Economics Working Paper Series

2016 - 11

Policy Evaluation with Interactive Fixed Effects

Mark K. Chan & Simon Kwok

June 2016

Policy Evaluation with Interactive Fixed Effects

Marc K. Chan
University of Technology Sydney

Simon Kwok
The University of Sydney *

June 2016

Abstract

We develop an alternative estimator for policy evaluation in the presence of interactive fixed effects. It extends Pesaran (2006)'s two-stage procedure to a difference-in-differences-type program evaluation framework, and extracts principal components from the control group to form factor proxies. Consistency and asymptotic distributions are derived under stationary factors, as well as nonstationary factors with any integration order. Simulation exercises demonstrate excellent performance of our estimator relative to existing methods. We present empirical results from microeconomic and macroeconomic applications. We find that our estimator generates the most robust treatment effect estimates, and our weights for control group units deliver strong economic interpretation regarding the nature of the underlying factors.

Keywords: program evaluation, interactive fixed effects, difference-in-differences

*Chan: University of Technology Sydney, UTS Business School, Sydney, Ultimo NSW 2007, Australia; Email: marc.chan@uts.edu.au. Kwok: School of Economics, Merewether Building, The University of Sydney, NSW 2006 Australia; Email: simon.kwok@sydney.edu.au.

1 Introduction

Panel data models with multifactor error structures ("interactive fixed effects") have become increasingly popular in recent years. In these models, the outcomes are influenced by unobserved common factors that vary over time. Heterogeneity in the sensitivity (factor loadings) to these factors across different entities can generate rich cross-sectional dependence in the data. Estimation procedures developed by Pesaran (2006) and Bai (2009) have resulted in many applications related to regional-level panel data.

Despite the increasing popularity of these methods, few studies have applied them directly to program evaluation – measurement of the impact of a particular policy intervention. One exception is Kim and Oka (2014), who use state-level panel data and the Bai (2009) estimator to study the effects of unilateral divorce law reforms on divorce rates in the US.[1] They point out that earlier studies are based on the difference-in-differences (DID) estimator, which does not fully account for the rich cross-sectional dependence and unobserved heterogeneity in the data. Their empirical analysis reconciles some of the mixed evidence in earlier studies, while also showing that their approach produces more robust estimates. Gobillon and Magnac (2016) clarify the identification conditions that allow the Bai (2009) estimator to be used in program evaluation. In Monte Carlo simulations, they find that Bai (2009) performs better than DID under certain factor specifications but worse in others. The estimator is used to evaluate the impact on local unemployment of an enterprise zone policy in France. They show that their earlier estimated impacts from DID (Gobillon, Magnac and Selod (2012)) are robust to interactive fixed effects.

In this paper, we develop an alternative approach for policy evaluation in the presence of interactive fixed effects. Our estimator is implemented in two stages. First, use the full panel of control group outcomes to extract principal components, which are in the form of weighted outcomes of control group units. These principal components serve as proxies for linear combinations of unobserved factors. Then, for each treatment group unit, estimate the unit-specific treatment effect in a time-series regression with a post-intervention period indicator augmented with the principal components.

[1] Among the numerous applications of Pesaran (2006), one example is Holly, Pesaran and Yamagata (2010), who study the determinants of house prices in a panel with 49 US states over 29 years. They find that real house prices have generally been rising in line with fundamentals (real incomes), but there are also some outlier states. They do not consider evaluation of particular housing policies.

Our estimator can be considered as an extension to Pesaran (2006). He uses cross-sectional averages of time-varying variables in the model as proxies for (linear combinations of) common factors, which are then used as augmented regressors in the second-stage regression. His estimator is not directly applicable here due to two complications: (i) the DID-type program evaluation framework is incompatible with some of the modeling assumptions; (ii) the treatment effect contaminates factor proxies that are formed by cross-sectional averages from all units. We propose constructing *principal components* from data on *control group units* to form factor proxies. The DID-type framework also prompts us to consider asymptotics in the dimensions of control group units (N_C), treatment group units (N_I), pre-intervention periods (T_0), and post-intervention periods (T_1). We derive the consistency and asymptotic distribution of our treatment effect estimator under stationary factors, as well as nonstationary factors with *any* integration order. The properties are obtained under large T, while N (both N_C and N_I) may be finite or infinite. By contrast, the theoretical analysis in Pesaran (2006) and Bai (2009) covered stationary factors under big N and T, and the Pesaran (2006) estimator was found to have the same properties when factors have integration order 1 (Kapetanios, Pesaran and Yamagata (2011)). Our principal component method is also carried out differently from Bai (2009). He uses a $T \times T$ covariance matrix to extract the principal components, which are used to estimate the factors and factor loadings explicitly.[2] By contrast, we extract the principal components from an $N_C \times N_C$ covariance matrix. The factors and factor loadings are treated as nuisance parameters and are not estimated separately. Our estimator possesses different asymptotic properties and more robust small sample performance relative to Bai's approach.

Through Monte Carlo simulation and two empirical applications, we compare the performance of our estimator with DID, Bai (2009), and "synthetic control" methods. Synthetic-control-style methods (e.g., Abadie, Diamond and Hainmueller (2010), Hsiao, Ching and Wan (2012)) have become popular as a tool for program evaluation in recent years. They match the outcome of the treatment unit via weighted control group units ("synthetic treatment unit") during *pre-intervention periods*. The weights are computed by constrained optimization in Abadie, Diamond and Hainmueller (2010), and by auxiliary regression in Hsiao, Ching

[2] He uses the whole panel of regression residuals to construct a $T \times T$ covariance matrix and extract principal components. The procedure is carried out iteratively to update the parameter estimates. The interactive-effect estimator is consistent and asymptotically normal under both large N and large T.

and Wan (2012). Assuming the weights are time-invariant, a counterfactual path of the treatment unit can be constructed in post-intervention periods. Synthetic control methods have been analyzed under stationary factors. Recent work has also shown that caution should be exercised due to implicit assumptions. For instance, Gobillon and Magnac (2016) derive the support conditions for Abadie, Diamond and Hainmueller (2010), in which the weights are restricted between 0 and 1. They show that bias can occur if characteristics (e.g., factor loadings) of the treatment unit is not a subset of the support of characteristics of control group units. Their simulation exercise confirms that the bias worsens when the supports of factor loadings become different between treatment and control units.

Our simulation exercise demonstrates excellent performance of our estimator relative to existing methods, across various specifications of factors and sample size (N_C, N_I, T_0, T_1). Our estimator generates the lowest empirical SD, especially in the presence of persistent stationary factors or nonstationary factors. We find that Hsiao, Ching and Wan (2012) is the second best performer, followed by DID and Bai (2009). Hsiao, Ching and Wan (2012) still performs relatively well under factors with integration order 1; by contrast, DID and Bai (2009) deteriorate quickly when factors are nonstationary. Our estimator is robust to nonstationary factors of higher integration orders, and is the only usable choice under such cases.

We present results from one macroeconomic and one microeconomic application: (i) effect of political and economic integration of Hong Kong with mainland China on the GDP growth of Hong Kong (see Hsiao, Ching and Wan (2012); (ii) effect of a large-scale tobacco control program on California's cigarette sales (see Abadie, Diamond and Hainmueller (2010)). In both applications, we find that our estimator generates the most robust estimates, and our weights for control group units deliver strong economic interpretation regarding the nature of the underlying factors. Although we find that results from different methods can be coherent in some scenarios, there are also scenarios in which the robustness of existing methods is undermined. For instance, in the GDP growth application, synthetic control fails to capture the factors underlying the Asian Financial Crisis, which occurred shortly after the policy intervention – the return of sovereignty of Hong Kong to China. This results in a spurious negative (but insignificant) effect of political integration. In the cigarette sales application, we find strong evidence for persistent, long-run trends in levels of cigarette sales. The existing methods tend to attribute these stochastic trends to the treatment effect, resulting in overestimation of the effect. By contrast, our estimator (under multiple principal components) finds a smaller treatment effect in level data, which is consistent with evidence from data on *yearly change* of cigarette sales.

The paper proceeds as follows. Section 2 discusses the model and estimator. Sections 3 and 4 report simulation and empirical results, respectively. Section 5 concludes.

2 The Model and Estimator

We have a panel consisting of N cross sectional units and T time periods. Suppose that the cross section consists of treatment group I with N_I treatment units, and control group C with N_C control units. There is a policy intervention occurring right after time period T_0, where $0 < T_0 < T$. The time horizon is splitted into pre-intervention periods $T^{pre} = \{1, \ldots, T_0\}$ and post-intervention periods $T^{post} = \{T_0 + 1, \ldots, T\}$.

If there is no policy intervention at all, all units would have the outcome of y_{it}^0, which is assumed to take the following multifactor structure

$$y_{it}^0 = \underset{1\times p}{\gamma_i'} \cdot \underset{p\times 1}{d_t} + \underset{1\times m}{\mu_i'} \cdot \underset{m\times 1}{f_t} + \epsilon_{it},$$

where d_t is a vector of p deterministic factors, f_t is a vector of m stochastic factors, and ϵ_{it} is an idiosyncratic noise. The factor loadings γ_i and μ_i are treated as fixed effects associated with the deterministic and stochastic factors, respectively.[3]

Now, with the policy in place, we assume that the ith treatment unit is subject to a treatment effect of Δ_i on top of y_{it}^0 during the post-intervention periods, with the observed outcome given by $y_{it}^0 + \Delta_i$; otherwise, the observed outcome is given by y_{it}^0 for the treatment units during the pre-intervention periods, and for all control units at all time periods. In the end, the econometrician observes a panel of $N \times T$ outcomes, given by

$$y_{it} = y_{it}^0 + \Delta_i 1_{\{i\in I\}} 1_{\{t>T_0\}}$$

[3] The model with fixed effect in levels is obtained as a special case when one of the components of d_t is set to 1.

for $i = 1 \ldots, N$ and $t = 1, \ldots, T$.

Assumption 1 (treatment status): For any i and t, we have (i) $E(\epsilon_{it} | 1_{\{i \in I\}}, \gamma_i, \mu_i, f_t) = 0$; and (ii) $0 < E(1_{\{i \in I\}}) < 1$.

Assumption 1(i) means that the treatment group indicator $1_{\{i \in I\}}$, the unobserved factors and factor loadings are jointly predetermined. Note that this assumption does not prohibit the presence of correlation between $1_{\{i \in I\}}$ and μ_i. Assumption 1(ii) restricts the probability of an individual being in the treatment group to be strictly between 0 and 1.[4]

Substituting y_{it}^0 into the equation, the observations y_{it} can be expressed as follows:

$$y_{it} = \underset{1 \times p}{\gamma_i'} \cdot \underset{p \times 1}{d_t} + \underset{1 \times m}{\mu_i'} \cdot \underset{m \times 1}{f_t} + \Delta_i 1_{\{i \in I\}} 1_{\{t > T_0\}} + \epsilon_{it}. \tag{1}$$

Our goal is to estimate the treatment effect Δ_i. Let us stack the control group observations into vector form:

$$\underset{N_C \times 1}{y_{Ct}} = \underset{N_C \times p}{\gamma_C'} \cdot \underset{p \times 1}{d_t} + \underset{N_C \times m}{\mu_C'} \cdot \underset{m \times 1}{f_t} + \epsilon_{Ct}. \tag{2}$$

Collecting all T observations into columns, we obtain the regression in matrix form:

$$\underset{N_C \times T}{y_C} = \underset{N_C \times p}{\gamma_C'} \cdot \underset{p \times T}{D} + \underset{N_C \times m}{\mu_C'} \cdot \underset{m \times T}{F} + \underset{N_C \times T}{\epsilon_C}. \tag{3}$$

For identification, we first consider the following assumptions on the common factors and factor loadings.[5]

Assumption 2 (common factors): (i) $d_t = [d_{1t}, \ldots, d_{pt}]'$ is a vector of p deterministic factors; (ii) $f_t = [f_{1t}, \ldots, f_{mt}]'$ is a vector of m stochastic factors. The stochastic factor f_{jt} is an $I(r_j)$ process with any fixed integration order $r_j \geq 0$, for $j = 1, \ldots, m$. The integration order can differ across factors. When $r_j = 0$ for all j, f_{jt} is a covariance stationary process; (iii) $rank(F) = m$; (iv) $1_{post}' M_F 1_{post} > 0$, where 1_{post} is the $T \times 1$ vector of post-intervention period indicators (consisting of T_0 zeros followed by T_1 ones), and $M_F = I_{T \times T} - F'(FF')^{-1}F$.

[4] Assumption 1 is common in the related literature, e.g., Hsiao et al. (2012) and Gobillon and Magnac (2016). Note that the latter treats the factors as given.

[5] When we discuss the principal components method in later sub-sections, we will provide sufficient conditions such that Assumption 3 holds.

Assumption 3 (factor loadings): For a given $N_C \times T$ panel of control group observations, we have $rank(\mu_C) = m$.

Assumption 2(ii) allows for arbitrary and heterogeneous integration orders on the stochastic factors. This is considerably more general than the assumptions in the interactive fixed effect literature (e.g., Pesaran (2006), Bai (2009), Kapetanios et al. (2011)) and in the program evaluation literature (e.g., Abadie et al. (2010), Hsiao et al. (2012), and Gobillon and Magnac (2016)).[6] Assumption 2(iii) makes sure that the m factors are linearly independent. This entails that the total number of sample periods T is no less than the number of factors m. Assumption 2(iv) is an identification condition which essentially requires that the vector of post-intervention period indicators is not in the space spanned by the m factors; otherwise, multicollinearity would occur, and the individual treatment effect is not identifiable. This assumption is similar to Assumption 5(a) of Pesaran (2006) and the condition in Proposition 3 of Bai (2009), but adapted to the individual treatment effect model (1). Gobillon and Magnac (2016) imposes the same identification condition.

Assumption 3 ensures sufficient variability in the factor loadings to allow for the identification of the m factors. This is often satisfied with large N_C in practice. This is the same as Assumption 4 in Hsiao et al (2012). Pesaran (2006) and Kapetanios et al (2011) require similar rank assumption for identification of individual-specific coefficients under the panel set-up, although they do not require this assumption for identifying the average of the individual-specific coefficients. It is possible to replace the rank condition in Assumption 3 by a weaker condition:

Assumption 3' (factor loadings): (i) For a given $N_C \times T$ panel of control group observations, we have $rank(\mu_C) = p$, where $0 < p \leq m$; (ii) If $p < m$, then for all $i = 1, \ldots N$, we have $\mu_i = R\xi_i$ for some $p \times 1$ vector ξ_i and a $m \times p$ matrix R.

Assumption 3' relaxes the full row rank condition of μ_C. The cost is that we can identify at most p (linearly independent) combinations out of the m unobserved factors, and we need the

[6] Kapetanios et al. (2011) derives asymptotic theories that extend Pesaran (2006)'s estimation method to the case with I(1) factors. Bai et al. (2009) considers the estimation of a panel cointegration model with I(1) factors. Abadie et al. (2010)'s synthetic control method assumes that the factors to be bounded uniformly from above, thus ruling out nonstationary factors. Hsiao et al. (2012) considers identification with arbitrary factors but the results on asymptotic distributions only apply under stationary factors.

proportionality condition of Assumption 3'(ii) to ensure identification. The proportionality condition essentially restricts that the factor loadings on the m factors remain to be of the same ratio (as defined in R) for all units.

Example: Suppose $m = 2$, $N_C = 3$, $N_I = 1$, $\mu_C = \begin{bmatrix} 1 & 2 \\ 1.1 & 2.2 \\ 1.5 & 3 \end{bmatrix}$, and $\mu_I = [1.2 \quad 2.4]$. Then $rank(\mu_C) = 1$, and assumption 3'(ii) is satisfied because we have $R = [1 \quad 2]'$ such that $\mu_i = R\xi_i$ for $i \in C$ and $i \in I$. Specifically, $\xi_1 = 1, \xi_2 = 1.1, \xi_3 = 1.5, \xi_4 = 1.2$.

2.1 Construction of Factor Proxies

To form factor proxies, we construct m linearly independent weighted cross-sectional means out of the control group observations y_{Ct}. Let $W = [w_1, \ldots, w_m]$ be an $N_C \times m$ weighting matrix, so that $W' y_{Ct} = [w_1' y_{Ct}, \ldots, w_m' y_{Ct}]$ contains m different weighted cross-sectional means of y_{Ct} at time t. The weighting matrix is common to all periods t, and satisfies the following assumption.

Assumption 4 (weighting matrix): Let w_{ij} be the $(i,j)^{th}$ element of W, an $N_C \times m$ weighting matrix. The weighting matrix must satisfy (i) $\sum_{i=1}^{N_C} w_{ij} = 1$; (ii) $\sum_{i=1}^{N_C} |w_{ij}| < K$ for some constant K; and (iii) no linear combinations of the columns of W are in the null space of μ_C.

Assumption 4(i)-(ii) are analogous to Assumption 5(ii)-(iii) in Pesaran (2006). When $m = 1$, one simple example that satisfies Assumption 4(i)-(ii) is the equal weighting scheme, with $w_{i1} = N_C^{-1}$ for all $i \in C$. Assumption 4(iii) is used for preserving the rank of μ_C (discussed later), and is often satisfied with large N_C in practice.

In practice, the number of factors m is unknown, but it is possible to estimate it empirically. The number of factors can be chosen empirically such that a pre-specified amount of variability of y_{Ct} is explained jointly by the chosen principal components. Even if the number of factors m is unknown and not estimated precisely, it is still possible to achieve consistent estimation without loss of asymptotic efficiency by choosing at least m principal components to proxy for the factors (Moon and Weidner, 2015).

In addition, we impose the following assumption on the idiosyncratic errors.

Assumption 5 (idiosyncratic errors): (i) The idiosyncratic error ϵ_{it} satisfies $E(\epsilon_{it}) = 0$ and $E(\epsilon_{it}^4) < \infty$ for all i and t. Also, defining $\sigma_{ij,st} := E(\epsilon_{is}\epsilon_{jt})$, we have $|\sigma_{ij,st}| \leq \bar{\sigma}_{ij}$ for all s, t and $|\sigma_{ij,st}| \leq \tau_{st}$ for all i, j such that, for each $k = 1, ..., m$,

$$\frac{1}{T}\sum_{i,j\in C}\sum_{s,t=1}^{T} w_{ik}\,|\sigma_{ij,st}| \leq M, \qquad \sum_{i,j\in C} w_{ik}\bar{\sigma}_{ij} \leq M, \quad \frac{1}{T}\sum_{s,t=1}^{T}\tau_{st} \leq M;$$

(ii) as $N_C \to \infty$, we have, for each $k = 1, ..., m$,

$$\sum_{i\in C} w_{ik}\bar{\sigma}_{ij} = O\left(\frac{1}{N_C}\right) \text{ for all } j \in C, \qquad \sum_{i,j=1,i\neq j}^{N_C} w_{ik}w_{jk}\bar{\sigma}_{ij} = O\left(\frac{1}{N_C}\right);$$

(iii) the errors ϵ_{is} are independent of the factors f_t and factor loadings γ_j and μ_j, for all i, j and s, t.

Assumption 5(i) allows for weak cross-sectional and serial dependence of ϵ_{it} and is more general than Assumption 2 of Pesaran (2006) and Assumption C of Bai (2009). In particular, it reduces to Bai's Assumption C (ii) under the special case when $w_{ik} = N_C^{-1}$ for all $i \in C$, and Pesaran's Assumption 2 when $w_{ik} = O(\frac{1}{N_C})$. Assumption 5(ii) controls the measurement error related to factor proxies.

2.1.1 Identification strategy

In this section, we will present the identification strategy for the individual treatment effect in (1). Then we will propose the regression model that implements this strategy. For expositional convenience, we will abstract from the presence of deterministic factors in the discussion.

The identification strategy is motivated from the fact that the control group outcome can be decomposed into the interactive fixed effect term and idiosyncratic error. With assumptions for the factor loadings and idiosyncratic errors, we can extract the dynamics of the common factors over both the pre- and post-intervention periods without being contaminated by treatment effects. We will show that the factor proxies constructed this way are consistent as the cross section of the control group sample grows. In this sense, the asymptotic argument for consistency is similar to that for the factor proxies in Pesaran (2006). Unlike other estimation methods for interactive fixed effect models that require large N and T (e.g., Bai (2009)), our factor proxies are consistent as $N_C \to \infty$, with T and N_I remaining finite.

For now, let us suppose that the number of factors is known.[7] By choosing a weighting matrix W that satisfies Assumption 4, the m weighted averages $w_1' y_{Ct}, \ldots, w_m' y_{Ct}$ can serve as a proxy for the m unobserved factors. Let us pre-multiply (2) by W' and we obtain

$$W' y_{Ct} = W' \mu_C' f_t + W' \epsilon_{Ct}. \tag{4}$$

Full rank case If the factor loading matrix μ_C is of full rank (Assumption 3) and the weighting matrix W is not in the null space of μ_C (Assumption 4(iii)), then

$$rank(\mu_C W) = m, \tag{5}$$

and hence the inverse $(\mu_C W)^{-1}$ exists. We can then express the vector of factors f_t into a linear combination of y_{Ct} and ϵ_{Ct}

$$f_t = (W' \mu_C')^{-1} W' y_{Ct} - (W' \mu_C')^{-1} W' \epsilon_{Ct}. \tag{6}$$

By Assumption 5, the variance of the kth weighted average of idiosyncratic errors is

$$\begin{aligned} Var(w_k' \epsilon_{Ct}) &= \sum_{i=1}^{N_C} w_{ik}^2 Var(\epsilon_{it}) + \sum_{i,j=1, i\neq j}^{N_C} w_{ik} w_{jk} Cov(\epsilon_{it}, \epsilon_{jt}) \\ &\leq \sum_{i=1}^{N_C} w_{ik}^2 \bar{\sigma}_{ii} + \sum_{i,j=1, i\neq j}^{N_C} w_{ik} w_{jk} \bar{\sigma}_{ij} \\ &= O\left(\frac{1}{N_C}\right). \end{aligned} \tag{7}$$

This implies that

$$f_t = (W' \mu_C')^{-1} W' y_{Ct} + O_p\left(\frac{1}{\sqrt{N_C}}\right),$$

and hence the linear combination of control group outcomes, $(W' \mu_C')^{-1} W' y_{Ct}$, serves as a good approximation of f_t when the control group is big enough.

Since the outcomes of the treatment group units $i \in I$ depend on the same factor vector

[7] The case with unknown number of factors will be discussed in the next sub-section.

f_t, we can substitute the factor vector expression (6) into (1) and obtain

$$y_{it} = \mu_i'(W'\mu_C')^{-1}W'y_{Ct} + \Delta_i 1_{\{t>T_0\}} + [\epsilon_{it} - \mu_i'(W'\mu_C')^{-1}W'\epsilon_{Ct}]. \tag{8}$$

This motivates us to consider the following regression for treatment unit $i \in I$,

$$y_{it} = \alpha_i' W' y_{Ct} + \delta_i 1_{\{t>T_0\}} + e_{it}. \tag{9}$$

Comparing with (1), the regression coefficients are given by

$$\alpha_i = (\mu_C W)^{-1}\mu_i,$$
$$\delta_i = \Delta_i.$$

This shows that it is possible to use m weighted averages $w_1'y_{Ct}, \ldots, w_m'y_{Ct}$ as factor proxies, provided that (5) is satisfied. The individual treatment effect Δ_i can be identified through the regression coefficient δ_i.

Note that unless we choose W very specifically such that $(\mu_C W)^{-1}$ is an identity matrix, each of the m weighted averages $w_1'y_{Ct}, \ldots, w_m'y_{Ct}$ merely represents a linear combination of the m unobserved factors $f_t = [f_{1t}, \ldots, f_{mt}]'$. The regression model also shows that μ_C and μ_I are in general not separately identified. However, there are exceptions under special cases. For instance, suppose $m = 1$ and we use equal weights (i.e., all elements in w_1 equal $1/N_C$). Then $\mu_C W = (1/N_C)\sum_{i\in C}\mu_i =: \bar{\mu}_C$. This implies $\alpha_i = (\mu_C W)^{-1}\mu_i = \frac{\mu_i}{\bar{\mu}_C}$ for $i \in I$, therefore we can identify the factor loading for each treatment group unit relative to the control group.

Rank deficient case Suppose μ_C is rank deficient in the sense that $rank(\mu_C) = p < m$. We can still identify p linear combinations of factors under the proportionality condition (Assumption 3'(ii)). Intuitively, to identify the p linear combinations out of m factors, we need p linearly independent proxies obtained from p different weighted averages of control group outcomes. The proportionality condition imposes $m - p$ restrictions on the factor loadings so that the remaining p free parameters can be used to identify the p loadings associated with the p linear combinations of factors.

More specifically, given that μ_C is of rank p, there exist a $m \times p$ matrix R with orthogonal columns and a $N_C \times p$ matrix W whose columns are orthogonal and sum to one, such that

$$rank(R'\mu_C W) = p, \tag{10}$$

which implies that the inverse $(R'\mu_C W)^{-1}$ exists. Using this fact and Assumption 3', we can identify the individual treatment effects. Hence the identification strategy is similar to those in the full rank case, and more details are presented in the Appendix.

Remark 1 *Compared with equation (8), the regression error in (9) is* $e_{it} = \epsilon_{it} - \mu_i'(W'\mu_C')^{-1}W'\epsilon_{Ct}$, *which involves the weighted averages of idiosyncratic noises* $W'\epsilon_{Ct}$ *from the control group. When approximating the factor* f_t *with* $W'y_{Ct}$, *there is measurement error, and the covariance between regression error* e_{it} *and factor proxy* $W'y_{Ct}$ *is:*

$$\begin{aligned}
Cov(e_{it}, W'y_{Ct}) &= Cov(-\alpha_i'W'\epsilon_{Ct}, W'\mu_C'f_t + W'\epsilon_{Ct}) + Cov(\epsilon_{it}, W'\mu_C'f_t + W'\epsilon_{Ct}) \\
&= Cov(-\alpha_i'W'\epsilon_{Ct}, W'\mu_C'f_t) + Cov(\epsilon_{it}, W'\mu_C'f_t) + \\
&\quad\; Cov(-\alpha_i'W'\epsilon_{Ct}, W'\epsilon_{Ct}) + Cov(\epsilon_{it}, W'\epsilon_{Ct}) \\
&= -\alpha_i'Var(W'\epsilon_{Ct}) + Cov(\epsilon_{it}, W'\epsilon_{Ct}).
\end{aligned}$$

In the second-to-last equality above, the first two terms are zero by the exogeneity of f_t, *while the remaining terms are non-zero in general. Note that the remaining terms vanish if the number of control units,* N_C, *tends to infinity, due to the fact that they are of order* $O\left(\frac{1}{N_C}\right)$ *by (7).*[8]

Remark 2 *In Pesaran (2006), the overall averages* $W'y_t$ *are used as factor proxies. Can we use this instead of the control group averages* $W'y_{Ct}$ *as the factor proxy? For simplicity, suppose there is only one factor* ($m = 1$), *and we set* $w_{i1} = \frac{1}{N}$ *for all* $i = 1, \ldots, N$. *The*

[8]Under Assumptions 4 and 5, both terms are $O\left(\frac{1}{N_C}\right)$ by (7) and the fact that $|Cov\left(\epsilon_{it}, \sum_{j=1}^{N_C} w_{jk}\epsilon_{jt}\right)| \leq \sum_{j=1}^{N_C} w_{jk}\bar{\sigma}_{ij} = O\left(\frac{1}{N_C}\right)$.

factor proxy now reduces to the simple average, given by

$$\begin{aligned}\bar{y}_t &= \bar{\mu} f_t + \frac{1}{N}\sum_{i=1}^{N}\Delta_i 1_{\{i\in I\}}1_{\{t>T_0\}} + \bar{\epsilon}_t \\ &= \bar{\mu} f_t + \frac{N_I}{N}\frac{1}{N_I}\sum_{i\in I}\Delta_i 1_{\{t>T_0\}} + \bar{\epsilon}_t \\ &= \bar{\mu} f_t + \frac{N_I}{N}\bar{\Delta} 1_{\{t>T_0\}} + \bar{\epsilon}_t,\end{aligned}$$

where $\bar{\Delta}$ is the sample average treatment effect among the treatment units. We readily see that the factor proxy is contaminated by $\bar{\Delta}$. The regression on $\bar{y}_t$ for the treatment unit $i \in I$ becomes

$$\begin{aligned}y_{it} &= \beta_i \bar{y}_t + \delta_i 1_{\{t>T_0\}} + e_{it} \\ &= \beta_i \bar{\mu} f_t + \beta_i \frac{N_I}{N}\bar{\Delta} 1_{\{t>T_0\}} + \delta_i 1_{\{t>T_0\}} + e_{it} + \beta_i \bar{\epsilon}_t \\ &= \beta_i \bar{\mu} f_t + \left(\beta_i \frac{N_I}{N}\bar{\Delta} + \delta_i\right) 1_{\{t>T_0\}} + (e_{it} + \beta_i \bar{\epsilon}_t).\end{aligned}$$

Comparing with the DGP (1), we see that

$$\begin{aligned}\mu_i &= \beta_i \bar{\mu}, \\ \Delta_i &= \delta_i + \beta_i \frac{N_I}{N}\bar{\Delta}.\end{aligned}$$

It is then clear that the individual treatment effect estimated from the regression on the overall averages is biased. The bias is given by $-\beta_i \frac{N_I}{N}\bar{\Delta}$, which vanishes if $N_I/N \to 0$.

Remark 3 *The DID-type framework (with N_C, N_I, T_0, T_1) is not compatible with the original Pesaran (2006) framework for a number of reasons. Adapted to our notation, his model is:*

$$\begin{aligned}y_{it} &= \gamma_i' d_t + \beta_i' x_{it} + \mu_i' f_t + \epsilon_{it}, \\ x_{it} &= \Gamma_i' d_t + M_i' f_t + v_{it},\end{aligned}$$

for $i = 1, ..., N$ and $t = 1, ..., T$.[9] *In his procedure, the first step is to form cross-sectional*

[9] The slope coefficients β_i follow a random coefficient model $\beta_i = \beta + \nu_i, \nu_i \sim IID(0, \Omega_\nu)$ (see Pesaran's Assumption 4 for more details).

averages $\bar{y}_t$ and $\bar{x}_t$. In the second step, for each i, regress y_{it} on x_{it}, $\bar{y}_t$ and $\bar{x}_t$ to obtain an estimate for β_i.

Our interaction term $1_{\{i\in I\}}1_{\{t>T_0\}}$ cannot be considered as x_{it} because x_{it} is determined by an idiosyncratic shock v_{it} which has positive variance across i and t (Pesaran's Assumption 2).[10] *Although we can express the interaction term as part of $\gamma_i' d_t$ in a parsimonious manner, the treatment effect will not be identified because γ_i is a nuisance parameter and is not estimated.*[11] *In addition, to avoid the issue in remark 2 we need to set the weight $w_i = 0$ for all treatment units $i \in I$. This violates Assumption 5(i) of Pesaran (2006) if the treatment group is a significant portion of all units.*

2.1.2 Choice of Weighting Matrix

How can we find the weighting matrix for a given data set? We propose obtaining W by applying principal component analysis (PCA) to the $N_C \times N_C$ square matrix $\frac{y_C y_C'}{T^{\bar{r}}}$, where $\bar{r} = \max(2r_1, ..., 2r_m, 1)$.[12] Provided that the observations in the control group have enough variability, so that the rank of the data matrix y_C of the control group outcomes is m, we can find m eigenvectors $w_1, \ldots, w_m$ of $\frac{y_C y_C'}{T^{\bar{r}}}$. The m eigenvectors are linearly independent, orthogonal to each other, and can be normalized so that all columns sum to one, in the sense that $\sum_{i=1}^{N_C} w_{ij} = 1$ for all $j = 1, \ldots, m$, and $w_j' w_k = 0$ for all $j \neq k$. The following proposition provides sufficient conditions such that the weighting matrix obtained from principal component analysis will satisfy rank conditions for identification:[13]

Proposition 1 *Suppose the $N_C \times T$ data matrix y_C is generated by equation (3), has rank*

[10] Even if we treat the interaction term as x_{it}, the estimation procedure will result in perfect multicollinearity. For instance, form the cross sectional average of the interaction term $\frac{1}{N}\sum_{i=1}^{N} 1_{\{i\in I\}}1_{\{t>T_0\}} = \frac{N_I}{N} 1_{\{t>T_0\}}$. For treatment unit $i \in I$, the regression is

$$y_{it} = \delta_i 1_{\{t>T_0\}} + \beta_i \bar{y}_t + \varphi_i \frac{N_I}{N} 1_{\{t>T_0\}} + e_{it},$$

which is subject to collinearity.

[11] For example, set $\gamma_i = [\Delta_1, ..., \Delta_{N_I}, 0, ..., 0]'$ and $d_t = 1_{\{t>T_0\}}$.

[12] The principal components remain identical regardless of the value of $\bar{r}$.

[13] The principal component method is not the only way to find the appropriate W. More generally, we may apply singular value decomposition. Identification is possible as long as one can find W that satisfies (5). Suppose Assumptions 2 and 3 hold, and that no linear combinations of the m factors are in the null space of μ_C. Then, by a standard result in linear algebra, we obtain $rank(\mu_C' F) = rank(F) - \dim\{N(\mu_C) \cap R(F)\} = m - 0 = m$ so that the $N_C \times N_C$ square matrix $\mu_C' F F' \mu_C$ is of rank m. There exists W such that $W' \mu_C' F F' \mu_C W$ is a diagonal matrix (whose diagonal entries are the singular values of the matrix $\mu_C' F F' \mu_C$) and is invertible. Following the arguments in the proof of the Proposition, we see that (5) is satisfied.

p, and Assumptions 2 and 5 hold. Let $\{w_1, \ldots, w_p\}$ be a set of p orthogonal eigenvectors obtained from principal component analysis on $\frac{y_C y_C'}{T^r}$. Let $W = [w_1, \ldots, w_p]$ be the $N_C \times p$ weighting matrix that satisfies Assumption 4. Then, as $N_C \to \infty$, one of the following results hold:

(a) When $p \geq m$, both Assumption 3 and the rank condition (5) hold, i.e., $rank(\mu_C) = m$ and $rank(\mu_C W) = m$;

(b) When $p < m$, there exists an $m \times p$ matrix R with orthogonal columns such that the rank condition (10) holds, i.e., $rank(R'\mu_C W) = p$.

Note that our principal component analysis is carried out on an $N_C \times N_C$ covariance matrix of control group observations. The principal components are treated as proxies for the linear combination of individual factors and factor loadings. By contrast, the principal component analysis in Bai (2009) is carried out on a $T \times T$ matrix, and the factors are identified under large T and N and appropriate normalizations (e.g., $FF'/T = I_m$, and $\mu\mu'$ being diagonal, where μ is an $m \times N$ matrix of factor loadings of all units); however, the normalization needs to be adjusted if some factors are nonstationary. The two estimators have different asymptotic properties and finite sample performance.[14]

2.2 Least Squares Estimation

Consider the regression for treatment group unit $i \in I$ (equation (9)):[15]

$$y_{it} = \alpha_i' W' y_{Ct} + \delta_i 1_{\{t > T_0\}} + e_{it}.$$

Let 1_{post} be the $T \times 1$ vector of post-intervention period indicators, consisting of T_0 zeros followed by T_1 ones. Denote $\bar{y}_C := y_C' W = [y_{C1}' W, \ldots, y_{CT}' W]$, a $T \times m$ matrix consisting of weighted averages of control group outcomes, and let $M_{\bar{y}_C} = I_{T \times T} - \bar{y}_C(\bar{y}_C'\bar{y}_C)^{-1}\bar{y}_C'$ be its orthogonal projection matrix. Also let $y_i = [y_{i1}, \ldots, y_{iT}]'$ and $e_i = [e_{i1}, \ldots, e_{iT}]'$. We may

[14] Our approach of extracting principal components from an $N_C \times N_C$ matrix offers computational convenience over Bai's (2009) method when $T > N_C$ (Stock and Watson (2002)).

[15] As described in earlier sections, the analysis and results still apply if we include deterministic factors to the model. In particular, an intercept can be added to the regression equation, capturing the fixed effect in levels.

thus represent the time series regression (9) in vector form,

$$y_i = \bar{y}_C \alpha_i + \delta_i 1_{post} + e_i. \tag{11}$$

The main object of study is the least squares estimator of the individual treatment effect δ_i, given by

$$\hat{\delta}_i = (1'_{post} M_{\bar{y}_C} 1_{post})^{-1} 1'_{post} M_{\bar{y}_C} y_i. \tag{12}$$

In relation to Pesaran (2006)'s correlated common effects (CCE) estimators, we refer to this estimator as the *CCE-DID* estimator. When we use principal component analysis instead of cross-sectional averages to form factor proxies, we refer to the estimator as the *CCEPC-DID* estimator.

2.3 Asymptotic distribution of the least squares estimator

We first discuss the asymptotic properties of the least squares estimator $\hat{\delta}_i$ when some of the factors are nonstationary. Then we will tackle the case when all factors are stationary.

2.3.1 Nonstationary case

Suppose $r_{\max} = \max(r_1, \ldots, r_m) > 0$. Let us define the $m \times m$ diagonal matrices:

$$\Upsilon = \begin{bmatrix} T^{r_1 \vee 0.5} & & & O \\ & T^{r_2 \vee 0.5} & & \\ & & \ddots & \\ O & & & T^{r_m \vee 0.5} \end{bmatrix},$$

$$\Upsilon_1 = \begin{bmatrix} T_1^{r_1 \vee 0.5} & & & O \\ & T_1^{r_2 \vee 0.5} & & \\ & & \ddots & \\ O & & & T_1^{r_m \vee 0.5} \end{bmatrix}.$$

Here $a \vee b = \max(a,b)$. Note that Υ is invertible. Define the limit

$$\Pi_r = \lim \Upsilon_1 \Upsilon^{-1} = \begin{bmatrix} \kappa^{r_1 \vee 0.5} & & & O \\ & \kappa^{r_2 \vee 0.5} & & \\ & & \ddots & \\ O & & & \kappa^{r_m \vee 0.5} \end{bmatrix}, \tag{13}$$

where $\kappa = \lim T_1/T$ as $T_1, T \to \infty$. Let us introduce the following probability limits, all of which contain elements that are $O_p(1)$ (see the Supplementary Results):

$$s_F' = \text{plim}\, \frac{1_{post}' F' \Upsilon_1^{-1}}{\sqrt{T_1}}, \tag{14}$$

$$s_{FF'} = \text{plim}\, \Upsilon^{-1} F F' \Upsilon^{-1}, \tag{15}$$

$$s_{Fe_i} = \text{plim}\, \Upsilon^{-1} F e_i, \tag{16}$$

$$\sigma_{e_i} = \lim \frac{1}{T_1} \sum_{s,t \in T^{post}} Cov(e_{is}, e_{it}). \tag{17}$$

Note that $s_{FF'}$ is invertible by Assumption 2(iii).

We have the following results.

Theorem 1 *Suppose $r_{\max} = \max(r_1, \ldots, r_m) > 0$ and that Assumptions 1-5 hold. Then, as $T_1, T \to \infty$ and $T_1/T \to \kappa \in [0,1]$, we obtain*

$$\sqrt{T_1}(\hat{\delta}_i - \delta_i) \xrightarrow{d} \frac{\sigma_{e_i} Z - s_F' \Pi_r s_{FF'}^{-1} \Pi_r s_{Fe_i}}{1 - s_F' \Pi_r s_{FF'}^{-1} \Pi_r s_F}, \tag{18}$$

where Z is a standard normal random variable, and the probability limits are defined in (13)-(17).

Corollary 1 *Under the same conditions as in the above theorem, we obtain*

$$\hat{\delta}_i - \delta_i \xrightarrow{a.s.} 0.$$

Remark 4 *When $\kappa = 0$, the limiting distribution of $\hat{\delta}_i$ is univariate normal regardless of the integration order(s), due to $\Pi_r = 0$. When $\kappa \neq 0$, the estimator $\hat{\delta}_i$ does* not *typically follow an asymptotically normal distribution.*

Remark 5 *The measurement error of factor proxies does not affect the limiting distribution of* $\hat{\delta}_i$ even under finite N_C*, as it is dominated in terms of stochastic order (see the discussion before (28) in the proof). Furthermore, for any given* N_C*, we may neglect the sample variations of the weighting matrix* W *obtained from principal component analysis on* $y_C y_C'/T^{\bar{r}}$*, as* $T_1, T \to \infty$ *(see section 6.3.4).*

2.3.2 Stationary case

Now consider the case when all m factors are stationary. The following theorem provides the asymptotic distribution of $\hat{\delta}_i$.

Theorem 2 *Suppose that all* $r_1 = \ldots = r_m = 0$*, and that Assumptions 1-5 hold. Let* $T_1, T \to \infty$ *and* $T_1/T \to \kappa \in [0,1]$*. Then we have the following results:*

(a) If $N_C \to \infty$*, then*

$$\sqrt{T_1}(\hat{\delta}_i - \delta_i) \xrightarrow{d} \frac{\sigma_{e_i} Z}{1 - \kappa E(f_t') E(f_t f_t')^{-1} E(f_t)},$$

(b) If $E(f_t) = 0$*, then*

$$\sqrt{T_1}(\hat{\delta}_i - \delta_i) \xrightarrow{d} \sigma_{e_i} Z.$$

Here Z *is a standard normal random variable, and* σ_{e_i} *is defined in (17).*

Corollary 2 *Under the same conditions as in the above theorem, we obtain*

$$\hat{\delta}_i - \delta_i \xrightarrow{a.s.} 0.$$

Remark 6 *Unlike the nonstationary case, we further need either* $N_C \to \infty$ *or* $E(f_t) = 0$ *to achieve consistency of* $\hat{\delta}_i$ *in the stationary case. They are necessary to eliminate the endogeneity bias due to measurement error of the factor proxies. However, just like the nonstationary case, the sample variability of the weighting matrix* W *does not affect the asymptotic distribution of* $\hat{\delta}_i$ *as* $T_1, T \to \infty$ *for any given* N_C*.*

Remark 7 *It is feasible to estimate the variance of* $\hat{\delta}_i$ *analytically in the stationary case.*

The asymptotic variance of $\hat{\delta}_i$ can be consistently estimated by

$$\widehat{Var}(\hat{\delta}_i) = \frac{1_{post}' M_{\bar{y}_C} \hat{\Sigma}_{e_i} M_{\bar{y}_C} 1_{post}}{(1_{post}' M_{\bar{y}_C} 1_{post})^2},$$

where $\hat{\Sigma}_{e_i}$ is the $T \times T$ sample autocovariance matrix of the regression residuals $\hat{e}_{i1}, \ldots, \hat{e}_{iT}$, where $\hat{e}_{it}$ is defined by $\hat{e}_{it} = y_{it} - \hat{\alpha}_i' W' y_{Ct} - \hat{\delta}_i 1_{\{t>T_0\}}$. If, for each i, the regression errors e_{it} form an iid sequence, then $\widehat{Var}(\hat{\delta}_i)$ is simplified to $s_{e_i}^2 / 1_{post}' M_{\bar{y}_C} 1_{post}$, where $s_{e_i}^2$ is the sample variance of the residuals, given by $s_{e_i}^2 = \frac{1}{T-1} \sum_{t=1}^{T} \hat{e}_{it}^2$.

Summarizing the results above, a nice feature of our estimator $\hat{\delta}_i$ is that it is $\sqrt{T_1}$ consistent regardless of the number of unobserved factors, the integration order(s), and whether or not there exists cointegration relationship among the factors. All asymptotic results hold when T_0 and N_I are finite or infinite. The results also hold under finite N_C if one of the following conditions are satisfied: (i) some factors are nonstationary; (ii) all factors are stationary and have zero mean.

3 Small Sample Properties of Estimators via Simulations

We compare the small sample properties of our estimator with existing methods. The data generating process (DGP) that we consider is

$$y_{it} = \delta_i \mathbf{1}\{i \in I\} \mathbf{1}\{t > T_0\} + \mu_i^{'} f_t + \epsilon_{it}, \tag{19}$$

with the treatment effect $\delta_i \sim N(\bar{\delta}, \sigma_\delta^2)$ and idiosyncratic error $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$. A variety of factor structures are examined. Table I considers stationary factors. The first DGP involves a single AR(1) factor:

$$f_{1t} = \rho_1 f_{1t-1} + u_{1t}, \tag{20}$$

where $u_{1t} \sim N(0, \sigma_{u1}^2)$. The second DGP involves three AR(1) factors:

$$f_{jt} = \rho_j f_{jt-1} + u_{jt}, \tag{21}$$

where $u_{jt} \sim N(0, \sigma_{uj}^2)$ for $j = 1, 2, 3$. We set $\bar{\delta} = 3$ and $\sigma_\delta = \sigma_\epsilon = 0.3$. In the one-factor case, we consider $\rho_1 = 0.5, \sigma_{u1} = 0.3$. In the three-factor case, $(\rho_j, \sigma_{uj}) = (0.5, 0.3)$, $(0.7, 0.5)$, $(0.9, 0.3)$, for $j = 1, 2, 3$, respectively.

Table II considers nonstationary factors. The first DGP involves three independent I(1) factors:

$$f_{jt} = f_{jt-1} + u_{jt}, \tag{22}$$

where $u_{jt} \sim N(0, \sigma_{uj}^2)$ and $\sigma_{uj} = 0.3$, 0.5, 0.1, for $j = 1, 2, 3$, respectively. The second DGP considers a model with three independent factors – one I(1), one I(2), and one I(3) factor.[16]

In all DGPs, the factor loadings among control units ($i \in C$) for the jth factor are $\mu_{ij} \sim N(\bar{\mu}_{Cj}, \sigma_{\mu_C}^2)$. The factor loadings among treatment units ($i \in I$) for the jth factor are $\mu_{ij} \sim N(\bar{\mu}_{Ij}, \sigma_{\mu_I}^2)$. We normalize $\bar{\mu}_{Cj} = 1$ for all j, and set $\sigma_{\mu_C} = 0.1$, $\sigma_{\mu_I} = 0.3$. In single-factor DGPs, $\bar{\mu}_{I1} = 1.5$. In three-factor DGPs, $\bar{\mu}_{Ij} = 1.3$, 1.5, 1.7 for $j = 1, 2, 3$, respectively.

We consider the following sizes of control and treatment units: $(N_C, N_I) = (5, 5)$, $(10, 10)$, $(25, 25)$, and $(50, 50)$. We also consider the following sizes of pre-intervention and post-intervention time periods: $(T_0, T_1) = (5, 5)$, $(10, 10)$, $(25, 25)$, and $(50, 50)$. Thus, the total number of observations ranges from $(5+5) \times (5+5) = 100$ to $(50+50) \times (50+50) = 10,000$.

Results from six estimators are reported: (1) the CCEPC-DID estimator based on three principal components; (2) the CCEPC-DID estimator based on one principal component; (3) the CCE-DID estimator, which uses equal weights among control units; (4) the Hsiao, Ching and Wan (2012) estimator (HCW thereafter), which constructs a "synthetic" treatment unit by regressing the outcome of a treatment unit on the outcomes of control units during the pre-intervention period; (5) the Bai (2009) estimator, which iteratively solves for factors and the treatment effect; (6) the standard difference-in-differences estimator.[17] In all cases, the mean bias and empirical standard deviation of the average treatment effect among treatment units are reported.

[16] In Appendix table A1, two extra cases are considered: (i) one AR(1) factor with $\rho_1 = 0.9$ and $\sigma_{u1} = 0.3$; (ii) one I(1) factor with $\sigma_{u1} = 0.3$.

[17] We do not report results from Abadie, Diamond and Hainmueller (2010)'s synthetic control estimator, as its small sample performance is dominated by the HCW estimator. For more discussions on its small sample properties, see Gobillon and Magnac (2016).

Table I compares the performance of the above estimators under the case of stationary factors. The left panel reports results from the one-factor model ($\rho = 0.5$). The mean bias is generally small among all estimators. Interestingly, the CCE-DID estimator, which uses equal weights among control units, has the lowest empirical SD among all estimators. The CCEPC-DID estimator comes second in performance, and the extra principal components have a small effect on the empirical SD. The HCW and DID estimators tie for the third place in terms of performance. The HCW estimator performs best when the number of pre-intervention periods (T_0) far exceeds the number of control units (N_C). When $T_0 < N_C$, the estimator can only be implemented by first choosing a subset of control units via a model selection procedure that they propose (AIC or AICC). The Bai (2009) estimator has the largest empirical SD.[18] There is also non-negligible bias when T is small. In addition, unlike the other estimators (esp. CCE), its performance improves with the size of T but not N.

The right panel reports results from the three-factor model ($\rho = 0.5, 0.7, 0.9$). The CCE-DID estimator remains the best performer, followed by the CCEPC-DID and HCW estimators. Despite the multi-factor structure, additional principal components do not lead to a lower empirical SD, as the most important linear combination of factors has been picked up by the first principal component. Relative to the above estimators, the Bai (2009) and DID estimators experience a large deterioration in performance under this factor structure.

Table II compares the performance of the estimators under the case of nonstationary factors. The left panel reports results from the model with three I(1) factors. There are notable differences from Table I. The CCEPC-DID estimator with three principal components becomes the best performer. Its empirical SD improves as T increases; by contrast, all the other estimators have worse empirical SD as T increases, due to inadequate incorporation of the nonstationary factors. The second-best performer is the CCEPC-DID estimator with one principal component, which outperforms the CCE-DID estimator. Surprisingly, the HCW estimator also performs well, although its asymptotic distribution is unknown under nonstationarity. The Bai (2009) and DID estimators have the worst performance, especially under large T.

The right panel reports results from the model with three factors of different integration

[18] When factors are i.i.d. normal, we find that Bai (2009) has a similar empirical SD as the other estimators. Gobillon and Magnac (2016) find that Bai (2009) and DID have similar performance when factors are i.i.d. uniform, and Bai (2009) performs better than DID under seasonal (sinusoidal) factors.

order (I(1), I(2) and I(3)). The CCEPC-DID estimator with three principal components remains the best performer, and it is robust across different sample sizes. In fact, it is the only estimator that keeps the empirical SD at reasonable levels. Generally, the other estimators are not practically useful due to large empirical SDs. One exception is when both N_C and N_I are large and $(T_0, T_1) = (5, 5)$. Under such cases, the CCEPC-DID (1PC) and CCE-DID estimators still generate reasonable empirical SDs.

Table III considers different sample balances between control and treatment units (N_C, N_I) $= (10, 40)$, $(40, 10)$, $(49, 1)$, as well as between pre-intervention and post-intervention periods $(T_0, T_1) = (10, 40)$, $(40, 10)$. The sample balance may affect the properties of the estimators. For instance, the asymptotic distribution of the CCEPC-DID and CCE-DID estimators depends on the ratio $\frac{T_1}{T_0}$, and the construction of factor proxies in finite samples depends on the size of N_C. The results suggest that both estimators are robust across different sample balances. The empirical SD is similar when $(T_0, T_1) = (10, 40)$ or $(40, 10)$. The empirical SD tends to be smaller when the number of control units is relatively large ($(N_C, N_I) = (40, 10)$), but becomes larger when the imbalance is extreme. The Bai (2009) estimator tends to favor a relatively large number of control units and post-intervention time periods. The HCW and DID estimators have similar performances under different sample balances.

4 Empirical Applications

4.1 Application 1: GDP Growth

We first apply our estimators to the data in Hsiao, Ching and Wan (2012), who study the effect of political and economic integration of Hong Kong with mainland China on the GDP growth of Hong Kong. Their original data set contains quarterly real GDP growth rates (year-on-year (YoY)) from 1993Q1 to 2008Q1 of 25 countries including Hong Kong. The sovereignty of Hong Kong was transferred from the UK to China on 1 July 1997. In the analysis of political integration, they use data from 1993Q1 to 2003Q4, which consists of 18 pre-intervention periods (1993Q1 - 1997Q2) and 26 post-intervention periods (1997Q3 - 2003Q4). Due to the relatively small number of pre-intervention periods, they use 10 countries from the original sample to form the pool of control units – China, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, Taiwan, Thailand, and the US. These countries are chosen due to their geographical or economic proximity to Hong Kong.

In the analysis of economic integration, they focus on the impact of the Closer Economic Partnership Arrangement (CEPA) between Hong Kong and mainland China, which started in January 2004. CEPA included a series of measures that strengthened the economic link between both regions in terms of reduction or elimination of tariffs and non-tariff barriers on trade in goods, liberalization of trade in services, and trade and investment facilitation. They use the full sample of countries from 1993Q1 to 2008Q1, which consists of 44 pre-intervention periods (1993Q1 - 2003Q4) and 17 post-intervention periods (2004Q1 - 2008Q1).[19]

Table IV compares the treatment effect estimates obtained from different methods.[20] The left panel reports the estimated effect of political integration, which provides mixed results. The CCEPC-DID estimates are around 0.01 when 1, 2, or 3 principal components are used, and they are statistically insignificant at the 10 percent level. When 5 principal components are used, the coefficient becomes 0.021 and statistically significant at the 5 percent level. The CCE-DID estimate is also positive at 0.01, but statistically insignificant. These results suggest that the transfer of sovereignty *increased* the YoY GDP growth of Hong Kong by 1 to 2 percentage points, although statistical significance depends on the number of principal components used.

We compute the HCW estimate using all countries from the control group donor pool. The estimated treatment effect has a mean of -0.036, with a SD of 0.089.[21] This suggests that the transfer of sovereignty reduced Hong Kong's GDP growth (although the SD is large). The Bai (2009) estimate is negative at -0.029 and statistically significant at the 1 percent level. By contrast, the conventional DID estimate is 0.001 and statistically insignificant.

The right panel reports the estimated effect of economic integration, which is more coherent across different methods. The CCEPC-DID and CCE-DID estimates range from 2.2 to 4.0 percentage points and are all statistically significant at the 1 percent level. The HCW

[19] In the economic integration sample, we discard Norway because its growth trajectory is very close to Hong Kong (closer than any other country in the sample). We suspect that this is unadjusted Hong Kong data. Hence we use 23 control countries in this sample.

[20] For the HCW (or synthetic control) method, we report the sample standard deviation (SD) of the treatment effect estimate over the post-intervention periods. For other methods, we compute the standard error by bootstrapping.

[21] The original estimate in their paper uses a model selection procedure to select a subset of countries from the control group donor pool. Using AIC, their estimate was -0.0403 with a SD of 0.0815.

estimate is 0.024, with a SD of 0.024.[22] The Bai (2009) and DID estimates are 0.035 and 0.032, respectively, and are both statistically significant at the 1 percent level. These results provide strong evidence that CEPA increased Hong Kong's GDP growth.

Figure 1 compares Hong Kong's actual GDP growth with predictions from the HCW and CCEPC-DID estimators. Figure 1(a) compares the results in political integration data. The actual growth was characterized by a large negative spike between 1998 and 2000. This was caused by the Asian Financial Crisis – shortly after the change of sovereignty in July 1997, the Crisis broke out in the second half of 1997, which resulted in severe economic downturn in the region.

The graph also shows markedly different predictions between HCW and CCEPC-DID estimators. The HCW estimator constructs a "synthetic" Hong Kong by regressing Hong Kong's GDP growth on the GDP growth of control group countries during the *pre-intervention* period. The slope coefficients are interpreted as weights of the control group countries. Assuming the weights are time-invariant, the counterfactual predicted growth of Hong Kong during the post-intervention period is constructed (extrapolated). The treatment effect is obtained from the average difference between counterfactual and actual growth during the post-intervention period. As the graph shows, the predictions match the data very well during the pre-intervention period. However, predicted growth continues to go up in 1998, while in reality Hong Kong's growth suffers a heavy blow from the Asian Financial Crisis. During the recovery between 2000 and 2003, the predicted and actual growth become close again. Because the predicted (counterfactual) growth is far above actual growth in 1998 and 1999, it explains why the estimated treatment effect was negative.

The CCEPC-DID estimator, on the other hand, uses *all* time periods of control group countries to construct factor proxies, which are in the form of weighted control units. Because information from all time periods are used, the model is able to pick up the most important factors during the entire period (not just the pre-intervention period), namely, systematic changes that occur during the Crisis in 1998-1999. As a result, predictions during the post-intervention period match the data very well, albeit at the cost of worse goodness-of-fit during the pre-intervention period. In addition, because the post-intervention period is within the estimation sample, the model generates two predictions during this period: (1) predicted

[22] Using AIC, their estimate was 0.0379 with a SD of 0.0151.

growth when there is a treatment effect; (2) counterfactual growth when there is no treatment effect. The former allows us to better assess the goodness-of-fit of the model to the data. These patterns in the graph explain why we obtained a small positive treatment effect from the CCEPC-DID estimator.

Figure 1(b) compares the results in economic integration data. In this data, because the Crisis occurred during the pre-intervention period, and the post-intervention period was relatively stable, the HCW estimator generates post-intervention predictions that are more in line with predictions from the CCEPC-DID estimator. This also explains why the treatment effect estimates are coherent across different methods in this data.

Table V reports the country weights for the first two principal components computed from the CCEPC-DID estimator. Each principal component consists of weighted control group countries that reflect (linear combinations of) the unobserved factors. In political integration data, the first principal component assigns the largest weight to China, Indonesia and Malaysia (0.151 each), followed by Singapore (0.127), Korea (0.121), Taiwan (0.087), Thailand (0.086), Philippines (0.083), US (0.036) and Japan (0.009). These weights correspond closely to the pattern of geographical and economic relationships in the region. They also reflect the impact of the Crisis on the most severely affected countries – Indonesia, Korea, Malaysia, Philippines, and Thailand. The second principal component assigns the largest positive weight to China and the largest negative weight to Indonesia. The heavy weight on Indonesia is due to the fact that it was most severely affected by the Crisis – the Indonesian rupiah depreciated against USD by over 80 percent from 1997 to 1998, followed by the fall of a three-decade long presidency by Suharto (and subsequent independence of East Timor). The heavy positive weight on China is probably due to its resilience to the Crisis and its unique development scenario relative to other countries in the region.

The country weights computed from the HCW estimator are different. In political integration data, the weights are: China (-0.018), Indonesia (-0.054), Japan (-0.619), Korea (-0.401), Malaysia (-0.050), Philippines (-0.115), Singapore (0.021), Taiwan (1.060), Thailand (-0.041), US (0.570).[23] Taiwan, Japan and the US have the largest absolute weights, while mainland China, which has a close economic relationship with Hong Kong, has the

[23] In contrast to Abadie, Diamond and Hainmueller (2010), their weights for control group units are not constrained between 0 and 1, and they do not need to sum to 1. When AIC is used, the weights in their paper were Japan (-0.69), Korea (-0.3767), US (0.8099), Philippines (-0.1624), and Taiwan (0.6189).

smallest absolute weight. Thus, it is more difficult to attach a direct economic interpretation to these weights – the weights are only chosen to mimic Hong Kong's GDP growth as closely as possible during the pre-intervention period.

In economic integration data, the pool of control countries is larger and the sample period is longer, hence the country weights can be different. Nevertheless, the first principal component assigns weights most heavily to Malaysia (0.106), China (0.103), Indonesia (0.095), Singapore (0.086) and Korea (0.067). The five countries with the smallest weights are Germany (0.007), Japan (0.008), Switzerland (0.013), Italy (0.014), and France (0.016). Interestingly, Anglo-Saxon countries have large weights among the Western countries: Australia (0.035), Canada (0.037), New Zealand (0.039), UK (0.034) and US (0.024). The second principal component gives the largest absolute weights to Indonesia (-0.448), Thailand (-0.338) and China (0.337). By contrast, in the HCW estimator, the five countries with the largest absolute weights are Austria (-0.9198), Germany (-0.7676), Netherlands (-0.6343), Finland (-0.6835), and Mexico (0.5773). The five countries with the smallest absolute weights are UK (0.000), Taiwan (0.002), Switzerland (0.006), France (0.043) and China (0.064).[24]

Figure 2 provides further evidence by plotting the principal components (i.e., weighted GDP growth of control group countries) over time. The patterns confirm the economic interpretation of the country weights discussed above. In political integration data (Figure 2(a)), the first principal component moves downward in 1998, then recovers in 1999 and 2000. The second principal component has a large upward spike in 1998-1999, which reflects the disproportionate impact on Indonesia during the Crisis (note that Indonesia's weight is negative). In economic integration data (Figure 2(b)), the patterns of the first two principal components are similar to Figure 2(a), while the third principal component exhibits wide swings in 1997-2001.

Finally, Table VI reports the proportion of variance of control group countries' GDP growth that can be explained by the principal components. In both political and economic integration data, around 99 percent of the variance can be explained by the first three principal components.

[24]When AIC is used, the weights in their paper were Austria (-1.2949), Germany (0.3552), Italy (-0.5768), Korea (0.3016), Mexico (0.234), Norway (0.2881), Switzerland (0.2436), Singapore (0.2222), and Philippines (0.1757).

4.2 Application 2: Cigarette Sales

We then apply our estimators to the data in Abadie, Diamond and Hainmueller (2010), who study the effect of a large-scale tobacco control program ("Proposition 99") on levels of California's cigarette sales. Their data set contains per-capita cigarette sales (packs) in 39 states including California, from 1970 to 2000. As an initiative statute in California, Proposition 99 was the first modern-time large-scale tobacco control program in the US. It was passed in November 1988 and went into effect in January 1989. The key element was a 25-cent increase of per pack state excise tax on the sale of tobacco cigarettes within California, with similar tax on other tobacco products. Revenue generated by the act was earmarked for various environmental and health care programs, and anti-tobacco advertisements. In the analysis, they define 1970-1988 as pre-intervention periods (19 years), and 1989-2000 as post-intervention periods (12 years). The control group consists of 38 states that did not adopt large-scale tobacco control programs between 1970 and 2000.

Table VII compares the treatment effect estimates obtained from different methods. The left panel reports the estimated effect of the program on the *level* of per-capita cigarette sales in California. This is the outcome variable analyzed by Abadie, Diamond and Hainmueller (2010). With one principal component, the CCEPC-DID estimate is -19.97 and statistically significant at the 1 percent level. However, the estimate becomes closer to zero when more principal components are used. When 10 principal components are used, the estimate is -5.39 and statistically significant at the 5 percent level. The other methods produce large negative estimates that are statistically significant at the 1 percent level. The CCE-DID estimate, which is based on equal weights among control states, is -20.62. The estimates obtained by synthetic control (Abadie, Diamond and Hainmueller (2010)), Bai (2009), and DID methods are -19.48, -57.23, and -27.35, respectively.[25]

As the simulation results in the Section 3 showed, sensitivity of CCEPC-DID estimates to the number of principal components may be an indication that nonstationary factors are present. Indeed, the graphs of the time series of cigarette sales in most states exhibit

[25] Similar in spirit to the HCW method, we use levels of per-capita cigarette sales in 1970, 1971, ..., 1988 as matching variables. In Abadie, Diamond and Hainmueller (2010), there are seven matching variables: per-capita cigarette sales in 1975, 1980, and 1988, and averages of Ln(GDP per capita), percent of the population aged 15-24, retail cigarette price, and per-capita beer consumption. When the above matching variables are used, the estimated treatment effect is -17.96 with a SD of 5.80.

substantial persistent long-run movements. To formally test for nonstationarity in cigarette sales data, we conduct the Breitung-Das (2005) panel unit root test on two subpanels as defined by US Census Regions: (i) West or Mid-West (20 states); (ii) South or Northeast (19 states).[26] In both panels, the test fails to reject the null hypothesis that there is unit root in a subset of states in the data. The p-values are 0.6751 and 0.9725, respectively. This suggests that stochastic trends are a prevalent feature in the data.

Nonstationarity suggests that the CCEPC-DID estimator with multiple principal components should generate the most plausible treatment effect estimate. The other methods may have overestimated the effect of the tobacco control program by attributing stochastic trends to the treatment effect. This also explains why the treatment effect estimate diminishes quickly when more principal components are used – multiple principal components are required to capture all the nonstationary factors, as the simulation results have shown.

We also report estimates from detrended cigarette sales data. The detrended data is constructed from subtracting per-capita cigarette sales by its cross-sectional mean. After detrending, the CCEPC-DID estimates become closer to zero when as few as two principal components are used. Nevertheless, with 10 principal components, the estimate is -4.69 and ramain statistically significant at the 5 percent level. The CCE-DID estimator becomes infeasible due to detrending, while the synthetic control and DID estimates are invariant to detrending. The Bai (2009) estimate becomes closer to zero at -39.73.

The right panel reports the estimated effect of the program on the *yearly change* of per-capita cigarette sales in California. The first-order difference removes nonstationarity from the series. In addition, this can be a reasonable specification if the price of cigarettes adjusts slowly to the tax increase, or anti-tobacco advertisements take time to generate an impact. The estimates are more coherent across different methods. The CCEPC-DID estimates lie between -1.69 to -1.90 (all statistically significant at the 1 percent level) when 1 to 5 principal components are used. At 10 principal components, the estimate is -0.74 and statistically insignificant. The CCE-DID, synthetic control, Bai and DID estimates are -1.19, -1.94, -3.01 and -0.75, respectively. Except for synthetic control, the other estimates are statistically significant at the 1 percent level. When the yearly change data are detrended, the CCEPC-DID estimates become closer to zero (ranging between -0.43 and -0.89) and the same is true

[26]The test allows for cross-sectional dependence and is based on $(T, N) \rightarrow_{seq} \infty$.

for the Bai estimate (-2.16). the synthetic control and DID estimates remain unchanged.

Comparing the estimation results from level and yearly change data, the CCEPC-DID method with multiple principal components yields the most coherent estimates between both data sets. For instance, consider the estimate of -0.74 (10 PCs) in yearly change data. This implies that during the 12 years of post-intervention period, the program changed per-capita cigarette sales by an average level of $-0.74 \times (1+2+...+12) \div 12 = -4.81$ packs. This is close to the estimate of -5.39 (10 PCs) in level data. By contrast, other methods yield less coherent estimates. The largest discrepancy comes from DID; from the estimate in yearly change data (-0.75), the implied change in average level is $-0.75 \times (1 + 2 + ... + 12) \div 12 = -4.88$ packs, which is substantially lower than the estimate of -27.35 in level data.

Figure 3 compares California's actual level of per-capita cigarette sales with predictions from synthetic control and CCEPC-DID methods. Actual per-capita cigarette sales were relatively stable at above 120 packs before 1980, and reduced substantially in the next 20 years to 40 packs in 2000. The rate of decline accelerated around 1988. The synthetic control method yields an excellent fit to the data during the pre-intervention period, and "synthetic California" (no treatment) is predicted to have around 80 packs of per-capita cigarette sales during the post-intervention period. When one principal component is used, predictions from the CCEPC-DID method fit the data reasonably well during the pre-intervention period. After intervention, the method generates both a with-treatment prediction and a counterfactual prediction (i.e., no treatment). The counterfactual prediction is very close to the prediction from synthetic control, while the with-treatment prediction underpredicts cigarette sales between 1989 and 1993 and overpredicts afterwards. When 10 principal components are used, predictions from the CCEPC-DID method are quite different. First, it has a close fit to the data in all periods. Second, the counterfactual prediction remains only around 5 packs above the actual value throughout the post-intervention period. Figures 4(a) and (b) compare California's actual *yearly change* of per-capita cigarette sales with predictions from synthetic control and CCEPC-DID methods. Both synthetic control and CCEPC-DID methods predict the pre-intervention outcomes reasonably well. The predictions from CCEPC-DID also do not change substantially when more principal components are used.

Table VIII report the state weights in the first two principal components from the CCEPC-DID estimator. If all states have equal weights, the weight should be $1 \div 38 = 0.0263$. In level data, the standard deviation of state weights in first principal component is merely 0.006. The first principal component assigns the largest weights to New Hampshire (0.048), Kentucky (0.041) and North Carolina (0.037). Interestingly, these states rank the top three in average per-capita cigarette sales during the sample period. In the data, the overall average level is 118.89 packs; the above states have an average level of 213, 187 and 164 packs, respectively. The lowest weights are assigned to Utah (0.014), New Mexico (0.019) and North Dakota (0.022). In the data, these states rank the bottom three in average per-capita cigarette sales (63, 84 and 98 packs, respectively). Therefore, the first principal component can be interpreted as the national trend of cigarette sales. The second principal component, on the other hand, assigns large positive weights to a number of adjacent states in the southern region (Alabama, Arkansas, Kentucky, Mississippi, South Carolina, Tennessee, West Virginia). These states were less subject to the national reduction in cigarette sales that started in the mid-1980s. This regional trend is picked up by the second principal component.

In yearly change data, the first principal component assigns the largest weights to New Hampshire (0.113), North Carolina (0.089) and Kentucky (0.073). Again, larger weights are assigned to states that have higher levels of per-capita cigarette sales, which also tend to have larger yearly changes in sales. The second principal component shows a different picture. Some adjacent states have different *signs* of weights – in particular, New Hampshire and Rhode Island (-0.917 vs 0.917), as well as Nevada, Idaho and Wyoming (-0.505 vs -0.453 vs 0.814). This reflects cross-border purchases – individuals can go to a neighboring state to purchase cigarettes, resulting in opposite yearly changes in cigarette sales between two adjacent states.

In contrast to the principal components above, the state weights from the synthetic control method have a less direct interpretation, although all weights are restricted between 0 and 1 and they sum to 1. For instance, in level data, the largest weights are assigned to Utah (0.394), Montana (0.232) and Nevada (0.205), and 32 states have a weight of 0.[27] The weights are chosen such that the weighted control units best mimic California during the pre-intervention period. In yearly change data, the weights are spread out among more states. The largest weights are assigned to Connecticut (0.173), Nevada (0.161) and Utah (0.14), and 28 states

[27] When the matching variables in Abadie, Diamond and Hainmueller (2010) are used, the state weights for synthetic California are Colorado (0.164), Connecticut (0.069), Montana (0.199), Nevada (0.234), Utah (0.334), other states zero.

have a weight of 0.

Figures 5(a) and (b) plot the principal components over time. In level data (Figure 5(a)), the first principal component, which represents the national trend, increases between 1970 and 1977, then declines afterwards. The second principal component, which represents the trend in the Southern region, increases from the 1970s and becomes steady starting from the mid-1990s. The third and fourth principal components also exhibit long-run tendencies. In yearly change data (Figure 5(b)), the first principal component captures the national trend – it is positive between 1970 and 1977, then becomes negative or close to zero in subsequent years. The second and third principal components tend to exhibit more short-term fluctuations.

Table IX reports the proportion of variance in cigarette sales of the control states that can be explained by the principal components. In level data, the first principal component explains 99.95 percent of the variance. The extremely high proportion provides further evidence that the principal component picks up stochastic trends, which dominate the stationary factors in the model. By contrast, in yearly change data, the first principal component explains 56.76 percent of the variance, and the first five principal components together explain 88 percent of the variance.

5 Conclusion

In this paper, we developed an estimator (CCE-DID / CCEPC-DID) for program evaluation when outcomes could be affected by unobserved common factors. The estimator was suitable for difference-in-differences (DID) style research designs. We derived consistency and asymptotic distributions under not only stationary factors but also nonstationary factors with any integration order. As such, our estimator broadens the scope of existing techniques such as Bai (2009) and synthetic control methods (Abadie, Diamond and Hainmueller (2010), Hsiao, Ching and Wan (2012)), which were limited to the case of stationary factors. Our estimator exhibited excellent small-sample performance and was robust across a variety of sample size and factor specifications, especially when stochastic trends were present. We applied our estimator to two data sets, one on GDP growth and another on cigarette sales. In the first application, our estimator found a weak positive effect of political integration on Hong Kong's GDP growth, after taking into account of the large shocks from the Asian Financial Crisis shortly after the policy intervention (i.e., return of sovereignty to China). In the second application, we found substantial evidence of stochastic trends in levels of cigarette sales. Our estimator revealed a smaller impact of California's tobacco control program on cigarette sales. By contrast, existing methods tended to attribute these stochastic trends to the treatment effect, resulting in overestimation of the effect of the program.

References

[1] 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 American Statistical Association, 105, 493–505.

[2] Bai, J. (2009): "Panel Data Models with Interactive Fixed Effects," Econometrica, 77, 1229–1279.

[3] Bai, J., C. Kao, and S. Ng (2009): "Panel Cointegration with Global Stochastic Trends," Journal of Econometrics, 149, 82-99.

[4] Breitung J. amd S. Das (2005): "Panel Unit Root Tests Under Cross-Sectional Dependence," Statistica Neerlandica, 59, 414-433.

[5] Gobillon L. and T. Magnac (2016): "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," Forthcoming in Review of Economics and Statistics.

[6] Gobillon L., T. Magnac, and H. Selod (2012): "Do Unemployed Workers Benefit From Enterprise Zones? The French Experience," Journal of Public Economics, 96, 881-892.

[7] Holly S., M.H. Pesaran, and T. Yamagata (2010): "A Spatio-Temporal Model of House Prices in the USA," Journal of Econometrics, 158, 160-173.

[8] 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.

[9] Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011): "Panels with Non-Stationary Multifactor Error Structures," Journal of Econometrics, 160, 326–348.

[10] 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.

[11] Moon, R.H. and M. Weidner (2015): "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," Econometrica, 83, 1543-1579.

[12] Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," Econometrica, 74, 967-1012.

[13] Stock, J. H. and M. W. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," Journal of American Statistical Association, 97, 1167–1179.

[14] Trefethen L. N., and D. Bau III (1997): Numerical Linear Algebra, SIAM, Philadelphia, USA.

6 Appendix

6.1 Identification under Assumption 3'

For our purpose, we will need the following linear algebra result on the reduced singular value decomposition of a rank-deficient rectangular matrix (e.g., p.28 of Trefethen and Bau, 1997).

Lemma 1 *(reduced SVD) Given an $m \times n$ matrix M that is of rank p, where $0 < p \leq \min(m, n)$, there exists an $m \times p$ matrix U, a $p \times p$ nonsingular matrix Σ, and an $n \times p$ matrix V, such that U and V are orthogonal and of full column rank, with $U'U = V'V = I$, and that $M = U\Sigma V'$. The latter implies that $\Sigma = U'MV$ is a diagonal, nonsingular matrix.*

Suppose the $m \times N_C$ matrix μ_C has rank $p \leq m$ (Assumption 3'(i)). By the singular value decomposition of the $m \times N_C$ matrix μ_C, there exists an $m \times p$ orthogonal matrix R (with $R'R = I_{p\times p}$), a $p \times p$ nonsingular diagonal matrix Σ, and an $N_C \times p$ orthogonal matrix $\tilde{W}$ (with $\tilde{W}'\tilde{W} = I_{p\times p}$) such that $\mu_C = R\Sigma\tilde{W}'$. This implies that $\Sigma = R'\mu_C\tilde{W}$ is a nonsingular diagonal matrix. Let S be a $p \times p$ diagonal matrix where the (k, k) element is the sum of the kth column of $\tilde{W}$. Note that S is nonsingular because $\tilde{W}$ has full column rank. Then the matrix defined by $W = \tilde{W}S^{-1}$ has columns that sum to one. Now, we may express $\mu_C = R\Sigma\tilde{W}' = R(\Sigma S)W' = R\tilde{\Sigma}W'$, where $\tilde{\Sigma} = \Sigma S$ is a nonsingular diagonal matrix.

Given that $rank(\mu_C) = p$, we can identify at most p linear combinations of the factors given by $R'F$, where R is determined by the left orthogonal matrix in the SVD of μ_C. We can correspondingly form p weighted averages of control group outcomes, where the weights are determined by the right orthogonal matrix in the SVD of μ_C. So we have

$$\begin{aligned}
W'y_{Ct} &= W'\mu_C' f_t + W'\epsilon_{Ct} \\
&= (S^{-1}\tilde{W}')(\tilde{W}\Sigma R')f_t + W'\epsilon_{Ct} \\
&= S^{-1}\Sigma R' f_t + W'\epsilon_{Ct}.
\end{aligned}$$

Since S and Σ are nonsingular and diagonal, it follows that $S^{-1}\Sigma = \Sigma S^{-1} = (R'\mu_C\tilde{W})S^{-1} = R'\mu_C W$ is invertible and so we obtain

$$\begin{aligned}
R'f_t &= (S^{-1}\Sigma)^{-1}W'(y_{Ct} - \epsilon_{Ct}) \\
&= \Sigma^{-1}SW'(y_{Ct} - \epsilon_{Ct}) \\
&= \Sigma^{-1}SW'y_{Ct} - \Sigma^{-1}SW'\epsilon_{Ct} \\
&= \Sigma^{-1}SW'y_{Ct} + O_p\left(\frac{1}{\sqrt{N_C}}\right),
\end{aligned}$$

so that $\Sigma^{-1}SW'y_{Ct}$ acts as a proxy for $R'f_t$ for large N_C. The last line follows from Assumption 5. Substituting into (1), and by Assumption 3'(ii)'s proportionality condition $\mu_i = R\xi_i$ (R is determined by the left orthogonal matrix in the SVD), we have, for each treated unit

$i \in I$,

$$\begin{aligned} y_{it} &= \mu_i' f_t + \Delta_i 1_{\{t>T_0\}} + \epsilon_{it} \\ &= \xi_i' R' f_t + \Delta_i 1_{\{t>T_0\}} + \epsilon_{it} \\ &= \xi_i' \Sigma^{-1} S W' y_{Ct} + \Delta_i 1_{\{t>T_0\}} + [\epsilon_{it} - \xi_i' \Sigma^{-1} S W' \epsilon_{Ct}]. \end{aligned}$$

This motivates the following regression for treatment unit $i \in I$,

$$y_{it} = \alpha_i' W' y_{Ct} + \delta_i 1_{\{t>T_0\}} + e_{it}.$$

Comparing the DGP with the regression of y_{it}, the regression coefficients are given by

$$\begin{aligned} \alpha_i &= S\Sigma^{-1}\xi_i = S(R'\mu_C \tilde{W})^{-1}\xi_i, \\ \delta_i &= \Delta_i. \end{aligned}$$

We therefore see that the individual treatment effect Δ_i can be identified through the regression coefficient δ_i under Assumption 3'.

6.2 Proof of Proposition 1

Assume that T is fixed throughout the proof. Suppose $w_1, \ldots, w_p$ are p eigenvectors of $\frac{y_C y_C'}{T^{\bar{r}}}$. The p eigenvectors are linearly independent, orthogonal to each other, and can be normalized so that all columns sum to one, in the sense that $\sum_{i=1}^{N_C} w_{ij} = 1$ for all $j = 1, \ldots, p$, and $w_j' w_k = 0$ for all $j \neq k$. Let $c_1, \ldots, c_p$ be the p largest eigenvalues, all of which are non-zero by the rank condition on y_C. Let C be the $p \times p$ diagonal matrix formed by these eigenvalues. By the definition of eigenvalue and eigenvector, we have

$$W' \frac{y_C y_C'}{T^{\bar{r}}} = CW'.$$

We may post-multiply the last line by W and obtain

$$W' \frac{y_C y_C'}{T^{\bar{r}}} W = CW'W.$$

It is easy to see that $CW'W$, and hence $W' y_C y_C' W$, is invertible. By Assumption 5, and using (4), we obtain

$$\begin{aligned} W' y_C y_C' W &= W' \mu_C' F F' \mu_C W + W' \epsilon_C \epsilon_C' W \\ &= W' \mu_C' F F' \mu_C W + O_p\left(\frac{1}{N_C}\right). \end{aligned} \tag{23}$$

Now, let $N_C \to \infty$ and discuss the full rank and rank deficient cases separately.

(a) When $p \geq m$, then noting that $\mu_C W$ is $m \times p$, it is immediate that $rank(\mu_C W) \leq m$ for large enough N_C. Now, suppose the contrary that $rank(\mu_C) < m$. Since $rank(FF') = m$ by Assumption 2, the rank of $W' \mu_C' F F' \mu_C W$ must be less than m, which is a contradiction. Therefore μ_C must have full rank, i.e., $rank(\mu_C) = m$. It then follows from Assumption 4(iii) that $rank(\mu_C W) = m$.

(b) When $p < m$, we see that $W' y_C y_C' W$ is of rank p (as both y_C and W are of rank p, and all the p eigenvalues are non-zero). It follows from (23) that $W' \mu_C' F F' \mu_C W$ is of rank p for large N_C, and under Assumption 4(iii), μ_C is also of rank p. Applying the reduced form SVD to μ_C (see the above lemma), there exists a $p \times p$ nonsingular diagonal matrix Σ, and an $m \times p$ matrix R and an $N_C \times p$ matrix W, where the columns of R and W are orthogonal, and the columns of W sum to one, such that $\mu_C = R \Sigma W'$. Pre-multiplying by R' and post-multiplying by W, we obtain $\Sigma = R' \mu_C W$, which is invertible and so $rank(R' \mu_C W) = p$ as $N_C \to \infty$.

6.3 Proof of Theorems 1 and 2

6.3.1 Preliminaries

Consider the following decomposition of $\hat{\delta}_i$, obtained by substituting (11) into (12):

$$\hat{\delta}_i - \delta_i = \left(1_{post}' M_{\bar{y}_C} 1_{post}\right)^{-1} 1_{post}' M_{\bar{y}_C} e_i. \tag{24}$$

The first task is to evaluate the right hand side of (24). Using (3) and the definition of $\bar{y}_C$, we obtain the weighted control group observations in matrix form

$$\underset{T \times m}{\bar{y}_C} := \underset{T \times N_C}{y_C'} \cdot \underset{N_C \times m}{W} = \underset{T \times m}{F'} \cdot \underset{m \times m}{\bar{\mu}_C} + \underset{T \times m}{\bar{\epsilon}_C},$$

where $\bar{\mu}_C = \mu_C W$ in the full rank case[28], and $\bar{\epsilon}_C = \epsilon_C' W$. For now, we treat W as given in the following proof. The effect of estimation error of W on $\hat{\delta}_i$ is discussed in section 6.3.4.

Now, let us compute the components in (24),

$$1_{post}' \bar{y}_C = 1_{post}' F' \bar{\mu}_C + 1_{post}' \bar{\epsilon}_C,$$

$$\bar{y}_C' \bar{y}_C = \bar{\mu}_C' F F' \bar{\mu}_C + \bar{\mu}_C' F \bar{\epsilon}_C + \bar{\epsilon}_C' F' \bar{\mu}_C + \bar{\epsilon}_C' \bar{\epsilon}_C,$$
$$\bar{y}_C' 1_{post} = \bar{\mu}_C' F 1_{post} + \bar{\epsilon}_C' 1_{post}.$$

Recall that the matrix $(\bar{\mu}_C)^{-1}$ exists in the full rank case. Now we have

$$\begin{aligned}
&1_{post}' M_{\bar{y}_C} 1_{post} \\
=&1_{post}' 1_{post} - 1_{post}' \bar{y}_C (\bar{y}_C' \bar{y}_C)^{-1} \bar{y}_C' 1_{post} \\
=&T_1 - [1_{post}' F' \bar{\mu}_C + 1_{post}' \bar{\epsilon}_C][\bar{\mu}_C' F F' \bar{\mu}_C + \bar{\mu}_C' F \bar{\epsilon}_C + \bar{\epsilon}_C' F' \bar{\mu}_C + \bar{\epsilon}_C' \bar{\epsilon}_C]^{-1} \\
&\times [\bar{\mu}_C' F 1_{post} + \bar{\epsilon}_C' 1_{post}] \\
=&T_1 - [1_{post}' F' + 1_{post}' \bar{\epsilon}_C \bar{\mu}_C^{-1}][F F' + F \bar{\epsilon}_C \bar{\mu}_C^{-1} + (\bar{\mu}_C')^{-1} \bar{\epsilon}_C' F' + (\bar{\mu}_C')^{-1} \bar{\epsilon}_C' \bar{\epsilon}_C \bar{\mu}_C^{-1}]^{-1} \\
&\times [F 1_{post} + (\bar{\mu}_C')^{-1} \bar{\epsilon}_C' 1_{post}].
\end{aligned} \tag{25}$$

[28] In the rank deficient case, simply set $\bar{\mu}_C = R' \mu_C W$, where R is an $m \times p$ matrix identified by the singular value decomposition of μ_C. The $p \times p$ square matrix $\bar{\mu}_C$ is invertible. In this case, the weighted control group observations can be expressed into $\bar{y}_C = y_C' \cdot W = (F'R) \cdot \bar{\mu}_C + \bar{\epsilon}_C$.

Also,

$$
\begin{aligned}
&1'_{post} M_{\bar{y}_C} e_i \\
=&1'_{post} e_i - 1'_{post} \bar{y}_C (\bar{y}'_C \bar{y}_C)^{-1} \bar{y}'_C e_i && (26) \\
=&1'_{post} e_i - [1'_{post} F' + 1'_{post} \bar{\epsilon}_C \bar{\mu}_C^{-1}][FF' + F\bar{\epsilon}_C \bar{\mu}_C^{-1} + (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C F' + (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C \bar{\epsilon}_C \bar{\mu}_C^{-1}]^{-1} \\
&\times [Fe_i + (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C e_i]. && (27)
\end{aligned}
$$

6.3.2 Nonstationary case

Suppose $r_{\min} = \min(r_1, \ldots, r_m) > 0$. Let us define the $m \times m$ diagonal matrices

$$
\tilde{\Upsilon} = \begin{bmatrix} T^{r_1} & & & O \\ & T^{r_2} & & \\ & & \ddots & \\ O & & & T^{r_m} \end{bmatrix},
$$

$$
\tilde{\Upsilon}_1 = \begin{bmatrix} T_1^{r_1} & & & O \\ & T_1^{r_2} & & \\ & & \ddots & \\ O & & & T_1^{r_m} \end{bmatrix}.
$$

Using the newly defined matrices, we apply the following normalizations:

$$
\begin{aligned}
\frac{1}{T_1} 1'_{post} M_{\bar{y}_C} 1_{post} =& 1 - \left[\frac{1'_{post} F' \tilde{\Upsilon}_1^{-1}}{\sqrt{T_1}} + \frac{1'_{post} \bar{\epsilon}_C}{\sqrt{T_1}} \bar{\mu}_C^{-1} \tilde{\Upsilon}_1^{-1} \right] \tilde{\Upsilon}_1 \tilde{\Upsilon}^{-1} \\
&\times \tilde{\Upsilon} [FF' + F\bar{\epsilon}_C \bar{\mu}_C^{-1} + (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C F' + (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C \bar{\epsilon}_C \bar{\mu}_C^{-1}]^{-1} \tilde{\Upsilon} \\
&\times \tilde{\Upsilon}^{-1} \tilde{\Upsilon}_1 \left[\tilde{\Upsilon}^{-1} \frac{F 1_{post}}{\sqrt{T_1}} + \tilde{\Upsilon}^{-1} (\bar{\mu}'_C)^{-1} \frac{\bar{\epsilon}'_C 1_{post}}{\sqrt{T_1}} \right],
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{\sqrt{T_1}} 1'_{post} M_{\bar{y}_C} e_i =& \frac{1'_{post} e_i}{\sqrt{T_1}} - \left[\frac{1'_{post} F' \tilde{\Upsilon}_1^{-1}}{\sqrt{T_1}} + \frac{1'_{post} \bar{\epsilon}_C}{\sqrt{T_1}} \bar{\mu}_C^{-1} \tilde{\Upsilon}_1^{-1} \right] \tilde{\Upsilon}_1 \tilde{\Upsilon}^{-1} \\
&\times \tilde{\Upsilon} [FF' + F\bar{\epsilon}_C \bar{\mu}_C^{-1} + (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C F' + (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C \bar{\epsilon}_C \bar{\mu}_C^{-1}]^{-1} \tilde{\Upsilon} \\
&\times \left[\tilde{\Upsilon}^{-1} F e_i + \tilde{\Upsilon}^{-1} (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C e_i \right].
\end{aligned}
$$

Given $r_{\min} > 0$, the first term strictly dominates the other terms in stochastic order in each of the six square brackets in the above two expressions (by lemmas 2-4 in the Supplementary Results). This is valid regardless of whether N_C is kept finite or tends to infinity. In particular, the measurement error due to factor proxies, as captured by the term $\tilde{\Upsilon}^{-1} (\bar{\mu}'_C)^{-1} \bar{\epsilon}'_C e_i$, is dominated by $\tilde{\Upsilon}^{-1}Fe_i$ in stochastic order. We can thus simplify the two expressions into

$$\begin{aligned}\frac{1}{T_1}1'_{post}M_{\bar{y}_C}1_{post} =& 1-\left[\frac{1'_{post}F'\tilde{\Upsilon}_1^{-1}}{\sqrt{T_1}}+o_p(1)\right]\tilde{\Upsilon}_1\tilde{\Upsilon}^{-1}\\ &\times[\tilde{\Upsilon}^{-1}FF'\tilde{\Upsilon}^{-1}+o_p(1)]^{-1}\\ &\times\tilde{\Upsilon}^{-1}\tilde{\Upsilon}_1\left[\tilde{\Upsilon}_1^{-1}\frac{F1_{post}}{\sqrt{T_1}}+o_p(1)\right],\end{aligned}\tag{28}$$

and

$$\begin{aligned}\frac{1}{\sqrt{T_1}}1'_{post}M_{\bar{y}_C}e_i =& \frac{1'_{post}e_i}{\sqrt{T_1}}-\left[\frac{1'_{post}F'\tilde{\Upsilon}_1^{-1}}{\sqrt{T_1}}+o_p(1)\right]\tilde{\Upsilon}_1\tilde{\Upsilon}^{-1}\\ &\times[\tilde{\Upsilon}^{-1}FF'\tilde{\Upsilon}^{-1}+o_p(1)]^{-1}\\ &\times\left[\tilde{\Upsilon}^{-1}Fe_i+o_p(1)\right],\end{aligned}\tag{29}$$

as $T_1 \to \infty$.

Let us define the following probability limits.

$$\tilde{s}'_F := \text{plim}\,\frac{1'_{post}F'\tilde{\Upsilon}_1^{-1}}{\sqrt{T_1}},\tag{30}$$

$$\tilde{s}_{FF'} := \text{plim}\,\tilde{\Upsilon}^{-1}FF'\tilde{\Upsilon}^{-1},\tag{31}$$

$$\tilde{s}_{Fe_i} := \text{plim}\,\tilde{\Upsilon}^{-1}Fe_i,\tag{32}$$

$$\sigma^2_{e_i} := \lim\frac{1}{T_1}\sum_{s,t\in T^{post}}Cov(e_{is},e_{it}).\tag{33}$$

By lemmas 2-4 in the Supplementary Results, all the terms defined in (30)-(33) contain elements that are $O_p(1)$.[29] When $\kappa = 0$, $\tilde{s}_F$ is $o_p(1)$ by Lemma 4. Note that $\tilde{s}_{FF'}$ is invertible by Assumption 2(iii).

The first term in the normalized numerator, $1'_{post}e_i/\sqrt{T_1}$, is $O_p(1)$ and converges in distribution to $\sigma_{e_i}Z$ by the Central Limit theorem as $T_1 \to \infty$. The quantity $\sigma^2_{e_i}$, defined in (33), is the long-run variance of e_{it}, computed over the post-intervention periods for each i. It is obtained from

$$\begin{aligned}\sigma^2_{e_i} &= \lim_{T_1\to\infty}Var\left(\frac{1}{\sqrt{T_1}}1'_{post}e_i\right)\\ &= \lim_{T_1\to\infty}\frac{1}{T_1}1'_{post}E(e_ie'_i)1_{post}\\ &= \lim_{T_1\to\infty}\frac{1}{T_1}\sum_{s,t\in T^{post}}Cov(e_{is},e_{it}).\end{aligned}$$

[29] For instance, to see that $\tilde{s}_{Fe_i}$ is $O_p(1)$, note that $\tilde{\Upsilon}^{-1}Fe_i = \tilde{\Upsilon}^{-1}F\epsilon_i - \tilde{\Upsilon}^{-1}F\epsilon'_CW(\mu_CW)^{-1}\mu_i$, both of which are $O_p(1)$ for finite N_C by lemma 3.

Define the limit

$$\tilde{\Pi}_r = \lim \tilde{\Upsilon}_1 \tilde{\Upsilon}^{-1} = \begin{bmatrix} \kappa^{r_1} & & & O \\ & \kappa^{r_2} & & \\ & & \ddots & \\ O & & & \kappa^{r_m} \end{bmatrix},$$

where $\kappa = \lim T_1/T$. It is then easy to see that

$$\begin{aligned} \operatorname{plim} \frac{1}{T_1} 1'_{post} M_{\bar{y}_C} 1_{post} &= 1 - \tilde{s}'_F \tilde{\Pi}_r \tilde{s}^{-1}_{FF'} \tilde{\Pi}_r \tilde{s}_F, \\ \operatorname{plim} \frac{1}{\sqrt{T_1}} 1'_{post} M_{\bar{y}_C} e_i &= \sigma_{e_i} Z - \tilde{s}'_F \tilde{\Pi}_r \tilde{s}^{-1}_{FF'} \tilde{\Pi}_r \tilde{s}_{Fe_i}. \end{aligned}$$

Suppose there exists $r_j = 0$ for some *but not all* $j = 1, \ldots, m$. In this case, we introduce two alternative normalizing matrices

$$\Upsilon = \begin{bmatrix} T^{r_1 \vee 0.5} & & & O \\ & T^{r_2 \vee 0.5} & & \\ & & \ddots & \\ O & & & T^{r_m \vee 0.5} \end{bmatrix},$$

$$\Upsilon_1 = \begin{bmatrix} T_1^{r_1 \vee 0.5} & & & O \\ & T_1^{r_2 \vee 0.5} & & \\ & & \ddots & \\ O & & & T_1^{r_m \vee 0.5} \end{bmatrix}.$$

Here $a \vee b = \max(a, b)$. Note that Υ is invertible. Now, those terms in (25) and (27) that contain $\bar{\epsilon}_C$ are strictly dominated by the first or second sample moments of the factor(s) with the maximal integration order. This is valid regardless of whether N_C is kept finite or tends to infinity. We obtain the stated results by following the previous arguments with all $\tilde{\Upsilon}$ replaced by Υ, all $\tilde{\Upsilon}_1$ replaced by Υ_1, $\tilde{\Pi}_r$ by $\Pi_r := \lim \Upsilon_1 \Upsilon^{-1}$, $\tilde{s}'_F$ by $s'_F := \operatorname{plim} 1'_{post} F' \Upsilon^{-1} / \sqrt{T_1}$, $\tilde{s}_{FF'}$ by $s_{FF'} := \operatorname{plim} \Upsilon^{-1} F F' \Upsilon^{-1}$, and $\tilde{s}_{Fe_i}$ by $s_{Fe_i} := \operatorname{plim} \Upsilon^{-1} F e_i$. By lemmas 2-4 in the Supplementary Results, the resulting terms after replacements contain elements that are at most $O_p(1)$.

6.3.3 Stationary case

Suppose that all $r_1 = \ldots = r_m = 0$, i.e., all m factors are stationary. The decomposition of $\hat{\delta}_i$ is slightly different from the nonstationary case. Starting from (24) and separating out the two terms in the numerator as in (26), we obtain

$$\hat{\delta}_i - \delta_i - b_i = \frac{1'_{post} e_i}{1'_{post} M_{\bar{y}_C} 1_{post}}, \tag{34}$$

where b_i is the asymptotic bias given by

$$b_i = \frac{-1'_{post} P_{\bar{y}_C} e_i}{1'_{post} M_{\bar{y}_C} 1_{post}},$$

and $P_{\bar{y}_C} = \bar{y}_C (\bar{y}'_C \bar{y}_C)^{-1} \bar{y}'_C$.

Let us first analyse the asymptotic bias. The denominator, after normalizing appropriately, becomes

$$\begin{aligned}
\frac{1}{T_1} 1'_{post} M_{\bar{y}_C} 1_{post} =& 1 - \frac{T_1}{T} \left[\frac{1'_{post} F'}{T_1} + \frac{1'_{post} \bar{\epsilon}_C}{T_1} \bar{\mu}_C^{-1} \right] \\
& \times \left[\frac{FF'}{T} + \frac{F\bar{\epsilon}_C}{T} \bar{\mu}_C^{-1} + (\bar{\mu}'_C)^{-1} \frac{\bar{\epsilon}'_C F'}{T} + (\bar{\mu}'_C)^{-1} \frac{\bar{\epsilon}'_C \bar{\epsilon}_C}{T} \bar{\mu}_C^{-1} \right]^{-1} \\
& \times \left[\frac{F 1_{post}}{T_1} + (\bar{\mu}'_C)^{-1} \frac{\bar{\epsilon}'_C 1_{post}}{T_1} \right].
\end{aligned}$$

Let us consider the limit of each term in the above expression as $T_1, T \to \infty$. By Assumption 2(ii), the first two population moments of f_t exist and remain constant over time, and hence their sample moments $F 1_{post}/T_1$ and FF'/T converge to $E(f_t)$ and $E(f_t f'_t)$, respectively. Next, the post-intervention sample mean $1'_{post} \bar{\epsilon}_C / T_1$ converges to $E(\bar{\epsilon}_{Ct})$, which is zero by Assumption 1(i). Third, the cross sample moment $F\bar{\epsilon}_C/T$ converges to $E(f_t \bar{\epsilon}_{Ct})$, which is zero by Assumption 1(i). Fourth, the second moment $\bar{\epsilon}'_C \bar{\epsilon}_C / T$ converges to the limiting $m \times m$ covariance matrix $V_{\bar{\epsilon}_C} := \text{plim } \frac{\bar{\epsilon}'_C \bar{\epsilon}_C}{T}$, with its (k, ℓ) element given by

$$v_{k\ell} = \lim_{T \to \infty} \frac{1}{T} \sum_{i,j \in C} \sum_{t=1}^{T} w_{ik} w_{j\ell} \sigma_{ij,tt}.$$

By Assumption 5(i), we see that $v_{k\ell}$ is $O(1)$ as $T \to \infty$. Note that the factor loading matrix $\bar{\mu}_C = \mu'_C W$ remains fixed for finite N_C. As a result, the denominator converges to, as $T_1, T \to \infty$,

$$1 - \kappa E(f'_t)[E(f_t f'_t) + (\bar{\mu}'_C)^{-1} V_{\bar{\epsilon}_C} \bar{\mu}_C^{-1}]^{-1} E(f_t). \tag{35}$$

The numerator, after appropriate normalization, becomes

$$\begin{aligned}
-\frac{1}{T_1} 1'_{post} P_{\bar{y}_C} e_i =& -\frac{1}{T_1} 1'_{post} \bar{y}_C (\bar{y}'_C \bar{y}_C)^{-1} \bar{y}'_C e_i \\
=& - \left[\frac{1'_{post} F'}{T_1} + \frac{1'_{post} \bar{\epsilon}_C}{T_1} \bar{\mu}_C^{-1} \right] \\
& \times \left[\frac{FF'}{T} + \frac{F\bar{\epsilon}_C}{T} \bar{\mu}_C^{-1} + (\bar{\mu}'_C)^{-1} \frac{\bar{\epsilon}'_C F'}{T} + (\bar{\mu}'_C)^{-1} \frac{\bar{\epsilon}'_C \bar{\epsilon}_C}{T} \bar{\mu}_C^{-1} \right]^{-1} \\
& \times \left[\frac{F e_i}{T} + (\bar{\mu}'_C)^{-1} \frac{\bar{\epsilon}'_C e_i}{T} \right].
\end{aligned}$$

The asymptotic limits in the first two brackets were the same as those in the denominator, so let us focus on the terms in the third square bracket. Recall that $e_i = [e_{i1}, \ldots, e_{iT}]'$ is the $T \times 1$ vector of regression errors given by $e_i = \epsilon_i - \bar{\epsilon}'_C (\bar{\mu}_C)^{-1} \mu_i$, where $\bar{\epsilon}_C = W' \epsilon_C$

$(m \times T)$ and $\bar{\mu}_C = \mu_C' W$ $(m \times m)$. The first term in the third square bracket, Fe_i/T, converges to zero, as we have $E(f_t e_{it}) = E(f_t \epsilon_{it}) - E(f_t \bar{\epsilon}_{Ct}')(\bar{\mu}_C)^{-1}\mu_i = 0$ by Assumption 1(i). This holds for finite N_C. The second term, however, tends to a non-zero limit, as $E(\bar{\epsilon}_{Ct} e_{it}) = E(\bar{\epsilon}_{Ct}\epsilon_{it}) - E(\bar{\epsilon}_{Ct}\bar{\epsilon}_{Ct}')(\bar{\mu}_C)^{-1}\mu_i \neq 0$ in general.

Combining the above results, the asymptotic bias of δ_i is given by

$$b_i = \frac{-E(f_t')[E(f_t f_t') + (\bar{\mu}_C')^{-1} V_{\bar{\epsilon}_C} \bar{\mu}_C^{-1}]^{-1} (\bar{\mu}_C')^{-1} E(\bar{\epsilon}_{Ct} e_{it})}{1 - \kappa E(f_t')[E(f_t f_t') + (\bar{\mu}_C')^{-1} V_{\bar{\epsilon}_C} \bar{\mu}_C^{-1}]^{-1} E(f_t)}.$$

Under scenario (a), From (7), we see that $\bar{\epsilon}_{Ct} = O_p\left(\frac{1}{\sqrt{N_C}}\right)$, so that $E(\bar{\epsilon}_{Ct} e_{it}) \to 0$ and $V_{\bar{\epsilon}_C} \to 0$ as $N_C \to \infty$. It follows that $b_i \to 0$ and its denominator converges to $1 - \kappa E(f_t') E(f_t f_t')^{-1} E(f_t)$.

Under scenario (b), we readily see that, when $E(f_t) = 0$, the asymptotic bias b_i is exactly equal to zero with its denominator equal to one. This holds for finite N_C.

As a result, the asymptotic bias is zero under both scenarios.

Finally, by normalizing the terms in (34) appropriately, we obtain

$$\sqrt{T_1}(\hat{\delta}_i - \delta_i) = \frac{\frac{1}{\sqrt{T_1}} 1_{post}' e_i}{\frac{1}{T_1} 1_{post}' M_{\bar{y}_C} 1_{post}}. \tag{36}$$

An application of the Central Limit theorem and Slutsky's theorem yields the stated result.

6.3.4 Estimation error of W

The above proof (for both the nonstationary and stationary cases) supposes that the $N_C \times m$ weighting matrix W is treated as given. In this sub-section we want to argue that the estimation error of W due to sample variation does not affect the asymptotic distribution of $\hat{\delta}_i$ for a given control group with a given size. Recall that the columns of W are the first m eigenvectors of the $N_C \times N_C$ sample covariance matrix $S_C := y_C y_C' / T^{\bar{r}}$. Denote Ω be the population counterpart of W, consisting of m columns of eigenvectors of the population covariance matrix $\Sigma_C := E(y_C y_C' / T^{\bar{r}})$. Let ω_{ik} be the (i,k) element of Ω. It is easy to see, by a strong law of large numbers and under Assumption 5(i), that $S_C = \Sigma_C + o_p(1)$ as $T \to \infty$. It follows that $W = \Omega + o_p(1)$ as $T \to \infty$.

In the nonstationary case, the only dominating term in (28) and (29) that involves W is the $m \times m$ matrix $\tilde{\Upsilon}^{-1} F e_i$. Using the definition of e_i, it is decomposed into $\tilde{\Upsilon}^{-1} F e_i = \tilde{\Upsilon}^{-1} F \epsilon_i - \tilde{\Upsilon}^{-1} F \bar{\epsilon}_C' (\bar{\mu}_C)^{-1} \mu_i$. Only the second term is a function of W. We may compute and decompose the (j,k) element of $F\bar{\epsilon}_C'$ as follows

$$\begin{aligned}
[F\bar{\epsilon}_C']_{(j,k)} &= [F\epsilon_C' W]_{(j,k)} \\
&= \sum_{i\in C} w_{ik} \sum_{t=1}^{T} f_{jt}\epsilon_{it} \\
&= \sum_{i\in C} \left[\omega_{ik} \sum_{t=1}^{T} f_{jt}\epsilon_{it} + (w_{ik}-\omega_{ik}) \sum_{t=1}^{T} f_{jt}\epsilon_{it} \right] \\
&= \sum_{i\in C} \left[\omega_{ik} \sum_{t=1}^{T} f_{jt}\epsilon_{it} + o_p(1) \right]
\end{aligned}$$

as $T\to\infty$. The last line follows from $w_{ik}-\omega_{ik}=o_p(1)$ for each i. In other words, for any given N_C, we have $F\epsilon_C' W = F\epsilon_C'\Omega + o_p(1)$ as $T\to\infty$. Similarly, we obtain $\bar{\mu}_C = \mu_C W = \mu_C\Omega + o_p(1)$ as $T\to\infty$.

For the stationary case, we may apply the same argument to the numerator and denominator of (36). The numerator can be decomposed into $\frac{1}{\sqrt{T_1}} 1_{post}' e_i = \frac{1}{\sqrt{T_1}} 1_{post}'\epsilon_i - \frac{1}{\sqrt{T_1}} 1_{post}'\bar{\epsilon}_C'(\bar{\mu}_C)^{-1}\mu_i$. By the same argument as before, we obtain $1_{post}'\bar{\epsilon}_C' = 1_{post}'\epsilon_C' W = 1_{post}'\epsilon_C'\Omega + o_p(1)$ as $T_1, T\to\infty$. Together with $\mu_C W = \mu_C\Omega + o_p(1)$, we obtain $\frac{1}{\sqrt{T_1}} 1_{post}' e_i = \frac{1}{\sqrt{T_1}} 1_{post}'\epsilon_i + o_p(1)$ as $T_1, T\to\infty$. As for the denominator, the dominating term that involves W is $(\bar{\mu}_C')^{-1}\frac{\bar{\epsilon}_C'\bar{\epsilon}_C}{T}\bar{\mu}_C^{-1}$. Now, the (k,ℓ) element of $\bar{\epsilon}_C'\bar{\epsilon}_C/T$ is

$$\begin{aligned}
\left[\frac{\bar{\epsilon}_C'\bar{\epsilon}_C}{T}\right]_{k,\ell} &= \left[\frac{W'\epsilon_C'\epsilon_C W}{T}\right]_{k,\ell} \\
&= \sum_{i,j\in C} w_{ik} w_{j\ell} \frac{\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}}{T} \\
&= \sum_{i,j\in C} [\omega_{ik}\omega_{j\ell} + \omega_{ik}(w_{j\ell}-\omega_{j\ell}) + \omega_{j\ell}(w_{ik}-\omega_{ik}) \\
&\quad + (w_{ik}-\omega_{ik})(w_{j\ell}-\omega_{j\ell})] \frac{\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}}{T} \\
&= \sum_{i,j\in C} [\omega_{ik}\omega_{j\ell} + o_p(1)] \frac{\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}}{T},
\end{aligned}$$

which implies that $\frac{W'\epsilon_C'\epsilon_C W}{T} = \frac{\Omega'\epsilon_C'\epsilon_C\Omega}{T} + o_p(1)$ as $T\to\infty$. In summary, the terms involving the estimation error of W are strictly dominated as $T_1, T\to\infty$.

6.4 Proof of Corollary 1 and 2

6.4.1 Nonstationary case

Again we focus on the decomposition (24), and then divide both the denominator (25) and the numerator (27) by T_1. The denominator converges in distribution to the same probability limit as before. By the strong law of large numbers, Assumption 1(i), and lemmas 2-4 in the Supplementary Results, both $\frac{1'_{post}e_i}{T_1}$ and $\frac{1'_{post}F'\Upsilon^{-1}}{T_1}$ are $o_p(1)$, and hence the numerator converges to zero almost surely as $T_1 \to \infty$.

6.4.2 Stationary case

The consistency follows readily from the proof of Theorem 2 (that $b_i \to 0$ or $b_i = 0$) and that $\frac{1}{T_1}1'_{post}e_i \xrightarrow{a.s.} 0$ by an application of the strong law of large numbers as $T_1 \to \infty$.

6.5 Supplementary Results

Let u_t be a covariance stationary process with mean 0 and variance σ_u^2. Let f_t be an $I(r)$ process (without drift), defined iteratively by the following: for $r = 0$,

$$f_t = u_t;$$

for $r = 1$,

$$f_t = f_{t-1} + u_t = \sum_{s=1}^{t} u_s;$$

for $r > 1$,

$$f_t = f_{t-1} + \pi_t = \sum_{s=1}^{t} \pi_s,$$

where π_t is an $I(r-1)$ process (without drift).

Lemma 2 *Let f_t be an $I(r)$ process (without drift). Then the following holds as $T \to \infty$:*
(i) $f_T = O_p(T^{r-0.5})$ for $r > 0$;
(ii) $\sum_{t=1}^{T} f_t = O_p(T^{r+0.5})$ for $r \geq 0$;
(iii) $\sum_{t=1}^{T} f_t^2 = \begin{cases} O_p(T^{2r}) & \text{for } r > 0, \\ O_p(T^{0.5}) & \text{for } r = 0. \end{cases}$

Lemma 3 *Suppose ε_t is a stationary process (i.e., $I(0)$) with mean zero and variance σ_ε^2. Assume that f_t and ε_t are independent. Then the following holds as $T \to \infty$:*
(i) if $r > 0$, then

$$\sum_{t=1}^{T} f_t \varepsilon_t = O_p(T^r);$$

(ii) if $r = 0$ and that (f_t, ε_t) is a strictly stationary bivariate process, then

$$\sum_{t=1}^{T} f_t \varepsilon_t = O_p(T^{0.5}).$$

Let $T = T_0 + T_1$, where T_0, T_1 are integers strictly between 0 and T. Let $T^{post} = \{T_0 + 1, T_0 + 2, \ldots, T\}$.

Lemma 4 *Let f_t be an $I(r)$ process (without drift), where $r > 0$. Let $T_1, T \to \infty$, and $T_1/T \to \kappa$. Then,*
(i) for $\kappa = 0$,

$$\frac{1}{T_0^{r-0.5} T_1} \sum_{t \in T^{post}} f_t = O_p(1);$$

(ii) for $\kappa \in (0, 1]$,

$$\frac{1}{T_1^{r+0.5}} \sum_{t \in T^{post}} f_t = O_p(1).$$

The result holds regardless of whether T_0 is finite or infinite.

Corollary 3 *Let f_t be an $I(0)$ process. Let $T_1 \to \infty$. Then,*

$$\frac{1}{\sqrt{T_1}} \sum_{t \in T^{post}} f_t = O_p(1).$$

Proof of Lemma 2 For part (i), let us prove the following statements: for $r > 0$,

$$\begin{aligned} f_t &= O_p(t^{r-0.5}), \\ Var(f_t) &= O(t^{2r-1}), \\ Cov(f_s, f_t) &= O(s^{2r-1}) \text{ for } s < t. \end{aligned}$$

We proceed by mathematical induction.

For $r = 1$,

$$f_t = \sum_{s=1}^{t} u_s.$$

So

$$\begin{aligned} Var(f_t) &= Var\left(\sum_{s=1}^{t} u_s\right) \\ &= \sum_{s=1}^{t} Var(u_s) \\ &= t\sigma_u^2 \\ &= O(t). \end{aligned}$$

It follows that

$$f_t = O_p(t^{0.5})$$

Also, for $s < t$,

$$\begin{aligned} Cov(f_s, f_t) &= Cov\left(f_s, f_s + \sum_{j=s+1}^{t} u_j\right) \\ &= Var(f_s) + Cov\left(f_s, \sum_{j=s+1}^{t} u_j\right) \\ &= Var(f_s) = O(s). \end{aligned}$$

Assume the results hold when f_t is $I(r)$. Now if f_t is $I(r+1)$,

$$f_t = \sum_{s=1}^{t} \pi_s$$

where π_t is an $I(r)$ process. By assumption, $\pi_t = O_p(t^{2r-1})$, $Var(\pi_t) = O(t^{2r-1})$ and $Cov(\pi_s, \pi_t) = O(s^{2r-1})$ for $s < t$. Now

$$\begin{aligned}
Var(f_t) &= Var\left(\sum_{i=1}^{t} \pi_i\right) \\
&= \sum_{i=1}^{t} Var(\pi_i) + 2\sum_{i=1}^{t-1}\sum_{j=i+1}^{t} Cov(\pi_i, \pi_j) \\
&= O\left(\sum_{s=1}^{t} i^{2r-1} + 2\sum_{i=1}^{t-1}\sum_{j=i+1}^{t} i^{2r-1}\right) \\
&= O\left(t^{2r+1} + 2\sum_{i=1}^{t-1}(t-i)\, i^{2r-1}\right) \\
&= O\left(t^{2r+1}\right).
\end{aligned}$$

Also, for $s < t$,

$$Cov(f_s, f_t) = Var(f_s) + Cov\left(\sum_{i=1}^{s} \pi_i, \sum_{j=s+1}^{t} \pi_j\right).$$

But

$$\begin{aligned}
Cov\left(\sum_{i=1}^{s} \pi_i, \sum_{j=s+1}^{t} \pi_j\right) &= \sum_{i=1}^{s}\sum_{j=s+1}^{t} Cov(\pi_i, \pi_j) \\
&= O\left(\sum_{i=1}^{s}\sum_{j=s+1}^{t} i^{2r-1}\right) \\
&= O\left(\sum_{i=1}^{s}(t-i)i^{2r-1}\right) \\
&= O\left(s^{2r+1}\right).
\end{aligned}$$

As a result,

$$Cov(f_s, f_t) = O\left(s^{2r+1}\right).$$

So the results hold for all non-negative integers r.

Part (ii) follows readily from part (i) for the case $r > 0$. For $r = 0$, the result is obtained from the Central Limit Theorem.

Let us turn to part (iii). For $r = 0$, the result is immediate from part (ii), by noting that f_t^2 is a stationary process. For $r > 0$, we obtain from part (i) that,

$$f_t = O_p(t^{r-0.5}),$$

which implies that

$$f_t^2 = O_p(t^{2r-1}).$$

Therefore, we have

$$\sum_{t=1}^{T} f_t^2 = O_p\left(\sum_{t=1}^{T} t^{2r-1}\right) = O_p(T^{2r}).$$

Proof of Lemma 3 The result for $r = 0$ is obvious by noting that $f_s\varepsilon_s$ is stationary. For $r > 0$, let us compute

$$Var\left(\sum_{s=1}^{t} f_s\varepsilon_s\right) = \sum_{s=1}^{t} Var(f_s\varepsilon_s) + 2\sum_{i=1}^{t-1}\sum_{j=1}^{t} Cov(f_i\varepsilon_i, f_j\varepsilon_j).$$

Since $E(f_t) = E(\varepsilon_t) = 0$ and that f_t and ε_t are independent, it follows that

$$Var(f_s\varepsilon_s) = Var(f_s)Var(\varepsilon_s) = \sigma_e^2 O(s^{2r-1}),$$

by the first lemma. Also, for $i < j$,

$$Cov(f_i\varepsilon_i, f_j\varepsilon_j) = E(f_i\varepsilon_i f_j\varepsilon_j) = E(f_i f_j)E(\varepsilon_i)E(\varepsilon_j) = 0.$$

So, we obtain

$$Var\left(\sum_{s=1}^{t} f_s\varepsilon_s\right) = O\left(\sum_{s=1}^{t} s^{2r-1}\right) = O(t^{2r}),$$

which implies that $\sum_{s=1}^{t} f_s\varepsilon_s = O_p(t^r)$.

Proof of Lemma 4 Let $T = T_0 + T_1$. If f_t is $I(1)$, we have

$$\begin{aligned}
\sum_{t=T_0+1}^{T} f_t &= \sum_{t=T_0+1}^{T} [f_{T_0} + (f_t - f_{T_0})] \\
&= T_1 f_{T_0} + \sum_{t=T_0+1}^{T} (f_t - f_{T_0}) \\
&= T_1 f_{T_0} + \sum_{t=T_0+1}^{T} \sum_{i=T_0+1}^{t} u_i \\
&= T_1 f_{T_0} + \sum_{t=1}^{T_1}\sum_{i=1}^{t} u_{T_0+i}.
\end{aligned}$$

If f_t is $I(2)$, we have

$$\begin{aligned}
\sum_{t=T_0+1}^{T} f_t &= T_1 f_{T_0} + \sum_{t=T_0+1}^{T} (f_t - f_{T_0}) \\
&= T_1 f_{T_0} + \sum_{t=T_0+1}^{T} \left[(t - T_0)\pi_{T_0} + \sum_{i=T_0+1}^{t} \sum_{j=T_0+1}^{i} u_j\right] \\
&= T_1 f_{T_0} + \frac{T_1(T_1+1)}{2}\pi_{T_0} + \sum_{t=1}^{T_1}\sum_{i=1}^{t}\sum_{j=1}^{i} u_{T_0+j}.
\end{aligned}$$

Repeating the above computations, we obtain the following lemma:

Lemma 5 *Let $\pi_{T_0}^{(j)}$ be an $I(j)$ process where $j > 0$. Let f_t be the $I(r)$ process defined as before.*

Suppose $r > 0$. Then we have

$$\sum_{t\in T^{post}} f_t = T_1 f_{T_0} + \frac{T_1(T_1+1)}{2!}\pi_{T_0}^{(r-1)} + \frac{T_1(T_1+1)(2T_1+1)}{3!}\pi_{T_0}^{(r-2)} + \cdots$$
$$+ P^{(r)}(T_1)\pi_{T_0}^{(1)} + \sum_{t=1}^{T_1}\sum_{i_1=1}^{t}\cdots\sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r},$$

where $P^{(r)}(T_1)$ is a polynomial in T_1 with maximum order r.

Now let us continue with the proof. Assume that $r > 0$. Then we see that

$$f_{T_0} = O_p(T_0^{r-0.5}),$$
$$\pi_{T_0}^{(j)} = O_p(T_0^{j-0.5}),$$

as $T_0 \to \infty$, or they are both $O_p(1)$ if T_0 is finite. Since u_{T_0+j} is $I(0)$, the process $\xi_j := \sum_{t=1}^{j}\sum_{i_1=1}^{t}\cdots\sum_{i_r=1}^{i_{r=1}} u_{T_0+i_r}$ is $I(r+1)$, and so $\xi_j = O_p(T_1^{r+0.5})$. Therefore,

$$\sum_{t\in T^{post}} f_t = O_p(T_1 T_0^{r-0.5}) + O_p(T_1^2 T_0^{r-1.5}) + O_p(T_1^3 T_0^{r-2.5}) + \cdots$$
$$+ O_p(T_1^r T_0^{0.5}) + O_p(T_1^{r+0.5}).$$

For $\kappa = 0$, we see that the correct normalizing factor is $\frac{1}{T_0^{r-0.5}T_1}$. Applying lemma 5, we have

$$\frac{1}{T_0^{r-0.5}T_1}\sum_{t\in T^{post}} f_t = \frac{1}{T_0^{r-0.5}} f_{T_0} + \frac{T_1+1}{2}\frac{1}{T_0^{r-0.5}}\pi_{T_0}^{(r-1)} + \frac{(T_1+1)(2T_1+1)}{6}\frac{1}{T_0^{r-0.5}}\pi_{T_0}^{(r-2)}$$
$$+ \cdots + P^{(r-1)}(T_1)\frac{1}{T_0^{r-0.5}}\pi_{T_0}^{(1)} + \frac{1}{T_0^{r-0.5}T_1}\sum_{t=1}^{T_1}\sum_{i_1=1}^{t}\cdots\sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r}$$
$$= \frac{1}{T_0^{r-0.5}} f_{T_0} + \frac{T_1+1}{2T_0}\frac{1}{T_0^{r-1.5}}\pi_{T_0}^{(r-1)} + \frac{(T_1+1)(2T_1+1)}{6T_0^2}\frac{1}{T_0^{r-2.5}}\pi_{T_0}^{(r-2)}$$
$$+ \cdots + \frac{P^{(r-1)}(T_1)}{T_0^{r-1}}\frac{1}{\sqrt{T_0}}\pi_{T_0}^{(1)} + \left(\frac{T_1}{T_0}\right)^{r-0.5}\frac{1}{T_1^{r+0.5}}\sum_{t=1}^{T_1}\sum_{i_1=1}^{t}\cdots\sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r}.$$

Except for the first term, all the terms on the right hand side vanish in the limit. As a result,

$$\frac{1}{T_0^{r-0.5}T_1}\sum_{t\in T^{post}} f_t = \frac{1}{T_0^{r-0.5}} f_{T_0} + o_p(1)$$
$$= O_p(1).$$

For $\kappa \in (0,1]$, we see that the correct normalizing factor is $\frac{1}{T_1^{r+0.5}}$. Applying lemma 5, we have

$$
\begin{aligned}
&\frac{1}{T_1^{r+0.5}} \sum_{t \in T^{post}} f_t \\
&= \frac{T_1}{T_1^{r+0.5}} f_{T_0} + \frac{T_1(T_1+1)}{2!} \frac{1}{T_1^{r+0.5}} \pi_{T_0}^{(r-1)} + \frac{T_1(T_1+1)(2T_1+1)}{3!} \frac{1}{T_1^{r+0.5}} \pi_{T_0}^{(r-2)} \\
&+ \cdots + P^{(r)}(T_1) \frac{1}{T_1^{r+0.5}} \pi_{T_0}^{(1)} + \frac{1}{T_1^{r+0.5}} \sum_{t=1}^{T_1} \sum_{i_1=1}^{t} \cdots \sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r} \\
&= \left(\frac{T_0}{T_1}\right)^{r-0.5} \frac{1}{T_0^{r-0.5}} f_{T_0} + \frac{1}{2!}\left(\frac{T_0}{T_1}\right)^{r-1.5} \frac{1}{T_0^{r-1.5}} \pi_{T_0}^{(r-1)} + \frac{1}{3!}\left(\frac{T_0}{T_1}\right)^{r-2.5} \frac{1}{T_0^{r-2.5}} \pi_{T_0}^{(r-2)} \\
&+ \cdots + \frac{1}{r!}\sqrt{\frac{T_0}{T_1}} \frac{1}{\sqrt{T_0}} \pi_{T_0}^{(1)} + \frac{1}{T_1^{r+0.5}} \sum_{t=1}^{T_1} \sum_{i_1=1}^{t} \cdots \sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r} + o_p(1) \\
&= \left(\frac{1-\kappa}{\kappa}\right)^{r-0.5} \frac{1}{T_0^{r-0.5}} f_{T_0} + \frac{1}{2!}\left(\frac{1-\kappa}{\kappa}\right)^{r-1.5} \frac{1}{T_0^{r-1.5}} \pi_{T_0}^{(r-1)} + \frac{1}{3!}\left(\frac{1-\kappa}{\kappa}\right)^{r-2.5} \frac{1}{T_0^{r-2.5}} \pi_{T_0}^{(r-2)} \\
&+ \cdots + \frac{1}{r!}\sqrt{\frac{1-\kappa}{\kappa}} \frac{1}{\sqrt{T_0}} \pi_{T_0}^{(1)} + \frac{1}{T_1^{r+0.5}} \sum_{t=1}^{T_1} \sum_{i_1=1}^{t} \cdots \sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r} + o_p(1).
\end{aligned}
$$

Note that all the $r+1$ terms explicitly displayed on the right are $O_p(1)$. In particular, when $\kappa = \infty$, the first r terms on the right vanish in the limit. As a result,

$$
\begin{aligned}
\frac{1}{T_1^{r+0.5}} \sum_{t \in T^{post}} f_t &= \frac{1}{T_1^{r+0.5}} \sum_{t=1}^{T_1} \sum_{i_1=1}^{t} \cdots \sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r} + o_p(1) \\
&= O_p(1).
\end{aligned}
$$

In summary, as $T_1, T \to \infty$, and $T_1/T \to \kappa$, we obtain the following results depending on the value of κ:

(i) for $\kappa = 0$,

$$
\begin{aligned}
\frac{1}{T_0^{r-0.5} T_1} \sum_{t \in T^{post}} f_t &= \frac{1}{T_0^{r-0.5}} f_{T_0} + o_p(1) \\
&= O_p(1);
\end{aligned}
$$

(ii) for $\kappa \in (0,1)$,

$$
\begin{aligned}
\frac{1}{T_1^{r+0.5}} \sum_{t \in T^{post}} f_t = &\left(\frac{1-\kappa}{\kappa}\right)^{r-0.5} \frac{1}{T_0^{r-0.5}} f_{T_0} + \frac{1}{2!}\left(\frac{1-\kappa}{\kappa}\right)^{r-1.5} \frac{1}{T_0^{r-1.5}} \pi_{T_0}^{(r-1)} + \frac{1}{3!}\left(\frac{1-\kappa}{\kappa}\right)^{r-2.5} \frac{1}{T_0^{r-2.5}} \pi_{T_0}^{(r-2)} \\
&+ \cdots + \frac{1}{r!}\sqrt{\frac{1-\kappa}{\kappa}} \frac{1}{\sqrt{T_0}} \pi_{T_0}^{(1)} + \frac{1}{T_1^{r+0.5}} \sum_{t=1}^{T_1} \sum_{i_1=1}^{t} \cdots \sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r} + o_p(1) \\
&= O_p(1);
\end{aligned}
$$

(iii) for $\kappa = 1$,

$$\frac{1}{T_1^{r+0.5}} \sum_{t \in T^{post}} f_t = \frac{1}{T_1^{r+0.5}} \sum_{t=1}^{T_1} \sum_{i_1=1}^{t} \cdots \sum_{i_r=1}^{i_{r-1}} u_{T_0+i_r} + o_p(1)$$
$$= O_p(1).$$

The last result (the $\kappa = 1$ case) holds regardless of whether T_0 is finite or infinite. If T_0 is finite, then all of the above expressions remain valid by setting $T_0 = 1$.

Proof of Corollary 3 The analysis is straightforward for $r = 0$, as the Central Limit Theorem implies that

$$\frac{1}{\sqrt{T_1}} \sum_{t \in T^{post}} f_t = O_p(1)$$

as long as $T_1 \to \infty$ (regardless of the value of κ).

TABLE I

SMALL SAMPLE PROPERTIES OF ESTIMATORS UNDER STATIONARY FACTORS[a]

	One AR(1) factor (ρ=0.5)							
	Mean Bias				Empirical SD			
(T_0, T_1)	5, 5	10, 10	25, 25	50, 50	5, 5	10, 10	25, 25	50, 50
(N_C, N_I)								
CCEPC-DID (3PCs)								
5, 5	0.16	-0.15	-0.01	0.04	0.21	0.14	0.07	0.05
10, 10	-0.11	0.14	-0.02	0.12	0.15	0.09	0.05	0.04
25, 25	0.04	0.00	0.02	-0.08	0.10	0.06	0.03	0.02
50, 50	-0.04	-0.02	0.12	0.06	0.07	0.04	0.02	0.02
CCEPC-DID (1PC)								
5, 5	0.16	-0.15	-0.02	0.04	0.18	0.13	0.07	0.05
10, 10	-0.10	0.14	-0.02	0.12	0.13	0.09	0.05	0.04
25, 25	0.04	0.00	0.02	-0.08	0.08	0.05	0.03	0.02
50, 50	-0.04	-0.02	0.12	0.06	0.06	0.04	0.02	0.02
CCE-DID								
5, 5	0.16	-0.15	-0.01	0.04	0.11	0.09	0.05	0.03
10, 10	-0.10	0.14	-0.02	0.12	0.08	0.06	0.03	0.02
25, 25	0.04	0.00	0.02	-0.08	0.05	0.04	0.02	0.02
50, 50	-0.04	-0.02	0.12	0.06	0.04	0.03	0.02	0.01
Hsiao et al. (2012)								
5, 5		-0.16	-0.02	0.04		0.20	0.08	0.05
10, 10			-0.02	0.12			0.06	0.04
25, 25				-0.08				0.03
50, 50								
Bai (2009)								
5, 5	-0.14	-0.29	-0.07	0.01	0.36	0.29	0.16	0.12
10, 10	-0.37	-0.01	-0.07	0.09	0.37	0.28	0.15	0.12
25, 25	-0.24	-0.14	-0.04	-0.11	0.38	0.27	0.17	0.12
50, 50	-0.31	-0.16	0.06	0.03	0.38	0.27	0.17	0.13
DID								
5, 5	0.16	-0.15	-0.01	0.04	0.18	0.20	0.09	0.06
10, 10	-0.10	0.14	-0.01	0.12	0.20	0.18	0.07	0.07
25, 25	0.05	0.00	0.02	-0.08	0.22	0.15	0.08	0.06
50, 50	-0.03	-0.02	0.12	0.06	0.21	0.15	0.08	0.07

	Three AR(1) factors (ρ_1=0.5, ρ_2=0.7, ρ_3=0.9)							
	Mean Bias				Empirical SD			
(T_0,T_1)	5, 5	10, 10	25, 25	50, 50	5, 5	10, 10	25, 25	50, 50
(N_C, N_I)								
CCEPC-DID (3PCs)								
5, 5	0.05	-0.36	0.21	0.24	0.37	0.14	0.24	0.09
10, 10	0.00	-0.02	-0.12	-0.06	0.22	0.23	0.12	0.14
25, 25	-0.07	0.03	-0.14	-0.08	0.20	0.13	0.10	0.05
50, 50	0.11	-0.06	0.02	0.03	0.16	0.14	0.10	0.06
CCEPC-DID (1PC)								
5, 5	0.04	-0.36	0.21	0.24	0.34	0.13	0.24	0.09
10, 10	0.00	-0.02	-0.12	-0.06	0.19	0.24	0.12	0.16
25, 25	-0.07	0.03	-0.14	-0.08	0.21	0.14	0.12	0.06
50, 50	0.10	-0.06	0.01	0.03	0.16	0.17	0.14	0.10
CCE-DID								
5, 5	0.05	-0.36	0.22	0.24	0.30	0.09	0.17	0.06
10, 10	0.00	-0.02	-0.12	-0.06	0.14	0.20	0.09	0.11
25, 25	-0.06	0.02	-0.13	-0.08	0.19	0.11	0.09	0.05
50, 50	0.10	-0.06	0.02	0.03	0.14	0.14	0.10	0.07
Hsiao et al. (2012)								
5, 5		-0.36	0.21	0.24		0.25	0.25	0.09
10, 10			-0.13	-0.06			0.14	0.14
25, 25				-0.08				0.06
50, 50								
Bai (2009)								
5, 5	-0.24	-0.44	0.17	0.24	0.92	0.85	0.85	0.70
10, 10	-0.28	-0.12	-0.17	-0.06	0.86	0.92	0.87	0.72
25, 25	-0.33	-0.08	-0.18	-0.09	0.87	0.89	0.88	0.67
50, 50	-0.19	-0.16	-0.03	0.02	0.89	0.89	0.93	0.70
DID								
5, 5	0.02	-0.34	0.21	0.25	0.61	0.43	0.50	0.39
10, 10	-0.01	0.00	-0.13	-0.05	0.50	0.56	0.49	0.49
25, 25	-0.08	0.04	-0.14	-0.08	0.52	0.53	0.49	0.37
50, 50	0.08	-0.04	0.01	0.04	0.54	0.52	0.60	0.43

[a] Number of replications=1000.

TABLE II

SMALL SAMPLE PROPERTIES OF ESTIMATORS UNDER NONSTATIONARY FACTORS[a]

Three I(1) factors

	Mean Bias				Empirical SD			
(T_0, T_1)	5, 5	10, 10	25, 25	50, 50	5, 5	10, 10	25, 25	50, 50
(N_C, N_I)								
<u>CCEPC-DID (3PCs)</u>								
5, 5	0.05	-0.35	0.21	0.26	0.33	0.17	0.22	0.56
10, 10	0.00	-0.01	-0.14	-0.06	0.23	0.17	0.32	0.31
25, 25	-0.07	0.02	-0.14	-0.08	0.18	0.13	0.18	0.12
50, 50	0.11	-0.06	0.02	0.02	0.15	0.11	0.13	0.12
<u>CCEPC-DID (1PC)</u>								
5, 5	0.06	-0.36	0.21	0.27	0.30	0.16	0.24	0.69
10, 10	0.00	-0.02	-0.14	-0.05	0.22	0.17	0.42	0.45
25, 25	-0.06	0.02	-0.15	-0.07	0.17	0.15	0.26	0.35
50, 50	0.10	-0.06	0.01	0.03	0.15	0.14	0.26	0.25
<u>CCE-DID</u>								
5, 5	0.07	-0.36	0.21	0.28	0.50	0.13	0.50	0.98
10, 10	0.00	-0.04	-0.15	-0.04	0.27	0.33	0.66	0.69
25, 25	-0.05	0.02	-0.15	-0.06	0.34	0.27	0.45	0.50
50, 50	0.11	-0.06	0.01	0.03	0.28	0.30	0.47	0.39
<u>Hsiao et al. (2012)</u>								
5, 5		-0.34	0.21	0.28		0.35	0.28	0.65
10, 10			-0.14	-0.05			0.39	0.39
25, 25				-0.08				0.18
50, 50								
<u>Bai (2009)</u>								
5, 5	-0.16	-0.40	0.12	0.22	1.35	1.85	2.53	4.36
10, 10	-0.21	-0.10	-0.25	-0.10	1.30	1.79	2.94	4.19
25, 25	-0.26	-0.05	-0.25	-0.11	1.27	1.86	2.91	4.19
50, 50	-0.12	-0.13	-0.11	-0.01	1.35	1.85	3.15	4.04
<u>DID</u>								
5, 5	0.07	-0.33	0.20	0.28	0.57	0.68	0.75	1.89
10, 10	0.02	0.00	-0.16	-0.04	0.54	0.62	1.14	1.70
25, 25	-0.04	0.05	-0.16	-0.07	0.50	0.71	1.05	1.63
50, 50	0.12	-0.04	-0.01	0.03	0.57	0.67	1.28	1.59

Three factors (I(1), I(2) and I(3))

	Mean Bias				Empirical SD			
(T_0, T_1)	5, 5	10, 10	25, 25	50, 50	5, 5	10, 10	25, 25	50, 50
(N_C, N_I)								
<u>CCEPC-DID (3PCs)</u>								
5, 5	0.03	-0.36	0.16	0.26	0.63	0.23	0.92	0.49
10, 10	0.01	-0.02	-0.14	-0.04	0.30	0.60	0.33	0.40
25, 25	-0.06	0.02	-0.15	-0.08	0.26	0.20	0.22	0.14
50, 50	0.10	-0.06	0.01	0.03	0.19	0.18	0.16	0.11
<u>CCEPC-DID (1PC)</u>								
5, 5	0.05	-0.37	0.00	0.37	0.73	0.25	6.01	3.99
10, 10	0.00	-0.07	-0.18	-0.37	0.25	1.67	1.05	12.77
25, 25	-0.05	-0.01	-0.23	-0.16	0.48	0.67	2.26	3.51
50, 50	0.11	-0.10	-0.10	-0.18	0.30	1.05	3.07	8.49
<u>CCE-DID</u>								
5, 5	0.04	-0.37	-0.11	0.37	1.00	0.28	7.77	4.39
10, 10	0.00	-0.06	-0.21	-0.42	0.32	2.15	1.63	16.44
25, 25	-0.06	-0.01	-0.28	-0.19	0.70	0.97	3.21	5.27
50, 50	0.10	-0.10	-0.16	-0.22	0.44	1.47	4.22	11.25
<u>Hsiao et al. (2012)</u>								
5, 5		-0.44	0.25	0.19		1.46	5.92	3.27
10, 10			-0.14	-0.17			1.91	3.00
25, 25				-0.09				2.34
50, 50								
<u>Bai (2009)</u>								
5, 5	-0.98	-2.4	9.2	67.7	39.6	140.1	1191	5808
10, 10	-0.90	-2.6	8.9	74.1	33.4	170.6	1172	6404
25, 25	-1.01	-2.4	9.0	63.5	36.3	158.3	1195	5483
50, 50	-0.84	-2.4	9.9	70.3	35.3	157.9	1304	6058
<u>DID</u>								
5, 5	-0.39	-1.1	5.2	31.1	23.3	57.8	602	2595
10, 10	-0.29	-1.2	4.5	42.2	16.0	88.1	543	3538
25, 25	-0.43	-1.0	4.6	30.4	19.5	78.4	561	2564
50, 50	-0.24	-1.1	5.8	37.1	18.4	79.5	687	3104

[a] Number of replications=1000.

TABLE III

SMALL SAMPLE PROPERTIES OF ESTIMATORS UNDER DIFFERENT SAMPLE BALANCES[a]

		Mean Bias					Empirical SD				
(N_C, N_I)	(T_0, T_1)	CCEPC-DID (3PCs)	CCE-DID	Hsiao et al. (2012)	Bai (2009)	DID	CCEPC-DID (3PCs)	CCE-DID	Hsiao et al. (2012)	Bai (2009)	DID
Stationary: One AR(1) factor (ρ=0.5):											
10,40	25,25	0.02	0.02	0.02	-0.03	0.02	0.05	0.03	0.06	0.21	0.09
40,10	25,25	0.16	0.16		0.10	0.16	0.03	0.02		0.13	0.07
49,1	25,25	0.00	0.00		-0.05	0.00	0.09	0.06		0.13	0.10
25,25	10,40	0.02	0.02		-0.04	0.02	0.04	0.02		0.12	0.12
25,25	40,10	0.02	0.02	0.02	-0.04	0.02	0.04	0.04	0.06	0.26	0.10
Stationary: Three AR(1) factors (ρ_1=0.5, ρ_2=0.7, ρ_3=0.9):											
10,40	25,25	-0.14	-0.13	-0.14	-0.19	-0.14	0.16	0.13	0.18	1.12	0.52
40,10	25,25	0.02	0.02		-0.04	0.01	0.09	0.08		0.74	0.48
49,1	25,25	-0.06	-0.05		-0.11	-0.06	0.12	0.08		0.60	0.35
25,25	10,40	-0.14	-0.13		-0.17	-0.12	0.12	0.07		0.58	0.59
25,25	40,10	-0.14	-0.13	-0.14	-0.19	-0.14	0.10	0.10	0.12	1.15	0.48
Nonstationary: Three I(1) factors:											
10,40	25,25	-0.14	-0.14	-0.14	-0.25	-0.16	0.21	0.44	0.25	3.19	0.98
40,10	25,25	0.02	0.00		-0.13	-0.01	0.17	0.54		2.77	1.06
49,1	25,25	-0.05	-0.07		-0.18	-0.07	0.17	0.40		2.35	0.82
25,25	10,40	-0.14	-0.15		-0.27	-0.18	0.17	0.42		2.07	1.05
25,25	40,10	-0.13	-0.13	-0.13	-0.37	-0.19	0.17	0.46	0.19	3.46	1.04
Nonstationary: Three factors (I(1), I(2) and I(3)):											
10,40	25,25	-0.16	-0.37	-0.13	10.90	4.92	0.29	5.62	2.20	1360	605
40,10	25,25	0.01	-0.11		7.99	4.64	0.18	2.85		1090	545
49,1	25,25	-0.06	-0.11		6.33	3.18	0.23	1.26		864	378
25,25	10,40	-0.14	-0.12		2.95	2.06	0.22	3.53		662	432
25,25	40,10	-0.13	-0.06	-0.11	22.02	9.17	0.23	5.06	0.44	1798	702

[a] Number of replications=1000.

TABLE IV

COMPARISON OF TREATMENT EFFECT ESTIMATES, GDP GROWTH DATA [a]

Estimator	Political Integration Data (93:Q1 - 03:Q4)			Economic Integration Data (93:Q1 - 08:Q1)		
CCEPC-DID						
Number of PCs						
1	0.010	(0.012)		0.025	(0.005)	***
2	0.011	(0.010)		0.026	(0.005)	***
3	0.011	(0.009)		0.028	(0.005)	***
5	0.021	(0.010)	**	0.022	(0.006)	***
10				0.040	(0.008)	***
CCE-DID	0.010	(0.012)		0.023	(0.005)	***
HCW (2012)	-0.036	(0.089)		0.024	(0.024)	
Bai (2009)	-0.029	(0.009)	***	0.035	(0.004)	***
DID	0.001	(0.008)		0.032	(0.003)	***

[a] Standard errors are in parentheses. The HCW estimate uses all countries from the control group donor pool and SD is in parentheses. *, Significant at the 10 percent level; **, Significant at the 5 percent level; ***, Significant at the 1 percent level.

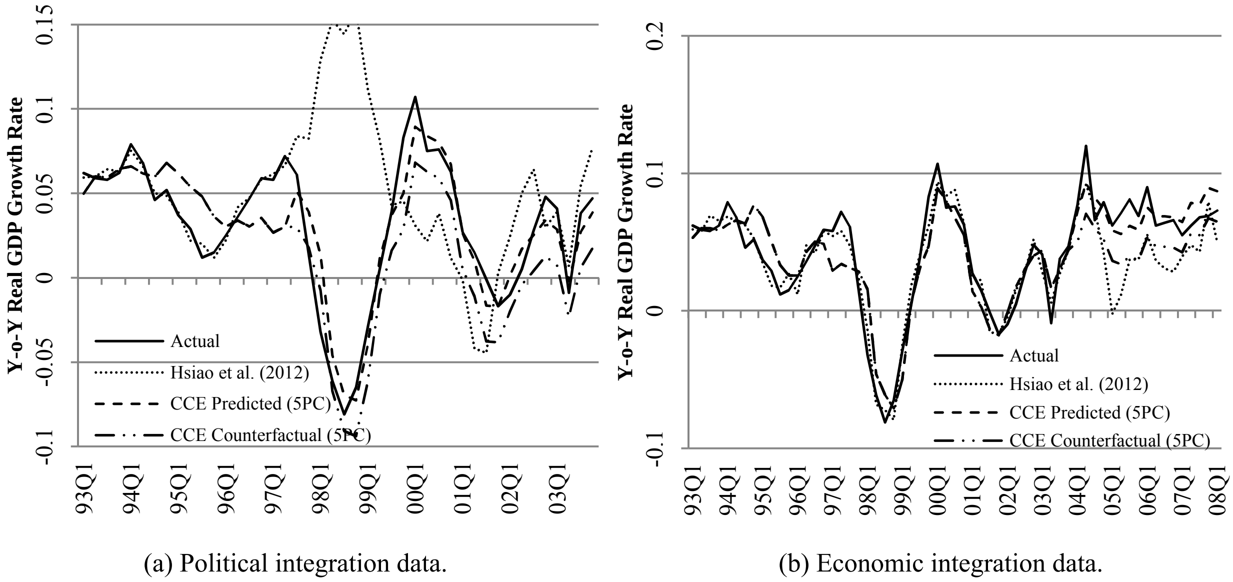


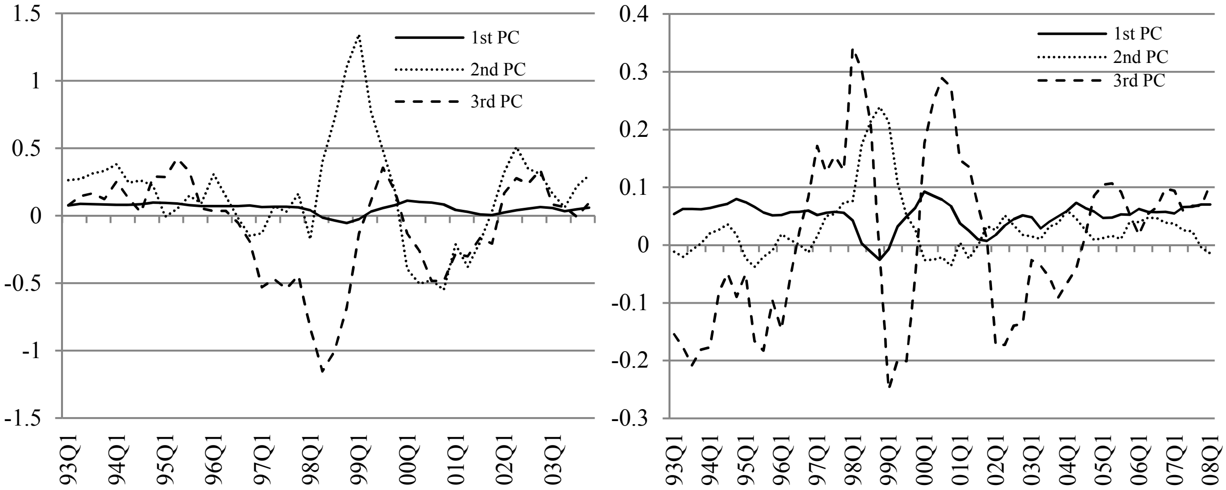
FIGURE 1. -- Actual Hong Kong GDP growth versus predictions from HCW and CCEPC-DID methods.

TABLE V

COUNTRY WEIGHTS FROM THE CCEPC-DID ESTIMATOR, GDP GROWTH DATA (FIRST TWO PRINCIPAL COMPONENTS)[a]

	Political Integration Data		Economic Integration Data			Political Integration Data		Economic Integration Data	
	PC1	PC2	PC1	PC2		PC1	PC2	PC1	PC2
Australia	-	-	0.035	0.217	Malaysia	0.151	-0.593	0.106	-0.194
Austria	-	-	0.017	0.104	Mexico	-	-	0.043	0.146
Canada	-	-	0.037	0.107	Netherlands	-	-	0.027	0.115
China	0.151	2.787	0.103	0.337	New Zealand	-	-	0.039	0.076
Denmark	-	-	0.023	0.052	Philippines	0.083	0.173	0.055	0.044
Finland	-	-	0.032	0.217	Singapore	0.127	1.161	0.086	0.037
France	-	-	0.016	0.117	Switzerland	-	-	0.013	0.086
Germany	-	-	0.007	0.039	Taiwan	0.087	1.707	0.057	0.198
Indonesia	0.151	-4.353	0.095	-0.448	Thailand	0.086	-0.616	0.062	-0.338
Italy	-	-	0.014	0.057	United Kingdom	-	-	0.034	0.174
Japan	0.009	-0.596	0.008	-0.087	United States	0.036	1.025	0.024	0.164
Korea	0.121	0.306	0.067	-0.218					

a For each principal component, all weights sum to one. PC1: first principal component; PC2: second principal component.



(a) Political integration data. (b) Economic integration data.

FIGURE 2. -- Principal components from the CCEPC-DID estimator, GDP growth data.

TABLE VI

PROPORTION OF VARIANCE EXPLAINED BY PRINCIPAL COMPONENTS, GDP GROWTH DATA[a]

	Political Integration Data	Economic Integration Data
1st PC	85.6980	94.0300
2nd PC	8.7698	2.8720
3rd PC	3.9839	2.0090
4th PC	1.0811	0.5911
5th PC	0.2375	0.2747

a Only the first five principal components are reported. Numbers are expressed in percent.

TABLE VII

COMPARISON OF TREATMENT EFFECT ESTIMATES, CIGARETTE SALES DATA [a]

	Level Data (1970-2000)						Yearly Change Data (1971-2000)					
Estimator	Not detrended			Detrended			Not detrended			Detrended		
CCEPC-DID												
Number of PCs used												
1	-19.97	(4.79)	***	-19.13	(3.41)	***	-1.69	(0.25)	***	-0.63	(0.13)	***
2	-15.29	(4.19)	***	-5.47	(3.80)		-1.90	(0.25)	***	-0.89	(0.20)	***
3	-5.85	(6.66)		-1.78	(4.91)		-1.77	(0.25)	***	-0.63	(0.23)	***
5	-1.23	(2.95)		-2.13	(2.40)		-1.71	(0.32)	***	-0.60	(0.22)	***
10	-5.39	(2.31)	**	-4.69	(2.11)	**	-0.74	(0.48)		-0.43	(0.29)	
CCE-DID	-20.62	(4.79)	***				-1.19	(0.25)	***			
Synthetic control	-19.48	(6.91)	***	-19.48	(6.91)	***	-1.94	(1.97)		-1.94	(1.97)	
Bai (2009)	-57.23	(3.97)	***	-39.73	(3.97)	***	-3.01	(0.14)	***	-2.16	(0.14)	***
DID	-27.35	(2.61)	***	-27.35	(2.61)	***	-0.75	(0.26)	***	-0.75	(0.26)	***

[a] Standard errors are in parentheses. The synthetic control estimate is based on Abadie et al. (2010) and SD is in parentheses. Detrended data are constructed by subtracting the variable by its cross-sectional mean each period. *, Significant at the 10 percent level; **, Significant at the 5 percent level; ***, Significant at the 1 percent level.

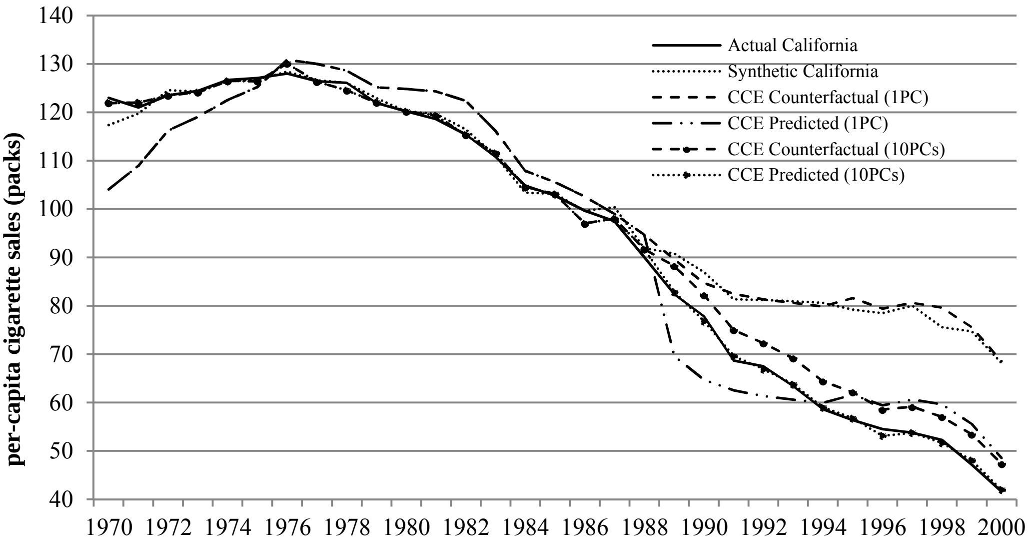
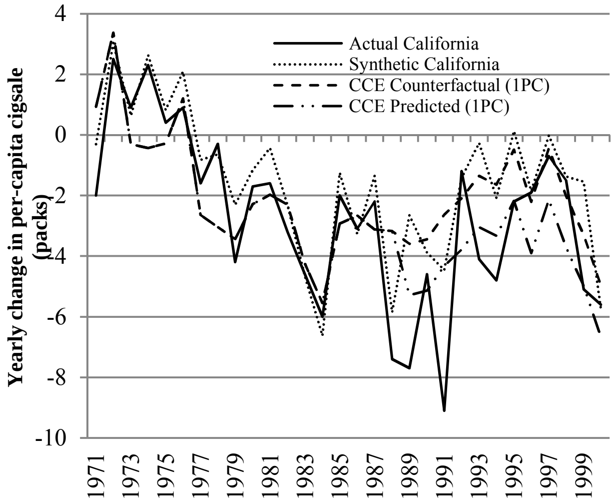
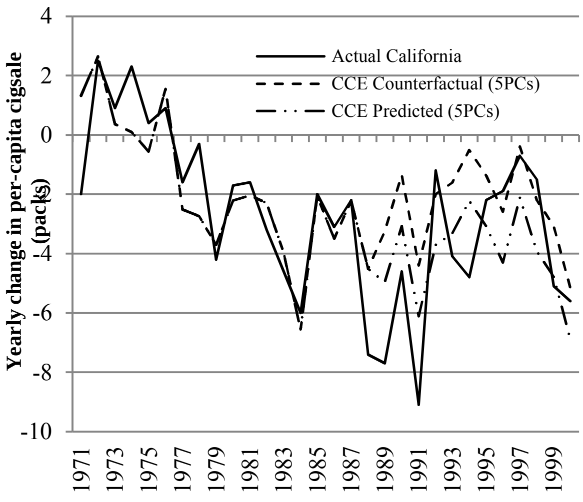


FIGURE 3. -- Actual level of cigarette sales versus predictions from synthetic control and CCEPC-DID methods.



(a) Synthetic control and CCEPC-DID (1PC).



(b) CCEPC-DID (5 PCs).

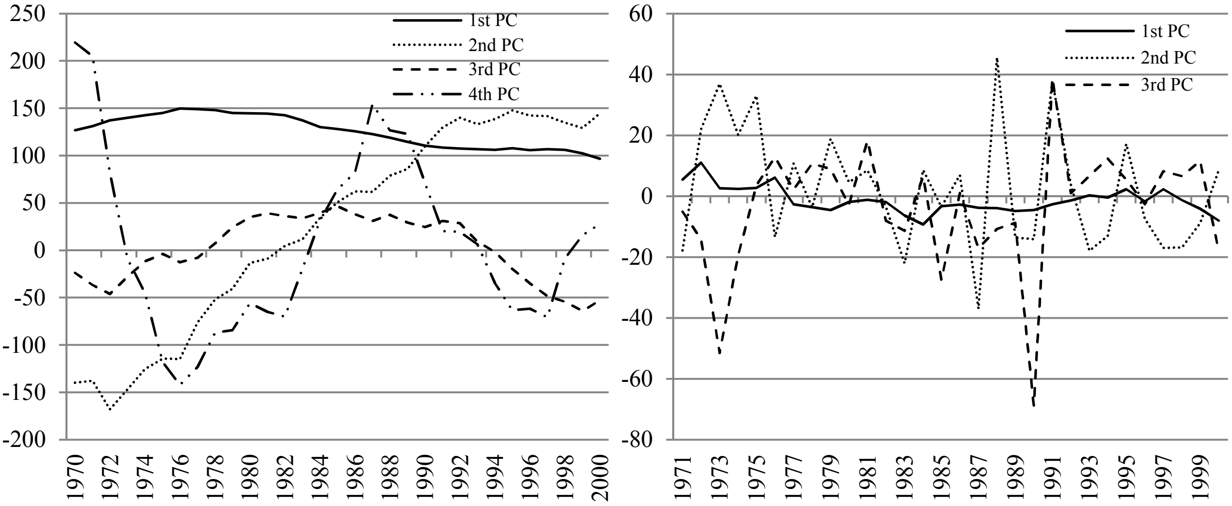
FIGURE 4. -- Actual yearly change of cigarette sales versus predictions from synthetic control and CCEPC-DID methods.

TABLE VIII

STATE WEIGHTS FROM THE CCEPC-DID ESTIMATOR (FIRST TWO PRINCIPAL COMPONENTS) [a]

	Level Data		Yearly Change Data			Level Data		Yearly Change Data	
	PC1	PC2	PC1	PC2		PC1	PC2	PC1	PC2
Alabama	0.024	0.464	0.016	0.000	Montana	0.023	-0.125	0.029	-0.027
Alaska	-	-	-	-	Nebraska	0.022	0.053	0.010	-0.020
Arizona	-	-	-	-	Nevada	0.034	-0.776	0.047	-0.505
Arkansas	0.025	0.522	0.003	0.261	New Hampshire	0.048	-0.984	0.113	-0.917
Colorado	0.024	-0.298	0.029	0.046	New Jersey	-	-	-	-
Connecticut	0.022	-0.110	-0.004	-0.107	New Mexico	0.019	-0.173	0.020	0.053
Delaware	0.031	0.224	0.005	-0.389	New York	-	-	-	-
D.C.	-	-	-	-	North Carolina	0.037	-0.651	0.089	0.243
Florida	-	-	-	-	North Dakota	0.022	-0.099	0.031	0.191
Georgia	0.026	0.250	0.009	-0.155	Ohio	0.026	0.291	0.006	-0.036
Hawaii	-	-	-	-	Oklahoma	0.026	0.101	0.033	-0.067
Idaho	0.022	-0.202	0.040	-0.453	Oregon	-	-	-	-
Illinois	0.025	-0.234	0.011	0.167	Pennsylvania	0.024	0.183	0.012	0.019
Indiana	0.031	0.286	0.044	0.259	Rhode Island	0.027	-0.206	0.029	0.917
Iowa	0.023	0.101	0.009	0.120	South Carolina	0.027	0.380	0.033	-0.104
Kansas	0.024	-0.009	0.023	0.242	South Dakota	0.022	0.156	0.023	0.101
Kentucky	0.041	0.385	0.073	0.311	Tennessee	0.026	0.633	0.015	-0.069
Louisiana	0.027	0.122	0.017	0.194	Texas	0.023	-0.186	0.012	-0.107
Maine	0.027	-0.036	0.019	0.011	Utah	0.014	-0.080	0.014	-0.172
Maryland	-	-	-	-	Vermont	0.030	-0.031	0.045	0.268
Massachusetts	-	-	-	-	Virginia	0.029	-0.019	0.038	-0.009
Michigan	-	-	-	-	Washington	-	-	-	-
Minnesota	0.022	0.005	-0.009	-0.432	West Virginia	0.025	0.323	0.030	0.073
Mississippi	0.024	0.355	0.025	0.039	Wisconsin	0.023	0.134	0.011	0.160
Missouri	0.028	0.370	0.015	0.075	Wyoming	0.029	-0.119	0.040	0.814

[a] For each principal component, all state weights sum to 1. PC1: first principal component; PC2: second principal component.



(a) Level data. (b) Yearly change data.

FIGURE 5. -- Principal components from the CCEPC-DID estimator, cigarette sales data.

TABLE IX

PROPORTION OF VARIANCE EXPLAINED BY PRINCIPAL COMPONENTS, CIGARETTE SALES DATA

	Level Data	Yearly Change Data
1st PC	99.9555	56.7611
2nd PC	0.0422	12.7650
3rd PC	0.0012	9.4671
4th PC	0.0006	4.9572
5th PC	0.0004	4.1041

a Only the first five principal components are reported. Numbers are expressed in percent.

APPENDIX TABLE A1

SMALL SAMPLE PROPERTIES OF ESTIMATORS, ADDITIONAL SCENARIOS[a]

Stationary: One AR(1) factor (ρ=0.9)

	Mean Bias				Empirical SD			
(T_0, T_1)	5, 5	10, 10	25, 25	50, 50	5, 5	10, 10	25, 25	50, 50
(N_C, N_I)								
CCEPC-DID (3PCs)								
5, 5	0.16	-0.15	-0.02	0.04	0.25	0.19	0.11	0.06
10, 10	-0.11	0.14	-0.02	0.12	0.18	0.12	0.06	0.04
25, 25	0.04	0.00	0.02	-0.08	0.12	0.07	0.04	0.03
50, 50	-0.04	-0.02	0.12	0.06	0.08	0.05	0.03	0.02
CCEPC-DID (1PC)								
5, 5	0.16	-0.15	-0.02	0.04	0.20	0.19	0.11	0.06
10, 10	-0.10	0.14	-0.02	0.12	0.15	0.11	0.06	0.04
25, 25	0.04	0.00	0.02	-0.08	0.09	0.07	0.04	0.03
50, 50	-0.04	-0.02	0.12	0.06	0.06	0.05	0.03	0.02
CCE-DID								
5, 5	0.15	-0.15	-0.02	0.04	0.14	0.12	0.07	0.04
10, 10	-0.10	0.14	-0.02	0.12	0.10	0.08	0.04	0.03
25, 25	0.04	0.00	0.02	-0.08	0.06	0.04	0.03	0.02
50, 50	-0.04	-0.02	0.12	0.06	0.05	0.03	0.02	0.01
Hsiao et al. (2012)								
5, 5		-0.16	-0.01	0.04		0.31	0.13	0.07
10, 10			-0.02	0.12			0.09	0.05
25, 25				-0.08				0.04
50, 50								
Bai (2009)								
5, 5	-0.16	-0.28	-0.07	0.02	0.40	0.64	0.59	0.49
10, 10	-0.39	-0.01	-0.07	0.11	0.43	0.62	0.58	0.53
25, 25	-0.26	-0.14	-0.04	-0.09	0.45	0.60	0.63	0.52
50, 50	-0.33	-0.15	0.06	0.05	0.45	0.60	0.64	0.54
DID								
5, 5	0.15	-0.15	-0.02	0.04	0.20	0.39	0.29	0.19
10, 10	-0.12	0.14	-0.02	0.13	0.23	0.37	0.24	0.26
25, 25	0.02	0.00	0.01	-0.07	0.25	0.32	0.30	0.25
50, 50	-0.05	-0.01	0.12	0.06	0.23	0.31	0.30	0.29

Nonstationary: One I(1) factor

	Mean Bias				Empirical SD			
(T_0, T_1)	5, 5	10, 10	25, 25	50, 50	5, 5	10, 10	25, 25	50, 50
(N_C, N_I)								
CCEPC-DID (3PCs)								
5, 5	0.17	-0.15	-0.02	0.03	0.25	0.21	0.13	0.09
10, 10	-0.11	0.13	-0.02	0.12	0.19	0.13	0.08	0.06
25, 25	0.04	0.00	0.02	-0.08	0.12	0.08	0.04	0.03
50, 50	-0.04	-0.02	0.12	0.06	0.09	0.06	0.03	0.02
CCEPC-DID (1PC)								
5, 5	0.17	-0.15	-0.01	0.03	0.21	0.20	0.13	0.09
10, 10	-0.11	0.13	-0.02	0.12	0.16	0.13	0.07	0.06
25, 25	0.04	0.00	0.02	-0.08	0.10	0.07	0.04	0.03
50, 50	-0.04	-0.02	0.12	0.06	0.07	0.05	0.03	0.02
CCE-DID								
5, 5	0.16	-0.15	-0.02	0.03	0.17	0.15	0.10	0.07
10, 10	-0.10	0.14	-0.02	0.12	0.11	0.09	0.06	0.05
25, 25	0.04	0.00	0.01	-0.08	0.08	0.06	0.04	0.03
50, 50	-0.04	-0.02	0.12	0.06	0.05	0.04	0.03	0.02
Hsiao et al. (2012)								
5, 5		-0.16	-0.01	0.03		0.36	0.18	0.12
10, 10			-0.02	0.12			0.13	0.09
25, 25				-0.08				0.06
50, 50								
Bai (2009)								
5, 5	-0.11	-0.26	-0.01	-0.06	0.55	1.11	1.41	1.99
10, 10	-0.34	0.01	-0.01	0.02	0.63	1.06	1.34	2.26
25, 25	-0.21	-0.12	0.02	-0.18	0.67	1.00	1.51	2.22
50, 50	-0.28	-0.13	0.13	-0.04	0.66	0.98	1.52	2.35
DID								
5, 5	0.17	-0.14	0.01	0.01	0.21	0.48	0.52	0.64
10, 10	-0.10	0.15	0.00	0.08	0.25	0.45	0.43	0.89
25, 25	0.05	0.01	0.04	-0.11	0.27	0.40	0.57	0.85
50, 50	-0.03	-0.01	0.14	0.02	0.25	0.38	0.56	0.98

[a] Number of replications = 1000.