SUPPLEMENT TO "FIXED EFFECTS DEMEANING IN THE PRESENCE OF INTERACTIVE EFFECTS IN TREATMENT EFFECTS REGRESSIONS AND ELSEWHERE"

Yana Petrova
Lund University

Joakim Westerlund*
Lund University
and
Deakin University

March 19, 2020

Abstract

This supplement provides (i) the formal assumptions discussed in Section 2 of the main paper, (ii) the asymptotic results described in the same section, and (iii) a Monte Carlo study.

A Assumptions

It is convenient to write the model given in equation (1) of the main paper in stacked form. It is given by

$$y_i = D\alpha_i + X_i\beta_i + F\gamma_i + \varepsilon_i, \tag{A.1}$$

where $y_i = [y_{i,1}, ..., y_{i,T}]'$ is $T \times 1$, $D = [D_1, ..., D_T]'$ is $T \times m$, $X_i = [X_{i,1}, ..., X_{i,T}]'$ is $T \times k$, $F = [F_1, ..., F_T]'$ is $T \times r$, and $\varepsilon_i = [\varepsilon_{i,1}, ..., \varepsilon_{i,T}]'$ is $T \times 1$.

The conditions that we will be working under are given in Assumptions ERR, RND, POS and MOM, which are similar to the conditions of Andrews (2005). The assumptions are stated

*Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.

in terms of cross-sectionally demeaned regressors. Let us therefore define $\overline{A} = N^{-1}\sum_{i=1}^{N} A_i$ and $A_i^* = A_i - \overline{A}$ for any matrix A_i. We also use $\mathcal{C}$ to denote the sigma-field generated by (D, F). Moreover, $\to_p$ and $\to_d$ signify convergence in probability and distribution, respectively. Finally, $\|A\| = \sqrt{\text{tr}\,(A'A)}$ denote the Frobenius (Euclidean) norm of any matrix A.

Assumption ERR.

(a) ε_i is conditionally independent across i given $\mathcal{C}$ with $E(\varepsilon_i|\mathcal{C}) = 0_{T\times 1}$.

(b) ε_i is conditionally uncorrelated with X_j given $\mathcal{C}$ for all i and j.

Assumption RND.

(a) $\beta_i = \beta + v_i$ and $\gamma_i = \gamma + \eta_i$, where v_i and η_i are conditionally independent across i given $\mathcal{C}$ with $E(v_i|\mathcal{C}) = 0_{k\times 1}$ and $E(\eta_i|\mathcal{C}) = 0_{r\times 1}$.

(b) v_i and η_j are conditionally uncorrelated of each other, as well as of X_n and ε_n given $\mathcal{C}$ for all i, j and n.

Before we make the next assumption, we need to introduce some additional notation. In particular, we define

$$\Psi = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(X_i^{*\prime} M_D X_i^*|\mathcal{C}), \tag{A.2}$$

$$R = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(X_i^{*\prime} M_D u_i u_i' M_D X_i^*|\mathcal{C}), \tag{A.3}$$

where $u_i = X_i v_i + F\eta_i + \varepsilon_i$ and $M_D = I_T - D(D'D)^{-1}D'$.

Assumption POS.

$$\widehat{\Psi} = \frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime} M_D X_i^* \to_p \Psi \tag{A.4}$$

as $N \to \infty$, where Ψ and R are positive definite almost surely (a.s.)

Assumption MOM. $E(\|M_D X_i^*\|^4) < \infty$ and $E(\|X_i^{*\prime} M_D u_i\|^4) < \infty$.

Assumptions ERR and RND are discussed in the Section 2 of the main paper. Assumption POS is a non-collinearity condition that rules out observed common factors that are included in both $X_{i,t}$ and D_t. The reason for distinguishing between D_t and $X_{i,t}$ is that while β_i is subject to the random coefficient condition in Assumption RND, α_i is not. Hence, unlike β_i and γ_i, α_i is not restricted in any way, but can be arbitrarily correlated with $X_{i,t}$. The "price" of this generality is that we cannot infer α_i, as D_t will be projected out prior to the estimation of β. This is also the reason for why the regressors in $X_{i,t}$ cannot be constant in i. We do, however, allow regressors in $X_{i,t}$ that are constant in t.

Assumption MOM is a high-level moment condition that is needed to establish both asymptotic mixed normality of the FE estimator and consistency of the estimated covariance matrix. Unlike in the bulk of the previous literature (see, for example, Chudik et al., 2011, and Pesaran, 2006), we do not require that v_i, η_i and $\varepsilon_{i,t}$ are independent of $X_{i,t}$ but only that they are uncorrelated with $X_{i,t}$. This is why Assumption MOM is stated in terms of $X_i^{*\prime} M_D u_i$. Under independence, Assumption MOM holds provided that v_i, η_i and $\varepsilon_{i,t}$ all have finite fourth-order moments.

A major difference when compared to the bulk of the existing large-T literature is that here we place no assumptions on the time series properties of F_t, D_t, $X_{i,t}$ and $\varepsilon_{i,t}$. Consider F_t. A standard assumption in the literature is that the limit of $T^{-1} \sum_{t=1}^{T} F_t F_t'$ is positive definite (see, for example, Bai, 2009, and Moon and Weidner, 2015), which rules out many empirically relevant cases, such as when F_t is trending. The assumptions considered here are more general in this regard and do not place any restrictions on the process generating F_t, which can be both deterministic and stochastic. The number of factors, r, is also not restricted in any way, which is quite different from the bulk of the existing literature where r is typically assumed to be known or accurately estimated (see Bai, 2009). The only restriction we make is that F_t must be independent of $\varepsilon_{i,t}$, v_i and η_i, which is standard in the literature. We similarly do not make any assumptions regarding the persistence of $X_{i,t}$ and $\varepsilon_{i,t}$.

B Asymptotic results

In this section, we begin by showing that the FE estimator is consistent and asymptotically normal as $N \to \infty$ for a fixed T, provided that Assumptions ERR, RND, POS and MOM are met. We then show that the same applies to PC.

The point that FE works in the presence of interactive effects has been made before by Cui et al. (2019), Kapetanios et al. (2019), and Westerlund (2019), but their results require that T is large. Gobillon and Magnac (2016) also comment on this possibility, but do not provide any formal results. Sarafidis and Wansbeek (2012) report Monte Carlo results showing that FE works well if the factor loadings are uncorrelated with the regressors.[1] However, no analytical results are provided. Andrews (2005) considers a pure cross-sectional model with common factors, which he estimates using OLS. According to the results, the estimator is consistent and asymptotically normal provided that the errors and regressors are uncorrelated. The author comments on the panel data case, but does not provide any results. Forchini and Peng (2016) consider a fixed-T panel data regression model that is similar to ours, which is again estimated using OLS. However, they require that the regressors have a factor structure, which is not necessary here.

While obviously related, the above cited work has different focus areas. In the present paper, we focus on fixed effects demeaning as a general, and empirically very attractive, device to increase the robustness not only of OLS but also of other estimation approaches, such as PC, a point that has been largely overlooked in the previous literature. Of course, in practice fixed effects are almost always included, and so our recommendation to demean is not very controversial but just supports the common practice. This is true when using OLS, but also when using PC. As pointed out in the main paper, PC does not require demeaning. In spite of this, demeaning is fairly common also in PC (see, for example, Gobillon and Magnac (2016), and Moon and Weidner, 2015). The reason is that if fixed effects are a part of the interactive effects, demeaning reduces the number of factors that has to be estimated. Hence, regardless of

[1] Similarly, Sarafidis and Robertson (2009) argue that demeaning can be useful to reduce the bias in GMM estimation of dynamic panel data models, and report some confirmatory Monte Carlo results.

whether one is using OLS or PC, demeaning is quite standard. Our fixed-T results complement the existing large-T results, and imply that demeaning is useful regardless of the size of T.

B.1 The FE estimator

The FE estimator described in the main paper is simply the OLS estimator applied to the data after subtracting the cross-sectional averages and is given by

$$\widehat{\beta}_{FE} = \left(\sum_{i=1}^{N} X_i^{*\prime} M_D X_i^* \right)^{-1} \sum_{i=1}^{N} X_i^{*\prime} M_D y_i^*. \tag{B.5}$$

Theorem B.1 below reports the asymptotic distribution of $\sqrt{N}(\widehat{\beta}_{FE} - \beta)$ under Assumptions ERR, RND, POS and MOM.

Theorem B.1. *Under Assumptions ERR, RND, POS and MOM, as $N \to \infty$,*

$$\sqrt{N}(\widehat{\beta}_{FE} - \beta) \to_d MN(0_{k\times 1}, \Psi^{-1} R \Psi^{-1}),$$

where $MN(\cdot, \cdot)$ signifies a mixed normal distribution.

Proof: By using the fact that $\beta_i = \beta + v_i$ and $\gamma_i = \gamma + \eta_i$, we obtain

$$y_i^* = D\alpha_i^* + X_i^* \beta + (X_i v_i)^* + F\eta_i^* + \varepsilon_i^* = D\alpha_i^* + X_i^* \beta + u_i^*, \tag{B.6}$$

where $u_i = X_i v_i + F\eta_i + \varepsilon_i$, which in turn implies

$$\begin{aligned} \sqrt{N}(\widehat{\beta}_{FE} - \beta) &= \left(\frac{1}{N} \sum_{i=1}^{N} X_i^{*\prime} M_D X_i^* \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i^{*\prime} M_D u_i^* \\ &= \widehat{\Psi}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i^{*\prime} M_D u_i - \widehat{\Psi}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i^{*\prime} M_D \overline{u} \\ &= \widehat{\Psi}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i, \end{aligned} \tag{B.7}$$

where $\xi_i = X_i^{*\prime} M_D u_i$, and the last equality is due to the fact that $\sum_{i=1}^{N} X_i^{*\prime} = 0_{k\times T}$. This last result is the key and provides intuition for why $\widehat{\beta}_{FE}$ works even though the interactive effects are misspecified, which is the same as in the large-T case considered by Westerlund (2019). We have assumed that the random components of β_i, γ_i and ε_i are all independent over i, mean zero, and uncorrelated with X_j for all i and j. The mean of β_i is the parameter of interest. Hence, if we can just eliminate γ it should be possible to exploit the assumed independence over i, and to obtain an asymptotic normal distribution for $\sqrt{N}(\widehat{\beta}_{FE} - \beta)$. One way to accomplish this goal is to demean X_i.

Let $\mathcal{F}_i$ be the sigma-field generated by $\mathcal{C}$ and $(\xi_1, ..., \xi_i)$. Then $\{(\xi_i, \mathcal{F}_i) : i \geq 1\}$ is a martingale difference sequence (MDS), because ξ_i is independent across i conditional on $\mathcal{C}$, and $E(\xi_i|\mathcal{F}_{i-1}) = E(\xi_i|\mathcal{C}) = 0_{k\times 1}$ (see, for example, Andrews, 2005, for a similar MDS construction). A conditional Lindeberg condition holds because ξ_i have four finite moments. Hence, letting

$$
\begin{aligned}
R &= \lim_{N\to\infty} E\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i\right)'\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(X_i^{*\prime} M_D u_i u_i' M_D X_i^* |\mathcal{C}), \qquad \text{(B.8)}
\end{aligned}
$$

by the MDS CLT given in Proposition A.1 of Magdalinos and Phillips (2009),

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i \to_d MN(0_{k\times 1}, R) \qquad \text{(B.9)}
$$

as $N \to \infty$, where $\to_d$ and $MN(\cdot,\cdot)$ signify convergence in distribution and a mixed normal distribution, respectively. Theorem B.1 is a direct consequence of this result and the fact that $\widehat{\Psi} \to_p \Psi$ as $N \to \infty$ by Assumption POS. ■

While the asymptotic distribution of $\sqrt{N}(\widehat{\beta}_{FE} - \beta)$ is conditional on $\mathcal{C}$, it is not difficult to show that consistency holds unconditionally. Indeed, by a law of large numbers for independent processes (see, for example, Lemma 1 of Andrews, 2005),

$$
\frac{1}{N}\sum_{i=1}^{N}\xi_i \to_p E(\xi_i|\mathcal{C}) = 0_{k\times 1} \qquad \text{(B.10)}
$$

as $N \to \infty$. Hence, since $\|\widehat{\Psi}^{-1}\| = O_p(1)$,

$$\|\widehat{\beta}_{FE} - \beta\| \leq \|\widehat{\Psi}^{-1}\| \left\| \frac{1}{N} \sum_{i=1}^{N} \xi_i \right\| = o_p(1). \tag{B.11}$$

As pointed out in the main paper, inference based on Theorem B.1 requires a consistent estimator of $\Psi^{-1} R \Psi^{-1}$. The estimator of Ψ is obviously given by $\widehat{\Psi}$. For the estimation of R, we use

$$\widehat{R} = \frac{1}{N} \sum_{i=1}^{N} X_i^{*\prime} M_D \widehat{u}_i \widehat{u}_i' M_D X_i^*, \tag{B.12}$$

where $\widehat{u}_i = M_D(y_i^* - X_i^* \widehat{\beta}_{FE})$. Theorem B.2 shows that $\widehat{\Psi}^{-1} \widehat{R} \widehat{\Psi}^{-1}$ is a consistent estimator for $\Psi^{-1} R \Psi^{-1}$.

Theorem B.2. *Under the conditions of Theorem B.1, as $N \to \infty$,*

$$\widehat{\Psi}^{-1} \widehat{R} \widehat{\Psi}^{-1} \to_p \Psi^{-1} R \Psi^{-1}.$$

Proof: From (B.6),

$$\begin{aligned} \widehat{u}_i &= M_D(y_i^* - X_i^* \widehat{\beta}_{FE}) = M_D[(X_i v_i)^* + F\eta_i^* + \varepsilon_i^* - X_i^* (\widehat{\beta}_{FE} - \beta)] \\ &= M_D[u_i - \overline{u} - X_i^* (\widehat{\beta}_{FE} - \beta)] = M_D u_i + O_p(N^{-1/2}), \end{aligned} \tag{B.13}$$

where the last equality holds because of Theorem B.1 and the fact that $\|\overline{u}\| = O_p(N^{-1/2})$ by the same MDS CLT arguments used in Proof of Theorem B.1. It follows that

$$\widehat{R} = \frac{1}{N} \sum_{i=1}^{N} X_i^{*\prime} M_D \widehat{u}_i \widehat{u}_i' M_D X_i^* = \frac{1}{N} \sum_{i=1}^{N} X_i^{*\prime} M_D u_i u_i' M_D X_i^* + O_p(N^{-1/2}) \to_p R \tag{B.14}$$

as $N \to \infty$. Hence, since $\widehat{\Psi} \to_p \Psi$ by Assumption POS, we can show that

$$\widehat{\Psi}^{-1} \widehat{R} \widehat{\Psi}^{-1} \to_p \Psi^{-1} R \Psi^{-1}, \tag{B.15}$$

and so we are done. ■

A major point about Theorem B.2 is that the covariance matrix of $\widehat{\beta}_{FE}$ is very easily estimable. This stands in sharp contrast to the large-T framework that typically involves some kind of heteroskedasticity and autocorrelation consistent (HAC) correction (see Bai, 2009), which is not only difficult to implement but is also known to lead to poor small-sample properties.

We now put Theorem B.2 to work in testing the null hypothesis $H_0 : H\beta = h$, where H is a $g \times k$ matrix of rank $g \leq k$ and h is a $g \times 1$ vector. Consider the Wald test statistic

$$W = N(H\widehat{\beta}_{FE} - h)'(H\widehat{\Psi}^{-1}\widehat{R}\widehat{\Psi}^{-1}H')^{-1}(H\widehat{\beta}_{FE} - h). \tag{B.16}$$

Suppose that H_0 is true. Then, because of the consistency of $\widehat{\Psi}^{-1}\widehat{R}\widehat{\Psi}^{-1}$ (Theorem B.2) and the asymptotic normality of $\sqrt{N}(H\widehat{\beta}_{FE} - h)$ under H_0 (Theorem B.1), we can show that

$$W = \sqrt{N}(H\widehat{\beta}_{FE} - h)'(H\Psi^{-1}R\Psi^{-1}H')^{-1}\sqrt{N}(H\widehat{\beta}_{FE} - h) + o_p(1) \to_d \chi^2(g) \tag{B.17}$$

as $N \to \infty$. If $g = 1$, then we can similarly show that

$$t = \frac{\sqrt{N}(H\widehat{\beta}_{FE} - h)}{\sqrt{H\widehat{\Psi}^{-1}\widehat{R}\widehat{\Psi}^{-1}H'}} = \frac{\sqrt{N}(H\widehat{\beta}_{FE} - h)}{\sqrt{H\Psi^{-1}R\Psi^{-1}H'}} + o_p(1) \to_d N(0,1) \tag{B.18}$$

as $N \to \infty$ under H_0.

B.2 The PC estimator

As mentioned in the main paper, the results of Section B.1 are not unique to the FE estimator but apply to all estimators of the same basic form and where the regressors satisfy the conditions that we here place on X_i. Let us now illustrate this using the PC estimator of Bai (2009). While not necessary, it is convenient to assume that r is known and that it is not larger than T, as this will allow us to invoke some of the results of Bai (2009).

The PC estimator of β is defined in equation (11) of Bai (2009). The demeaned version of this estimator is given by

$$\widehat{\beta}_{PC} = \left(\sum_{i=1}^{N} X_i^{*\prime} M_W X_i^* \right)^{-1} \sum_{i=1}^{N} X_i^{*\prime} M_W y_i^*. \tag{B.19}$$

where $W = [D, \widehat{F}]$ with $\widehat{F}$ being an $T \times r$ matrix of estimated PC factors. The definition of $\widehat{F}$ is given in equation (12) of Bai (2009), which in our case reads

$$\frac{1}{NT} \sum_{i=1}^{N} M_D (y_i - X_i^* \widehat{\beta}_{PC})(y_i - X_i^* \widehat{\beta}_{PC})' M_D \widehat{F} = \widehat{F} V_N, \tag{B.20}$$

where V_N is a diagonal matrix that consists of the r largest eigenvalues of the $T \times T$ matrix $(NT)^{-1} \sum_{i=1}^{N} M_D (y_i - X_i^* \widehat{\beta}_{PC})(y_i - X_i^* \widehat{\beta}_{PC})' M_D$, arranged in decreasing order.

We begin by analyzing $\widehat{F}$. By using the same arguments as in Proof of Proposition 1 of Bai (2009), we can show that $\|\beta - \widehat{\beta}_{PC}\| = o_p(1)$. By using this and the fact that

$$M_D (y_i - X_i^* \widehat{\beta}_{PC}) = M_D X_i^* (\beta - \widehat{\beta}_{PC}) + M_D u_i^* \tag{B.21}$$

by (B.6), we can show that

$$\begin{aligned} \widehat{F} V_N &= \frac{1}{NT} \sum_{i=1}^{N} M_D (y_i - X_i^* \widehat{\beta}_{PC})(y_i - X_i^* \widehat{\beta}_{PC})' M_D \widehat{F} \\ &= \frac{1}{NT} \sum_{i=1}^{N} M_D u_i^* u_i^{*\prime} M_D \widehat{F} + o_p(1) \\ &= Q \widehat{F} + o_p(1), \end{aligned} \tag{B.22}$$

where

$$Q = \lim_{N \to \infty} \frac{1}{NT} \sum_{i=1}^{N} E(M_D u_i u_i' M_D | \mathcal{C}). \tag{B.23}$$

This is a new eigenvalue-eigenvector relation, where asymptotically each column of $\widehat{F}$ is an eigenvector of Q. The columns of $\widehat{F}$ are therefore asymptotically equal to the first r eigenvectors associated with the first r largest eigenvalues of Q, which are the limits of the eigenvalues that sit on the main diagonal of V_N. Let F^0 be the limit of $\widehat{F}$. It follows that

$$\|\widehat{F}V_N - QF^0\| = o_p(1). \tag{B.24}$$

Since $P_{\widehat{F}} = P_{\widehat{F}V_N}$, similarly to the proof of Lemma A.7 of Bai (2009), this last result implies

$$\|M_{\widehat{F}} - M_{QF^0}\| = \|P_{\widehat{F}} - P_{QF^0}\| = \|P_{\widehat{F}V_N} - P_{QF^0}\| = o_p(1), \tag{B.25}$$

and so we obtain

$$\|M_W - M_{W^0}\| = o_p(1), \tag{B.26}$$

where $W^0 = [D, QF^0]$. It is important to note that $\widehat{F}$ is not consistent for (the space spanned by) F, as this requires T to be large. It is, however, asymptotically uncorrelated of v_i, η_i and ε_i, which is enough for our purposes, as we will now demonstrate.

Let us now consider $\widehat{\beta}_{PC}$. Making use of (B.6),

$$\sqrt{N}(\widehat{\beta}_{PC} - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime} M_W X_i^*\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{*\prime} M_W u_i^*. \tag{B.27}$$

By adding and subtracting, the numerator can be written as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{*\prime} M_W u_i^* = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{*\prime} M_{W^0} u_i^* + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{*\prime}(M_W - M_{W^0}) u_i^*. \tag{B.28}$$

Denote by X_i^j the j-th column of X_i . In this notation, the j-th row of the last term on the right can be written as

$$\begin{aligned} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{j*\prime}(M_W - M_{W^0}) u_i^* &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \text{tr}\left[(M_W - M_{W^0}) u_i^* X_i^{j*\prime}\right] \\ &\leq \text{tr}\left[(M_W - M_{W^0})^2\right]^{1/2} \text{tr}\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i^* X_i^{j*\prime}\right)^2\right]^{1/2} \\ &= o_p(1), \end{aligned} \tag{B.29}$$

where the inequality holds because $[\text{tr}\,(A'B)]^2 \leq \text{tr}\,(A'A)\text{tr}\,(B'B)$, and the last equality is due to $\text{tr}\,[(M_W - M_{W^0})^2] = \|M_W - M_{W^0}\| = o_p(1)$. It follows that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{*\prime} M_W u_i^* = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{*\prime} M_{W^0} u_i^* + o_p(1). \tag{B.30}$$

We can similarly show that

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime} M_W X_i^* &= \frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime} M_{W^0} X_i^* + \frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime}(M_W - M_{W^0}) X_i^* \\ &= \frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime} M_{W^0} X_i^* + o_p(1),\end{aligned} \tag{B.31}$$

which in turn implies

$$\sqrt{N}(\widehat{\beta}_{PC} - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime} M_{W^0} X_i^*\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^{*\prime} M_{W^0} u_i^* + o_p(1). \tag{B.32}$$

As alluded to in the above, $M_{W^0} X_i$ is uncorrelated of v_i, η_i and ε_i. Hence, if we assume that

$$\frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime} M_{W^0} X_i^* \to_p \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(X_i^{*\prime} M_{W^0} X_i^* | \mathcal{C})$$

as $N \to \infty$, where the limiting matrix is positive definite a.s., such that Assumption POS holds, then all the conditions of Section A are met. The asymptotic results reported in Section B.1 for FE therefore apply also to PC.

B.3 The Wald test for uncorrelated coefficients

As discussed in the main paper, the Wald test given in (B.16) can be used to test the Assumption RND (b) requirement that errors v_i and η_i should be uncorrelated with $X_{i,t}$. In the literature it is very common to assume that any correlation is driven by $\overline{X}_i = T^{-1}\sum_{t=1}^{T} X_{i,t}$ (see Hsiao, 2003, chapter 4.3, for a detailed discussion), and therefore so shall we. The formal conditions are stated in Assumptions TEST, POS′ and MOM′ below, where $Z_i = [M_D X_i^*, (\overline{X}_i' \otimes M_D X_i)^*, (\overline{X}_i^{*\prime} \otimes \widehat{F})]$ and $\widehat{F} = M_D(\overline{y} - \overline{X}\widehat{\beta}_{FE})$.

Assumption TEST.

(a) $\beta_i = \beta + v_i$ and $\gamma_i = \gamma + \eta_i$ with $\gamma \neq 0_{r\times 1}$ and

$$\begin{bmatrix} v_i \\ \eta_i \end{bmatrix} = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} \overline{X}_i + \begin{bmatrix} w_i \\ z_i \end{bmatrix},$$

where Λ_1 and Λ_2 are $k \times k$ and $r \times k$, respectively.

(b) w_i and z_i are conditionally independent across i given $\mathcal{C}$ with $E(w_i|\mathcal{C}) = 0_{k\times 1}$ and $E(z_i|\mathcal{C}) = 0_{r\times 1}$.

(c) w_i and z_i are conditionally uncorrelated of each other as well as of $\overline{X}_j$ and ε_j given $\mathcal{C}$ for all i and j.

Let

$$\Phi = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(Z_i^{*\prime} Z_i^{*}|\mathcal{C}), \tag{B.33}$$

$$S = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(Z_i^{*\prime} M_D e_i e_i' M_D Z_i^{*}|\mathcal{C}), \tag{B.34}$$

where $e_i = X_i w_i + F z_i + \varepsilon_i$.

Assumption POS′.

$$\widehat{\Phi} = \frac{1}{N}\sum_{i=1}^{N} Z_i^{*\prime} Z_i^{*} \to_p \Phi \tag{B.35}$$

as $N \to \infty$, where Φ and S are positive definite a.s.

Assumption MOM′. $E(\|Z_i^{*}\|^4) < \infty$ and $E(\|Z_i^{*\prime} M_D e_i\|^4) < \infty$.

The null hypothesis of interest is given by $H_0 : \Lambda_1 = 0_{k\times k}$ and $\Lambda_2 = 0_{r\times k}$, which can be tested using the following version of the Wald test considered in Section B.1 of this supplement and that is described in Section 2 of the main paper:

$$W_{RND} = N(H_{RND}\widehat{\theta}_{FE})'(H_{RND}\widehat{\Phi}^{-1}\widehat{S}\widehat{\Phi}^{-1}H'_{RND})^{-1}H_{RND}\widehat{\theta}_{FE}, \tag{B.36}$$

where $H_{RND} = [0_{(k+1)k\times k}, I_{(k+1)k}]$ and $\widehat{\theta}_{FE}$ is the FE slope estimator in a regression of $M_D y_i^*$ onto Z_i. Also,

$$\widehat{S} = \frac{1}{N}\sum_{i=1}^{N} Z_i' M_D \widehat{e}_i \widehat{e}_i' M_D Z_i, \tag{B.37}$$

where $\widehat{e}_i = M_D(y_i^* - Z_i^*\widehat{\theta}_{FE})$. We would like to point out here that, despite the notation, $\widehat{F}$ is not intended as an estimator of F. In fact, $\widehat{F}$ may not even have the same dimension as F. However, we can show that asymptotically $\widehat{F}$ is going to be highly correlated with $F\gamma$, which is enough to ensure that the test is consistent.

Theorem B.3. *Suppose that Assumptions ERR, POS' and TEST hold. Then, under H_0, as $N \to \infty$,*

$$W_{RND} \to_d \chi^2[(k+1)k].$$

Proof: Assumption TEST (a) can be inserted into (B.6), giving

$$\begin{aligned} y_i^* &= D\alpha_i^* + X_i^*\beta + (X_i v_i)^* + F\eta_i^* + \varepsilon_i^* \\ &= D\alpha_i^* + X_i^*\beta + (X_i\Lambda_1\overline{X}_i)^* + F\Lambda_2\overline{X}_i^* + u_i^* \\ &= D\alpha_i^* + X_i^*\beta + (\overline{X}_i' \otimes X_i)^*\lambda_1 + (\overline{X}_i^{*\prime} \otimes F)\lambda_2 + e_i^*, \end{aligned} \tag{B.38}$$

where $\lambda_1 = \operatorname{vec}\Lambda_1$ and $\lambda_2 = \operatorname{vec}\Lambda_2$. Because w_i and z_i have exactly the same properties as v_i and η_i under uncorrelatedness, e_i^* will behave just as u_i^* in (B.6). We can, therefore, think of (B.38) as an augmented version of (B.6) with $(\overline{X}_i' \otimes X_i)^*$ and $(\overline{X}_i^{*\prime} \otimes F)$ as additional regressors. The problem is that F is unobserved, which means that (B.38) is not really feasible. In order to account for this, note that in analogy to (B.13),

$$M_D(y_i - X_i\widehat{\beta}_{FE}) = M_D[u_i - X_i(\widehat{\beta}_{FE} - \beta)] = M_D(u_i - X_i\widehat{\Psi}^{-1}\overline{\xi}), \tag{B.39}$$

implying that

$$\widehat{F} = M_D(\overline{y} - \overline{X}\widehat{\beta}_{FE}) = M_D(\overline{u} - \overline{X}\widehat{\Psi}^{-1}\overline{\xi}), \tag{B.40}$$

which shows that $\widehat{F}$ is correlated with F, unless of course $\overline{\gamma} = 0_{r\times 1}$, which is ruled out by Assumption TEST (a).

Let $Z_i = [M_D X_i^*, (\overline{X}_i' \otimes M_D X_i)^*, (\overline{X}_i^{*\prime} \otimes \widehat{F})]$ and $\theta = [\beta', \lambda_1', \lambda_2']'$. Under H_0 the FE estimator $\widehat{\theta}_{FE}$ of θ can be written as

$$\begin{aligned}\sqrt{N}(\widehat{\theta}_{FE} - \theta) &= \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' Z_i\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i M_D e_i^* \\ &= \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' Z_i\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i M_D e_i,\end{aligned} \tag{B.41}$$

where the first equality is due to the fact that $\Lambda_2 = 0_{r\times k}$ under H_0, while second equality is due to $\sum_{i=1}^{N} Z_i = 0_{T\times(k+2)k}$. As for the remaining term on the right-hand side, by the MDS CLT (see Proof of Theorem B.1),

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i M_D e_i \to_d MN(0_{(k+2)k\times 1}, S) \tag{B.42}$$

as $N \to \infty$, where

$$S = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(Z_i' M_D e_i e_i' M_D Z_i | \mathcal{C}).$$

It follows that if we let

$$
\begin{aligned}
\widehat{\Phi} &= \frac{1}{N}\sum_{i=1}^{N} Z_i' Z_i, \\
\Phi &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(Z_i' M_D Z_i|\mathcal{C}),
\end{aligned}
$$

then, under H_0,

$$
\sqrt{N}(\widehat{\theta}_{FE} - \theta) = \widehat{\Phi}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i M_D e_i \to_d MN(0_{(k+2)k\times 1}, \Phi^{-1} S \Phi^{-1}). \tag{B.43}
$$

Let $\widehat{e}_i = M_D(y_i^* - Z_i^*\widehat{\theta}_{FE})$. By using the same steps as in Proof of Theorem B.2, it is not difficult to show that under H_0,

$$
\widehat{S} = \frac{1}{N}\sum_{i=1}^{N} Z_i' M_D \widehat{e}_i \widehat{e}_i' M_D Z_i \to_p S. \tag{B.44}
$$

Hence, letting $H_{RND} = [0_{(k+1)k\times k}, I_{(k+1)k}]$, we have that under H_0,

$$
\begin{aligned}
W_{RND} &= N(H_{RND}\widehat{\theta}_{FE})'(H_{RND}\widehat{\Phi}^{-1}\widehat{S}\widehat{\Phi}^{-1}H_{RND}')^{-1}H_{RND}\widehat{\theta}_{FE} \\
&= (\sqrt{N}H_{RND}\widehat{\theta}_{FE})'(H_{RND}\Phi^{-1}S\Phi^{-1}H_{RND}')^{-1}\sqrt{N}H_{RND}\widehat{\theta}_{FE} + o_p(1) \\
&\to_d \chi^2[(k+1)k]
\end{aligned} \tag{B.45}
$$

as $N \to \infty$. ■

Theorem B.3 confirms that asymptotically W_{RND} is correctly sized under the null hypothesis. Under the alternative hypothesis $\Lambda_1 \neq 0_{k\times k}$ or $\Lambda_2 \neq 0_{r\times k}$, or both, in which case $\sqrt{N}(\widehat{\theta}_{FE} - \theta) = O_p(\sqrt{N})$ implying $W_{RND} = O_p(N)$. The power of the test therefore approaches one as $N \to \infty$. It is therefore consistent. It is important to point out, thought, that this supposes that the Assumption TEST is satisfied, so that the correlation between v_i and η_i on the one hand and $X_{i,t}$ on the other hand is in fact driven by $\overline{X}_i$. A more general approach would be to test if v_i and η_i are uncorrelated with $X_{i,1}, ..., X_{i,T}$ (see Hsiao, 2003, chapter 4.3). However, this approach is feasible only if T is really small, and unreported Monte Carlo results confirm that the performance depends critically on T. In this paper, we therefore focus on $\overline{X}_i$, which is a restriction. Moreover, any dependence on $\overline{X}_i$ must be linear, for otherwise the test is unlikely to detect it.

C Monte Carlo study

This section presents our Monte Carlo results, which are divided into three parts. The full set of results is very large. In Section C.1, we therefore present a small but representative subset of the results for the case when $X_{i,t}$ is a treatment dummy. If the purpose is to just get a feeling for the results, it is enough to read this section. Sections C.2 and C.3 contain additional results. In particular, while Section C.2 report some results that where omitted from Section C.1, Section C.3 presents results for a model with more general types of regressors.

C.1 Main results

We consider the same three estimators as in Section 3 of the main paper, which are implemented in exactly the same way as in that same section.[2] The data generating process is similar to the one used by Gobillon and Magnac (2016), and can be seen as a restricted version of (A.1) that sets $\beta = D_t = r = 1$, $\alpha_i \sim N(1,1)$, $v_i \sim N(0,1)$ and $X_{i,t} = B_i C_t$, where $B_i = 1(i \leq N_0)$ and $C_t = 1(t \geq T_0)$. Two specifications of N_0 are considered; $N_0 = 0.1N$, which reflects the empirical illustration in Section 3 of the main paper, and $N_0 = 0.5N$, as in, for example, Friedberg (1998), Kim and Oka (2014), and Wolfers (2006). In both cases, $T_0 = 0.5T$. Also, following studies such as Chudik et al. (2011), Kapetanios et al. (2011), and Pesaran (2006), $\varepsilon_{i,t}$ is allowed to be both serially correlated and heteroskedastic through the following autoregressive

[2]In their Monte Carlo study, Gobillon and Magnac (2016) assume that the number of factors are known, which is never the case in practice. By contrast, the use of information criteria to select the number of common factors is very common in the empirical literature. Our use of the CP criterion of Bai (2009) reflects this. The results reported here should therefore be highly relevant for applied work.

(AR) specification:

$$\varepsilon_{i,t} = \rho \varepsilon_{i,t-1} + u_{i,t}, \tag{C.46}$$

where $\rho = 0.5$, $u_{i,t} \sim N(0, \sigma_i^2)$, $\sigma_i^2 \sim U(1,2)$ and $\varepsilon_{i,0} = 0$. While we do not comment on this in the main paper, the empirical results suggest that the estimated factors can be well described by highly persistent AR processes. Motivated in part by this, in part by existing Monte Carlo studies (see, for example, Chudik and Pesaran, 2015, Moon and Weidner, 2015, and Pesaran, 2006), F_t is generated as

$$F_t = (1 - \phi) + \phi F_{t-1} + e_t, \tag{C.47}$$

where $\phi = 0.8$, $e_t \sim N(0,1)$ and $F_0 = 0$. As for γ_i, similarly to Gobillon and Magnac (2016), we set $\gamma_i = \gamma + \mu \cdot 1(i \leq N_0) + N(0,1)$, where $\gamma = 1$ and $\mu \in \{0,1\}$ determines the size of the break in the mean. If $\mu = 0$, there is no break and as a consequence γ_i is independent of $X_{i,t}$. If, on the other hand, $\mu = 1$, the mean is breaking, and therefore γ_i is correlated with $X_{i,t}$ and the correlation between the two is 0.3.

We focus on the bias, the root mean square error (RMSE), and the 5% size of a double-sided t-test for testing $\beta = 1$. We also report 5% rejection frequencies for the Wald test for uncorrelated loadings. When $\mu = 0$, these rejection frequencies represent size, whereas when $\mu = 1$, they represent power. All results are based on 1,000 replications of samples with $T \in \{6, 10, 20\}$ and $N \in \{30, 50, 100, 200\}$.

Tables 1 and 2 report the results for the case without and with a break in the mean of the loadings, respectively. We begin by considering the former set of results. The first thing to note about Table 1 is the poor performance of the PC estimator based on raw data. The bias is generally at or above one, which means that the bias as a percentage of the size of the true coefficient ($\beta = 1$) is close to 100%. The bias and RMSE do come down with increases in N and T, as expected given the existing joint limit theory of the PC estimator (see Bai, 2009, and Moon and Weidner, 2015), albeit only very slowly. Demeaning removes most of this poor performance.

Indeed, the demeaned PC estimator is essentially unbiased and the RMSE decreases very fast with increases in T and especially in N. As expected, the FE estimator also performs well. In fact, the performance of FE is almost indistinguishable from that of demeaned PC. The only notable difference is that the RMSE is generally slightly higher for PC than for FE, although the difference gets smaller as N and T grow.

The fact that PC and FE tend to perform very similarly only under demeaning suggests that it is not the augmentation by the estimated PC factors that drives the results, but rather the demeaning. The intuition is that if T is small, accurate estimation of the factors is not possible, and therefore the factor estimation error has a dominating effect. On the other hand, we know from Section B that demeaning works even if T is small, provided that the uncorrelatedness condition is met. In this case, the estimated PC factors are just redundant regressors, whose inclusion should be asymptotically irrelevant, although in small samples it is expected to lead to variance inflation.

Another difference between demeaned PC and FE is that the size distortions are generally much higher for the former estimator than for the latter. We also see that while decreasing in T, for PC the distortions have a tendency to accumulate and to become very serious as N increases. This is in agreement with the results reported by Moon and Weidner (2015) (see also Chudik et al., 2011), who show that T as large as 300 may be needed for the distortions of PC to go away. The FE results look much better. In particular, while there are some distortions among the smaller values of N when $N_0 = 0.1N$, these disappear very quickly as N increases, and when $N_0 = 0.5N$ size accuracy is almost perfect.[3] These results for the FE-based t-test are reflected also in the Wald test for uncorrelated loadings, which generally performs well, except when $N_0 = 0.1N$ and N is small.

The introduction of a break in the mean of the loadings generally leads to a substantial drop in performance for all three estimators. This can be seen by comparing the results reported in Table 1 with those reported in Table 2. We also see that while demeaned PC and FE generally

[3]There are some minor distortions also when $N_0 = 0.5N$. These are, however, generally not larger than that they can be attributed to simulation uncertainty. Indeed, with 1,000 replications the 95% confidence interval for the size of the 5% level tests studied here (in %) is [3.6, 6.4].

perform very similarly when T is small, as T increases so does the relative performance of PC. This is expected, because the accuracy of the estimated PC factors is increasing in T. Hence, when T is small accurate estimation of the factors is not possible, and so the performance is again driven by the demeaning. As T increases, however, the accuracy of the factor estimates increases. At this point, the estimated factors stop being just redundant regressors, and so the relative performance of PC increases. Demeaning is still the key, though, which is obvious from the difference in the PC results depending on whether the data have been demeaned.

An important observation from Table 2 is that, except perhaps for the smallest values of N, the Wald test for uncorrelated loadings has good discriminatory power against the type of mean breaks considered here, which is expected given the discussion in Section 3 of the main paper. We also see that power increases steadily as N grows. This result is reassuring because we know from before that undetected breaks can have a substantial effect on performance. The results reported in Table 2 suggest that in large-N samples breaks are very likely to be detected. If the Wald test does not reject, we use FE, while if the test rejects, we use demeaned PC, which is relatively more accurate, especially if T is large.

The above conclusions apply not only to the particular setup considered here but also to all variations of it that we have considered, the results of which are again reported in Sections C.2 and C.3. In Section C.2, we take the same data generating process as here but vary the persistence of $\varepsilon_{i,t}$, the size of the breaks in γ_i, and the specification of F_t. While the results obviously differ, the conclusions do not. As an indication of this, we compute the correlation between all the bias results for FE and demeaned PC, on the one hand, and demeaned PC and PC based on raw data, on the other hand. While the former correlation is 0.90 (0.97) with 10% (50%) treated units, the latter correlation is -0.48 (-0.59). In other words, the performance of PC is driven mainly by the demeaning. Given that the theoretical results of Section B are not restricted to the treatments effects case, in Section C.3 we consider different specifications of $X_{i,t}$. Again, the conclusions are unaffected.

C.2 Additional results for the treatment effects case

As already mentioned, the results reported in the previous section are just a small fraction of the full set of results. The basic data generating process considered in this section is the same as before, except that we now consider more variations of it.

V1. The heterogeneity of the slope, β_i: We consider $\beta_i \sim N(1,1)$, as in Section C.1, and $\beta_i = \beta = 1$.

V2. The persistence of $\epsilon_{i,t}$, as measured by ρ: While in Section C.1, $\rho = 0.5$, here we consider $\rho \in \{0, 0.5, 1\}$. Hence, $\varepsilon_{i,t}$ can be serially uncorrelated ($\rho = 0$), persistent but stationary ($\rho = 0.5$), or unit root non-stationary ($\rho = 1$). Following Gobillon and Magnac (2016), we also consider $\epsilon_{i,t} \sim U(-\sqrt{3}, \sqrt{3})$.

V3. The size of the break in the factor loading, γ_i, as measured by μ: While in Section C.1, $\mu \in \{0, 1\}$, here we consider $\mu \in \{0, 0.5, 1\}$, similarly to Gobillon and Magnac (2016).

V4. The mean of the factor loadings, γ: While in Section C.1, $\gamma = 1$, here we set $\gamma = 0$.

V5. The number of factors, r: While in Section C.1, $r = 1$, here we consider $r \in \{2, 3\}$.

V6. The factor, F_t: Motivated by the empirical illustration, in Section C.1 we generated F_t as a highly persistent AR process. Here, we consider three additional specifications; $F_t \sim U(0.5, 1.5)$, $F_t \sim N(1, 1)$ and $F_t = 5 \cdot \sin(\pi t/T)$, where the first and third are taken from Gobillon and Magnac (2016).

V1–V6 are all the variations that we have considered. However, since some of the results were very similar, we do not report the results for all the parameterizations. For example, since the results for $\beta_i = 1$ resembled those for $\beta_i \sim N(1, 1)$, here we only report results for the latter parametrization. Similarly, the results for $\rho = 1$ ($\epsilon_{i,t} \sim U(-\sqrt{3}, \sqrt{3})$) were similar to those obtained for $\rho = 0.5$ ($\rho = 0$), and so we only report the results for $\rho \in \{0, 0.5\}$. We also do not report the results for $\mu = 0.5$ and $F_t \sim N(1, 1)$, as the conclusions were the same as for $\mu = 1$ and $F_t \sim U(0.5, 1.5)$, respectively. The rest of the data generating process is the same as in Section C.1, including the values of N and T that we consider, and the number of replications.

Tables 3–15 contain the results, and have the same structure as Tables 1 and 2. The conclusions are consistent with those reached in Section C.1. In particular, PC based on raw data generally leads to the worst performance by far. Demeaning leads to a marked improvement. Indeed, provided that the loadings are not breaking, the demeaned PC estimator is essentially unbiased and the RMSE decreases very fast with increases in T and especially in N. The FE estimator performs very similarly to demeaned PC. A break in the loadings leads to a marked loss of performance. This is true for all three estimators; however, it is only the demeaned PC results that have a clear tendency to improve as N increases, which is again consistent with the results reported in Section C.1.

Comparing across the different parameterizations, we see that while the persistence of the regression error affects the performance of PC based on raw data, the demeaned PC and FE estimators are basically unaffected. Similarly, while the results depend on the specification of the factors, specially when the loadings are breaking, the effect is largest for non-demeaned PC. The specification that leads to the worst performance is $F_t = 5 \cdot \sin(\pi t/T)$ when the loadings are breaking. The performance is, however, not all that different from the highly persistent AR case considered in the paper, and so the conclusions are the same. As expected, while the bias is unaffected, increasing the number of common factors leads to an increase in RMSE.

As with the estimators, the performance of the Wald test for uncorrelated slopes is consistent with the one reported in the main paper. Size accuracy is good. There are some distortions when $N_0 = 0.1N$, but these disappear quite quickly as N grows. Power is also good and it increases with the sample size. The Wald test supposes that Assumption TEST is met, which in turn requires that the mean of the loadings is nonzero. In order to assess the effect of a violation of this assumption, in Table 15 we report some results for the case when $\gamma = 0$. We see that the results are almost identical to those reported in Tables 1 and 2 for the case when $\gamma = 1$, suggesting that nonzero mean loading condition is not essential for test performance.

In contrast to the other tables, in addition to PC and FE, Table 15 reports the results obtained when applying OLS directly to the raw data. This is to illustrate that demeaning can be costly if the mean of the loadings is zero, in which case demeaning is no longer necessary for OLS to be consistent and asymptotically normal. In agreement with this, we see that OLS based on raw data generally leads to the lowest RMSE when $\gamma = 0$. The difference is, however, very small. Moreover, while the RMSE goes down, the size distortions go up when the demeaning is removed. Hence, even if it is known that the loadings have zero mean, it is not clear that one would prefer to apply OLS to the raw data.

C.3 Results with general regressors

In Sections C.1 and C.2, the data generating process is tailored to the treatment effects example, and resembles the one in Gobillon and Magnac (2016). As we pointed out earlier, however, our theoretical results are by no means restricted to the treatments effects case, but applies more broadly when estimating panel regressions with interactive effects. In this section, we therefore consider more types of regressors. The data generating process can be seen as version of the one used earlier. Just as before, we set $D_t = k = r = 1$, $\alpha_i \sim N(1,1)$ and

$$F_t = (1-\phi) + \phi F_{t-1} + e_t, \tag{C.48}$$

where $\phi = 0.8$, $F_0 = 0$ and $e_t \sim N(0,1)$. In the main paper, we allowed the support of γ_i to differ between the treated and non-treated units, which in turn induced a correlation with $X_{i,t}$. In this section, we instead set $\beta_i = 1 + v_i$ and $\gamma_i = 1 + \eta_i$, where v_i and η_i are allowed to depend on $\overline{X}_i$, as in Assumption TEST. Specifically,

$$\begin{bmatrix} v_i \\ \eta_i \end{bmatrix} = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} \overline{X}_i + \begin{bmatrix} w_i \\ z_i \end{bmatrix}, \tag{C.49}$$

where $(w_i, z_i)' \sim N(0_{2\times 1}, I_2)$. If v_i (η_i) is uncorrelated with $\overline{X}_i$, we set $\Lambda_1 = 0$ ($\Lambda_2 = 0$), whereas if v_i (η_i) is correlated with $\overline{X}_i$, then we set $\Lambda_1 = 0.5$ ($\Lambda_2 = 0.5$). As for $X_{i,t}$, we consider four experiments.

X1. $X_{i,t} = v_{i,t}$, where $v_{i,t} = 0.5 v_{i,t-1} + u_{i,t} + \sum_{j=i+1}^{K} 0.5(u_{i-j,t} + u_{i+j,t})$ with $K = 10$, $v_{i,0} = 0$ and $u_{i,t} \sim N(0,1)$.

X2. $X_{i,t} = \Gamma_i F_t + v_{i,t}$, where $\Gamma_i \sim N(1,1)$ and $v_{i,t}$ is as in X1.

X3. $X_{i,t} = 0.5X_{i,t-1} + \Gamma_i F_t + v_{i,t}$, where $X_{i,0} = 0$, and Γ_i and $v_{i,t}$ are the same as in X2 and X1, respectively.

X4. $X_{i,t} = B_i C_t$, where $B_i = 1(i \leq 0.5N)$ and $C_t = 1(t \geq 0.5T)$.

The data generating process of $v_{i,t}$ is taken from Bai and Ng (2002), and is chosen to showcase the generality of the types of regressors that can be accommodated. In particular, by generating the data in this way, $v_{i,t}$ is not only correlated over time, but also correlated with $2K$ of its neighbouring cross-section units. The correlation is, however, only weak.[4] Hence, since in this case $X_{i,t} = v_{i,t}$, in X1 $X_{i,t}$ is weakly correlated. By contrast, in X2, $X_{i,t}$ has an interactive effects specification, and so the cross-section dependence is of the strong type. Moreover, because $\times$ F_t and $v_{i,t}$ are generated as AR processes, $X_{i,t}$ has a moving average (MA) representation. By contrast, in X3, we generate $X_{i,t}$ as an ARMA process. Finally, in X4, $X_{i,t}$ is generated as a treatment dummy. Note that $\overline{X}_i = 0.5B_i$. Hence, by making η_i correlated with $\overline{X}_i$ we achieve the same goal as when changing the support of γ_i. Just as before, the results are based on making 1,000 replications of samples where $T \in \{6, 10, 20\}$ and $N \in \{30, 50, 100, 200\}$.

The main purpose of the exercise is to assess the small-sample accuracy of our theoretical predictions. Hence, unlike in Sections C.1 and C.2, here we are not particularly interested in the performance of the PC approach used by Gobillon and Magnac (2016), but focus on the FE estimator based on demeaning and its infeasible counterpart based on taking F_t as known. As in the main paper, we report bias, root mean squared error (RMSE), and the size of a double-sided t-test for testing $\beta = 1$.

The W_{RND} test for uncorrelated coefficients is also simulated. Three versions are considered, one for each of the hull hypotheses of $\Lambda_1 = 0$, $\Lambda_2 = 0$ and $\Lambda_1 = \Lambda_2 = 0$. As mentioned in Section A, provided that Assumption POS is met, for estimation purposes there is no need for any additional restrictions on the types of regressors that can be included in $X_{i,t}$. However, because of the way that $\overline{X}_i$ enters the augmented test regression, when using W_{RND} Assumption POS' has to be met. In particular and as we point out in Section 3 of the main paper, in X4,

[4]See Chudik et al. (2011) for a discussion of the concepts of weak and strong cross-section dependence.

we have $M_D X_i^* = (B_i - \overline{B})(C - \overline{C}1_{T\times 1})$ and $(\overline{X}_i' \otimes M_D X_i)^* = \overline{C}(B_i - \overline{B})(C - \overline{C}1_{T\times 1})$, where $C = [C_1, ..., C_T]'$ with $\overline{B}$ and $\overline{C}$ being the averages of B_i and C_t, respectively. Hence, $M_D X_i^*$ is proportional to $(\overline{X}_i' \otimes M_D X_i)^*$, which means that $\widehat{\Phi}$ is singular, leading to a violation of Assumption POS′. The solution to this problem is to simply drop $(\overline{X}_i' \otimes M_D X_i)^*$ from the test regression, and to base the test on the significance of $(\overline{X}_i^{*\prime} \otimes \widehat{F})$ only, which implies that the Wald test only has power in the direction of $\Lambda_2 \neq 0$. Hence, in X4, we only test if $\Lambda_2 = 0$.

Table 16 reports the X1 results for different combinations of N, T, Λ_1 and Λ_2. As expected given our asymptotic results, we see that when $\Lambda_1 = \Lambda_2 = 0$ the results for the two estimators are almost indistinguishable. This is true regardless of whether one is looking at the bias, RMSE or size. Because $\Lambda_1 = \Lambda_2 = 0$ the rejection frequencies of all three Wald tests represent size. Consistent with this we see that, while there are some distortions among the smaller values of N, especially for the joint test, as N increases the rejection frequencies approach the nominal 5% level. By contrast, when one of Λ_1 or Λ_2 is nonzero, then the rejection frequencies of the relevant tests increase markedly, and they continue to increase as N grows. Interestingly, the estimators continue to perform very similarly with only minor differences, although the performance of both estimators deteriorates quite substantially when $\Lambda_1 = 0.5$. The fact that the infeasible estimator works well when $\Lambda_2 = 0.5$ but not when $\Lambda_1 = 0.5$ is expected, because conditioning on F_t only takes care of the problem if it is the loadings that are correlated. Hence, the problems experienced by the FE estimator when $\Lambda_1 = 0.5$ are by no means unique, but apply to all known estimators, including PC. It is therefore quite reassuring to note that the Wald test seems to have high power in the direction of $\Lambda_1 \neq 0$.

The results reported in Tables 17 and 18 for X2 and X3 are qualitatively very similar to those reported in Table 16 for X1, and therefore the conclusions are the same. First, while the effect of $\Lambda_2 = 0.5$ on the FE estimator is generally larger than before, the estimators still perform quite similarly when $\Lambda_1 = 0$, especially when N is large. This is true when looking at bias and RMSE. However, if we look at size, we see that FE can be quite distorted, and that it is only when both N and T are large that the distortions come down to acceptable levels. Second, none of the estimators work when $\Lambda_1 = 0.5$. Third, the Wald tests seem to work well with decent size accuracy and high power in the direction of both $\Lambda_1 \neq 0$ and $\Lambda_2 \neq 0$.

The results for X4 are reported in Table 19. The first thing to note is that the bias and RMSE results are higher than in X1–X3, which is partly expected given the relatively low variation in $X_{i,t}$. However, the results do improve with increases in T and particularly in N. The effect of $\Lambda_1 = 0.5$ is not as pronounced as before, which is again partly expected given the relatively low variation in $X_{i,t}$. We also see that in contrast to before, now the size distortions of the t-tests are increasing in N. The Wald test for testing if $\Lambda_2 = 0$ performs as expected, with rejection frequencies that are close to 5% when $\Lambda_2 = 0$ and well above 5% when $\Lambda_2 = 0.5$. Note also that in treatment effects models with the treatment dummy as the only regressor, we cannot infer whether or not $\Lambda_1 = 0$. This problem is, however, less of an issue in models with multiple regressors, as here $M_D X_i^*$ and $(\overline{X}_i' \otimes M_D X_i)^*$ are unlikely to be perfectly collinear.

References

Andrews, D. W. K. (2005). Cross-Section Regression With Common Shocks. *Econometrica 73*, 1551–1585.

Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica 77*, 1229–1279.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica 70*, 191–221.

Chudik, A. and M. H. Pesaran (2015). Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. *Journal of Econometrics 188*, 393–420.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and Strong Cross-Section Dependence and Estimation of Large Panels. *Econometrics Journal 14*.

Cui, G., K. Hayakawa, S. Nagata, and T. Yamagata (2019). A Robust Approach to Heteroskedasticity, Error Serial Correlation and Slope Heterogeneity for Large Linear Panel Data Models with Interactive Effects. ISER DP No. 1037.

Forchini, G. and B. Peng (2016). A Conditional Approach to Panel Data Models with Common Shocks. *Econometrics 4*, 1–12.

Friedberg, L. (1998). Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data. *American Economic Review 88*, 608–627.

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

Hsiao, C. (2003). *Analysis of Panel Data* (2 ed.). Econometric Society Monographs. Cambridge University Press.

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

Kapetanios, G., L. Serlenga, and Y. Shin (2019). Testing for Correlated Factor Loadings in Cross Sectionally Dependent Panels. SERIES Working papers N. 02/2019.

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.

Magdalinos, T. and P. C. Phillips (2009). Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors. *Econometric Theory 25*, 482–526.

Moon, H. R. and M. Weidner (2015). Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects. *Econometrica 83*, 1543–1579.

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

Sarafidis, V. and D. Robertson (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *Econometrics Journal 12*(1), 62–81.

Sarafidis, V. and T. Wansbeek (2012). Cross-Sectional Dependence in Panel Data Analysis. *Econometric Reviews 31*, 483–531.

Westerlund, J. (2019). On Estimation and Inference in Heterogeneous Panel Regressions with Interactive Effects. *Journal of Time Series Analysis 40*, 852–857.

Wolfers, J. (2006). Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *American Economic Review 96*, 1802–1820.

Table 1: Simulation results without break in the loadings ($\mu = 0$).

		Bias			RMSE			5% size			
N	T	PC	PC*	FE	PC	PC*	FE	PC	PC*	FE	WAL
				10% treated units ($N_0 = 0.1N$)							
30	6	1.078	−0.044	−0.012	2.20	1.54	1.44	28.1	32.1	23.4	21.0
50	6	0.955	−0.025	−0.020	1.94	1.17	1.07	29.2	27.6	12.6	13.4
100	6	0.798	0.014	−0.034	1.47	0.81	0.78	32.7	25.4	9.1	8.2
200	6	0.706	−0.008	−0.005	1.24	0.54	0.53	38.1	23.5	6.5	8.0
30	10	1.377	0.064	0.050	2.48	1.48	1.46	42.6	42.6	22.4	20.4
50	10	1.117	0.058	0.044	2.04	1.11	1.11	39.5	33.2	13.7	14.3
100	10	0.895	−0.059	−0.051	1.59	0.72	0.79	41.9	23.2	9.9	10.1
200	10	0.865	−0.007	−0.004	1.46	0.51	0.56	51.1	19.8	6.6	8.4
30	20	1.221	−0.069	−0.032	2.25	1.41	1.33	54.1	53.9	20.7	21.3
50	20	0.898	−0.054	−0.048	1.75	0.93	1.06	49.6	38.5	14.4	13.3
100	20	0.793	−0.017	−0.024	1.52	0.66	0.78	47.9	30.5	8.4	8.0
200	20	0.582	−0.004	−0.037	1.11	0.47	0.57	50.6	27.2	7.3	9.1
				50% treated units ($N_0 = 0.5N$)							
30	6	1.406	−0.015	−0.020	2.05	0.81	0.70	32.7	22.4	6.7	6.4
50	6	1.400	−0.006	0.001	2.01	0.64	0.61	39.5	19.3	5.6	4.5
100	6	1.326	0.039	0.025	1.93	0.46	0.43	46.5	17.1	4.8	5.8
200	6	1.185	−0.001	−0.017	1.77	0.32	0.28	58.8	12.5	4.6	4.9
30	10	1.676	−0.028	−0.010	2.28	0.85	0.80	48.1	32.5	7.9	7.2
50	10	1.577	0.010	0.017	2.15	0.62	0.58	52.8	20.1	7.1	7.0
100	10	1.571	−0.014	−0.015	2.15	0.44	0.46	60.0	12.2	6.0	6.4
200	10	1.532	−0.002	−0.004	2.06	0.30	0.30	73.7	10.1	4.6	5.4
30	20	1.828	0.069	0.020	2.36	0.87	0.79	73.1	43.2	7.3	7.4
50	20	1.687	0.038	0.018	2.20	0.65	0.62	68.7	33.4	6.0	5.6
100	20	1.650	−0.001	0.012	2.13	0.42	0.42	75.2	19.5	4.6	4.3
200	20	1.597	0.017	−0.002	2.05	0.29	0.31	80.6	14.4	3.4	5.1

Notes: "PC*" and "PC" refer to the PC estimators based on the demeaned and raw data, respectively, "FE" is the OLS estimator based on the demeaned data, and "WAL" refers to the rejection frequency of the Wald test for testing the null hypothesis that the loadings are uncorrelated with $\overline{X}_i$. μ refers to the size of the break in the mean of the factor loading (γ_i). If $\mu = 0$, the rejection frequencies of the Wald test represent size, while if $\mu = 1$ they represent power. "RMSE" refers to the root mean square error, and the reported sizes are for a double-sided t-test for testing $\beta = 1$. N_0 refers to the number of treated units.

Table 2: Simulation results with a break in the loadings ($\mu = 1$).

		Bias			RMSE			5% size			
N	T	PC	PC*	FE	PC	PC*	FE	PC	PC*	FE	WAL
					10% treated units ($N_0 = 0.1N$)						
30	6	2.773	0.217	0.291	4.23	1.73	1.76	35.7	34.7	29.8	39.5
50	6	2.424	0.267	0.274	3.77	1.58	1.72	41.4	34.9	27.7	37.0
100	6	2.266	0.278	0.394	3.49	1.21	1.48	50.4	34.9	30.9	51.7
200	6	2.044	0.204	0.345	3.27	0.92	1.34	59.4	36.5	41.8	76.3
30	10	3.123	0.223	0.346	4.41	1.77	1.90	52.4	46.4	27.7	39.1
50	10	3.229	0.248	0.489	4.49	1.46	1.76	50.8	35.6	27.0	41.6
100	10	2.829	0.134	0.415	4.04	1.00	1.59	61.5	30.2	33.5	55.2
200	10	2.618	0.116	0.428	3.73	0.68	1.41	71.9	26.7	47.2	81.9
30	20	3.178	0.088	0.303	4.42	1.70	1.98	65.5	55.9	32.1	39.4
50	20	3.202	0.090	0.363	4.35	1.26	1.83	69.7	45.0	29.7	46.4
100	20	2.638	0.084	0.323	3.84	0.82	1.64	65.8	38.4	36.9	57.5
200	20	2.125	0.032	0.385	3.44	0.51	1.56	65.3	27.3	51.3	84.0
					50% treated units ($N_0 = 0.5N$)						
30	6	3.170	0.199	0.242	4.36	1.40	1.47	30.7	31.8	28.8	48.7
50	6	3.253	0.277	0.346	4.29	1.20	1.30	38.0	30.6	37.4	65.9
100	6	3.007	0.274	0.341	4.18	1.16	1.32	46.8	37.8	54.9	89.2
200	6	2.946	0.300	0.387	4.14	1.09	1.33	60.2	47.2	72.4	98.9
30	10	3.758	0.374	0.456	4.75	1.51	1.63	44.4	45.7	32.2	56.0
50	10	3.686	0.300	0.385	4.69	1.34	1.53	46.8	41.0	42.6	78.3
100	10	3.598	0.195	0.364	4.51	0.98	1.40	57.3	33.2	60.2	95.1
200	10	3.397	0.189	0.387	4.42	0.81	1.39	67.5	38.3	74.3	99.9
30	20	3.830	0.236	0.388	4.64	1.38	1.58	70.5	54.9	34.3	58.4
50	20	3.804	0.164	0.286	4.57	1.18	1.53	69.9	49.5	47.7	79.8
100	20	3.746	0.134	0.338	4.54	0.88	1.48	71.9	40.1	67.1	97.3
200	20	3.666	0.105	0.287	4.43	0.68	1.33	78.8	36.4	78.3	100.0

Notes: See Table 1 for an explanation.

Table 3: Simulation results with $\mu = 0$, $F_t \sim U(0.5, 1.5)$ and $\rho = 0.5$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
		10% treated units ($N_0 = 0.1N$)									
30	6	0.058	1.392	0.063	1.33	2.01	0.99	31.6	36.2	21.8	18.6
50	6	−0.009	1.150	−0.017	1.03	1.62	0.80	27.9	28.3	14.8	12.5
100	6	0.010	1.095	0.000	0.73	1.36	0.55	22.6	36.2	9.9	7.3
200	6	0.005	1.013	−0.007	0.51	1.20	0.38	21.1	49.7	7.0	6.9
30	10	0.066	1.415	0.018	1.30	1.96	0.93	42.0	45.2	20.8	20.3
50	10	−0.010	1.175	−0.020	1.01	1.58	0.72	28.4	38.9	13.7	11.6
100	10	−0.005	0.990	−0.016	0.69	1.22	0.53	19.7	43.8	9.9	8.3
200	10	−0.005	0.941	−0.006	0.49	1.09	0.35	16.9	61.7	6.3	7.6
30	20	0.004	1.050	0.002	1.32	1.69	0.80	50.0	53.9	21.9	19.2
50	20	0.030	0.956	0.049	1.09	1.46	0.63	44.5	54.8	11.7	11.9
100	20	0.017	0.752	0.012	0.69	1.11	0.44	32.8	53.2	8.9	8.8
200	20	−0.009	0.720	−0.008	0.46	0.94	0.31	26.0	64.9	5.9	6.2
		50% treated units ($N_0 = 0.5N$)									
30	6	−0.009	1.821	0.007	0.73	1.95	0.54	22.3	45.1	7.1	7.3
50	6	−0.035	1.727	−0.021	0.54	1.81	0.41	16.4	49.7	5.8	5.3
100	6	−0.008	1.742	−0.012	0.39	1.79	0.30	11.7	67.6	5.1	4.5
200	6	−0.001	1.752	−0.004	0.29	1.78	0.21	10.0	87.8	5.3	4.9
30	10	0.014	1.889	0.035	0.70	1.99	0.51	27.8	64.4	7.4	8.0
50	10	0.002	1.789	−0.001	0.57	1.86	0.38	18.2	70.3	5.2	6.6
100	10	−0.012	1.745	0.001	0.42	1.79	0.28	11.8	84.2	5.7	6.4
200	10	−0.002	1.743	−0.008	0.28	1.77	0.20	8.1	96.3	5.0	5.2
30	20	−0.006	1.822	−0.010	0.85	1.92	0.45	45.3	86.7	8.0	8.1
50	20	0.004	1.816	0.021	0.65	1.88	0.35	33.5	94.1	6.6	5.5
100	20	0.013	1.820	−0.002	0.46	1.85	0.23	20.9	98.0	4.3	7.5
200	20	0.003	1.822	−0.006	0.30	1.84	0.17	12.2	99.4	4.0	5.4

Notes: See Table 1 for an explanation.

Table 4: Simulation results with $\mu = 1$, $F_t \sim U(0.5, 1.5)$ and $\rho = 0.5$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
10% treated units ($N_0 = 0.1N$)											
30	6	0.039	3.227	0.021	1.27	3.63	0.98	31.1	49.1	21.7	27.3
50	6	−0.004	3.112	0.023	1.02	3.44	0.81	28.6	55.6	15.3	26.4
100	6	−0.012	3.069	−0.012	0.76	3.27	0.60	26.4	75.8	13.0	30.8
200	6	−0.002	3.043	0.016	0.62	3.16	0.44	30.4	94.5	10.6	49.6
30	10	−0.039	3.459	−0.004	1.30	3.76	0.96	40.6	65.1	22.2	31.1
50	10	−0.015	3.110	−0.004	1.07	3.36	0.73	32.1	73.3	14.3	31.0
100	10	0.004	3.065	−0.031	0.79	3.21	0.56	25.5	91.6	12.1	42.5
200	10	0.028	3.105	0.009	0.60	3.19	0.42	25.7	99.1	11.1	66.8
30	20	0.002	3.215	0.023	1.40	3.59	0.85	55.6	83.3	22.7	33.0
50	20	0.057	3.065	0.029	1.15	3.37	0.66	47.3	88.4	15.9	42.3
100	20	−0.022	3.018	0.002	0.85	3.23	0.48	40.9	95.5	10.3	55.5
200	20	0.011	2.961	0.015	0.69	3.19	0.36	45	95.3	9.8	81.2
50% treated units ($N_0 = 0.5N$)											
30	6	0.039	3.956	0.034	0.76	4.01	0.58	24.1	42.7	9.5	27.2
50	6	−0.004	3.936	−0.007	0.64	3.98	0.48	18.8	40.0	9.3	35.4
100	6	0.011	3.917	0.008	0.47	3.94	0.38	16.3	54.7	12.6	59.5
200	6	−0.009	3.928	−0.002	0.42	3.95	0.32	21.6	70.8	21.1	83.8
30	10	0.003	3.946	0.009	0.74	3.99	0.53	29.6	59.1	8.6	33.0
50	10	−0.002	3.896	−0.024	0.63	3.93	0.43	22.1	61.4	8.2	49.3
100	10	−0.008	3.913	−0.009	0.54	3.93	0.34	21.5	71.2	11.8	74.6
200	10	0.015	3.914	−0.007	0.48	3.93	0.27	25.0	84.5	14.6	95.7
30	20	0.031	3.954	0.021	0.95	3.99	0.46	50.1	88.2	7.9	44.7
50	20	−0.028	3.931	−0.005	0.83	3.96	0.37	46.6	90.4	9.1	62.1
100	20	0.024	3.940	−0.010	0.75	3.95	0.27	47.0	91.8	7.8	91.0
200	20	0.015	3.964	0.011	0.73	3.97	0.22	53.8	95.0	11.9	99.2

Notes: See Table 1 for an explanation.

Table 5: Simulation results with $\mu = 0$, $F_t \sim U(0.5, 1.5)$ and $\rho = 0$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
10% treated units ($N_0 = 0.1N$)											
30	6	0.009	0.764	0.004	1.20	1.54	0.79	45.6	42.3	23.8	18.9
50	6	−0.024	0.325	−0.012	0.95	1.02	0.60	40.7	33.2	13.2	14.5
100	6	−0.019	−0.025	−0.038	0.67	0.48	0.42	33.1	23.7	8.3	8.0
200	6	−0.014	−0.014	0.000	0.49	0.29	0.30	26.3	22.8	6.9	7.3
30	10	0.055	0.817	0.035	1.14	1.53	0.71	59.9	50.5	22.6	20.4
50	10	−0.007	0.240	0.003	0.87	0.88	0.55	48.8	36.7	15.6	11.1
100	10	0.001	−0.015	−0.008	0.47	0.39	0.38	32.0	30.0	9.2	9.4
200	10	−0.009	−0.028	−0.011	0.29	0.27	0.28	29.9	30.5	7.2	7.0
30	20	−0.024	0.146	0.004	0.79	0.86	0.64	48.8	44.4	22.6	18.6
50	20	0.005	0.005	0.006	0.53	0.52	0.48	43.9	41.9	14.1	14.6
100	20	0.004	−0.008	0.007	0.35	0.34	0.35	41.0	41.6	9.0	8.7
200	20	−0.013	−0.030	−0.014	0.25	0.25	0.25	40.5	42.0	6.2	7.2
50% treated units ($N_0 = 0.5N$)											
30	6	0.023	1.137	0.009	0.60	1.50	0.30	47.4	38.1	5.6	6.4
50	6	−0.015	1.475	0.009	0.54	1.73	0.24	52.7	44.0	7.0	7.9
100	6	0.007	0.137	−0.002	0.25	0.51	0.18	20.6	14.4	6.6	6.9
200	6	0.024	0.430	0.006	0.49	0.91	0.24	38.9	23.0	6.3	6.2
30	10	−0.005	0.190	0.002	0.33	0.59	0.18	31.1	14.8	5.0	6.2
50	10	−0.004	0.000	−0.002	0.13	0.11	0.13	7.1	5.3	5.2	5.3
100	10	−0.002	0.010	0.006	0.36	0.20	0.17	36.5	7.5	5.7	4.7
200	10	0.002	0.000	−0.001	0.15	0.11	0.13	11.4	6.7	5.2	6.4
30	20	0.000	0.000	−0.001	0.10	0.08	0.10	9	7.3	6.7	5.1
50	20	−0.005	0.002	0.003	0.24	0.11	0.12	31.5	9.3	6.5	5.4
100	20	−0.002	−0.001	−0.003	0.10	0.08	0.09	8.7	6.0	5.5	4.3
200	20	0.004	0.004	0.003	0.07	0.06	0.07	6.9	6.6	4.8	5.7

Notes: See Table 1 for an explanation.

Table 6: Simulation results with $\mu = 1$, $F_t \sim U(0.5, 1.5)$ and $\rho = 0$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
				10% treated units ($N_0 = 0.1N$)							
30	6	−0.054	2.376	−0.006	1.21	3.17	0.82	49.0	40.1	24.6	28.4
50	6	−0.004	1.222	−0.023	1.02	2.24	0.66	41.1	35.2	16.5	27.5
100	6	−0.020	0.001	−0.006	0.76	0.59	0.50	33.7	26.2	13.4	36.0
200	6	0.025	−0.018	0.024	0.57	0.30	0.37	33.1	22.0	12.8	52.6
30	10	0.000	3.007	0.011	1.13	3.52	0.72	59.0	49.3	23.7	34.3
50	10	0.008	0.940	0.013	0.87	1.97	0.57	48.1	37.7	16.6	35.0
100	10	0.025	−0.014	0.010	0.55	0.42	0.44	37.7	28.9	12.6	48.7
200	10	−0.006	−0.045	−0.013	0.34	0.28	0.33	34.7	27.7	13.2	71.9
30	20	−0.009	0.592	0.021	0.78	1.64	0.64	47.4	44.6	23.3	40.5
50	20	−0.011	−0.015	−0.015	0.58	0.61	0.52	45.6	39.0	16.2	46.1
100	20	0.007	−0.018	0.017	0.40	0.37	0.38	46.4	40.4	13.6	65.1
200	20	−0.005	−0.040	−0.012	0.27	0.24	0.27	40.9	36.8	8.8	89.6
				50% treated units ($N_0 = 0.5N$)							
30	6	0.004	3.754	0.014	0.72	3.93	0.47	28.8	23.2	12.8	28.2
50	6	−0.018	3.163	0.007	0.59	3.60	0.38	23.7	20.2	11.0	40.6
100	6	0.019	1.647	0.015	0.42	2.64	0.32	13.5	17.5	17.1	64.3
200	6	0.018	0.174	0.000	0.36	1.14	0.27	10.4	25.8	26.6	85.3
30	10	−0.034	3.947	−0.012	0.77	4.00	0.40	43.3	31.4	10.9	40.2
50	10	−0.010	3.435	0.008	0.69	3.73	0.33	37.0	22.2	11.6	61.2
100	10	−0.019	2.121	0.011	0.65	2.94	0.27	33.9	24.8	15.2	84.7
200	10	0.044	0.504	−0.007	0.83	1.59	0.23	66.0	36.3	26.0	98.7
30	20	0.021	3.412	−0.004	0.81	3.72	0.33	51.9	50.1	8.1	54.2
50	20	−0.011	2.388	0.003	0.91	3.13	0.26	61.6	48.3	9.1	77.5
100	20	−0.062	0.947	0.011	1.03	2.06	0.21	88.3	43.3	13.3	96.7
200	20	−0.048	0.009	−0.001	1.08	0.85	0.17	97.5	54.7	16.5	100.0

Notes: See Table 1 for an explanation.

Table 7: Simulation results with $\mu = 0$, F_t as in the main paper and $\rho = 0$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
		10% treated units ($N_0 = 0.1N$)									
30	6	0.103	0.789	0.054	1.38	2.06	1.24	44.2	39.5	22.2	19.7
50	6	0.042	0.467	0.046	1.13	1.55	0.99	38.6	31.1	15.1	14.4
100	6	0.008	0.110	0.003	0.80	0.70	0.72	30.4	24.6	10.4	8.5
200	6	−0.014	0.075	0.009	0.53	0.52	0.51	26.6	26.0	7.7	6.3
30	10	0.100	0.974	0.043	1.30	2.17	1.33	58.8	44.7	23.3	22.2
50	10	−0.046	0.290	−0.035	0.89	1.34	1.01	43.5	33.2	16.2	14.0
100	10	−0.023	0.079	−0.028	0.48	0.73	0.78	28.6	28.4	10.2	11.6
200	10	0.002	0.039	0.014	0.31	0.43	0.53	27.8	31.4	7.5	7.6
30	20	−0.011	0.253	−0.087	0.86	1.35	1.34	46.9	42.2	23.8	21.4
50	20	0.015	0.145	0.038	0.53	1.03	1.07	37.8	38.2	15.9	14.0
100	20	−0.021	0.035	−0.015	0.37	0.63	0.70	39.8	39.1	8.3	10.8
200	20	0.000	0.006	0.005	0.26	0.28	0.51	36.5	36.2	7.2	7.4
		50% treated units ($N_0 = 0.5N$)									
30	6	0.000	1.299	0.015	0.79	2.08	0.66	29.1	33.8	6.1	6.3
50	6	−0.033	1.035	−0.014	0.67	1.86	0.54	22.4	30.6	7.1	5.0
100	6	−0.014	0.673	−0.001	0.46	1.47	0.37	14.2	34.9	5.0	5.1
200	6	0.002	0.320	0.005	0.32	1.11	0.24	8.2	39.8	4.3	4.7
30	10	0.017	1.588	−0.001	0.78	2.17	0.64	42.0	41.0	6.3	5.9
50	10	0.005	1.257	−0.019	0.65	2.03	0.54	30.8	34.9	5.9	6.5
100	10	0.015	0.818	−0.007	0.46	1.69	0.38	17.0	33.8	5.8	5.6
200	10	−0.016	0.526	−0.006	0.26	1.36	0.27	10.0	42.4	5.5	5.5
30	20	−0.025	1.419	−0.026	0.74	2.16	0.73	47.0	48.4	6.8	6.1
50	20	0.021	0.993	0.023	0.51	1.82	0.53	32.6	44.2	6.8	7.0
100	20	−0.004	0.732	−0.003	0.28	1.61	0.40	24.1	43.4	6.4	5.6
200	20	−0.003	0.599	0.002	0.14	1.52	0.28	22.9	46.9	5.9	4.9

Notes: See Table 1 for an explanation.

Table 8: Simulation results with $\mu = 1$, F_t as in the main paper and $\rho = 0$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
10% treated units ($N_0 = 0.1N$)											
30	6	0.312	2.087	0.345	1.78	3.72	1.76	45.1	39.7	30.6	38.0
50	6	0.320	1.132	0.391	1.53	2.78	1.51	39.3	31.3	30.1	37.4
100	6	0.232	0.506	0.292	1.24	1.86	1.40	37.7	29.0	34.2	55.3
200	6	0.195	0.268	0.340	0.95	1.14	1.27	36.8	31.4	52.6	79.0
30	10	0.187	2.639	0.327	1.63	4.22	1.78	62.8	42.7	31.2	42.9
50	10	0.145	1.580	0.316	1.22	3.31	1.76	46.4	32.2	33.2	42.7
100	10	−0.021	0.771	0.287	0.51	2.37	1.53	32.1	29.6	36.1	58.7
200	10	−0.005	0.355	0.363	0.33	1.44	1.41	29.5	33.3	56.4	83.8
30	20	0.074	1.420	0.435	1.01	3.25	1.92	47.0	35.5	34.4	42.4
50	20	−0.002	0.954	0.354	0.55	2.67	1.76	41.6	34.8	32.9	43.8
100	20	0.012	0.701	0.315	0.38	2.26	1.58	39.6	36.6	40.6	62.2
200	20	−0.010	0.332	0.286	0.26	1.55	1.49	38.1	37.8	60.3	85.5
50% treated units ($N_0 = 0.5N$)											
30	6	0.351	3.137	0.410	1.45	4.29	1.47	36.8	29.9	36.9	51.0
50	6	0.241	2.517	0.295	1.21	3.87	1.28	32.9	27.5	46.8	70.2
100	6	0.261	1.928	0.296	1.17	3.45	1.23	30.6	30.7	63.7	90.2
200	6	0.248	1.300	0.305	1.11	2.82	1.22	28.7	40.1	78.6	98.4
30	10	0.238	3.455	0.315	1.44	4.59	1.52	54.8	32.0	40.1	61.1
50	10	0.238	3.255	0.339	1.27	4.47	1.46	47.7	26.0	52.5	74.9
100	10	0.137	2.896	0.375	0.90	4.23	1.40	34.0	28.4	70.1	97.5
200	10	0.058	2.002	0.288	0.53	3.56	1.35	25.6	36.1	82.8	99.9
30	20	0.113	3.527	0.302	1.22	4.48	1.52	56.1	39.8	42.7	64.6
50	20	0.106	3.171	0.241	0.79	4.29	1.40	42.0	38.8	52.2	80.0
100	20	0.013	2.880	0.322	0.42	4.22	1.44	33.7	34.1	74.0	98.0
200	20	−0.013	2.800	0.366	0.18	4.21	1.37	32.1	37.9	84.8	100.0

Notes: See Table 1 for an explanation.

Table 9: Simulation results with $\mu = 0$, $F_t = 5 \cdot \sin(\pi t/T)$ and $\rho = 0.5$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
10% treated units ($N_0 = 0.1N$)											
30	6	−0.046	2.643	−0.062	1.42	3.28	1.42	34.4	35.7	21.8	20.0
50	6	−0.009	2.467	−0.030	1.15	2.93	1.11	29.7	43.9	12.2	13.5
100	6	−0.025	2.294	−0.018	0.78	2.60	0.78	26.9	69.0	9.4	9.7
200	6	0.020	2.104	0.018	0.56	2.34	0.55	28.0	87.9	6.2	8.0
30	10	−0.006	3.194	0.047	1.29	3.75	1.12	39.9	53.2	21.6	21.4
50	10	−0.032	3.036	−0.027	0.99	3.49	0.89	30.7	60.6	15.1	11.5
100	10	−0.006	2.900	−0.042	0.68	3.15	0.59	21.1	83.7	8.9	9.2
200	10	0.005	2.848	0.010	0.49	2.99	0.43	19.3	97.3	6.1	5.6
30	20	0.024	3.272	0.022	1.30	3.98	0.89	51.0	68.8	20.1	18.4
50	20	−0.044	3.184	0.006	1.00	3.71	0.69	42.4	76.8	13.5	11.9
100	20	0.014	3.165	−0.008	0.68	3.53	0.48	31.6	88.8	9.5	7.8
200	20	−0.002	2.960	−0.009	0.45	3.29	0.35	23.7	91.9	7.4	6.8
50% treated units ($N_0 = 0.5N$)											
30	6	0.001	3.213	−0.009	0.82	3.32	0.81	23.7	27.2	7.5	6.9
50	6	0.039	3.181	0.050	0.65	3.24	0.64	18.0	29.3	5.8	6.6
100	6	0.005	3.177	0.005	0.47	3.21	0.44	18.9	41.5	5.1	5.2
200	6	−0.005	3.198	−0.014	0.32	3.21	0.33	14.3	63.9	6.8	5.5
30	10	0.051	3.613	0.027	0.76	3.71	0.61	29.4	43.2	5.2	6.5
50	10	0.024	3.612	0.000	0.59	3.67	0.50	17.7	40.2	7.6	6.8
100	10	0.017	3.578	0.002	0.39	3.61	0.33	10.3	41.5	4.4	4.5
200	10	0.000	3.568	0.001	0.30	3.58	0.24	10.5	55.1	5.5	4.9
30	20	0.004	3.933	−0.002	0.82	4.03	0.47	44.7	62.7	6.9	7.1
50	20	−0.023	3.871	0.002	0.63	3.93	0.36	32.7	58.2	5.4	5.6
100	20	0.000	3.888	−0.007	0.43	3.92	0.26	19.1	61.0	5.4	5.6
200	20	−0.001	3.902	0.006	0.29	3.92	0.18	10.7	65.8	4.6	5.4

Notes: See Table 1 for an explanation.

Table 10: Simulation results with $\mu = 1$, $F_t = 5 \cdot \sin(\pi t/T)$ and $\rho = 0.5$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
				10% treated units ($N_0 = 0.1N$)							
30	6	−1.349	6.246	−1.749	2.01	6.53	2.23	42.2	27.0	45.1	53.2
50	6	−1.184	6.118	−1.678	1.74	6.31	2.03	39.7	32.0	45.1	59.6
100	6	−0.943	6.201	−1.632	1.33	6.30	1.80	48.1	50.3	58.9	81.3
200	6	−0.821	6.210	−1.684	1.05	6.26	1.77	61.4	69.2	87.9	97.5
30	10	−0.660	7.036	−1.034	1.45	7.28	1.50	41.9	36.4	35.7	52.0
50	10	−0.503	7.042	−0.996	1.17	7.21	1.30	35.6	35.7	33.1	58.8
100	10	−0.304	7.174	−1.000	0.79	7.25	1.17	26.6	40.4	43.7	82.9
200	10	−0.201	7.053	−1.011	0.54	7.09	1.10	23.8	62.5	66.8	98.1
30	20	−0.192	7.628	−0.496	1.29	7.88	0.98	53.0	48.3	28.2	51.3
50	20	−0.148	7.730	−0.526	1.04	7.89	0.86	43.8	47.1	22.8	61.7
100	20	−0.026	7.777	−0.465	0.67	7.86	0.68	29.9	47.0	22.1	83.6
200	20	−0.037	7.788	−0.502	0.50	7.83	0.61	29.0	52.3	33.7	98.4
				50% treated units ($N_0 = 0.5N$)							
30	6	−1.448	6.552	−1.655	1.71	6.60	1.83	43.7	11.3	57.8	74.5
50	6	−1.377	6.549	−1.634	1.56	6.58	1.75	47.7	10.1	74.2	90.7
100	6	−1.382	6.541	−1.690	1.49	6.55	1.75	66.1	11.0	96.7	99.8
200	6	−1.240	6.516	−1.667	1.31	6.52	1.70	90.7	18.7	100.0	100.0
30	10	−0.791	7.277	−0.977	1.10	7.33	1.15	45.5	15.1	39.3	73.1
50	10	−0.682	7.266	−0.977	0.94	7.29	1.09	38.2	12.6	54.0	90.7
100	10	−0.555	7.289	−1.001	0.73	7.30	1.05	33.6	8.4	84.9	99.8
200	10	−0.485	7.304	−1.008	0.59	7.31	1.04	43.3	9.6	99.3	100.0
30	20	−0.247	7.835	−0.500	0.86	7.88	0.69	46.9	25.7	21.0	78.6
50	20	−0.262	7.840	−0.502	0.70	7.87	0.62	35.0	16.1	31.7	92.6
100	20	−0.197	7.829	−0.490	0.50	7.84	0.55	24.6	15.2	48.0	99.6
200	20	−0.197	7.851	−0.508	0.38	7.86	0.54	18.9	16.4	80.1	100.0

Notes: See Table 1 for an explanation.

Table 11: Simulation results with $\mu = 0$, $F_t = 5 \cdot \sin(\pi t/T)$ and $\rho = 0$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
		10% treated units ($N_0 = 0.1N$)									
30	6	0.014	1.878	0.005	1.37	2.80	1.28	44.8	36.8	22.7	20.3
50	6	0.051	0.964	0.047	1.06	1.95	0.98	40.5	35.5	13.4	14.3
100	6	0.007	0.308	−0.004	0.75	0.99	0.70	31.1	32.4	9.5	8.6
200	6	−0.002	0.105	−0.012	0.54	0.47	0.51	30.1	29.0	8.0	6.7
30	10	0.028	2.552	0.048	1.13	3.50	0.88	60.9	41.6	20.1	18.7
50	10	0.036	1.453	0.000	0.85	2.60	0.70	46.4	34.7	14.6	12.2
100	10	−0.021	0.483	0.023	0.45	1.40	0.51	32.6	37.2	9.5	7.8
200	10	−0.004	0.121	0.000	0.28	0.52	0.35	32.8	37.7	7.5	6.4
30	20	0.018	2.054	0.045	0.75	3.37	0.72	49.9	38.8	22.4	19.8
50	20	−0.008	1.243	−0.018	0.51	2.52	0.55	42.6	43.0	14.0	12.4
100	20	0.003	0.281	0.014	0.36	1.03	0.39	42.2	46.3	9.1	8.7
200	20	−0.006	0.090	−0.007	0.26	0.39	0.29	43.2	46.8	7.2	6.8
		50% treated units ($N_0 = 0.5N$)									
30	6	−0.046	3.172	−0.028	0.85	3.28	0.73	32.0	21.6	8.8	7.5
50	6	−0.008	3.167	0.013	0.66	3.24	0.58	23.9	12.0	7.6	6.2
100	6	0.024	3.164	0.027	0.47	3.21	0.40	13.0	14.2	6.2	4.9
200	6	−0.016	3.136	−0.006	0.35	3.18	0.29	7.7	17.5	4.2	4.0
30	10	0.002	3.615	0.020	0.74	3.71	0.49	44.1	18.8	5.5	6.5
50	10	0.033	3.626	0.024	0.63	3.68	0.39	30.7	8.6	5.9	6.1
100	10	0.005	3.637	−0.007	0.47	3.67	0.27	16.3	5.2	5.2	5.4
200	10	−0.005	3.632	−0.005	0.31	3.66	0.19	7.1	6.1	4.4	3.9
30	20	−0.012	3.953	−0.006	0.74	4.05	0.37	48.3	19.1	8.1	7.1
50	20	−0.009	3.903	−0.003	0.57	3.97	0.27	36.9	12.8	6.4	5.0
100	20	−0.015	3.838	−0.013	0.33	3.91	0.19	24.3	10.5	5.4	6.4
200	20	−0.002	3.475	−0.002	0.14	3.71	0.14	26.0	19.9	5.4	5.5

Notes: See Table 1 for an explanation.

Table 12: Simulation results with $\mu = 1$, $F_t = 5 \cdot \sin(\pi t/T)$ and $\rho = 0$.

		Bias			RMSE			5% size			
N	T	PC*	PC	FE	PC*	PC	FE	PC*	PC	FE	WAL
				10% treated units ($N_0 = 0.1N$)							
30	6	−1.255	5.938	−1.750	1.98	6.40	2.17	50.1	17.5	47.2	57.3
50	6	−1.073	5.345	−1.666	1.65	5.99	1.92	49.1	18.2	51.8	62.5
100	6	−1.029	4.304	−1.674	1.43	5.17	1.81	46.3	35.0	70.6	84.0
200	6	−0.804	1.885	−1.642	1.18	3.16	1.71	45.0	51.0	91.8	98.4
30	10	−0.393	7.209	−1.044	1.27	7.51	1.40	61.7	9.6	41.8	54.2
50	10	−0.161	6.836	−1.020	0.93	7.23	1.25	50.1	7.0	43.3	62.0
100	10	−0.008	4.862	−0.986	0.50	5.97	1.12	35.1	18.2	54.7	83.9
200	10	0.021	1.095	−0.990	0.28	2.55	1.05	31.7	44.0	80.1	98.8
30	20	−0.017	7.046	−0.505	0.81	7.66	0.88	53.2	9.5	31.2	54.2
50	20	0.000	5.609	−0.497	0.52	6.70	0.73	45.4	14.6	26.6	63.6
100	20	0.001	1.996	−0.497	0.36	3.84	0.63	42.1	36.9	31.6	87.7
200	20	−0.001	0.240	−0.505	0.25	0.68	0.58	44.5	52.3	49.8	99.0
				50% treated units ($N_0 = 0.5N$)							
30	6	−1.490	6.513	−1.702	1.74	6.57	1.85	50.7	6.1	66.8	76.7
50	6	−1.490	6.507	−1.682	1.66	6.55	1.77	47.7	3.4	86.6	93.4
100	6	−1.533	6.528	−1.659	1.62	6.57	1.70	46.7	2.7	98.6	99.8
200	6	−1.547	6.479	−1.665	1.59	6.54	1.69	55.9	9.1	100.0	100.0
30	10	−0.694	7.333	−0.975	1.05	7.37	1.09	56.4	3.0	53.1	76.9
50	10	−0.672	7.305	−1.003	0.95	7.33	1.07	47.1	1.0	74.3	93.1
100	10	−0.650	7.289	−1.004	0.85	7.31	1.04	38.9	1.0	96.1	100.0
200	10	−0.699	7.276	−1.006	0.82	7.31	1.03	45.0	7.4	99.9	100.0
30	20	−0.201	7.787	−0.494	0.78	7.85	0.61	50.7	2.3	33.7	76.7
50	20	−0.215	7.812	−0.497	0.62	7.86	0.57	41.8	1.3	47.0	93.2
100	20	−0.184	7.640	−0.488	0.43	7.77	0.52	29.6	3.3	73.4	100.0
200	20	−0.033	7.354	−0.495	0.22	7.60	0.51	33.8	25.6	95.7	100.0

Notes: See Table 1 for an explanation.

Table 13: Simulation results with $r \in \{2,3\}$, $\mu = 0$ and $N_0 = 0.1N$.

		Bias			RMSE			5% size			
N	T	PC	PC*	FE	PC	PC*	FE	PC	PC*	FE	WAL
					Two factors ($r = 2$)						
30	6	0.537	−0.025	−0.013	2.17	1.50	1.41	28.3	30.7	19.1	20.2
50	6	0.552	−0.026	−0.046	1.65	1.18	1.07	31.5	26.0	13.6	13.0
100	6	0.421	−0.048	−0.030	1.41	0.80	0.77	32.0	22.2	8.5	7.8
200	6	0.402	−0.018	−0.016	1.23	0.55	0.53	38.3	21.2	5.7	5.2
30	10	0.701	0.077	0.090	2.20	1.49	1.45	37.6	42.2	20.9	18.3
50	10	0.539	0.009	−0.021	1.91	1.12	1.15	37.8	32.1	13.1	12.4
100	10	0.467	−0.031	−0.025	1.33	0.71	0.78	38.2	23.2	7.7	8.0
200	10	0.413	0.020	0.020	1.21	0.48	0.56	43.8	18.6	7.1	8.8
30	20	0.642	0.051	0.079	1.91	1.32	1.37	52.6	50.6	17.8	18.9
50	20	0.375	−0.001	−0.031	1.37	0.92	1.11	43.7	36.1	12.6	12.9
100	20	0.253	−0.016	−0.016	0.98	0.64	0.81	41.4	32.1	9.3	8.1
200	20	0.164	−0.001	−0.007	0.76	0.42	0.54	39.1	26.4	7.3	5.5
					Three factors ($r = 3$)						
30	6	0.509	−0.097	−0.050	2.37	1.67	1.65	32.3	31.0	19.3	18.2
50	6	0.566	−0.004	0.027	2.05	1.22	1.23	31.3	27.1	13.5	14.4
100	6	0.467	0.022	0.010	1.62	0.87	0.86	35.0	21.1	9.4	9.1
200	6	0.464	−0.020	0.003	1.56	0.58	0.62	42.2	21.7	7.1	6.6
30	10	0.691	0.023	0.034	2.61	1.56	1.69	41.4	40.0	20.1	19.8
50	10	0.658	0.015	0.019	2.12	1.08	1.21	40.7	30.2	12.9	13.4
100	10	0.430	0.018	0.006	1.69	0.76	0.91	41.4	24.7	9.9	9.8
200	10	0.458	−0.009	0.002	1.32	0.50	0.64	44.6	20.4	6.8	6.5
30	20	0.570	−0.011	−0.025	2.08	1.31	1.56	51.1	45.4	18.4	19.2
50	20	0.363	−0.023	−0.006	1.51	0.91	1.26	44.2	39.1	13.1	12.3
100	20	0.189	−0.017	−0.022	0.95	0.61	0.92	39.7	30.4	8.6	7.8
200	20	0.145	−0.015	0.022	0.73	0.41	0.65	40.4	29.9	7.0	7.1

Notes: See Table 1 for an explanation.

Table 14: Simulation results with with $r \in \{2, 3\}$, $\mu = 1$ and $N_0 = 0.1N$.

		Bias			RMSE			5% size			
N	T	PC	PC*	FE	PC	PC*	FE	PC	PC*	FE	WAL
					Two factors ($r = 2$)						
30	6	1.672	0.036	0.003	4.23	2.09	2.18	37.6	38.7	36.5	55.7
50	6	1.343	−0.083	−0.084	3.67	1.73	1.99	39.1	35.4	37.4	62.4
100	6	1.510	0.155	0.124	3.55	1.50	1.86	48.5	37.4	44.3	83.0
200	6	1.170	0.031	0.025	3.13	1.32	1.77	53.3	44.7	59.0	95.4
30	10	1.816	−0.059	0.028	4.22	2.09	2.39	48.4	52.6	40.6	63.9
50	10	1.705	0.024	0.010	4.01	1.64	2.13	48.2	42.4	39.8	74.9
100	10	1.551	0.035	0.117	3.68	1.12	1.92	53.1	33.7	51.1	91.9
200	10	1.327	0.012	0.035	3.27	0.88	1.93	62.1	35.8	62.6	99.6
30	20	1.427	−0.027	−0.040	3.58	1.81	2.37	57.8	56.9	38.5	65.3
50	20	1.458	0.000	0.016	3.41	1.44	2.29	57.2	45.6	40.3	82.4
100	20	1.016	−0.014	−0.036	2.73	0.79	2.08	52.1	38.8	52.7	96.5
200	20	0.945	0.014	0.065	2.58	0.53	2.00	54.0	34.7	66.7	99.5
					Three factors ($r = 3$)						
30	6	1.833	0.102	0.102	5.08	2.39	2.53	39.4	40.7	38.2	61.4
50	6	1.624	0.054	0.076	4.50	2.01	2.31	44.2	39.6	39.6	70.4
100	6	1.187	−0.061	−0.023	4.01	1.86	2.30	51.1	39.9	50.5	89.1
200	6	1.192	0.008	−0.042	3.80	1.59	2.13	59.4	48.6	62.8	96.9
30	10	1.959	0.090	0.067	5.21	2.36	2.73	47.1	52.0	41.7	68.1
50	10	1.589	0.013	0.001	4.61	1.83	2.54	50.9	45.9	40.2	82.9
100	10	1.540	−0.027	0.113	4.31	1.29	2.37	55.7	40.0	51.7	96.3
200	10	1.384	0.034	0.019	4.00	1.02	2.29	66.0	43.5	64.1	99.2
30	20	1.523	0.001	−0.113	4.27	1.94	2.84	61.5	57.8	42.8	77.3
50	20	1.420	−0.034	0.013	4.02	1.29	2.64	56.4	46.1	44.0	88.2
100	20	1.178	−0.014	−0.010	3.43	0.88	2.52	53.2	38.2	56.0	98.8
200	20	0.935	0.004	0.021	2.91	0.60	2.49	55.0	39.2	68.1	100.0

Notes: See Table 1 for an explanation.

Table 15: Simulation results with $\gamma = 0$ and $N_0 = 0.1N$.

		Bias				RMSE				5% size				
N	T	PC	PC*	FE	OLS	PC	PC*	FE	OLS	PC	PC*	FE	OLS	WAL
Without a break in the loadings ($\mu = 0$)														
30	6	−0.026	−0.031	−0.002	−0.004	1.32	1.32	1.20	1.16	21.80	30.2	19.9	24.1	19.4
50	6	0.056	0.030	0.009	0.019	0.96	1.10	0.91	0.89	17.80	28.3	12.2	13.9	14.1
100	6	−0.013	−0.012	−0.027	−0.025	0.64	0.75	0.66	0.63	12.90	23.9	9.7	10.3	8.4
200	6	0.007	0.021	0.026	0.025	0.46	0.54	0.47	0.45	11.40	24.5	8	7.9	6.3
30	10	0.018	0.007	0.026	0.013	1.19	1.36	1.23	1.17	27.70	40	21.4	26.5	19.6
50	10	0.015	0.014	0.036	0.030	0.88	0.99	0.92	0.88	21.60	28.2	12.8	15.9	12.8
100	10	0.036	0.018	−0.009	0.007	0.62	0.72	0.65	0.62	15.90	22.2	9.3	9.6	8.6
200	10	−0.002	0.004	0.004	0.006	0.41	0.48	0.48	0.47	13	16.9	8.5	8.6	8.5
30	20	0.031	0.028	0.029	0.032	1.14	1.36	1.17	1.14	40.9	50.8	22.2	26.3	22.2
50	20	−0.032	−0.037	−0.036	−0.026	0.83	0.93	0.93	0.90	35.2	37.4	14.3	16.4	13.4
100	20	0.009	0.010	0.021	0.025	0.56	0.64	0.63	0.61	28	30.4	9.6	8.9	8.9
200	20	0.001	−0.010	0.010	0.013	0.39	0.46	0.45	0.43	25.5	27.4	6.6	6.6	6.6
With a break in the loadings ($\mu = 1$)														
30	6	0.606	−0.129	−0.129	−0.125	1.88	1.67	1.69	1.68	29.90	34.1	30.4	35.6	32.6
50	6	0.608	0.000	0.005	0.006	1.51	1.36	1.46	1.45	26.90	29.4	28.3	31.8	35.3
100	6	0.527	0.040	0.018	0.021	1.19	1.14	1.36	1.36	28.00	32.3	35.9	39.5	50.5
200	6	0.509	0.081	0.051	0.051	1.06	0.89	1.29	1.29	35.70	36.4	51.6	51.8	71.1
30	10	0.759	0.042	0.071	0.054	1.88	1.67	1.77	1.71	40.10	46.8	31.8	35.9	37.6
50	10	0.523	−0.006	−0.093	−0.089	1.47	1.33	1.67	1.63	31.30	36.2	32.8	35	43.2
100	10	0.517	0.030	0.082	0.087	1.15	0.88	1.38	1.37	32.3	29.2	38.6	41.1	60
200	10	0.442	0.017	−0.008	−0.013	0.96	0.63	1.36	1.36	39.7	25.1	53.3	57.3	81.1
30	20	0.580	0.097	0.020	0.053	1.63	1.50	1.82	1.78	48.6	56.2	36.1	40.1	42.5
50	20	0.448	0.020	0.025	0.027	1.23	1.04	1.58	1.56	42.4	40.4	33.9	36.2	50.1
100	20	0.356	0.025	−0.038	−0.036	0.94	0.71	1.50	1.49	40.7	33.4	42.2	46.1	69.4
200	20	0.239	−0.025	−0.050	−0.054	0.71	0.52	1.48	1.48	38.4	29.7	59.9	60.5	89.7

Notes: "OLS" refers to the OLS estimator based on raw data. See Table 1 for an explanation of the rest.

Table 16: Simulation results for X1.

		Bias		RMSE		5% size		Wald		
N	T	FE	INF	FE	INF	FE	INF	W_1	W_2	$W_{1\&2}$
No correlation: $\Lambda_1 = \Lambda_2 = 0$										
30	6	0.189	0.165	0.242	0.209	7.8	7.9	9.2	9.6	12.8
50	6	0.142	0.137	0.179	0.172	7.4	7.1	9.8	6.1	10.3
100	6	0.099	0.096	0.124	0.121	5.3	4.8	7.8	5.2	7.4
200	6	0.068	0.069	0.084	0.085	5.1	4.0	6.4	6.5	6.6
30	10	0.186	0.164	0.235	0.206	8.4	7.2	10.9	10.1	14.0
50	10	0.135	0.127	0.173	0.161	7.2	6.4	8.3	6.1	11.1
100	10	0.092	0.089	0.117	0.114	6.4	6.4	6.1	6.3	8.5
200	10	0.066	0.066	0.083	0.082	5.0	5.0	5.9	6.6	6.8
30	20	0.164	0.153	0.204	0.191	7.4	6.9	10.4	9.2	13.3
50	20	0.122	0.118	0.153	0.148	5.7	5.4	7.7	5.9	9.0
100	20	0.089	0.087	0.111	0.109	5.6	5.3	8.4	6.8	8.5
200	20	0.062	0.062	0.077	0.077	5.8	5.6	4.9	6.1	5.8
Loading correlation: $\Lambda_1 = 0$, $\Lambda_2 = 0.5$										
30	6	0.203	0.165	0.262	0.209	10.6	7.9	10.1	44.8	45.5
50	6	0.160	0.137	0.205	0.172	10.4	7.1	10.5	69.6	67.5
100	6	0.116	0.096	0.147	0.121	9.8	4.8	8.7	90.1	89.1
200	6	0.081	0.069	0.101	0.085	9.6	4.0	6.9	98.8	98.1
30	10	0.196	0.164	0.247	0.206	8.9	7.2	11.2	48.1	48.2
50	10	0.151	0.127	0.192	0.161	10.4	6.4	9.7	74.2	72.0
100	10	0.103	0.089	0.130	0.114	8.6	6.4	6.4	93.8	92.1
200	10	0.074	0.066	0.092	0.082	7.9	5.0	6.8	99.7	99.2
30	20	0.169	0.153	0.212	0.191	8.3	6.9	10.1	44.3	42.7
50	20	0.128	0.118	0.160	0.148	6.8	5.4	7.8	70.3	68.3
100	20	0.093	0.087	0.116	0.109	6.4	5.3	7.8	93.4	90.6
200	20	0.065	0.062	0.082	0.077	6.8	5.6	5.0	99.8	99.4
Slope correlation: $\Lambda_1 = 0.5$, $\Lambda_2 = 0$										
30	6	0.663	0.650	0.839	0.818	57.4	58.1	72.6	9.4	72.2
50	6	0.555	0.563	0.703	0.709	58.4	57.3	93.3	6.2	91.7
100	6	0.422	0.428	0.522	0.530	59.1	56.9	100.0	5.5	99.9
200	6	0.294	0.302	0.373	0.385	51.2	50.1	100.0	6.7	100.0
30	10	0.537	0.532	0.676	0.670	50.9	52.8	63.4	10.0	62.6
50	10	0.453	0.453	0.557	0.559	56.6	56.3	88.7	6.3	86.0
100	10	0.324	0.325	0.409	0.411	51.2	49.9	99.6	7.0	99.2
200	10	0.246	0.248	0.308	0.311	52.4	52.6	100.0	5.7	100.0
30	20	0.402	0.404	0.501	0.503	44.3	46.7	51.8	8.7	48.7
50	20	0.337	0.341	0.420	0.423	45.4	43.3	78.3	5.8	73.5
100	20	0.255	0.257	0.317	0.319	47.4	47.8	98.4	5.1	97.1
200	20	0.186	0.186	0.234	0.235	46.5	47.1	100.0	5.8	100.0

Notes: "FE" and "INF" refer to the fixed effects OLS estimator and the infeasible OLS estimator based on knowing F_t. "W_1", "W_2" and "$W_{1\&2}$" refer to the Wald tests for testing $\Lambda_1 = 0$, $\Lambda_2 = 0$ and $\Lambda_1 = \Lambda_2 = 0$, respectively.

Table 17: Simulation results for X2.

		Bias		RMSE		5% size		Wald		
N	T	FE	INF	FE	INF	FE	INF	W_1	W_2	$W_{1\&2}$
No correlation: $\Lambda_1 = \Lambda_2 = 0$										
30	6	0.212	0.176	0.268	0.221	8.7	8.0	11.4	9.5	15.6
50	6	0.141	0.136	0.182	0.171	6.3	5.9	10.1	7.1	12.8
100	6	0.107	0.100	0.133	0.126	6.0	5.5	9.5	8.4	11.6
200	6	0.075	0.071	0.093	0.089	6.1	4.7	6.3	5.7	6.9
30	10	0.189	0.158	0.241	0.197	7.9	5.3	10.6	7.4	13.4
50	10	0.134	0.123	0.169	0.153	6.1	5.4	8.9	7.6	10.9
100	10	0.097	0.091	0.121	0.114	5.7	5.3	6.6	6.4	6.9
200	10	0.068	0.063	0.085	0.080	5.5	5.0	5.8	6.0	7.0
30	20	0.189	0.152	0.238	0.190	8.3	5.9	12.6	8.4	16.2
50	20	0.135	0.122	0.171	0.153	6.5	6.4	9.0	7.8	12.4
100	20	0.095	0.088	0.118	0.110	6.3	5.9	8.4	5.9	9.1
200	20	0.064	0.060	0.080	0.075	5.0	3.7	5.2	7.0	7.1
Loading correlation: $\Lambda_1 = 0$, $\Lambda_2 = 0.5$										
30	6	0.270	0.176	0.349	0.221	16.2	8.0	14.0	54.5	55.8
50	6	0.187	0.136	0.259	0.171	14.5	5.9	12.2	71.7	71.0
100	6	0.143	0.100	0.188	0.126	15.3	5.5	10.6	88.4	88.2
200	6	0.114	0.071	0.157	0.089	19.8	4.7	7.4	97.8	97.7
30	10	0.249	0.158	0.333	0.197	16.4	5.3	13.2	59.8	60.1
50	10	0.182	0.123	0.240	0.153	14.4	5.4	10.9	73.4	71.0
100	10	0.141	0.091	0.182	0.114	16.6	5.3	7.2	94.2	92.6
200	10	0.106	0.063	0.141	0.080	18.9	5.0	6.5	98.9	98.9
30	20	0.244	0.152	0.315	0.190	15.7	5.9	15.4	56.8	57.9
50	20	0.180	0.122	0.230	0.153	15.6	6.4	11.0	74.3	71.7
100	20	0.130	0.088	0.167	0.110	16.0	5.9	8.0	94.4	92.7
200	20	0.111	0.060	0.146	0.075	24.0	3.7	5.1	99.9	99.4
Slope correlation: $\Lambda_1 = 0.5$, $\Lambda_2 = 0$										
30	6	0.908	0.811	1.170	1.014	59.9	62.6	85.6	8.9	86.2
50	6	0.808	0.742	1.030	0.937	65.2	64.6	97.0	8.3	97.3
100	6	0.686	0.634	0.880	0.796	68.4	68.3	100.0	7.2	99.9
200	6	0.627	0.571	0.804	0.720	75.9	73.6	100.0	5.0	100.0
30	10	0.878	0.754	1.120	0.941	63.5	64.4	84.8	7.9	82.7
50	10	0.767	0.687	0.964	0.850	68.9	67.5	96.8	8.9	96.0
100	10	0.684	0.611	0.860	0.764	72.9	71.4	99.9	6.3	99.9
200	10	0.625	0.556	0.786	0.695	79.0	76.4	100.0	5.5	100.0
30	20	0.767	0.634	0.975	0.795	61.3	60.5	79.3	9.2	78.8
50	20	0.673	0.579	0.851	0.729	65.0	61.6	91.0	7.6	88.9
100	20	0.627	0.551	0.786	0.688	74.0	71.6	99.7	7.3	99.3
200	20	0.614	0.533	0.758	0.655	81.7	79.9	100.0	6.1	100.0

Notes: See Table 16 for an explanation.

Table 18: Simulation results for X3.

N	T	Bias FE	Bias INF	RMSE FE	RMSE INF	5% size FE	5% size INF	Wald W_1	Wald W_2	Wald $W_{1\&2}$
		No correlation: $\Lambda_1 = \Lambda_2 = 0$								
30	6	0.228	0.183	0.290	0.232	9.8	8.6	13.6	10.2	17.8
50	6	0.152	0.140	0.195	0.176	6.4	6.6	12.3	6.7	13.1
100	6	0.115	0.105	0.143	0.132	6.9	6.0	9.8	7.6	10.8
200	6	0.082	0.076	0.103	0.095	6.6	5.3	7.0	5.1	6.6
30	10	0.207	0.169	0.267	0.212	9.8	7.6	12.5	8.1	15.6
50	10	0.145	0.130	0.183	0.162	6.4	5.5	10.1	7.2	12.0
100	10	0.105	0.097	0.131	0.121	6.7	5.4	7.2	6.5	7.7
200	10	0.074	0.070	0.093	0.087	6.1	6.1	6.4	5.8	7.3
30	20	0.202	0.158	0.255	0.197	8.3	5.6	13.1	8.5	17.2
50	20	0.143	0.128	0.182	0.160	6.0	6.5	11.1	7.5	12.1
100	20	0.100	0.092	0.125	0.115	6.4	5.4	9.0	6.2	10.2
200	20	0.068	0.064	0.086	0.079	5.0	4.1	5.9	6.4	6.6
		Loading correlation: $\Lambda_1 = 0$, $\Lambda_2 = 0.5$								
30	6	0.325	0.183	0.418	0.232	21.7	8.6	18.6	71.0	72.6
50	6	0.244	0.140	0.332	0.176	20.3	6.6	15.3	85.4	85.7
100	6	0.195	0.105	0.262	0.132	24.8	6.0	12.9	94.2	93.9
200	6	0.169	0.076	0.235	0.095	32.8	5.3	10.5	97.7	97.7
30	10	0.301	0.169	0.407	0.212	20.6	7.6	17.2	76.7	77.5
50	10	0.231	0.130	0.309	0.162	21.1	5.5	14.5	84.9	84.1
100	10	0.184	0.097	0.245	0.121	24.9	5.4	10.7	95.6	95.9
200	10	0.147	0.070	0.200	0.087	29.5	6.1	9.5	98.4	98.4
30	20	0.291	0.158	0.376	0.197	19.8	5.6	18.6	71.3	73.2
50	20	0.219	0.128	0.282	0.160	20.5	6.5	13.9	87.2	85.5
100	20	0.161	0.092	0.210	0.115	22.1	5.4	9.3	97.9	97.8
200	20	0.145	0.064	0.195	0.079	30.6	4.1	7.3	99.2	99.2
		Slope correlation: $\Lambda_1 = 0.5$, $\Lambda_2 = 0$								
30	6	1.635	1.478	2.145	1.862	68.8	71.9	96.8	8.8	97.0
50	6	1.482	1.395	1.915	1.767	71.6	72.2	99.9	7.9	100.0
100	6	1.266	1.246	1.655	1.585	74.1	75.3	100.0	7.1	100.0
200	6	1.144	1.097	1.482	1.397	79.1	77.9	100.0	5.6	100.0
30	10	1.682	1.463	2.194	1.849	73.7	72.5	96.4	8.5	96.9
50	10	1.466	1.337	1.873	1.674	76.1	75.5	99.7	8.2	99.9
100	10	1.312	1.226	1.671	1.546	79.6	77.2	100.0	5.5	100.0
200	10	1.204	1.142	1.539	1.442	82.3	81.3	100.0	5.2	100.0
30	20	1.529	1.279	1.968	1.616	72.3	72.4	94.8	9.8	94.4
50	20	1.329	1.164	1.700	1.472	75.5	75.2	98.7	7.1	98.2
100	20	1.249	1.130	1.573	1.413	81.6	80.6	100.0	7.3	100.0
200	20	1.220	1.097	1.517	1.357	87.4	86.0	100.0	6.7	100.0

Notes: See Table 16 for an explanation.

Table 19: Simulation results for X4.

		Bias		RMSE		5% size		Wald		
N	T	FE	INF	FE	INF	FE	INF	W_1	W_2	$W_{1\&2}$
No correlation: $\Lambda_1 = \Lambda_2 = 0$										
30	6	0.536	0.386	0.686	0.494	6.2	8.4	6.7	–	–
50	6	0.416	0.302	0.544	0.387	6.1	6.1	5.6	–	–
100	6	0.288	0.212	0.373	0.276	4.4	6.1	5.2	–	–
200	6	0.207	0.151	0.269	0.194	7.7	5.2	5.9	–	–
30	10	0.561	0.352	0.730	0.448	8.5	9.3	6.5	–	–
50	10	0.411	0.272	0.529	0.344	6.3	6.0	5.0	–	–
100	10	0.276	0.183	0.366	0.231	5.2	5.0	5.4	–	–
200	10	0.205	0.124	0.266	0.158	4.3	4.7	5.1	–	–
30	20	0.522	0.309	0.689	0.390	7.2	8.3	7.5	–	–
50	20	0.395	0.237	0.515	0.298	5.3	6.9	5.6	–	–
100	20	0.271	0.164	0.362	0.207	6.0	5.7	4.8	–	–
200	20	0.200	0.117	0.267	0.149	5.5	6.5	5.1	–	–
Loading correlation: $\Lambda_1 = 0$, $\Lambda_2 = 0.5$										
30	6	0.595	0.386	0.776	0.494	8.6	8.4	10.3	–	–
50	6	0.470	0.302	0.615	0.387	9.0	6.1	11.5	–	–
100	6	0.361	0.212	0.473	0.276	9.8	6.1	16.2	–	–
200	6	0.311	0.151	0.408	0.194	17.3	5.2	28.8	–	–
30	10	0.599	0.352	0.800	0.448	9.5	9.3	10.1	–	–
50	10	0.482	0.272	0.630	0.344	9.4	6.0	13.6	–	–
100	10	0.363	0.183	0.488	0.231	11.0	5.0	19.8	–	–
200	10	0.311	0.124	0.415	0.158	17.6	4.7	35.2	–	–
30	20	0.565	0.309	0.748	0.390	9.2	8.3	12.8	–	–
50	20	0.461	0.237	0.610	0.298	8.4	6.9	14.6	–	–
100	20	0.373	0.164	0.499	0.207	13.2	5.7	21.4	–	–
200	20	0.323	0.117	0.431	0.149	19.3	6.5	37.4	–	–
Slope correlation: $\Lambda_1 = 0.5$, $\Lambda_2 = 0$										
30	6	0.585	0.445	0.746	0.563	9.3	12.9	6.7	–	–
50	6	0.464	0.365	0.598	0.460	8.7	12.8	5.6	–	–
100	6	0.353	0.297	0.445	0.366	12.3	16.3	5.2	–	–
200	6	0.306	0.268	0.373	0.316	18.9	30.7	5.9	–	–
30	10	0.599	0.410	0.774	0.512	10.2	12.9	6.5	–	–
50	10	0.461	0.337	0.588	0.419	8.8	13.5	5.0	–	–
100	10	0.359	0.289	0.450	0.348	11.8	22.0	5.4	–	–
200	10	0.301	0.259	0.369	0.296	20.4	32.2	5.1	–	–
30	20	0.547	0.366	0.716	0.457	10.0	13.0	7.5	–	–
50	20	0.439	0.320	0.559	0.388	9.2	15.8	5.6	–	–
100	20	0.342	0.274	0.426	0.323	12.8	22.8	4.8	–	–
200	20	0.294	0.251	0.363	0.286	22.0	39.0	5.1	–	–

Notes: See Table 16 for an explanation.