifying assumptions are satisfied for both). These properties suggest that PCDID-MG is more useful in applied research.

For completeness, the Online Appendix also considers the case of homogeneous ITET. Interestingly, we find that both estimators identify the ATET under similar assumptions. Also, PCDID-MG becomes asymptotically *less* efficient than the pooled estimator. These findings consolidate the idea that the properties of estimators depend heavily on the extent of treatment effect heterogeneity in the model. On a side note, if researchers are worried about unparallel trends but (unrealistically) *not* treatment effect heterogeneity, the pooled estimator can readily incorporate non-binary and/or multiple treatments, via running a panel multiple regression with the treatment variable(s) and factor proxies as covariates.

4.4 Parallel Trend Test under the Factor Structure

The techniques from PCDID approach enable us to develop a test of what we call "weak parallel trends (PTW)" under the functional form specification in equation (5):

$$\textbf{PTW}: \mathrm{E}(\mu_i | i \in C) = \mathrm{E}(\mu_i | i \in E) =: \mu_0 \text{ for some finite and non-zero vector } \mu_0,$$

which posits that the expected factor loadings are the same between control and treated units. We first perform a factor-augmented time-series regression for each $j \in E$:

$$y_{jt} = b_{0j} + b_{1j}' x_{jt} + \delta_j 1_{\{t > T_{0j}\}} + a_j \bar{u}_{Ct} + e_{jt}. \qquad (15)$$

where $\bar{u}_{Ct} := \frac{1}{N_C} \sum_{i \in C} \hat{u}_{it}$ is the cross-sectional average of control panel residuals $\hat{u}_{it}$ (see step 1 of the PCDID procedure in Section 4.1). Note that $\bar{u}_{Ct}$ is a scalar term, even though there may be multiple factors in the model. After obtaining the OLS estimator $\hat{a}_j$ of a_j, we define the *Alpha statistic* as the simple mean-group estimator

$$\hat{a}^{mg} := \frac{1}{N_E} \sum_{j \in E} \hat{a}_j.$$

To motivate the test, note that any $\ell \times 1$ factor loading of treated unit j can be uniquely represented (upon normalization) by $\mu_j = \alpha_j E(\mu_i | i \in C) + v_j$ where α_j is a *scalar* and v_j is an $\ell \times 1$ vector. Under PTW, we have $\alpha := E(\alpha_j | j \in E) = 1$ and $E(v_j | j \in E) = 0$ (a zero vector); the converse is also true.[10] The asymptotic result that $\hat{a}^{mg}$ estimates $\alpha = 1$ under PTW forms the basis of *Alpha test*.

In Online Appendix, we show that the Alpha test is powerful against various departures from PTW. Its test performance (including size and power) is also not affected by the specific normalization applied to the factors and factor loadings. This follows from the rotational invariance of the factor structure $\mu_i' f_t$ (i.e., for any

$\ell \times \ell$ invertible matrix *R*, we have $\mu_i' f_t = \mu_i' R R^{-1} f_t$). This holds for all $i \in C \cup E$ because PCDID applies a common normalization to all treated and control units.

We state the asymptotic result below. The assumptions, which are standard regularity conditions, are detailed in the Appendix for space reasons. Define $\tilde{G} = \{\tilde{G}_i\}_{i \in E} = \{1_{post,i}, F, X_i\}_{i \in E}$.

Theorem 4 (**Alpha test**). Suppose Assumptions E, F, FLM, AI $_\alpha$, IE and MM hold. Then, under Assumption PTW, we have the following results as $T, N_E, N_C \to \infty$ jointly and $\frac{T}{N_C} \to 0$:

(a) $\hat{a}^{mg} \xrightarrow{p} 1$.

(b) $\sqrt{N_E}\,\bar{\varphi}_{N_E,T}^{-1}(\hat{a}^{mg} - 1) \xrightarrow{d} N(0,1)$ if additionally Assumption FLM2 holds, where $\bar{\varphi}_{N_E,T}^2 := Var\left(\frac{1}{\sqrt{N_E}} \sum_{i \in E} h_{Ti}'(\mu_i - \mu_0) \,|\, \tilde{G}\right)$ and $h_{Ti} = \frac{F' M_{[1_{post,i}, X_i]} F \mu_0}{\mu_0' F' M_{[1_{post,i}, X_i]} F \mu_0}$.

The asymptotic normality result (and the $\sqrt{N_E}$ -rate) is robust to a wide range of factor dynamics, including nonstationary processes. As in PCDID-MG, we may form the *studentized* test statistic $\frac{\hat{a}^{mg} - 1}{\sqrt{Var(\hat{a}^{mg})}}$ using a suitable nonparametric variance estimator, e.g., $Var(\hat{a}^{mg}) = \frac{1}{N_E(N_E - 1)} \sum_{j \in E} (\hat{a}_j - \hat{a}^{mg})^2$ when the factor loadings μ_j are uncorrelated across $j \in E$.

A rejection in the Alpha test may imply that the interactive effects model is misspecified. In a fully nonparametric setting, Callaway and Sant'Anna (2018) show that the parallel trend assumption (PTA) in DID analysis is, in general, untestable. Importantly, they also show that a stronger, augmented version of this assumption is testable. In our notations, their *augmented PTA* assumes that the potential outcome $y_{it}(0)$ satisfies conditions akin to
$E(y_{it}(0) - y_{i,t-1}(0) \,|\, X, i \in E) = E(y_{it}(0) - y_{i,t-1}(0) \,|\, X, i \in C)$ *for each* $t = 2,..,T$,

whereas the PTA assumes this for each $t > T_0$ only. Our PTW is testable in the sense that under the factor structure, it can be viewed as a special case of the augmented PTA (see footnote 3 and Online Appendix).

4.5 Comparison of PCDID with Other Approaches

The PCDID method resembles SC, unconfoundedness and matrix completion approaches in that they allow for unparallel trends via factor models. PCDID uses all periods of data from treated units in a factor-augmented regression to estimate its target parameter (e.g., equation (8)). The factor proxies are constructed from data on control units ($i \in C$) only. The other approaches first estimate the missing potential outcome ($\hat{y}_{it}(0)$) of treated units after intervention ($i \in E$ and $t > T_{0i}$) using data from treated units before intervention ($i \in E$ and $t \leq T_{0i}$) and control units ($i \in C$); then, they compute $\hat{\delta}_{it} := y_{it}(1) - \hat{y}_{it}(0)$.

For illustration, we briefly compare the PCDID and GSC methods below (full details in Online Appendix). Xu (2017)'s GSC estimator is particularly relevant to ours, as it: (i) estimates factors and factor loadings using the control observations; (ii) uses these estimates and the pre-intervention periods of treated units to estimate the factor loadings of treated units, and (iii) uses all the above estimates to predict $\hat{y}_{it}(0)$ of treated units after intervention. Consider the DGP in equation (1). Without losing the main intuition, let our estimand be the ITET $\bar{\Delta}_i := E(\Delta_{it} \mid t > T_0)$ for a given $i \in E$. Suppose the DGP is given by (1), where f_t is a stationary, mean-zero factor satisfying $Var(f_t \mid t \leq T_0) = Var(f_t \mid t > T_0) > 0$, and $\tilde{\epsilon}_{it}$ are *iid* with mean zero and independent of f_t. Importantly, we allow f_t to be correlated with the intervention dummy $1_{\{t > T_0\}}$, so that $b_0 := E(f_t \mid t \leq T_0)$ and $b_1 := E(f_t \mid t > T_0)$ may be different. Let $\hat{\delta}_i^{gsc} := \frac{1}{T_1} \sum_{t > T_0} \hat{\delta}_{it}^{gsc}$ denote the GSC estimator.

It can be shown that GSC uses stronger assumption than PCDID on identification, and both approaches are numerically different. In large samples, PCDID is at least as efficient as GSC, and the former is more efficient when f_t is correlated with $1_{\{t > T_0\}}$. The last result is summarized below.

Proposition 1 (**Asymptotic efficiency of PCDID and GSC estimators**). Suppose $0 < \kappa_i < 1$ and the asymptotic identification condition of $\hat{\delta}_i^{gsc}$ holds (condition (4) in Online Appendix). Then, under the above DGP, $\hat{\delta}_i$ (defined in (10)) is asymptotically at least as efficient as $\hat{\delta}_i^{gsc}$. The two estimators are equally asymptotically efficient iff $b_0 = b_1 = 0$.

The intuition is that GSC discards data in the post-intervention subsample which contain useful information about the correlation between factors and intervention dummy (see (ii) above). This leads to efficiency loss, which becomes bigger as the proportion of post-intervention periods to be discarded is higher.

PCDID has an efficiency advantage relative to other approaches. However, its regression-based approach suggests that, unlike other approaches, assumptions on treatment effect *dynamics* are needed to ensure that the estimator converges to a well-defined causal parameter of interest. To understand this, recall that PCDID uses both pre- and post-intervention periods to carry out factor-augmented regressions. Misspecified TE dynamics may result in correlation between factor proxy $\hat{f}_t$ and regression error e_{it} in equation (8), leading to violation of the control function principle and incorrect counterfactuals.

We can modify the PCDID regressions to relax some restrictions on TE dynamics (full details in Online Appendix). Suppose the TE dynamics depends on time-varying covariates and/or factor structure (violation of **E(i), IE**), i.e., $\Delta_{it} = \breve{\beta}_i' x_{it} + \breve{\varsigma}_i + \breve{\mu}_i' f_t + \tilde{\Delta}_{it}$ where $\tilde{\Delta}_{it}$ is the idiosyncratic TE component. Substituting it into (5) yields a reduced-form model

$$y_{it} = \breve{\varsigma}_i 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + (\beta_i' + \breve{\beta}_i' 1_{\{i \in E\}} 1_{\{t > T_{0i}\}}) x_{it} + \varsigma_i + (\mu_i' + \breve{\mu}_i' 1_{\{i \in E\}} 1_{\{t > T_{0i}\}}) f_t + \epsilon_{it} \quad (16)$$

Note that y_{it}'s sensitivity to x_{it} and f_t changes after policy intervention. This motivates an extended form of the PCDID regression in (8) for $i \in E$ using data from $t = 1,..,T$:

$$y_{it} = b_{0i} + \delta_i 1_{\{t>T_{0i}\}} + a_{1i}' \hat{f}_t + a_{2i}' 1_{\{t>T_{0i}\}} \hat{f}_t + b_{1i}' x_{it} + b_{2i}' 1_{\{t>T_{0i}\}} x_{it} + e_{it}, \qquad (17)$$

where the extra regressors are $1_{\{t>T_{0i}\}} \hat{f}_t$ and $1_{\{t>T_{0i}\}} x_{it}$, ensuring that we have the correct control functions ($a_{1i}' \hat{f}_t$) to construct counterfactuals. The estimated ITET is $\hat{\delta}_i + \frac{1}{T-T_{0i}} \sum_{t=T_{0i}+1}^{T} (\hat{b}_{2i}' x_{it} + \hat{a}_{2i}' \hat{f}_t)$. This extended form is made more flexible at the expense of efficiency. In practice, we recommend (8) unless there are strong economic reasons to believe that the sensitivities change after policy intervention.

Suppose the TE exhibits time trends (violation of **E(i), AI(ii), IE**). We can estimate PCDID regressions with extra regressor(s) that control for such trends:

$$y_{it} = b_{0i} + \delta_i 1_{\{t>T_{0i}\}} + \breve{\delta}_i' z_{it} 1_{\{t>T_{0i}\}} + a_i' \hat{f}_t + b_{1i}' x_{it} + e_{it}, \qquad (18)$$

where z_{it} may consist of step and power functions of time (e.g., $1_{\{t>T_{0i}+j\}}$ for some $j \geq 1$; $(t-T_{0i})^m$ for some $m \geq 1$) and observed exogenous covariates. This approach can incorporate the TE dynamics as long as the specification is known to the modeller.

Overall, we recommend using PCDID when the TE dynamics are relatively simple. If researchers suspect that the TE dynamics are complicated, they should use other approaches that do not place restrictions on TE dynamics.

5 Small Sample Properties of Estimators via Simulations

5.1 Baseline specifications and results

All DGPs follow the general form $y_{it} = \Delta_{it} 1_{\{i \in E\}} 1_{\{t>T_{0i}\}} + \beta_i' x_{it} + \varsigma_i + \mu_i' f_t + \tilde{\epsilon}_{it}$. We first discuss the DGPs for ITET estimation, and then the DGPs for ATET estimation. The DGPs for ITET estimation set $N_E = 1$, $T_{0i} = \frac{T}{2}, \Delta_{it} = 3, \varsigma_i = 0$ and $x_{it} = 0$ (relaxed below). The idiosyncratic error exhibits serial correlation as well as heteroscedasticity across i; specifically, $\tilde{\epsilon}_{it} = \rho_\epsilon \tilde{\epsilon}_{i,t-1} + v_{it} h_i$ where

$\rho_\epsilon = 0.1$, $\nu_{it} \sim N(0, 0.01(1-\rho_\epsilon^2))$ is iid across *i* and *t*, and $h_i \sim unif(0.5, 1.5)$ is iid across *i*. Three interactive effect scenarios are considered:[11]

(A) Stationary factors: For each of the three factors *j* = 1, 2, 3, $f_{jt} = \phi_j + \eta_j 1_{\{t>\frac{T}{2}\}} + \rho_j f_{jt-1} + u_{jt}$ with i.i.d. $u_{jt} \sim N(0, \sigma_{uj}^2)$, $\phi_1 = \phi_2 = \phi_3 = 0$, $\eta_1 = \eta_2 = \eta_3 = 0$, $\rho_1 = 0.5$, $\rho_2 = 0.7$, $\rho_3 = 0.9$, $\sigma_{u1}^2 = 0.0675$, $\sigma_{u2}^2 = 0.0459$ and $\sigma_{u3}^2 = 0.0171$. Hence all three factors are AR(1) and have variance 0.09.

(B) Stationary factors with break: same as A except that $\eta_2 = 1.2$. Hence factor 2 is AR(1) except for a jump at $t = \frac{T}{2} + 1$. Factor 2 is positively correlated with $1_{\{t>T_{0i}\}}$ but not collinear because it is time-varying for all *t* (**Assumption AI(i)** is satisfied).

(C) Nonstationary factors with drift: same as A except that $\phi_1 = 0.1$, $\rho_1 = \rho_2 = \rho_3 = 1$, $\sigma_{u1} = 0.3$, $\sigma_{u2} = 0.5$, and $\sigma_{u3} = 0.1$. Hence all three factors are I(1), and factor 1 drifts upward by 0.1 per period.

In all scenarios, the factor loading of unit *i* for factor *j* is distributed as $\mu_{ij} \sim N(m_{Cj}, \sigma_{\mu_C}^2)$ if $i \in C$, and $\mu_{ij} \sim N(m_{Ej}, \sigma_{\mu_E}^2)$ if $i \in E$. We set $(m_{C1}, m_{C2}, m_{C3}) = (1, 0.9, 0.8)$, $(m_{E1}, m_{E2}, m_{E3}) = (1.2, 1.4, 1.6)$, $\sigma_{\mu_C} = \sigma_{\mu_E} = 0.3$ so that the average factor loadings are higher among treated units than control units (μ_i is positively correlated with $1_{\{i \in E\}}$). Yet in Scenario A, $\mu_i' f_t$ remains uncorrelated with $1_{\{i \in E\}} 1_{\{t>T_{0i}\}}$ because f_t is uncorrelated with $1_{\{t>T_{0i}\}}$. In scenario B, $\mu_i' f_t$ is positively correlated with $1_{\{i \in E\}} 1_{\{t>T_{0i}\}}$ due to factor 2. In scenario C, $\mu_i' f_t$ tends to be larger among treated units and in later time periods (due to factor 1), but its correlation with $1_{\{i \in E\}} 1_{\{t>T_{0i}\}}$ is undefined due to unit roots in f_t.

We compare seven estimators: (i) PCDID, (ii) DID with two-way FE ("DID-2wfe"), (iii) DID with FE and unit-specific cubic time trend ("DID-trend"; e.g., Wooldridge (2005)), (iv) Bai (2009), (v) GSC (Xu (2017)), (vi) stepwise GM

(Gobillon and Magnac (2016)), (vii) nuclear norm matrix completion ("MC-NNM"; Athey et al. (2018)). Details of these estimators are in Online Appendix.[12] PCDID, BAI, GSC and GM assume there are three factors, whereas MC-NNM computes the rank (or number of factors) automatically. We consider 10 different (N, T) combinations, and compute the bias and standard deviation (SD) of each estimator based on 1000 replications.

Table I reports the ITET estimation results. Overall, PCDID yields the best performance in terms of bias and SD. Other methods have relatively uneven performance across scenarios and sample size. DID-2wfe performs worst, followed by DID-trend. Even in Scenario A where DID-2wfe exhibits no bias, it is still less efficient than PCDID as it does not control for the factor structure. MC-NNM outperforms DID methods in scenario A, which is consistent with Athey et al. (2018). It is biased heavily upward in B and C, even though the estimated rank is close to the true number of factors when sample size is large. Both GSC and GM have a high SD when T is small. When T is large, their performance are similar to PCDID in A, but still worse in B and C. Compared to PCDID, BAI has worse bias overall but slightly better SD in B. The average number of iterations required for convergence is nontrivial, ranging from 8 to 120 depending on scenario and sample size, and non-convergence is quite common.

In the DGPs for ATET estimation, we set $N_E = N_C = \frac{N}{2}$. In addition, we set $\Delta_{it} = 3 + \tilde{\Delta}_i + 0.25(\mu_{i1} - 1.2) + \rho_\Delta \Delta_{i,t-1} + u_{\Delta,it}$, where $\tilde{\Delta}_i \sim N(0,1)$, $\rho_\Delta = 0.1$ and $u_{\Delta,it} \sim N(0, 0.01(1-\rho_\Delta^2))$. Therefore, the treatment effect varies cross-sectionally and over time, and it is correlated with the unobserved loading of the first factor μ_{i1}. We compare the simple mean-group estimator (PCDID-MG) with the other estimators. Table II reports the ATET estimation results when the policy intervention date is homogeneous ($T_{0i} = \frac{T}{2}$). The relative advantage of PCDID is larger compared to ITET estimation – PCDID-MG outperforms the other estimators in all scenarios and sample sizes. The next-best performers are GSC and GM, both having a large SD when T is small. Somewhat surprisingly, numerical convergence in BAI becomes strenuous in all scenarios. DID-2wfe, DID-trend and MC-NNM are satisfactory in scenario A only.

Additional results are reported in Online Appendix: (1) ATET results under staggered adoption, e.g., units with higher TE are treated earlier; (2) Performance of PCDID when the number of factor proxies is chosen by original and conservative implementations of the GR test; (3) Performance of PCDID when there are exogenous covariates or covariates endogenous to the factor structure. The results remain qualitatively similar and lend support to the robustness of our approach.

5.2 Inference on ITET/ATET and parallel trend test

We examine the finite-sample performance of PCDID inference procedures for ITET and ATET. The DGPs are the same as in Table I and II, respectively (except we set $\rho_\epsilon = 0$ in the DGP for ITET inference). The null hypothesis is set at the DGPs' true value ($\delta_0 = 3$). We use 1000 replications and a nominal size of 5%. We compare four procedures. The first two are based on the full sample only, while the other two involve 199 bootstrap samples per replication: (1) **TrueF**: assume factors are observed in PCDID estimation (infeasible) and compute t-statistic, reject if $|t| \geq 1.96$. (2) **Asym**: use factor proxies in PCDID estimation and compute t-statistic, reject if $|t| \geq 1.96$. (3) **b-t**: same as Asym, but reject if $t \leq c_{0.025}$ or $t \geq c_{0.975}$ where $c_{0.025}, c_{0.975}$ are percentiles of the bootstrap distribution of t-statistics. (4) **b-se**: same as Asym, but the standard error in the t-statistic formula is obtained from bootstrap samples.

Comparing Asym with TrueF will give the relative performance when factors are estimated instead of known (e.g., Gonçalves and Perron (2014)). Because Asym is asymptotically valid (see Theorems 1,2,3), comparing b-t/b-se with Asym will show whether bootstrapping yields reasonable results relative to this baseline approach. For ITET inference, a mix of wild and stationary bootstrap (on residuals) is used; for ATET inference, wild bootstrap is used. See Online Appendix for details on bootstrap sample construction. In TrueF, Asym and b-t, we compute t-statistics based on analytical standard errors. For ITET inference, it is $\frac{\hat{\delta}_i - \delta_0}{se(\hat{\delta}_i)}$ where $se(\hat{\delta}_i)$ is obtained from the classical standard error formula based on the time-series regression for unit *i*.[13] For ATET inference, it is $\frac{\hat{\delta}^{mg} - \delta_0}{se(\hat{\delta}^{mg})}$ with the nonparametric estimator $se(\hat{\delta}^{mg}) = \sqrt{\frac{1}{N_E(N_E-1)}\sum_{i\in E}(\hat{\delta}_i - \hat{\delta}^{mg})^2}$ (see discussion after Theorem 3).

Table III reports the simulation results. In ITET inference (left panel), TrueF has a rejection rate close to the nominal size of 5%, as expected. Asym's rejection rate is similar to TrueF's, especially when the sample is large; however, it tends to overreject when $N << T$. B-t performs similarly to Asym, and has better performance when $N << T$. B-se has a lower rejection rate (i.e., more conservative) compared to b-t. In ATET inference (right panel), the inference procedures perform better. TrueF, Asym, b-t and b-se all yield similar results in general, with rejection rates close to 5% when the sample is large, and Asym and b-t having less severe overrejection when $N << T$.

In Online Appendix, we examine the rejection rates, power and rotational invariance of the parallel trend alpha test, showing that its theoretical properties are confirmed by simulations.

6 Effect of Waiver Programs on Welfare Caseloads

DID regressions have been widely used for examining the effects of various welfare reforms on labor market outcomes (see e.g., Chan and Moffitt (2018) for a recent survey). In this illustration, we examine the effects of waiver programs on caseloads in the Aid to Families with Dependent Children (AFDC) program. Providing cash assistance to low-income female-headed families, AFDC had been one of the largest means-tested transfer programs in the US. In the 1990s, many states sought waivers from the federal government, which allowed them to deviate from federal AFDC rules. The majority of the states had a waiver in place when AFDC was replaced by the Temporary Assistance for Needy Families (TANF) program in 1997.

Our analysis builds on Ziliak et al. (2000), who found mixed evidence regarding the effects of waivers. Our focus is to apply PCDID, GSC and DID regressions (taking into account of clustering) and compare the results. Our data set covers 50 states plus the District of Columbia for 117 months from Oct1986 to Jun1996. Because PCDID is robust to the presence of nonstationarity trends, we do not first-difference the data in baseline specifications (see Online Appendix for a full analysis of first-differenced data as well as robustness checks). As in Ziliak et al. (2000), we use the waiver approval date to define policy intervention. Specifically, we define T_{0i} as the approval date of state *i*'s work requirement waiver or time limit waiver, whichever is earlier.[14] The policy intervention exhibits staggered adoption with T_{0i} as early as mid-1992 (see Figure A2). There are 20 control states, which have neither of the waivers by the end of the sample period.

We perform PCDID estimation separately in four different samples, all of which have T = 117 periods: (1) Control plus all treated states $(N_C, N_E) = (20, 31)$; (2) Control plus 10 Southern treated states $(N_C, N_E) = (20, 10)$; (3) Control plus 21 non-Southern treated states $(N_C, N_E) = (20, 21)$; (4) Control plus Wyoming $(N_C, N_E) = (20, 1)$. Summary statistics indicate that the average characteristics of treated and control states are generally similar, while individual states, e.g., Wyoming, can be quite different from the average (see Table A7). The Southern states are defined as states in the South Census region (18 in total); we will examine whether the policy had larger effects in the South, where the welfare system was relatively stringent.

We apply PCDID to the model $y_{it} = \Delta_{it} 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + \beta_i' x_{it} + \varsigma_i + \mu_i' f_t + \tilde{\epsilon}_{it}$, where y_{it} is the log of per-capita welfare caseload in state *i* at time *t*, Δ_{it} is the treatment effect, T_{0i} is the waiver approval date, x_{it} is a vector of six time-varying covariates including the maximum combined real AFDC/Food Stamp benefits for a family of three, state unemployment rate, state log employment-to-population ratio, and calendar quarter dummies; ς_i is the state fixed effect, $\mu_i' f_t$ is the interactive effects structure, and $\tilde{\epsilon}_{it}$ is the idiosyncratic error. We use PCDID-MG ($\hat{\delta}^{mg} := \frac{1}{N_E}\sum_{i \in E}\hat{\delta}_i$) for ATET ($\bar{\bar{\Delta}} := E(\bar{\Delta}_i \mid i \in E)$) estimation in samples 1,2,3, and PCDID ($\hat{\delta}_i$) for ITET ($\bar{\Delta}_i := E(\Delta_{it} \mid t > T_{0i})$) estimation in sample 4.[15] Our preferred inference criteria are analytical standard errors (Asym) and p-values from bootstrapped t-statistics (b-t); for completeness we also report bootstrapped standard errors (b-se), which tend to be more conservative (see Section 5).

In sample 4 we have $\frac{\sqrt{T}}{N_c} \approx \frac{10.82}{20} = 0.54$, and in samples 1-3 we have $\frac{T}{N_c} = \frac{117}{20} = 5.85$. How do they reconcile with the asymptotic ratio conditions in Theorem 1 and 2, respectively? Because we only have a single finite sample, there is no way to judge whether it is a good approximation to those conditions. Nonetheless, the simulations in Section 5.1 suggest that the asymptotic bias of our estimators is modest even when T is 20 times as large as N_c (as shown in Panel B in Table I and II, the worst bias is around 2% of the true value in DGP; the bias of our estimators is also smaller relative to other methods). In practice, the presence of serial correlation can also reduce the effective sample size in T, particularly when persistent and stochastic trends are present.

Panel A of Table IV reports the baseline results, which are based on 4 PCs (see Figure A3 for PC plots). In sample 1 (all treated states), the policy intervention coefficient implies that waivers reduced per-capita welfare caseload by 1.7% on average. The effects are larger among Southern treated states (sample 2, -2.4%) than other treated states (sample 3, -1.3%). In Wyoming (sample 4), waivers reduced the per-capita caseload by a larger magnitude of 11.4% on average. In samples 1,2,4, the policy coefficients are statistically significant at either 1% or 5% level by the Asym and b-t criteria.

We use the estimates to compute the fraction of decline in per-capita caseload in treated states between Jan93 (t=76) and Jun96 (t=117) that can be explained by the policy intervention. Specifically, we compute $1-\frac{\bar{y}_{E,117}^{cf}-\bar{y}_{E,76}^{cf}}{\bar{y}_{E,117}^{pred}-\bar{y}_{E,76}^{pred}}$ where $\bar{y}_{E,t}^{cf}$ denotes the counterfactual average per-capita caseload in treated states at period t assuming no reform at all (i.e., set $T_{0i}=\infty \ \forall i$) and $\bar{y}_{E,t}^{pred}$ denotes the predicted/fitted caseload when T_{0i} is the same as in the data. Among all treated states, the waivers explained 6.88% of the drop in caseloads between Jan93 and Jun96. This proportion is 10.41% among southern treated states and 24.86% in Wyoming.

To provide visual evidence, Figure 1 plots the actual, predicted and counterfactual caseloads over time. While the figure compares average caseloads in control and treated states, note that the trends in individual states are far more heterogeneous. Overall, both control and treated caseloads rose sharply between 1990-93 and then dropped sharply afterwards. The caseload trends, which are stochastic according to visual inspection and unit root tests, exhibit subtle differences. The trajectories of control and treated caseloads tend to narrow before 1993 (pre-intervention) and then diverge afterwards (see also Figure A4). In addition, caseloads peaked earlier among treated states. These potentially unparallel and stochastic trends motivate the use of our method. For example, in Wyoming (sample 4) we see an abrupt drop in actual treated caseload at the policy intervention date, and the model predicts a counterfactual path that is higher but still diverges from control caseloads. As a robustness check, we also use the Alpha test to formally examine whether PTW is violated (bottom of table). While the test fails to reject the null of PTW, note that under stochastic trends, even the slightest violation of PTW can result in a divergent nuisance term that is detrimental to DID regressions (Section 2). On this basis we still recommend using PCDID to control for potentially unparallel and stochastic trends.

Panel B reports results from extra specifications using sample 1. We first examine sensitivity to the number of factors. The recursive version of the GR test recommends 3 factors. We find that the coefficients tend to stabilize when 3 or more PCs are used. To examine whether there are substantial TE dynamics, we add a stepwise covariate, e.g., $\mathbf{1}\{t > T_{0i} + 3\}$ to the PCDID regression, and find that its coefficient is close to zero. Next, we consider a GSC estimator that uses pre-intervention data from treated units to predict counterfactual post-intervention outcomes. We align its estimation/inference procedure and target parameter (ATET) to facilitate comparison with PCDID. In GSC estimation, we have to drop the welfare benefit covariate as the estimates are unstable otherwise. The ATET estimate is -0.033, which is almost double to that in PCDID. To explain this discrepancy, we compare how the average caseloads in treated states are driven by the estimated factor structure ($\mu_i^{'} \hat{f}_t$) in PCDID and GSC models. We find that PC2 drives a much slower caseload decline between 1993-96 in the GSC model relative to the PCDID model (Figure 2). This results in a smoother decline in the counterfactual path and magnifies both the TE dynamics and ATET estimate in the GSC model.

Nevertheless, both PCDID and GSC predict a limited role of waivers in explaining the caseload decline between 1993-96 (6.88% and 14.05%, respectively). The PCDID estimate is more conservative and is likely to be more credible for the following reason. The reforms were implemented when both control and treated caseloads started to reverse trends, i.e., the unobserved factors are temporally correlated with policy intervention. This is a scenario where PCDID exhibits a clear advantage over GSC in terms of finite sample bias (e.g., “Scenario B” in Section 5) and asymptotic efficiency (Proposition 1). In that scenario, it is crucial to estimate the factor structure accurately, for which PCDID is advantageous because it uses the entire sample period for estimation. Indeed, in our data, GSC seems to have underestimated the sensitivity of caseloads in treated states to PC2 (Figures 2 and A3). More generally, we recommend using PCDID when the trends are potentially correlated with policy interventions.

In Panel C, we use sample 1 to estimate: (1) DID regression with state-specific time trends $y_{it} = \delta 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + b_1^{'} x_{it} + b_{0i} + \sum_{m=1}^{M} a_{im} t^m + e_{it}$ where M = 3 (cubic) or M = 4 (quartic) (e.g., Wooldridge (2005)); (2) DID-2wfe regression $y_{it} = \delta 1_{\{i \in E\}} 1_{\{t > T_{0i}\}} + b_1^{'} x_{it} + b_{0i} + \tau_t + e_{it}$. Cluster wild bootstrapped standard errors are reported (see Online Appendix). In (1) where we impose restrictive trend specifications relative to PCDID, the policy effect is close to zero and statistically insignificant. In (2), the policy coefficient of -0.054 implies that waivers explained 24% of caseload decline in 1993-96. Interestingly, this estimate is similar to the 1-PC PCDID specification (-0.056, not shown). Overall, the DID results appear to be sensitive to the specification.

7 Conclusions

In this paper, we developed a class of factor-augmented regression estimators (PCDID) for treatment effect estimation. PCDID was similar in spirit to the control function approach. It used factor proxies constructed from control units to control for unobserved trends, assuming that the unobservables followed an interactive effects structure. The estimation and inference procedures were relatively straightforward. After defining the key causal parameters of interest, ITET and ATET, we showed that the basic PCDID estimator targeted the ITET, whereas the simple mean-group (PCDID-MG) and pooled estimators targeted the ATET. We showed consistency and asymptotic normality of these estimators under minimal assumptions on the trend specification. We provided inference procedures based on the asymptotic normality results. We developed a parallel trend Alpha test (assuming an interactive effects structure) as a byproduct of the PCDID approach.

In Monte Carlo simulations, which focused on scenarios with unparallel and/or nonstationary trends, we illustrated that PCDID had an advantage in terms of bias and efficiency. In the application, we found that waiver programs (work requirements and time limits) reduced welfare caseload per capita by an average of 1.7% among all treated states (or a 6.88% of the caseload decline in 93Q1-96Q2) based on the PCDID-MG estimator. While the GSC estimator yielded larger estimates (3.3%), PCDID was more credible because the programs were implemented when caseloads (in control and treated states) started to peak and trend downward – a scenario where PCDID could estimate the factor structure more accurately. The results from classical DID approaches were not robust, due to the fact that the caseloads exhibited potentially unparallel and stochastic trends.

References

ABADIE, A., A. DIAMOND, AND J. HAINMUELLER (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505.

ABRAHAM, S. AND L. SUN (2018): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Working Paper*.

AHN, S. AND A. HORENSTEIN (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227.

AHN, S., Y. LEE, AND P. SCHMIDT (2001): "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects," *Journal of Econometrics*, 101, 219–255.

——— (2013): "Panel Data Models with Multiple Time-Varying Individual Effects," *Journal of Econometrics*, 174, 1–14.

ARKHANGELSKY, D., S. ATHEY, D. HIRSHBERG, G. IMBENS, AND S. WAGER (2019): " Synthetic Difference in Differences," *Working Paper*.

ATHEY, S., M. BAYATI, N. DOUDCHENKO, G. IMBENS, AND K. KHOSRAVI (2018): " Matrix Completion Methods for Causal Panel Data Models," *Working Paper*.

ATHEY, S. AND G. IMBENS (2018): "Design-Based Analysis in Difference-in-Differences Settings with Staggered Adoption," *Working Paper*.

BAI, J. (2004): "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics*, 122, 137–183.

——— (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

BAI, J. AND S. NG (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, 1133–1150.

BEN-MICHAEL, E., A. FELLER, AND J. ROTHSTEIN (2018): "The Augmented Synthetic Control Method," *Working Paper*.

BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387–422.

BORUSYAK, K. AND X. JARAVEL (2017): "Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume," *Working Paper*.

CALLAWAY, B. AND P. SANT'ANNA (2018): "Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment," Working Paper.

CHAN, M. K. (2013): "A Dynamic Model of Welfare Reform," *Econometrica*, 81, 941–1001.

CHAN, M. K. AND R. A. MOFFITT (2018): "Welfare Reform and the Labor Market," *Annual Review of Economics*, 10, 347–381.

CHEN, L., J. DOLADO, AND J. GONZALO (2014): "Detecting Big Structural Breaks in Large Factor Models," *Journal of Econometrics*, 180, 30–48.

CHERNOZHUKOV, V., I. FERNANDEZ-VAL, J. HAN, AND W. NEWEY (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580.

CHERNOZHUKOV, V., K. WÜTHRICH, AND Y. ZHU (2018): "Inference on Average Treatment Effects in Aggregate Panel Data Settings," *Working Paper*.

DE CHAISEMARTIN, C. AND X. D'HAULTFŒUILLE (2018): "Two-way fixed effects estimators with heterogeneous treatment effects," *Working Paper*.

DEJONG, R. AND J. DAVIDSON (2000): "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes," *Econometric Theory*, 16, 621–642.

ERGEMEN, Y. (2019): "System Estimation of Panel Data Models Under Long-Range Dependence," *Journal of Business & Economic Statistics*, 37, 13–26.

ERGEMEN, Y. AND C. VELASCO (2017): "Estimation of Fractionally Integrated Panels with Fixed Effects and Cross-Section Dependence," *Journal of Econometrics*, 196, 248–258.

FERMAN, B. AND C. PINTO (2018): "Synthetic Controls with Imperfect Pre-Treatment Fit," *Working Paper*.

GOBILLON, L. AND T. MAGNAC (2016): "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *The Review of Economics and Statistics*, 98, 535–551.

GONÇALVES, S. AND B. PERRON (2014): "Bootstrapping Factor-Augmented Regression Models," *Journal of Econometrics*, 182, 156–173.

GOODMAN-BACON, A. (2018): "Difference-in-Differences with Variation in Treatment Timing," Working Paper.

GREENAWAY-MCGREVY, R., C. HAN, AND D. SUL (2012): "Asymptotic Distribution of Factor Augmented Estimators for Panel Regression," *Journal of Econometrics*, 169, 48–53.

HOLTZ-EAKIN, D., W. NEWEY, AND H. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–1395.

HSIAO, C., H. S. CHING, AND S. K. WAN (2012): "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," *Journal of Applied Econometrics*, 27, 705–740.

KAPETANIOS, G., M. PESARAN, AND T. YAMAGATA (2011): "Panels with non-stationary multifactor error structures," *Journal of Econometrics*, 160, 326–348.

KIM, D. AND T. OKA (2014): "Divorce Law Reforms and Divorce Rates in the USA: An Interactive Fixed-Effects Approach," *Journal of Applied Econometrics*, 29, 231–245.

MOON, H. R. AND M. WEIDNER (2015): "Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543–1579.

PARK, J. AND P. PHILLIPS (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468–497.

PESARAN, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

PHILLIPS, P. AND H. MOON (1999): "Linear Regression Limit Theorems for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

STOCK, J. AND M. WATSON (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

WESTERLUND, J. AND J. URBAIN (2015): "Cross-Sectional Averages Versus Principal Components," *Journal of Econometrics*, 185, 372–377.

WOOLDRIDGE, J. (2005): "Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models," *The Review of Economics and Statistics*, 87, 385–390.

XU, Y. (2017): "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25, 57–76.

ZILIAK, J. P., D. N. FIGLIO, E. E. DAVIS, AND L. S. CONNOLLY (2000): "Accounting for the Decline in AFDC Caseloads: Welfare Reform or the Economy?" *Journal of Human Resources*, 35, 570–586.

Appendix: Some Key Assumptions

This section contains some key assumptions for the theoretical results. The full list of assumptions can be found in the Online Appendix. Define $\tilde{G} := \{\tilde{G}_i\}_{i\in E} = \{1_{post,i}, F, X_i\}_{i\in E}$.

Assumption FL (factor loadings): Assumption FL$_C$ holds. In addition,

(i) For all $i \in E$, $E\|\mu_i\|^2 \leq c$ for some constant $c > 0$.

(ii) $\text{plim}_{N_E \to \infty} \frac{1}{N_E} \mu_E^{'} \mu_E$ is positive definite.

Assumption AI$_{mg}$ (asymptotic identification, simple mean-group estimator):

(i) $\text{plim}_{N_E, T \to \infty} \inf_{i \in E} \frac{1}{T} 1_{post,i}^{'} M_{G_i} 1_{post,i}$ exists and is strictly positive $a.s.$.

(ii) $\zeta^2 := \text{plim}_{N_E, T \to \infty} E\left(\left\| \frac{1}{\sqrt{N_E T}} \sum_{i \in E} 1_{post,i}^{'} M_{G_i} \epsilon_i \right\|^2 \middle| \tilde{G} \right)$ exists and is strictly positive $a.s.$.

Assumption RT$_{mg}$ (treatment effects, simple mean-group estimator): Let $\upsilon_i := \bar{\Delta}_i - \bar{\bar{\Delta}}$. The following conditions are satisfied:

(i) For some $p > 2$, there exists $0 < c < \infty$ such that $E|\upsilon_i|^p \leq c$ for all $i \in E$.

(ii) υ_i is a mixing process with mixing coefficient ϕ of size $-p/2(p-1)$ for $p \geq 2$, or α of size $-p/(p-2)$, $p > 2$.

(iii) $\lim_{N_E \to \infty} Var(N_E^{-1/2} \sum_{i \in E} \upsilon_i)$ exists and is strictly positive.

Assumption AI$_\alpha$ (asymptotic identification, Alpha test): Let r be the normalization order of $F\mu_0$ such that $\|F\mu_0\|^2 / T^{2r} = O_p(1)$ as $T \to \infty$. The following conditions hold:

(i) $\text{plim}_{N_E, T \to \infty} \inf_{i \in E} \frac{1}{T^{2r}} \mu_0^{'} F^{'} M_{[1_{post,i}, X_i]} F \mu_0$ exists and is strictly positive $a.s.$.

(ii) $\text{plim}_{N_E, T \to \infty} E\left(\left\| \frac{1}{\sqrt{N_E} T^r} \sum_{i \in E} \mu_0^{'} F^{'} M_{[1_{post,i}, X_i]} F(\mu_i - \mu_0) \right\|^2 \middle| \tilde{G} \right)$ exists and is strictly positive $a.s.$.

Assumption FLM (mixing factor loadings, Alpha test): The following conditions are satisfied:

(i) For some $p > 1$, there exists $0 < c < \infty$ such that $E(\|\mu_i\|^p) \leq c$ for all $i \in C \cup E$.

(ii) $\{\mu_i : i \in C\}$ and $\{\mu_i : i \in E\}$ are mixing sequences with mixing coefficients ϕ of size $-p/(2p-1)$ for $p \geq 1$, or α of size $-p/(p-1)$ for $p > 1$.

Assumption FLM2 (conditional mixing factor loadings, treated units, Alpha test): The following conditions are satisfied:

(ii) For each $i \in E$, $E(\mu_i \mid \tilde{G}) = \mu_0 \quad a.s.$.

(ii) For some $p > 2$, there exists $0 < c < \infty$ such that $E(\|\mu_i\|^p \mid \tilde{G}) \leq c$ for all $i \in E$.

(iii) Conditional on $\tilde{G}, \{\mu_i : i \in E\}$ is a mixing sequence with mixing coefficients ϕ of size $-p/2(p-1)$ for $p \geq 2$, or α of size $-p/(p-2)$ for $p > 2$.

Notes

1 Similar to PCDID estimators, the CCE estimators are robust to different factor dynamics including stationary (Pesaran (2006)) and unit-root processes (Kapetanios et al. (2011)). Weighted cross-sectional averages are allowed for CCE estimators; see Westerlund and Urbain (2015) and Greenaway-McGrevy et al. (2012).

2 Quasi-differencing methods have been developed for fixed-T models assuming that the factors are unknown (e.g., Holtz-Eakin et al. (1988), Ahn et al. (2001), Ahn et al. (2013)). These methods require stronger assumptions on idiosyncratic errors (e.g., iid) or require exclusion restrictions.

3 In a nonparametric framework, parallel trend holds when, given each $t > T_0$, the potential outcome $y_{it}(0)$ satisfies conditions akin to $E(y_{it}(0) - y_{i,t-1}(0) \mid i \in E) = E(y_{it}(0) - y_{i,t-1}(0) \mid i \in C)$; see Callaway and Sant'

Anna (2018) for details. Given t and $t-1$, the factor structure implies $E(y_{it}(0)-y_{i,t-1}(0)\,|\,i\in E)=E(\mu_i'\,|\,i\in E)(f_t-f_{t-1})$ and $E(y_{it}(0)-y_{i,t-1}(0)\,|\,i\in C)=E(\mu_i'\,|\,i\in C)(f_t-f_{t-1})$, hence the interest in PTW.

[4] We do not impose the time-homogeneity condition on period-specific disturbances ($\mu_i f_t+\epsilon_{it}$ in our context) as in the semi-/non-parametric panel data models of Chernozhukov et al. (2013). This also highlights the importance of imposing a (factor) structure on the disturbances. Nonetheless, the unobserved μ_i and f_t involve minimal assumptions; see **Assumptions F** and **FL$_C$**.

[5] See also Wooldridge (2005) and Pesaran (2006), where x_{it} is correlated with other regressors as well as unit-specific β_i. While Wooldridge shows that an FE estimator may consistently estimate the population average of β_i, he does not consider a factor structure, a key feature in Pesaran (2006) and our model. As in the 2wfe literature, time-invariant covariates are subsumed into fixed effect ς_i.

[6] Moon and Weidner (2015) establishes consistency of the least squares estimator of the regression coefficient when the number of factors used in estimation is at least ℓ.

[7] The asymptotic normality result hinges on the pivotal nature of the studentized statistic and a standard conditioning argument that applies to the case with nonstationary regressors (Park and Phillips (1988)). This is a non-trivial result; e.g., Chernozhukov et al. (2018) requires stationarity as the key assumption underlying their inference procedure.

[8] Although we are unable to provide formal proofs, the asymptotic normality result provides support for bootstrapping the t-statistic. By contrast, the limiting distribution of the non-studentized statistic may vary discontinuously with the serial dependence properties of the factors (e.g., stationary, near unit-root and unit-root processes). Hence we do not recommend bootstrapping the PCDID estimate directly.

[9] Although we only focus on the ATET, other estimands are available, such as the conditional moments and quantiles of ITET among $i \in E$. These estimands can help unveil the distributional features of ITET.

[10] Suppose α = 1 and $E(v_j \mid j \in E) = 0$. It follows that $E(\mu_j' \mid j \in E) f_t = E(\mu_i' \mid i \in C) f_t$ for all t, or in vector form: $F[E(\mu_j \mid j \in E) - E(\mu_i \mid i \in C)] = 0$ (see Online Appendix). That F has full column rank ℓ (**Assumption F(ii)**) implies $E(\mu_j \mid j \in E) = E(\mu_i \mid i \in C) = \mu_0$ where $\mu_0 \neq 0$, and hence PTW is satisfied.

[11] Existing studies have typically considered factors in simpler forms, e.g., iid factor or deterministic sinusoid functions (Bai (2009), Gobillon and Magnac (2016)). For space reasons, we do not report scenarios that involve deterministic trends; such cases are trivial and are encompassed by PCDID as a special case. We also do not report scenarios that involve a mix of nonstationary and stationary factors because the results look similar to the scenario with nonstationary factors only.

[12] Doubly weighted methods such as Arkhangelsky et al. (2019) numerically solves for vertical and horizontal weights with non-negativity constraints. Hence it can be viewed as an extension of Abadie et al. (2010) where vertical nonnegative weights are involved. Nonnegative weights require strong support conditions as derived in Gobillon and Magnac (2016).

[13] As mentioned above, the DGP for ITET inference sets $\rho_\epsilon = 0$; see Gonçalves and Perron (2014) for a similar setup. The formula can be replaced by the Newey-West HAC estimator if $\rho_\epsilon \neq 0$.

[14] We follow the definitions of variables in Ziliak et al. (2000). Work requirements and time limits are the key components of welfare reform and they unambiguously reduce welfare participation. For simplicity, we do not consider earnings disregards and parental responsibility waivers, which yield mixed

evidence in the literature, e.g. see Chan (2013) and Chan and Moffitt (2018) for details.

[15] We include some covariates that are potentially endogenous to the factor structure, e.g., unemployment rate. PCDID is robust to such covariates (see Section 4.1 and Online Appendix). We perform fixed effect estimation of y_{it} on x_{it} in the control panel (2,340 observations) to obtain residuals, which are then used for constructing factor proxies $\hat{f}_t$ by PCA.

TABLE I: SMALL SAMPLE PROPERTIES OF ESTIMATORS, ITET ESTIMATION[a]

		PCDID		DID-2wfe		DID-trend		BAI				GSC		GM		MC-NNM		
N_E+N_C	T	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Avg #Iter		Bias	SD	Bias	SD	Bias	SD	Avg Rank
Panel A: Stationary factors																		
51	10	0.00	0.15	0.00	0.47	-0.01	0.33	-0.02	0.19	*23*		0.01	0.59	0.01	0.57	0.01	0.33	*1.7*
51	20	0.00	0.09	-0.01	0.46	-0.02	0.31	-0.01	0.07	*13*	b	0.00	0.14	0.00	0.10	-0.01	0.26	*1.5*
51	50	0.00	0.05	-0.02	0.41	0.00	0.36	-0.02	0.19	*11*		0.00	0.06	0.00	0.04	-0.01	0.13	*3.0*
6	100	0.00	0.09	0.00	0.27	0.02	0.37	-0.10	0.24	*40*		0.00	0.09	0.00	0.08	0.01	0.18	*1.1*
11	100	0.00	0.05	0.00	0.28	0.03	0.37	-0.05	0.22	*26*		0.00	0.05	0.00	0.04	0.00	0.16	*1.3*
26	100	0.00	0.03	0.00	0.27	0.02	0.36	-0.03	0.23	*13*		0.00	0.03	0.00	0.02	0.00	0.14	*1.7*
6	10	0.01	0.24	0.02	0.48	0.01	0.34	-0.17	0.39	*81*		0.01	0.55	0.01	0.55	0.01	0.34	*2.2*
11	20	0.00	0.13	-0.02	0.47	0.00	0.33	-0.07	0.29	*42*		0.00	0.17	0.00	0.16	-0.01	0.28	*1.5*
26	50	-0.01	0.05	-0.02	0.42	-0.01	0.37	-0.03	0.23	*15*		-0.01	0.06	0.00	0.04	0.00	0.15	*2.6*
51	100	0.00	0.03	0.00	0.27	0.02	0.34	-0.01	0.18	*8*		0.00	0.03	0.00	0.02	0.00	0.11	*2.8*
Panel B: Stationary factors with break																		
51	10	-0.01	0.23	1.27	0.87	0.43	0.42	-0.02	0.20	*49*	b	-0.01	2.25	-0.01	2.25	2.00	1.19	*1.9*
51	20	0.00	0.16	1.56	1.08	0.59	0.49	-0.02	0.15	*49*	b	-0.01	0.68	-0.01	0.68	2.63	1.22	*2.0*
51	50	0.00	0.12	1.77	1.19	1.05	0.77	-0.02	0.11	*64*	b	0.01	0.36	0.01	0.36	3.22	1.18	*2.3*
6	100	0.06	0.34	1.85	1.28	1.43	1.04	-0.10	0.23	*105*		0.09	0.68	0.09	0.67	4.25	1.20	*1.6*
11	100	0.02	0.18	1.94	1.27	1.49	1.01	-0.06	0.14	*97*	b	0.05	0.39	0.05	0.39	4.07	1.21	*1.7*
26	100	0.01	0.12	2.00	1.20	1.52	0.97	-0.03	0.11	*84*	b	0.01	0.27	0.01	0.27	3.99	1.21	*1.7*
6	10	0.03	0.32	1.23	0.96	0.42	0.43	-0.09	0.42	*120*		0.05	2.65	0.05	2.65	2.16	1.20	*2.7*
11	20	0.01	0.20	1.53	1.08	0.59	0.50	-0.02	0.21	*76*		0.07	0.82	0.07	0.82	2.78	1.19	*2.2*
26	50	0.00	0.14	1.89	1.17	1.11	0.74	-0.03	0.13	*69*	b	0.02	0.38	0.02	0.38	3.44	1.17	*2.3*
51	100	0.01	0.11	1.96	1.21	1.50	0.95	-0.03	0.10	*78*	b	0.01	0.24	0.01	0.24	3.41	1.10	*2.9*
Panel C: Nonstationary factors with drift																		
51	10	0.01	0.17	0.13	0.66	-0.03	0.45	0.00	0.17	*34*		0.03	0.66	0.03	0.66	0.17	0.80	*2.2*
51	20	-0.01	0.11	0.18	0.93	-0.04	0.49	-0.01	0.11	*29*	b	0.00	0.22	0.00	0.22	0.27	1.09	*2.1*
51	50	0.00	0.07	0.53	1.53	0.02	0.67	-0.01	0.07	*31*	b	0.00	0.15	0.00	0.15	0.70	1.39	*2.5*
6	100	-0.01	0.19	0.98	2.44	-0.03	0.99	-0.03	0.14	*67*	b	-0.03	0.33	-0.03	0.33	2.94	3.50	*2.4*
11	100	-0.01	0.09	0.98	2.43	-0.02	0.98	-0.01	0.08	*46*	b	-0.02	0.21	-0.02	0.20	2.01	2.74	*2.4*
26	100	0.00	0.06	0.87	2.39	-0.02	0.99	-0.01	0.06	*35*	b	-0.01	0.13	-0.01	0.13	1.40	2.08	*2.4*

6	10		-0.02	0.22		0.13	0.63		0.00	0.44		-0.12	0.42	*94*		0.01	1.23		0.01	1.23		0.29	1.07	*2.6*	
11	20		-0.01	0.15		0.20	0.94		-0.04	0.48		-0.04	0.26	*50*		-0.01	0.26		-0.01	0.25		0.32	1.16	*2.4*	
26	50		-0.01	0.09		0.51	1.52		0.00	0.72		-0.01	0.07	*34*	[b]	-0.02	0.17		-0.02	0.17		0.74	1.49	*2.5*	
51	100		0.00	0.05		0.95	2.41		-0.01	1.02		-0.01	0.05	*32*	[b]	-0.01	0.12		-0.01	0.11		1.31	1.98	*2.4*	

[a] See Section 5.1 and Online Appendix 3 for details of DGPs and estimators. PCDID: basic PCDID estimator. DID-2wfe: Two way fixed effects estimator. DID-trend: DID with unit-specific cubic trend. BAI: Bai (2009)'s iterative estimator. GSC: Xu (2017)'s generalized synthetic control estimator. GM: Gobillon and Magnac (2016)'s stepwise estimator. MC-NNM: Athey et al. (2018)'s nuclear norm matrix completion estimator. SD: empirical standard deviation of estimator. Avg #Iter: average number of iterations used. Avg Rank: Average matrix rank computed. Panels A, B and C refer to scenario A, B and C in Section 5.1. Number of replications=1000. In all specifications, N_E=1 and T_0=T_1=T/2. PCDID, BAI, GSC and GM assume 3 factors.

[b] Numerical convergence is attained in all replications. See Section 5.1 and Online Appendix 3.3 for numerical convergence criteria.

TABLE II: SMALL SAMPLE PROPERTIES OF ESTIMATORS, ATET ESTIMATION[a]

		PCDID		DID-2wfe		DID trend		BAI			GSC		GM		MC-NNM		
N_E+N_C	T	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Avg #Iter	Bias	SD	Bias	SD	Bias	SD	Avg Rank
Panel A: Stationary factors																	
100	10	-0.01	0.15	0.00	0.44	-0.01	0.27	-0.34	0.49	*82*	0.00	0.26	0.00	0.26	0.00	0.48	*1.5*
100	20	0.00	0.15	-0.02	0.45	-0.01	0.27	-0.36	0.57	*113*	0.00	0.15	0.00	0.15	-0.02	0.41	*1.1*
100	50	0.00	0.15	-0.02	0.40	-0.01	0.34	-0.35	0.52	*97*	0.00	0.15	0.00	0.15	-0.04	0.25	*1.7*
10	100	0.01	0.45	0.01	0.51	0.02	0.56	-0.36	0.69	*179*	0.01	0.45	0.01	0.45	0.01	0.48	*1.1*
20	100	0.02	0.32	0.01	0.41	0.04	0.45	-0.36	0.66	*142*	0.02	0.32	0.02	0.32	0.02	0.38	*1.1*
50	100	0.00	0.21	0.00	0.32	0.02	0.37	-0.34	0.57	*107*	0.00	0.21	0.00	0.21	-0.01	0.27	*1.1*
10	10	0.02	0.48	0.02	0.62	0.01	0.53	-0.32	0.62	*181*	0.02	0.62	0.02	0.62	0.02	0.64	*2.2*
20	20	0.01	0.35	-0.02	0.55	-0.01	0.42	-0.34	0.66	*169*	0.01	0.36	0.01	0.36	-0.01	0.46	*1.6*
50	50	0.01	0.19	0.00	0.42	0.01	0.36	-0.33	0.57	*116*	0.01	0.20	0.01	0.20	-0.03	0.26	*2.4*
100	100	0.01	0.14	0.01	0.28	0.03	0.34	-0.37	0.47	*72*	0.01	0.14	0.01	0.14	-0.01	0.23	*1.1*
Panel B: Stationary factors with break																	
100	10	0.00	0.16	1.24	0.46	0.41	0.27	-0.34	0.69	*203*	0.03	1.13	0.03	1.13	2.61	0.95	*2.2*
100	20	0.00	0.15	1.54	0.50	0.59	0.29	-0.35	0.76	*238*	0.02	0.27	0.02	0.27	3.37	0.69	*2.2*
100	50	0.00	0.15	1.80	0.44	1.06	0.36	-0.26	0.70	*263*	0.01	0.18	0.01	0.18	3.82	0.51	*2.3*
10	100	0.07	0.52	1.86	0.89	1.42	0.78	-0.24	0.66	*203*	0.13	0.69	0.13	0.69	4.62	0.74	*2.0*
20	100	0.04	0.33	1.91	0.63	1.47	0.58	-0.20	0.57	*196*	0.06	0.41	0.06	0.41	4.40	0.59	*2.3*
50	100	0.00	0.21	1.90	0.47	1.45	0.45	-0.20	0.51	*186*	0.02	0.24	0.02	0.24	4.15	0.47	*2.2*
10	10	0.06	0.50	1.28	0.76	0.45	0.55	-0.29	0.74	*223*	0.15	2.61	0.15	2.61	2.69	1.10	*2.7*
20	20	0.02	0.37	1.53	0.70	0.59	0.45	-0.32	0.76	*235*	0.07	0.63	0.07	0.63	3.54	0.80	*2.1*
50	50	0.02	0.21	1.81	0.53	1.07	0.40	-0.23	0.68	*245*	0.03	0.27	0.03	0.27	4.05	0.54	*2.2*
100	100	0.01	0.14	1.91	0.37	1.46	0.38	-0.19	0.48	*176*	0.01	0.16	0.01	0.16	3.39	0.39	*3.0*
Panel C: Nonstationary factors with drift																	
100	10	-0.01	0.15	0.12	0.54	-0.02	0.38	-0.27	0.64	*143*	-0.03	0.41	-0.03	0.41	0.54	1.68	*2.1*
100	20	0.00	0.15	0.18	0.75	-0.03	0.43	-0.21	0.94	*239*	0.00	0.16	0.00	0.16	0.78	1.99	*2.5*
100	50	0.00	0.15	0.51	1.12	0.02	0.57	-0.17	1.30	*342*	0.00	0.15	0.00	0.15	1.98	2.60	*3.1*
10	100	-0.01	0.48	0.92	1.90	0.00	0.98	-0.23	1.36	*299*	-0.02	0.53	-0.02	0.53	5.34	4.80	*2.7*
20	100	0.01	0.32	0.98	1.82	0.00	0.91	-0.22	1.42	*335*	0.00	0.35	0.00	0.35	5.01	4.52	*3.0*
50	100	0.00	0.21	0.95	1.66	-0.01	0.84	-0.23	1.48	*359*	-0.01	0.22	-0.01	0.22	4.82	4.33	*2.9*

10	10		0.00	0.49		0.15	0.75		0.01	0.60		-0.28	0.74	*212*		0.00	0.67		0.00	0.67		0.59	1.85	*2.7*	
20	20		-0.01	0.35		0.19	0.81		-0.03	0.54		-0.23	0.96	*249*		0.00	0.38		0.00	0.38		0.86	2.17	*2.7*	
50	50		0.01	0.20		0.53	1.14		0.03	0.59		-0.22	1.27	*332*		0.00	0.21		0.00	0.21		2.28	2.91	*3.1*	
100	100		0.01	0.14		0.98	1.60		-0.01	0.83		-0.22	1.42	*361*		0.01	0.15		0.01	0.15		3.96	3.70	*3.1*	

[a] See Section 5.1 and Online Appendix 3 for details of DGPs and estimators. PCDID: PCDID simple mean-group (PCDID-MG) estimator. DID-2wfe: Two way fixed effects estimator. DID-trend: DID with unit-specific cubic trend. BAI: Bai (2009)'s iterative estimator. GSC: Xu (2017)'s generalized synthetic control estimator. GM: Gobillon and Magnac (2016)'s stepwise estimator. MC-NNM: Athey et al. (2018)'s nuclear norm matrix completion estimator. SD: empirical standard deviation of estimator. Avg #Iter: average number of iterations used. Avg Rank: Average matrix rank computed. Panels A, B and C refer to scenarios A, B and C in Section 5.1. Number of replications=1000. In all specifications, N_E=N_c=N/2 and T_0=T_1=T/2. PCDID, BAI, GSC and GM assume 3 factors. In BAI, numerical convergence is attained in some replications only. See Online Appendix 3.3 for numerical convergence criteria.

TABLE III: PCDID INFERENCE PROCEDURES, REJECTION RATE (%)[a]

DGPs for ITET estimation (Table I)						DGPs for ATET estimation (Table II)					
Basic PCDID estimator, ITET inference (H_0: ITET=3)						PCDID-MG estimator, ATET inference (H_0: ATET=3)					
N_E+N_C	T	TrueF	Asym	b-t	b-se	N_E+N_C	T	TrueF	Asym	b-t	b-se
Panel A: Stationary factors											
51	10	4.4	5.5	3.5	3.7	100	10	5.4	6.5	5.3	6.5
51	20	4.2	6.2	6.1	4.5	100	20	5.0	5.3	5.2	5.6
51	50	4.1	5.0	6.9	3.3	100	50	5.7	5.9	6.6	5.6
6	100	4.6	22.6	10.6	3.8	10	100	5.1	5.4	7.3	3.5
11	100	5.4	14.3	6.8	1.0	20	100	3.9	3.9	4.9	4.3
26	100	4.8	6.3	3.5	1.1	50	100	5.1	5.3	5.4	4.9
6	10	4.8	8.8	6.6	6.4	10	10	5.6	6.5	10.4	5.2
11	20	4.8	9.5	7.3	4.6	20	20	5.5	6.7	7.9	8.9
26	50	5.0	7.8	7.0	3.3	50	50	4.8	4.7	4.9	5.6
51	100	5.1	4.7	4.7	3.1	100	100	4.5	4.8	4.5	4.9
Panel B: Stationary factors with break											
51	10	4.9	5.9	4.7	5.5	100	10	5.5	7.6	5.3	5.4
51	20	6.2	7.2	7.0	5.4	100	20	5.0	6.4	4.8	5.0
51	50	5.1	4.9	5.3	3.6	100	50	6.0	6.0	4.9	4.8
6	100	5.0	32.4	13.6	9.7	10	100	5.2	7.2	7.8	2.3
11	100	5.8	13.9	6.5	2.9	20	100	4.9	5.9	5.5	3.9
26	100	6.7	8.1	7.0	3.5	50	100	5.2	5.3	5.1	5.1
6	10	4.7	5.3	4.2	3.6	10	10	5.9	7.6	9.4	4.0
11	20	4.0	7.1	6.4	3.2	20	20	6.2	7.9	7.7	7.0
26	50	5.8	6.3	7.2	4.5	50	50	4.6	5.2	4.3	4.1
51	100	4.9	5.0	5.1	3.7	100	100	4.7	5.5	5.2	4.7
Panel C: Nonstationary factors with drift											
51	10	4.1	6.0	4.5	5.7	100	10	5.4	6.6	5.9	7.3
51	20	4.8	5.6	6.5	5.3	100	20	4.7	6.2	5.8	5.9
51	50	6.6	6.2	6.7	5.5	100	50	5.8	5.7	5.2	5.3
6	100	3.6	26.6	7.9	3.3	10	100	5.0	6.2	7.8	2.4
11	100	5.3	15.0	9.8	3.6	20	100	4.0	4.0	4.8	4.0
26	100	4.8	5.9	7.2	2.5	50	100	4.8	5.1	5.4	5.0
6	10	5.3	6.8	5.5	3.7	10	10	5.6	6.9	9.5	5.1
11	20	3.2	8.4	6.4	4.3	20	20	6.5	7.6	8.1	7.9
26	50	4.8	6.6	8.0	4.6	50	50	5.0	5.4	4.8	5.7
51	100	4.8	4.9	5.9	3.8	100	100	4.6	4.6	4.9	4.6

[a] See Section 5.2 for details of inference procedures. All DGPs are the same as in Table I and II, respectively (except setting ρ_ϵ=0 in the DGPs for ITET inference). The null hypothesis is set at the DGPs' true value. TrueF: assume factors are observed in PCDID estimation (infeasible) and compute t-statistic, reject if $|t|>=1.96$. Asym: use 3 factor proxies in PCDID estimation and compute t-statistic, reject if $|t|>=1.96$. b-t: same as Asym, but reject if $t<=c_{0.025}$ or $t>=c_{0.975}$ where $c_{0.025}$, $c_{0.975}$ are percentiles of the bootstrap distribution of t-statistics. b-se: same as Asym, but the standard error in the t-statistics formula is obtained from bootstrap samples. See Section 5.2 for analytical standard errrors used in TrueF, Asym and b-t. See Online Appendix 6 for details on bootstrap sample construction. Number of replications=1000. Bootstrap repetitions = 199. T_0=T_1=T/2. In ITET inference, N_E=1; in ATET inference, N_E=N_C=N/2. The nominal size is 5%.

TABLE IV: EFFECTS OF WELFARE WAIVERS ON WELFARE CASELOADS[a]

	Panel A: PCDID (#PC=4)[b]				Panel B: Extra specs (all treated states)[b]			Panel C: DID regressions (all treated states)[c]		
	ALL treated states	Southern treated states	Non-southern treated states	Wyoming (ITET)[d]	PCDID (#PC=3)	PCDID (#PC=4)	GSC (#PC=4)	Unit-specific time polynomial: Quartic	Unit-specific time polynomial: Cubic	Two-way fixed effect
Intervention dummy ($1\{t>T_{0i}\}$)	-0.017	-0.024	-0.013	-0.114	-0.018	-0.013	-0.033	-0.007	-0.008	-0.054
asym-se	(0.007)	(0.007)	(0.010)	(0.029)	(0.008)	(0.006)	(0.011)			
b-se	(0.009)	(0.012)	(0.011)	(0.051)	(0.010)	(0.008)	(0.009)	(0.010)	(0.013)	(0.032)
b-t-pval	{0.030}	{0.000}	{0.241}	{0.010}	{0.050}	{0.020}	{0.000}			
Stepwise TE dynamics ($1\{t>T_{0i}+3\}$)						-0.000				
asym-se						(0.007)				
b-se						(0.017)				
b-t-pval						{0.975}				
Max monthly welfare ben. ($100)	0.014	0.046	-0.001	-0.010	0.035	0.013		-0.003	-0.003	0.020
asym-se	(0.008)	(0.019)	(0.004)	(0.020)	(0.016)	(0.008)				
b-se	(0.016)	(0.033)	(0.009)	(0.027)	(0.016)	(0.015)		(0.002)	(0.002)	(0.014)
b-t-pval	{0.121}	{0.000}	{0.834}	{0.754}	{0.010}	{0.191}				
State unemployment rate (%)	0.021	0.016	0.023	-0.030	0.029	0.021	0.018	0.007	0.011	0.023
asym-se	(0.004)	(0.006)	(0.004)	(0.008)	(0.005)	(0.004)	(0.004)			
b-se	(0.007)	(0.007)	(0.008)	(0.014)	(0.007)	(0.007)	(0.007)	(0.002)	(0.003)	(0.008)
b-t-pval	{0.000}	{0.000}	{0.000}	{0.020}	{0.000}	{0.000}	{0.000}			
Ln(state empl. to popn ratio)	0.058	0.070	0.052	-0.622	-0.128	0.043	-0.016	-0.338	-0.198	-1.346
asym-se	(0.129)	(0.236)	(0.157)	(0.357)	(0.134)	(0.130)	(0.144)			
b-se	(0.164)	(0.241)	(0.167)	(0.502)	(0.170)	(0.168)	(0.176)	(0.087)	(0.110)	(0.450)
b-t-pval	{0.724}	{0.744}	{0.764}	{0.231}	{0.412}	{0.804}	{0.915}			
% of predicted change in caseload explained by reform (Jan93-Jun96)	6.88%	10.41%	5.11%	24.86%	6.89%	5.95%	14.05%	2.88%	3.29%	24.00%

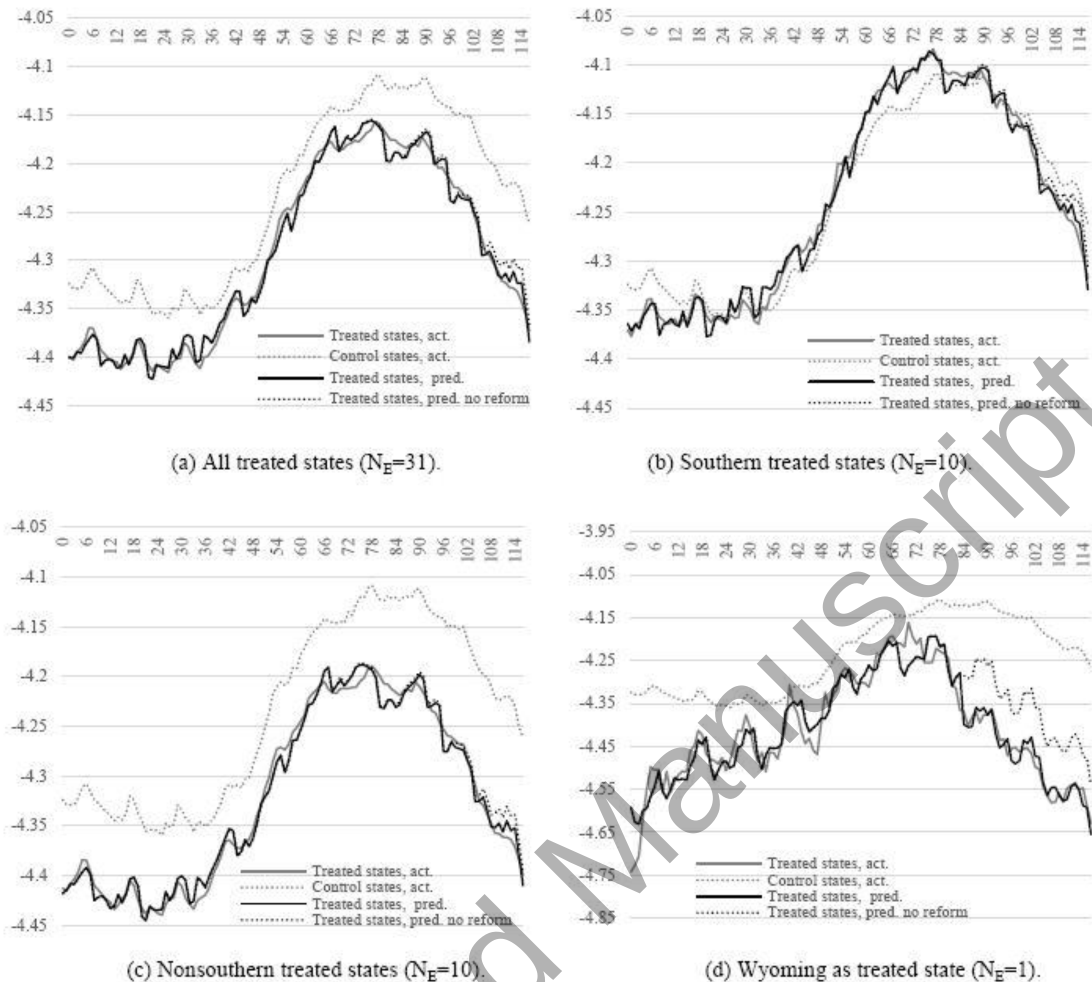
Alpha statistic (raw)	0.992		1.189		0.898		-														
asym-se	(0.138)		(0.183)		(0.183)		-														
b-t-pval	{0.824}		{0.362}		{0.864}		-														
Number of treated states	31		10		21		1		31		31		31			31		31		31	

[a] The sample period is from Oct86 to Jun96. There are 20 control states. Policy intervention dummy: =1 if a work requirement or time limit waiver is approved/implemented in the state, =0 otherwise. Max monthly welfare ben.: maximum combined real AFDC/Food Stamp benefits for a family of three (in $100). Except the 2wfe specifications, covariates include calendar quarter dummies.

[b] Standard errors from the asymptotic formula (asym-se) and bootstrapping the coefficient (b-se) are reported in parentheses. P-values from bootstrapping the t-statistic (b-t-pval) are reported in curly brackets. See Online Appendix 6 for details of bootstrap sample construction. Bootstrap repetitions=199.

[c] Wild cluster bootstrapped standard errors are in parentheses. See Online Appendix 6 for details. Bootstrap repetitions=199.

[d] For ITET, the asymptotic standard error is obtained from the Newey-West HAC estimator with $T^{1/4} \approx 3$ lags.



(a) All treated states (N_E=31).

(b) Southern treated states (N_E=10).

(c) Nonsouthern treated states (N_E=10).

(d) Wyoming as treated state (N_E=1).

Figure 1. Actual and predicted welfare caseloads, PCDID model with 4 PCs.

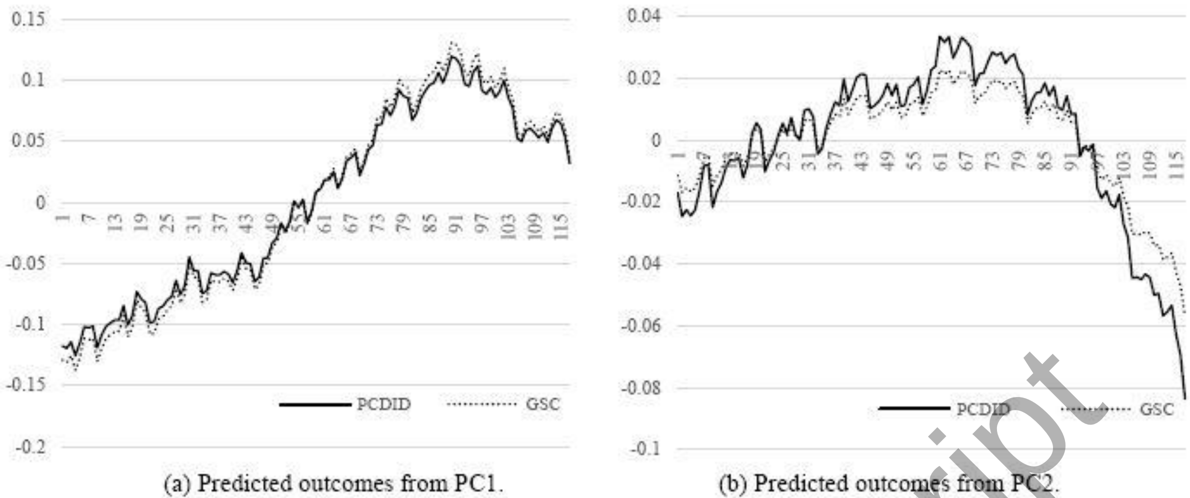


Figure 2. Predicted trends of average caseloads in treated states from the estimated factor structure.