Testing for correlation between the regressors and factor loadings in heterogeneous panels with interactive effects

George Kapetanios[1] · Laura Serlenga[2,4] · Yongcheol Shin[3]

Abstract

A large literature on modelling cross-section dependence in panels has been developed through interactive effects. However, there are areas where research has not really caught on yet. One such area is the one concerned with whether the regressors are correlated with factor loadings or not. This is an important issue because if the regressors are uncorrelated with loadings, we can simply use the consistent two-way fixed effects (FE) estimator without employing any more sophisticated econometric methods such as the principal component (PC) or the common correlated effects estimators. We explore this issue, which has received surprisingly little attention and propose a Hausman-type test to address the matter. Further, we develop two nonparametric variance estimators for the FE and PC estimators as well as their difference, that are robust to the presence of heteroscedasticity, autocorrelation and slope heterogeneity. Under the null hypothesis of no correlation between the regressors and loadings the proposed test follows the χ^2 distribution asymptotically. Monte Carlo simulation results confirm satisfactory size and power performance of the test even in small samples. Finally, we provide extensive empirical evidence in favour of uncorrelated factor loadings. In this situation, the FE estimator would provide a simple and robust estimation strategy which is invariant to nontrivial computational issues associated with the PC estimator.

✉ Yongcheol Shin
  yongcheol.shin@york.ac.uk

1 King's College London, London, UK

2 University of Bari, Bari, Italy

3 University of York, York, UK

4 European Commission, JRCIspra, Italy

1 Introduction

Panel data models have been increasingly popular in applied economics and finance, due to their ability to model various sources of heterogeneity. A standard practice is to impose strong restrictions on error cross-section dependence (CSD). This takes the form of independence across individual units under the fixed effects model whilst a common time effect severely restricts the nature of CSD under the random effects specification. However, most recently, a large number of studies have developed econometric methodologies for modelling CSD, mainly through the structure of interactive effects (hereafter IE) introducing heterogeneous unobserved factors into the error components and allowing for a richer cross-sectional covariance structure.

In this framework, conventional wisdom has been that the standard two-way fixed effects (FE) estimator would be inconsistent, due to ignoring the potential endogeneity arising from the correlation between the regressors and factors and/or factor loadings (e.g. Bai 2009). Hence, two leading approaches have proposed in the literature, see Chudik and Pesaran (2015) for a survey. The first, based on the principal component (PC) estimation, estimates the factors jointly and iteratively with the main slope parameters, see Bai (2009) and Moon and Weidner (2015), Fernandez-Val and Weidner (2016) and Charbonneau (2017) for extensions. The second approach, advanced by Pesaran (2006), treats factors as nuisance terms, and removes their effects through proxying them by the cross-section averages of the dependent and independent variables. This is referred to as the common correlated effects (CCE) estimator. A growing number of extensions have been developed by Kapetanios et al. (2011), Chudik and Pesaran (2015), Westerlund and Urbain (2015) and Petrova and Westerlund (2020).

In the empirical work, the CCE estimator is mostly used as this is easier to implement with respect to the PC. Indeed, a common practice is to apply the CCE estimator after detecting the existence of strong CSD by the Pesaran (2015) CD test, see Mastromarco et al. (2016), Holly et al. (2010), Baltagi and Li (2014), among others.

This paper contributes to this literature by raising some important issues which might be considered relevant for practitioners. We start by highlighting a simple fact that the FE estimator is not always inconsistent even in the presence of IE. If the regressors are correlated with factors but uncorrelated with loadings, then the FE estimator is shown to be consistent, which has also been noted earlier by Coakely et al. (2006), Sarafidis and Wansbeek (2012) and Westerlund (2019a). In such a case, we formally show that the FE estimator is consistent and asymptotically normally distributed. But, the variance estimator provided by the standard FE estimation will be invalid due to the presence of remaining zero-mean IE in the error components. Hence, we provide two consistent nonparametric variance estimators that are robust to the presence of heteroscedastic and serially-correlated disturbances as well as the slope parameters heterogeneity. Via Monte Carlo studies, we find that FE and CCE

estimators display a similar and satisfactory performance when the regressors are correlated with factors but uncorrelated with loadings. Further, the coverage rate of the FE estimator evaluated using nonparametric variance estimators reaches the nominal 95%. The performance of both CCE and FE estimators worsens significantly if the regressors are correlated with loadings, which is in line with Westerlund and Urbain (2013). As expected, the performance of the PC estimator is not unduly affected by the presence of correlation between the regressors and loadings.

Furthermore, we point out that in the specification tests that have been proposed in the literature to testing the presence of the CSD or IE, e.g. Pesaran (2015), Sarafidis et al. (2009), Bai (2009) and Castagnetti et al. (2015), the rejection of the null hypothesis does not always imply that the FE estimator is inconsistent under the alternative model with IE. For instance, Sarafidis et al. (2009) maintain an assumption that factor loadings (between equations for the dependent variable and the regressors) are uncorrelated even under the alternative. More importantly, we show that the Hausman test for the null hypothesis of the two way additive fixed effects against the alternative hypothesis of IE proposed by Bai (2009), would be inconsistent against the alternative, especially if the regressors are uncorrelated with loadings. This suggests that the presence of no correlation between the regressors and loadings emerges as an influential but under-appreciated feature of the panel data model with IE. For large T, in order to avoid any potential omitted variables bias, it is natural to allow for the regressors to be correlated with unobserved factors. But, it still remains the important issue to test whether the regressors are correlated with loadings or not in practice.

Despite a growing number of studies on modelling CSD through IE, it is rather surprising to find that the literature has been silent on investigating the important issue of testing the validity of correlation between the regressors and factor loadings in panels with IE. This is the important hypothesis to be tested because if the loadings are uncorrelated with the regressors, we can just use the simple but consistent FE estimator. In what follows we develop the Hausman-type test that determines the validity of whether the regressors are correlated with loadings. Both the FE and PC estimators are consistent under the null hypothesis of uncorrelated factor loadings whilst only the latter is consistent under the alternative hypothesis. Our proposed test is different from the Hausman test developed by Bai (2009), because our null hypothesis is subsumed under his alternative model with IE. As a result, the FE estimator is not necessarily more efficient than the PC estimator under the null hypothesis. Based on this idea, we develop two nonparametric variance estimators for the difference between the FE and PC estimators, that are shown to be robust to the presence of heteroscedasticity, autocorrelation and slope heterogeneity. We derive that the proposed test statistic follows the χ^2 distribution asymptotically. Monte Carlo simulation results confirm that the size and the power performance of the test is quite satisfactory even in small samples.

Finally, our most important contribution is the provision of extensive empirical evidence that regressors are uncorrelated with factor loadings, in many panel datasets employed in the literature. We find that the null hypothesis of the regressors being uncorrelated with factor loadings, is not rejected in thirteen out of fourteen datasets considered. Next, we find that Bai's Hausman test rejects the null of additive effects model against the alternative of IE only once whilst the CD test by Pesaran (2015)

strongly rejects the null of weak CSD for all the datasets. Such conflicting results could provide an additional support for our main findings that the regressors are indeed uncorrelated with factor loadings even in cross-sectionally correlated panels with IE, in which case we show that Bai's Hausman test is inconsistent. Furthermore, the FE estimator is invariant to any complex issues related to selecting the number of unobserved factors incorrectly which would significantly affect the performance of PC estimators (Moon and Weidner 2015), and to employing the inconsistent initial estimates which may not guarantee the convergence of the iterative PC estimator (Hsiao 2018). In this regard, the FE estimation combined with nonparametric variance estimators will provide the simple and robust approach, avoiding uncertainty in specifying and estimating nuisance parameters for potential efficiency gain. This suggests that the FE estimator can still be of considerable applicability in a wide variety of cross-sectionally correlated panel data, especially if the regressors are found to be uncorrelated with factor loadings, the validity of which can be easily verified by our proposed test.

The paper proceeds as follows. Section 2 describes the model and highlights that the FE estimator is still consistent in panels with IE, under the condition that the regressors are uncorrelated with factor loadings. Section 3 develops the Hausman-type test for the validity of correlated factor loadings, which is the important hypothesis to be tested. Section 4 employs a range of Monte Carlo simulations to investigate the finite sample performance of the alternative estimators and the proposed test statistic. Section 5 presents empirical evidence documenting that the null hypothesis of the regressors uncorrelated with factor loadings, is not rejected for thirteen out of fourteen datasets. Section 6 offers some concluding remarks. Mathematical proofs, the data descriptions and additional empirical results are relegated to Appendices. Additional simulation results can be found in Online Supplement.

2 The model

Consider the following heterogeneous panel data model with IE:

$$y_{it} = \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it} \tag{1}$$

where y_{it} is the dependent variable of the ith cross-sectional unit in period t, $\boldsymbol{x}_{it}$ is the $k \times 1$ vector of covariates with $\boldsymbol{\beta}_i$ the $k \times 1$ vector of parameters, and ε_{it}'s are idiosyncratic errors. $\boldsymbol{f}_t$ is an $r \times 1$ vector of unobserved common factors while $\boldsymbol{\gamma}_i$ is an $r \times 1$ vector of random heterogeneous loadings.

We make the following assumptions:

Assumption A (i) ε_{it} is independently distributed across i with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_{\varepsilon_i}^2$ and $E\left(\varepsilon_{it}^{8+\delta}\right) < \infty$ for some $\delta > 0$. Each ε_{it} follows a linear process with absolutely summable autocovariances such that $\lim_{T\to\infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} |E(\varepsilon_{is}\varepsilon_{it})|^{1+\delta} < \infty$, $E\left|N^{-1/2}\sum_{i=1}^{N} [\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})]\right|^4 < \infty$ for all t, s, and $\lim_{T,N\to\infty} T^{-2}N^{-1} \sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{w=1}^{T} |Cov(\varepsilon_{is}\varepsilon_{it}, \varepsilon_{ir}\varepsilon_{iw})| < \infty$. The largest eigenvalue of

$E\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i'\right)$ is bounded uniformly in every i and t, where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. ε_{it} is independent of $\boldsymbol{x}_{js}$, $\boldsymbol{\gamma}_j$ and $\boldsymbol{f}_s$ for all i, j, s and t.

(ii) $\boldsymbol{f}_t$ is covariance stationary with finite mean and variance, $\boldsymbol{\Sigma}_f$ with $E\left(\left\|\boldsymbol{f}_t\right\|^4\right) < \infty$ where $\boldsymbol{\Sigma}_f$ is an $r \times r$ positive definite matrix.

(iii) $\boldsymbol{\gamma}_i$ is *iid* across i with finite mean and variance, $\boldsymbol{\Sigma}_\gamma$, where $\boldsymbol{\Sigma}_\gamma$ is an $r \times r$ positive definite matrix. $\boldsymbol{\gamma}_i$ are independent of ε_{jt} and $\boldsymbol{f}_t$ for all i, j and t.

(iv) The $k \times 1$ vector of $\boldsymbol{\beta}_i$ are generated as $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$. $\boldsymbol{\eta}_i$ is independent across i with $E\left(\boldsymbol{\eta}_i\right) = 0$ and $E\left(\boldsymbol{\eta}_i \boldsymbol{\eta}_i'\right) = \Omega_{\eta\eta,i}$, where $\Omega_{\eta\eta,i}$ is a positive definite matrix uniformly for every i. $E\left\|\boldsymbol{\eta}_i\right\|^4 \leq \Delta < \infty$ and $\|\boldsymbol{\beta}\| < \infty$. $\boldsymbol{\eta}_i$ is independent of ε_{it} and $\boldsymbol{\gamma}_i$.

Assumption A is standard in the literature, see Bai (2009), Karabiyik et al. (2017) and Cui et al. (2019) (CHNY, hereafter).

For a consistent estimation of the parameters in (1), we need to first account for unobserved factors, and then estimate $\boldsymbol{\beta}$ by applying panel estimators to (1) with defactored variables. On the basis of this idea, two popular approaches have been proposed. The common correlated effects (CCE) estimator advanced by Pesaran (2006), imposes that $\boldsymbol{x}_{it}$ share the same factors, $\boldsymbol{f}_t$

$$\boldsymbol{x}_{it} = \boldsymbol{\Gamma}_i' \boldsymbol{f}_t + \boldsymbol{v}_{it} \tag{2}$$

where $\boldsymbol{\Gamma}_i$ an $r \times k$ matrix of random heterogeneous loadings and $\boldsymbol{v}_{it}$ are idiosyncratic errors, and proposes to approximate $\boldsymbol{f}_t$ by the cross-section averages of the dependent and independent variables. Next, Bai (2009) allows $\boldsymbol{x}_{it}$ to be arbitrarily correlated with both $\boldsymbol{\gamma}_i$ and $\boldsymbol{f}_t$, and proposes the iterative principal component (PC) approach that estimates the factors jointly and iteratively with the slope parameters. The validity of the CCE approach depends crucially upon whether an appropriate rank condition, that has to be assumed, holds. Westerlund and Urbain (2015) argue that the issue of correctly selecting the number of factors, r in the PC estimation, is essentially the same as the issue of satisfying the condition, $r \leq k + 1$ in CCE estimation. Further, it is shown that both estimators involve bias terms, which do not disappear unless $N/T \to 0$. The finite sample performance of the two approaches has been intensively investigated. The earlier studies by Kapetanios and Pesaran (2005) and Chudik et al. (2011), provide Monte Carlo evidence in favour of the CCE estimator, which is partly due to uncertainty associated with estimating the true number of unobserved factors in the PC estimation. Further, Westerlund and Urbain (2015) show that the performance of the PC estimator is sensitive to the value of β. For $\beta = 0$, the PC estimator outperforms CCE, while for $\beta \neq 0$, the CCE estimator tends to outperform.

However, we find that the performance of the two-way fixed effect (FE) estimator has not been widely investigated. Exceptions include the studies by Coakely et al. (2006), Sarafidis and Wansbeek (2012) and Westerlund (2019a). This simply reflects the conventional view that the FE estimator would be inconsistent in the presence of IE, due to ignoring endogeneity stemming from the correlation between regressors and factors/loadings. We aim to challenge this maintained view. For large T, suppose that $\boldsymbol{f}_t$ may represent the unobserved common policy or globalisation trend, and $\boldsymbol{\gamma}_i$ are the heterogeneous individual responses (parameters).

In practice, it is important to test the validity of whether $\boldsymbol{x}_{it}$ are correlated with $\boldsymbol{\gamma}_i$ or not. Formally, we set the null and alternative hypothesis as follows:

$$H_0 : \boldsymbol{x}_{it} \text{ uncorrelated with } \boldsymbol{\gamma}_i \tag{3}$$

$$H_1 : \boldsymbol{x}_{it} \text{ correlated with } \boldsymbol{\gamma}_i \tag{4}$$

Under Assumptions A(ii) and (iii), we can express $\boldsymbol{\gamma}_i' \boldsymbol{f}_t$ in (1) by[1]

$$\boldsymbol{\gamma}_i' \boldsymbol{f}_t = \mu + \alpha_i + \theta_t + \mathring{\boldsymbol{\gamma}}_i' \dot{\boldsymbol{f}}_t \tag{5}$$

where $\mu = \bar{\boldsymbol{\gamma}}' \bar{\boldsymbol{f}}$, $\alpha_i = \boldsymbol{\gamma}_i' \bar{\boldsymbol{f}}$, $\theta_t = \bar{\boldsymbol{\gamma}}' \boldsymbol{f}_t$, $\mathring{\boldsymbol{\gamma}}_i = \boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}}$ and $\dot{\boldsymbol{f}}_t = \boldsymbol{f}_t - \bar{\boldsymbol{f}}$ with $\bar{\boldsymbol{\gamma}} = N^{-1} \sum_{i=1}^N \boldsymbol{\gamma}_i$ and $\bar{\boldsymbol{f}} = T^{-1} \sum_{t=1}^T \boldsymbol{f}_t$. Using (5) in (1), we have:

$$y_{it} = \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \mu + \alpha_i + \theta_t + \mathring{\boldsymbol{\gamma}}_i' \dot{\boldsymbol{f}}_t + \varepsilon_{it} \tag{6}$$

This transformation clearly shows that the panel data model with nonzero-mean IE, $\boldsymbol{\gamma}_i' \boldsymbol{f}_t$ in (1) can be equally expressed as the 2-way fixed effects panel data model with zero-mean IE, $\mathring{\boldsymbol{\gamma}}_i' \dot{\boldsymbol{f}}_t$ in (6).[2] Next, applying the 2-way within transformation to (6) to obtain[3]

$$\ddot{y}_{it} = \boldsymbol{\beta}_i' \ddot{\boldsymbol{x}}_{it} + \ddot{u}_{it}, \ \ddot{u}_{it} = \mathring{\boldsymbol{\gamma}}_i' \dot{\boldsymbol{f}}_t + \ddot{\varepsilon}_{it} \tag{7}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}$ with $y_{i.} = T^{-1} \sum_{t=1}^T y_{it}$, $y_{.t} = N^{-1} \sum_{i=1}^N y_{it}$, $\bar{y}_{..} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T y_{it}$, and similarly for $\ddot{x}_{it}$ and $\ddot{\varepsilon}_{it}$.

Under Assumption A and (3), it is easily seen by the independence of $\boldsymbol{\gamma}_i$ from all other random quantities in the model and $E\left(\mathring{\boldsymbol{\gamma}}_i\right) = E\left(\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}}\right) = 0$ that $\ddot{\boldsymbol{x}}_{it}$ is uncorrelated with the composite error, $\ddot{u}_{it} = \mathring{\boldsymbol{\gamma}}_i' \dot{\boldsymbol{f}}_t + \ddot{\varepsilon}_{it}$ in (7), provided $\boldsymbol{x}_{it}$ are strictly

---

[1] Hsiao (2018) argues that the assumption of zero mean for $\boldsymbol{\gamma}_i$ or $\boldsymbol{f}_t$ often used as normalization conditions, is not innocuous. With the mean zero assumption for $\boldsymbol{\gamma}_i$, the cross-sectional mean equation of (1)

$$\bar{y}_t = \boldsymbol{\beta}' \bar{\boldsymbol{x}}_t + \bar{\varepsilon}_t, \quad t = 1, \dots, T$$

no longer involves $\boldsymbol{\gamma}_i' \boldsymbol{f}_t$, where $\bar{y}_t = N^{-1} \sum_{i=1}^N y_{it}$, and similarly for $\bar{\boldsymbol{x}}_t$ and $\bar{\varepsilon}_t$. Thus the least squares regression of $\bar{y}_t$ on $\bar{\boldsymbol{x}}_t$ is consistent and asymptotically normally distributed as $T \to \infty$. Similarly, under the mean zero assumption for $\boldsymbol{f}_t$, the individual time series mean equation

$$\bar{y}_i = \bar{\boldsymbol{x}}_i' \boldsymbol{\beta} + \bar{\varepsilon}_i, \quad i = 1, \dots, N,$$

does not involve $\boldsymbol{\gamma}_i' \boldsymbol{f}_t$, where $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$, and similarly for $\bar{\boldsymbol{x}}_i$ and $\bar{\varepsilon}_i$. The least squares regression of $\bar{y}_i$ on $\bar{\boldsymbol{x}}_i$ can yield consistent and asymptotically normally distributed estimator of $\boldsymbol{\beta}$ if N is large. In this regard, we consider the general case with $E\left(\boldsymbol{\gamma}_i\right) \neq 0$ and $E\left(\boldsymbol{f}_t\right) \neq 0$.

[2] This may suggest that the additive case with $\boldsymbol{\gamma}_i = (\alpha_i, 1)'$ and $\boldsymbol{f}_t = (1, \theta_t + \mu)'$ may not always be the special case of the interactive effects, $\boldsymbol{\gamma}_i' \boldsymbol{f}_t$, as argued by Bai (2009), because we have both 2-way effects, $\alpha_i + \theta_t$ and zero-mean interactive effects, $\mathring{\boldsymbol{\gamma}}_i' \dot{\boldsymbol{f}}_t$ in (6).

[3] More precisely, we obtain: $\ddot{y}_{:it} = \boldsymbol{\beta}_i' \ddot{\boldsymbol{x}}_{it} + \ddot{u}_{it} + \ddot{\zeta}_{it}$ where $\ddot{\zeta}_{it} = (1 - \frac{1}{N}) \sum_{i=1}^N \boldsymbol{\eta}_i' (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i.})$. Under the assumption that $\boldsymbol{\eta}_i$ and $\boldsymbol{x}_{it}$ are uncorrelated, it follows that $\ddot{\zeta}_{it} \to_p 0$. When evaluating $\ddot{\zeta}_{it}$ numerically for the different combinations of (N, T), say $(N, T) = \{(25, 25), (50, 50), (100, 100)\}$, we find that its value is almost zero in all cases.

exogenous with respect to ε_{it} because

$$E\left(\ddot{\boldsymbol{x}}'_{it}\mathring{\boldsymbol{\gamma}}'_i\dot{\boldsymbol{f}}_t\right) = E\left\{\ddot{\boldsymbol{x}}'_{it}\dot{\boldsymbol{f}}'_t E\left(\mathring{\boldsymbol{\gamma}}_i|\ddot{\boldsymbol{x}}_{it}, \dot{\boldsymbol{f}}_t\right)\right\} = 0. \tag{8}$$

See also Section 5 in Hsiao (2018). Therefore, under the null hypothesis, (3), we can apply the two-way FE estimation to (1) and obtain a consistent estimator of $\boldsymbol{\beta}$ from (7). Conversely, if $\boldsymbol{x}_{it}$ and $\boldsymbol{\gamma}_i$ are correlated, it is clear that $E\left(\ddot{u}_{it}\boldsymbol{x}_{it}\right) \neq 0$ so that the FE estimator is inconsistent. Notice that the consistency of the FE estimator requires only $\boldsymbol{\gamma}_i$ to be uncorrelated with $\boldsymbol{x}_{it}$, but this is implicitly a maintained assumption in the CCE literature.[4] A further possibility that we do not entertain is that $\boldsymbol{x}_{it}$ contains a different set of factors to that entering y_{it} directly and that the two sets of factors are uncorrelated. This points out the symmetry of the role of loadings and factors in the IE setting. Then, (8) may hold even if (3) does not. However, we view this setting as too unlikely to be of interest.

The two-way FE estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N}\ddot{X}'_i\ddot{X}_i\right)^{-1}\sum_{i=1}^{N}\ddot{X}'_i\ddot{\boldsymbol{y}}_i \tag{9}$$

where $\ddot{X}_i = \left(\ddot{\boldsymbol{x}}_{i1}, \ldots, \ddot{\boldsymbol{x}}_{iT}\right)'$ and $\ddot{\boldsymbol{y}}_i = \left(\ddot{y}_{i1}, \ldots, \ddot{y}_{iT}\right)'$. As $\ddot{u}_{it}$ in (7) still contains zero-mean IE, $\mathring{\boldsymbol{\gamma}}'_i\dot{\boldsymbol{f}}_t$, the standard variance estimator for $\hat{\boldsymbol{\beta}}_{FE}$ will be invalid. Thus, we propose the two consistent variance estimators, which are also robust to the heteroscedasticity and the serial-correlation as well as the slope heterogeneities. The first is the nonparametric variance estimator, similarly applied in deriving the variance of the CCE estimator by Pesaran (2006):

$$\hat{\boldsymbol{V}}^{NON}\left(\hat{\boldsymbol{\beta}}_{FE}\right)$$
$$= \left(\sum_{i=1}^{N}\ddot{X}'_i\ddot{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\left(\ddot{X}'_i\ddot{X}_i\right)\left(\hat{\boldsymbol{\beta}}_{FE,i} - \bar{\boldsymbol{\beta}}_{FE}\right)\left(\hat{\boldsymbol{\beta}}_{FE,i} - \bar{\boldsymbol{\beta}}_{FE}\right)'\left(\ddot{X}'_i\ddot{X}_i\right)\right)\left(\sum_{i=1}^{N}\ddot{X}'_i\ddot{X}_i\right)^{-1} \tag{10}$$

where $\hat{\boldsymbol{\beta}}_{FE,i} = \left(\ddot{X}'_i\ddot{X}\right)^{-1}\ddot{X}'_i\ddot{\boldsymbol{y}}_i$ and $\bar{\boldsymbol{\beta}}_{FE} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{FE,i}$. Next, we consider the following heteroscedasticity and autocorrelation robust variance estimator (see CHNY):

---

[4] Pesaran (2006) implicitly assumes that the factor loadings $\boldsymbol{\gamma}_i$ in (1) and $\boldsymbol{\Gamma}_i$ in (2), are uncorrelated. Bai (2009) discusses this implication in detail, and shows via simulations that the CCE estimator is biased when $\boldsymbol{x}_{it}$ is correlated with both $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$. See also Westerlund and Urbain (2013) for the simulation evidence showing that the CCE estimator performs poorly when factor loadings are correlated. Remark 2 in Westerlund and Urbain (2013) also questions the uncorrelated factor loadings assumption by arguing that a common shock that has a positive effect on savings, should have negative effects on interest rates. However, their discussion relates to the sign of the average effect of common shocks or the sign of the cross-section mean of loadings. Since the independence assumption does not restrict the sign of these means, the relevance of such a relaxation would be somewhat questionable.

$$\hat{\boldsymbol{V}}^{HAC}\left(\hat{\boldsymbol{\beta}}_{FE}\right) = \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \hat{\boldsymbol{u}}_{FE,i} \hat{\boldsymbol{u}}_{FE,i}' \ddot{\boldsymbol{X}}_i\right) \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i\right)^{-1} \quad (11)$$

where $\hat{\boldsymbol{u}}_{FE,i} = \ddot{\boldsymbol{y}}_i - \ddot{\boldsymbol{X}}_i \hat{\boldsymbol{\beta}}_{FE}$.

We show that $\hat{\boldsymbol{\beta}}_{FE}$ is consistent and follows the normal distribution asymptotically under the null, (3). The result holds for both homogeneous and heterogeneous $\boldsymbol{\beta}$.

Theorem 1 *Under Assumption A and under (3), as $N, T \to \infty$,*

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) \to_d N\left(0_{k \times 1}, \boldsymbol{\Psi}_{FE}^{-1} \boldsymbol{R}_{FE} \boldsymbol{\Psi}_{FE}^{-1}\right) \quad (12)$$

where $\boldsymbol{\Psi}_{FE} = \lim_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\frac{\ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i}{T}\right)$. Considering $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$, $\boldsymbol{R}_{FE} = \boldsymbol{R}_{1,FE} + \boldsymbol{R}_{2,FE}$ where

$$\boldsymbol{R}_{1,FE} = \lim_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\frac{\ddot{\boldsymbol{X}}_i' \dot{\boldsymbol{F}}}{T} \mathring{\boldsymbol{\gamma}}_i \mathring{\boldsymbol{\gamma}}_i' \frac{\dot{\boldsymbol{F}}' \ddot{\boldsymbol{X}}_i}{T}\right) \quad (13)$$

$$\boldsymbol{R}_{2,FE} = \lim_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\frac{\ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i}{T} \boldsymbol{\eta}_i \boldsymbol{\eta}_i' \frac{\ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i}{T}\right), \quad (14)$$

and $\dot{\boldsymbol{F}} = \left(\dot{\boldsymbol{f}}_1, \dots, \dot{\boldsymbol{f}}_T\right)'$. Furthermore,

$$\hat{\boldsymbol{V}}^{NON}\left(\hat{\boldsymbol{\beta}}_{FE}\right)^{-1/2} \left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) \to_d N\left(0, \boldsymbol{I}_k\right) \text{ and } \hat{\boldsymbol{V}}^{HAC}\left(\hat{\boldsymbol{\beta}}_{FE}\right)^{-1/2}$$
$$\left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) \to_d N\left(0, \boldsymbol{I}_k\right). \quad (15)$$

If $\boldsymbol{\eta}_i = 0$, $\forall i$, then (12) and (15) continue to hold with $\boldsymbol{R}_{FE} = \boldsymbol{R}_{1,FE}$.

3 The Hausman-type test

A number of specification tests have been proposed to test the presence of the CSD or the multiplicative IE in panels. The most popular test is the cross-section dependence (CD) test statistic proposed by Pesaran (2015), that is increasingly applied to the residuals of regression models for use as an ex-post diagnostic tool. However, the CD test fails to reject the null hypothesis of no error CSD when the factor loadings have zero means, implying that the CD test will display very poor power when it is applied to cross-sectionally demeaned data. Furthermore, the residual-based CD test has been shown to often reject the null hypothesis of no remaining CSD in the case of the CCE estimator (e.g. Mastromarco et al. 2016). Juodis and Reese (2018) show that the application of the CD test to regression residuals obtained from IE models

introduces a bias term of order $\sqrt{T}$, rendering an erroneous rejection of the null.[5] Sarafidis et al. (2009) propose an alternative testing procedure for the null hypothesis of homogeneous factor loadings against the alternative of heterogeneous loadings after estimating a linear dynamic panel data model by GMM. This approach is valid only when N is large relative to T, but it can be applied to testing for any remaining error CSD after including time dummies. But, they maintain an assumption that the loadings between equations for y and x are uncorrelated (see their Assumption 5(b)).

The PC estimator is consistent both under models with two-way additive (fixed) effects and under models with IE, but less efficient than the FE estimator under the null model with additive effects only. But, the FE estimator is inconsistent under the alternative model with IE. Following this idea, Bai (2009, Section 9) advances the following Hausman test for testing the null of additive effects, *i.e.* $\gamma_i' f_t = \alpha_i + \theta_t$ against the alternative of IE[6]:

$$H_B = \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC_B}\right)' (V_B)^{-1} \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC_B}\right), \tag{16}$$

where $\hat{\boldsymbol{\beta}}_{PC_B}$ is the iterative PC estimator proposed by Bai (2009), $V_B = \widetilde{Var}\left(\hat{\boldsymbol{\beta}}_{PC_B}\right) - \widetilde{Var}\left(\hat{\boldsymbol{\beta}}_{FE}\right)$, $\widetilde{Var}\left(\hat{\boldsymbol{\beta}}_{FE}\right)$ is the the standard variance estimator provided by the two-way FE estimation, and $\widetilde{Var}\left(\hat{\boldsymbol{\beta}}_{PC_B}\right)$ is the analytic (sandwich-form) variance estimator, which takes into account unknown form of heteroscedastic and autocorrelated errors. Bai (2009) derives that $H_B \to_d \chi_k^2$ under the null.[7] Westerlund (2019b) proposes the alternative Hausman test statistic obtained by replacing the PC estimator with the CCE estimator, H_W.

The conventional wisdom is that if the null hypothesis of no error CSD is rejected, the FE estimator would be biased due to the potential endogeneity arising from the correlation between the regressors and unobserved factors and/or loadings.

In empirical applications we apply the CD test and Bai's Hausman test to the number of datasets that have been employed in the literature, and find that the CD test strongly rejects the null hypothesis of weak error CSD while the H_B test rarely rejects the null of additive-effects. The results of the CD test confirm the presence of strong CSD while the latter indicates the absence of IE. This suggests that the FE estimator is consistent (and potentially efficient). However, if the regressors are uncorrelated with loadings, the H_B test is inconsistent against the alternative model.

The results of the Monte Carlo simulation (in Section S1 in the Online Supplement) clearly demonstrate the limitation of applying the H_B in practice because it cannot

---

[5] Nonetheless, the CD test may be used as a model-selection tool, with a reduction in the absolute value of the CD test statistic typically being interpreted as an indication of an improved model specification.

[6] Focusing on the special cases, Castagnetti et al. (2015) propose two tests for the null of no factor structure: one for the null that factor loadings are cross sectionally homogeneous, and another for the null that common factors are homogeneous over time. Using extremes of the estimated loadings and common factors, they show that their statistics follow an asymptotic Gumbel distribution under the null. Furthermore, they show that the average-type statistics diverge under the null while the Hausman test is inconsistent.

[7] See Sections 3.2 and 9 in Bai (2009) for details.

distinguish between panels with the 2-way additive fixed effects and panels with IE where the regressors are uncorrelated with loadings.[8]

The above discussion suggests that the null hypothesis of the absence of correlation between the regressors and factor loadings emerges as an influential but underappreciated feature of the panel data model with IE.

We have shown that the presence of IE does not always imply that the FE estimator is inconsistent. In particular, the FE estimator is still consistent under the null (3), even though the regressors are correlated with factors. In this case we may prefer to use the simple FE estimator, which is invariant to any complex issues related to selecting the number of unobserved factors incorrectly which would significantly affect the performance of PC estimators (Moon and Weidner 2015), and to employing the inconsistent initial estimates which may not guarantee the convergence of the interactive PC estimator (Hsiao 2018).

In this regard, it is surprising to find that the literature has been silent on investigating the important issue of testing if the regressors are correlated with loadings or not in panels with IE. For large T context, it is natural to allow for x_{it} to be correlated with f_t to avoid any omitted variables bias. It still remains the important issue to test whether x_{it} are correlated with γ_i. Given the pervasive evidence of cross sectionally dependent errors in panels (Pesaran 2015), as the main contribution, we proceed to develop a novel Hausman-type test that investigates the validity of the null hypothesis, (3). In the model (1), recall that the PC estimator is consistent under the null, (3) and under the alternative, (4) whereas the FE estimator is consistent only under the null, (3). Following this idea, we propose the Hausman-type test based on the difference between the FE and PC estimators as follows:

$$H = \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC} \right)' V^{-1} \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC} \right) \tag{17}$$

where $\hat{\boldsymbol{\beta}}_{PC}$ is the bias corrected PC estimator to be defined in (18) below, and $V = Var\left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC} \right) = Var\left(\hat{\boldsymbol{\beta}}_{FE} \right) + Var\left(\hat{\boldsymbol{\beta}}_{PC} \right) - Cov\left(\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{PC} \right) - Cov\left(\hat{\boldsymbol{\beta}}_{PC}, \hat{\boldsymbol{\beta}}_{FE} \right)$. Notice that the FE estimator is not necessarily more efficient than the PC estimator under the null, which implies that

$$Var\left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC} \right) \neq Var\left(\hat{\boldsymbol{\beta}}_{FE} \right) - Var\left(\hat{\boldsymbol{\beta}}_{PC} \right)$$

---

[8] We investigate this issue by examining the the finite sample performance of the CD and the H_B tests for the heterogeneous panel data with the multiplicative IE. We allow the regressors to be uncorrelated with loadings under Experiment 1 while they are correlated under Experiment 2. In both experiments we maintain that the regressors are correlated with factors. As expected, the CD test results display that the null of weak residual CSD is strongly rejected for all the data generating process (DGP), correctly indicating the presence of IE under both Experiments. However, we find that the H_B test is consistent only under Experiment 2. Surprisingly, the H_B test does not display any power against the DGP under Experiment 1, where its rejection probability is close to, and sometimes lower than, the nominal size, especially in the presence of serially correlated errors. This is mainly because the FE estimator is still consistent under Experiment 1.

in contrast to the well-established finding in Hausman (1978). Hence, our proposed test is not exactly the Hausman test. We interpret the Hausman-type test in (17) as a test for the null hypothesis, (3) in heterogeneous panels with IE, (1).

Before developing the asymptotic theory for the Hausman-type statistic, we describe the asymptotic distribution of the bias-corrected PC estimator given by

$$\hat{\boldsymbol{\beta}}_{PC} = \tilde{\boldsymbol{\beta}}_{PC} - \frac{1}{N}\hat{\boldsymbol{B}}_{NT} - \frac{1}{T}\hat{\boldsymbol{C}}_{NT} \tag{18}$$

where the $\tilde{\boldsymbol{\beta}}_{PC}$ is the PC estimator obtained by iteratively solving the set of nonlinear equations:

$$\tilde{\boldsymbol{\beta}}_{PC} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{y}_i \text{ and}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_{PC}\right)\left(\boldsymbol{y}_i - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_{PC}\right)'\right]\hat{\boldsymbol{F}} = \hat{\boldsymbol{F}}V_{NT}$$

where $\boldsymbol{M}_{\hat{F}} = \boldsymbol{I}_T - \hat{\boldsymbol{F}}\left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\hat{\boldsymbol{F}}'$, V_{NT} is the diagonal matrix that consists of the r largest eigenvalues of the above matrix in the brackets arranged in a decreasing order, $\hat{\boldsymbol{F}}$ is $\sqrt{T}$ times the corresponding eigenvectors, and $\frac{1}{N}\hat{\boldsymbol{B}}_{NT}$ and $\frac{1}{T}\hat{\boldsymbol{C}}_{NT}$ are the bias correction terms derived in CHNY (see Appendix 9 for details).

Next, similar to the nonparametric and HAC variance estimators developed for the FE estimator, we propose two versions of the robust variance estimator for the PC estimator as follows[9]

$$\hat{\boldsymbol{V}}^{NON}\left(\hat{\boldsymbol{\beta}}_{PC}\right)$$
$$= \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N}\left(\boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)\left(\tilde{\boldsymbol{\beta}}_{PC,i} - \tilde{\boldsymbol{\beta}}_{PC}\right)\left(\tilde{\boldsymbol{\beta}}_{PC,i} - \tilde{\boldsymbol{\beta}}_{PC}\right)'\left(\boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)\right)$$
$$\left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)^{-1} \tag{19}$$

where $\tilde{\boldsymbol{\beta}}_{PC,i} = \left(\boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{y}_i$, and

$$\hat{\boldsymbol{V}}^{HAC}\left(\hat{\boldsymbol{\beta}}_{PC}\right) = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \hat{\boldsymbol{X}}_i' \hat{\boldsymbol{u}}_{PC,i} \hat{\boldsymbol{u}}_{PC,i}' \hat{\boldsymbol{X}}_i\right) \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)^{-1} \tag{20}$$

---

[9] The asymptotic validity of these estimators is verified in Theorem 2. Through the stochastic simulations (see Sect. 4), we find that the finite sample performance of both estimators is satisfactory and close to each other. Given the popularity of the HAC variance estimator, we propose to apply it in practice (see Sect. 5).

where $\hat{\boldsymbol{u}}_{PC,i} = \boldsymbol{y}_i - \hat{\boldsymbol{X}}_i \hat{\boldsymbol{\beta}}_{PC}$.

We provide the asymptotic distribution of the $\hat{\boldsymbol{\beta}}_{PC}$ estimator in Theorem 2.

Theorem 2 *Suppose that Assumption A holds. Considering,* $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$, *as* $N, T \to \infty$,

$$\sqrt{N} \left(\hat{\boldsymbol{\beta}}_{PC} - \boldsymbol{\beta} \right) \to_d N \left(0_{k \times 1}, \boldsymbol{\Psi}_{PC}^{-1} \boldsymbol{R}_{1,PC} \boldsymbol{\Psi}_{PC}^{-1} \right) \tag{21}$$

where $\boldsymbol{\Psi}_{PC} = \lim_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} E \left(\frac{\boldsymbol{V}_i' \boldsymbol{V}_i}{T} \right)$ *with* $\boldsymbol{V}_i = (\boldsymbol{v}_{i1}, \dots, \boldsymbol{v}_{iT})'$ *defined in* (33) *in Appendix 7, and*

$$\boldsymbol{R}_{1,PC} = \lim_{N,T \to \infty} N^{-1} \sum_{i=1}^{N} E \left(\frac{\boldsymbol{V}_i' \boldsymbol{V}_i}{T} \boldsymbol{\eta}_i \boldsymbol{\eta}_i' \frac{\boldsymbol{V}_i' \boldsymbol{V}_i}{T} \right) \tag{22}$$

Furthermore,

$$\hat{\boldsymbol{V}}^{NON} \left(\hat{\boldsymbol{\beta}}_{PC} \right)^{-1/2} \left(\hat{\boldsymbol{\beta}}_{PC} - \boldsymbol{\beta} \right) \to_d N \left(0, \boldsymbol{I}_k \right) \text{ and}$$

$$\hat{\boldsymbol{V}}^{HAC} \left(\hat{\boldsymbol{\beta}}_{PC} \right)^{-1/2} \left(\hat{\boldsymbol{\beta}}_{PC} - \boldsymbol{\beta} \right) \to_d N \left(0, \boldsymbol{I}_k \right). \tag{23}$$

It is worth noting in the homogeneous case with $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i that while the FE estimator is $\sqrt{N}$-consistent, the PC estimator can achieve a faster rate of convergence as it completely removes the effect of the unobserved factors, asymptotically. Further, the rate of convergence of the FE estimator is also shared by the CCE estimator, if the rank condition in Pesaran (2006) does not hold. Such condition cannot be ascertained but needs to be assumed, in which case the FE and CCE estimators have comparable theoretical properties. Nevertheless, the superiority of the PC estimator does not necessarily extend to its small sample properties as we examine in Monte Carlo study below.

Having established that the two versions of the robust estimator can consistently standardise the estimator, we propose to estimate $Cov \left(\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{PC} \right)$ by[10]

$$\hat{\boldsymbol{C}}^{NON} \left(\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{PC} \right)$$
$$= \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i \right)^{-1} \left(\sum_{i=1}^{N} \left(\ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i \right) \left(\hat{\boldsymbol{\beta}}_{FE,i} - \hat{\boldsymbol{\beta}}_{FE} \right) \left(\tilde{\boldsymbol{\beta}}_{PC,i} - \tilde{\boldsymbol{\beta}}_{PC} \right)' \left(\boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i \right) \right)$$
$$\left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i \right)^{-1}$$

$$\hat{\boldsymbol{C}}^{HAC} \left(\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{PC} \right) = \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i \right)^{-1} \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \hat{\boldsymbol{u}}_{FE,i} \hat{\boldsymbol{u}}_{PC,i}' \hat{\boldsymbol{X}}_i \right) \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i \right)^{-1}.$$

---

[10] Following Bai (2009), we have also employed the analytic (sandwich-form) variance estimator of $\boldsymbol{V}$, taking into account unknown form of heteroscedastic and autocorrelated errors. After conducting the preliminary simulations, we find that the two robust estimators perform more satisfactory.

Accordingly, we define two operating versions of the Hausman-type statistic by

$$H^{NON} = \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC}\right)' \left(\hat{\boldsymbol{V}}^{NON}\right)^{-1} \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC}\right) \tag{24}$$

$$H^{HAC} = \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC}\right)' \left(\hat{\boldsymbol{V}}^{HAC}\right)^{-1} \left(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{PC}\right) \tag{25}$$

where

$$\hat{\boldsymbol{V}}^{NON} = \hat{\boldsymbol{V}}^{NON}\left(\hat{\boldsymbol{\beta}}_{FE}\right) + \hat{\boldsymbol{V}}^{NON}\left(\hat{\boldsymbol{\beta}}_{PC}\right) - 2\hat{\boldsymbol{C}}^{NON}\left(\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{PC}\right) \tag{26}$$

$$\hat{\boldsymbol{V}}^{HAC} = \hat{\boldsymbol{V}}^{HAC}\left(\hat{\boldsymbol{\beta}}_{FE}\right) + \hat{\boldsymbol{V}}^{HAC}\left(\hat{\boldsymbol{\beta}}_{PC}\right) - 2\hat{\boldsymbol{C}}^{HAC}\left(\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{PC}\right) \tag{27}$$

We provide the main result in the following Theorem.

Theorem 3 *Under Assumption* A, *as* $N, T \to \infty$,

$$H^j \to_d \chi^2_k \text{ for } j = NON, HAC$$

H^j follows the χ^2_k distribution even though the rate of convergence of the PC estimator is $\sqrt{NT}$ while the FE estimator is $\sqrt{N}$-consistent. This follows from the use of the robust covariance estimators that properly normalise the test statistic as shown in Appendix 7.

Next, notice that our proposed test, (17), is fundamentally different from Bai's Hausman test, (16), because our null hypothesis, (3) is subsumed under his alternative model with IE, as is clearly demonstrated in (6). Furthermore, Bai's test will be consistent only if the regressors are correlated with both factors and loadings. Importantly, Bai's Hausman test will be inconsistent if $\boldsymbol{x}_{it}$ are uncorrelated with γ_i in (1), which is mainly because the FE estimator is still consistent under (3). This suggests that the non-rejection of the null by the Bai's test is not informative because it cannot distinguish between the panel data model with the 2-way additive fixed effects only and the model with IE where the regressors are uncorrelated with loadings. See the Online Supplement for the simulation evidence. In the empirical applications below we find that Bai's Hausman test rarely rejects the null of additive effects model against the alternative of IE even though the CD test strongly rejects the null of weak CSD for all the datasets. Such conflicting results may suggest that the regressors are indeed uncorrelated with factor loadings even in the panels with IE, which could provide the support for the usefulness of our proposed test.

4 Monte Carlo simulations

4.1 Review of previous studies

Westerlund and Urban (2013) find that the CCE estimator does not perform well in the presence of correlated factor loadings, especially if the full rank condition is

not satisfied. Karabiyik et al. (2017) discuss the role of the rank condition in the CCE estimation, and show that the second moment matrix of the estimated factors becomes asymptotically singular if the number of factors is strictly less than the number of dependent and independent variables, invalidating the key arguments commonly applied to establish the asymptotic theory. Westerlund and Urbain (2015) provide a formal comparison between the CCE and PC estimators by employing the same data generating process (DGP)[11] and show that the two estimators are asymptotically equivalent only if $N/T \to 0$ whereas their asymptotic distributions are no longer equivalent if $N/T \to \tau > 0$, especially in terms of asymptotic biases.

Though a number of papers have examined the small sample performance of the CCE and PC estimators, we find that only two studies by Sarafidis and Wansbeek (2012) and Westerlund (2019a), have explicitly analysed the performance of the FE estimator in the presence of CSD. Assuming the homogeneous parameters with $N = 100$ and $T = 50$, Sarafidis and Wansbeek (2012) compare the performance of the FE, CCE and PC estimators. If the factor loadings between the equations for y and x are uncorrelated and the rank condition is satisfied, they find that all three estimators perform well in terms of bias and RMSE. If the factor loadings are correlated, however, the FE estimator is severely biased. The CCE estimator is substantially biased if the rank condition is violated. As expected, the performance of the PC estimator is not significantly affected by the presence of correlated factor loadings.

Recently, Westerlund (2019a) shows that the FE estimator can be consistent even in the presence of IE, because both FE and CCE estimators belong to a class of estimators that satisfy a zero sum restriction. But, he maintains the assumption that factor loadings are uncorrelated in which case he demonstrates that the performance of the FE and CCE estimators is satisfactory.

4.2 Monte Carlo design

We generate the data as follows:

$$y_{it} = \beta_i x_{it} + \gamma_{1i} f_{1t} + \gamma_{2i} f_{2t} + \varepsilon_{it}, \tag{28}$$
$$x_{it} = \Gamma_{1i} f_{1t} + \Gamma_{2i} f_{2t} + u_{it}, \tag{29}$$

where $(f_{1t}, f_{2t}, \varepsilon_{it}, u_{it})'$ are drawn from the multivariate normal distribution with zero means and covariance matrix, $\Sigma_i = diag\left(\sigma_{f1}^2, \sigma_{f2}^2, \sigma_{\varepsilon_i}^2, \sigma_{u_i}^2\right) = I_4$. We follow Pesaran (2006) and Westerlund and Urbain (2013), and generate the factor loadings, $(\gamma_{1i}, \gamma_{2i})$ and $(\Gamma_{1i}, \Gamma_{2i})$ as follows:

- Experiment 1 with uncorrelated factor loadings and the full rank in which case $\gamma_{1i} \sim iidN(1, 1)$, $\gamma_{2i} \sim iidN(0, 1)$, $\Gamma_{1i} \sim iidN(0, 1)$, $\Gamma_{2i} \sim iidN(1, 1)$ such that $E\begin{pmatrix} \gamma_{1i} & \gamma_{2i} \\ \Gamma_{1i} & \Gamma_{2i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

---

[11] The DGP and the estimators are not identical to what have proposed by Pesaran (2006) and Bai (2009).

- Experiment 2 with uncorrelated factor loadings and the rank deficiency in which case $\gamma_{1i} \sim iidN(1,1)$, $\gamma_{2i} \sim iidN(0,1)$, $\Gamma_{1i} \sim iidN(1,1)$, $\Gamma_{2i} \sim iidN(0,1)$, such that $E\begin{pmatrix} \gamma_{1i} & \gamma_{2i} \\ \Gamma_{1i} & \Gamma_{2i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$.
- Experiment 3 with correlated factor loadings and the full rank in which case: $\gamma_{1i} = \gamma_1 + \upsilon_{1i}$, $\gamma_{2i} = \gamma_2 + \upsilon_{2i}$, $\Gamma_{1i} = \Gamma_1 + \upsilon_{1i}$, and $\Gamma_{2i} = \Gamma_2 + \upsilon_{2i}$ with $\gamma_1 = 1$, $\gamma_2 = 0$, $\Gamma_1 = 2$, $\Gamma_2 = 1$ and $(\upsilon_{1i}, \upsilon_{2i}) \sim iidN(0, I_2)$, such that $E\begin{pmatrix} \gamma_{1i} & \gamma_{12} \\ \Gamma_{1i} & \Gamma_{12} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$
- Experiment 4 with correlated factor loadings and the rank deficiency in which case $\gamma_{1i} \sim iidN(1,1)$, $\gamma_{2i} \sim iidN(0,1)$, $\gamma_{1i} = \Gamma_{1i}$ and $\gamma_{2i} = \Gamma_{2i}$ such that $E\begin{pmatrix} \gamma_{1i} & \gamma_{2i} \\ \Gamma_{1i} & \Gamma_{2i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$.

We specify the main slope parameter as $\beta_i = 1 + \eta_i$, $\eta_i \sim iidN(0, 0.04)$ and consider the following combination of $(N, T) = 20, 30, 50, 100, 200$, setting the number of replications at $R = 1,000$.[12]

4.3 The small sample performance of FE, CCE and PC estimators

We examine the finite sample performance of the following estimators: the two-way fixed effect (FE) estimator, $\hat{\beta}_{FE}$, the CCE estimator by Pesaran (2006), $\hat{\beta}_{CCE}$, and the bias corrected PC estimators proposed by CHNY, $\hat{\beta}_{PC}$. We consider both pooled and mean group estimator except for $\hat{\beta}_{PC}$ (see Appendix 7 for details). Notice that consistency of the PC estimator depends crucially upon correctly selecting the number of unobserved factors (Moon and Weidner 2015). In this regard, to address uncertainty associated with the selection criteria, we initially consider the two information criteria, denoted IC_{p1} and AIC_1, proposed by Bai and Ng (2002). Overall, we find that the PC estimator using IC_{p1} outperforms that with AIC_1, and we only report the results based on IC_{p1}.

We report the following summary statistics:

- Bias: $\hat{\beta}_R - \beta_0$, where β_0 is a true parameter value and $\hat{\beta}_R = R^{-1} \sum_{r=1}^{R} \hat{\beta}_r$ is the mean coefficient across R replications.

- RMSE: the root mean square error estimated by $\sqrt{R^{-1} \sum_{r=1}^{R} \left(\hat{\beta}_r - \beta_0 \right)^2}$.

Table 1 shows the simulation results for Experiment 1 with the full rank and uncorrelated factor loadings. The biases of all estimators are mostly negligible even in small samples with the FE performing slightly worse than other estimators when $N = 20$. The results for RMSEs display qualitatively similar patterns. RMSEs of CCE and PC estimators are lower than those of the FE and decline as N or T grows. On the other hand, the RMSE of the FE estimator improves only with N. Finally, biases and

---

[12] We have considered the cases with homogeneous β's and obtained qualitatively similar results, which are reported in the Online Supplement. We have also explored the experiments under the stronger parameter heterogeneity by generating $\eta_{ik} \sim iidN(0, 0.25)$ as wells as $\eta_{ik} \sim iidN(0, 1)$, and still obtained qualitatively similar results (unreported to save space). Finally, we obtained qualitatively similar results for serially correlated factors. These results are available upon request.

RMSEs of the pooled and mean group estimators display almost identical patterns. The relative performance of FE, CCE and PC estimators is generally in line with the simulation results reported in Chudik et al. (2011), Sarafidis and Wansbeek (2012) and CHNY.

The important exception is the poor performance of the PC estimator using AIC_1.[13] In this case the biases are substantial in small samples. They decline only if both N and T become large. Further, their RMSEs are much larger than those of the other estimators and decrease only if N and T are large. This demonstrates the influence of the estimated number of factors for the PC estimator. Given that information criteria have very variable performance, this is a problematic issue for PC estimators in which case the FE estimator can make an operational alternative.

Table 2 presents simulation results for Experiment 2 where factor loadings are uncorrelated but the rank condition is violated. The performance of the CCE estimators tends to slightly deteriorate, both bias and RMSE of the CCE estimator are higher than in the case with the full rank. The performance of the CCE estimator improves slowly with N only, suggesting that the rank deficiency may slow down its performance. On the other hand, the bias and the RMSE of the PC and FE estimators do not appear to be affected by the rank deficiency. Finally, we find that the mean group estimator performs slightly better than the pooled estimator in small samples.

Table 3 shows the results for Experiment 3 with correlated loadings and full rank. Now, only the FE estimator is severely biased. Next, the biases of the CCE estimator are not negligible for small N, but its performance improves sharply with N, a consistent finding with Westerlund and Urbain (2013), who note that 'the problem with correlated loadings goes away if the rank condition is satisfied'. The overall performance of the PC estimator is qualitatively similar to the previous cases, confirming that it is still consistent with both N and T.

Table 4 presents the simulation results for Experiment 4 with correlated loadings and the rank deficiency. Both CCE and FE estimators are severely biased, confirming our theoretical prediction that both estimators are inconsistent in the presence of correlated factors loadings as also discussed in Sarafidis and Wansbeek (2012) and Westerlund and Urbain (2013). On the other hand, the performance of the PC estimators is qualitatively similar to those presented in Table 2.

Overall, our results show that, when the factor loadings are uncorrelated, all the estimators show a similar and satisfactory performance, suggesting that the FE estimator can produce reliable results even in the presence of IE. When factor loadings are correlated, however, the FE estimator becomes severely biased and the performance of the CCE estimator tends to worsen. Only under the full rank condition, the performance of the CCE improves with N. The performance of the bias-corrected PC estimator is qualitatively similar across all four experiments.

4.4 The performance of the Hausman-type test statistic

We examine the small sample performance of the H test statistics, under the above four experiments, considering the following combination of $(N, T) =$

---

[13] For a complete comparison we report the simulation results based on AIC_1 in the Online Supplement.

Table 1 Simulation results for Experiment 1 with uncorrelated loadings and the full rank

T/N	β_{CCEP}					β_{CCEMG}				
	20	30	50	100	200	20	30	50	100	200
Bias										
20	− 0.0038	− 0.0013	0.0015	− 0.0001	− 0.0003	− 0.0041	− 0.0026	0.0023	0.0000	− 0.0002
30	− 0.0037	0.0019	0.0021	− 0.0003	− 0.0007	− 0.0027	0.0017	0.0017	− 0.0008	− 0.0008
50	0.0010	0.0008	0.0001	0.0014	0.0007	0.0006	0.0016	0.0002	0.0015	0.0006
100	0.0019	0.0002	− 0.0008	− 0.0004	0.0005	0.0015	0.0003	− 0.0007	− 0.0004	0.0005
200	− 0.0011	0.0008	− 0.0007	− 0.0003	0.0003	− 0.0006	0.0007	− 0.0006	− 0.0005	0.0003
RMSE										
20	0.0730	0.0609	0.0444	0.0319	0.0229	0.0741	0.0613	0.0454	0.0322	0.0234
30	0.0670	0.0546	0.0407	0.0281	0.0195	0.0660	0.0546	0.0410	0.0282	0.0194
50	0.0598	0.0461	0.0347	0.0255	0.0181	0.0582	0.0458	0.0343	0.0254	0.0179
100	0.0538	0.0416	0.0311	0.0224	0.0163	0.0525	0.0407	0.0312	0.0222	0.0162
200	0.0545	0.0401	0.0294	0.0213	0.0151	0.0529	0.0397	0.0293	0.0212	0.0150
	β_{FEP}					β_{FEMG}				
Bias										
20	0.0008	0.0042	0.0006	0.0012	− 0.0004	− 0.0012	0.0002	− 0.0004	0.0014	− 0.0016
30	− 0.0073	− 0.0022	0.0010	0.0009	0.0004	− 0.0017	− 0.0023	0.0034	− 0.0007	0.0004
50	− 0.0011	− 0.0001	− 0.0017	0.0006	0.0029	0.0022	0.0053	− 0.0002	0.0003	0.0031
100	0.0052	− 0.0001	− 0.0021	− 0.0004	0.0015	0.0073	0.0010	− 0.0017	− 0.0007	0.0001
200	− 0.0037	0.0010	0.0011	0.0029	− 0.0003	− 0.0014	0.0009	− 0.0001	0.0022	0.0005

Table 1 continued

	β_{FEP}				β_{FEMG}			
RMSE								
20	0.1285	0.0849	0.0578	0.0418	0.1230	0.1035	0.0808	0.0558
30	0.1253	0.0816	0.0579	0.0405	0.1198	0.0992	0.0760	0.0538
50	0.1251	0.0798	0.0559	0.0390	0.1152	0.0938	0.0749	0.0530
100	0.1230	0.0790	0.0544	0.0380	0.1114	0.0887	0.0715	0.0489
200	0.1255	0.0778	0.0533	0.0378	0.1135	0.0913	0.0717	0.0485

	β_{PC}			
Bias				
20	0.0029	−0.0002	−0.0008	0.0000
30	0.0045	−0.0023	−0.0004	0.0004
50	−0.0012	0.0002	0.0017	0.0000
100	−0.0019	−0.0005	−0.0008	−0.0002
200	−0.0006	0.0019	−0.0006	−0.0004
RMSE				
20	0.0773	0.0476	0.0327	0.0229
30	0.0685	0.0405	0.0277	0.0193
50	0.0604	0.0346	0.0247	0.0174
100	0.0515	0.0320	0.0229	0.0162
200	0.0491	0.0306	0.0210	0.0146

CCEP and CCEMG are the pooled and mean group common correlated estimators by Pesaran (2006); FEP and FEMG denote the pooled and mean group two-way fixed effects estimators. PC is the iterative principal component estimator in CHNY. The PC estimator is bias-corrected and evaluated using the IC_{p1} criterion by Bai and Ng (2002)

Table 2 Simulation results for Experiment 2 with uncorrelated loadings and the rank deficiency

T/N	β_{CCEP}					β_{CCEMG}				
	20	30	50	100	200	20	30	50	100	200
Bias										
20	− 0.0025	− 0.0035	0.0001	− 0.0015	0.0013	0.0011	− 0.0054	− 0.0008	− 0.0012	0.0020
30	0.0052	0.0011	0.0015	− 0.0015	0.0018	0.0018	0.0010	0.0013	− 0.0016	0.0013
50	0.0010	0.0015	− 0.0010	0.0001	0.0025	− 0.0013	0.0014	0.0001	0.0001	0.0023
100	− 0.0024	0.0010	− 0.0029	− 0.0001	0.0016	− 0.0018	− 0.0001	− 0.0019	0.0005	0.0010
200	0.0010	− 0.0054	0.0005	0.0005	− 0.0011	− 0.0010	− 0.0048	0.0001	0.0010	− 0.0006
RMSE										
20	0.1168	0.0906	0.0703	0.0495	0.0351	0.1033	0.0819	0.0648	0.0443	0.0322
30	0.1107	0.0956	0.0691	0.0492	0.0347	0.0947	0.0802	0.0591	0.0428	0.0299
50	0.1071	0.0830	0.0668	0.0481	0.0339	0.0900	0.0717	0.0555	0.0402	0.0276
100	0.1022	0.0833	0.0657	0.0474	0.0320	0.0843	0.0700	0.0535	0.0384	0.0271
200	0.0980	0.0840	0.0665	0.0452	0.0333	0.0833	0.0694	0.0538	0.0367	0.0276

T/N	β_{FEP}					β_{FEMG}				
	20	30	50	100	200	20	30	50	100	200
Bias										
20	− 0.0009	− 0.0052	− 0.0012	− 0.0006	0.0006	− 0.0001	− 0.0032	− 0.0008	0.0004	0.0005
30	0.0052	− 0.0019	0.0017	0.0005	0.0006	0.0027	− 0.0003	0.0022	0.0006	0.0013
50	0.0037	0.0056	− 0.0018	0.0020	0.0014	0.0034	0.0032	− 0.0036	0.0011	0.0009
100	− 0.0021	0.0004	0.0012	− 0.0010	0.0011	− 0.0005	0.0017	0.0011	0.0009	0.0003
200	− 0.0018	0.0011	0.0031	0.0002	− 0.0007	− 0.0033	− 0.0009	0.0025	0.0003	0.0000

Table 2 continued

	β_{FEP}					β_{FEMG}				
RMSE										
20	0.1370	0.1054	0.0801	0.0593	0.0403	0.1301	0.0974	0.0767	0.0550	0.0398
30	0.1339	0.1106	0.0791	0.0573	0.0396	0.1214	0.1031	0.0735	0.0551	0.0366
50	0.1207	0.0994	0.0778	0.0571	0.0398	0.1140	0.0927	0.0716	0.0521	0.0377
100	0.1191	0.0985	0.0770	0.0537	0.0377	0.1082	0.0901	0.0687	0.0502	0.0346
200	0.1175	0.1015	0.0785	0.0541	0.0391	0.1111	0.0932	0.0716	0.0502	0.0362

	β_{PC}				
Bias					
20	− 0.0026	− 0.0003	− 0.0035	− 0.0017	0.0004
30	− 0.0009	− 0.0005	− 0.0029	0.0009	− 0.0009
50	− 0.0031	− 0.0041	0.0008	− 0.0008	0.0001
100	− 0.0016	− 0.0017	− 0.0014	0.0008	0.0000
200	− 0.0014	0.0003	− 0.0025	− 0.0009	− 0.0007
RMSE					
20	0.0814	0.0630	0.0460	0.0323	0.0228
30	0.0674	0.0541	0.0397	0.0292	0.0203
50	0.0599	0.0473	0.0370	0.0245	0.0182
100	0.0521	0.0425	0.0320	0.0231	0.0157
200	0.0484	0.0399	0.0309	0.0218	0.0149

See notes to Table 1

Table 3 Simulation results for Experiment 3 with correlated loadings and the full rank

T/N	β_{CCEP}					β_{CCEMG}				
	20	30	50	100	200	20	30	50	100	200
Bias										
20	0.0728	0.0484	0.0301	0.0158	0.0086	0.0678	0.0459	0.0282	0.0155	0.0089
30	0.0727	0.0468	0.0297	0.0130	0.0088	0.0675	0.0440	0.0283	0.0130	0.0083
50	0.0750	0.0502	0.0287	0.0150	0.0074	0.0683	0.0471	0.0274	0.0145	0.0073
100	0.0746	0.0474	0.0304	0.0152	0.0073	0.0682	0.0444	0.0291	0.0149	0.0071
200	0.0767	0.0513	0.0298	0.0144	0.0075	0.0700	0.0482	0.0287	0.0141	0.0074
RMSE										
20	0.1121	0.0808	0.0566	0.0346	0.0249	0.1060	0.0791	0.0570	0.0352	0.0250
30	0.1007	0.0738	0.0500	0.0330	0.0227	0.0955	0.0710	0.0493	0.0326	0.0225
50	0.1011	0.0711	0.0465	0.0291	0.0185	0.0940	0.0678	0.0450	0.0288	0.0184
100	0.0959	0.0644	0.0443	0.0273	0.0175	0.0886	0.0613	0.0432	0.0269	0.0175
200	0.0984	0.0675	0.0439	0.0261	0.0168	0.0901	0.0644	0.0428	0.0259	0.0167
	β_{FEP}					β_{FEMG}				
Bias										
20	0.6517	0.6528	0.6545	0.6564	0.6574	0.5413	0.5372	0.5364	0.5378	0.5389
30	0.6528	0.6562	0.6591	0.6606	0.6602	0.5371	0.5381	0.5383	0.5378	0.5383
50	0.6586	0.6568	0.6571	0.6629	0.6637	0.5384	0.5360	0.5344	0.5393	0.5388
100	0.6589	0.6596	0.6619	0.6625	0.6637	0.5403	0.5386	0.5393	0.5379	0.5385
200	0.6589	0.6623	0.6626	0.6624	0.6651	0.5396	0.5411	0.5387	0.5365	0.5382

Table 3 continued

	β_{FEP}					β_{FEMG}				
RMSE										
20	0.6595	0.6586	0.6588	0.6595	0.6602	0.5492	0.5435	0.5410	0.5408	0.5415
30	0.6587	0.6612	0.6623	0.6629	0.6621	0.5438	0.5434	0.5416	0.5401	0.5402
50	0.6643	0.6607	0.6599	0.6647	0.6650	0.5445	0.5403	0.5373	0.5411	0.5400
100	0.6639	0.6628	0.6642	0.6639	0.6644	0.5456	0.5422	0.5418	0.5393	0.5393
200	0.6635	0.6657	0.6648	0.6634	0.6657	0.5446	0.5446	0.5409	0.5377	0.5389

	β_{PC}				
Bias					
20	0.0012	−0.0028	−0.0014	−0.0018	−0.0010
30	−0.0056	−0.0048	0.0000	−0.0007	−0.0012
50	−0.0055	−0.0040	−0.0019	−0.0013	−0.0004
100	−0.0043	−0.0041	−0.0024	−0.0016	−0.0005
200	−0.0042	−0.0035	−0.0016	−0.0021	−0.0005
RMSE					
20	0.0759	0.0613	0.0453	0.0329	0.0234
30	0.0653	0.0518	0.0403	0.0286	0.0193
50	0.0572	0.0453	0.0362	0.0254	0.0180
100	0.0528	0.0421	0.0333	0.0233	0.0160
200	0.0500	0.0391	0.0312	0.0214	0.0150

See notes to Table 1

Table 4 Simulation results for Experiment 4 with correlated loadings and the rank deficiency

T/N	β_{CCEP}					β_{CCEMG}				
	20	30	50	100	200	20	30	50	100	200
Bias										
20	0.4731	0.4771	0.4785	0.4777	0.4758	0.3478	0.3478	0.3434	0.3405	0.3384
30	0.4750	0.4766	0.4839	0.4857	0.4879	0.3469	0.3412	0.3453	0.3447	0.3429
50	0.4843	0.4839	0.4880	0.4895	0.4867	0.3472	0.3450	0.3458	0.3447	0.3395
100	0.4882	0.4939	0.4882	0.4873	0.4923	0.3518	0.3475	0.3429	0.3396	0.3425
200	0.4842	0.4887	0.4909	0.4925	0.4947	0.3466	0.3454	0.3444	0.3426	0.3427
RMSE										
20	0.4926	0.4925	0.4902	0.4873	0.4856	0.3647	0.3607	0.3528	0.3470	0.3445
30	0.4915	0.4875	0.4927	0.4919	0.4934	0.3602	0.3499	0.3521	0.3491	0.3465
50	0.4960	0.4932	0.4949	0.4941	0.4906	0.3563	0.3526	0.3511	0.3482	0.3422
100	0.4990	0.5008	0.4933	0.4906	0.4947	0.3609	0.3535	0.3468	0.3421	0.3441
200	0.4946	0.4955	0.4950	0.4951	0.4963	0.3549	0.3512	0.3477	0.3447	0.3439

	β_{FEP}					β_{FEMG}				
	20	30	50	100	200	20	30	50	100	200
Bias										
20	0.6495	0.6548	0.6516	0.6571	0.6558	0.5387	0.5403	0.5353	0.5375	0.5357
30	0.6504	0.6564	0.6600	0.6605	0.6617	0.5358	0.5382	0.5412	0.5399	0.5383
50	0.6563	0.6556	0.6608	0.6643	0.6628	0.5359	0.5361	0.5398	0.5418	0.5376
100	0.6601	0.6632	0.6591	0.6618	0.6654	0.5416	0.5399	0.5365	0.5377	0.5400
200	0.6576	0.6603	0.6618	0.6653	0.6665	0.5383	0.5379	0.5388	0.5388	0.5389

Table 4 continued

	β_{FEP}					β_{FEMG}				
RMSE										
20	0.6583	0.6608	0.6561	0.6603	0.6584	0.5483	0.5470	0.5402	0.5407	0.5383
30	0.6579	0.6612	0.6635	0.6627	0.6634	0.5435	0.5434	0.5447	0.5422	0.5400
50	0.6618	0.6599	0.6636	0.6660	0.6640	0.5417	0.5407	0.5427	0.5436	0.5389
100	0.6651	0.6665	0.6615	0.6632	0.6663	0.5471	0.5436	0.5389	0.5392	0.5408
200	0.6629	0.6635	0.6639	0.6664	0.6672	0.5434	0.5414	0.5409	0.5400	0.5396

	β_{PC}									
Bias										
20	0.0013	−0.0012	−0.0011	−0.0005	0.0002					
30	−0.0028	0.0002	−0.0001	−0.0007	−0.0009					
50	−0.0031	−0.0026	−0.0008	−0.0005	−0.0003					
100	−0.0045	−0.0020	−0.0009	−0.0009	0.0004					
200	−0.0019	−0.0018	−0.0013	−0.0008	−0.0005					
RMSE										
20	0.0738	0.0597	0.0460	0.0323	0.0227					
30	0.0638	0.0530	0.0404	0.0277	0.0205					
50	0.0575	0.0473	0.0350	0.0254	0.0175					
100	0.0518	0.0403	0.0307	0.0232	0.0161					
200	0.0500	0.0384	0.0306	0.0209	0.0154					

See notes to Table 1

Table 5 Size and power of the H^{NON} statistic and coverage rates at 95% level for heterogeneous βs, $\beta_i = 1 + \eta_i$, $\eta_i \sim iidN(0, 0.04)$ and no serial correlation

T/N	Experiment 1					T/N	Experiment 3				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.061	0.053	0.050	0.058	0.057	50	1	1	1	1	1
100	0.054	0.063	0.047	0.054	0.049	100	1	1	1	1	1
150	0.079	0.067	0.062	0.064	0.045	150	1	1	1	1	1
200	0.065	0.062	0.058	0.057	0.047	200	1	1	1	1	1
500	0.058	0.053	0.066	0.051	0.057	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.942	0.948	0.942	0.940	0.946	50	0	0	0	0	0
100	0.951	0.938	0.950	0.943	0.955	100	0	0	0	0	0
150	0.939	0.941	0.950	0.938	0.954	150	0	0	0	0	0
200	0.936	0.946	0.949	0.945	0.945	200	0	0	0	0	0
500	0.944	0.946	0.935	0.944	0.947	500	0	0	0	0	0
	Coverage rates β_{PC}						*Coverage rates β_{PC}*				
50	0.918	0.923	0.940	0.948	0.935	50	0.924	0.930	0.938	0.946	0.943
100	0.913	0.937	0.945	0.949	0.952	100	0.834	0.936	0.943	0.946	0.952
150	0.926	0.933	0.945	0.936	0.948	150	0.809	0.777	0.938	0.944	0.947
200	0.899	0.930	0.934	0.942	0.948	200	0.884	0.856	0.815	0.947	0.942
500	0.928	0.928	0.936	0.944	0.947	500	0.896	0.913	0.901	0.907	0.952

T/N	Experiment 2					T/N	Experiment 4				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.061	0.054	0.047	0.046	0.046	50	1	1	1	1	1
100	0.063	0.057	0.056	0.066	0.043	100	1	1	1	1	1
150	0.069	0.069	0.045	0.050	0.047	150	1	1	1	1	1
200	0.064	0.062	0.061	0.060	0.057	200	1	1	1	1	1
500	0.073	0.049	0.074	0.051	0.042	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.931	0.951	0.950	0.953	0.952	50	0	0	0	0	0
100	0.928	0.943	0.937	0.939	0.941	100	0	0	0	0	0
150	0.940	0.928	0.954	0.952	0.948	150	0	0	0	0	0
200	0.927	0.948	0.945	0.945	0.944	200	0	0	0	0	0
500	0.930	0.948	0.936	0.954	0.956	500	0	0	0	0	0

Table 5 continued

T/N	Experiment 2					T/N	Experiment 4				
	50	100	150	200	500		50	100	150	200	500
	Coverage rates β_{PC}						*Coverage rates β_{PC}*				
50	0.906	0.927	0.942	0.955	0.955	50	0.919	0.946	0.939	0.946	0.944
100	0.915	0.938	0.937	0.932	0.941	100	0.865	0.935	0.946	0.945	0.950
150	0.909	0.926	0.933	0.954	0.939	150	0.889	0.875	0.935	0.942	0.949
200	0.921	0.934	0.936	0.946	0.945	200	0.910	0.906	0.911	0.942	0.935
500	0.931	0.932	0.935	0.935	0.947	500	0.910	0.924	0.932	0.935	0.951

FE denotes the two-way fixed effect estimators; PC is the iterative bias corrected principal component estimator in CHNY. The PC estimator is evaluated using the IC_{p1} criterion by Bai and Ng (2002). H^{NON} is the H-statistic defined in (24)

50, 100, 150, 200, 500. To construct the H statistic, we consider the difference between the FE estimator, $\boldsymbol{\beta}_{FE}$ and the bias corrected PC estimator, $\boldsymbol{\beta}_{PC}$ standardised respectively by both versions of robust variance estimator, denoted NON and HAC.[14] We examine size and power of the H statistic, but we also report the coverage rates for the three estimators. We consider slope heterogeneity such as $\beta_i = \beta + \eta_i$, $\eta_i \sim N(0, 0.04)$ and serially correlated errors given by

$$\varepsilon_{it} = \rho_\varepsilon \varepsilon_{i,t-1} + v_{\varepsilon_{it}} \text{ and } u_{it} = \rho_u u_{i,t-1} + v_{uit} \text{ with } \rho_\varepsilon = \rho_u = 0 \text{ or } 0.5,$$

where $(v_{\varepsilon_{it}}, v_{uit})'$ are drawn from the bivariate normal distribution with zero means and covariance matrix, $diag(\sigma^2_{v_{\varepsilon i}}, \sigma^2_{v_{ui}}) = \boldsymbol{I}_2$. Hence, we examine the following two cases:

Case 1: Heterogeneous βs and no serial correlation; see Tables 5 and 6.

Case 2: Heterogeneous βs and serial correlation; see Tables 7 and 8.

Overall, the test performance of the H statistics reported in Tables 5, 6, 7 and 8, is satisfactory and qualitatively similar in terms of the empirical size and power. This confirms that all the estimators are consistent under the null with and without serial correlation. Furthermore, the satisfactory coverage rates revealed by the three estimators demonstrate that both nonparametric and HAC variance estimators are also robust to serial correlation.

In Experiments 1 and 2, the sizes of both H^{NON} and H^{HAC} tests approach the nominal level (0.05) in most cases as the sample size rises. The power of the H test is always one under Experiments 3 and 4. In particular, when the regressors are uncorrelated with factor loadings, $\boldsymbol{\beta}_{FE}$ is shown to be consistent and its coverage rate reaches the nominal 95% in Experiments 1 and 2, irrespective of the rank condition. In Experiments 3 and 4 when loadings are correlated with the regressor, however, $\boldsymbol{\beta}_{FE}$

---

[14] In what follows, we apply the bias corrected PC estimators using IC_{p1}. We have also investigated the performance of the H statistics using the uncorrected PC estimators, and obtained qualitatively similar results.

Table 6 Size and power of the H^{HAC} statistic and coverage rates at 95% level for heterogeneous βs, $\beta_i = 1 + \eta_i$, $\eta_i \sim iidN(0, 0.04)$ and no serial correlation

T/N	Experiment 1					T/N	Experiment 3				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.076	0.057	0.052	0.064	0.044	50	1	1	1	1	1
100	0.070	0.068	0.044	0.049	0.040	100	1	1	1	1	1
150	0.069	0.054	0.046	0.054	0.053	150	1	1	1	1	1
200	0.076	0.071	0.053	0.055	0.044	200	1	1	1	1	1
500	0.064	0.078	0.061	0.067	0.053	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.931	0.933	0.941	0.928	0.949	50	0	0	0	0	0
100	0.926	0.936	0.943	0.955	0.954	100	0	0	0	0	0
150	0.932	0.940	0.952	0.940	0.949	150	0	0	0	0	0
200	0.931	0.951	0.940	0.947	0.957	200	0	0	0	0	0
500	0.929	0.939	0.943	0.940	0.951	500	0	0	0	0	0
	Coverage rates β_{PC}						*Coverage rates β_{PC}*				
50	0.907	0.931	0.944	0.946	0.949	50	0.922	0.932	0.940	0.940	0.960
100	0.927	0.928	0.928	0.950	0.950	100	0.616	0.928	0.950	0.944	0.952
150	0.916	0.935	0.957	0.940	0.945	150	0.828	0.746	0.960	0.944	0.960
200	0.918	0.931	0.939	0.954	0.936	200	0.850	0.880	0.828	0.938	0.948
500	0.923	0.925	0.932	0.945	0.951	500	0.922	0.918	0.914	0.890	0.964

T/N	Experiment 2					T/N	Experiment 4				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.074	0.074	0.066	0.066	0.050	50	1	1	1	1	1
100	0.064	0.052	0.042	0.052	0.040	100	1	1	1	1	1
150	0.062	0.042	0.058	0.058	0.040	150	1	1	1	1	1
200	0.086	0.066	0.072	0.046	0.048	200	1	1	1	1	1
500	0.062	0.074	0.078	0.052	0.046	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.944	0.934	0.934	0.936	0.940	50	0	0	0	0	0
100	0.944	0.942	0.942	0.942	0.952	100	0	0	0	0	0
150	0.936	0.952	0.942	0.940	0.948	150	0	0	0	0	0
200	0.916	0.940	0.942	0.948	0.948	200	0	0	0	0	0
500	0.944	0.930	0.920	0.930	0.952	500	0	0	0	0	0

Table 6 continued

T/N	Experiment 2					T/N	Experiment 4				
	50	100	150	200	500		50	100	150	200	500
	Coverage rates β_{PC}						*Coverage rates β_{PC}*				
50	0.916	0.956	0.952	0.936	0.944	50	0.932	0.936	0.957	0.941	0.948
100	0.928	0.946	0.944	0.934	0.966	100	0.880	0.944	0.943	0.942	0.954
150	0.900	0.932	0.946	0.952	0.948	150	0.897	0.901	0.944	0.943	0.940
200	0.908	0.936	0.940	0.928	0.936	200	0.877	0.901	0.926	0.948	0.956
500	0.914	0.948	0.940	0.952	0.926	500	0.926	0.935	0.939	0.931	0.946

See notes to Table 5. H^{HAC} is the H-statistic defined in (25)

is significantly biased and displays a zero coverage rate. The coverage rates of the bias-corrected PC estimator tend to 95% under all four experiments.[15]

We have also considered the cases with homogeneous β's and obtained qualitatively similar results, which are reported in the Online Supplement.

4.5 The pretest estimator

The estimated number of factors can influence the performance of the PC estimator considerably, and this issue needs to be handled carefully. The previous literature has not provided clear evidence on what is the best course of action to choose the number of factors. In this regard, we propose a pretest estimator which is constructed as follows. The pretest estimator, denoted $\hat{\beta}_{pretest}$, selects either the FE or the PC estimator depending on the Hausman-type test results. To be more specific, we first evaluate the H^{NON} and H^{HAC} statistics. If the null hypothesis, (3) is not rejected, then we select $\hat{\beta}_{pretest} = \hat{\beta}_{FE}$ while, if the null is rejected, we set $\hat{\beta}_{pretest} = \hat{\beta}_{PC}$.

In the Online Supplement we have examined the finite sample performance of this pretest estimator under the same four experiments considered above. Its overall performance is satisfactory in terms of bias and RMSE, irrespective of whether factor loadings are correlated or not. This suggests that such an estimator has considerable potential as it alleviates the issue of selecting the number of factors, especially in the case where the regressors are found to be uncorrelated with factor loadings in practice.

5 Empirical applications

We investigate the empirical relevance of the null hypothesis of no correlation between the regressors and factor loadings by applying our proposed statistics H^{HAC} defined

---

[15] We have also examined the coverage rates for estimators using the analytic variance estimator derived in Bai (2009, Section 9). We find that, when errors are serially correlated and conditionally heteroscedastic and/or βs are heterogeneous, coverage rates of the PC estimators are mostly well below the nominal level. Similar findings are reported in Chudik et al. (2011) and Sarafidis and Wansbeek (2012). This demonstrates an importance of using the robust variance estimators for a reliable inference.

in (25) to fourteen datasets.[16] The details of the data and the empirical specifications are provided in Appendix 8.

The Cobb–Douglas production function The first application comprises five different cases—the OECD members ($N = 26$, $T = 41$, Mastromarco et al. 2016), the 20 Italian regions ($N = 20$, $T = 21$), the 48 U.S. States ($N = 48$, $T = 17$) and the aggregate sectorial data for manufacturing from developed and developing countries ($N = 25$, $T = 25$). Following the economic growth literature, we estimate the Cobb–Douglas production function by the FE and PC estimators and then apply our proposed Hausman-type test. For OECD, the output is measured by the per capita GDP while the regressor is the capital-labour ratio. For the Italian regions, output is the per capita value added while for the U.S. application, the output is the per capita gross State product, with the same regressor. In the fourth application, the output is measured as the aggregated manufacturing sector value-added of OECD countries, see Eberhardt and Teal (2019). In the fifth application, the production function is augmented by the R&D stock expenditure, and the output is the aggregate sectorial value added for manufacturing, see Eberhardt et al. (2013).

The gravity model of bilateral trade flows Next, we consider the estimation of a gravity model of the bilateral trade flows for the EU14 countries, counting $N = 91$ pairs from 1960 to 2008 ($T = 49$). Here, we follow Serlenga and Shin (2007) and estimate the gravity panel data regression, in which the bilateral trade flow is set as a function of GDP, countries' similarity, relative factor endowment, the real exchange rate as well as the trade union and common currency dummies.

The gasoline demand function This application aims at estimating the price and income elasticity of gasoline demand. In particular, we focus on estimating the demand function for gasoline using the data from Liu (2014), which contains quarterly data for the 50 States in the U.S. over the period 1994–2008 ($N = 50$, $T = 60$).

Housing prices We estimate the income elasticity of real housing prices from 1975 to 2010. We consider two datasets; the first data from Holly et al. (2010) covers the 49 U.S. States ($N = 49$, $T = 36$) while the second covers the 384 Metropolitan Statistical Areas ($N = 384$, $T = 36$) obtained from Baltagi and Li (2014).

Technological spillovers on productivity We consider two applications. First, we estimate the effects of domestic and foreign R&D on TFP controlling for the human capital. We use a balanced panel of 24 OECD countries over the period 1971–2004 ($N = 24$ and $T = 34$), see Coe et al. (2009) and Ertur and Musolesi (2017). In the second application we explore the channels through which technological investments affect the productivity performance of industrialised economies by estimating the productivity effects of R&D and Information and Communication Technologies (ICT), controlling for the inputs accumulation as labour and (non-ICT) capital for OECD industries. We use a balanced panel of 49 high-tech industries over the period 1977–2006 ($N = 53$ and $T = 30$) from Pieri et al. (2018).

Health care expenditure and income We estimate the relationship between healthcare expenditure and income after controlling for public expenditure over total health expenditure. We consider a panel of 167 countries covering the period 1995–2012 ($N = 167$ and $T = 18$), see Baltagi et al. (2017).

---

[16] The evidence provided by H_{NON} is qualitatively similar, as shown in Appendix 10.

Table 7 Size and power of the H^{NON} statistic and coverage rates at 95% level for heterogeneous βs, $\beta_i = 1 + \eta_i$, $\eta_i \sim iidN(0, 0.04)$ and serial correlation, $\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it} + v_{\varepsilon it}$, $u_{it} = \rho_u u_{it} + v_{uit}$, $\rho_\varepsilon = \rho_u = 0.5$

T/N	Experiment 1					T/N	Experiment 3				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.068	0.057	0.054	0.043	0.054	50	1	1	1	1	1
100	0.061	0.049	0.065	0.062	0.046	100	1	1	1	1	1
150	0.073	0.057	0.052	0.049	0.042	150	1	1	1	1	1
200	0.059	0.055	0.066	0.048	0.051	200	1	1	1	1	1
500	0.060	0.062	0.054	0.061	0.057	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.944	0.948	0.945	0.951	0.948	50	0	0	0	0	0
100	0.942	0.952	0.938	0.949	0.955	100	0	0	0	0	0
150	0.932	0.943	0.945	0.957	0.949	150	0	0	0	0	0
200	0.941	0.941	0.931	0.947	0.952	200	0	0	0	0	0
500	0.926	0.937	0.952	0.934	0.939	500	0	0	0	0	0
	Coverage rates β_{PC}						*Coverage rates β_{PC}*				
50	0.926	0.914	0.942	0.941	0.941	50	0.933	0.933	0.946	0.948	0.936
100	0.922	0.948	0.951	0.943	0.937	100	0.925	0.948	0.936	0.934	0.944
150	0.919	0.924	0.936	0.945	0.941	150	0.924	0.943	0.937	0.947	0.957
200	0.919	0.929	0.944	0.945	0.935	200	0.927	0.937	0.934	0.942	0.954
500	0.916	0.945	0.938	0.914	0.952	500	0.915	0.926	0.949	0.958	0.943

T/N	Experiment 2					T/N	Experiment 4				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.064	0.054	0.062	0.057	0.055	50	1	1	1	1	1
100	0.056	0.068	0.055	0.05	0.044	100	1	1	1	1	1
150	0.061	0.053	0.051	0.052	0.051	150	1	1	1	1	1
200	0.067	0.05	0.064	0.048	0.054	200	1	1	1	1	1
500	0.072	0.062	0.065	0.062	0.053	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.938	0.933	0.942	0.946	0.953	50	0	0	0	0	0
100	0.943	0.943	0.933	0.937	0.955	100	0	0	0	0	0
150	0.943	0.945	0.956	0.949	0.946	150	0	0	0	0	0
200	0.941	0.951	0.935	0.957	0.943	200	0	0	0	0	0
500	0.924	0.944	0.949	0.938	0.950	500	0	0	0	0	0

Table 7 continued

T/N	Experiment 2					T/N	Experiment 4				
	50	100	150	200	500		50	100	150	200	500
	Coverage rates β_{PC}						*Coverage rates* β_{PC}				
50	0.915	0.931	0.945	0.950	0.938	50	0.934	0.940	0.923	0.936	0.952
100	0.918	0.938	0.935	0.933	0.960	100	0.904	0.936	0.946	0.933	0.955
150	0.916	0.945	0.943	0.930	0.935	150	0.920	0.946	0.945	0.959	0.953
200	0.903	0.931	0.936	0.945	0.957	200	0.920	0.943	0.939	0.948	0.953
500	0.927	0.950	0.954	0.944	0.952	500	0.913	0.944	0.940	0.937	0.951

See notes to Table 5

Demographic and business cycle volatility. We estimate the impact of the age composition of the labor force on business cycle volatility. We employ a balanced panel dataset for 51 countries over the period 1957–2000 ($N = 51$ and $T = 44$) provided by Everaert and Vierke (2016).

Carbon emissions and trade We explore the nexus between carbon emissions and trade using a balanced panel of 32 OECD countries over the period 1990–2013 ($N = 32$ and $T = 24$), see Liddle (2018).

In Table 9, we present the estimation and test results. First of all, the test results by H_{HAC} provide a surprisingly convincing evidence that the null hypothesis of the regressors being uncorrelated with factor loadings, is not rejected (even at 1% significance level) in thirteen out of fourteen datasets considered.[17] We also report the results for the CD test proposed by Pesaran (2015), which tests the null of no (weak) CSD against the alternative of strong CSD, and the Hausman test proposed by Bai (2009), H_B in (16) and the Hausman test proposed by Westerlund (2019b), H_W, which test the null of additive-effects against the alternative of IE. The CD test strongly rejects the null hypothesis for all the datasets whilst both H_B rejects only once the null hypothesis of additive-effects model, at 10% significance level, and the H_W test reject three times. These test results are rather in conflict, since the former suggests the presence of CSD while the latter suggests no IE. As highlighted in Sect. 2, however, the rejection of CD test does not always imply that the FE estimator is biased in panels with IE. Further, in Sect. 3, we show that the H_B test has no power against the alternative model with IE, especially if the regressors are uncorrelated with factor loadings. Indeed, such conflicting results can provide support for our main test results that the regressors are indeed uncorrelated with factor loadings in the panels with IE.

Next, we turn to the slope estimates provided by both FE and PC estimators, and find that they are mostly significant. Their magnitudes and signs are relatively similar to each other, and consistent with theoretical predictions. There is only an exception reported in the gravity model of international trade.[18]

---

[17] The results of H_{NON} are shown in Appendix 10 and Table 10.

[18] Notice that the FE estimation tends to produce substantially large coefficient on GDP, which has been widely reported in the literature.

Table 8 Size and power of the H^{HAC} statistic and coverage rates at 95% for heterogeneous βs, $\beta_i = 1 + \eta_i$, $\eta_i \sim iidN(0, 0.04)$ and serial correlation, $\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it} + v_{\varepsilon it}, u_{it} = \rho_u u_{it} + v_{uit}, \rho_\varepsilon = \rho_u = 0.5$

T/N	Experiment 1					T/N	Experiment 3				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.068	0.074	0.045	0.054	0.053	50	1	1	1	1	1
100	0.068	0.053	0.054	0.052	0.045	100	1	1	1	1	1
150	0.070	0.064	0.056	0.053	0.047	150	1	1	1	1	1
200	0.053	0.060	0.058	0.059	0.043	200	1	1	1	1	1
500	0.068	0.055	0.062	0.059	0.053	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.937	0.941	0.940	0.947	0.950	50	0	0	0	0	0
100	0.929	0.938	0.947	0.943	0.950	100	0	0	0	0	0
150	0.924	0.933	0.940	0.947	0.954	150	0	0	0	0	0
200	0.943	0.943	0.943	0.928	0.956	200	0	0	0	0	0
500	0.924	0.944	0.952	0.931	0.946	500	0	0	0	0	0
	Coverage rates β_{PC}						*Coverage rates β_{PC}*				
50	0.923	0.941	0.951	0.938	0.939	50	0.900	0.932	0.947	0.942	0.959
100	0.909	0.937	0.939	0.944	0.960	100	0.928	0.939	0.931	0.954	0.941
150	0.916	0.911	0.936	0.937	0.953	150	0.916	0.925	0.945	0.938	0.945
200	0.911	0.931	0.938	0.950	0.959	200	0.930	0.939	0.946	0.949	0.938
500	0.906	0.937	0.937	0.941	0.932	500	0.926	0.952	0.926	0.938	0.937
T/N	Experiment 2					T/N	Experiment 4				
	50	100	150	200	500		50	100	150	200	500
	Size						*Power*				
50	0.078	0.054	0.057	0.051	0.050	50	1	1	1	1	1
100	0.075	0.051	0.052	0.059	0.063	100	1	1	1	1	1
150	0.064	0.064	0.042	0.054	0.053	150	1	1	1	1	1
200	0.091	0.044	0.060	0.051	0.036	200	1	1	1	1	1
500	0.086	0.055	0.066	0.040	0.049	500	1	1	1	1	1
	Coverage rates β_{FE}						*Coverage rates β_{FE}*				
50	0.929	0.948	0.937	0.936	0.952	50	0	0	0	0	0
100	0.918	0.942	0.951	0.942	0.925	100	0	0	0	0	0
150	0.943	0.942	0.959	0.937	0.952	150	0	0	0	0	0
200	0.914	0.954	0.942	0.946	0.957	200	0	0	0	0	0
500	0.925	0.941	0.938	0.954	0.952	500	0	0	0	0	0
	Coverage rates β_{PC}						*Coverage rates β_{PC}*				
50	0.918	0.942	0.937	0.928	0.934	50	0.929	0.917	0.935	0.942	0.940
100	0.910	0.936	0.925	0.936	0.941	100	0.937	0.929	0.938	0.952	0.947
150	0.920	0.930	0.938	0.942	0.947	150	0.926	0.940	0.943	0.945	0.936
200	0.907	0.929	0.949	0.945	0.940	200	0.918	0.925	0.944	0.946	0.945
500	0.922	0.932	0.941	0.944	0.952	500	0.930	0.940	0.945	0.944	0.948

See notes to Table 6

Combining all the above test and estimation results, we come to a conclusion that the regressors are uncorrelated with factor loadings in many cross-sectionally correlated panels with IE in practice. In this situation, the FE estimation can produce consistent estimator. We emphasise that the FE estimator is invariant to any complex issues related to selecting the number of unobserved factors incorrectly which would significantly affect the performance of PC estimators (Moon and Weidner 2015), and to employing the inconsistent initial estimates which may not guarantee the convergence of the iterative PC estimator (Hsiao 2018). This suggests that the FE estimator can still be of considerable applicability in a wide variety of cross-sectionally correlated panel data with IE, especially if the regressors are found to be uncorrelated with factor loadings, the validity of which can be easily verified by our proposed test.

6 Conclusions

A large strand of the literature on panel data has focused on analysing CSD, based on the error components model with IE, which is implicitly understood to bias the conventional two-way FE estimator, due to the potential endogeneity arising from the correlation between regressors and factors/loadings. Two main approaches have been advocated to deal with this issue: the CCE estimator by Pesaran (2006) and the PC estimator by Bai (2009).

In this paper we have shown that the panel data model with IE can be encompassed by the standard two-way error components model if the regressors are correlated with factors but uncorrelated with the loadings. This suggests that the null hypothesis of no correlation between the regressors and factor loadings emerges as an influential but under-appreciated feature of the panel data model with IE. We propose the Hausman-type test, which follows the χ^2 distribution asymptotically under the null hypothesis. Monte Carlo simulation results confirm that the size and the power of the proposed test is quite satisfactory even in small samples.

Finally, we apply the proposed tests to a number of existing panel datasets, and find strong evidence in favor of the regressors uncorrelated with factor loadings in nine of ten datasets. In this situation, the FE estimator would provide a simple and robust estimation strategy in practice by avoiding nontrivial computational issues associated with the PC estimator, the performance of which relies crucially upon applying the complex bias-corrections and using reliable information criteria correctly selecting the number of unobserved factors.

We conclude by noting a couple of avenues for future research. A natural but challenging extension is to develop the LM-type test which does not require us to estimate the PC estimator at all. Next, it is worthwhile to develop the Hausman-type test in the dynamic heterogeneous panel data model with IE.

Table 9 Empirical applications to fourteen different datasets

		β_{FE}	se_{HAC}	β_{PC}	se_{HAC}	H_{HAC}	H_B	H_W	CD
Production function									
OECD	$\beta_{\frac{k}{T}}$	0.589	0.015	0.633	0.012	0.385	0.003	0.710	24.27
						(0.535)	(0.954)	(0.399)	(0.001)
ITA	$\beta_{\frac{k}{T}}$	0.537	0.035	0.343	0.016	0.921	0.004	0.707	16.75
						(0.337)	(0.946)	(0.401)	(0.001)
US	$\beta_{\frac{k}{T}}$	0.165	0.011	0.106	0.004	0.671	0.001	0.039	14.35
						(0.412)	(0.978)	(0.841)	(0.001)
UNIDO	$\beta_{\frac{k}{T}}$	0.637	0.022	0.563	0.036	0.258	0.008	4.091	22.08
						(0.612)	(0.929)	(0.043)	(0.001)
Production and *R&D*									
$k = 3$	β_l	0.603	0.013	0.487	0.010	0.898	2.175	2.476	12.65
	β_k	0.490	0.020	0.462	0.012	(0.825)	(0.536)	(0.479)	(0.001)
	β_{rd}	0.068	0.006	0.097	0.007				
Gravity model									
$k = 6$	β_{gdp}	3.299	0.039	1.808	0.010	21.09	11.41	24.48	97.3
	β_{rer}	0.037	0.004	− 0.041	0.006	(0.001)	(0.076)	(0.001)	(0.001)
	β_{sim}	1.362	0.022	1.231	0.057				
	β_{rlf}	0.033	0.002	0.014	0.005				
	β_{cee}	0.291	0.004	0.369	0.006				
	β_{emu}	0.254	0.006	0.179	0.006				
Gasoline demand									
$k = 2$	β_p	− 0.161	0.007	− 0.101	0.002	1.349	0.001	0.296	85.59
	β_{inc}	0.436	0.020	0.392	0.012	(0.501)	(0.999)	(0.862)	(0.001)

Table 9 continued

		β_{FE}	se_{HAC}	β_{PC}	se_{HAC}	H_{HAC}	H_B	H_W	CD
Income elasticity									
US States	β_{inc}	0.832	0.030	0.634	0.011	1.354	0.361	0.738	161.5
						(0.244)	(0.547)	(0.391)	(0.001)
US MSAs	β_{inc}	0.595	0.004	0.558	0.002	0.173	0.127	14.21	85.52
						(0.677)	(0.721)	(0.001)	(0.001)
Technological spillovers									
OECD countries									
$k = 3$	β_{sd}	0.057	0.006	0.087	0.003	0.571	0.002	1.657	5.762
	β_{sf}	0.139	0.061	0.113	0.044				
	β_{hc}	0.112	0.018	0.013	0.002	(0.903)	(0.999)	(0.646)	(0.001)
OECD industries	β_{l}	0.154	0.019	0.185	0.023				
$k = 4$	β_{nitc}	0.491	0.026	0.512	0.040				
	β_{itc}	0.024	0.008	0.084	0.006	1.416	0.974	4.121	6.447
	β_{rd}	0.027	0.003	0.031	0.008	(0.841)	(0.913)	(0.389)	(0.001)
Health	β_{gdp}	0.721	0.006	0.732	0.007	0.0219	0.003	3.676	132.9
$k = 2$	β_{pe}	0.019	0.002	0.021	0.009	(0.989)	(0.998)	(0.159)	(0.001)
Volatility	β_{s}	−0.617	0.370	−0.683	0.391	0.001	0.003	0.161	44.43
						(0.967)	(0.955)	(0.688)	(0.001)
Carbon emission	β_{gdp}	0.429	0.023	0.427	0.020				
$k = 4$	β_{tsh}	−0.106	0.015	−0.096	0.013				
	β_{ish}	−0.123	0.040	−0.125	0.066	0.239	0.168	2.323	12.21
	β_{fsh}	0.866	0.037	0.987	0.053	(0.993)	(0.996)	(0.676)	(0.001)

FE denotes the two-way fixed effects estimator. PC is the iterative principal component estimator in CHNY. The PC estimator is bias-corrected. Applying IC_{p1} criterion by Bai and Ng (2002), we extract the four factors in the gravity dataset while we estimate two factors in all other dataset. se_{HAC} are the standard errors evaluated by the non parametric HAC variance-covariance estimators. H_{HAC} are the H-statistic defined in (25) with the p-values inside (.). H_B is the Hausman test statistic proposed by Bai (2009); H_W is the Hausman test statistic proposed by Westerlund (2019b) while the CD Pesaran (2015) is applied to the one way fixed effects residuals

Declarations

Conflict of interest   The authors declare no conflict of interest

7 Appendix: Proofs

7.1 Preliminary lemmas

We first provide two Lemmas that extend the Law of Large Numbers and Central Limit Theorem to cover the martingale difference sequence for the panel data. We define the concept of spatial martingale difference arrays as follows: Let $W_{i,T}$ for $i = 1, \ldots, N$ and $T = 1, \ldots, T$, be arrays of matrices of random variables. Define the σ-field generated by $W_{j,T}$ for $j = 1, \ldots, N, j \neq i$, as $\mathcal{F}_{-i}$. Then, $W_{i,T}$ is a spatial martingale difference array if $E\left(W_{i,T}|\mathcal{F}_{-i}\right) = 0$, $i = 1, \ldots, N$. It is clear that the resulting sequence is a martingale difference array sequence for any ordering of the random matrices $W_{i,T}$.

Lemma 4   *Let $W_{i,T}$ and $\mu_{i,T}$ for $i = 1, \ldots, N$ and $T = 1, \ldots$, be arrays of matrices of random variables and constants such that $W_{i,T} - \mu_{i,T}$ is a spatial martingale difference array where $\sup_{i,T} E \left\| W_{i,T} \right\|^{1+\delta} < \infty$ for some $\delta > 0$. Then, as $(N, T) \to_j \infty$,*

$$N^{-1} \sum_{i=1}^{N} \left(W_{i,T} - \mu_{i,T} \right) \to_p 0.$$

Proof   By Theorem 12.11 of Davidson (1994), if $\sup_{i,T} E \left\| W_{i,T} \right\|^{1+\delta} < \infty$, then

$$\lim_{M \to \infty} \sup_{i,T} E \left(\left\| W_{i,T} - \mu_{i,T} \right\| I_{\{ \| W_{i,T} - \mu_{i,T} \| > M \}} \right) = 0,$$

which is a generalisation of uniform integrability to arrays. Then, the result follows immediately by Corollary 19.9 of Davidson (1994). □

Lemma 5   *Let $w_{i,T}$ and $\mu_{i,T}$, for $i = 1, \ldots, N$ and $T = 1, \ldots$, be arrays of vectors of random variables and constants such that $w_{i,T} - \mu_{i,T}$ is a spatial martingale difference*

array where $E\left[\left(w_{i,T} - \mu_{i,T}\right)\left(w_{i,T} - \mu_{i,T}\right)'\right] = \Sigma_{i,T}$, *and* $\sup_{i,T} E \|w_{iT}\|^{2+\delta} < \infty$ *for some* $\delta > 0$. *Assume that* $\Sigma = \lim_{N,T\to\infty} N^{-1} \sum_{i=1}^{N} \Sigma_{i,T}$ *is positive definite and* $\sup_{N,T} N^{-1} \sum_{i=1}^{N} \Sigma_{i,T} < \infty$. *Then, as* $(N, T) \to_j \infty$,

$$N^{-1} \sum_{i=1}^{N} (w_{iT} - \mu_{iT}) \to_d N(0, \Sigma). \tag{30}$$

Proof By Theorem 12.11 of Davidson (1994), if $\sup_{i,T} E \left\|w_{i,T}\right\|^{2+\delta} < \infty$, we obtain the uniform integrability condition,

$$\lim_{M\to\infty} \sup_{i,T} E\left(\left\|W_{i,T} - \mu_{i,T}\right\| I_{\{\|W_{i,T}-\mu_{i,T}\|>M\}}\right) = 0.$$

Together with $\sup_{N,T} N^{-1} \sum_{i=1}^{N} \Sigma_{i,T} < \infty$, this implies that the Lindeberg condition holds by Theorem 23.18 of Davidson (1994). Then, by Theorem 23.16 of Davidson (1994), it follows that

$$\max_{i,T} N^{-1} (w_{iT} - \mu_{iT}) \to_p 0. \tag{31}$$

Together with $\sup_{i,T} E \left\|w_{i,T}\right\|^{2+\delta} < \infty$, (31) implies (30) by Theorem 24.3 of Davidson (1994). □

7.2 Proof of Theorem 1

Considering $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$, we have

$$\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta} = \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i\right)^{-1} \sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \left(\ddot{\boldsymbol{X}}_i \boldsymbol{\eta}_i + \dot{\boldsymbol{F}} \mathring{\boldsymbol{\gamma}}_i + \ddot{\boldsymbol{\varepsilon}}_i\right)$$

$$= \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i\right)^{-1} \sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \left(\ddot{\boldsymbol{X}}_i \boldsymbol{\eta}_i + \dot{\boldsymbol{F}} \mathring{\boldsymbol{\gamma}}_i + \boldsymbol{\varepsilon}_i\right) + o_p(1). \tag{32}$$

Using Lemma 4, it is easily seen that as $(N, T) \to_j \infty$,

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i}{T} \to_p \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\frac{\ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i}{T}\right) = \boldsymbol{\Psi}_{FE}$$

Next, by the independence of $\boldsymbol{\gamma}_i$ and $\boldsymbol{\eta}_i$ each other and from $\boldsymbol{X}_i$ and $\boldsymbol{F}$ across i, and using the fact that $E\left(\mathring{\boldsymbol{\gamma}}_i\right) = E(\boldsymbol{\eta}_i) = 0$, it follows that $\ddot{\boldsymbol{X}}_i' \left(\dot{\boldsymbol{F}} \mathring{\boldsymbol{\gamma}}_i + \ddot{\boldsymbol{\varepsilon}}_i\right)$ is a spatial martingale difference but also a martingale difference sequence for any ordering across i. To see this, for any ordering over i, we have:

$$E\left[\ddot{\boldsymbol{X}}_i' \left(\ddot{\boldsymbol{X}}_i \boldsymbol{\eta}_i + \dot{\boldsymbol{F}} \mathring{\boldsymbol{\gamma}}_i + \boldsymbol{\varepsilon}_i\right) | \ddot{\boldsymbol{X}}_j, \dot{\boldsymbol{F}}, \mathring{\boldsymbol{\gamma}}_j, \boldsymbol{\eta}_j\right]$$

$$= E\left[\ddot{X}_i'\ddot{X}_i\eta_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] + E\left[\ddot{X}_i'\dot{F}\mathring{\pmb{\gamma}}_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] + E\left[\ddot{X}_i'\varepsilon_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right]$$

$$= E\left[\ddot{X}_i'\ddot{X}_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] E\left[\pmb{\eta}_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right]$$

$$+ E\left[\ddot{X}_i'\dot{F}|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] E\left[\mathring{\pmb{\gamma}}_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right]$$

$$+ E\left[\ddot{X}_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] E\left[\varepsilon_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] \text{ for } j \neq i$$

Since

$$E\left[\pmb{\eta}_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] = E\left[\mathring{\pmb{\gamma}}_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] = E\left[\varepsilon_i|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] = 0 \text{ for } j \neq i$$

hence

$$E\left[\ddot{X}_i'\left(\ddot{X}_i\eta_i + \dot{F}\mathring{\pmb{\gamma}}_i + \varepsilon_i\right)|\ddot{X}_j, \dot{F}, \mathring{\pmb{\gamma}}_j, \pmb{\eta}_j\right] = 0 \text{ for } j \neq i$$

which proves the martingale difference property. Notice that we repeatedly use the fact that the product of a stochastic process with a second process, that is independent over its index as well as of the first process, is a martingale difference process. Next, we note that $\left\{\frac{\ddot{X}_i'\varepsilon_i}{\sqrt{T}}\right\}_{i=1}^N$, $\left\{\frac{\ddot{X}_i'\dot{F}\mathring{\pmb{\gamma}}_i}{T}\right\}_{i=1}^N$ and $\left\{\frac{\ddot{X}_i'\ddot{X}_i\eta_i}{T}\right\}_{i=1}^N$ are spatial martingale difference series. Notice that $\left(\sum_{i=1}^N \ddot{X}_i'\ddot{X}_i\right)^{-1}\sum_{i=1}^N \ddot{X}_i'\varepsilon_i = O_p\left(\frac{1}{\sqrt{NT}}\right)$, is of the smaller probability order of magnitude than the other two terms in the RHS of (32). Therefore, it follows that as $(N, T) \to_j \infty$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\ddot{X}_i'\dot{F}\mathring{\pmb{\gamma}}_i}{T} \to_d N\left(0, \pmb{R}_{1,FE}\right)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\ddot{X}_i'\left(\ddot{X}_i\eta_i + \dot{F}\mathring{\pmb{\gamma}}_i\right)\dot{X}_i'}{T} \to^d N\left(0, \pmb{R}_{1,FE} + \pmb{R}_{2,FE}\right).$$

where $\pmb{R}_{1,FE}$ are $\pmb{R}_{2,FE}$ defined in (13) and (14). This proves (12) in Theorem 1. We will prove (15) in the proof of Theorem 3.

The results for the case of $\eta = 0$, $\forall i$, follows straightforwardly from the above analysis.

7.3 Proof of Theorem 2

Let R_{NT} denote terms of the lower order of probability than the leading terms. By Theorem 6 in CHNY, considering $\pmb{\beta}_i = \pmb{\beta} + \pmb{\eta}_i$, we have:

$$\hat{\pmb{\beta}}_{PC} - \pmb{\beta} = \left(\sum_{i=1}^N V_i'V_i\right)^{-1}\sum_{i=1}^N V_i'V_i\eta_i + R_{NT}$$

CHNY assume that x_{it} follows a linear factor structure, (2), which can be expressed as

$$X_i = F\Gamma_i + V_i, \tag{33}$$

where $X_i = (x_{i1}, \ldots, x_{iT})'$, $F = (f_1, \ldots, f_T)'$ and $V_i = (v_{i1}, \ldots, v_{iT})'$. See also Assumptions B1-B5 in CHNY.

Note that $\sum_{i=1}^{N} V_i' V_i \eta_i = O_p\left(\sqrt{NT}\right)$ and $\sum_{i=1}^{N} V_i' \varepsilon_i = O_p\left(\sqrt{NT}\right)$. Using Lemmas 4 and 5, it follows that as $(N, T) \to_j \infty$,

$$\left(\frac{1}{N} \sum_{i=1}^{N} \frac{V_i' V_i}{T}\right)^{-1} \to_p \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\frac{V_i' V_i}{T}\right) = \Psi_{PC}$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{V_i' V_i \eta_i}{T} \to_d N\left(0, R_{1,PC}\right)$$

where $R_{1,PC}$ is defined in (22). This proves (21). (23) follows by the proof of Theorem 3.

7.4 Proof of Theorem 3

Given Theorems 1 and 2, it suffices to derive the equivalence and consistency of the two robust covariance estimators for $\hat{\beta}_{FE}$ and $\hat{\beta}_{PC}$, which are given by (10), (11), (19) and (20), respectively. Rewrite them compactly as

$$V^{NON}\left(\hat{\beta}_{FE}\right) = \hat{\Psi}_{FE}^{-1} \hat{R}_{FE}^{NON} \hat{\Psi}_{FE}^{-1}; \quad V^{HAC}\left(\hat{\beta}_{FE}\right) = \hat{\Psi}_{FE}^{-1} \hat{R}_{FE}^{HAC} \hat{\Psi}_{FE}^{-1}$$

$$V^{NON}\left(\hat{\beta}_{PC}\right) = \hat{\Psi}_{PC}^{-1} \hat{R}_{PC}^{NON} \hat{\Psi}_{PC}^{-1}; \quad V^{HAC}\left(\hat{\beta}_{PC}\right) = \hat{\Psi}_{PC}^{-1} \hat{R}_{PC}^{HAC} \hat{\Psi}_{PC}^{-1}$$

where $\hat{X}_i = M_{\hat{F}} X_i$,

$$\hat{\Psi}_{FE} = \sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i, \quad \hat{\Psi}_{PC} = \sum_{i=1}^{N} \hat{X}_i' \hat{X}_i = \sum_{i=1}^{N} X_i' M_{\hat{F}} X_i,$$

$$\hat{R}_{FE}^{NON} = \sum_{i}^{N} \left(\ddot{X}_i' \ddot{X}_i\right) \left(\hat{\beta}_{FE,i} - \hat{\beta}_{FE}\right) \left(\hat{\beta}_{FE,i} - \hat{\beta}_{FE}\right)' \left(\ddot{X}_i' \ddot{X}_i\right),$$

$$\hat{R}_{PC}^{NON} = \sum_{i}^{N} X_i' M_{\hat{F}} X_i \left(\tilde{\beta}_{PC,i} - \tilde{\beta}_{PC}\right) \left(\tilde{\beta}_{PC,i} - \tilde{\beta}_{PC}\right)' X_i' M_{\hat{F}} X_i,$$

$$\hat{R}_{FE}^{HAC} = \sum_{i}^{N} \ddot{X}_i' \hat{u}_{FE,i} \hat{u}_{FE,i}' \ddot{X}_i, \quad \hat{R}_{PC}^{HAC} = \sum_{i}^{N} \hat{X}_i' \hat{u}_{PC,i} \hat{u}_{PC,i}' \hat{X}_i.$$

Finally, we define:

$$\hat{C}^{NON}\left(\hat{\beta}_{FE},\hat{\beta}_{PC}\right) = \hat{\Psi}_{FE}^{-1}\hat{R}_{FE,PC}^{NON}\hat{\Psi}_{PC}^{-1}; \ \hat{C}^{HAC}\left(\hat{\beta}_{FE},\hat{\beta}_{PC}\right) = \hat{\Psi}_{FE}^{-1}\hat{R}_{FE,PC}^{HAC}\hat{\Psi}_{PC}^{-1}$$

where

$$\hat{R}_{FE,PC}^{NON} = \sum_{i}^{N}\ddot{X}_{i}'\ddot{X}_{i}\left(\hat{\beta}_{FE,i} - \hat{\beta}_{FE}\right)\left(\tilde{\beta}_{PC,i} - \tilde{\beta}_{PC}\right)'\hat{X}_{i}'\hat{X}_{i},$$

$$\hat{R}_{FE,PC}^{HAC} = \sum_{i}^{N}\ddot{X}_{i}'\hat{u}_{FE,i}\hat{u}_{PC,i}'\hat{X}_{i}$$

To establish that the two covariance estimators are (asymptotically) equivalent, we need to show:

$$\hat{R}_{FE}^{NON} = \hat{R}_{FE}^{HAC} + R_{NT} \tag{34}$$

$$\hat{R}_{PC}^{NON} = \hat{R}_{PC}^{HAC} + R_{NT} \tag{35}$$

$$\hat{R}_{FE,PC}^{NON} = \hat{R}_{FE,PC}^{HAC} + R_{NT} \tag{36}$$

where R_{NT} denotes terms of the lower order of probability than the leading terms. We focus on the PC estimator in (35). First, consider $\hat{R}_{PC}^{HAC}$ and notice that

$$\hat{X}_{i}'\hat{u}_{PC,i} = \hat{X}_{i}'\left(\hat{u}_{PC,i} + \hat{X}_{i}\left(\hat{\beta}_{PC} - \beta\right)\right).$$

By Theorem 11 in CHNY, it follows that as $(N, T) \to \infty$ and $\frac{T}{N} \to c \in (0, \Delta]$ with $\Delta < \infty$,

$$\sum_{i=1}^{N}\hat{X}_{i}'\hat{u}_{PC,i}\hat{u}_{PC,i}'\hat{X}_{i} = \sum_{i=1}^{N}V_{i}'u_{PC,i}u_{PC,i}'V_{i} + R_{NT}$$

where $u_{PC,i} = X_{i}\eta_{i} + \varepsilon_{i}$. Then, we have

$$\hat{R}_{PC}^{HAC} = \sum_{i=1}^{N}V_{i}'V_{i}\eta_{i}\eta_{i}'V_{i}'V_{i} + \sum_{i=1}^{N}V_{i}'\varepsilon_{i}\varepsilon_{i}'V_{i} + R_{NT}. \tag{37}$$

Next, it is easily seen that

$$\tilde{\beta}_{PC,i} - \beta = \left(\hat{X}_{i}'\hat{X}_{i}\right)^{-1}\hat{X}_{i}'\varepsilon_{i} + \eta_{i}$$

$$\tilde{\beta}_{PC} - \beta = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\hat{X}_{i}'\hat{X}_{i}\right)^{-1}\hat{X}_{i}'\varepsilon_{i} + \eta_{i}\right]$$

Then,

$$
\begin{aligned}
\hat{X}'_i \hat{X}_i \left(\tilde{\beta}_{PC,i} - \tilde{\beta}_{PC} \right) &= \hat{X}'_i \hat{X}_i \left(\tilde{\beta}_{PC,i} - \beta_i + \beta_i - \beta + \beta - \tilde{\beta}_{PC} \right) \\
&= \hat{X}'_i \hat{X}_i \left(\tilde{\beta}_{PC,i} - \beta_i \right) + \hat{X}'_i \hat{X}_i \eta_i + \hat{X}'_i \hat{X}_i \left(\beta - \tilde{\beta}_{PC} \right) \\
&= \hat{X}'_i \varepsilon_i + \hat{X}'_i \hat{X}_i \eta_i + \hat{X}'_i \hat{X}_i \left(\beta - \tilde{\beta}_{PC} \right)
\end{aligned}
\tag{38}
$$

By Theorem 11 in CHNY, we then obtain:

$$
\hat{R}_{PC}^{NON} = \sum_{i=1}^{N} V'_i V_i \eta_i \eta'_i V'_i V_i + \sum_{i=1}^{N} V'_i \varepsilon_i \varepsilon'_i V_i + R_{N,T}.
\tag{39}
$$

This proves (35).

Noticing that both $\frac{V'_i \varepsilon_i \varepsilon'_i V_i}{T} - E \left(\frac{V'_i \varepsilon_i \varepsilon'_i V_i}{T} \right)$ and $\frac{V'_i V_i}{T} \eta_i \eta'_i \frac{V'_i V_i}{T} - E \left(\frac{V'_i V_i}{T} \eta_i \eta'_i \frac{V'_i V_i}{T} \right)$ are iid and martingale difference processes over i and by Lemma 4, we have:

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{V'_i V_i}{T} \eta_i \eta'_i \frac{V'_i V_i}{T} \to_p \lim_{NT,\to\infty} \frac{1}{N} \sum_{i=1}^{N} E \left(\frac{V'_i V_i}{T} \eta_i \eta'_i \frac{V'_i V_i}{T} \right)
$$

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{V'_i \varepsilon_i \varepsilon'_i V_i}{T} \to_p \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} E \left(\frac{V'_i \varepsilon_i \varepsilon'_i V_i}{T} \right)
$$

This provides consistency of both variance estimators.

Along similar lines to (38), it is straightforward to prove (34) and (36) because

$$
\left(\ddot{X}'_i \ddot{X}_i \right) \left(\hat{\beta}_{FE,i} - \hat{\beta}_{FE} \right) = \ddot{X}'_{ii} \varepsilon_i + \ddot{X}'_i \ddot{X}_i \eta_i + \ddot{X}'_i \dot{F} \mathring{\gamma}_i + \ddot{X}'_i \ddot{X}_i \left(\beta - \hat{\beta}_{FE} \right),
$$

and the term $\ddot{X}'_i \dot{F} \mathring{\gamma}_i$ can be analysed similarly to $\ddot{X}'_i \ddot{X}_i \eta_i$. Using these results, it readily follows that

$$
\left(\hat{\beta}_{FE} - \hat{\beta}_{PC} \right)' \left(\hat{V}^{NON} \right)^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{PC} \right) \sim \chi_k^2
$$

$$
\left(\hat{\beta}_{FE} - \hat{\beta}_{PC} \right)' \left(\hat{V}^{HAC} \right)^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{PC} \right) \sim \chi_k^2
$$

where $\hat{V}^{NON}$ and $\hat{V}^{HAC}$ are defined in (26) and (27), respectively.

8 Appendix: The data and empirical specifications

We describe the empirical specifications and the data in details. For the production function, we estimate the following panel data regression:

$$\ln\left(\frac{Y}{L}\right)_{it} = \beta \ln\left(\frac{K}{L}\right)_{it} + e_{it}, \ e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it} \tag{40}$$

The first group consists of 26 OECD countries; Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Israel, Italy, Japan, Korea, Mexico, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Turkey, the U.K. and the U.S. The data is collected from PWT 7.0 and covers the period 1970–2010. Y is GDP measured in million U.S. \$ at the 2005 price, K the capital measured in millions U.S. \$, constructed using the perpetual inventory method (PIM), and L the labour measured as the total employment in thousands. The second group contains the EU27 countries; Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and the U.K. The data are extracted from PWT 9.0 over the period 1990–2015 and the definition of the variables, Y, K and L is the same as above. The third group includes 20 Italian regions over the period 1995–2016; Piemonte, Valle d'Aosta, Liguria, Lombardia, Trentino Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, Toscana, Umbria, Marche, Lazio, Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia and Sardegna. Due to the data availability, we construct Y by the value added measured in million Euros at the 2010 price, L by the total employment in thousands, and K by Gross Fixed Capital Formation in millions Euros. The data, gathered from ISTAT, covers the period 1995 to 2000. The fourth data taken from Munnell (1990), comprises the 48 U.S. states and covers the period, 1970–1986. Y is the per capita gross state product, K is the private capital computed by apportioning Bureau of Economic Analysis (BEA) national stock estimates, and L is the number of employers in thousands in non-agricultural payrolls. The fifth application employs the aggregate sectorial data for manufacturing from developed and developing countries for the period 1970–2002, collected from UNIDO by Eberhardt and Teal (2019). We extract a balanced panel of 25 countries with 25 time periods from 1970 to 1995, where we cover Australia, Belgium, Brazil, Colombia, Cyprus, Ecuador, Egypt, Spain, Finland, Fiji, France, Hungary, Indonesia, India, Italy, Korea, Malta, Norway, Panama, Philippines, Poland, Portugal, Singapore, the USA and Zimbabwe.

The production function is augmented with R&D in the sixth application. From the data provided by Eberhardt et al. (2013), we extract a balanced panel of 82 country-industry units representing manufacturing industries across ten OECD economies (Denmark, Finland, Germany, Italy, Japan, Netherlands, Portugal, Sweden, the United Kingdom, and the US) from 1980 to 2005. We consider an augmented Cobb-Douglas production function:

$$\ln Y_{it} = \beta_l \ln L_{it} + \beta_k \ln K_{it} + \beta_{rd} \ln RD_{it} + e_{it}, \ e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it} \qquad (41)$$

where Y measured as deflated value added, L is the total number of hours worked by persons engaged, K is the total tangible assets by book value and RD is the R&D stock expenditure. See Eberhardt et al. (2013) for details.

Next, we consider the gravity model specifications for the bilateral trade flows given by

$$\begin{aligned} \ln (trade_{it}) \ = \ & \beta_{gdp} \ln (gdp_{it}) + \beta_{rer} \ln (rer_{it}) + \beta_{sim} \ln (sim_{it}) + \beta_{rlf} \ln (rlf_{it}) \\ & + \beta_{cee} cee_{it} + \beta_{euro} euro_{it} + e_{it}, e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it} \end{aligned} \qquad (42)$$

Here, $trade_{it}$ is the sum of bilateral import flows ($import_{odt}$) and export flows ($export_{odt}$) measured in million U.S. dollars at the 2000 price with o and d denoting the origin and the destination country, gdp_{it} is the sum of gdp_{ot} and gdp_{dt} both of which are measured as the gross domestic product at the 2000 dollar price, $rer_{it} = ner_{odt} \times xpi_{US}$ is the real exchange rate measured in the 2000 dollar price, where ner_{hft} is the bilateral nominal exchange rate normalised in terms of the U.S. \$, sim is a measure of similarity in size constructed by $sim_{it} = \left[1 - \left(\frac{gdp_{ot}}{gdp_{ot} + gdp_{dt}} \right)^2 - \left(\frac{gdp_{dt}}{gdp_{ot} + gdp_{dt}} \right)^2 \right]$, and $rlf_{it} = |pgdp_{ot} - pgdp_{dt}|$ measures countries' difference in relative factor endowment where $pgdp$ is per capita GDP. cee and $euro$ represent dummies equal to one when countries of origin and destination both belong to the European Economic Community and share the euro as common currency, respectively. The data are collected from the IMF Direction of Trade Statistics, and covers the period, 1960–2008. We consider a sample of 91 country-pairs amongst the EU14 member countries (Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden and the U.K.).

Further, we estimate the gasoline demand function by

$$\ln (q_{it}) = \beta_p \ln (p_{it}) + \beta_{inc} \ln (inc_{it}) + e_{it}, \ e_{it} = \alpha_i + \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it} \qquad (43)$$

where gasoline consumption, q_{it}, is approximated as monthly sales volumes of motor gasoline, per capita per day; p_{it} is the after tax gasoline prices computed by adding the state/federal tax rates to the motor gasoline sales to end user price and inc_{it} represent the quarterly personal disposable income. Prices, income, and tax rates are converted to constant 2005 dollars using GDP implicit price deflators. The source of data is Liu (2014).

Finally, we estimate the income elasticity of real house price using the specification:

$$\ln (p_{it}) = \beta_{inc} \ln (inc_{it}) + e_{it}, \ e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it}, \qquad (44)$$

where p_{it} is the housing price index and inc_{it} is the real per capital income. We consider two annual datasets. The first sample from Holly et al. (2010) consists of the panel data for 48 U.S. States (excluding Alaska and Hawaii) plus the District of Columbia ($N = 49$) over the period 1975–2010. The second sample taken from

Baltagi and Li (2014), contains a panel data of 384 Metropolitan Statistical Areas over the period 1975–2010.

Technological knowledge spillovers We estimate the effect of technological knowledge spillovers on total factor productivity (TFP). Following Coe et al. (2009), we estimate the effect of domestic and foreign $R\&D$ capital stocks on TFP, controlling for the impact of human capital, using the following regression:

$$tfp_{it} = \beta_{sd} s_{it}^d + \beta_{sf} s_{it}^f + \beta_{hc} hc_{it} + e_{it}, \ e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it}$$

where tfp_{it} is the TFP, s_{it}^d is domestic R&D capital stocks, s_{it}^f is foreign R&D capital stocks and hc_{it} is human capital. TFP is defined as the log of output minus a weighted average of labor and capital inputs; domestic and foreign R&D capital stocks are measured in U.S. \$ at 2000 prices and PPP exchange rates. We use a balanced panel of 24 OECD countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, Korea, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the U.K. and the U.S.) observed over the period 1971–2004. The measures of TFP and R&D capital stock come from Coe et al. (2009) (source: OECD Economic Outlook) while the average number of years of schooling used to measure human capital is taken from (Ertur and Musolesi 2017).

Next, we explore the channels through which technological investment affects the productivity performance of industrialized economies using industry-level data, extracted from the EU KLEMS database and the OECD ANBERD database. The sample includes fourteen OECD countries (Austria, Belgium, Denmark, Germany, Spain, Finland, France, Ireland, Italy, Japan, Netherlands, Sweden, the UK and the US). In particular, we estimate the productivity effects of R&D and ICT identifying three channels of transmission: input accumulation, technological change and spillovers by the following specification

$$\ln(va_{it}) = \beta_l \ln(l_{it}) + \beta_{nitc} \ln(nonITC_{it}) + \beta_{itc} \ln(ITC_{it}) + \beta_{rd} \ln(rd_{it})$$
$$+ e_{it}, \ e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it}$$

where va_{it} is value added; l_{it} is the labour input measured as the number of hours worked; rd_{it} is the R&D input measured as the cumulative value of industry research expenses; ICT (ITC_{it}) and non-ICT ($nonITC_{it}$) assets are built from annual investment flows by means of the perpetual inventory method and adopting an asset-specific rate of geometric depreciation. We extract a balanced panel of 49 high-tech industries, from 1977 to 2006, from the dataset provided by Pieri et al. (2018).

Health care expenditure and income We estimate the relationship between healthcare expenditure and income using data on 167 countries over the period 1995–2012 by

$$\ln(h_{it}) = \beta_{gdp} \ln(gdp_{it}) + \beta_{pe} \ln(pe_{it}) + e_{it}, \ e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it}$$

where h_{it} is per-capita health spending and gdp_{it} is per capita GDP and pe_{it} is public-health expenditure rate. The data on per-capita health expenditure and per-capita GDP

are expressed at constant 2005 PPP prices. The public-health expenditure rate is defined as the percentage of public expenditure over total health expenditure. The source is the World Bank and the dataset is taken from Baltagi et al. (2017).

Demographic and business cycle volatility We estimate the impact of the age composition of the labor force on business cycle volatility. Jaimovich and Siu (2009) argue that a significant fraction of the run-up of US volatility in the mid-1960s and of the marked decline since the mid-1980s, known as the Great Moderation, is accounted for by long swings in the age composition of the US population induced by the baby boom and subsequent baby bust. Jaimovich and Siu (2009) define the volatile-age labor force share, s_{it} as the fraction of the 15- to 64-year-old labor force accounted for by those aged 15–29 and 60–64 and linked it to the time-varying standard deviation of output σ_{it} in the following benchmark regression:

$$\sigma_{it} = \beta_s s_{it} + \varepsilon_{it}.$$

They show that shifts in the volatile-age share variable s_{it} have a large and significant effect on cyclical volatility in the G7 countries from 1963 to 1999. By using a different dataset and allowing for cross-sectional dependence Everaert and Vierke (2016) find a much lower effect. Here, we use the balanced panel dataset for 51 countries over the period 1957–2000 taken from Everaert and Vierke (2016) where the GDP data (taken from the Penn World Table) are used to calculate output volatility, which is defined as the 9-year rolling standard deviation of logged annual GDP. The demographic data are taken from the United Nations World Population Prospects.

Trade and carbon emissions We explore the nexus between carbon emissions and trade in the 32 OECD countries (Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, the UK and the US). Following Liddle (2018), we estimate the following regression:

$$co_{it} = \beta_{gdp} gdp_{it} + \beta_{tsh} tsh_{it} + \beta_{ish} ish_{it} + \beta_{fsh} fsh_{it} + e_{it}, \ e_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it}$$

where co_{it} is the consumption-based per capita carbon emissions, gdp_{it} the real GDP per capita, tsh_{it} the trade (import plus export) expressed as a percentage of the GDP, ish_{it} the industry value added expressed as a percentage of the GDP and fsh_{it} the fossil fuel energy consumption as a share of total energy consumption. Data sources are the World Bank's World Development Indicators and the Global Carbon Project. We extract a balanced panel of 32 OECD countries from 1990 to 2013, from the dataset in Liddle (2018).

9 Appendix: The bias corrected PC estimator

The bias corrected estimator proposed by Cui et al. (2019) is given by

$$\hat{\boldsymbol{\beta}}_{PC} = \tilde{\boldsymbol{\beta}}_{PC} - \frac{1}{N}\hat{\boldsymbol{B}}_{NT} - \frac{1}{T}\hat{\boldsymbol{C}}_{NT}$$

where the estimator for $(\boldsymbol{\beta}_{PC}, \boldsymbol{F})$ denoted as $(\tilde{\boldsymbol{\beta}}_{PC}, \hat{\boldsymbol{F}})$ is the solution of the set of nonlinear equations

$$\tilde{\boldsymbol{\beta}}_{PC} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}} \boldsymbol{y}_i \text{ and}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_{PC}\right)\left(\boldsymbol{y}_i - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_{PC}\right)'\right]\hat{\boldsymbol{F}} = \hat{\boldsymbol{F}}\boldsymbol{V}_{NT}$$

where $\boldsymbol{M}_{\hat{F}} = \boldsymbol{I}_T - \hat{\boldsymbol{F}}\left(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}\right)^{-1}\hat{\boldsymbol{F}}'$, $\boldsymbol{V}_{NT}$ is the diagonal matrix that consists r largest eigenvalues of the above matrix in the brackets arranged in a decreasing order and $\hat{\boldsymbol{F}}$ is $\sqrt{T}$ times the corresponding eigenvectors. The bias correction term is given by

$$\hat{\boldsymbol{B}}_{NT} = -\left(\frac{1}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{Z}}_i\boldsymbol{M}_{\hat{F}}\hat{\boldsymbol{Z}}_i'\right)^{-1}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{Z}}_i\hat{\boldsymbol{F}}\hat{\boldsymbol{\Upsilon}}_\gamma^{-1}\hat{\gamma}_i\hat{u}_{it}^2$$

where $\hat{\boldsymbol{Z}}_i = \boldsymbol{X}_i - \frac{1}{N}\sum_{j=1}^{N}\hat{a}_{ij}\boldsymbol{X}_j$ with $\hat{a}_{ij} = \hat{\gamma}_i'\hat{\boldsymbol{\Upsilon}}_\gamma^{-1}\hat{\gamma}_j$, $\hat{\boldsymbol{\Upsilon}}_\gamma = (\hat{\Gamma}'\hat{\Gamma}/N)$, $\hat{\Gamma} = (\hat{\gamma}_1, \hat{\gamma}_2, \dots \hat{\gamma}_N)'$ and

$$\hat{\boldsymbol{C}}_{NT} = -\left(\frac{1}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{Z}}_i\boldsymbol{M}_{\hat{F}}\hat{\boldsymbol{Z}}_i'\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{X}}_i\boldsymbol{M}_{\hat{F}}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{F}}\hat{\boldsymbol{\Upsilon}}_\gamma^{-1}\hat{\gamma}_i$$

where $\hat{\boldsymbol{\Omega}} = diag\left(\frac{1}{N}\sum_{j=1}^{N}\hat{u}_{j1}^2, \dots, \frac{1}{N}\sum_{j=1}^{N}\hat{u}_{jT}^2\right)$.

10 Appendix: Further empirical results

See Table 10.

Table 10 Empirical applications to fourteen different datasets

		β_{FE}	se_{NON}	β_{PC}	se_{NON}	H_{NON}
Production function						
OECD	$\beta_{\frac{k}{l}}$	0.589	0.015	0.633	0.012	0.499
						(0.479)
ITA	$\beta_{\frac{k}{l}}$	0.537	0.042	0.343	0.016	0.671
						(0.413)

Table 10 continued

		β_{FE}	se_{NON}	β_{PC}	se_{NON}	H_{NON}
US	$\beta_{\frac{k}{l}}$	0.165	0.011	0.106	0.042	0.647 (0.421)
UNIDO	$\beta_{\frac{k}{l}}$	0.637	0.028	0.563	0.036	0.193 (0.659)
Production and $R\&D$	β_l	0.603	0.018	0.487	0.008	
$k = 3$	β_k	0.490	0.025	0.462	0.014	0.183
	β_{rd}	0.068	0.007	0.097	0.006	(0.980)
Gravity model	β_{gdp}	3.299	0.050	1.808	0.010	
$k = 6$	β_{rer}	0.037	0.006	-0.041	0.006	
	β_{sim}	1.362	0.036	1.231	0.041	
	β_{rlf}	0.033	0.006	0.014	0.005	
	β_{cee}	0.291	0.006	0.369	0.006	19.20
	β_{emu}	0.254	0.008	0.179	0.006	(0.003)
Gasoline demand	β_p	-0.161	0.008	-0.101	0.002	1.404
$k = 2$	β_{inc}	0.436	0.022	0.392	0.011	(0.495)
Income elasticity						
US States	β_{inc}	0.832	0.032	0.634	0.011	1.404 (0.495)
US MSAs	β_{inc}	0.595	0.003	0.568	0.002	0.172 (0.677)
Technological spillovers						
OECD countries	β_{sd}	0.057	0.007	0.087	0.006	
$k = 3$	β_{sf}	0.139	0.041	0.133	0.057	0.981
	β_{hc}	0.112	0.063	0.013	0.074	(0.805)
OECD industries	β_l	0.154	0.033	0.185	0.025	
$k = 4$	β_{nitc}	0.491	0.041	0.512	0.034	
	β_{itc}	0.024	0.008	0.084	0.007	0.694
	β_{rd}	0.027	0.004	0.031	0.011	(0.951)
Health	β_{gdp}	0.721	0.008	0.732	0.0073	0.017
$k = 2$	β_{pe}	0.019	0.009	0.021	0.0107	(0.991)
Volatility	β_s	-0.617	0.373	-0.683	0.391	0.001 (0.967)
Carbon emission	β_{gdp}	0.429	0.023	0.427	0.020	
$k = 4$	β_{tsh}	-0.106	0.016	-0.096	0.014	
	β_{ish}	-0.123	0.045	-0.125	0.066	0.227
	β_{fsh}	0.866	0.039	0.987	0.058	(0.994)

H_{NON} are the H-statistic defined in (24) with the p-values inside (.). se_{NON} are the standard errors evaluated by the non parametric NON variance-covariance estimators, see notes to Table 9

References

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

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

Baltagi BH, Li J (2014) Further evidence on the spatio-tempral model oh house prices in the Unite States. J Appl Econom 29(3):515–522

Baltagi BH, Lagravinese R, Moscone F, Tosetti E (2017) Health care expenditure and income: a global perspective. Health Econ 26(7):863–874

Castagnetti C, Rossi E, Trapani L (2015) Testing for no factor structures: on the use of Hausman-type statistics. Econ Lett 130:66–68

Charbonneau KB (2017) Multiple fixed effects in binary response panel data models. Econom J 20(3):S1–S13

Chudik A, Pesaran MH (2015) Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. J Econom 188(2):393–420

Chudik A, Pesaran MH, Tosetti E (2011) Weak and strong cross section dependence and estimation of large panels. Econom J 14(1):C45–C90

Coakely J, Fuertes AM, Smith R (2006) Unobserved heterogeneity in panel time series models. Comput Stat Data Anal 50(1):2361–2380

Coe DT, Helpman E, Hoffmaister AW (2009) International R&D spillovers and institutions. Eur Econ Rev 53(7):723–741

Cui G, Hayakawa K, Nagata S, Yamagata T (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

Davidson J (1994) Stochastic limit theory. Oxford University Press, Oxford

Eberhardt M, Teal F (2019) The magnitude of the task ahead: macro implications of heterogeneous technology. Rev Income Wealth 66:334–360

Eberhardt M, Helmers C, Strauss H (2013) Do spillovers matter when estimating private returns to R&D? Rev Econ Stat 95(2):436–448

Ertur C, Musolesi A (2017) Weak and strong cross-sectional dependence: a panel data analysis of international technology diffusion. J Appl Econom 32(3):477–503

Everaert G, Vierke H (2016) Demographics and business cycle volatility: A spurious relationship? J Appl Econom 31(7):1467–1477

Fernandez-Val I, Weidner M (2016) Individual and time effects in nonlinear panel models with large N, T. J Econom 192(1):291–312

Hausman JA (1978) Specification tests in econometrics. Econometrica 46(6):1251–1271

Holly S, Pesaran MH, Yamagata T (2010) A spatio-temporal model of house prices in the USA. J Econom 158(1):160–173

Hsiao C (2018) Panel models with interactive effects. J Econom 206(2):645–673

Jaimovich N, Siu HE (2009) The young, the old, and the restless: demographics and business cycle volatility. Am Econ Rev 99(3):804–26

Juodis A, Reese S (2018) The incidental parameters problem in testing for remaining cross-section correlation. Papers arXiv:1810.03715

Kapetanios G, Pesaran MH (2005) Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns. CESifo working paper series 1416, CESifo Group Munich

Kapetanios G, Pesaran MH, Yamagata T (2011) Panels with non-stationary multifactor error structures. J Econom 160(2):326–348

Karabiyik H, Reese S, Westerlund J (2017) On the role of the rank condition in CCE estimation of factor-augmented panel regressions. J Econom 197(1):60–64

Liddle B (2018) Consumption-based accounting and the trade-carbon emissions nexus. Energy Econ 69:71–78

Liu W (2014) Modeling gasoline demand in the united states: a flexible semiparametric approach. Energy Econ 45:244–253

Mastromarco C, Serlenga L, Shin Y (2016) Modelling technical efficiency in cross sectionally dependent stochastic frontier panels. J Appl Econom 31(1):281–297

Moon HR, Weidner M (2015) Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83(4):1543–1579

Munnell A (1990) How does public infrastructure affect regional economic performance? Conf Ser [Proc] 34:69–112

Pesaran MH (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4):967–1012

Pesaran MH (2015) Testing weak cross-sectional dependence in large panels. Econom Rev 34(6):1089–1117

Petrova Y, Westerlund J (2020) Fixed effects demeaning in the presence of interactive effects in treatment effects regressions and elsewhere. J Appl Econom 35:960–964

Pieri F, Vecchi M, Venturini F (2018) Modelling the joint impact of R&D and ICT on productivity: a frontier analysis approach. Res Policy 47(9):1842–1852

Sarafidis V, Wansbeek T (2012) Cross-sectional dependence in panel data analysis. Econom Rev 31(5):483–531

Sarafidis V, Yamagata T, Robertson D (2009) A test of cross section dependence for a linear dynamic panel model with regressors. J Econom 148(2):149–161

Serlenga L, Shin Y (2007) Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. J Appl Econom 22(2):361–381

Westerlund J (2019a) On estimation and inference in heterogeneous panel regressions with interactive effects. J Time Ser Anal 40(5):852–857

Westerlund J (2019b) Testing additive versus interactive effects in fixed-ICT panels. Econ Lett 174(C):5–8

Westerlund J, Urbain J-P (2013) On the estimation and inference in factor-augmented panel regressions with correlated loadings. Econ Lett 119(3):247–250

Westerlund J, Urbain J-P (2015) Cross-sectional averages versus principal components. J Econom 185(2):372–377