Two-Way Fixed Effects versus Panel Factor Augmented Estimators: Asymptotic Comparison among Pre-testing Procedures

Minyu Han Jihun Kwak Donggyu Sul

University of Texas at Dallas

Incomplete

February 16, 2019

Abstract

Empirical researchers may wonder whether or not a two-way fixed effects estimator (with individual and year fixed effects) is good enough to isolate the influence of common shocks on the estimation of slope coefficients. Otherwise, they need to run the so-called panel factor augmented regressions instead. There are two pre-testing procedures available in the literature: The use of the number of factors and the direct testing of estimated factor loading coefficients. This paper compares the two pre-testing methods asymptotically. Under the alternative of the heterogeneous factor loadings, both pre-testing procedures suggest to use the commonly correlated effects (CCE) estimator. Meanwhile under the null of the homogenous factor loadings, the pre-testing method used by the number of factors always suggests more efficient estimations. By comparing asymptotic variances, this paper finds that when the slope coefficients are homogeneous with homogenous factor loadings, the two-way fixed effects estimation is more efficient than the CCE estimation. Meanwhile the slope coefficients are heterogeneous with homogenous factor loadings, the CCE estimation is, surprisingly, more efficient than the two-way fixed effects estimation. By means of Monte Carlo simulations, we verify the asymptotic claims. We demonstrate how to use the two pre-testing methods by taking an empirical example.

JEL Classification Number: C33

Keywords: Two-way fixed effects estimator, Panel factor augmented estimator, Factor number, Maximum test, CCE estimator

1 Introduction

The following two-way fixed effects (TFE) regression has been the most popularly used panel model.

$$y_{it} = a_i + \beta' x_{it} + F_t + \epsilon_{it}, \tag{1}$$

where a_i is an individual fixed effect for $i = 1, ..., n$, and F_t is a common shock to all individuals at time $t = 1, .., T$, which is called a year or time fixed effect. If the common shock F_t, which can cause the cross-sectional dependence among y_{it}, influences on each individual differently, the two-way fixed effects regression is not good enough to isolate the effect from the common shock. In this case, the factor augmented regression is used instead.

$$y_{it} = a_i + \beta' x_{it} + \gamma_i' F_t + \varepsilon_{it}, \tag{2}$$

where F_t is no longer a single factor but can be a $(r \times 1)$ vector of latent common factors, and γ_i is a $(r \times 1)$ vector of factor loadings. More importantly, the $(1 \times k)$ vector of regressors x_{it} may be possibly sharing the same common factors. That is, the regressors can be modelled by

$$x_{it} = b_i + \Gamma_i F_t + \Psi_i G_t + x_{it}^o, \tag{3}$$

where G_t is a $(1 \times m)$ vector of other common factors, Γ_i is a $(k \times r)$ matrix of factor loadings, Ψ_i is a $(k \times m)$ matrix of factor loadings, and x_{it}^o is a pure idiosyncratic term.

When the factor loading coefficients in (2) are heterogeneous ($\gamma_i \neq \gamma$), the TFE estimator has the following two problems: First, when Γ_i is correlated with γ_i, the TFE estimator becomes inconsistent since x_{it} is correlated with ϵ_{it}. Second, even when Γ_i is not correlated with γ_i, a typical panel robust variance is no longer consistent due to the existence of the cross-sectional dependence. The solution is rather simple. Once including the common factors as additional regressors, one can exclude the source of cross-sectional correlation from the estimation. Under some regularity conditions, the latent common components $\gamma_i' F_t$ can be approximated as the linear combination among the sample cross-sectional averages of x_{it} and y_{it}. The so-called 'Commonly Correlated Effects' (CCE hereafter) estimator, which was simple and intuitive estimation method proposed by Pesaran (2006), has been popularly used in practice and STATA commands are available online. Along with the CCE estimator, empirical researchers also have used 'Iterative Effect' (henceafter IE) developed by Bia (2009). The IE estimator approximates the latent common factors to regression errors by using Principal Component (PC) estimation. See Reese and Westerlund (2018) and Hayakawa, Nagata and Yamagata (2018) for more recent reference, and Chudik and Pesaran (2015) for a recent survey on this literature.

Even though there are many good estimators available and more applied researchers have considered panel factor augmented estimators as the alternative of the TFE estimator, in practice empirical researchers have still used two-way fixed effects estimation. Naturally, to encourage empirical researchers to use panel factor augmented estimators, pre-testing procedures become desirable.

The purpose of this paper is to provide '*asymptotic*' comparison among pre-testing procedures. There are broadly three types pre-tests available in the literature. The first two types are proposed by Bai (2009): Hausman type test and the use of the number of common factors. The Hausman type test examines whether or not panel factor augmented estimators share the same probability limit of the TFE estimators. A pre-test with a fixed T is considered by Westerlund (2019). However as CRT (2015a) point out, the Hausman type test may fail when Γ_i is not correlated with γ_i. In this case, the TFE shares the same probability limit with the CCE or IE estimator. The second type method is the use of the number of common factors. As Bai (2009) and Parker and Sul (2016) point out, the TFE residual does not include any significant factors if $\gamma_i = \gamma$ for all i since the time fixed effects successfully eliminate unknown common factors. Throughout the paper, we will call this method 'BPS'. The last method is a direct testing proposed by CRT (2015b). The CRT method tests whether or not the maximum of estimated $\hat{\gamma}_i$ is significantly different from sample cross-sectional averages.

We starts to review three types of pre-tests by comparing pros and cons of each pre-test. Among them, the direct testing method suggest by CRT (2015b) is the most powerful test in the sense that even when $\gamma_i = \gamma$ for all i except for just one unit, the direct testing method detects this deviation since this CRT method is based on the maximum value of the estimated $\hat{\gamma}_i$. We introduce this local heterogeneity more formally to evaluate these tests asymptotically. Interestingly, we find that BPS method always leads to more efficient (in terms of asymptotic variance) estimation.

Even though it is not directly related to this literature, the literature of testing cross-sectional dependence is indirectly related. See Pesaran (2004, 2015), Ng (2006), Pesaran, Ullah and Yamagata (2008), Sarafidis, Yamagata and Robertson (2009), Baltagi, Feng and Kao (2011), Sarafidis and Wansbeek (2012) and Baltagi, Kao and Na (2013) for recent references.

The rest of the paper is organized as follows. Section 2 provides a short review and the notion of local heterogeneity of factor loadings. We also provide a formal pre-testing procedure for BPS method. Asymptotic results under homogeneity and heterogeneity of slope coefficents are discussed in Section 3. Key theorems and some important remarks are provided. Section 4 includes Monte Carlo results and one empirical example. Section 5 concludes. All technical proofs are in Appendix.

2 Extant Pre-Testing Procedures

This section provides a short review on extant pre-testing procedures for panel factor augmented regressions first, and then discusses how to evaluate each pre-testing procedure, next. We start this section by asking the following question: Should a panel factor augmented regression run when regression errors are cross-sectionally dependent? If so, all tests for cross-sectional dependence can be treated as pre-testing for a panel factor augmented regression. Suppose that regression errors are spatially correlated. Then a good test statistic for testing the cross-sectional dependence must detect this spatial correlated, but it does not automatically imply that the factor augmented regression is the right one to run. In the sense, we consider the following three pre-testing procedures in this section.

2.1 Hausman-type Test

Here we assume the true data generating process of y_{it} is given by

$$y_{it} = a_i + \beta' x_{it} + u_{it}, \text{ with } u_{it} = \gamma_i' F_t + \varepsilon_{it}. \tag{4}$$

If γ_i is correlated with Γ_i, which is the vector of factor loadings in regressors in (3), then the regressors, x_{it}, are correlated with the regression error, u_{it} even when both γ_i and Γ_i have zero means. Define $\tilde{y}_{it}$ as the deviation from its time series mean. Further define $\dot{y}_{it}$ and $\dot{x}_{it}$ as

$$\dot{y}_{it} = \tilde{y}_{it} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \tilde{y}_{it}, . \quad \dot{x}_{it} = \tilde{x}_{it} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \tilde{x}_{it}.$$

That is, $\dot{y}_{it} = y_{it} - T^{-1} \sum_{t=1}^{T} y_{it} - n^{-1} \sum_{i=1}^{n} y_{it} + n^{-1} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{it}$. Then the two-way fixed effects regression can be rewritten as

$$\dot{y}_{it} = \beta' \dot{x}_{it} + \dot{u}_{it} \text{ with } \dot{u}_{it} = \left(\gamma_i - \frac{1}{n} \sum\nolimits_{i=1}^{n} \gamma_i \right) \tilde{F}_t + \dot{\varepsilon}_{it}. \tag{5}$$

If $\gamma_i \neq \gamma$, then the TFE estimator in (5) becomes inconsistent since $\dot{x}_{it}$ is still correlated with $\dot{u}_{it}$. Meanwhile either the CCE or the IE estimator is consistent. Bai (2009) points out this difference, and proposes a Hausman-type test to detect whether or not $\gamma_i = \gamma$. Westerlund (2019) extends this test to the case where the time series observation is small.

However, this test is not airtight in the sense that the TFE can be consistent even when $\gamma_i \neq \gamma$. If γ_i is not correlated with Γ_i, or simply regressors do not have any common factors, then both TFE and factor augmented estimators are consistent. CRT (2015a) point out this issue, and formally show that the Hausman-type test for testing $\gamma_i = \gamma$ is not consistent asymptotically. Therefore, we do not consider this test in this paper.

2.2 CRT Test

The next test is the maximum value test proposed by CRT (2015b). Interestingly, CRT (2015b) consider only the case of heterogeneous slope coefficients. Instead of (2), CRT consider the following factor augmented regression.

$$y_{it} = a_i + \beta_i' x_{it} + \gamma_i' F_t + \varepsilon_{it} \tag{6}$$

It is important to note that the CRT test becomes valid only without imposing the homogeneity restriction of β_i. The reason is straightforward. When $\beta_i = \beta$ for all i, it does not matter whether or not one imposes the homogeneity of β_i in (6). A serious problem eixsts when $\beta_i \neq \beta$. Assume we impose the homogeneity restriction under $\beta_i \neq \beta$ for some i. Then the factor augmented regression in (6) becomes

$$y_{it} = a_i + \beta' x_{it} + u_{it}, \text{ with } u_{it} = (\beta_i - \beta)' x_{it} + \gamma_i' F_t + \varepsilon_{it}. \tag{7}$$

Substituting (3) into (7) results in

$$u_{it} = \left[(\beta_i - \beta)' \Gamma_i + \gamma_i' \right] F_t + (\beta_i - \beta)' \Lambda_i G_t + (\beta_i - \beta)' x_{it}^o + \varepsilon_{it}. \tag{8}$$

Even when $\gamma_i = \gamma$, the regression error, u_{it}, includes the heterogeneous factor loadings of Γ_i and Λ_i since $\beta_i \neq \beta$. Hence for the consistency of the test, the homogeneity restriction should not be imposed in (6).

Here we provide a step by step procedure for CRT test for the null hypothesis of[1]

$$\mathcal{H}_0 : \gamma_i = \gamma. \tag{9}$$

Step 1 Obtain the regression residuals $\hat{u}_{it} = y_{it} - \hat{a}_i - \hat{\beta}_i' x_{it}$ from the following regression. That is, run the CCE estimation for each i.

$$y_{it} = a_i + \beta_i' x_{it} + \delta_{x,i}' \bar{x}_t + \delta_{y,i} \bar{y}_t + \epsilon_{it}, \tag{10}$$

where $\bar{x}_t = n^{-1} \sum_{i=1}^n x_{it}$ and $\bar{y}_t = n^{-1} \sum_{i=1}^n y_{it}$.

Step 2 Assume a single factor, and estimate γ_i by using PC analysis. Let $\hat{\gamma}_i$ be the PC estimator. Then construct the following Mahalanobis distance.[2]

$$\mathcal{O}_i = \left(\hat{\gamma}_i - \hat{\mu}_\gamma \right)^2 / \hat{\Sigma}_\gamma, \tag{11}$$

where $\hat{\mu}_r$ and $\hat{\Sigma}_\gamma$ are sample mean and variance of $\hat{\gamma}_i$. That is,

$$\hat{\mu}_r = \frac{1}{n} \sum\nolimits_{i=1}^n \hat{\gamma}_i, \text{ and } \hat{\Sigma}_\gamma = \frac{1}{n-1} \sum\nolimits_{i=1}^n \left(\hat{\gamma}_i - \hat{\mu}_r \right)^2.$$

Step 3 Construct the following max-type test given by

$$\mathcal{S}_{\gamma,nT} = T \cdot \max_{1 \le i \le n} \left[\mathcal{O}_i \right]. \tag{12}$$

[1] Note that CRT (2015b) also propose a pre-test for $F_t = F$ for all t. The procedure is exactly identical, but here we do not consider this test joinly since in practice, the null hypothesis of $\gamma_i = \gamma$ becomes of interest.

[2] Mahalanobis distance is a well known statistic to measure the degree of outlyingness. As $\hat{\gamma}_i$ departs further from it's center or central location, the outlyingness approaches to infinity. There are many statistical outlyingness functions available. See Zuo and Serfling (2000) for more discussions.

CRT (2015b) shows that $\mathcal{S}_{\gamma,nT}$ has a Gumbel distribution. The critical value, $c_{\alpha n}$, can be calculated by

$$c_{\alpha n} = 2\ln n - \ln\ln n - 2\ln\Gamma\left(1/2\right) - \ln\left|\ln\left(1-\alpha\right)\right|^2, \tag{13}$$

where $\Gamma(\cdot)$ is a gamma function, and α is the significance level.

2.3 BPS Procedure

Both Bai (2009) and Parker and Sul (2016) use the estimated number of common factors to evaluate the homogeneous factor loadings. This method can be used for a single panel data or panel regressions with multiple regressors. Consider a single panel data case first. Suppose that a panel data w_{it} follows a single factor structure given in

$$w_{it} = a_i + \gamma_i F_t + w_{it}^o, \tag{14}$$

where w_{it}^o is a pure idiosyncratic term. Then observe this.

$$\dot{w}_{it} = \left(\gamma_i - \frac{1}{n}\sum\nolimits_{i=1}^n \gamma_i\right)\left(F_t - \frac{1}{T}\sum\nolimits_{t=1}^T F_t\right) + \dot{w}_{it}^o. \tag{15}$$

The homogeneity of γ_i leads to $\dot{w}_{it} = \dot{w}_{it}^o$. Define $\#\left(w_{it}\right)$ and $\hat{\#}\left(w_{it}\right)$ as the true and the estimated number of common factors to w_{it}, respectively. Then it becomes obvious that

$$\#\left(w_{it}\right) = 1 \;\&\; \#\left(\dot{w}_{it}\right) = 0. \tag{16}$$

Hence following to Bai and Ng (2002), as $n, T \to \infty$,

$$\Pr\left[\hat{\#}\left(w_{it}\right) = 1\right] = 1 \quad \& \quad \Pr\left[\hat{\#}\left(\dot{w}_{it}\right) = 0\right] = 1, \tag{17}$$

with a proper information criterion. It is worth noting that this method does not require to estimate γ_i.

In a regression setting, this method can be easily implemented as well. Here we propose the following two-step procedure.

Step 1 Run the following two-way fixed effects regression with the homogeneity restriction on β_i.

$$\dot{y}_{it} = \beta'\dot{x}_{it} + u_{it}. \tag{18}$$

Get the residuals, $\hat{u}_{it} = \dot{y}_{it} - \hat{\beta}_{\text{tfe}}'\dot{x}_{it}$, where $\hat{\beta}_{\text{tfe}}$ is the LS estimator in (18).

Step 2 Use Bai and Ng's (2002, BN hereafter) IC_2 criterion to estimate the number of common factors with $\hat{u}_{it}$.

From (8), it is easy to show only with $\gamma_i = \gamma$ and $\beta_i = \beta$ for all i, the number of common factors with u_{it} becomes zero.

$$\#(u_{it}) = 0 \text{ if } \gamma_i = \gamma \text{ \& } \beta_i = \beta. \tag{19}$$

Otherwise, the true number of common factors with u_{it} becomes a non-zero constant. That is,

$$\#(u_{it}) \geq 1 \text{ if either } \gamma_i \neq \gamma \text{ or } \beta_i \neq \beta. \tag{20}$$

It is because BPS method is not directly testing the null of $\gamma_i = \gamma$, but just focusing on whether or not u_{it} has a factor structure. If $\hat{\#}(\hat{u}_{it}) > 0$, then the following CCE type regression should be run.

$$y_{it} = \begin{cases} a_i + \beta_i' x_{it} + \delta_{x,i}' \bar{x}_t + \delta_{y,i} \bar{y}_t + \epsilon_{it} & \text{for CCE MG}, \\ a_i + \beta' x_{it} + \delta_{x,i}' \bar{x}_t + \delta_{y,i} \bar{y}_t + \epsilon_{it} & \text{for CCEP}. \end{cases} \tag{21}$$

If $\hat{\#}(\hat{u}_{it}) = 0$, then it implies that both $\beta_i = \beta$ and $\gamma_i = \gamma$ asymptotically. Hence in this case, the two-way fixed regression in (18) or (1) should be run for the pooled estimator. For the MG estimation, one can run the following regression.

$$\dot{y}_{it} = \beta_i' \dot{x}_{it} + \epsilon_{it}, \tag{22}$$

2.4 Summary and Resulting Estimators

We consider the following two cases separately: Pooled and MG estimation. Except for a few, almost all of empirical studies have considered pooled estimation. Consider the following two choices as we discussed in Introduction.

$$\text{Pooled Case:} \quad y_{it} = \begin{cases} a_i + \beta' x_{it} + F_t + \epsilon_{it} \\ a_i + \beta' x_{it} + \delta_{x,i}' \bar{x}_t + \delta_{y,i} \bar{y}_t + \epsilon_{it} \end{cases} \tag{23}$$

Theoretically, however, there is no particular reason to avoid the MG estimation. Alternatively, researchers may be interested in individual estimator for the slope coefficient. In this case, the following two choices are considered.

$$\text{MG Case:} \quad \begin{cases} a_i + \beta_i' x_{it} + F_t + \epsilon_{it} \\ a_i + \beta_i' x_{it} + \delta_{x,i}' \bar{x}_t + \delta_{y,i} \bar{y}_t + \epsilon_{it} \end{cases} \tag{24}$$

The first and second regressions for each case yields two-way fixed effects (TFE) and CCE estimators, respectively. Let $\hat{\beta}_{\text{tfe},i}$ be the LS estimator in the first regression, and $\hat{\beta}_{\text{cce},i}$ be the LS estimator in the second regression in (24). Then the TFE MG and CCE MG estimators can be constructed by taking the sample cross-sectional averages of $\hat{\beta}_{\text{tfe},i}$ and $\hat{\beta}_{\text{cce},i}$, respectively.

Table 1 shows the results of the BPS and the CRT methods under four different conditions. Since the BPS method imposes the homogeneity restriction on the slope coefficients, meanwhile the CRT method puts the heterogeneity restriction, the pre-testing results do not alter whether or not empirical researchers are interested either in pooled or MG estimation. There are two differences between the BPS and the CRT methods. Table 1 shows the first difference between the two pre-tests. When either $\beta_i \neq \beta$ or $\gamma_i \neq \gamma$, the BPS method always recommends the CCE estimator. Meanwhile the CRT method precisely differentiates the heterogeneous γ_i from the case of the homogeneous factor loadings. Hence the first difference between the two pre-tests happens when $\beta_i \neq \beta$ but $\gamma_i = \gamma$. If empirical researchers are interested in pooling regressions, then the BPS method provides a 'correct' answer in this case since the regression error, u_{it}, in (18) includes more than a single factor as it is shown in (20). When the MG estimation becomes of interest, the situation becomes converted. The CRT method assists a 'correct' guide under the case of $\beta_i \neq \beta$. However, it does not imply that the TFE estimator in the case of $\beta_i \neq \beta$ and $\gamma_i = \gamma$ is more efficient than the CCE-MG estimator. We will investigate this case asymptotically in the next section.

The second difference between the two pre-tests is not shown in Table 1. Precisely speaking, the BPS method is not a test, but just an identification procedure since the BPS method utilizes BN's IC_2 criterion. As $n, T \to \infty$, the probability to select the correct number of common factors becomes unity. Meanwhile the CRT method is a well constructed test, so that it makes a mistake with α times, where α is the significance level. This difference is minor, but in Monte Carlo simulation, this difference matters slightly.

Table 1: Pre-Testing Results Under Various DGPs

Conditions	BPS	CRT
$\beta_i = \beta$ & $\gamma_i = \gamma$	TFE	TFE
$\beta_i = \beta$ & $\gamma_i \neq \gamma$	CCE	CCE
$\beta_i \neq \beta$ & $\gamma_i = \gamma$	CCE	TFE
$\beta_i \neq \beta$ & $\gamma_i \neq \gamma$	CCE	CCE

In the next section, we will provide asymptotic comparison between the two pre-tests.

3 Asymptotic Comparison

We first consider the case of $\beta_i = \beta$ for all i. In the next subsection, we consider the heterogeneity of β_i. As we discussed in the previous section, the results of the asymptotic comparison are hinging on the assumption of the slope coefficients. Since it is unknown whether or not $\beta_i = \beta$, the overall comparison will be made in the end of this section.

We take the following assumptions.

Assumption 1 (Common factors) *The unobserved common factors, F_t and G_t, are $(r \times 1)$ and $(m \times 1)$ covariance stationary, with absolute summable autocovariance, distributed independently of ε_{it} and x_{is}^o for all i, t and s.*

Assumption 2 (Individual-Specific Error)

(i) $\mathrm{E}\Big(\varepsilon_{it} x_{js}^o\Big) = 0$, for all i, j, s and t.

(ii) For each i, x_{it}^o follows linear stationary processes with absolute summable autocovariance.

Assumption 3 (Factor Loadings) *The unobserved factor loadings, γ_i, Γ_i and Ψ_i, are independently and identically distributed across i, and of individual specific errors ε_{jt} and x_{jt}^o, the common factors, F_t and G_t for all i, j and t with fixed means γ, Γ and Ψ, respectively, and finite variances.*

Assumption 4 (Serial and Cross-Sectional Weak Dependence and Heteroskedasticity)

(i) $\mathbb{E}(\varepsilon_{it}) = 0$ *and* $\mathbb{E}\left|\varepsilon_{it}\right|^{12} \leq M$.

(ii) $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ *for all* t, s *and* $|\sigma_{ij,ts}| \leq \tau_{ts}$ *for all* i, j *such that*

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{\sigma}_{ij} \leq M,\ \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts} \leq M,\ \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}|\sigma_{ij,ts}| \leq M.$$

(iii) *For every* t *and* s, $\mathbb{E}\left|n^{-1/2}\sum_{i=1}^{n}\left[\varepsilon_{it}\varepsilon_{js} - \mathbb{E}(\varepsilon_{it}\varepsilon_{js})\right]\right|^4 \leq M$.

(iv) *Moreover*

$$T^{-2}n^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\mathrm{cov}\left(\varepsilon_{it}\varepsilon_{js}, \varepsilon_{jp}\varepsilon_{jq}\right)\right| \leq M,$$

$$T^{-1}n^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\left|\mathrm{cov}\left(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ls}\varepsilon_{ms}\right)\right| \leq M.$$

Assumption 5 (Rank Condition) *The total number of common factors in the regression error, u_{it}, is less than or equal to $k+1$, where k is the number of regressors.*

Assumption 6 (Homogeneous Slope Coefficients) *Under homogeneity,*

$$\beta_i = \beta,$$

where $\|\beta\| < M$.

Assumptions 1 and 2 allow for serial and cross-sectional dependences in both common factors and individual-specific errors. Assumption 3 entails the factor loadings, with non-zero fixed means, to be strong in the sense of Chudik, Pesaran and Tosetti (2011). Assumptions 1 through 3 are fairly general since the case in which the error components might be correlated with the regressor x_{it} are not excluded. Assumption 4 allows weak serial and cross-sectional correlation for ε_{it}. Assumption 6 restricts β_i to be homogeneous.

We define the two pooled estimators:

$$\hat{\beta}_{\text{tfe,p}} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{y}_{it}\right), \tag{25}$$

$$\hat{\beta}_{\text{cce,p}} = \left(\sum_{i=1}^{n}X_i'M_zX_i\right)^{-1}\left(\sum_{i=1}^{n}X_i'M_zY_i\right), \tag{26}$$

where $X_i = [x_{i1},\cdots,x_{iT}]'$, $Y_i = [y_{i1},\cdots,y_{iT}]'$, $z_{it} = [y_{it},\ x_{it}]'$, $M_z = I_T - \bar{Z}(\bar{Z}'\bar{Z})^{-}\bar{Z}'$, $\bar{Z} = [\bar{z}_1,\cdots,\bar{z}_T]'$, and $\bar{z}_1 = n^{-1}\sum_{i=1}^{n}z_{i1}$. Note that $(\bar{Z}'\bar{Z})^{-}$ is the generalized inverse of $\bar{Z}'\bar{Z}$.

The MG estimators are defined as

$$\hat{\beta}_{\text{tfe,mg}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{\text{tfe},i} \text{ with } \hat{\beta}_{\text{tfe},i} = \left(\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\left(\sum_{t=1}^{T}\dot{x}_{it}\dot{y}_{it}\right), \tag{27}$$

$$\hat{\beta}_{\text{cce,mg}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{\text{cce},i} \text{ with } \hat{\beta}_{\text{cce},i} = \left(X_i'M_zX_i\right)^{-1}\left(X_i'M_zY_i\right). \tag{28}$$

3.1 Under Homogeneity of Slope Coefficients

When $\beta_i = \beta$, both the BPS and the CRT methods provide the same answer as Table 1 showed. To differentiate the outcomes of the two pre-tests under $\beta_i = \beta$, we introduce the notion of a local heterogeneity of factor loadings.

Definition 1 (Local-Heterogeneity of γ_i) *The $(r \times 1)$ factor loading vector γ_i is locally-heterogeneous such that*

$$\gamma_i = \gamma + \tau_i, \qquad \tau_i \sim iid(0, \Omega_{\epsilon,i}) \tag{29}$$

where

$$\Omega_{\epsilon,i} = \begin{cases} 0 \text{ or } \tau_i = 0 & \text{if } i \in \mathcal{G} \\ \Omega_0 \text{ or } \tau_i \neq 0 & \text{if } i \in \mathcal{G}^c \end{cases} \tag{30}$$

where the number of individuals in $\mathcal{G}^c$ is fixed υ, which is not dependent on n.

Here we consider a case where $\gamma_i \neq \gamma$ for a few individuals. The local heterogeneity implies the weak factor in Chudik and Pesaran's (2011) sense if $\gamma = 0$. Note that

$$\mathbb{E}\frac{1}{n}\sum_{i=1}^{n}(\gamma_i - \gamma)^2 = \Omega_0/n. \tag{31}$$

The condition in (31) states that the common factors F_t are weak factors if $\gamma = 0$. Meanwhile the weak factors do not imply the local heterogeneity. For example, Reese and Westerlund (2015) consider the following notion of the weak factors when $\gamma = 0$.

$$\tau_i = \tau_i^o / n^\alpha \text{ with } \alpha \in [0,1] \text{ and } \epsilon_i^o = O_p(1). \tag{32}$$

Under (32), as $n \to \infty$, the maximum of γ_i also converges to zero (or $\gamma = 0$). In this case, CRT's (2015b) max-type test fails.[3] Because of the same reason, CRT (2015b) assume no weak factor given in (32).

Next, we will study the asymptotic behavior of the BPS and the CRT pre-testing methods under the local heterogeneity.

Theorem 1: (Consistency of Tests for Local Heterogeneity of Factor Loadings) *Under local heterogeneity of γ_i in (30) and Assumption 1-6,*

(i) as $n, T \to \infty$ with $T/n \to 0$,

$$\lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 1, \text{ and} \tag{33}$$

(ii) as $n, T \to \infty$ with $T/n^{5/3} \to 0$ and $n/T^3 \to 0$,

$$\lim_{n,T\to\infty} \Pr(\mathcal{S}_{\gamma,nT} > c_{\alpha,n}) = 1 \tag{34}$$

The technical proof of Theorem 1 is in Appendix A. Here we provide an intuitive explanation. As Parker and Sul (2016) showed, Bai and Ng's (2002) information criteria (IC) are not precise enough to detect weak factors. Under the local heterogeneity, the remaining common factor in the regression residuals, $\hat{u}_{it} = \dot{y}_{it} - \hat{\beta}'_{\text{tfe,p}} \dot{x}_{it}$, contain only a weak factor. Since only a few individuals are influenced by the common factor, Bai and Ng's IC cannot detect the presence of the common factor even with very large n and T. Interestingly when the factor loadings to the regression error, γ_i, is correlated with the factor loadings to the regressors, Γ_i, the TFE estimator has the bias, which is an $O_p(1/n)$ term under the local heterogeneity. Due to this bias, we need the condition of $T/n \to 0$. Meanwhile the CRT test is based on the maximum value of Mahalanobis distances. The maximum value is, of course, very sensative to non-zero ϵ_i in (30). Hence the CRT method detects the local heterogeneity very precisely as $n, T \to \infty$.

Next, we provide an important remark regarding dynamic panel regressions.

[3] To see this, assume that $\gamma_i = \gamma + \epsilon_i$, with $\epsilon_i = O_p\left(n^{-1/2}\right)$. Then as $n, T \to \infty$ with $T/n \to 0$, the following condition becomes

$$\frac{T}{\ln n} \|\boldsymbol{\epsilon}_i\|^2 = \frac{T}{n \ln n} O_p(1) \to 0,$$

which implies the failure of Theorem 3 in CRT (2015b).

Remark 1 (Dynamic Panel Regression): A latent model can be written as following:

$$y_{it} = a_i + \rho y_{it-1} + \lambda_i' F_t + \varepsilon_{it}, \tag{35}$$

or

$$y_{it} = a_i (1 - \rho L)^{-1} + \lambda_i' F_t (1 - \rho L)^{-1} + \varepsilon_{it} (1 - \rho L)^{-1},$$

where L is a lag operator. Let $\hat{\rho}_{\text{fe}}$ be the one-way fixed effect or within group (WG) estimator. From the direct calculation, as long as the pooled estimator is used, we can show that

$$\tilde{y}_{it} - \hat{\rho}_{\text{fe}} \tilde{y}_{it-1} = \tilde{\varepsilon}_{it} + \lambda_i' \tilde{F}_t + (\rho - \hat{\rho}_{\text{fe}}) \tilde{y}_{it-1} = \tilde{\varepsilon}_{it}^* + \lambda_i' \tilde{F}_t^*, \tag{36}$$

where $\tilde{F}_t^* = \tilde{F}_t + (\rho - \hat{\rho}_{\text{fe}}) \sum_{j=0}^{\infty} \rho^j \tilde{F}_{t-1-j}$, and $\tilde{\varepsilon}_{it}^* = \tilde{\varepsilon}_{it} + (\rho - \hat{\rho}_{\text{fe}}) \sum_{j=0}^{\infty} \rho^j \tilde{\varepsilon}_{it-1-j}$. Therefore the number of common factors is not influenced by the WG estimation. Hence Theorem 1 holds continuously.

Under the suitable conditions, both the TFE and the CCE estimators are consistent, and asymptotically the modified regressors are independent of the modified regression errors. The main difference between the two variances comes from the asymptotic covariance of the modified regressors. Interestingly, the CCE estimation cleans up the common components of the regressors effectively by projecting out the cross-sectional averages of y_{it} and x_{it}. The covariance matrix with the remained terms becomes asymptotically equivalent to the covariance matrix with the idiosyncratic terms of x_{it}. Meanwhile the TFE estimation does not effectively eliminate the common components of x_{it} if the factor loadings to x_{it} are strongly heterogeneous, which results in larger covariance matrix of the modified regressors. This difference makes that the TFE estimator becomes more efficient than the CCE estimator in general. Only when the two-way within group transformation cleans up the common components of x_{it} effectively, the relative variance ratio becomes unity asymptotically. To be specific, we derive the asymptotic variances of the CCE and the TFE estimators under the local heterogeneity. Since both estimators are consistent under the local heterogeneity, it is not hard to show that the asymptotic variance ratio of the CCE to the TFE residuals becomes unity as $n, T \to \infty$ with $T/n \to 0$. Meanwhile the probability limits of the denominator terms for the CCE is smaller than that of TFE estimators in general. Let $\dot{X}_i = [\dot{x}_{i1}, \cdots, \dot{x}_{iT}]'$, $X_i^o = [x_{i1}^o, \cdots, x_{iT}^o]'$. Further define $\Omega_{\text{cce,p}}$ and $\Omega_{\text{tfe,p}}$ as

$$\Omega_{\text{cce,p}} = \mathbb{E} \frac{1}{nT} \sum\nolimits_{i=1}^{n} X_i^{o\prime} \Omega_{\varepsilon,i} X_i^o, \ \& \ \Omega_{\text{tfe,p}} = \mathbb{E} \frac{1}{nT} \sum\nolimits_{i=1}^{n} \dot{X}_i' \Omega_{\varepsilon,i} \dot{X}_i, \tag{37}$$

and

$$Q_{\text{cce,p}} = \text{plim}_{n,T\to\infty} \frac{1}{nT} \sum\nolimits_{i=1}^{n} X_i^{o\prime} X_i^o, \ \& \ Q_{\text{tfe,p}} = \text{plim}_{n,T\to\infty} \frac{1}{nT} \sum\nolimits_{i=1}^{n} \dot{X}_i' \dot{X}_i. \tag{38}$$

Theorem 2 (Comparison of Asymptotic Variances) *Under Assumption 1-6, as $n, T \to \infty$ with $T/n \to 0$,*

(i) *the asymptotic variances of TFE and CCE pooled estimators are given by*

$$V_{\ell,\mathrm{p}} = Q_{\ell,\mathrm{p}}^{-1} \Omega_{\ell,\mathrm{p}} Q_{\ell,\mathrm{p}}^{-1}, \tag{39}$$

where $\ell \in \{\text{cce, tfe}\}$.

(ii) *Under i.i.d. assumption of ε_{it} over i and t, the relative variance ratio satisfies*

$$V_{\mathrm{cce,p}} \geq V_{\mathrm{tfe,p}} \tag{40}$$

The proof of Theorem 2 is straightforward, hence it is omitted. Note that Pesaran (2006) already showed the asymptotic variance of the CCE pooled estimator. Here we merely changed the weight function in Pesaran's (2006) case. The result for the asymptotic variance of the TFE estimator may be new, but nothing special. When ε_{it} is not i.i.d over t, it is still easy to show that Theorem 2 (ii) holds. Note that $Q_{\mathrm{cce,p}} \leq Q_{\mathrm{tfe,p}}$ always as long as $\Gamma_i \neq \Gamma \neq 0$. Also it is easy to show that $\Omega_{\mathrm{cce,p}} \leq \Omega_{\mathrm{tfe,p}}$ since $\dot{X}_i = \dot{X}_i^o + \tilde{\Gamma}_i \tilde{F}_t + \tilde{\Psi}_i \tilde{G}_t$ so that $\mathbb{E}\dot{X}_i \varepsilon_i \varepsilon_i' \dot{X}_i \leq \mathbb{E}\dot{X}_i^o \varepsilon_i \varepsilon_i' \dot{X}_i^o = \mathbb{E}X_i^o \varepsilon_i \varepsilon_i' X_i^o$ as $n, T \to \infty$. When ε_{it} is not i.i.d. over i, it is not easy to show that Theorem 2 (ii) holds unless we know the weak dependence structure. The equality holds only when $\tilde{\Gamma}_i = \tilde{\Psi}_i = 0$ for all i.

Next, we provide a couple of important remarks.

Remark 2 (Requirement of the T/n ratio) There are some special cases where the CCE pooled estimator becomes inconsistent. See Westerlund and Urbain (2014) for more detailed discussions even under the presence of strong factors. For more detailed conditions under strong factors, see Westerlund and Urbain (2018). Meanwhile see Reese and Westerlund (2015) for the case of weak factors. When either the rank condition or the T/n ratio does not hold, the CCE estimator becomes seriously biased. In practice, it is not known whether or not $T/n \to 0$ since both T and n is fixed. If $n < T$ or $n \simeq T$, then using lower time frequency data, for example annual data rather than month data, leads to decrease the number of time series observations, but keep the entire time length.

Remark 3 (Asymptotic variance of the MG estimators) It is well known that the pooled estimator can be re-written as

$$\hat{\beta}_{\ell,\mathrm{p}} = \frac{\sum_{i=1}^n W_{\ell,i} \hat{\beta}_{\ell,i}}{\sum_{i=1}^n W_{\ell,i}} \text{ with } \ell \in \{\text{cce,tfe}\}, \tag{41}$$

where the weight function $W_{\ell,i}$ is given by

$$W_{\ell,i} = \begin{cases} T^{-1} X_i^{o\prime} X_i^o & \text{if } \ell = \text{cce} \\ T^{-1} \dot{X}_i' \dot{X}_i & \text{if } \ell = \text{tfe} \end{cases} . \tag{42}$$

When $\beta_i = \beta$ for all i, it is easy to show that the asymptotic variance of the CCE MG estimator is relatively larger than that of the TFE MG estimator under i.i.d assumption of ε_{it} over t and i.

We combine the result of Theorem 1 with Theorem 2 together. Define the pooled BPS estimator as

$$\hat{\beta}_{\text{BPS,p}} = \begin{cases} \hat{\beta}_{\text{tfe,p}} & \text{if } \hat{\#}(\hat{u}_{it}) = 0 \\ \hat{\beta}_{\text{cce,p}} & \text{o.w.} \end{cases} . \tag{43}$$

Alternatively, we can define the BPS and CRT estimators as

$$\hat{\beta}_{\text{m,p}} = \omega_{\text{m}} \hat{\beta}_{\text{tfe,p}} + (1 - \omega_{\text{m}}) \hat{\beta}_{\text{cce,p}}, \text{ with m} \in \{\text{BPS, CRT}\}$$

where $\omega_{BPS} = 1[\hat{\#}(\hat{u}_{it}) = 0]$ and $\omega_{CRT} = 1\left(S_{\gamma,NT} \le c_{\alpha,N}\right)$. Note that $1\left(\cdot\right)$ is an indicator function. The asymptotic variances of the BPS and CRT estimators can be written as

$$V(\hat{\beta}_{\text{m,p}}) = \omega_{\text{m}} V_{\text{tfe,p}} + (1 - \omega_{\text{m}}) V_{\text{cce,p}}.$$

Similarly we can define the MG BPS and CRT estimators as follow.

$$\hat{\beta}_{\text{m,mg}} = \omega_{\text{m}} \hat{\beta}_{\text{tfe,mg}} + (1 - \omega_{\text{m}}) \hat{\beta}_{\text{cce,mg}}, \text{ with m} \in \{\text{BPS, CRT}\}.$$

Note that the indicator function is not dependent on the choice of the MG or pooled estimation. Now, we are ready to propose the following Theorem.

Theorem 3 (Asymptotic Comparison under Homogeneous Slope Coefficients) *Under Assumptions 1-6,* as $n, T \to \infty$ *with* $T/n \to 0$ and $n/T^3 \to 0$,

$$V_{\text{CRT,p}} \ge V_{\text{BPS,p}}, \quad \& \quad V_{\text{CRT,mg}} \ge V_{\text{BPS,mg}}. \tag{44}$$

Note that Theorem 3 holds when $\beta_i = \beta$ by Assumption 6. The equality holds if $\omega_{BPS} = \omega_{CRT}$. Note that there are two cases that the equality holds always. The first case is when regressors have same or zero factor loadings ($\tilde{\Gamma}_i = \tilde{\Psi}_i = 0$ for all i). In this case, the BPS estimator becomes equivalent to the CRT estimator. The second case is when $\gamma_i \neq \gamma$ for all i. In this case, the equality holds since the power of the CRT test is always unity. Meanwhile under the null of $\gamma_i = \gamma$, the variance of the CRT estimator is always greater than that of the BPS estimator since $\omega_{CRT} = 1$ with $\alpha\%$ times. Lastly, under the local heterogeneity such that $\gamma_i \neq \gamma$ for some i, the inequality holds since asymptotically ω_{CRT} converges to unity, but ω_{BPS} converges to zero.

Next, we consider the case where $\beta_i \neq \beta$.

3.2 Under Heterogeneity of Slope Coefficients

Since the slope coefficients are heterogeneous, we need to change Assumption 6 to 6A.

Assumption 6A (Heterogeneous Slope Coefficients) *Under heterogeneity,*

$$\beta_i = \beta + \eta_i, \text{ with } \eta_i \sim iid\left(0, \Omega_\eta\right), \tag{45}$$

where $\|\beta\| < M$, $\|\Omega_\eta\| < M$, Ω_η is a $k \times k$ symmetric nonnegative definite matrix, and the random deviations η_i are distributed independently of γ_j, Γ_i, Ψ_i, ε_{jt}, v_{jt} for all i and j.

Assumption 6A is the standard assumption for the heterogeneous slope coefficients. Note that we need particularly the independence between η_i and the second central moments of regressors, which can be interpreted as the independence between η_i and Γ_i or Ψ_i. Otherwise, any pooled estimator leads to inconsistency due to the correlation between weights and η_i in (41).

It is very important to note that even when $\gamma_i = \gamma$ for all i, the regression error has heterogeneous factor loading coefficients under the presence of $\beta_i \neq \beta$ for all i. To see this, assume $\gamma_i = \gamma = 1$ and rewrite the true data generating process in (1) as

$$y_{it} = a_i + \beta_i' x_{it} + F_t + \varepsilon_{it}. \tag{46}$$

Imposing the homogeneity restriction on the slope coefficients leads to

$$y_{it} = a_i + \beta' x_{it} + F_t + e_{it}, \text{ with } e_{it} = \eta_i' \Gamma_i F_t + \eta_i' \Psi_i G_t + \eta_i' x_{it}^o + \varepsilon_{it}. \tag{47}$$

Only when $\Gamma_i = \Psi_i = 0$ for all i, the regression error, e_{it}, does not have any factor structure. Otherwise, the factor loadings with e_{it} are always heterogeneous under $\beta_i \neq \beta$ even when $\gamma_i = \gamma$.

Suppose that empirical researchers are interested only in pooled estimators. Then the CRT's procedure becomes invalid in this case since the CRT's max-type test examines only whether or not $\gamma_i = \gamma$. Assumption 6A does not allow any dependence between η_i and factor loadings of Γ_i or Ψ_i. If η_i is correlated with either of them, then the TFE estimator becomes inconsistent. In this case, the CRT method may leads to inconsistent estimation if Assumption 6A is violated. What if we impose the homogeneity restriction of β_i in (10)? Unfortunately, this restriction may solves this issue, but in this case the CRT's test does not examine the homogeneity of factor loadings of γ_i any more, which is the original purpose of the CRT's test. Interestingly, the BPS method works consistently under the heterogeneity of β_i. Since the BPS method is based on the estimated number of common factors from the regression residuals, this method always suggests to run the CCE regression as long as $\beta_i \neq \beta$, regardless of $\gamma_i = \gamma$. As shown in Table 1, this is the reason that the two methods suggest different solutions when $\beta_i \neq \beta$ but $\gamma_i = \gamma$.

The following lemma shows when the TFE estimator has a different limiting distribution from the CCE pooled estimator under $\beta_i \neq \beta$.

Lemma 1: (Asymptotic Difference between TFE and CCE Pooled Estimators under Heterogeneous Slope Coefficients) *Under Assumption 1-5 and 6A, If either* $\Gamma_i \neq 0$ *or* $\Psi_i \neq 0$, *but* $\gamma_i = \gamma$,

(i) *then as* $n, T \to \infty$ with $T/n \to 0$,

$$\left\| \hat{\beta}_{\text{cce,p}} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \beta_i \right\| = O_p\left(n^{-1/2} T^{-1/2}\right), \tag{48}$$

(ii) *as* $n, T \to \infty$,

$$\left\| \hat{\beta}_{\text{tfe,p}} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \beta_i \right\| = O_p\left(n^{-1/2}\right) + O_p\left(n^{-1/2} T^{-1/2}\right). \tag{49}$$

See Appendix C for the detailed proof of Lemma 1. Here we provides intuitive explanations about the asymptotic difference between the TFE and the CCE pooled estimators under $\beta_i \neq \beta$. Under cross-sectional independence or the case of $\Gamma_i = \Psi_i = \gamma_i = 0$, the within group estimator shares the same limiting distribution of the sample mean of β_i. When the rank condition is satisfies, the cross-sectional averages of y_{it} and x_{it} eliminates the common components effectively, which leads to the same conclusion under the conditions of $\Gamma_i \neq 0$ or $\Psi_i \neq 0$. Lemma 1 (i) shows that the CCE pooled estimator shares the same limiting distribution of the sample mean of β_i. Note that $\hat{\beta}_{\text{cce,p}} - \beta = O_p\left(n^{-1/2}\right)$, and $n^{-1} \sum_{i=1}^{n} \beta_i - \beta = O_p\left(n^{-1/2}\right)$. Meanwhile when $\Gamma_i \neq 0$ or $\Psi_i \neq 0$, but $\beta_i \neq \beta$, imposing the homogeneity restriction of β_i make the regression error in (47) included the common factors. Even though, the TFE estimator is consistent because of Assumption 6A or independence between η_i and the rest of factor loadings, but the TFE estimator is not efficient. The limiting distribution of the TFE estimator is much similar to the CCE pooled estimator when the rank condition is not satisified, since the new factor loadings $\eta_i' \Gamma_i$ and $\eta_i' \Psi_i$ are not correlated with Γ_i and Ψ_i. Lemma 1 (ii) in (49) reflects this fact, and shows that the TFE estimator is less efficient than the CCE pooled estimator. Of course, when both $\Gamma_i = \Psi_i = 0$ and $\gamma_i = \gamma$, then as $n, T \to \infty$

$$\left\| \hat{\beta}_{\text{tfe,p}} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \beta_i \right\| = \left\| \hat{\beta}_{\text{cce,p}} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \beta_i \right\| = O_p\left(n^{-1/2} T^{-1/2}\right). \tag{50}$$

Now we are ready to present the following Corollary.

Corollary 3-1: (Asymptotic Comparison for Pooled Estimators under Heterogeneous Slope Coefficients) *Under Assumptions 1-5 and 6A,* as $n, T \to \infty$ *with* $T/n \to 0$ and $n/T^3 \to 0$,

$$V_{\mathrm{CRT,p}} \geq V_{\mathrm{BPS,p}}. \tag{51}$$

The proof of Corollary 3-1 is straightforward, hence it is omitted. Table 1 shows that when $\gamma_i \neq \gamma$ but $\beta_i \neq \beta$, the BPS method always suggests the CCE, meanwhile the CRT method suggests the CCE estimator for $(1-\alpha)\%$ times and the TFE estimator for $\alpha\%$ times. Note that the TFE estimator in this case is still consistent since the correlation between common factors are washed out by taking off the cross-sectional averages under Assumption 6A. If Assumption 6A violates, that is, if either γ_i is correlated with Γ_i, or η_i is correlated with any of factor loadings, then the TFE estimator becomes inconsistent. However even under Assumption 6A, the TFE estimator is inefficient compared with the CCE estimator as Lemma 1 (1) shows. Hence in the case of $\gamma_i \neq \gamma$ but $\beta_i \neq \beta$, the BPS method becomes more effective. When $\gamma_i = \gamma$ but $\beta_i \neq \beta$, the CRT method always suggests the TFE estimation, but the BPS method always suggests the CCE estimation asymptotically. Hence from Lemma 1, the BPS method leads to more an efficient estimation.

Since the MG estimation is sometimes robust than the pooled estimation as Lee and Sul (2019) suggest, there is little theoretical justification to use a pooled estimation. In fact, the MG estimation is also one of ways to pool the cross and time series information as shown in (41). Only the difference between the MG and pooled estimators are weight functions: The MG estimation assigns an equal weight, $1/n$, meanwhile the pooled estimation assigns heavier weights if the variances of regressors are larger.

Next, we consider the case where the MG estimation becomes of interest to empirical researchers. Suppose that the CRT's test does not reject the null of $\gamma_i = \gamma$. Then the TFE MG estimator in (27) is expected to use. That is, the following regression is supposed to be run instead.

$$y_{it} = a_i + \beta_i' x_{it} + F_t + \varepsilon_{it}. \tag{52}$$

Interestingly, it is not straightforward to run (52). The typical two-way fixed effects transformation leads to

$$\dot{y}_{it} = \beta_i' \dot{x}_{it} + e_{it}, \tag{53}$$

where

$$e_{it} = \xi_{it} + \dot{\varepsilon}_{it}, \text{ with } \xi_{it} = \beta_i' \frac{1}{n} \sum\nolimits_{i=1}^{n} \tilde{x}_{it} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \beta_i' \tilde{x}_{it}.$$

Since the cross-sectional average of $\tilde{x}_{it}$ approximates the common factors to $\tilde{x}_{it}$, ξ_{it} can be treated as additional common components in the modified error term, e_{it} in (53). The existence of ξ_{it} influences on the asymptotic variance of the TFE estimator.

Lemma 2: (Asymptotic Difference between TFE and CCE MG Estimators under Heterogeneous Slope Coefficients) *Under Assumption 1-5 and 6A, If either* $\Gamma_i \neq 0$ *or* $\Psi_i \neq 0$, *but* $\gamma_i = \gamma$, *as* $n, T \to \infty$,

$$\left\| \hat{\beta}_{\text{tfe,mg}} - \frac{1}{n} \sum\nolimits_{i=1}^{n} \beta_i \right\| = O_p\left(n^{-1/2}\right) + O_p\left(n^{-1/2} T^{-1/2}\right). \tag{54}$$

Appendix C provides the proof of Lemma 2. There are various other ways to reduce the asymptotic variance. For example, an iterative method must work in this case. Let $\hat{\beta}_i^1$ be the first stage estimator for each i based on (53). Next, estimate the common factor by taking the cross-sectional average of the following residuals.

$$\hat{F}_{t,c} = \frac{1}{n} \sum\nolimits_{i=1}^{n} \left(\tilde{y}_{it} - \hat{\beta}_i^{1\prime} \tilde{x}_{it} \right). \tag{55}$$

Next let $\hat{\beta}_i^2$ be the second stage estimator for each i in the following regression.

$$\tilde{y}_{it} - \hat{F}_{t,c} = \beta_i' \tilde{x}_{it} + \text{error}_{it} \tag{56}$$

Repeating (55) and (56) until the LS estimator converges. This estimator is almost equivalent to the IE estimator proposed by Bai (2009). Instead of the PC estimation for F_t, here we use the cross-sectional average of the residuals. However we do not consider this estimator further simply because this new iterative estimator cannot be viewed as a TFE estimator anymore.

Corollary 3-2: (Asymptotic Comparison for MG Estimators under Heterogeneous Slope Coefficients) *Under Assumptions 1-5 and 6A,* as $n, T \to \infty$,

$$V_{\text{CRT,mg}} \geq V_{\text{BPS,mg}}.$$

The proof Corollary 3-2 is rather minor, and hence it is omitted. Constrast to Corollary 3-1, here the T/n ratio requirement is not needed simply because $\beta_i \neq \beta$. When $\gamma_i \neq \gamma$, both the CRT and the BPS methods suggest to use the CCE-MG estimation. Also when regressors do not include any common factors, or factor loading coefficients are homogeneous across i, Lemma 2 does not hold, and it is easy to show that $\left\| \hat{\beta}_{\text{tfe,mg}} - n^{-1} \sum_{i=1}^{n} \beta_i \right\| = O_p\left(n^{-1/2} T^{-1/2}\right)$. Hence in this case, the TFE-MG estimator becomes asymptotically equivalent to the CCE-MG estimator. Except for these two cases, the BPS method always leads to more efficient estimation.

4 Monte Carlo Simulations

This section examines theoretical findings of this paper, and investigates how the pre-testing methods perform in finite samples by means of Monte Carlo simulations. The data generating process (DGP) is given by

$$y_{it} = \sum_{j=1}^{2} \beta_{j,i} x_{j,it} + \gamma_i F_t + \varepsilon_{it},$$

where each regressor has the following factor structure.

$$x_{j,it} = \lambda_{j,i} F_t + \delta_{j,i} G_t + x_{j,it}^o \text{ for } j = 1, 2.$$

Based on restrictions on factor loadings with $x_{j,it}$, the following two cases are considered: $\lambda_{ji} \neq 0$, $\delta_{ji} \neq 0$ v.s. $\lambda_{ji} = \delta_{ji} = 0$. In the first case, both regressors have two common factors. λ_{ji} is possibly correlated with γ_i. The second case does not allow any cross sectional dependence in the regressors. All commom factors, ε_{it}, and $x_{j,it}^o$ are drawn from $iid\mathcal{N}(0,1)$, factor loadings are drawn from $iid\mathcal{N}(1,1)$. Here we report only the first case to save the space. All other simulation results are reported online.

We compare the finite sample performances of the following three estimators: BPS, CRT and CCE pooled and MG estimators. Note that the CCE estimator is robust compared with the BPS or the CRT estimator since the factor augmented regression nests the TFE regression. We first consider the finite sample performances of three estimators under the case of the homogeneous slope coefficients.

We set $\beta_{1i} = \beta_{2i} = 1$. Table 2 shows the finite sample performances of three estimators when $\gamma_i = \gamma$. As we discussed in Section 2, IC_2 always selects the correct number of common factors. Surprisingly even with small n and T, IC_2 never fails. Meanwhile the $\mathcal{S}_{\gamma,nT}$ statistic shows somewhat a mild size distortion with small n. The nominal size used in the test is 5%. With $n = 25$, the size of the test is slowly decreasing over T, but never reach at the 5% level even with $T = 200$. However as n increases, the size distortion quickly disappears. With $n = T = 200$, the CRT test shows little size distortion. As Theorem 3 states, the variance of the BPS pooled estimator is always the smallest among three pooled estimators. Only when the regressors do not have any factor structure or $\lambda_{ji} = \delta_{ji} = 0$, the variances of other pooled estimators are similar to the variance of the BPS pooled estimator. See the online supplementary appendix for more detailed evidences. Meanwhile the variance of the BPS-MG estimator is more or less similar to that of the CRT-MG etimator. The CCE pooled and MG estimators are robust but least efficient.

Table 3 reports the case of $\gamma_i \neq \gamma$. Evidently, both the BPS and the CRT methods detect this case precisely, which leads to the relative variance ration becomes unity. Also note that in this case, both the BPS and CRT methods always suggest the CCE estimation. Hence the relative variance ratio of the CCE pooled to the BPS pooled estimator becomes unity. The similar finding is observed for the case of the MG estimation.

Table 4 displays the case of the local heterogeneous factor loadings. Only one factor loading is different from the rest. As Theorem 1 shows, the CRT detects this case precisely even with large n. As either n or $T \to \infty$, the rejection rate becomes unity. Meanwhile the BPS method fails to detect the local heterogeneity, so that the BPS method always suggests the TFE estimation. As Theorem 2 states, in this case, the variance of the TFE estimator is smaller than that of the CCE estimator. Meanwhile the CRT method is suggesting the CCE estimation more as $n, T \to \infty$. Hence asymptotically the variance of the BPS estimator is smaller than either CCE or CRT estimator. By combining all results from Table 2, 3 and 4, we can confirm our theoretical findings in Theorem 3.

Next, we investigate the finite sample performance under heterogeneous slope coefficients. Table 5 reports the case where $\beta_i \neq \beta$ but $\gamma_i = \gamma$. As shown in Table 1, the BPS method suggests the CCE estimator, meanwhile the CRT method leads to the TFE estimator. As $n, T \to \infty$ jointly, the CRT method selects the TFE estimation more. As shown in Lemmas 1 and 2, both the BPS pooled or MG estimator is more efficient than the CRT pooled or MG estimator.

Table 6 shows the case where $\beta_i \neq \beta$ and $\gamma_i \neq \gamma$. In this case, both pre-testing procedures suggest the CCE estimator. Hence the variance ratio becomes unity even with small n and T.

5 Conclusion

This paper compared the effectiveness of the two pre-testing procedures – BPS and CRT methods – asymptotically, and showed that the BPS method is more effective. When the slope coefficients are homogeneous, the BPS and the CRT methods are basically same except for the case of the local heterogeneity of the factor loadings. Of course, the CRT method is based on a max-type test so that it allows some minor mistakes under the homogeneous factor loadings. Surprisingly, when the slope coefficients are heterogeneous, the BPS always suggests to run correctly a specified regression. Meanwhile the original CRT method fails to suggest under the homogeneous factor loading case. We did not consider to alter the original CRT method in this paper, which does not impose the homogeneous restriction on the slope coefficients, but if the restriction is imposed, then the modified CRT method restores the virtue except for the local heterogeneity case.

Nonetheless, the finding of this paper is helpful for empirical researchers. After a TFE regression is run, a simple BPS procedure can be run to check whether or not a factor augmented regression is needed to run.

References

[1] Bai, J., (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica*, 77, 1229-1279.

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

[3] Baltagi, B. H., Feng, Q., and Kao, C., (2011). Testing for Sphericity in a Fixed Effects Panel Data Model. *The Econometrics Journa*l, 14(1), 25-47.

[4] Baltagi, B. H., Kao, C., and Na, S., (2013). Testing for Cross-sectional Dependence in a Panel Factor Model Using the Wild Bootstrap F Test. *Statistical Papers*, 54(4), 1067-1094.

[5] Castagnetti, C., Rossi, E., and Trapani, L., (2015a). Inference on Factor Structures in Heterogeneous Panels. *Journal of Econometrics*, 184(1), 145-157.

[6] Castagnetti, C., Rossi, E., and Trapani, L., (2015b). Testing for No Factor Structures: On the Use of Hausman-type Statistics. *Economics Letters*, 130(C), 66-68.

[7] Chudik, A., and Pesaran, M. H., (2015). Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. *Journal of Econometrics*, 188(2), 393-420.

[8] Chudik, A., Pesaran, M. H., and Tosetti, E., (2011). Weak and Strong Cross-section Dependence and Estimation of Large Panels. *The Econometrics Journal*, 14(C), 45-90.

[9] Hayakawa, K., Nagata, S. and Yamagata, T., (2018). A Robust Approach to Heteroskedasticity, Error Serial Correlation and Slope Heterogeneity for Large Linear Panel Data Models with Interactive Effects. *SSRN Electronic Journal*.

[10] Karabiyik, H., Urbain, J-P., and Westerlund, J., (2018). CCE Estimation of Factor-Augmented Regression Models with more Factors than Observables. *Journal of Applied Econometrics*, (Online: 30 SEP), 1–17.

[11] Lee, Y. and Sul, D. (2019). Robust Pooled Estimation: Depth Weighted Mean Group Estimation. Mimeo, Syracuse University.

[12] Ng, S., (2006). Testing Cross-Section Correlation in Panel Data Using Spacings. *Journal of Business & Economic Statistics*, 24(1), 12-23.

[13] Parker, J., and Sul, D., (2016). Identification of Unknown Common Factors: Leaders and Followers. *Journal of Business & Economic Statistics*, 34:2, 227-239.

[14] Pesaran, M. H., (2004). General Diagnostic Tests for Cross Section Dependence in Panels. No 1240, *IZA Discussion Papers*, Institute for the Study of Labor (IZA).

[15] Pesaran, M. H., (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4), 967-1012.

[16] Pesaran, M. H., (2015). Testing Weak Cross-Sectional Dependence in Large Panels. *Econometric Reviews*, 34:6-10, 1089-1117.

[17] Pesaran, M. H., Ullah, A., and Yamagata, T., (2008). A Bias-adjusted LM Test of Error Cross-section Independence. *The Econometrics Journal*, 11: 105-127.

[18] Reese, S., and Westerlund, J., (2015). Panicca: Panic on Cross-Section Averages. *Journal of Applied Econometrics*, (Online: 26 AUG), 1-21.

[19] Reese, S., and Westerlund, J., (2018). Estimation of Factor-Augmented Panel Regressions with Weakly Influential Factors. *Econometric Reviews*, 37(5), 401-465.

[20] Sarafidis, V., and Wansbeek, T., (2012). Cross-Sectional Dependence in Panel Data Analysis. *Econometric Reviews*, 31(5), 483-531.

[21] Sarafidis, V., Yamagata, T., and Robertson, D., (2009). A Test of Cross Section Dependence for a Linear Dynamic Panel Model with Regressors. *Journal of Econometrics*, 148(2), 149-161.

[22] Westerlund, J., (2019). Testing Additive versus Interactive Effects in Fixed-T panels. *Economics Letters*, 174, 5-8.

[23] Westerlund, J., and Urbain, J-P., (2013). On the Estimation and Inference in Factor-augmented Panel Regressions with Correlated Loadings. *Economics Letters*, 119(3), 247-250.

[24] Zuo, Y., & Serfling, R. (2000). General notions of statistical depth function. *Annals of statistics*, 461-482.

Technical Appendix

Appendix A: Proof of Theorem 1

Part I:

First, we show

$$\mathbb{E}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\tilde{\gamma}_i'\tilde{\gamma}_i\right)=O_p\left(n^{-1}\right),$$

under the local heterogeneity defined in Definition 1. Assume without loss of generality that the number of individuals in $\mathcal{G}^c$, $\upsilon=1$, such that

$$\gamma_i=\begin{cases}\gamma & \text{if } i<n\\ \gamma+\tau_n \text{ with } \tau_n\sim iid\left(0,\Omega_0\right) & \text{if } i=n\end{cases}.$$

Then the cross-sectional mean of γ_i becomes

$$\frac{1}{n}\sum\nolimits_{i=1}^{n}\gamma_i=\gamma+\frac{1}{n}\sum\nolimits_{i=1}^{n}\tau_i=\gamma+\frac{\tau_n}{n},$$

and the demeaned factor loading is given by

$$\tilde{\gamma}_i=\gamma_i-\frac{1}{n}\sum\nolimits_{i=1}^{n}\gamma_i=\begin{cases}-\dfrac{\tau_n}{n} & \text{if } i<n\\ \dfrac{n-1}{n}\tau_n & \text{if } i=n\end{cases},$$

in light of which, the following holds,

$$\mathbb{E}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\tilde{\gamma}_i'\tilde{\gamma}_i\right)=\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n-1}\frac{1}{n^2}\tau_n'\tau_n+\left(\frac{n-1}{n}\right)^2\tau_n'\tau_n\right]=\frac{1}{n}\Omega_0+O_p\left(n^{-2}\right). \tag{57}$$

Next, we derive the order of residual, $\hat{u}_{it}$, obtained using BPS method. Define

$$\hat{u}_{it}=\dot{y}_{it}-\hat{\beta}_{\mathrm{tfe,p}}'\dot{x}_{it}=\left(\beta-\hat{\beta}_{\mathrm{tfe,p}}\right)'\dot{x}_{it}+\dot{u}_{it}, \tag{58}$$

where

$$\dot{u}_{it}=\tilde{\gamma}_i'\tilde{F}_t+\dot{\varepsilon}_{it}.$$

Consider the first term in (58). The TFE pooled estimator is given by

$$\begin{aligned}\hat{\beta}_{\mathrm{tfe,p}}-\beta &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\left(\dot{\varepsilon}_{it}+\tilde{F}_t'\tilde{\gamma}_i\right)\\ &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\tilde{\Gamma}_i'\tilde{F}_t+\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^{o}\right)\left(\dot{\varepsilon}_{it}+\tilde{F}_t'\tilde{\gamma}_i\right)\\ &= \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{\varepsilon}_{it}+I+II\right],\end{aligned}$$

where

$$\begin{aligned} I &= \frac{1}{nT}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{n}\tilde{\Gamma}_i'\tilde{F}_t\tilde{F}_t'\tilde{\gamma}_i, \\ II &= \frac{1}{nT}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{n}\left[\left(\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^o\right)\tilde{F}_t'\tilde{\gamma}_i\right]. \end{aligned}$$

Note that $\mathbb{E}\dot{x}_{it}\dot{\varepsilon}_{js}=\mathbb{E}\left[\left(\tilde{\Gamma}_i'\tilde{F}_t+\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^o\right)\dot{\varepsilon}_{js}\right]=0$ for all i,j,s,t, so

$$\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{\varepsilon}_{it}=O_p\left((nT)^{-1/2}\right).$$

Next, consider I. By Assumption 1, we have $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'\to^p\Sigma_F$. If Γ_i is correlated with γ_i, such that $\Gamma_i=q\gamma_i+\Gamma_i^o$, where Γ_i^o is independent of γ_i, I is biased, and the order of which is given by

$$\begin{aligned} \mathbb{E}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\tilde{\Gamma}_i'\tilde{\gamma}_i\right) &= \frac{1}{n}\sum\nolimits_{i=1}^{n}\mathbb{E}\left[\left(q\tilde{\gamma}_i+\tilde{\Gamma}_i^o\right)'\tilde{\gamma}_i\right] \\ &= q\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n-1}\tilde{\gamma}_i'\tilde{\gamma}_i+\tilde{\gamma}_n'\tilde{\gamma}_n\right] \\ &= q\frac{1}{n}\Omega_0+O\left(n^{-2}\right)=O\left(n^{-1}\right) \end{aligned}$$

which is the same as the bias of CCEP estimator. If Γ_i is not correlated with γ_i, such that

$$\mathbb{E}\left(\frac{1}{nT}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{n}\tilde{\Gamma}_i'\tilde{\gamma}_i\right)=0,$$

then the following holds,

$$\begin{aligned} \mathbb{E}\left\|\frac{1}{n}\sum\nolimits_{i=1}^{n}\tilde{\Gamma}_i'\tilde{\gamma}_i\right\|^2 &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[\tilde{\Gamma}_i'\tilde{\gamma}_i\tilde{\gamma}_i'\tilde{\Gamma}_i\right] \\ &= \frac{1}{n^2}\left\{\sum_{i=1}^{n-1}\mathbb{E}\left[\tilde{\Gamma}_i'\left(\frac{1}{n^2}\right)\Omega_0\tilde{\Gamma}_i\right]+\mathbb{E}\left[\tilde{\Gamma}_n'\left(\frac{n-1}{n}\right)^2\Omega_0\tilde{\Gamma}_n\right]\right\} \\ &= O\left(\frac{1}{n^2}\right), \end{aligned}$$

which implies $I=O_p\left(n^{-1}\right)$.

For II, first note that $\mathbb{E}(II) = 0$ if Ψ_i is independent of γ_i. Then it holds that

$$\begin{aligned}
& \mathbb{E}\left\|\frac{1}{nT}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{n}\left(\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^o\right)\tilde{F}_t'\tilde{\gamma}_i\right\|^2 \\
= & \frac{1}{n^2T^2}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{n}\mathbb{E}\left\{\left(\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^o\right)\tilde{F}_t'\tilde{\gamma}_i\tilde{\gamma}_i'\tilde{F}_t\left(\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^o\right)'\right\} \\
= & \frac{1}{n^2T^2}\sum\nolimits_{t=1}^{T}\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1}\left(\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^o\right)\tilde{F}_t'\left(\frac{1}{n^2}\epsilon_n'\epsilon_n\right)\tilde{F}_t\left(\tilde{\Psi}_i'\tilde{G}_t+\dot{x}_{it}^o\right)'\right] \\
& +\frac{1}{n^2T^2}\sum\nolimits_{t=1}^{T}\mathbb{E}\left[\left(\tilde{\Psi}_n'\tilde{G}_t+\dot{x}_{nt}^o\right)\tilde{F}_t'\left(\frac{n-1}{n}\right)^2\epsilon_n'\epsilon_n\tilde{F}_t\left(\tilde{\Psi}_n'\tilde{G}_t+\dot{x}_{nt}^o\right)'\right] \\
= & O\left(\frac{1}{n^3T}\right)+O\left(\frac{1}{n^2T}\right)=O\left(\frac{1}{n^2T}\right).
\end{aligned}$$

Putting all together, we have

$$\hat{\beta}_{\text{tfe,p}}-\beta=O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{n\sqrt{T}}\right). \tag{59}$$

This implies that we need the $T/n \to 0$ condition for the consistency of TFE estimator under the local heterogeneity. That is,

$$\sqrt{nT}(\hat{\beta}_{\text{tfe,p}}-\beta)=O_p\left(\sqrt{T/n}\right)+O_p(1)+O_p\left(1/\sqrt{n}\right).$$

For the second term in (57), it holds that

$$\dot{u}_{it}=\begin{cases} -\dfrac{1}{n}\tau_n'\tilde{F}_t+\dot{\varepsilon}_{it} & \text{if } i<n \\ \dfrac{n-1}{n}\tau_n'\tilde{F}_t+\dot{\varepsilon}_{it} & \text{if } i=n \end{cases}. \tag{60}$$

Define $\kappa = \hat{\#}\left(\hat{u}_{it}(\kappa)\right)$. Combining (59) and (60) yields

$$\hat{u}_{it}(\kappa)=\begin{cases} -\dfrac{1}{n}\tau_n'\tilde{F}_t+\dot{\varepsilon}_{it}+O_p\left(1/\sqrt{nT}\right)+O_p\left(n^{-1}\right) & \text{if } i<n \\ \dfrac{n-1}{n}\tau_n'\tilde{F}_t+\dot{\varepsilon}_{it}+O_p\left(1/\sqrt{nT}\right)+O_p\left(n^{-1}\right) & \text{if } i=n \end{cases}. \tag{61}$$

Last, we need to show that under the local heterogeneity of γ_i,

$$\lim_{n,T\to\infty}\Pr\left[\hat{\#}\left(\hat{u}_{it}(\kappa)\right)=0\right]=1,$$

We shall prove for all $0<\kappa\le\kappa_{\max}$,

$$\lim_{n,T\to\infty}\Pr\left[IC_2(\kappa)<IC_2(0)\right]=0,$$

where

$$IC_2(\kappa)=\ln\left(\hat{V}(\kappa)\right)+\kappa\left(\frac{n+T}{nT}\right)\ln\left(\min[n,T]\right),\ IC_2(0)=\ln\left(\hat{V}(0)\right),$$

$$\hat{V}(\kappa) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{u}_{it}^{2}(\kappa)\,, \text{ and } \hat{V}(0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{\varepsilon}_{it}^{2}.$$

Note that the penalty function of $IC_2(\kappa)$ satisfies that

$$\begin{aligned} & \left(\frac{n+T}{nT}\right)\ln\left(\min[n,T]\right) \\ = & \left(\frac{1}{n}+\frac{1}{T}\right)\ln\left(\min[n,T]\right) > O_p\left(\frac{1}{\min[n,T]}\right). \end{aligned}$$

Hence, to show

$$\begin{aligned} & \lim_{n,T\to\infty}\Pr\left[IC_2(\kappa) < IC_2(0)\right] \\ = & \lim_{n,T\to\infty}\Pr\left[\ln\left(\hat{V}(\kappa)\right) - \ln\left(\hat{V}(0)\right) + \kappa\left(\frac{1}{n}+\frac{1}{T}\right)\ln\left(\min[n,T]\right) < 0\right] = 0, \end{aligned}$$

it suffices to show that

$$\ln\left(\hat{V}(\kappa)\right) - \ln\left(\hat{V}(0)\right) \le O_p\left(\frac{1}{\min[n,T]}\right).$$

Also note that

$$\ln\left(\hat{V}(\kappa)\right) - \ln\left(\hat{V}(0)\right) = \ln\left(\frac{\hat{V}(\kappa)}{\hat{V}(0)}\right) \le \frac{\hat{V}(\kappa)}{\hat{V}(0)} - 1 = \frac{\hat{V}(\kappa) - \hat{V}(0)}{\hat{V}(0)}$$

for $\left(\hat{V}(\kappa)/\hat{V}(0)\right) > 0$. Since $\hat{V}(0) = O_p(1)$, it is sufficient to show that

$$\hat{V}(\kappa) - \hat{V}(0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{u}_{it}^{2}(\kappa) - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{\varepsilon}_{it}^{2} \le O_p\left(\frac{1}{\min[n,T]}\right).$$

Substituting (61) into $\hat{V}(\kappa)$ gives us

$$\begin{aligned} \hat{V}(\kappa) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{u}_{it}^{2}(\kappa) \\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{\varepsilon}_{it}^{2} + O_p\left((nT)^{-1}\right) + O_p\left(n^{-2}\right) \\ &\quad + \frac{1}{nT}\sum_{t=1}^{T}\left[\sum_{i=1}^{n-1}\frac{1}{n^2}\left(\tau_n'\tilde{F}_t\right)^2 + \left(\frac{n-1}{n}\right)^2\left(\tau_n'\tilde{F}_t\right)^2\right] \\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{\varepsilon}_{it}^{2} + O_p\left((nT)^{-1}\right) + O_p\left(n^{-2}\right) + O_p\left(n^{-1}\right) \end{aligned}$$

Therefore

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{u}_{it}^{2}(\kappa) - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{\varepsilon}_{it}^{2} = O_p\left(n^{-1}\right) + O_p\left((nT)^{-1}\right) \le O_p\left(\frac{1}{\min[n,T]}\right).$$

Part II

See the proof of Theorem 3 in CRT (2015b).

Appendix B: Proof of Theorem 3

There are three sub-cases: Under the null, alternative and local heterogeneity. We consider each case separately, and then combine them together later.

Case A: Under the null of $\gamma_i = \gamma$ As $n, T \to \infty$, it is easy to show that

$$\lim_{n,T\to\infty} \omega_{BPS} = \lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 1.$$

Meanwhile as $n, T \to \infty$ with $T/n^{5/3} \to 0$ and $n/T^3 \to 0$, CRT (2015b) showed that

$$\lim_{n,T\to\infty} \omega_{CRT} = \lim_{n,T\to\infty} \Pr(\mathcal{S}_{\gamma,nT} \le c_{\alpha,n}) = \alpha.$$

Hence

$$\lim_{n,T\to\infty} \hat{\beta}_{\text{BPS,p}} = \hat{\beta}_{\text{tfe,p}}, \;\; \lim_{n,T\to\infty} \hat{\beta}_{\text{CRT,p}} = \alpha\hat{\beta}_{\text{tfe,p}} + (1-\alpha)\,\hat{\beta}_{\text{cce,p}}.$$

Since $V_{\text{cce,p}} \ge V_{\text{tfe,p}}$ in this case, the following inequality holds.

$$V_{\text{BPS,p}} \le V_{\text{CRT,p}}$$

Similarly, we can show that

$$\lim_{n,T\to\infty} \hat{\beta}_{\text{BPS,mg}} = \hat{\beta}_{\text{tfe,mg}}, \;\; \lim_{n,T\to\infty} \hat{\beta}_{\text{CRT,mg}} = \alpha\hat{\beta}_{\text{tfe,mg}} + (1-\alpha)\,\hat{\beta}_{\text{cce,mg}},$$

and

$$V_{\text{BPS,mg}} \le V_{\text{CRT,mg}}$$

Case B: Under the alternative In this case, both the BPS and CRT methods suggest the CCE estimation. Hence we have

$$V_{\text{BPS,p}} = V_{\text{CRT,p}}, \;\&\; V_{\text{BPS,mg}} = V_{\text{CRT,mg}}$$

Case C: Under the local heterogeneity Under the local heterogeneity, as $n, T \to \infty$ with $T/n \to 0$, the BPS method suggests asymptotically,

$$\lim_{n,T\to\infty} \omega_{BPS} = \lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 1,$$

meanwhile as $n, T \to \infty$ with $T/n^{5/3} \to 0$ and $n/T^3 \to 0$, the CRT method suggests

$$\lim_{n,T\to\infty} \omega_{CRT} = \lim_{n,T\to\infty} \Pr(\mathcal{S}_{\gamma,nT} \leq c_{\alpha,n}) = 0.$$

Hence it is easy to show that

$$\lim_{n,T\to\infty} \hat{\beta}_{\text{BPS,p}} = \hat{\beta}_{\text{tfe,p}}, \quad \lim_{n,T\to\infty} \hat{\beta}_{\text{CRT,p}} = \hat{\beta}_{\text{cce,p}}.$$

Similary,

$$\lim_{n,T\to\infty} \hat{\beta}_{\text{BPS,mg}} = \hat{\beta}_{\text{tfe,mg}}, \quad \lim_{n,T\to\infty} \hat{\beta}_{\text{CRT,mg}} = \hat{\beta}_{\text{cce,mg}}.$$

Therefore

$$V_{\text{BPS,p}} \leq V_{\text{CRT,p}}, \ \& \ V_{\text{BPS,mg}} \leq V_{\text{CRT,mg}}$$

Combining all three cases, we can verify (44).

Appendix C:

Proof of Lemma 1

Let $P_t = [F_t, G_t]'$ and $\Lambda_i = [\Gamma_i', \Psi_i']$. Then we rewrite the panel regression as

$$y_{it} = a_i + \beta' x_{it} + F_t + e_{it},$$

$$x_{it} = \Lambda_i P_t + x_{it}^o.$$

and

$$e_{it} = \eta_i' x_{it} + \varepsilon_{it}.$$

$$\hat{\beta}_{\text{tfe,p}} - \frac{1}{n}\sum\nolimits_{i=1}^n \beta_i = \hat{\beta}_{\text{tfe,p}} - \beta - \frac{1}{n}\sum\nolimits_{i=1}^n (\beta_i - \beta) = \hat{\beta}_{\text{tfe,p}} - \beta - \frac{1}{n}\sum\nolimits_{i=1}^n \eta_i.$$

$$\begin{aligned} & \hat{\beta}_{\text{tfe,p}} - \beta - \frac{1}{n}\sum\nolimits_{i=1}^n \eta_i \\ = & \left(\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it}\dot{x}_{it}'\right)^{-1} \left(\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it}\dot{e}_{it}\right) - \frac{1}{n}\sum_{i=1}^n \eta_i \\ = & \left(\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it}\dot{x}_{it}'\right)^{-1} \left[\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it}\dot{e}_{it} - \left(\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it}\dot{x}_{it}'\right)\frac{1}{n}\sum_{i=1}^n \eta_i\right] \end{aligned}$$

Consider the numerator term first.

$$
\begin{aligned}
& \sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{e}_{it}-\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)\frac{1}{n}\sum_{i=1}^{n}\eta_i \\
= & \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\dot{x}_{it}\dot{e}_{it}-\dot{x}_{it}\dot{x}_{it}'\frac{1}{n}\sum_{i=1}^{n}\eta_i\right] \\
= & \sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\left[\tilde{x}_{it}'\eta_i-\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{it}'\eta_i+\dot{\varepsilon}_{it}-\left(\tilde{x}_{it}'-\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{it}'\right)\frac{1}{n}\sum_{i=1}^{n}\eta_i\right] \\
= & \sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\left[\tilde{x}_{it}'\tilde{\eta}_i-\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{it}'\tilde{\eta}_i+\dot{\varepsilon}_{it}\right]
\end{aligned}
$$

where $\tilde{\eta}_i=\eta_i-\frac{1}{n}\sum_{i=1}^{n}\eta_i$. Let

$$
\hat{\beta}_{\mathrm{tfe,p}}-\beta-\frac{1}{n}\sum_{i=1}^{n}\eta_i=\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\left[I+II+III\right],
$$

where

$$
I=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{\Lambda}_i\tilde{P}_t\tilde{P}_t'\left(\Lambda_i'\tilde{\eta}_i-\frac{1}{n}\sum_{i=1}^{n}\Lambda_i'\tilde{\eta}_i\right).
$$

Let $\tilde{\vartheta}_i=\Lambda_i'\tilde{\eta}_i-\frac{1}{n}\sum_{i=1}^{n}\Lambda_i'\tilde{\eta}_i$. Without a loss of generality, we may let $\frac{1}{T}\sum_{t=1}^{T}\tilde{P}_t\tilde{P}_t'=I_{r+m}$, and introduce an invertable rotating matrix H. As CRT (2015b) did, we further let $H=I$ for the notational simplicity. Then we have

$$
I=\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{\Lambda}_i\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{P}_t\tilde{P}_t'\right)\tilde{\vartheta}_i\right]=\frac{1}{n}\sum_{i=1}^{n}\tilde{\Lambda}_i\tilde{\vartheta}_i=O_p\left(\frac{1}{\sqrt{n}}\right),
$$

since $\mathbb{E}\tilde{\Lambda}_i\tilde{\vartheta}_i=0$ and their second moments are finite by Assumption 3 and 6A.

Next, consider II. By the mutual independence of Λ_i, P_t, x_{js}^o and η_m for all i,j,m,t and s, it follows that

$$
\mathbb{E}\left(II\right)=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\left[\dot{x}_{it}^o\tilde{P}_t'\left(\Lambda_i'\tilde{\eta}_i-\frac{1}{n}\sum_{i=1}^{n}\Lambda_i'\tilde{\eta}_i\right)\right]=0,
$$

and that the second moment of the II term is finite, which leads to

$$
II=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\dot{x}_{it}^o\left(\tilde{P}_t'\Lambda_i'\tilde{\eta}_i-\frac{1}{n}\sum_{i=1}^{n}\tilde{P}_t'\Lambda_i'\tilde{\eta}_i\right)\right]=O_p\left(\frac{1}{\sqrt{nT}}\right).
$$

Similarly, x_{it} is independent of $\eta_j'x_{js}^o$ and ε_{js} for all i,j,t and s, it holds that

$$
III=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\left(\varphi_{it}+\dot{\varepsilon}_{it}\right)=O_p\left(\frac{1}{\sqrt{nT}}\right),
$$

where $\varphi_{it} = \tilde{x}^{o\prime}_{it}\tilde{\eta}_i - \frac{1}{n}\sum_{i=1}^n \tilde{x}^{o\prime}_{it}\tilde{\eta}_i$.

Next, we need to show that $\hat{\beta}_{\text{cce,p}} - \frac{1}{n}\sum_{i=1}^n \beta_i = O_p\left(n^{-1/2}T^{-1/2}\right)$, which was already shown in the eq. (57) in Pesaran (2006). When the rank condition is satisfied, the first term disappears. Only when the rank condition is not satisfied, the first term should be included. In this case, the CCE pooled estimator is not efficient compared with the TFE estimator.

Proof of Lemma 2

Let $\gamma_i = \gamma = 1$ without a loss of generality. The TFE transformation leads to

$$\dot{y}_{it} = \beta_i'\dot{x}_{it} + w_{it},$$

where w_{it} is defined as

$$w_{it} = \xi_{it} + \dot{\varepsilon}_{it}, \text{ with } \xi_{it} = \beta_i'\frac{1}{n}\sum\nolimits_{i=1}^n \tilde{x}_{it} - \frac{1}{n}\sum\nolimits_{i=1}^n \beta_i'\tilde{x}_{it}.$$

Note that

$$\hat{\beta}_{\text{tfe,mg}} - \frac{1}{n}\sum_{i=1}^n \beta_i = \frac{1}{nT}\sum\nolimits_{i=1}^n \left(\frac{1}{T}\sum\nolimits_{t=1}^T \dot{x}_{it}\dot{x}_{it}'\right)^{-1}\left(\sum\nolimits_{t=1}^T \dot{x}_{it}w_{it}\right). \tag{62}$$

Consider the numerator term first.

$$\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \dot{x}_{it}w_{it} = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \dot{x}_{it}\left[\left(\frac{1}{n}\sum\nolimits_{i=1}^n \tilde{x}_{it}'\right)\eta_i - \frac{1}{n}\sum\nolimits_{i=1}^n \tilde{x}_{it}'\eta_i + \dot{\varepsilon}_{it}\right].$$

Note that

$$\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_{it}'\right)\eta_i - \frac{1}{n}\sum_{i=1}^n \tilde{x}_{it}'\eta_i = \tilde{P}_t'\left(\bar{\Lambda}_n'\eta_i - \frac{1}{n}\sum_{i=1}^n \Lambda_i'\eta_i\right) + \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}^o_{it}\right)'\eta_i - \frac{1}{n}\sum_{i=1}^n \tilde{x}^{o\prime}_{it}\eta_i,$$

where $\bar{\Lambda}_n = \frac{1}{n}\sum_{i=1}^n \Lambda_i$. Rewrite (62) as

$$\hat{\beta}_{\text{tfe,mg}} - \frac{1}{n}\sum_{i=1}^n \beta_i = \left(\frac{1}{T}\sum_{t=1}^T \dot{x}_{it}\dot{x}_{it}'\right)^{-1}\left[I + II + III\right],$$

where

$$\begin{aligned}
I &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \tilde{\Lambda}_i\tilde{P}_t\tilde{P}_t'\left(\bar{\Lambda}_n'\eta_i - \frac{1}{n}\sum_{i=1}^n \Lambda_i'\eta_i\right),\\
II &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \dot{x}^o_{it}\tilde{P}_t'\left(\bar{\Lambda}_n'\eta_i - \frac{1}{n}\sum_{i=1}^n \Lambda_i'\eta_i\right),\\
III &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \dot{x}_{it}\left(\Theta_{it} + \dot{\varepsilon}_{it}\right),
\end{aligned}$$

$\varrho_i = \bar{\Lambda}_n' \eta_i - \frac{1}{n}\sum_{i=1}^n \Lambda_i' \eta_i$, and $\Theta_{it} = \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_{it}^o\right)' \eta_i - \frac{1}{n}\sum_{i=1}^n \tilde{x}_{it}^{o\prime} \eta_i$. Note that

$$\bar{\Lambda}_n' \tilde{\eta}_i = \left(\frac{1}{n}\sum_{i=1}^n \Lambda_i\right) \tilde{\eta}_i = O_p(1),$$

and by independence of Λ_i and η_j for all i and j,

$$\frac{1}{n}\sum_{i=1}^n \tilde{\Lambda}_i' \eta_i = O_p\left(\frac{1}{\sqrt{n}}\right),$$

then

$$\begin{aligned}
\varrho_i &= \bar{\Lambda}_n' \eta_i - \frac{1}{n}\sum_{i=1}^n \Lambda_i' \eta_i = \bar{\Lambda}_n' \eta_i - \bar{\Lambda}_n' \frac{1}{n}\sum_{i=1}^n \eta_i + \bar{\Lambda}_n' \frac{1}{n}\sum_{i=1}^n \eta_i - \frac{1}{n}\sum_{i=1}^n \Lambda_i' \eta_i \\
&= \bar{\Lambda}_n' \tilde{\eta}_i - \frac{1}{n}\sum_{i=1}^n \tilde{\Lambda}_i' \tilde{\eta}_i = O_p(1).
\end{aligned}$$

Following the similar strategy adopted in the Proof of Lemma 1, we can show that

$$I = \frac{1}{n}\sum_{i=1}^n \left[\tilde{\Lambda}_i \left(\frac{1}{T}\sum_{t=1}^T \tilde{P}_t \tilde{P}_t'\right) \varrho_i\right] = O_p\left(\frac{1}{\sqrt{n}}\right),$$

since $\mathbb{E}\tilde{\Lambda}_i \varrho_i = 0$ and their second moments are finite by Assumption 3 and 6A.

Next, consider II. By the mutual independence of Λ_i, P_t, x_{js}^o and η_m for all i, j, m, t and s, and the bounded second moment of II, it follows that,

$$II = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left[\dot{x}_{it}^o \tilde{P}_t' \left(\bar{\Lambda}_n' \tilde{\eta}_i - \frac{1}{n}\sum_{i=1}^n \tilde{\Lambda}_i' \eta_i\right)\right] = O_p\left(\frac{1}{\sqrt{nT}}\right).$$

Last, since x_{it} is independent of $\eta_j' x_{js}^o$ and ε_{js} for all i, j, t and s, it holds that

$$III = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it} \left(\Theta_{it} + \dot{\varepsilon}_{it}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right),$$

Therefore,

$$\hat{\beta}_{\text{tfe,mg}} - \frac{1}{n}\sum_{i=1}^n \beta_i = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right).$$

Table 2: Finite sample peformances of pre-testing procedures
under homogeneous factor loadings and slope coefficients

		Frequencies*		Pooled Case Variance Comparison			MG Case Variance Comparison		
n	T	$\text{IC}_2 > 0$	$\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	$V_{\text{BPS,p}}$	$\frac{V_{\text{CRT,p}}}{V_{\text{BPS,p}}}$	$\frac{V_{\text{CCE,p}}}{V_{\text{BPS,p}}}$	$V_{\text{BPS,mg}}$	$\frac{V_{\text{CRT,mg}}}{V_{\text{BPS,mg}}}$	$\frac{V_{\text{CCE,mg}}}{V_{\text{BPS,mg}}}$
25	25	0.000	0.255	0.304	1.516	2.997	0.530	1.281	2.040
25	50	0.000	0.185	0.148	1.345	2.709	0.239	1.172	1.803
25	100	0.000	0.159	0.072	1.319	2.736	0.113	1.159	1.805
25	200	0.000	0.139	0.036	1.333	2.778	0.055	1.164	1.855
50	25	0.000	0.210	0.143	1.531	3.245	0.250	1.264	2.144
50	50	0.000	0.131	0.071	1.268	2.972	0.120	1.125	1.875
50	100	0.000	0.099	0.035	1.200	2.971	0.058	1.086	1.845
50	200	0.000	0.090	0.017	1.118	2.882	0.028	1.036	1.786
100	25	0.000	0.174	0.070	1.414	3.257	0.128	1.203	2.078
100	50	0.000	0.107	0.035	1.229	3.057	0.058	1.103	1.966
100	100	0.000	0.084	0.017	1.176	3.118	0.029	1.069	1.897
100	200	0.000	0.079	0.009	1.111	2.889	0.014	1.071	1.857
200	25	0.000	0.178	0.036	1.417	3.333	0.066	1.182	2.106
200	50	0.000	0.097	0.017	1.235	3.059	0.028	1.107	1.964
200	100	0.000	0.075	0.008	1.250	3.250	0.014	1.071	1.929
200	200	0.000	0.056	0.004	1.250	3.250	0.007	1.000	1.857

Note: *) The nominal size equals 5%. All variances are multiplied by 10^3.

Table 3: Finite sample peformances of pre-testing procedures
under heterogeneous factor loadings but homogeneous slope coefficients

		Frequencies		Pooled Case Variance Comparison			MG Case Variance Comparison		
n	T	$\text{IC}_2 > 0$	$\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	$V_{\text{BPS,p}}$	$\frac{V_{\text{CRT,p}}}{V_{\text{BPS,p}}}$	$\frac{V_{\text{CCE,p}}}{V_{\text{BPS,p}}}$	$V_{\text{BPS,mg}}$	$\frac{V_{\text{CRT,mg}}}{V_{\text{BPS,mg}}}$	$\frac{V_{\text{CCE,mg}}}{V_{\text{BPS,mg}}}$
25	25	0.990	1.000	1.043	0.998	0.998	1.199	0.997	0.997
25	50	1.000	1.000	0.465	1.000	1.000	0.495	1.002	1.002
25	100	1.000	1.000	0.229	1.000	1.000	0.236	1.000	1.000
25	200	1.000	1.000	0.109	1.000	1.000	0.110	1.000	1.000
50	25	1.000	1.000	0.482	1.000	1.000	0.565	1.000	1.000
50	50	1.000	1.000	0.223	1.000	1.000	0.239	1.000	1.000
50	100	1.000	1.000	0.107	1.000	1.000	0.111	1.000	1.000
50	200	1.000	1.000	0.054	1.000	1.000	0.054	1.000	1.000
100	25	1.000	1.000	0.241	1.000	1.000	0.279	1.000	1.000
100	50	1.000	1.000	0.109	1.000	1.000	0.116	1.000	1.000
100	100	1.000	1.000	0.052	1.000	1.000	0.054	1.000	1.000
100	200	1.000	1.000	0.025	1.000	1.000	0.026	1.000	1.000
200	25	1.000	1.000	0.121	1.000	1.000	0.140	1.000	1.000
200	50	1.000	1.000	0.055	1.000	1.000	0.059	1.000	1.000
200	100	1.000	1.000	0.026	1.000	1.000	0.027	1.000	1.000
200	200	1.000	1.000	0.013	1.000	1.000	0.013	1.000	1.000

Table 4: Finite sample peformances of pre-testing procedures
under local heterogenous factor loadings and homogeneous slope coefficients

		Frequencies		Pooled Case Variance Comparison			MG Case Variance Comparison		
n	T	$\text{IC}_2 > 0$	$\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	$V_{\text{BPS,p}}$	$\frac{V_{\text{CRT,p}}}{V_{\text{BPS,p}}}$	$\frac{V_{\text{CCE,p}}}{V_{\text{BPS,p}}}$	$V_{\text{BPS,mg}}$	$\frac{V_{\text{CRT,mg}}}{V_{\text{BPS,mg}}}$	$\frac{V_{\text{CCE,mg}}}{V_{\text{BPS,mg}}}$
25	25	0.000	0.845	0.549	1.590	1.689	0.705	1.451	1.545
25	50	0.000	0.953	0.360	1.189	1.172	0.400	1.123	1.133
25	100	0.000	0.993	0.298	0.695	0.681	0.269	0.796	0.788
25	200	0.000	0.999	0.248	0.415	0.415	0.208	0.505	0.505
50	25	0.000	0.823	0.198	2.116	2.384	0.300	1.657	1.837
50	50	0.000	0.950	0.122	1.721	1.762	0.155	1.432	1.465
50	100	0.000	0.994	0.088	1.136	1.136	0.092	1.130	1.130
50	200	0.000	1.000	0.066	0.742	0.742	0.062	0.806	0.806
100	25	0.000	0.784	0.087	2.333	2.655	0.142	1.761	1.951
100	50	0.000	0.934	0.048	2.063	2.125	0.066	1.621	1.667
100	100	0.000	0.990	0.031	1.742	1.742	0.038	1.447	1.447
100	200	0.000	1.000	0.021	1.190	1.190	0.022	1.182	1.182
200	25	0.000	0.754	0.039	2.564	3.051	0.067	1.836	2.075
200	50	0.000	0.915	0.020	2.500	2.650	0.031	1.742	1.839
200	100	0.000	0.987	0.012	2.167	2.167	0.017	1.588	1.588
200	200	0.000	0.999	0.007	1.857	1.857	0.009	1.444	1.444

Table 5: Finite sample peformances of pre-testing procedures
under homogeneous factor loadings but heterogeneous slope coefficients

				Pooled Case			MG Case		
		Frequencies		Variance Comparison			Variance Comparison		
n	T	$\text{IC}_2 > 0$	$\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	$V_{\text{BPS,p}}$	$\frac{V_{\text{CRT,p}}}{V_{\text{BPS,p}}}$	$\frac{V_{\text{CCE,p}}}{V_{\text{BPS,p}}}$	$V_{\text{BPS,mg}}$	$\frac{V_{\text{CRT,mg}}}{V_{\text{BPS,mg}}}$	$\frac{V_{\text{CCE,mg}}}{V_{\text{BPS,mg}}}$
25	25	1.000	0.320	22.86	1.525	0.999	20.82	1.247	1.000
25	50	1.000	0.251	21.57	1.666	1.000	20.45	1.288	1.000
25	100	1.000	0.246	21.37	1.657	1.000	20.76	1.254	1.000
25	200	1.000	0.295	20.25	1.577	1.000	19.93	1.203	1.000
50	25	1.000	0.253	11.92	1.647	1.000	10.63	1.246	1.000
50	50	1.000	0.171	10.50	1.767	1.000	10.02	1.276	1.000
50	100	1.000	0.153	10.41	1.803	1.000	10.06	1.243	1.000
50	200	1.000	0.147	10.40	1.836	1.000	10.32	1.300	1.000
100	25	1.000	0.197	5.603	1.801	1.000	5.077	1.321	1.000
100	50	1.000	0.124	5.311	1.911	1.000	5.030	1.286	1.000
100	100	1.000	0.095	5.164	2.025	1.000	4.983	1.298	1.000
100	200	1.000	0.092	5.078	1.968	1.000	5.000	1.303	1.000
200	25	1.000	0.179	2.932	1.796	1.000	2.667	1.272	1.000
200	50	1.000	0.100	2.694	1.952	1.000	2.562	1.310	1.000
200	100	1.000	0.072	2.644	1.985	1.000	2.578	1.263	1.000
200	200	1.000	0.063	2.591	1.969	1.000	2.557	1.282	1.000

Table 6: Finite sample peformances of pre-testing procedures under heterogeneous factor loadings and slope coefficients

				Pooled Case			MG Case		
		Frequencies		Variance Comparison			Variance Comparison		
n	T	$\mathrm{IC}_2 > 0$	$\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	$V_{\mathrm{BPS,p}}$	$\frac{V_{\mathrm{CRT,p}}}{V_{\mathrm{BPS,p}}}$	$\frac{V_{\mathrm{CCE,p}}}{V_{\mathrm{BPS,p}}}$	$V_{\mathrm{BPS,mg}}$	$\frac{V_{\mathrm{CRT,mg}}}{V_{\mathrm{BPS,mg}}}$	$\frac{V_{\mathrm{CCE,mg}}}{V_{\mathrm{BPS,mg}}}$
25	25	1.000	1.000	26.205	1.000	1.000	21.588	1.000	1.000
25	50	1.000	1.000	23.696	1.000	1.000	20.466	1.000	1.000
25	100	1.000	1.000	23.292	1.000	1.000	20.290	1.000	1.000
25	200	1.000	1.000	22.985	1.000	1.000	19.830	1.000	1.000
50	25	1.000	1.000	12.315	1.000	1.000	10.546	1.000	1.000
50	50	1.000	1.000	11.690	1.000	1.000	10.443	1.000	1.000
50	100	1.000	1.000	11.089	1.000	1.000	10.289	1.000	1.000
50	200	1.000	1.000	10.579	1.000	1.000	9.826	1.000	1.000
100	25	1.000	1.000	6.128	1.000	1.000	5.275	1.000	1.000
100	50	1.000	1.000	5.476	1.000	1.000	5.073	1.000	1.000
100	100	1.000	1.000	5.164	1.000	1.000	4.909	1.000	1.000
100	200	1.000	1.000	5.292	1.000	1.000	5.162	1.000	1.000
200	25	1.000	1.000	3.061	1.000	1.000	2.671	1.000	1.000
200	50	1.000	1.000	2.678	1.000	1.000	2.543	1.000	1.000
200	100	1.000	1.000	2.604	1.000	1.000	2.504	1.000	1.000
200	200	1.000	1.000	2.538	1.000	1.000	2.503	1.000	1.000