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

Minyu Han
University of Texas at Dallas

Jihun Kwak
Hanyang University

Donggyu Sul
University of Texas at Dallas

February 9, 2021

Abstract

Empirical researchers may wonder whether or not a two-way fixed effects estimator (with individual and period fixed effects) is sufficiently sophisticated to isolate the influence of common shocks on the estimation of slope coefficients. If it is not, practitioners need to run the so-called panel factor augmented regression instead. There are two pre-testing procedures available in the literature: the use of the estimated number of factors and the direct test of estimated factor loading coefficients. This paper compares the two pre-testing methods asymptotically. Under the presence of the heterogeneous factor loadings, both pre-testing procedures suggest using the Common Correlated Effects (CCE) estimator. Meanwhile, when factor loadings are homogeneous, the pre-testing method utilizing the estimated number of factors always suggests more efficient estimation methods. By comparing asymptotic variances, this paper finds that when the slope coefficients are homogeneous with homogeneous factor loadings, the two-way fixed effects estimation is more efficient than the CCE estimation. However, when the slope coefficients are heterogeneous with homogeneous 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 through the use of an empirical example.

JEL Classification Number: C33

Keywords: two-way fixed effects estimator, panel factor augmented estimator, factor number, maximum test, CCE estimator

*Helpful comments on the original version were received from Alexander Chudik, Yoonseok Lee, and participants in Asian Econometric Society Group meeting and Midwest Econometric meeting. We thank valuable comments from the Editor, the Associate Editor and two anonymous referees. Also, special thanks goes to Ryan Greenaway-McGrevy for editorial help. Corresponding author: Jihun Kwak (jihunkwak.us@gmail.com).

1 Introduction

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

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

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 period or time fixed effect. If the common shock F_t, which can cause cross-sectional dependence among y_{it}, influences each individual differently, the TFE regression is not sufficiently sophisticated to isolate the heterogeneous effect from the common shock. In this case, the following factor augmented regression is used instead:

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

where F_t does not need to be a single common shock and 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 $(k \times 1)$ vector of regressors x_{it} may share the same common factors. That is, the regressors can be modeled by

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

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

When the true factor loadings in (2) are homogeneous ($\gamma_i = \gamma$) but we run (2), the resulting estimator of β is less efficient. 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, the typical panel robust variance estimator is no longer consistent due to the existence of the cross-sectional dependence. The solution is rather simple. Once the common factors are included as additional regressors, one can exclude the source of cross-sectional correlation from estimation. Under some regularity conditions, the latent common components $\gamma_i' F_t$ can be approximated as a linear combination of the sample cross-sectional averages of x_{it} and y_{it}. The so-called Common Correlated Effects (CCE hereafter) estimator, a simple and intuitive estimation method proposed by Pesaran (2006), has been frequently used in practice. Along with the CCE estimator, empirical researchers have also used the Interactive Fixed Effects estimator (IE hereafter) developed by Bai (2009). The IE estimator approximates the latent common factors to regression errors by using the Principal Components (PC) estimation. See Reese and Westerlund (2018) and Hayakawa, Nagata, and Yamagata (2018) for more recent reference, and Chudik and Pesaran (2013) for a survey on this literature.

There are broadly three types of pre-tests available in the literature. The first two types are proposed by Bai (2009): a 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 Castagnetti, Rossi, and Trapani (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 method is based on estimation 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 within group transformation successfully eliminates unknown common factors. Throughout the paper, we will call this method the BPS method. The last method is a direct test proposed by Castagnetti, Rossi, and Trapani (2015b, CRT hereafter). The CRT method tests whether or not the maximum of estimated $\hat{\gamma}_i$ is significantly different from the sample cross-sectional average of it.

Table 1: Efficiency Gain from Correct Pre-Testing ($T = 25$)

$y_{it} = x_{it} + u_{it}$, $u_{it} = F_t + \varepsilon_{it}$, $x_{it} = \psi_i F_t + x_{it}^o$, $\psi_i \sim \mathcal{N}\left(s, s^2\right)$

		$\mathrm{V}(\hat{\beta}_{\mathrm{cce,p}})/\mathrm{V}(\hat{\beta}_{\mathrm{tfe,p}})$				
n	$\mathrm{V}(\hat{\beta}_{\mathrm{tfe,p}})$	$s = 0$	$s = 1$	$s = 2$	$s = 5$	$s = 10$
25	1.738	1.080	2.036	2.954	5.698	10.17
50	0.873	1.093	2.112	3.078	5.937	10.62
100	0.416	1.089	2.036	2.954	5.734	10.30
200	0.207	1.101	2.162	3.225	6.189	11.45
500	0.081	1.099	2.171	3.143	6.286	11.00

Since the CCE estimator is consistent regardless of whether $\gamma_i \neq \gamma$ and the computational cost is minor, one may think that the TFE estimator does not need to be considered if the efficiency gain by using the TFE is small. We will show the efficiency gain in detail in the Monte Carlo simulation section later, but here we provide evidence of why we need a pre-testing procedure. Table 1 reports the efficiency gain, which is measured by the relative variance ratio of the TFE estimator, $\hat{\beta}_{\mathrm{tfe,p}}$, to the CCE estimator, $\hat{\beta}_{\mathrm{cce,p}}$. The underlying data generating process is given by $y_{it} = x_{it} + u_{it}$, $u_{it} = F_t + \varepsilon_{it}$, $x_{it} = \psi_i F_t + x_{it}^o$, $\psi_i \sim N\left(s, s^2\right)$, and other random variables are generated from a standard normal distribution. See (23) in the next section for the formula of the pooled CCE regression. As s increases, the regressor x_{it} is more cross-sectionally dependent. Both the TFE and the CCE estimators are not biased in this case. However, as shown in Table 1, the TFE becomes more efficient than the CCE as s increases. This is because CCE estimation eliminates all factors from y_{it} and x_{it}, so that the variance of $\hat{\beta}_{\text{cce,p}}$ is asymptotically independent of s. Meanwhile, the TFE eliminates the homogeneous factor in the regression errors, u_{it}, and regressor, x_{it}, but does not remove the heterogeneous common components in x_{it}. As a result, as s increases, the variance of the modified regressor increases, so that the TFE becomes more efficient.

The purpose of this paper is to provide asymptotic analyses of pre-testing procedures when the slope coefficients are either heterogeneous or homogeneous across cross-sectional units. To evaluate each pre-testing procedure, we derive asymptotic variances of estimators suggested by pre-testing procedures. Under homogeneity of the slope coefficients, both the BPS and the CRT methods detect precisely whether the factor loadings are homogeneous. However, the CRT test allows a false rejection of the null with probability α, which is equivalent to the size of the test. Due to this minor difference, the BPS method leads to more efficient estimation than the CRT method.

Under heterogeneity of the slope coefficients ($\beta_i \neq \beta$), the two pre-testing procedures lead to different results even when $\gamma_i = \gamma$. The BPS procedure requires the homogeneity restriction on the slope coefficients. Suppose that $\beta_i \neq \beta$, but one imposes the homogeneity restriction on the slope coefficients. In this case, the regression error includes the additional term of $(\beta_i - \beta)' x_{it}$. If regressors, x_{it}, have heterogeneous factor loadings, then the regression error includes heterogeneous factor loadings as well. Hence, the BPS method suggests the factor augmented estimation in (2) is preferrable even when $\gamma_i = \gamma$. Meanwhile, the CRT method examines whether the estimated factor loadings to the regression residuals are homogeneous, where the regression residuals are estimated with the heterogenous slope coefficients. If the homogeneity restriction on the slope coefficients is imposed, it is impossible to identify the source of heterogeneous factor loadings in the regression errors. Hence, the CRT method tests only whether $\gamma_i = \gamma$ regardless of $\beta_i \neq \beta$. If $\gamma_i = \gamma$, the CRT method suggests the TFE estimation in (1) even when $\beta_i \neq \beta$. As we mentioned earlier, the homogeneity restriction on β_i leads to heterogeneous factor loadings in the regression error. As a result, the CRT method may lead to inconsistent estimation when $\beta_i \neq \beta$ but $\gamma_i = \gamma$.

The literature of testing cross-sectional dependence is also indirectly relevant. 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 the BPS method. Asymptotic results under homogeneity and heterogeneity of slope coefficients 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 the Appendix.

2 Extant Pre-Testing Procedures

This section provides a short review on extant pre-testing procedures for panel factor augmented regressions, and discusses how to evaluate each pre-testing procedure.

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 of y_{it} from its time series mean. For example, $\tilde{y}_{it} = y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$, and $\tilde{x}_{it} = x_{it} - T^{-1}\sum_{t=1}^{T} x_{it}$. 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 TFE 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 number of time series observations 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. Castagnetti, Rossi, and Trapani (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 second test is a maximum value test proposed by CRT. Instead of (2), CRT consider the following factor augmented regression.

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

The basic idea of the CRT test is straightforward. The null hypothesis of the CRT test[1] is given by

$$\mathcal{H}_0 : \gamma_i = \gamma. \tag{7}$$

If $\gamma_i = \gamma$, then the number of common factors becomes one. For example, consider the case of the two factors, and rewrite $\gamma_i' F_t$ as $\gamma_{1i}F_{1t} + \gamma_{2i}F_{2t}$. If $\gamma_{1i} = \gamma_1$ and $\gamma_{2i} = \gamma_2$, then $\gamma_{1i}F_{1t} + \gamma_{2i}F_{2t} = \gamma_1 F_{1t} + \gamma_2 F_{2t} = F_{3t}$ where F_{3t} is a $(T \times 1)$ vector. If the estimated number of the common factors is more than one, then the null hypothesis is naturally rejected.

Here we provide a step-by-step procedure for the CRT test.

Step 1: Define $\hat{e}_{it} = y_{it} - \hat{a}_i - \hat{\beta}_i' x_{it}$ as the residual from the CCE regression for each i.

$$y_{it} = a_i + \beta_i' x_{it} + \delta_{x,i}' \bar{x}_t + \delta_{y,i} \bar{y}_t + u_{it}, \tag{8}$$

where $\bar{x}_t = n^{-1} \sum_{i=1}^n x_{it}$ and $\bar{y}_t = n^{-1} \sum_{i=1}^n y_{it}$. Alternatively, one can run the IE regression augmented with the PC estimators of F_t proposed by Song (2013).

Step 2: Set the number of common factors $r = 1$, and estimate γ_i by applying the PC estimator to $\hat{e}_{it}$.[2] Let $\hat{\gamma}_i$ be the PC estimator. Then construct the following Mahalanobis distance.[3]

$$\mathcal{O}_i = \left(\hat{\gamma}_i - \hat{\mu}_\gamma\right)^2 / \hat{\Sigma}_\gamma, \tag{9}$$

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

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

Step 3: Construct the following max-type test given by

$$\mathcal{S}_{\gamma,nT} = T \cdot \max_{1 \le i \le n} \mathcal{O}_i. \tag{10}$$

CRT show that the limiting distribution of $\mathcal{S}_{\gamma,nT}$ becomes a Gumbel distribution. The exact critical value, $c_{\alpha n}$, can be calculated by

$$c_{\alpha n} = 2 \ln n - \ln \ln n - 2 \ln \Gamma(1/2) - \ln \left| \ln (1 - \alpha) \right|^2, \tag{11}$$

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

[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 jointly since in practice, the null hypothesis of $\gamma_i = \gamma$ becomes of interest.

[2] As CRT (2015b) claim, there is no reason to test the null of homogeneous factor loadings when the number of common factors is more than one. To see this, let $u_{it} = \gamma_{1i}F_{1t} + \gamma_{2i}F_{2t} + \varepsilon_{it}$. Suppose that $\gamma_{1i} = \gamma_1$ and $\gamma_{2i} = \gamma_2$ for all i. Then u_{it} has a single factor, or $u_{it} = F_t + \varepsilon_{it}$ with $F_t = \gamma_1 F_{1t} + \gamma_2 F_{2t}$. If $\gamma_{1i} = \gamma_1$ but $\gamma_{2i} \neq \gamma_2$, then u_{it} has two factors.

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

Here we address an important issue. Suppose that one imposes the homogeneity restriction on (6). Then the factor augmented regression in (6) becomes

$$y_{it} = a_i + \beta' x_{it} + u_{it}, \text{ with } u_{it} = (\beta_i - \beta)' x_{it} + \gamma_i' F_t + \varepsilon_{it}. \tag{12}$$

Substituting (3) into (12) results in

$$u_{it} = \left[(\beta_i - \beta)' \Gamma_i' + \gamma_i' \right] F_t + (\beta_i - \beta)' \Psi_i' G_t + (\beta_i - \beta)' x_{it}^o + \varepsilon_{it}. \tag{13}$$

When $\beta_i = \beta$ for all i, it does not matter whether one imposes the homogeneity of β_i on (6) since $u_{it} = \gamma_i' F_t + \varepsilon_{it}$. A serious problem exists when $\beta_i \neq \beta$. When $\beta_i \neq \beta$ for some i, the regression error includes extra terms of $(\beta_i - \beta)' \Gamma_i' F_t + (\beta_i - \beta)' \Psi_i' G_t + (\beta_i - \beta)' x_{it}^o$. In this case, the CRT test is no longer directly testing the homogeneity of γ_i. That is, the CRT test fails to perform its own purpose, which is directly testing the homogeneity of γ_i. If $\beta_i \neq \beta$, but the homogeneous restriction is imposed on (6), then the CRT test will reject the null $(\gamma_i = \gamma)$, even when $\gamma_i = \gamma$.

2.3 BPS Procedure

Both Bai (2009) and Parker and Sul (2016) use the estimated number of common factors to evaluate the homogeneity of factor loadings. Suppose that a panel data w_{it} follows a single factor structure[4] given in

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

where w_{it}^o is a pure idiosyncratic term. Taking off the time series and cross-sectional averages yields

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

The homogeneity of γ_i leads to $\dot{w}_{it} = \dot{w}_{it}^o$. Define $\#(\tilde{w}_{it})$ and $\hat{\#}(\tilde{w}_{it})$ as the true and estimated number of common factors to $\tilde{w}_{it} = w_{it} - T^{-1} \sum_{t=1}^{T} w_{it}$, respectively. Then it becomes obvious that

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

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

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

with a proper information criterion.

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

[4] As we mentioned earlier, if the number of common factors to w_{it} is more than one, the factor loadings are heterogeneous.

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

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

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

Step 2: Use BN's IC_2 criterion to estimate the number of common factors with $\hat{u}_{it}$.[5]

From (13), it is easy to show only with $\gamma_i = \gamma$ and $\beta_i = \beta$ for all i, the number of common factors with $\dot{u}_{it}$ becomes zero.

$$\#\left(\dot{u}_{it}\right) = 0 \text{ if } \gamma_i = \gamma \;\&\; \beta_i = \beta. \tag{19}$$

Otherwise, the true number of common factors with $\dot{u}_{it}$ becomes a non-zero constant. That is,

$$\#\left(\dot{u}_{it}\right) \geq 1 \text{ if either } \gamma_i \neq \gamma \text{ or } \beta_i \neq \beta. \tag{20}$$

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

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

Note that instead of the CCEP and CCEMG, the pooled IE estimation by Bai (2009) and the heterogeneous IE estimator by Song (2013) can be used in (21), respectively.

If $\hat{\#}\left(\hat{u}_{it}\right) = 0$, then it implies that both $\beta_i = \beta$ and $\gamma_i = \gamma$. Hence in this case, the TFE regression in (18) or (1) should be run for the pooled estimation. 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 empirical studies have considered pooled estimation. Consider the following two choices we discussed in the Introduction.

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

Alternatively, researchers may be interested in an individual-specific estimator for the slope coefficient. In this case, the following two choices are considered.

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

[5] Sul (2019) reports that BN's IC_2 criterion performs best among other criteria considered by Bai and Ng (2002).

The first and second regressions for each case yield 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.[6]

Table 2 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, and the CRT method does not impose the restriction, the pre-testing results do not alter whether or not empirical researchers are interested in either the 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 asymptotically. Meanwhile the CRT method precisely differentiates the heterogeneous γ_i from the case of the homogeneous factor loadings. Hence, the two pre-tests recommend different outcomes when $\beta_i \neq \beta$ but $\gamma_i = \gamma$. The CRT procedure suggests TFE, while the BPS recommends CCE. If empirical researchers are interested in pooling regressions, then the BPS method provides a 'correct' answer in this case since the regression error, $\dot{u}_{it}$, 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.

Table 2: 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

The second difference between the two pre-tests is not shown in Table 2. 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 of selecting a correct number of common factors becomes unity. Meanwhile the CRT method is a well constructed test, so that it makes a mistake with probability α, where α is the significance level. This difference is minor, but in the Monte

[6] Note that instead of the CCE estimators, one may consider IE least squares estimators suggested by Bai (2009). However, in this paper we consider only the CCE estimators to avoid any issues related to weak factors. When both x_{it} and u_{it} in (4) have weak factors, it is well known that Bai's estimator becomes inconsistent. Meanwhile, the CCE estimator is still consistent in this case.

Carlo simulation, this difference matters somewhat significantly.

In the next section, we will provide asymptotic comparisons 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 case in which $\beta_i \neq \beta$ for some i. As we discussed in the previous section, the results of the asymptotic comparisons are hinging on the assumption of the slope coefficients. Since it is unknown whether or not $\beta_i = \beta$, an overall comparison will be made at the end of this section.

We take the following assumptions.

Assumption 1 (Common Factors)

(i) $\exists M > 0$, $\mathbb{E}\left\|F_t\right\|^{12} < M$ and $\mathbb{E}\left\|G_t\right\|^{12} < M$.

(ii) *The unobserved common factors, F_t and G_t, are distributed independently of ε_{it} and x_{is}^o for all i, t and s.*

(iii) *As $T \to \infty$, $T^{-1}\sum_{t=1}^{T} F_t F_t' \to^p \Sigma_F > 0$ and $T^{-1}\sum_{t=1}^{T} G_t G_t' \to^p \Sigma_G > 0$ for some $r \times r$ and $m \times m$ matrices Σ_F and Σ_G.*

Assumption 2 (Individual-Specific Error) $\exists M > 0$,

(i) $\mathbb{E}\left\|x_{it}^o\right\|^{12} < M$.

(ii) $\mathbb{E}\left(\varepsilon_{it} x_{js}^o\right) = 0$, *for all i, j, s and t.*

(iii) $\mathbb{E}\left|\sum_{t=1}^{T} x_{it}\varepsilon_{it}\right|^r \leq M\mathbb{E}\left|\sum_{t=1}^{T}\left(x_{it}\varepsilon_{it}\right)^2\right|^{r/2}$ for all i, $r < 6$.

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) $\exists M, M_1 > 0$,

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

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

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

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

(iv)

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

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

(v) $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\varepsilon_{it} - n^{-1}\sum_{i=1}^{n}\varepsilon_{it} - T^{-1}\sum_{t=1}^{T}\varepsilon_{it} + (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}\right)^2 \to^p M_1 > 0.$

(vi) $\mathbb{E}\left|\sum_{t=1}^{T}\varepsilon_{it}\right|^r \le M\mathbb{E}\left|\sum_{t=1}^{T}\varepsilon_{it}\right|^{r/2}$ for all i, $r < 12$; $\mathbb{E}\left|\sum_{i=1}^{n}\varepsilon_{it}\right|^r \le M\mathbb{E}\left|\sum_{i=1}^{n}\varepsilon_{it}\right|^{r/2}$ for all t, $r < 12$.

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 $\|\beta\| < M$.

Assumption 7 (Identification)

(i) *Let* $X_i = [x_{i1}, \cdots, x_{iT}]'$, $Y_i = [y_{i1}, \cdots, y_{iT}]'$, $z_{it} = [y_{it}, x_{it}']'$, $M_z = I_T - \bar{Z}(\bar{Z}'\bar{Z})^{-}\bar{Z}'$, $\bar{Z} = [\bar{z}_1, \cdots, \bar{z}_T]'$, and $\bar{z}_t = n^{-1}\sum_{i=1}^{n} z_{it}$, *where* $(\bar{Z}'\bar{Z})^{-}$ *is the generalized inverse of* $\bar{Z}'\bar{Z}$*. The* $k \times k$ *matrices* $(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'$, $T^{-1}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'$, $T^{-1}X_i'M_zX_i$ *and* $(nT)^{-1}\sum_{i=1}^{n}X_i'M_zX_i$ *are full rank.*

(ii) Let $F = (F_1, \cdots, F_T)'$, $G = (G_1, \cdots, G_T)'$, $P = (F, G)$, $M_P = I_T - P(P'P)^{-1}P'$, and $M_{X_i} = I_T - X_i(X_i'X_i)^{-1}X_i'$. The $k \times k$ matrices $T^{-1}(X_i'M_PX_i)$ and $T^{-1}(P'M_{X_i}P)$ are full rank.

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 et al. (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. Assumption 7 (i) rules out the possibility that the defactored regressors become rank deficient. The existence of the 12−th moment of F_t, G_t, x_{it}^o and ε_{it} is required to establish the consistency of the principal component estimator $\hat{\gamma}_i$ in Step 2 of the CRT test. Assumption 7 (ii) ensures the identification of γ_i in this step. Further, note that two additional assumptions are required for the consistency of the CRT test. See Appendix A for the additional conditions.

We define the two pooled estimators as

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

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

The MG estimators are defined as

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

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

3.1 Under the Homogeneity of Slope Coefficients

When $\beta_i = \beta$, both the BPS and the CRT methods provide the same answer as Table 1 showed. In practice, it is more realistic that $\gamma_i \neq \gamma$ for a few individuals. Also, the case where the variation of γ_i is small enough not to influence the consistency of the TFE cannot be ruled out. To consider these cases formally, we define the following notion of local heterogeneity.

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

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

where

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

where the number of individuals in $\mathcal{G}^c$ is a fixed number 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 if $\gamma = 0$. Note that

$$\mathbb{E}\frac{1}{n}\sum\nolimits_{i=1}^{n}\left(\gamma_i - \gamma\right)^2 = \upsilon\Omega_0/n. \tag{31}$$

This implies that as $n \to \infty$, the variance of γ_i goes to zero. The condition in (31) states that the common factors F_t are weak factors if $\gamma = 0$. Meanwhile the weak factors do not imply local heterogeneity. For example, Reese and Westerlund (2015) consider the following notion of the weak factors when $\gamma = 0$.

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

Under (32), as $n \to \infty$, the maximum of τ_i also converges to zero (or $\gamma = 0$). In this case, CRT's max-type test fails.[7] Because of the same reason, CRT assume no weak factor given in (32).

Next, we will study the asymptotic behaviors 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 Assumptions 1-7,*

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

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

(ii) *additionally, if Assumptions 8-10 hold, 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 and Assumptions 8-9 are given in Appendix A. Here we provide an intuitive explanation. Under the homogeneity of factor loadings, the residuals, $\hat{u}_{it}$, do not include any factor so that the estimated number of common factors becomes zero. The point of interest here is whether a few factor loadings are different from the rest of them. As Parker and Sul (2016) showed, BN's (2002) information criteria (IC) are not precise enough to detect weak factors. Under the local heterogeneity (29) and (30), the demeaned factor loading is given by

$$\breve{\gamma}_i = \gamma_i - \frac{1}{n}\sum\nolimits_{i=1}^{n}\gamma_i = \begin{cases} O_p\left(n^{-1}\right) & \text{if } i \in \mathcal{G} \\ O_p(1) & \text{if } i \in \mathcal{G}^c \end{cases}.$$

[7] 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

$$\frac{T}{\ln n}\left\|\epsilon_i\right\|^2 = \frac{T}{n\ln n}O_p(1) \to 0,$$

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

As a result, $n^{-1}\sum_{i=1}^{n}\breve{\gamma}_i\breve{\gamma}_i' = O_p\left(n^{-1}\right)$, which is too small for IC to detect. Moreover, 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 an $O_p\left(n^{-1}\right)$ bias under the local heterogeneity. Hence, the regression residuals contain only weak factors:

$$\hat{u}_{it} = \dot{y}_{it} - \hat{\beta}_{\text{tfe,p}}'\dot{x}_{it} = (\beta - \hat{\beta}_{\text{tfe,p}})'\dot{x}_{it} + \breve{\gamma}_i'\tilde{F}_t + \dot{\varepsilon}_{it}.$$

Bai and Ng's IC cannot detect any weak common factor even with very large n and T. Meanwhile, the CRT test is based on the maximum value of Mahalanobis distances. The maximum value is, of course, very sensitive to a non-zero τ_i in (30). Even when only one of γ_i is different from the rest of $\gamma_i = \gamma$, the CRT method rejects the null of homogenous factor loadings. Hence, the CRT method detects the local heterogeneity precisely as $n, T \to \infty$.

Next, we compare the asymptotic variances of the TFE and the CCE estimators under the homogeneity of γ_i. Under suitable conditions, both the TFE and the CCE estimators are consistent since asymptotically the modified regressors are independent of the modified regression errors. Let $\dot{X}_i = [\dot{x}_{i1}, \cdots, \dot{x}_{iT}]'$, $X_i^o = [x_{i1}^o, \cdots, x_{iT}^o]'$ and $\varepsilon_i = [\varepsilon_{i1}, \cdots, \varepsilon_{iT}]'$. Further define $\Omega_{\text{cce,p}}$ and $\Omega_{\text{tfe,p}}$ as

$$\Omega_{\text{cce,p}} = \mathbb{E}\frac{1}{nT}\sum_{i=1}^{n} X_i^{o\prime}\varepsilon_i\varepsilon_i' X_i^o, \ \& \ \Omega_{\text{tfe,p}} = \mathbb{E}\frac{1}{nT}\sum_{i=1}^{n} \dot{X}_i'\varepsilon_i\varepsilon_i'\dot{X}_i, \tag{35}$$

and

$$Q_{\text{cce,p}} = \text{plim}_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n} X_i^{o\prime}X_i^o, \ \& \ Q_{\text{tfe,p}} = \text{plim}_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n} \dot{X}_i'\dot{X}_i. \tag{36}$$

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 a larger covariance matrix of the modified regressors. This difference makes the TFE estimator more efficient than the CCE estimator in general. Only when the two-way within group transformation eliminates the common components of x_{it} effectively, does the discrepancy between two variances become zero asymptotically.

Define the asymptotic variances of TFE and CCE pooled estimators as

$$V_{\ell,\text{p}} = Q_{\ell,\text{p}}^{-1}\Omega_{\ell,\text{p}}Q_{\ell,\text{p}}^{-1}, \tag{37}$$

for $\ell \in \{\text{cce, tfe}\}$. It is easy to show that as $n, T \to \infty$, under i.i.d. assumption of ε_{it} over i and t, the difference between $V_{\text{cce,p}}$ and $V_{\text{tfe,p}}$ becomes non-negative definite. That is,

$$V_{\text{cce,p}} - V_{\text{tfe,p}} \geq 0. \tag{38}$$

The equality holds only when the factor loadings to x_{it} are homogeneous.

In Theorem 2, we compare 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 probability limits of the denominator terms for the CCE are smaller than those of TFE estimators in general as $n, T \to \infty$ with $T/n \to 0$.

Theorem 2 (Comparison of Asymptotic Variances) *Assume that Assumptions 1-7 hold, and further assume that ε_{it} is i.i.d. over i and t. As $n, T \to \infty$ with $T/n \to 0$, the asymptotic variances satisfy that*

$$V_{\text{cce,p}} - V_{\text{tfe,p}} \geq 0. \tag{39}$$

See Appendix B for the proof of Theorem 2. Note that Pesaran (2006) already showed the asymptotic variance of the CCE pooled estimator. Here we fix the weight function in Pesaran (2006) at $1/n$. 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 i, it is not easy to show that Theorem 2 holds unless we know the weak dependence structure. The equality holds only when $\breve{\Gamma}_i = \breve{\Psi}_i = 0$ for all i.

Next, we consider the mean group estimators,

$$\hat{\beta}_{\ell,\text{mg}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_{\ell,i}, \text{ with } \ell \in \{\text{cce,tfe}\}.$$

It is well known that the pooled estimator can be re-written as

$$\hat{\beta}_{\ell,\text{p}} = \left(\sum_{i=1}^{n} W_{\ell,i}\right)^{-1}\left(\sum_{i=1}^{n} W_{\ell,i}\hat{\beta}_{\ell,i}\right), \text{ with } \ell \in \{\text{cce,tfe}\}, \tag{40}$$

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

$$W_{\ell,i} = \begin{cases} T^{-1}X_i^{o\prime}X_i^{o} & \text{if } \ell = \text{cce} \\ T^{-1}\dot{X}_i^{\prime}\dot{X}_i & \text{if } \ell = \text{tfe} \end{cases}. \tag{41}$$

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 i and t.

Finally, we combine the results of Theorem 1 and Theorem 2 together. Define the BPS and CRT estimators as

$$\hat{\beta}_{\text{BPS}} = \begin{cases} \hat{\beta}_{\text{tfe,p}} & \text{if } \hat{\#}(\hat{u}_{it}) = 0 \\ \hat{\beta}_{\text{cce,p}} & \text{if } \hat{\#}(\hat{u}_{it}) \neq 0 \end{cases}, \text{ and } \hat{\beta}_{\text{CRT}} = \begin{cases} \hat{\beta}_{\text{tfe,p}} & \text{if } \mathcal{S}_{\gamma,nT} \leq c_{\alpha n} \\ \hat{\beta}_{\text{cce,p}} & \text{if } \mathcal{S}_{\gamma,nT} > c_{\alpha n} \end{cases},$$

or alternatively they can be rewritten as

$$\hat{\beta}_{\text{m,p}} = \omega_{\text{m}}\hat{\beta}_{\text{tfe,p}} + (1-\omega_{\text{m}})\hat{\beta}_{\text{cce,p}}, \text{ with m} \in \{\text{BPS, CRT}\}, \tag{42}$$

where $\omega_{BPS} = 1[\hat{\#}(\hat{u}_{it}) = 0]$ and $\omega_{CRT} = 1\left(\mathcal{S}_{\gamma,nT} \le c_{\alpha n}\right)$. Note that $1\left(\cdot\right)$ is an indicator function, so that $\hat{\beta}_{\text{m,p}}$ is not a weighted average of $\hat{\beta}_{\text{tfe,p}}$ and $\hat{\beta}_{\text{cce,p}}$. The asymptotic variances of the BPS and CRT estimators can be written as

$$V(\hat{\beta}_{\text{m,p}}) = \omega_{\text{m}} V_{\text{tfe,p}} + (1-\omega_{\text{m}}) V_{\text{cce,p}}. \tag{43}$$

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

$$\hat{\beta}_{\text{m,mg}} = \omega_{\text{m}} \hat{\beta}_{\text{tfe,mg}} + (1-\omega_{\text{m}}) \hat{\beta}_{\text{cce,mg}}, \text{ with m} \in \{\text{BPS, CRT}\}. \tag{44}$$

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 1-10,* as $n, T \to \infty$ *with* $T/n \to 0$ and $n/T^3 \to 0$,

$$V_{\text{CRT,p}} - V_{\text{BPS,p}} \ge 0, \quad \& \quad V_{\text{CRT,mg}} - V_{\text{BPS,mg}} \ge 0. \tag{45}$$

See Appendix C for the proof of Theorem 3. Note that Theorem 3 holds when $\beta_i = \beta$ by Assumption 6. The equality holds if $\omega_{BPS} = \omega_{CRT}$. There are two cases in which the equality always holds. The first case is when regressors have the same or zero factor loadings ($\breve{\Gamma}_i = \breve{\Psi}_i = 0$ for all i). In this case, the BPS estimator becomes equivalent to the CRT estimator. The second case is when $\gamma_i \ne \gamma$ for all i. In this case, the equality holds since the power of the CRT test becomes unity as $n, T \to \infty$. 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 probability α. Lastly, under the local heterogeneity of γ_i, the inequality holds since asymptotically ω_{CRT} converges to unity, but ω_{BPS} converges to zero.

The next subsection considers the case where $\beta_i \ne \beta$.

3.2 Under the Heterogeneity of Slope Coefficients

To investigate the heterogeneous slope coefficients case, we change Assumption 6 to 6A.

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

$$\beta_i = \beta + \eta_i, \text{ with } \eta_i \sim iid\left(0, \Omega_\eta\right), \tag{46}$$

where $\|\beta\| < M$*,* $\|\Omega_\eta\| < M$*,* Ω_η *is a* $k \times k$ *symmetric non-negative definite matrix, and*

(ii) *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 a standard assumption for the heterogeneous slope coefficients. Note that we particularly need the independence between η_i and $x_{it}x_{it}'$. Otherwise, any pooled estimator leads to inconsistency due to the correlation between the weights in (40) and β_i.

It is very important to note that if $\beta_i \neq \beta$ for some i, the regression error has heterogeneous factor loading coefficients even when $\gamma_i = \gamma$ for all i. Suppose that empirical researchers are interested only in pooled estimators even with $\beta_i \neq \beta$. Imposing the homogeneity restriction on the slope coefficients leads to

$$y_{it} = a_i + \beta' x_{it} + u_{it}, \text{ with } u_{it} = \left(\eta_i'\Gamma_i' + \gamma_i'\right) F_t + \eta_i'\Psi_i' G_t + \xi_{it}, \tag{47}$$

where $\eta_i = \beta_i - \beta$ and $\xi_{it} = \eta_i' x_{it}^o + \varepsilon_{it}$.

Recall that the CRT method in (6) is based on the panel regression without imposing the homogeneity restriction of $\beta_i = \beta$. In contrast, the BPS method requires imposing the homogeneity restriction. As shown in Table 2, when $\beta_i \neq \beta$, the BPS suggests the CCE regardless of $\gamma_i \neq \gamma$. It is natural since u_{it} in (47) has heterogeneous factor loadings as long as $\beta_i \neq \beta$. Meanwhile, the CRT's max-type test examines only whether or not $\gamma_i = \gamma$. Even when $\gamma_i = \gamma$, as shown in (47), the regression error, u_{it}, includes multiple factors if either $\Gamma_i \neq 0$ or $\Psi_i \neq 0$. If Assumption 6A (ii) is violated, then the TFE pooled estimator becomes inconsistent, so that the CRT method leads to an inconsistent estimation. Only when $\Gamma_i = \Gamma$ and $\Psi_i = \Psi$ for all i, does the TFE pooled estimator become consistent. However, even in this case, the CCE pooled estimator could be more efficient. To see this, let $M_t = \Gamma' F_t + \Psi' G_t$, and rewrite (47) as $\dot{y}_{it} = \beta' \dot{x}_{it} + \dot{u}_{it}$, with $\dot{u}_{it} = \left(\beta_i - n^{-1}\sum_{i=1}^n \beta_i\right)' \tilde{M}_t + \dot{\xi}_{it}$. The TFE error, $\dot{u}_{it}$, still has a single factor $\tilde{M}_t$. Meanwhile, the CCE error, ξ_{it}, does not have any factor structure.

If either $\Gamma_i \neq \Gamma$ or $\Psi_i \neq \Psi$, but Assumption 6A (ii) holds, then the TFE pooled estimator becomes consistent. However, as shown in Appendix D, the TFE pooled estimator may not be efficient compared with the CCE pooled estimator. See next Monte Carlo simulation section for more detailed discussions.

Next, we consider the MG estimation, which is an alternative way to pool the cross-sectional and time series information as shown in (40). The only difference between the MG and pooled estimators is weight functions[8]: the MG estimation assigns an equal weight, $1/n$, meanwhile the pooled estimation assigns heavier weights if the variances of regressors are larger. We consider the case where the MG estimation becomes of interest to empirical researchers. Suppose that the

[8] See Lee and Sul (2020b) for the asymptotic comparison between the MG and the conventional pooled estimations.

CRT's test does not reject the null of $\gamma_i = \gamma$. Then the TFE MG estimator in (27) is expected to be used. That is, the following regression is supposed to be run.

$$y_{it} = a_i + \beta_i' x_{it} + F_t + \varepsilon_{it}. \tag{48}$$

Interestingly, it is not straightforward to run (48). The typical two-way fixed effects transformation leads to

$$\dot{y}_{it} = \beta_i' \dot{x}_{it} + e_{it}, \tag{49}$$

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 (49). The existence of ξ_{it} influences the asymptotic variance of the TFE MG estimator. There are various other ways to reduce the asymptotic variance. For example, an iterative method might work in this case. Let $\hat{\beta}_i^1$ be the first stage estimator for each i based on (49). Next, estimate the common factor by taking the cross-sectional average of the following residuals.

$$\hat{F}_{t,c} = \frac{1}{n} \sum\nolimits_{i=1}^{n} \left(\tilde{y}_{it} - \hat{\beta}_i^{1\prime} \tilde{x}_{it} \right). \tag{50}$$

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

$$\tilde{y}_{it} - \hat{F}_{t,c} = \beta_i' \tilde{x}_{it} + \text{error}_{it} \tag{51}$$

Repeating (50) and (51) 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 MG estimator anymore.

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

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

$$y_{it} = a_i + \rho y_{it-1} + \lambda_i' F_t + \varepsilon_{it}, \tag{52}$$

or

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

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

$$\hat{u}_{it} = \tilde{y}_{it} - \hat{\rho}_{\text{fe}} \tilde{y}_{it-1} = \tilde{\varepsilon}_{it} + \lambda_i' \tilde{F}_t + (\rho - \hat{\rho}_{\text{fe}}) \tilde{y}_{it-1} = \tilde{\varepsilon}_{it}^* + \lambda_i' \tilde{F}_t^*, \tag{53}$$

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 as $n, T \to \infty$,

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

Meanwhile if $\lambda_i \neq \lambda$, then it is easy to show that

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

Appendix E provides a general proof of remark 1 in the dynamic panel models with weakly exogenous (or predetermined) regressors. Remark 1 states that the BPS method works continuously in dynamic panel regressions. However, to achieve a more efficient estimation, one should use a bias corrected estimation method. See Chudik et al. (2018) for more discussion.

Lastly, we discuss how to implement the BPS method in unbalanced panels in Appendix F. Specific Stata estimation methods and commands are also described.

4 Monte Carlo Simulations and an Empirical Example

This section consists of two subsections. The first subsection examines theoretical findings of this paper, and investigates how the pre-testing methods perform in finite samples by means of Monte Carlo simulations. The second subsection demonstrates the usefulness of the suggested method with violent crime rates across 48 US contiguous states.

4.1 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: $\lambda_{j,i} \neq 0$, $\delta_{j,i} \neq 0$ v.s. $\lambda_{j,i} = \delta_{j,i} = 0$. In the first case, both regressors have two common factors. The second case does not allow any cross sectional dependence in the regressors. All common factors, ε_{it}, $x_{j,it}^o$ are drawn from $\mathcal{N}(0, 1)$, and factor loadings are drawn from $\mathcal{N}(1, 1)$. 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 in the case of the homogeneous slope coefficients.

We set $\beta_{1,i} = \beta_{2,i} = 1$. Table 3 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 a somewhat 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 reaches 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 we discussed in Introduction and Theorem 3, the variance of the BPS pooled estimator is always the smallest among three pooled estimators when the factor loadings in regressors, $x_{j,it}$, are heterogeneous (or the signal-to-noisy ratio is strong). Only when the regressors do not have any factor structure ($\lambda_{j,i} = \delta_{j,i} = 0$) or the signal-to-noisy rate is very weak, the variances of other pooled estimators are similar to the variance of the BPS pooled estimator. See the online supplementary appendices for more detailed evidence. Meanwhile, the variance of the BPS MG estimator is more or less similar to that of the CRT MG estimator. The CCE pooled and MG estimators are robust but the least efficient.

Table 4 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 becoming 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. A similar finding is observed for the case of the MG estimation.

Table 5 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 the CCE or CRT estimator. By combining all results from Table 3, 4 and 5, we can confirm our theoretical findings in Theorem 3.

Next, we investigate the finite sample performance under heterogeneous slope coefficients. Table 6 reports the case where $\beta_i \neq \beta$ but $\gamma_i = \gamma$. As shown in Table 3, the BPS method suggests the

CCE estimator, while 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 in Appendix D, both the BPS pooled and MG estimators are more efficient than the CRT pooled and MG estimators.

Table 7 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.

4.2 Empirical Example: Determination of Violent Crimes

This subsection demonstrates how to use the pre-testing procedures studied in the previous section in practice. Levitt (1997) investigated the determinants of the change in the violent crimes (C_{it}) across cities in the US. The main variable of interest was the number of sworn police officers per capita. Figure 1 shows national trends of violent crime rates and the number of sworn police officers per capita. We take the logarithm of all variables. As shown in Figure 1, the crime rates had been increasing until 1991, and have been declining continuously since then. Meanwhile, the number of police officers had been continuously increased until 2010.

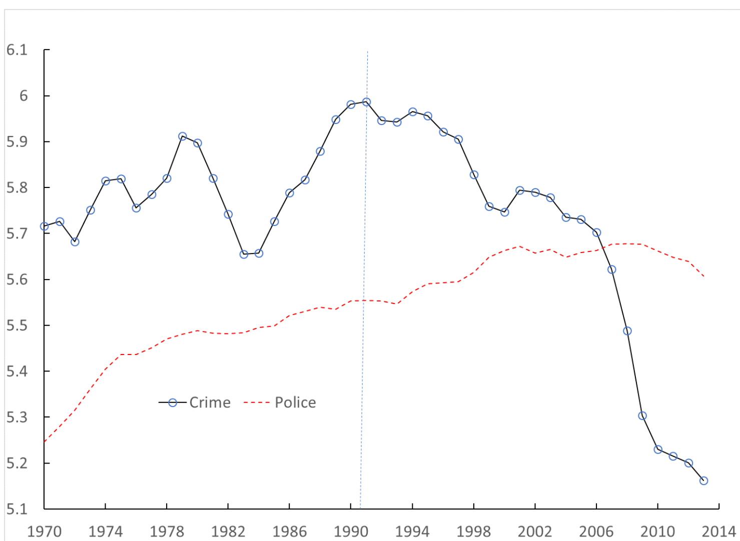

Figure 1: National Trends of Violent Crime and the number of Police Offiers

Levitt used several control variables. Furthermore, Levitt used contemporaneous variables as regressors, which may invite the non-zero correlations between regressors and regression errors. To avoid such endogeneity, Levitt used various election variables as IVs. Here we use pre-determined variables to detour this issue. Among various control variables, we select only two control variables: unemployment rates (U_{it}) and the percentage of the black population (B_{it}). Adding more control variables does not change the results dramatically. The annual violent crimes and the number of sworn police officers (P_{it}) across 48 contiguous states from 1970 to 2013 are collected from the uniform crime statistics reported by the FBI. Unemployment rates and black population are collected from the Bureau of Economic Analysis and Census survey, respectively. The following two modified regressions are run.

$$\Delta \ln C_{it} = a_i + F_t + \beta_{1,i}\Delta \ln P_{it-1} + \beta_{2,i}\Delta \ln U_{it-1} + \beta_{3,i}\Delta \ln B_{it-1} + \varepsilon_{1,it}, \quad (54)$$

$$\Delta \ln C_{it} = a_i + \beta_{1,i}\Delta \ln P_{it-1} + \beta_{2,i}\Delta \ln U_{it-1} + \beta_{3,i}\Delta \ln B_{it-1} + \lambda_i' \bar{F}_{n,t} + \varepsilon_{2,it}, \quad (55)$$

where $\bar{F}_{n,t}$ is a vector of the cross-sectional averages of regressand and regressors. We follow the CRT's procedure from (8) to (11), and get the CRT's test statistics: $\mathcal{S}_{\gamma,nT} = 13.99$, but the 5% critical value is 11.28. Hence we reject the null of the homogeneous factor loadings. The CRT procedure suggests using the CCE estimation. Meanwhile, with homogeneous restrictions on all slope coefficients except for a_i, the residuals from (54) do not have any factor so that the BPS' procedure suggests running TFE regressions.

Table 8: Comparison between TFE and CCE Estimations

Sample: 1970-2013	WG				MG			
Variables	TFE	*t-ratio*	CCE	*t-ratio*	TFE	*t-ratio*	CCE	*t-ratio*
$\Delta \ln P_{it-1}$	0.018	0.204	0.022	0.256	-0.060	-0.875	-0.057	-0.763
$\Delta \ln U_{it-1}$	**-0.073**	-3.894	**-0.070**	-3.786	**-0.070**	-3.516	**-0.076**	-3.647
$\Delta \ln B_{it-1}$	**-0.748**	-3.015	-0.538	-1.464	**-0.634**	-2.241	-0.458	-1.044
Sample: 1991-2013								
$\Delta \ln P_{it-1}$	0.119	1.526	0.102	1.381	0.088	1.127	0.088	0.944
$\Delta \ln U_{it-1}$	-0.027	-0.857	-0.024	-0.653	-0.023	-0.684	-0.040	-1.082
$\Delta \ln B_{it-1}$	-0.363	-0.668	0.468	0.590	0.167	0.286	0.668	0.686

Table 8 reports the pooled and mean group TFE and CCE estimates. Interestingly, regardless of the choice of estimations, the estimates of the slope coefficient, β_1, on $\Delta \ln P_{it-1}$ become insignificant. For the WG estimation, the panel robust covariance estimation is used for the $t-$ratios. The WG estimators for β_1 are all positive, but the MG estimators are all negative. But they are not significantly different from zero. Note that Levitt (1997) reported the negative, but significant estimated slope coefficient on $\Delta \ln P_{it-1}$. Nonetheless the difference between the TFE and CCE estimation can be found on the estimated β_3. The TFE estimates are slightly larger in absolute value: $\hat{\beta}_{\text{tfe,p}}$ is -0.748, but $\hat{\beta}_{\text{cce,p}}$ is -0.538. Similarly $\hat{\beta}_{\text{tfe,mg}}$ is -0.634, but $\hat{\beta}_{\text{cce,mg}}$ is just -0.458. Both CCE pooled and mean group estimates are insignificant.

Since the negative correlation between the crime rates and the number of sworn police officers is found after 1991, we consider the sub-sample period from 1991 to 2013. We re-examined the two pre-testing procedures. The BPS' procedure continuously suggests the TFE. The CRT's test statistics are now changed to 7.66, so that the CRT test suggests the TFE as well. However, we cannot find any empirical evidence that the growth rate of the number of police officers is negatively correlated with the growth rate of the crime rates. In fact, the point estimate of $\hat{\beta}_{1,\mathrm{tfe,p}}$ is positive, and more importantly, becomes significant at the 20% level. Since the FBI crime rates are all rates reported by local and federal agencies, this result may be interpreted that the more police officers are hired, the more criminals are captured.

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 the 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 running a correctly specified regression. Meanwhile, the original CRT method fails to suggest one under the homogeneous factor loadings case. We did not consider altering 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.

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 be 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). Testing for No Factor Structures: On the Use of Hausman-type Statistics. *Economics Letters*, 130(C), 66-68.

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

[7] Chudik, A., and Pesaran, M. H. (2013). Large Panel Data Models with Cross-Sectional Dependence: A Survey. *CESifo Working Paper Series No. 4371*.

[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] Chudik, A., Pesaran, M. H., and Yang, J-C. (2018). Half-Panel Jackknife Fixed-Effects Estimation of Linear Panels with Weakly Exogenous Regressors. *Journal of Applied Econometrics*, 33, 816–836.

[10] Hansen, B. E. (2020). Econometrics [Unpublished manuscript]. Department of Economics, University of Wisconsin, Madison, WI.

[11] 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*.

[12] Jolliffe, I. T. (2002). Principal Component Analysis. New York: Springer-Verlag.

[13] Lee, Y., and Sul, D. (2020a). Depth-Weighted Estimation of Heterogenous Agent Panel Data Models. Mimeo, Syracuse University.

[14] Lee, Y., and Sul, D. (2020b). Trimmed Mean Group Estimation. Mimeo, University of Texas at Dallas.

[15] Levitt, S. (1997). Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *American Economic Review*, 87(3), 270-90.

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

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

[18] 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).

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

[20] Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics.* New York: Oxford University Press.

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

[22] 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.

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

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

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

[26] 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.

[27] Song, M. (2013). Essays on Large Panel Data Analysis (Doctoral dissertation). Department of Economics, Columbia University, New York, NY.

[28] Stock, J. H., and Watson, M. W. (1998). Diffusion Indexes (NBER Working Paper 6702). Cambridge, MA: National Bureau of Economic Research.

[29] Sul, D. (2019). *Panel Data Econometrics: Common Factor Analysis for Empirical Researchers.* New York: Routledge.

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

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

Technical Appendix

Appendix A: Proof of Theorem 1

The consistency of the CRT test requires the following conditions.

Assumption 8 (Serial Dependence)

(i) Let $\delta > 0$ and $\alpha \in (1, +\infty)$. ε_{it}, F_t, and x_{it} are $L_{2+\delta}$-NED (Near Epoch Dependent) of size α on a uniform mixing base $\{v_t\}_{t=-\infty}^{+\infty}$ of size $-q/(q-2)$ and $q > \frac{2\alpha-1}{\alpha-1}$.

(ii) Let $V_{iT}^{F\varepsilon} := T^{-1}\mathbb{E}\left[\left(\sum_{t=1}^{T} F_t\varepsilon_{it}\right)\left(\sum_{t=1}^{T} F_t\varepsilon_{it}\right)'\right]$. $V_{iT}^{F\varepsilon} > 0$ uniformly in T, and as $T \to \infty$, $V_{iT}^{F\varepsilon} \longrightarrow V_i^{F\varepsilon}$ with $\left\|V_i^{F\varepsilon}\right\| < \infty$. The same holds for $V_{iT}^{x\varepsilon} := T^{-1}\mathbb{E}\left[\left(\sum_{t=1}^{T} x_{it}\varepsilon_{it}\right)\left(\sum_{t=1}^{T} x_{it}\varepsilon_{it}\right)'\right]$, $V_{iT}^{Fx} := T^{-1}\mathbb{E}\left(\bar{\omega}_{iT}^{Fx}\bar{\omega}_{iT}^{Fx\prime}\right)$ with $\bar{\omega}_{iT}^{Fx} = vec\left(\sum_{t=1}^{T} F_t x_{it}'\right) - \mathbb{E}\left[vec\left(\sum_{t=1}^{T} F_t x_{it}'\right)\right]$, and $V_{iT}^{xx} := T^{-1}\mathbb{E}\left(\bar{\omega}_{iT}^{xx}\bar{\omega}_{iT}^{xx\prime}\right)$ with $\bar{\omega}_{iT}^{xx} = vec\left(\sum_{t=1}^{T} x_{it}x_{it}'\right) - \mathbb{E}\left[vec\left(\sum_{t=1}^{T} x_{it}x_{it}'\right)\right]$.

(iii) Let $\omega_{kt}^{F\varepsilon}$ be the kth element of $F_t\varepsilon_{it}$ and define $S_{kT,m}^{F\varepsilon} := \sum_{t=m+1}^{m+T}\omega_{kt}^{F\varepsilon}$. There exists a positive definite matrix $\bar{\Omega}^{F\varepsilon} = \left\{\varpi_{kh}^{F\varepsilon}\right\}$ such that $T^{-1}\left|\mathbb{E}\left(S_{kT,m}^{F\varepsilon}S_{hT,m}^{F\varepsilon}\right) - \varpi_{kh}^{F\varepsilon}\right| \leq MT^{-\psi}$, for all k and h and uniformly in m, with $\psi > 0$. The same holds for $x_{it}\varepsilon_{it}$.

Assumption 9 (Cross Sectional Dependence) It holds that $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\mathbb{E}\left(\varepsilon_{it}\varepsilon_{js}\right)\right|\ln n \longrightarrow 0$ as $n, T \to \infty$ for all $i \neq j$.

Assumptions 8 and 9 are identical to Assumptions 6-7 in CRT (2015b).

Part I: Proof of Theorem 1 (i)

First, we show

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\breve{\gamma}_i'\breve{\gamma}_i\right) = O\left(n^{-1}\right),$$

under the local heterogeneity defined in Definition 1. Without loss of generality, assume 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 demeaned factor loading is given by

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

in light of which, the following holds,

$$\mathbb{E}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\breve{\gamma}_i'\breve{\gamma}_i\right)=\frac{1}{n}\Omega_0+O_p\left(n^{-2}\right). \tag{56}$$

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

$$\hat{u}_{it}=\dot{y}_{it}-\hat{\beta}_{\text{tfe,p}}'\dot{x}_{it}=\left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\dot{x}_{it}+\dot{u}_{it}, \tag{57}$$

where

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

Consider the first term in (57). The TFE pooled estimator is given by

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

where

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

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

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

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. Then I is biased, and the order of which is given by

$$\mathbb{E}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\breve{\Gamma}_i'\breve{\gamma}_i\right)=q\frac{1}{n}\Omega_0+O\left(n^{-2}\right)=O\left(n^{-1}\right)$$

which is the same as the bias of CCEP estimator. Let $\Sigma_{\breve{\Gamma}}=\mathbb{E}\left(\breve{\Gamma}_i\breve{\Gamma}_i'\right)$, which is bounded by Assumption 3. If Γ_i is not correlated with γ_i, such that

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

then the following holds,

$$\mathbb{E}\left\|\frac{1}{n}\sum\nolimits_{i=1}^{n}\breve{\Gamma}_i'\breve{\gamma}_i\right\|^2=\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left(\breve{\gamma}_i'\breve{\Gamma}_i\breve{\Gamma}_i'\breve{\gamma}_i\right)=O\left(n^{-2}\right),$$

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

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

Putting all together, we have

$$\hat{\beta}_{\text{tfe,p}}-\beta=\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{\varepsilon}_{it}+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{n\sqrt{T}}\right). \tag{58}$$

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

$$\sqrt{nT}(\hat{\beta}_{\text{tfe,p}}-\beta)=\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}\dot{\varepsilon}_{it}+O_p\left(\sqrt{T/n}\right)+O_p\left(1/\sqrt{n}\right).$$

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

$$\dot{u}_{it}=\begin{cases} -\dfrac{1}{n}\tau_n'\tilde{F}_t+\dot{\varepsilon}_{it} & \text{if } i<n \\ \dfrac{n-1}{n}\tau_n'\tilde{F}_t+\dot{\varepsilon}_{it} & \text{if } i=n \end{cases}. \tag{59}$$

Let $\tilde{P}_t'=\left(\tilde{F}_t',\tilde{G}_t'\right)$.

$$\begin{aligned}
\hat{u}_{it} &= \begin{cases} \left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\left(\breve{\Gamma}_i'\tilde{F}_t+\breve{\Psi}_i'\tilde{G}_t\right)-\dfrac{1}{n}\tau_n'\tilde{F}_t+\left[\dot{\varepsilon}_{it}+\left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\dot{x}_{it}^o\right] & \text{if } i<n, \\ \left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\left(\breve{\Gamma}_i'\tilde{F}_t+\breve{\Psi}_i'\tilde{G}_t\right)+\dfrac{n-1}{n}\tau_n'\tilde{F}_t+\left[\dot{\varepsilon}_{it}+\left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\dot{x}_{it}^o\right] & \text{if } i=n, \end{cases} \\
&=: \Lambda_i'\tilde{P}_t+v_{it},
\end{aligned} \tag{60}$$

with

$$\begin{aligned}
\Lambda_i' &= \begin{cases} \left[\left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\breve{\Gamma}_i'-\dfrac{1}{n}\tau_n' \quad \left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\breve{\Psi}_i'\right] & \text{if } i<n, \\ \left[\left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\breve{\Gamma}_i'+\dfrac{n-1}{n}\tau_n' \quad \left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\breve{\Psi}_i'\right] & \text{if } i=n, \end{cases} \\
v_{it} &= \dot{\varepsilon}_{it}+\left(\beta-\hat{\beta}_{\text{tfe,p}}\right)'\dot{x}_{it}^o.
\end{aligned}$$

Eq. (58) and the local heterogeneity of the factor loadings imply that

$$\Lambda_i'\Lambda_i = \begin{cases} O_p\left(\frac{1}{n^2}\right) + O_p\left(\frac{1}{nT}\right) & \text{if } i < n, \\ O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) + \left(\frac{n-1}{n}\right)^2 \tau_n'\tau_n & \text{if } i = n. \end{cases} \tag{61}$$

Define $\kappa = \hat{\#}(\hat{u}_{it})$. Last, we need to show that under the local heterogeneity of γ_i,

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

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

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

where

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

Define the eigenvalues of a $n \times n$ matrix A as $\psi_1(A), \cdots, \psi_n(A)$, ordered from largest to smallest. Then we have

$$\hat{V}(0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{u}_{it}^2 = \sum_{j=1}^{\kappa}\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right),$$

$$\hat{V}(\kappa) = \sum_{j=\kappa+1}^{n}\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right),$$

by optimal algebraic properties of principal components (Jolliffe, 2002, pp. 13-15), with $\hat{u} = (\hat{u}_1, \cdots, \hat{u}_T)'$ and $\hat{u}_t = (\hat{u}_{1t}, \cdots, \hat{u}_{nt})'$. The penalty function of $IC_2(\kappa)$ satisfies that

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

Hence, it suffices to show that

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

By Assumption 4, $\hat{V}(0) = O_p(1)$ and is bounded away from zero,

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

for $\left(\hat{V}(\kappa)/\hat{V}(0)\right) > 0$. It is thus sufficient to show that

$$\begin{aligned} \hat{V}(\kappa) - \hat{V}(0) &= \sum_{j=\kappa+1}^{n}\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right) - \sum_{j=1}^{n}\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right) \\ &= -\sum_{j=1}^{\kappa}\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right) \leq O_p\left(\frac{1}{\min[n,T]}\right). \end{aligned}$$

Rewrite the residual given in (60) in matrix form

$$\hat{u} = v + \tilde{P}\Lambda',$$

where $\tilde{P} = \left(\tilde{P}_1, \cdots, \tilde{P}_T\right)'$, $\Lambda = (\Lambda_1, \cdots, \Lambda_n)'$, $v = (v_1, \cdots, v_T)'$, and $v_t = (v_{1t}, \cdots, v_{nt})'$. Note that for $T \times n$ matrices A and B for some $1 \leq T \leq n$, by the singular value version of Weyl inequality, we have

$$\sigma_{i+j-1}(A+B) \leq \sigma_i(A) + \sigma_j(B), \text{ for } 1 \leq i,\ j,\ \ (i+j-1) \leq T,$$

where $\sigma_i(\cdot)$ denotes the ith singular value. Let $i = 1$, $A = v$ and $B = \tilde{P}\Lambda'$, for $j = 1, \cdots, \kappa$,

$$\sigma_j(\hat{u}) \leq \sigma_1(v) + \sigma_j\left(\tilde{P}\Lambda'\right).$$

Since $\sigma_i(A) = \sqrt{\psi_i(A'A)}$,

$$\sqrt{\psi_j(\hat{u}'\hat{u})} \leq \sqrt{\psi_1(v'v)} + \sqrt{\psi_j\left(\Lambda\tilde{P}'\tilde{P}\Lambda'\right)}.$$

Hence,

$$\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right) \leq \psi_j\left(\frac{\Lambda\tilde{P}'\tilde{P}\Lambda'}{nT}\right) + \psi_1\left(\frac{v'v}{nT}\right) + 2\sqrt{\psi_j\left(\frac{\Lambda\tilde{P}'\tilde{P}\Lambda'}{nT}\right)}\sqrt{\psi_1\left(\frac{v'v}{nT}\right)}.$$

Following eq. (61),

$$\psi_j\left(\frac{\Lambda\tilde{P}'\tilde{P}\Lambda'}{nT}\right) = O_p\left(\frac{1}{n}\right).$$

By Assumption 4,

$$\psi_1\left(\frac{v'v}{nT}\right) = \psi_1\left(\frac{\dot{\varepsilon}'\dot{\varepsilon}}{nT}\right) + O_p\left(\frac{1}{n^2}\right) + O_p\left(\frac{1}{nT}\right) = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{T}\right).$$

Then for a fixed $\kappa \leq \kappa_{\max}$,

$$\sum\nolimits_{j=1}^{\kappa}\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right) \leq O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{T}\right).$$

Therefore,

$$\hat{V}(\kappa) - \hat{V}(0) = -\sum\nolimits_{j=1}^{\kappa}\psi_j\left(\frac{\hat{u}'\hat{u}}{nT}\right) \leq O_p\left(\frac{1}{\min[n,T]}\right).$$

Q.E.D.□

Part II: Proof of Theorem 1 (ii)

As long as $\gamma_i \neq \gamma$ for any i, the local heterogeneity implies the alternative in CRT (2015b). See the proof of Theorem 3 in CRT (2015b).
Q.E.D.□

Appendix B: Proof of Theorem 2

Let $\dot{X}_i = [\dot{x}_{i1}, \cdots, \dot{x}_{iT}]'$, $X_i^o = [x_{i1}^o, \cdots, x_{iT}^o]'$, $\tilde{F} = \left[\tilde{F}_1, ..., \tilde{F}_T\right]'$, $\tilde{G} = \left[\tilde{G}_1, ..., \tilde{G}_T\right]'$ and $\varepsilon_i = [\varepsilon_{i1}, \cdots, \varepsilon_{iT}]'$. Under the local heterogeneity, we have

$$\sqrt{nT}(\hat{\beta}_{\text{tfe,p}} - \beta) = \left(\frac{1}{nT}\sum\nolimits_{i=1}^{n} \dot{X}_i'\dot{X}_i\right)^{-1} \frac{1}{\sqrt{nT}}\sum\nolimits_{i=1}^{n} \dot{X}_i'\dot{\varepsilon}_i + O_p\left(\sqrt{T/n}\right) + O_p\left(1/\sqrt{n}\right).$$

By definition,

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

and

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

where $\dot{X}_i = \tilde{F}\breve{\Gamma}_i + \tilde{G}\breve{\Psi}_i + \dot{X}_i^o$. Note that

$$\begin{aligned}\Omega_{\text{tfe,p}} &= \mathbb{E}\frac{1}{nT}\sum\nolimits_{i=1}^{n} \dot{X}_i'\varepsilon_i\varepsilon_i'\dot{X}_i \\ &= \mathbb{E}\frac{1}{nT}\sum\nolimits_{i=1}^{n}\left(\breve{\Gamma}_i'\tilde{F}'\varepsilon_i\varepsilon_i'\tilde{F}\breve{\Gamma}_i + \breve{\Psi}_i'\tilde{G}'\varepsilon_i\varepsilon_i'\tilde{G}\breve{\Psi}_i + \dot{X}_i^{o\prime}\varepsilon_i\varepsilon_i'\dot{X}_i^o\right),\end{aligned}$$

and

$$\begin{aligned}Q_{\text{tfe,p}} &= \text{p}\lim_{n,T\to\infty}\frac{1}{nT}\sum\nolimits_{i=1}^{n} \dot{X}_i'\dot{X}_i \\ &= \text{p}\lim_{n,T\to\infty}\frac{1}{nT}\sum\nolimits_{i=1}^{n}\left(\breve{\Gamma}_i'\tilde{F}'\tilde{F}\breve{\Gamma}_i + \breve{\Psi}_i'\tilde{G}'\tilde{G}\breve{\Psi}_i + \dot{X}_i^{o\prime}\dot{X}_i^o\right).\end{aligned}$$

Hence, as $n, T \to \infty$ with $T/n \to 0$

$$V_{\text{cce,p}} - V_{\text{tfe,p}} = Q_{\text{cce,p}}^{-1}\Omega_{\text{cce,p}}Q_{\text{cce,p}}^{-1} - Q_{\text{tfe,p}}^{-1}\Omega_{\text{tfe,p}}Q_{\text{tfe,p}}^{-1}.$$

Assume ε_{it} is i.i.d. over i and t, and let $\mathbb{E}\varepsilon_i\varepsilon_i' = \sigma_\varepsilon^2 I$. Then, it is easy to show that

$$Q_{\text{cce,p}}^{-1}\Omega_{\text{cce,p}}Q_{\text{cce,p}}^{-1} - Q_{\text{tfe,p}}^{-1}\Omega_{\text{tfe,p}}Q_{\text{tfe,p}}^{-1} = \sigma_\varepsilon^2(Q_{\text{cce,p}}^{-1} - Q_{\text{tfe,p}}^{-1}) \geq 0.$$

The equality holds when $\Gamma_i = \Psi_i = 0$.
Q.E.D.□

Appendix C: 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 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

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

Hence

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

Since $V_{\text{cce,p}} - V_{\text{tfe,p}} \geq 0$ in this case, the following inequality holds.

$$V_{\text{BPS,p}} - V_{\text{CRT,p}} \leq 0.$$

Similarly, we can show that

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

and

$$V_{\text{BPS,mg}} - V_{\text{CRT,mg}} \leq 0.$$

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 $n, T \to \infty$, the BPS method suggests,

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

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

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

Similarly,

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

Therefore

$$V_{\text{BPS,p}} - V_{\text{CRT,p}} \leq 0, \;\&\; V_{\text{BPS,mg}} - V_{\text{CRT,mg}} \leq 0.$$

Combining all three cases, we can verify (45).

Q.E.D.$\square$

Appendix D: Asymptotic Comparison between TFE and CCE Estimators When $\beta_i \neq \beta$

We establish the following lemma. We assume the independence of η_i and $x_{it}x_{it}'$, and we derive the limiting distributions of TFE and CCE pooled estimators when $\beta_i \neq \beta$ but $\gamma_i = \gamma$.

Lemma 1: (Asymptotic Distributions of TFE and CCE Pooled Estimators under Heterogeneous Slope Coefficients) *Under Assumption 1-5, 6A and 7, if either $\Gamma_i \neq 0$ or $\Psi_i \neq 0$, but $\gamma_i = \gamma$, as $n, T \to \infty$,*

(i)

$$\sqrt{n}\left(\hat{\beta}_{\text{cce,p}} - \beta\right) \longrightarrow^d \mathcal{N}\left(0, \Omega_{\text{cce,p}}\right),$$

where $\Omega_{\text{cce,p}} = Q_{x^o}^{-1}\Omega_{x^o,\eta}Q_{x^o}^{-1}$, $\Omega_{x^o,\eta}$ and Q_{x^o} are defined in (62).

(ii)

$$\sqrt{n}\left(\hat{\beta}_{\text{tfe,p}} - \beta\right) \to^d \mathcal{N}\left(0, \Omega_{\text{tfe,p}}\right),$$

where $\Omega_{\text{tfe,p}} = Q_{\dot{x}}^{-1}\Omega_{\dot{x}\tilde{x},\eta}Q_{\dot{x}}^{-1}$, $Q_{\dot{x}}$ and $\Omega_{\dot{x}\tilde{x},\eta}$ are defined in (63) and (64).

Proof of Lemma 1 (i) Let $Q_{x^o,i}$ $=\text{plim}_{T\to\infty} T^{-1}X_i^{o\prime}M_P X_i^o$. If $\beta_i = \beta + \eta_i$ and the factor loadings $\gamma_i = \gamma$ for all i, as shown in Proof of Theorem 3 in Pesaran (2006), $\hat{\beta}_{\text{cce,p}}$ can be written as

$$\hat{\beta}_{\text{cce,p}} - \beta = \left(\frac{1}{n}\sum\nolimits_{i=1}^{n} Q_{x^o,i}\right)^{-1}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} Q_{x^o,i}\eta_i\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right).$$

As $n, T \to \infty$,

$$\sqrt{n}\left(\hat{\beta}_{\text{cce,p}} - \beta\right) \longrightarrow^d \mathcal{N}\left(0, \Omega_{\text{cce,p}}\right),$$

$$\Omega_{\text{cce,p}} = Q_{x^o}^{-1}\Omega_{x^o,\eta}Q_{x^o}^{-1}, \tag{62}$$

where $\Omega_{x^o,\eta} = n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(Q_{x^o,i}\Omega_\eta Q_{x^o,i}\right)$, $\Omega_\eta = \mathbb{E}\left(\eta_i\eta_i'\right)$, and Q_{x^o} $=\text{plim}_{n,T\to\infty}\left(n^{-1}\sum_{i=1}^{n} Q_{x^o,i}\right)$.

Proof of Lemma 1 (ii) If $\gamma_i = \gamma$, we rewrite the panel regression as

$$y_{it} = a_i + \beta' x_{it} + \eta_i' x_{it} + \gamma F_t + \varepsilon_{it}.$$

After the within transformation,

$$\dot{y}_{it} = \beta'\dot{x}_{it} + \eta_i'\tilde{x}_{it} - \left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\eta_i'\tilde{x}_{it}\right) + \dot{\varepsilon}_{it},$$

with $\tilde{x}_{it} = x_{it} - T^{-1}\sum_{t=1}^{T} x_{it}$. The TFE pooled estimator is given by

$$\hat{\beta}_{\text{tfe,p}} - \beta = \left(\frac{1}{nT}\sum\nolimits_{i=1}^{n}\sum\nolimits_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1} \times$$
$$\left\{\frac{1}{nT}\sum\nolimits_{i=1}^{n}\sum\nolimits_{t=1}^{T}\dot{x}_{it}\left[\tilde{x}_{it}'\eta_i - \left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\tilde{x}_{it}'\eta_i\right)\right]\right\} + O_p\left(\frac{1}{\sqrt{nT}}\right)$$

By Assumption 7 and the WLLN, as $n, T \to \infty$,

$$\text{plim}_{n,T\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{x}_{it}\dot{x}_{it}' = Q_{\dot{x}}, \tag{63}$$

where $Q_{\dot{x}}$ is a $k \times k$ positive definite matrix. By the independence of $x_{it}x_{it}'$ and η_i for all i and j, we have

$$\mathbb{E}\left\{\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{x}_{it} \left(\tilde{x}_{it}'\eta_i - \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{x}_{it}'\eta_i\right)\right)\right\} = 0.$$

Note that $n^{-1}\sum_{i=1}^{n} \tilde{x}_{it}'\eta_i = O_p\left(n^{-1/2}\right)$. The variance is thus given by

$$\begin{aligned} & \text{Var}\left\{\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{x}_{it} \left[\tilde{x}_{it}'\eta_i - \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{x}_{it}'\eta_i\right)\right]\right\} \\ = \;& n^{-2}\sum_{i=1}^{n} \mathbb{E}\left(Q_{\dot{x}\tilde{x},i}\Omega_\eta Q_{\dot{x}\tilde{x},i}\right) + O_p\left(n^{-2}\right), \end{aligned}$$

where $\Omega_\eta = \mathbb{E}\left(\eta_i\eta_i'\right) \geq 0$, and $Q_{\dot{x}\tilde{x},i} = \text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \dot{x}_{it}\tilde{x}_{it}'$. By CLT, as $n, T \to \infty$,

$$\sqrt{n}\left(\hat{\beta}_{\text{tfe,p}} - \beta\right) \to^d \mathcal{N}\left(0, \Omega_{\text{tfe,p}}\right), \tag{64}$$

with $\Omega_{\text{tfe,p}} = Q_{\dot{x}}^{-1}\Omega_{\dot{x}\tilde{x},\eta}Q_{\dot{x}}^{-1}$, and $\Omega_{\dot{x}\tilde{x},\eta} = n^{-1}\sum_{i=1}^{n} \mathbb{E}\left(Q_{\dot{x}\tilde{x},i}\Omega_\eta Q_{\dot{x}\tilde{x},i}\right)$. Q.E.D.□

Asymptotic Variance Comparison We can further decompose $\Omega_{\text{cce,p}}$ and $\Omega_{\text{tfe,p}}$ as follows:

$$\Omega_{\text{cce,p}} = \Omega_\eta + Q_{x^o}^{-1}\left\{n^{-1}\sum_{i=1}^{n} \mathbb{E}\left[\left(Q_{x^o,i} - Q_{x^o}\right)\Omega_\eta\left(Q_{x^o,i} - Q_{x^o}\right)\right]\right\}Q_{x^o}^{-1},$$

$$\begin{aligned} \Omega_{\text{tfe,p}} = \;& \Omega_\eta + Q_{\dot{x}}^{-1}\left\{n^{-1}\sum_{i=1}^{n} \mathbb{E}\left[\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\Omega_\eta\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\right]\right\}Q_{\dot{x}}^{-1} \\ & + Q_{\dot{x}}^{-1}\left[n^{-1}\sum_{i=1}^{n} \mathbb{E}\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\right]\Omega_\eta + \Omega_\eta\left[n^{-1}\sum_{i=1}^{n} \mathbb{E}\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\right]Q_{\dot{x}}^{-1}. \end{aligned}$$

Suppose that $Q_{x^o,i} = Q_{x^o}$ for all i, it is easy to show that as $n, T \to \infty$, $\Omega_{\text{cce,p}} \to \Omega_\eta$. Next, observe this.

$$\begin{aligned} & \Omega_{\text{tfe,p}} - \Omega_{\text{cce,p}} \\ = \;& Q_{\dot{x}}^{-1}\left\{n^{-1}\sum_{i=1}^{n} \mathbb{E}\left[\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\Omega_\eta\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\right]\right\}Q_{\dot{x}}^{-1} \\ & + Q_{\dot{x}}^{-1}\left[n^{-1}\sum_{i=1}^{n} \mathbb{E}\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\right]\Omega_\eta + \Omega_\eta\left[n^{-1}\sum_{i=1}^{n} \mathbb{E}\left(Q_{\dot{x}\tilde{x},i} - Q_{\dot{x}}\right)\right]Q_{\dot{x}}^{-1} \\ = \;& A + B + C. \end{aligned}$$

Recall that $\breve{\Upsilon}_i = \Upsilon_i - n^{-1}\sum_{i=1}^{n} \Upsilon_i$ and $\Upsilon_i' = \left[\Gamma_i', \Psi_i'\right]$. By WLLN,

$$Q_{\dot{x}\tilde{x},i} = \text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \left(\breve{\Upsilon}_i'\tilde{P}_t + \dot{x}_{it}^o\right)\left(\Upsilon_i'\tilde{P}_t + \tilde{x}_{it}^o\right)' = \breve{\Upsilon}_i'\Sigma_{\tilde{P}}\Upsilon_i + Q_{\dot{x}^o\tilde{x}^o,i},$$

$$Q_{\dot{x}} = \text{plim}_{n,T\to\infty}\left(nT\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} \left(\breve{\Upsilon}_i'\tilde{P}_t + \dot{x}_{it}^o\right)\left(\breve{\Upsilon}_i'\tilde{P}_t + \dot{x}_{it}^o\right)' = \mathbb{E}\left(\breve{\Upsilon}_i'\Sigma_{\tilde{P}}\breve{\Upsilon}_i\right) + Q_{\dot{x}^o},$$

with $Q_{\dot{x}^o\tilde{x}^o,i}$ $=\text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}\dot{x}_{it}^{o}\tilde{x}_{it}^{o\prime}$, $Q_{\dot{x}^o}$ $=\text{plim}_{n,T\to\infty}(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{x}_{it}^{o}\dot{x}_{it}^{o\prime}$, and $\Sigma_{\tilde{P}}$ $=$ $\mathbb{E}\left(\tilde{P}_t\tilde{P}_t'\right)$ being a $(r+m)\times(r+m)$ positive definite matrix. Since we assume that $Q_{\dot{x}^o,i}=Q_{\dot{x}^o}$, then

$$
\begin{aligned}
Q_{\dot{x}^o\tilde{x}^o,i}-Q_{\dot{x}^o} &= Q_{\dot{x}^o\tilde{x}^o,i}-Q_{\dot{x}^o,i}+\left(Q_{\dot{x}^o,i}-Q_{\dot{x}^o}\right)\\
&= \text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}\dot{x}_{it}^{o}\left(n^{-1}\sum_{j=1}^{n}\tilde{x}_{jt}^{o\prime}\right)=O\left(n^{-1}\right),
\end{aligned}
$$

$$
Q_{\dot{x}\tilde{x},i}-Q_{\dot{x}}=\breve{\Upsilon}_i'\Sigma_{\tilde{P}}\Upsilon_i-\mathbb{E}\left(\breve{\Upsilon}_i'\Sigma_{\tilde{P}}\breve{\Upsilon}_i\right)+O\left(n^{-1}\right)\neq 0,\text{ if }\breve{\Upsilon}_i\neq 0\text{ for some }i.
$$

As $n,T\to\infty$,

$$
A=Q_{\dot{x}}^{-1}\left\{n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\left(Q_{\dot{x}\tilde{x},i}-Q_{\dot{x}}\right)\Omega_\eta\left(Q_{\dot{x}\tilde{x},i}-Q_{\dot{x}}\right)\right]\right\}Q_{\dot{x}}^{-1}\geq 0.
$$

Moreover,

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(Q_{\dot{x}\tilde{x},i}-Q_{\dot{x}}\right) &= n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\breve{\Upsilon}_i'\Sigma_{\tilde{P}}\Upsilon_i-\mathbb{E}\left(\breve{\Upsilon}_i'\Sigma_{\tilde{P}}\breve{\Upsilon}_i\right)\right]+O\left(\frac{1}{n}\right)\\
&= n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\breve{\Upsilon}_i'\Sigma_{\tilde{P}}\left(n^{-1}\sum_{j=1}^{n}\Upsilon_j\right)\right]+O\left(\frac{1}{n}\right)\\
&= n^{-2}\sum_{i=1}^{n}\mathbb{E}\left(\breve{\Upsilon}_i'\Sigma_{\tilde{P}}\Upsilon_i\right)+O\left(\frac{1}{n}\right)=O\left(\frac{1}{n}\right),
\end{aligned}
$$

which implies that as $n,T\to\infty$,

$$
\begin{aligned}
B &= \Omega_\eta\left[n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(Q_{\dot{x}\tilde{x},i}-Q_{\dot{x}}\right)\right]Q_{\dot{x}}^{-1}\to 0,\\
C &= Q_{\dot{x}}^{-1}\left[n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(Q_{\dot{x}\tilde{x},i}-Q_{\dot{x}}\right)\right]\Omega_\eta\to 0.
\end{aligned}
$$

Therefore, as $n,T\to\infty$,

$$
\Omega_{\text{tfe,p}}-\Omega_{\text{cce,p}}=A+B+C\to A\geq 0.
$$

If $Q_{x^o,i}\neq Q_{x^o}$ for some i, then it is not straightforward to compare two variances mathematically. We investigate this issue by means of Monte Carlo simulations.

Appendix E: Proof of Remark 1

In this part, we establish the consistency of the BPS method in dynamic panel data models with weakly exogeneous regressors. Consider the following DGP:

$$
y_{it}=a_i+\beta x_{it-1}+\lambda_i F_t+u_{it},
$$

where we assume

$$
\begin{aligned}
x_{it} &= \lambda_{x,i}F_t+\phi_i G_t+x_{it}^{o},\\
x_{it}^{o} &= c_i+\rho x_{it-1}^{o}+\varepsilon_{it},
\end{aligned}
$$

and

$$u_{it} = \delta \varepsilon_{it} + \epsilon_{it},$$

where $\varepsilon_{it} \sim iid\left(0, \sigma_{\varepsilon}^2\right)$, $\epsilon_{it} \sim iid\left(0, \sigma_{\epsilon}^2\right)$ and $\mathbb{E}\left(\varepsilon_{it}\epsilon_{js}\right) = 0$ for all i, j, t, and s. Note that we can consider a more complicated data generating process, but the main result does not change. Furthermore, the current setting has been popularly used in this literature. Let $\hat{\beta}_{\text{fe}}$ be the one way fixed effects or WG estimator. We consider the following cases:

For the first four cases ($\beta_i = \beta$), we have

$$\hat{\beta}_{\text{fe}} - \beta = \delta \frac{\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it-1}\dot{\varepsilon}_{it}}{\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it-1}^2} + \frac{\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it-1}\dot{\epsilon}_{it}}{\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it-1}^2}.$$

Case 1: When $\lambda_i = \lambda$, $\lambda_{x,i} = \lambda_x$, and $\phi_i = \phi$ for all i

$$\text{plim}_{n\to\infty}\left(\hat{\beta}_{\text{fe}} - \beta\right) = -\delta\frac{1+\rho}{T} + O_p\left(n^{-1/2}T^{-1/2}\right).$$

Then we have

$$\hat{u}_{it} = \dot{u}_{it} + \left(\beta - \hat{\beta}_{\text{fe}}\right)\dot{x}_{it-1}^o = \dot{u}_{it} + O_p\left(T^{-1}\right) + O_p\left((nT)^{-1/2}\right).$$

Let $\hat{u} = \left(\hat{u}_1, \cdots, \hat{u}_T\right)'$, and $\hat{u}_t = \left(\hat{u}_{1t}, \cdots, \hat{u}_{nt}\right)'$. As in the Proof of Theorem 1, we can show that

$$\hat{V}(\kappa) - \hat{V}(0) = -\frac{1}{nT}\sum_{j=1}^{\kappa} \psi_j\left(\hat{u}'\hat{u}\right) \leq O_p\left(\frac{1}{\min[n,T]}\right).$$

Case 2: When $\lambda_i = \lambda$ for all i, but $\lambda_{x,i} \neq \lambda_x$ and $\phi_i \neq \phi$ for some i

Note that

$$\frac{\text{p}\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it-1}\dot{\varepsilon}_{it}}{\text{p}\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it-1}^2} \neq -\left(\frac{1+\rho}{T}\right),$$

since

$$\text{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \dot{x}_{it-1}^2 = \sigma_{\lambda_x}^2\sigma_F^2 + \sigma_{\phi}^2\sigma_G^2 + \sigma_{x^o}^2 \neq \sigma_{x^o}^2,$$

where

$$\begin{aligned}
\sigma_{\lambda_x}^2\sigma_F^2 &= \text{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \lambda_{x,i}^2 F_{t-1}^2, \\
\sigma_{\phi}^2\sigma_G^2 &= \text{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \phi_i^2 G_{t-1}^2,
\end{aligned}$$

and

$$\sigma_{x^o}^2 = \text{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \left(x_{it-1}^o\right)^2.$$

Then we have

$$\hat{u}_{it} = \dot{u}_{it} + \left(\beta - \hat{\beta}_{\text{fe}}\right) \dot{x}_{it-1} = \dot{u}_{it} + O_p\left(T^{-1}\right) + O_p(n^{-1/2}T^{-1/2}).$$

Hence, it is straightforward to show

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

Case 3: When $\lambda_i \neq \lambda$, $\lambda_{x,i} = \lambda_x$, and $\phi_i = \phi$ for all i

In this case, we have

$$\hat{\beta}_{\text{fe}} - \beta = -\delta\frac{1+\rho}{T} + O_p\left(n^{-1/2}T^{-1/2}\right),$$

then we have

$$\hat{u}_{it} = \dot{u}_{it} + \left(\lambda_i - n^{-1}\sum\nolimits_{i=1}^{n} \lambda_i\right) \tilde{F}_t + \left(\beta - \hat{\beta}_{\text{fe}}\right) \dot{x}^{o}_{it-1}.$$

Hence the estimated number of common factor becomes non-zero.

Case 4: When $\lambda_i \neq \lambda$, $\lambda_{x,i} \neq \lambda_x$, and $\phi_i \neq \phi$ for some i

In this case, we have

$$\hat{u}_{it} = \dot{u}_{it} + \left(\lambda_i - n^{-1}\sum\nolimits_{i=1}^{n} \lambda_i\right) \tilde{F}_t + \left(\beta - \hat{\beta}_{\text{fe}}\right) \dot{x}_{it-1}.$$

Hence the estimated number of common factor must be greater than zero.

Case 5: When β_i is heterogeneous ($\beta_i = \beta + \eta_i$)

Note that

$$\begin{aligned} y_{it} &= a_i + \beta x_{it-1} + \lambda_i F_t + \left(\beta_i - \beta\right) x_{it-1} + u_{it} \\ &= a_i + \beta x_{it-1} + \lambda_i F_t + \eta_i x_{it-1} + u_{it}, \end{aligned}$$

$$\hat{u}_{it} = \dot{u}_{it} + \breve{\lambda}_i \tilde{F}_t + \eta_i \tilde{x}_{it-1} - \left(n^{-1}\sum\nolimits_{i=1}^{n} \eta_i \tilde{x}_{it-1}\right) + \left(\beta - \hat{\beta}_{\text{fe}}\right) \dot{x}_{it-1}.$$

Hence, the estimated number of common factor is always greater than zero.

When ρ is heterogeneous, an imposing homogeneous restriction on ρ induces a factor structure in $\hat{u}_{it}$. In this case, the estimated number of common factor becomes non-zero even when $\lambda_i = \lambda$, $\lambda_{x,i} = \lambda_x$, and $\phi_i = \phi$ for all i. Table E-1 shows the results of an asymtotic factor number in various cases.

Table E-1: Summary of The Results with Various Cases

			Aymptotic Factor Number
$\beta_i = \beta$	$\lambda_i = \lambda$	$\lambda_{x,i} = \lambda_x$, $\phi_i = \phi$	$\lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 1$
	$\lambda_i = \lambda$	$\lambda_{x,i} \neq \lambda_x$ or $\phi_i \neq \phi$	$\lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 1$
	$\lambda_i \neq \lambda$	$\lambda_{x,i} = \lambda_x$, $\phi_i = \phi$	$\lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 0$
	$\lambda_i \neq \lambda$	$\lambda_{x,i} \neq \lambda_x$ or $\phi_i \neq \phi$	$\lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 0$
$\beta_i \neq \beta$	No restriction		$\lim_{n,T\to\infty} \Pr[\hat{\#}(\hat{u}_{it}) = 0] = 0$

Appendix F: Unbalanced Panel

Here we only extend the application of the BPS method to unbalanced panels. Let n_t be the number of cross-sectional units observed in period t. For $i = 1, \cdots, n_t$ and $t = 1, \cdots, T$, we can modify the proposed two-step procedure as follows:

Step 1: Hansen (2020, p. 620) discusses how to estimate $\hat{\beta}_{\text{tfe}}$ for the unbalanced panel data. Here we briefly describe his procedure. Let τ_t be a set of T dummy variables where the t-th element of τ_t is equal to one, otherwise zero. Instead of (18), run the following one-way fixed effects regression.

$$\tilde{y}_{it} = \beta' \tilde{x}_{it} + \tilde{\tau}_t' F^o + u_{it}, \tag{65}$$

where $\tilde{y}_{it} = y_{it} - T^{-1} \sum_{t=1}^T y_{it}$, $\tilde{x}_{it} = x_{it} - T^{-1} \sum_{t=1}^T x_{it}$, and $\tilde{\tau}_t = \tau_t - T^{-1} \sum_{t=1}^T \tau_t$. This produces estimates of the slope coefficients $\hat{\beta}_{\text{tfe}}$ and the time effects $\hat{F}^o$. Next, get the residuals, $\hat{u}_{it} = \tilde{y}_{it} - \hat{\beta}_{\text{tfe}}' \tilde{x}_{it} - \tilde{\tau}_t' \hat{F}^o$.

Step 2 This algorithm is introduced in Appendix B in Bai (2009) and can be implemented by `regife` in Stata, which computes both the factor and the factor loadings $\hat{\gamma}_i^*$ and $\hat{F}_t^*$. Obtain a new balanced panel data matrix[1] with

$$\hat{u}_{it}^* = \begin{cases} \hat{\gamma}_i^{*\prime} \hat{F}_t^* & \text{if } \hat{u}_{it} \text{ is missing,} \\ \hat{u}_{it} & \text{o.w.} \end{cases}$$

[1]See Appendix A in Stock and Watson (1998) for more discussions about different imputing methods of dealing with specific data irregularities.

Use BN's IC_2 criterion to estimate the number of common factors with $\hat{u}_{it}^*$.

If $\#\left(\hat{u}_{it}^*\right) = 0$, then the regression in (65) should be run. Otherwise, the factor-augmented regressions should be considered. Pesaran (2015, p. 793) provides detailed procedures on how to deal with unbalanced panel data. Also, note that the CCE estimation in unbalanced panels can be implemented using the Stata package `xtdcce2`.

Table 3: Finite sample performances of pre-testing procedures under homogeneous factor loadings and slope coefficients

n	T	Frequencies*		Pooled Case Variance Comparison			MG Case Variance Comparison		
		$\mathrm{IC}_2 > 0$	$\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	$V_{\mathrm{BPS,p}}$	$\frac{V_{\mathrm{CRT,p}}}{V_{\mathrm{BPS,p}}}$	$\frac{V_{\mathrm{CCE,p}}}{V_{\mathrm{BPS,p}}}$	$V_{\mathrm{BPS,mg}}$	$\frac{V_{\mathrm{CRT,mg}}}{V_{\mathrm{BPS,mg}}}$	$\frac{V_{\mathrm{CCE,mg}}}{V_{\mathrm{BPS,mg}}}$
25	25	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 4: Finite sample performances of pre-testing procedures
under heterogeneous factor loadings but homogeneous slope coefficients

		Frequencies		Pooled Case Variance Comparison			MG Case 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	0.990	1.000	1.006	0.990	0.990	1.150	0.989	0.989
25	50	1.000	1.000	0.475	1.000	1.000	0.489	1.000	1.000
25	100	1.000	1.000	0.262	1.000	1.000	0.253	1.000	1.000
25	200	1.000	1.000	0.155	1.000	1.000	0.140	1.000	1.000
50	25	1.000	1.000	0.474	1.000	1.000	0.555	1.000	1.000
50	50	1.000	1.000	0.271	1.000	1.000	0.231	1.000	1.000
50	100	1.000	1.000	0.111	1.000	1.000	0.112	1.000	1.000
50	200	1.000	1.000	0.059	1.000	1.000	0.058	1.000	1.000
100	25	1.000	1.000	0.238	1.000	1.000	0.277	1.000	1.000
100	50	1.000	1.000	0.106	1.000	1.000	0.114	1.000	1.000
100	100	1.000	1.000	0.053	1.000	1.000	0.054	1.000	1.000
100	200	1.000	1.000	0.026	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 5: Finite sample performances of pre-testing procedures
under local heterogeneous factor loadings and homogeneous slope coefficients

n	T	Frequencies		Pooled Case Variance Comparison			MG Case Variance Comparison		
		$\text{IC}_2 > 0$	$\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	$V_{\text{BPS,p}}$	$\frac{V_{\text{CRT,p}}}{V_{\text{BPS,p}}}$	$\frac{V_{\text{CCE,p}}}{V_{\text{BPS,p}}}$	$V_{\text{BPS,mg}}$	$\frac{V_{\text{CRT,mg}}}{V_{\text{BPS,mg}}}$	$\frac{V_{\text{CCE,mg}}}{V_{\text{BPS,mg}}}$
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 6: Finite sample performances of pre-testing procedures under homogeneous factor loadings but heterogeneous 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	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 7: Finite sample performances of pre-testing procedures under heterogeneous factor loadings and slope coefficients

n	T	Frequencies $\mathrm{IC}_2 > 0$	Frequencies $\mathcal{S}_{\gamma,nT} > c_{\alpha,n}$	Pooled Case Variance Comparison $V_{\mathrm{BPS,p}}$	$\frac{V_{\mathrm{CRT,p}}}{V_{\mathrm{BPS,p}}}$	$\frac{V_{\mathrm{CCE,p}}}{V_{\mathrm{BPS,p}}}$	MG Case Variance Comparison $V_{\mathrm{BPS,mg}}$	$\frac{V_{\mathrm{CRT,mg}}}{V_{\mathrm{BPS,mg}}}$	$\frac{V_{\mathrm{CCE,mg}}}{V_{\mathrm{BPS,mg}}}$
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