# Factor-Augmented Nonstationary Panels with Multiple Structural Changes<sup>\*</sup>

Badi H. Baltagi<sup>†</sup> Qu Feng <sup>‡</sup> Wei Wang<sup>§</sup>

15 September 2023

#### Abstract

Nonstationary panels have been widely used in empirical studies in economics, especially in macroeconomics and finance. This paper considers multiple structural changes in nonstationary heterogeneous panels with common factors. Kapetanios, Pesaran, Yamagata (2011) showed that unobserved nonstationary factors can be proxied by cross-sectional averages of observable data. This means that unobserved error factors can be treated as additional regressors, and different break points in slopes and error factor loadings can be considered as multiple breaks in linear regression models with panel data. Therefore, we generalize the least squares approach by Bai and Perron (1998) to nonstationary panels and show that the break points in both slopes and error factor loadings can be consistently estimated for two important cases involving i) nonstationary factors and ii) nonstationary regressors considered by Phillips and Moon (1999). Monte Carlo simulations are conducted to study the performance of the main results in finite samples.

**Keywords:** Nonstationary Panels, Multiple Structural Changes, Heterogeneity, Common Factors, Common Correlated Effects.

JEL Classification: C13, C23, C33, C38

<sup>\*</sup>An earlier version of this paper was presented at the International Association of Applied Econometrics (IAAE 2023) Annual Conference in Oslo, The International Panel Data Conference (IPDC 2023) in Amsterdam, the Asian Meeing of the Econometric Society (2022) in Tokyo; Asian Meeing of the Econometric Society (2022) in China; SETA (2022) in Seoul. Constructive comments from the audiences at these conferences are highly acknowledged.

<sup>&</sup>lt;sup>†</sup>Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020, USA, Email Address: bbaltagi@syr.edu.

<sup>&</sup>lt;sup>‡</sup>Economics Division, School of Social Sciences, Nanyang Technological University. SHHK-04-84, 48 Nanyang Avenue, Singapore 639818. Email Address: qfeng@ntu.edu.sg.

<sup>&</sup>lt;sup>§</sup>Zhongtai Securities Institute for Financial Studies, Shandong University. 27 Shanda South Road, Jinan, China, 250100. Email Address: wangwei\_0115@outlook.com.

### 1 Introduction

Nonstationary panel data regression models allowing for cross-sectional dependence using a factor structure in the errors continue to be the focus of a lot of theoretical as well as empirical studies in econometrics. See Hsiao (2018) who provides a very detailed and insightful review of some of the main modeling and estimation approaches in the factoraugmented panel data literature. Also, Feng and Kao (2020) for a textbook treatment of this subject focusing on three main approaches for factor-augmented panel data models. These include Pesaran's (2006) common correlated effects (CCE) approach, Bai's (2009) iterated principal components (IPC) approach, and the likelihood approaches proposed by Bai and Li (2014) and advocated by Hsiao (2018). More recently, the transformed approach developed by Hsiao, Shi, Zhou (2021) shows very good properties in dealing with error factors in panel data models.

This paper contributes to the nonstationary panels with common factors literature. It is motivated by Bai and Kao (2006) who consider a panel cointegration model with stationary factors, which are allowed to be correlated with the regressors.  $\sqrt{nT}$ -consistent fully modified (2sFM) estimators of the slope parameters are derived. In a panel cointegration model with nonstationary factors  $y_{it} = x'_{it}\beta + \gamma'_i f_t + \varepsilon_{it}$  considered by Bai, Kao and Ng (2009),  $f_t$  are treated as parameters, and  $y_{it}$  cointegrates with  $x_{it}$  and  $f_t$  with coefficients  $(1, -\beta', \gamma'_i)$ . The IPC approach is applied to deal with unobserved factors, as in Bai (2009), and  $\sqrt{nT}$ -consistent continuously updated bias-corrected (CupBC) and continuously updated fully modified (CupFM) estimators of the slope parameters  $\beta$  are proposed. Recently, Huang, Jin, and Su (2020) and Huang, Jin, Phillips, Su (2021) introduce heterogeneity modelled as a latent group structure in the slope parameters in a panel cointegration model with nonstationary factors, thus adding two features of heterogeneity and cross-section dependence to the nonstationary panel literature. A penalized principal component estimation, which is an iterative procedure between penalized regression and principal component analysis (PCA), is proposed to consistently estimate group membership and the slope parameters. Different from the homogeneous panel literature considered above, Kapetanios, Pesaran, and Yamagata (2011, KPY hereafter) estimate a model of *heterogeneous* panels with nonstationary factors. They find that the CCE approach proposed by Pesaran (2006) is still valid for I(1) factors. In addition, Holly,

Pesaran and Yamagata (2010) apply these methods to examine empirical features of the US housing markets.<sup>1</sup>

Following Huang et al. (2021) and Dong et al. (2021), this paper adds heterogeneity to the literature by considering multiple structural changes in nonstationary panels with common factors. Specifically, we consider multiple breaks in slopes and error factor loadings in heterogeneous panels with nonstationary regressors and factors. As such, this paper enriches the literature of nonstationary panels by accommodating two additional empirical features of multiple structural changes and cross-sectional dependence. As in Pesaran (2006), Kapetanios, Pesaran, Yamagata (KPY), unobserved nonstationary factors can be proxied by cross-sectional averages of observable data. Thus, unobserved error factors can be treated as additional regressors, and different break points in slopes and error factor loadings can be considered as multiple breaks in linear regression models with panel data. Therefore, we generalize the least squares approach by Bai and Perron (1998) to nonstationary panels and show that the break points in both slopes and error factor loadings can be consistently estimated. In addition, different from KPY, we also consider the case of nonstationary regressors after the CCE transformation. This model can be considered as an extension of Phillips and Moon (1999, Section 5) to the case of allowing for an error factor structure and multiple breaks in slopes. Similarly, a T-consistent estimator of the heterogeneous slope parameters is obtained.

There have been important work on estimating and testing for multiple structural changes in the time series literature and here we briefly review some of the classic papers. Simultaneous estimation using least squares include Bai and Perron (1998) and Mohitosh and Perron (2008) and the sequential approach examined by Bai (1997) and Pang et al. (2021), to mention a few. Bai and Perron (2003) provide a dynamic programming algorithm to reduce the complexity of computation. Likelihood approaches are used by Bai (2000) in vector autoregressive models (VAR) and by Qu and Perron (2007) in multivariate regression models. Maheu and Song (2018) use a Bayesian approach to estimate multiple structural breaks in VAR and other multivariate models. Regarding testing for multiple structural breaks, in addition to Bai and Perron (1998) and Qu and Perron (2007), Wald-type tests are considered by Kejriwal and Perron (2008), and a nonparametric maximum

<sup>&</sup>lt;sup>1</sup>Dong, Gao and Peng (2021) propose a general model of nonstationary panels by considering varyingcoefficient slopes and factor loadings.

likelihood approach is proposed by Zou et al. (2014). More recently, Oka and Perron (2018) and Bergamelli et al. (2019) propose a multiple hypothesis testing approach in cointegrating regressions.

Estimation of structural breaks in panels has attracted a lot of attention since the important paper by Bai (2010). Kim (2011) estimate a common deterministic trend break for large panels with nonstationary or stationary error. Baltagi, Feng and Kao (2016, 2019, BFK hereafter) extend Pesaran's (2006) heterogenous panels to the cases of common breaks in slopes with exogenous and endogenous regressors. Baltagi, Kao and Wang (2015) consider interactive fixed effects in errors of heterogeneous panels, instead of commonly correlated factors. Baltagi, Kao and Liu (2017) consider the estimation of break point in simple nonstationary panels. These models mainly focus on the case of a single common break. Li, Qian and Su (2017) propose adaptive group fused LASSO (AGFL) in panels with multiple breaks in slopes, with and without interactive effects, respectively. Lumsdaine, Okui and Wang (2023) consider the estimation of panel group structure models with structural breaks. Kaddoura and Westerlund (2023) consider the estimation of panel data models with multiple structural breaks when time dimension is fixed.

Recently, Karavias, Narayan and Westerlund (2023) consider a single break in stationary homogeneous panels with interactive effects, and Ditzen, Karavias and Westerlund (2023) extend the analysis to the case of multiple breaks. Unlike these two papers, we focus on nonstationary heterogeneous panels and nonstationary factors with multiple breaks. In addition, multiple breaks in factor loadings are also considered in our paper. Thus, our model can be applied to empirical research using aggregate level data over a long period, e.g., climate change analysis.

Since we are studying breaks in error factor loadings, this paper is also related to the literature on structural instability in factor models considered by Stock and Watson (2009), and extensively studied by Breitung and Eickmeier (2011), Chen, Dolado and Gonzalo (2014), Yamamoto and Tanaka (2015), and Cheng, Liao and Schorfheide (2016). Recent advancements in this direction also include Baltagi, Kao and Wang (2017), Bai, Han and Shi (2020), and Duan, Bai and Han (2023). In addition, Baltagi, Kao and Wang (2021) and Ma and Tu (2023) and allow for multiple breaks in the loading.

The paper is organized as follows. Section 2 introduces the model of nonstationary

panels with common factors and multiple structural changes in slopes and error factor loadings. Section 3 presents the main ideas for estimation. Asymptotic properties of the estimators are derived in Section 4. In Section 5, we consider the special case of nonstationary regressors after the CCE transformation. Monte Carlo simulations are conducted in Section 6. Section 7 provides concluding remarks. The mathematical proofs are relegated to the Appendix.

**Notation:** For any matrix or vector A, the Frobenius norm of A is defined as  $||A|| = \sqrt{tr(AA')}$ .  $(N,T) \to \infty$  denotes N and T tend to infinity simultaneously.  $[\cdot]$  is the greatest integer function. Stochastic processes such as Brownian motion W(r) on [0,1] are written as W, integrals such as  $\int_c^d W(r) dr$  as  $\int_c^d W$  and stochastic integrals  $\int_c^d W(r) dW(r)$  as  $\int_c^d W dW$ .  $B_{\omega}$  denotes the Brownian motion with covariance matrix  $\Sigma_{\omega}$ . " $\Rightarrow$ " denotes weak convergence.

### 2 Model

By extending Pesaran's (2006) influential framework to the nonstationary case, KPY consider the following heterogeneous panel regression with nonstationary factors:

$$y_{it} = x'_{it}\beta_i + \gamma'_i f_t + \varepsilon_{it}, \ i = 1, ..., N; \ t = 1, ..., T,$$
(1)

where  $x_{it}$  is a  $p \times 1$  vector of explanatory variables with heterogeneous slopes  $\beta_i$ ,  $\varepsilon_{it}$  is the idiosyncratic error, independent of  $x_{it}$ , and  $\gamma_i$  is the corresponding loading vector.<sup>2</sup> The  $q \times 1$  vector of unobserved factors  $f_t$  follow I(1) processes,

$$f_t = f_{t-1} + \varphi_t, \tag{2}$$

 $\varphi_t$  is the idiosyncratic error.  $x_{it}$  follow an I(1) processes under the Assumption of commonly correlated effects,

$$x_{it} = \Gamma'_i f_t + v_{it},\tag{3}$$

where  $\Gamma_i$  is an  $q \times p$  factor loading matrix.  $v_{it}$  is a  $p \times 1$  vector of disturbances. Thus,  $y_{it}$  is also nonstationary. KPY show that the CCE approach is robust to nonstationary

<sup>&</sup>lt;sup>2</sup>The fixed effects model can be considered as a special case when the first component of  $x_{it}$  is 1 and the other components of the slope parameters  $\beta_i$  are homogeneous. We examine the performance of the break estimators in a fixed effects model in the Monte Carlo experiments.

factors.  $v_{it}$  is assumed to be I(0) as in KPY, in what we call Case 1 in this and the next section. Case 2 assumes  $v_{it}$  to be I(1) and this is studied in Section 5.

This paper considers multiple structural breaks in slopes  $\beta_i$  and error factor loadings  $\gamma_i$  in KPY's model (1) above:

$$y_{it} = x'_{it}\beta_i(\mathcal{K}_0) + \gamma'_i f_t + \varepsilon_{it}, i = 1, ..., N; t = 1, ..., T.$$
(4)

Common breaks in slopes  $\beta_i(\mathcal{K}_0)$  could arise due to technological progress or major policy shifts in a long time horizon. Assume there are  $m_0$  breaks in the slope parameters.<sup>3</sup> As in Bai and Perron (1998),  $\mathcal{K}_0$  denotes an  $m_0$ -partition  $(K_{0,1}, ..., K_{0,m_0})$ , and the value of the slopes  $\beta_i(\mathcal{K}_0)$  vary across  $m_0 + 1$  different regimes, i.e.,

$$\beta_i(\mathcal{K}_0) = \begin{cases} \beta_{i1}, & t = 1, \dots, K_{0,1}, \\ \vdots \\ \beta_{i,m_0+1}, & t = K_{0,m_0} + 1, \dots, T. \end{cases}$$

This model generalizes the analysis of stationary panels with a single break in slopes by BFK (2016, 2019) to nonstationary panels with multiple structural breaks. Thus, additional technical challenges are involved in the derivations of asymptotic properties of estimators with nonstationary data in the case of multiple breaks.

Similarly, factor loadings  $\gamma_i$  could also suffer from structural changes often seen in the macroeconomic literature (Stock and Watson, 2009). Assume there are  $m_1$  breaks in the error factor loadings with an  $m_1$ -partition  $\mathcal{K}_1 = (K_{1,1}, ..., K_{1,m_1})$ ,

$$\gamma_i(\mathcal{K}_1) = \begin{cases} \gamma_{i1}, & t = 1, \dots, K_{1,1}, \\ \vdots \\ \gamma_{i,m_1+1}, & t = K_{1,m_1} + 1, \dots, T \end{cases}$$

The model becomes

$$y_{it} = x'_{it}\beta_i(\mathcal{K}_0) + \gamma_i(\mathcal{K}_1)'f_t + \varepsilon_{it}, \ i = 1, ..., N; \ t = 1, ..., T.$$
(5)

In addition, the nonstationary  $f_t$  and  $x_{it}$  follow processes (2) and (3). We suppress the superscript 0 in the true values of  $\mathcal{K}_0$  and  $\mathcal{K}_1$  for now. Breaks  $\mathcal{K}_1$  in error factor loadings are allowed to have overlaps with breaks  $\mathcal{K}_0$  in the slopes. Different from breaks  $\mathcal{K}_0$  in

<sup>&</sup>lt;sup>3</sup>To accommodate the case of partial structural changes in the slopes considered by Bai and Perron (1998),  $w'_{it}\alpha_i$  can be added to the right-hand side of (4) to denote the regressors and their corresponding slopes that are constant over time.

slopes to model the changes in long-run structural relationship between y and x, breaks  $\mathcal{K}_1$  in error loadings  $\gamma_i$  can be considered equivalent to instability of variance of errors  $\gamma'_i f_t + \varepsilon_{it}$  in (4), or changes in error factor variance with constant loadings.

In the special case of  $m_0 = 2, m_1 = 1$ , of model (5), we assume  $K_{0,1} < K_{0,2} < K_{1,1}$ , without loss of generality. Thus, three breaks  $K_{0,1}, K_{0,2}, K_{1,1}$  split the sample into 4 regimes:

$$y_{it} = \begin{cases} x'_{it}\beta_{i1} + \gamma'_{i1}f_t + \varepsilon_{it}, & t = 1, ..., K_{0,1} \\ x'_{it}\beta_{i2} + \gamma'_{i1}f_t + \varepsilon_{it}, & t = K_{0,1} + 1, ..., K_{0,2} \\ x'_{it}\beta_{i3} + \gamma'_{i1}f_t + \varepsilon_{it}, & t = K_{0,2} + 1, ..., K_{1,1} \\ x'_{it}\beta_{i3} + \gamma'_{i2}f_t + \varepsilon_{it}, & t = K_{1,1} + 1, ..., T, \end{cases}$$
(6)

each of which can be considered the same as KPY. This is also the case when there are multiple breaks in slopes and error factor loadings, i.e.,  $m_0 > 1$ ,  $m_1 > 1$ . We follow KPY and use the CCE approach to deal with unobserved nonstationary factors  $f_t$ . In this model, the parameters to be estimated include the slopes  $\beta_i(\mathcal{K}_0)$  and the break points  $\mathcal{K}_0, \mathcal{K}_1$ .

Like estimating break point  $\mathcal{K}_0$  in slopes, estimating  $\mathcal{K}_1$  in factor loadings is equally important. As pointed out in the growing literature since Stock and Watson (2009), the structural instability in the factor structure could have implications on the accuracy of forecasting and number of estimated factors. In our model (5) ignoring the break  $\mathcal{K}_1$  in  $\gamma_i$  could bias the estimates of the factor loadings in empirical studies, e.g., US housing markets by Holly, Pesaran and Yamagata (2010). In addition, when the focus is on  $\varepsilon_{it}$ , e.g., testing for remaining cross-sectional dependence in  $\varepsilon_{it}$  (Juodis and Reese, 2022), estimating  $\mathcal{K}_1$  is necessary for obtaining a consistent estimate of  $\varepsilon_{it}$ .

Compared with Bai, Kao and Ng's (2009) model of panel cointegration with nonstationary factors, our model (5) adds two new empirical features: heterogeneous slopes and structural breaks in slopes and factor loadings. Structural breaks here can be regarded as a different way of modeling parameter heterogeneity from the latent group structure considered by Huang et al. (2021). Besides, we apply the CCE approach to deal with unobserved factors, instead of the IPC approach used in the two papers above. In addition, different from BFK's (2016, 2019) models of a common structural break in heterogeneous panels with exogenous and endogenous regressors, this paper focuses on *multiple* breaks and *nonstationary* factors and regressors. In line with Bai, Kao and Ng (2009),  $f_t$  are treated as additional explanatory variables, instead of an error component in (5). Thus  $\mathcal{K}_0$  and  $\mathcal{K}_1$  are considered as multiple breaks in a linear regression and are estimated by least squares as proposed by Bai and Perron (1998).

As in the literature on nonstationary panels with factors, the major challenge in estimating our model (5) lies in the unobserved factors. In this paper, we adopt the CCE approach proposed by Pesaran (2006) and examined by KPY in the case of nonstationary factors. To simplify the analysis, we follow Stock and Watson's (2016, p.429) idea of using the cross-sectional averages of  $x_{it}$ ,  $\bar{x}_t = \frac{1}{N} \sum_{i=1}^{N} x_{it}$ , instead of those of  $y_{it}$  and  $x_{it}$ , to proxy for  $f_t$  in this paper.<sup>4</sup> The cross-sectional average of  $x_{it}$  in (3),

$$\bar{x}_t = \bar{\Gamma}' f_t + \bar{v}_t, \ \bar{\Gamma} = \frac{1}{N} \sum_{i=1}^N \Gamma_i \text{ and } \bar{v}_t = \frac{1}{N} \sum_{i=1}^N x v_{it}$$

When  $\overline{\Gamma}$  is of full rank  $(q \leq p)$ , like OLS,

$$f_t = (\bar{\Gamma}\bar{\Gamma}')^{-1}\bar{\Gamma}(\bar{x}_t - \bar{v}_t).$$
(7)

Since  $\bar{v}_t \to 0$  as  $N \to \infty$ , it is also asymptotically valid to use  $\bar{x}_t$  as observable proxies for nonstationary  $f_t$ ,

$$f_t - (\bar{\Gamma}\bar{\Gamma}')^{-1}\bar{\Gamma}\bar{x}_t \xrightarrow{p} 0 \text{ as } N \to \infty.$$
 (8)

Hence, the idea of CCE is being used for nonstationary factors in each regime.<sup>5</sup>

Using (7) for  $f_t$ , (5) can be written as

$$y_{it} = x'_{it}\beta_i(\mathcal{K}_0) + f'_t\gamma_i(\mathcal{K}_1) + \varepsilon_{it}$$
  
$$= x'_{it}\beta_i(\mathcal{K}_0) + [(\bar{\Gamma}\bar{\Gamma}')^{-1}\bar{\Gamma}(\bar{x}_t - \bar{v}_t)]'\gamma_i(\mathcal{K}_1) + \varepsilon_{it}$$
  
$$= x'_{it}\beta_i(\mathcal{K}_0) + \bar{x}'_t\gamma_i^*(\mathcal{K}_1) + \varepsilon_{it}^*, \qquad (9)$$

where  $\gamma_i^*(\mathcal{K}_1) = \bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1}\gamma_i(\mathcal{K}_1)$  and  $\varepsilon_{it}^* = \varepsilon_{it} - \bar{v}'_t\bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1}\gamma_i(\mathcal{K}_1)$ . Thus, by proxying  $f_t$  with observables, equation (9) can be regarded as a panel data regression with multiple common breaks  $\mathcal{K}_0, \mathcal{K}_1$  in slopes  $\beta_i$  and  $\gamma_i^*$ . In the special case of no breaks  $\mathcal{K}_1$  in loadings in model (4),  $\gamma_i^*(\mathcal{K}_1)$  in equation (9) becomes  $\gamma_i^* = \bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1}\gamma_i$ . In this paper, we consider

<sup>5</sup>As in KPY, when the rank condition holds, there is no need to estimate the number of error factors.

<sup>&</sup>lt;sup>4</sup>Karavias et al. (2023) use this proxy for  $f_t$ . BFK (2019) focus on estimating a single break point in heterogeneous slopes using the cross-sectional average  $(y_{it}, x_{it})$  to proxy for  $f_t$  and treat the error factor structure as nuisance parameters. This paper also estimates break points in error factor loadings  $\mathcal{K}_1$  along with  $\mathcal{K}_0$ . To simplify the analysis, we use the cross-sectional average  $x_{it}$  to proxy for  $f_t$ . In additional Monte Carlo simulations, we use the cross-sectional average  $(y_{it}, x_{it})$  to proxy for  $f_t$  and similar results are obtained.

the general model (5) and use least squares proposed by Bai and Perron (1998) to estimate break points ( $\mathcal{K}_0, \mathcal{K}_1$ ), slopes  $\beta_i(\mathcal{K}_0)$  and their cross-sectional averages.

**Remark 1**: Breitung and Eickmeier (2011) point out that the structural breaks in the factor loadings can be captured by inflating the number of factors in the PCA estimation. However, the inflated number of factors may fail the rank condition required by the CCE approach above. This implies that using the cross-sectional averages does not necessarily capture the inflated number of factors. As shown in the next section, our estimator of  $\mathcal{K}_0$  and  $\beta_i(\mathcal{K}_0)$  can be robust to the breaks  $\mathcal{K}_1$  in error factor structure in a simultaneous estimation approach. Identifying the breaks  $\mathcal{K}_1$  can be separately achieved if the rank condition is not satisfied with inflated number of factors.<sup>6</sup>

### 3 Estimation

To simplify notation, let  $z_{it} = (x'_{it}, \bar{x}'_t)', \ \delta_i(\mathcal{K}_0, \mathcal{K}_1) = (\beta_i(\mathcal{K}_0)', \gamma_i^*(\mathcal{K}_1)')'$ . Thus, equation (9) above can be written as

$$y_{it} = z'_{it}\delta_i(\mathcal{K}_0, \mathcal{K}_1) + \varepsilon^*_{it}.$$
(10)

We rearrange the  $m_0 + m_1$  breaks  $\mathcal{K}_0, \mathcal{K}_1$  in time line as  $\{\mathcal{K}^0\} = \{\mathcal{K}_0, \mathcal{K}_1\} = \{k_1^0, k_2^0, ..., k_m^0\}$ with  $m = m_0 + m_1$ . Superscript 0 denotes for true values of breaks. After reparameterization, model (10) can be considered as a panel data regression with multiple structural changes in slopes:

$$y_{it} = z'_{it}\delta_{ij} + \varepsilon^*_{it}, t = k^0_{j-1} + 1, \dots, k^0_j,$$
(11)

where j = 1, ..., m + 1, and  $k_0^0 = 0, k_{m+1}^0 = T$ .

**Remark 2:** Equation (11) can be considered as a panel data version of the multiple structural change model considered by Bai and Perron (1998) using nonstationary data. It also extends the stationary panel data model with one common break in BFK (2016) to the case of multiple common breaks with nonstationary data.

**Remark 3:** The intuition on identifying break points in this literature apply here as well. First, as pointed out by Bai (1997) and Bai and Perron (1998), the key information to identify the break points in time series regressions depend on the break magnitude

<sup>&</sup>lt;sup>6</sup>In this case, we can use partitioned regression to consistently estimate  $\mathcal{K}_0$  and  $\beta_i(\mathcal{K}_0)$  first when the rank condition is satisfied with a small number of factors. After  $\hat{\mathcal{K}}_0$  and  $\hat{\beta}_i(\hat{\mathcal{K}}_0)$  are obtained, PCA or other methods can be applied to identify the factor structure and the breaks in loadings in errors  $f'_t \gamma_i(\mathcal{K}_1) + \varepsilon_{it}$  estimated by  $y_{it} - x'_{it}\hat{\beta}_i(\hat{\mathcal{K}}_0)$ .

and the variance of the regressors relative to the variance of the errors. Second, in panels with mean shifts or (trend) stationary regressors, Bai (2010), Kim (2011) and BFK (2016) show that the break magnitude increases with N under the common break assumption. Thus the break point can be consistently estimated in panels as  $(N,T) \rightarrow \infty$ . Third, Baltagi, Kao and Liu (2017), Pang Du and Chong (2021) show that using nonstationary regressors, the variance of the regressors increases with T, implying that it is easier to identify break points in regressions using nonstationary rather than stationary regressors.

Define  $Y_i = (y_{i1}, \dots, y_{iT})'$ ,  $\delta_i = (\delta'_{i1}, \dots, \delta'_{i,m+1})'$ ,  $\underline{Z}_i(\mathcal{K}^0) = diag(Z_{i1}, \dots, Z_{i,m+1})$  with  $Z_{ij} = (z_{i,k_{j-1}^0+1}, \dots, z_{i,k_j^0})'$ ,  $j = 1, \dots, m+1$  and  $\varepsilon_i^* = (\varepsilon_{i1}^*, \dots, \varepsilon_{iT}^*)'$ . Thus, equation (11) can be written in matrix form: for  $i = 1, \dots, N$ ,

$$Y_i = \underline{Z}_i(\mathcal{K}^0)\delta_i + \varepsilon_i^* \tag{12}$$

For possible breaks  $\mathcal{K} = m$ -partition  $(k_1, ..., k_m)$ , the OLS estimator of  $\delta_i$  is  $\hat{\delta}_i(\mathcal{K}) = [\underline{Z}_i(\mathcal{K})'\underline{Z}_i(\mathcal{K})]^{-1}\underline{Z}_i(\mathcal{K})'Y_i$ , and the corresponding sum of squared residuals is

$$SSR_i(\mathcal{K}) = \left[Y_i - \underline{Z}_i(\mathcal{K})\hat{\delta}_i(\mathcal{K})\right]' \left[Y_i - \underline{Z}_i(\mathcal{K})\hat{\delta}_i(\mathcal{K})\right]$$

Thus, the OLS estimator of  $\mathcal{K}^0 = (k_1^0, ..., k_m^0)$  is defined as

$$\hat{\mathcal{K}} = (\hat{k}_1, ..., \hat{k}_m) = \arg\min_{(k_1, ..., k_m)} \frac{1}{NT} \sum_{i=1}^N SSR_i(\mathcal{K}).$$
(13)

Due to the computation complexity  $O_p(mT^2)$  of the grid search algorithm, obtaining  $(\hat{k}_1, ..., \hat{k}_m)$  by solving (13) is generally very time consuming when  $m \ge 3$  and T is large. In practice, we recommend the dynamic programming algorithm proposed by Bai and Perron (2003).

In this paper, we assume that m is known. This assumption can be relaxed by following the idea of sequential estimation based on parameter-consistancy tests by Bai and Perron (1998). Alternatively, m can be determined by an information criterion approach with a penalty factor related to m as in Boldea et al. (2020) who consider a fixed effects panel data model with multiple breaks.

Next, we consider the estimation of  $\beta_i(\mathcal{K}_0)$ . Denote  $X_i = (x_{i1}, \cdots, x_{iT})'$  and  $\bar{X} = (\bar{x}_1, \cdots, \bar{x}_T)'$ . Stacking the time dimension of equation (9) in matrix form gives

$$Y_i = X_i \beta_i(\mathcal{K}_0) + \bar{X} \gamma_i^*(\mathcal{K}_1) + \varepsilon_i^*.$$

Reparameterizing  $\underline{X}_{i}(\mathcal{K}_{0}) = diag(X_{i1}, X_{i2}, \cdots, X_{i,m_{0}+1})$  with  $\underbrace{X_{i1}}_{K_{0,1} \times p} = (x_{i1}, \dots, x_{i,K_{0,1}})',$  $\underbrace{X_{i2}}_{(K_{0,2}-K_{0,1}) \times p} = (x_{i,K_{0,1}+1}, \dots, x_{i,K_{0,2}})', \cdots, \underbrace{X_{i,m_{0}+1}}_{(T-K_{0,m_{0}}) \times p} = (x_{i,K_{0,m_{0}}+1}, \dots, x_{iT})' \text{ and } b_{i} = (\beta'_{i1}, \cdots, \beta'_{i,m_{0}+1})'$ gives

$$Y_i = \underline{X}_i(\mathcal{K}_0)b_i + \bar{X}\gamma_i^*(\mathcal{K}_1) + \varepsilon_i^*.$$
(14)

By partitioned regression in equation (14):

$$\hat{b}_{i} = \hat{b}_{i} \left( \hat{\mathcal{K}}_{0} \right) = \left[ \underline{X}_{i} (\hat{\mathcal{K}}_{0})' M_{\bar{X}} \underline{X}_{i} (\hat{\mathcal{K}}_{0}) \right]^{-1} \underline{X}_{i} \left( \hat{\mathcal{K}}_{0} \right)' M_{\bar{X}} Y_{i}, \tag{15}$$

where  $M_{\bar{X}} = I - \bar{X} \left( \bar{X}' \bar{X} \right)^{-1} \bar{X}'$ . Similarly, the mean of  $b_i$  can also be consistently estimated by the following mean-group estimator

$$\hat{b}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{b}_{i} = \frac{1}{N} \sum_{i=1}^{N} \left[ \underline{X}_{i}(\hat{\mathcal{K}}_{0})' M_{\bar{X}} \underline{X}_{i}(\hat{\mathcal{K}}_{0}) \right]^{-1} \underline{X}_{i} \left( \hat{\mathcal{K}}_{0} \right)' M_{\bar{X}} Y_{i}.$$
(16)

The partitioned regression (15) suggests that the CCE transformed regressors  $M_{\bar{X}} X_i(\hat{\mathcal{K}}_0)$ become stationary after partialling out I(1)  $f_t$  in the case of stationary  $v_{it}$ . This leads to  $\sqrt{T}$ -consistent  $\hat{b}_i$  as shown in the next Section. By contrast, when  $v_{it}$  follows an I(1)process,  $M_{\bar{X}} \underline{X}_i(\hat{\mathcal{K}}_0)$  remains nonstationary. In this case,  $y_{it}$  and  $x_{it}$  are cointegrated after dealing with the unobserved factors in each regime, and T-consistency of  $b_i$  can be obtained. This is different from the setup in KPY. We will consider I(0)  $v_{it}$  as Case 1 in Section 4, and I(1)  $v_{it}$  as Case 2 in Section 5.

#### Main Results 4

#### 4.1 Assumptions

The following assumptions are needed for establishing the asymptotic properties of the break and slope estimators above.

**Assumption 1**  $k_j^0 = [\lambda_j^0 T]$  with  $\lambda_j^0 \in (0, 1), j = \{1, \dots, m\}.$ 

Assumption 2  $Rank(\bar{\Gamma}) = q \leq p$ .

Assumption 3 Factor loadings  $\gamma_i(\mathcal{K}_1)$  and  $\Gamma_i$  are independent and identically distributed (IID) across *i*, and independent of  $\varepsilon_{jt}$ ,  $v_{jt}$  and  $f_t$  for all *i*, *j*, *t*. Assume  $\gamma_i(\mathcal{K}_1) = \gamma(\mathcal{K}_1) + \eta_i$ ,  $\eta_i \sim IID(0, \Sigma_{\gamma})$  and  $\Gamma_i = \Gamma + \xi_i, \ \xi_i \sim IID(0, \Omega_{\xi}), \ i = 1, ...N$ , where the means  $\gamma, \ \Gamma$  are non-zero and fixed and the variances  $\Omega_{\eta}$ ,  $\Omega_{\xi}$  are finite.

Assumption 4 For i = 1, ..., N,  $b_i = b + v_{b,i}, v_{b,i} \sim iid(0, \Sigma_b)$ , where  $b = (\beta'_1, \beta'_2, \cdots, \beta'_{m_0+1})'$ ,  $v_{b,i} = (v'_{\beta_1,i}, v'_{\beta_2,i}, \cdots v'_{\beta_{m_0+1},i})'$  and  $\Sigma_b = diag(\Sigma_{\beta_1}, \Sigma_{\beta_2}, \cdots, \Sigma_{\beta_{m_0+1}})$  for i = 1, 2, ..., N, where  $\|b\| < \infty$ ,  $\|\Sigma_b\| < \infty$ , and the random deviations  $v_{b,i}$  are independent of  $x_{it}$  and  $\varepsilon_{jt}$ for all i, j and t.

**Assumption 5**  $\varphi_t$  is an vector of  $L_{2+\vartheta}$ ,  $\vartheta > 0$  and stationary near epoque dependent process of size 1/2, on some  $\alpha$ -mixing process of size  $-(2+\vartheta)/\vartheta$  and independent of  $v_{jt}$ and  $\varepsilon_{jt}$  for all i, j, t.

**Assumption 6** Matrices  $\frac{1}{Th} \sum_{t=1}^{h} f_t f'_t$ ,  $\frac{1}{Th} \sum_{t=T-h+1}^{T} f_t f'_t$ , and  $\frac{1}{Th} \sum_{t=\lambda_j T+1}^{\lambda_j T+h} f_t f'_t$ , for  $j = \{1, \dots, m\}$ , have minimum eigenvalues bounded away from zero in probability for all  $1 \le h \le T$ .

Assumption 7 (i) Matrices  $\frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{h} z_{it} z'_{it}$ ,  $\frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=T-h+1}^{T} z_{it} z'_{it}$ ,  $\frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=\lambda_j T+1}^{\lambda_j T+h} z_{it} z'_{it}$ and  $\frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=\lambda_j T+1}^{\lambda_j T+h} z_{it} z'_{it}$ , for  $j = \{1, \dots, m\}$ , have minimum eigenvalues bounded away from zero in probability for  $1 \le h \le T$ ; (ii) for each t,  $\frac{1}{N} \sum_{i=1}^{N} z_{it} z'_{it}$  is stochastically bounded as  $N \to \infty$ .

Assumption 8 (i) The disturbances  $\varepsilon_{it}$ , i = 1, ..., N, are cross-sectionally independent; (ii) For each series i,  $\varepsilon_{it}$  is independent of  $\varphi_{t'}$  for all t and t'; (iii) errors  $\varepsilon_{is}$  and  $v_{jt}$  are independent for all i, j, s, t; (iv)  $\varepsilon_{it}$  is a stationary process with absolute summable autocovariances, such that  $\varepsilon_{it} = \sum_{l=0}^{\infty} a_{il}\zeta_{i,t-l}$ , where  $\{\zeta_{it}, t = 1, ..., T\}$  are IID random variables with zero mean and have a finite fourth-order moments. Assume  $0 < Var(\varepsilon_{it}) =$  $\sum_{l=0}^{\infty} a_{il}^2 = \sigma_i^2 < \infty$ . (v) for the  $T \times 1$  vector  $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{i,T})'$ ,  $Var(\varepsilon_i) = \Sigma_{\varepsilon,i}$  and  $0 < ||\Sigma_{\varepsilon,i}|| < \infty$ .

Assumption 9 (i) The disturbances  $v_{it}$ , i = 1, ..., N, are cross-sectionally independent; (ii) For each series i,  $v_{it}$  is independent of  $\varphi_{t'}$  for all t and t'; (iii)  $v_{it}$  are linear stationary processes with zero mean and absolute summable autocovariances,  $v_{it} = \sum_{l=0}^{\infty} \Xi_{il} v_{i,t-l}$ , where  $(\zeta_{it}, v'_{it})'$  are  $(p+1) \times 1$  vectors of IID random variables with variance-covariance matrix  $I_{p+1}$  and has a finite fourth-order moments, and  $Var(v_{it}) = \sum_{l=0}^{\infty} \Xi_{il} \Xi'_{il} = \Sigma_{v,i}$ , and  $0 < ||\Sigma_{v,i}|| < \infty$ . Assumption 10 For i = 1, ..., N,  $\frac{1}{T} \underline{X}_i(\mathcal{K}_0)' M_{\overline{X}} \underline{X}_i(\mathcal{K}_0)$  is nonsingular, and their inverses have finite second-order moments, and  $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{v,i}^{N}$  is nonsingular.

For  $j = \{1, \dots, m\}$ , define  $\phi_{N,j} = \sum_{i=1}^{N} (\delta_{i,j+1} - \delta_{ij})' (\delta_{i,j+1} - \delta_{ij})$  in equation (11) as the magnitude of common breaks in panels.

# Assumption 11 $\phi_{N,j} \to \infty$ , $\frac{T}{N}\phi_{N,j} \to \infty$ , as $(N,T) \to \infty$ for $j = \{1, \dots, m\}$ .

Assumption 1 is common in the time series and panel data literature of structural changes, e.g., Bai (1997), Bai and Perron (1998), Bai (2010), BFK (2016, 2019). It rules out the case that true breaks happen on the boundary of the observed time period. It also implies that there are sufficient number of observations between breaks for large sample approximation. However, Bai (2010) pointed out that the common breaks close to the boundary are allowed in a panel mean shift model when  $T/N \rightarrow 0$ . To simplify our proofs, we adopt this convenient assumption. We explore the performance of our break estimator in the case of boundary breaks in Monte Carlo experiments.

Assumption 2 on the rank condition guarantees that equation (7) is valid, see Pesaran (2006) and KPY who discuss the situation of rank deficiency. This assumption can be relaxed to accommodate more empirical situations. For example, Karabiyik, Urbain and Westerlund (2019) consider the case of p < q. When p < q, additional exogenous co-variates should be included to proxy the unobserved error factors. Karabiyik, Reese and Westerlund (2017) explore the case when using too many observables causes the second moment matrix of the estimated factors to become asymptotically singular. Juodis, Karabiyik and Westerlund (2021) establish the theory of CCE allowing common factors to be correlated with the regressors. Our theoretical results can be extended to the case of rank deficiency by following the papers mentioned above. We will explore the performance of the estimators in case that Assumption 2 is not satisfied in the Monte Carlo simulations.

Assumptions 3, and 7 are borrowed from BFK (2016, 2019). Assumptions 4, 5 on random coefficients, and 10 on the identification condition for the individual slopes are borrowed from KPY. Under Assumptions 8 and 9, the idiosyncratic errors  $\varepsilon_{it}$  and  $v_{it}$  follow a general linear stationary process with heteroscedasticity and autocorrelation for each *i*. Assumption 11 specifies the relationship between T/N and the magnitude of breaks  $\phi_{N,j}$ , j = 1, ..., m.  $\phi_{N,j}$  can grow slower or faster than N, depending on the relative rate of T/N. The condition on the magnitude of breaks in Assumption 11 generalizes Assumption 2 in stationary panels considered by BFK (2019) to the multiple breaks case.

Under these assumptions, we can show that the multiple breaks are estimated consistently, as summarized in the following theorem:

**Theorem 1** Under Assumptions 1-11, 
$$\lim_{(N,T)\to\infty} P\left(\hat{k}_j = k_j^0\right) = 1, j = \{1, \cdots, m\}$$

The rate of convergence and the distribution of the estimated structural breaks in stationary or nonstationary homogeneous panels have been discussed by Bai (2010), Baltagi, Kao and Liu (2017) and others. As pointed out by Bai (2020), Theorem 1 implies a degenerate limiting distribution for  $\hat{k}_j$ . To obtain a non-degenerate distribution, a different framework of shrinking magnitude of breaks is usually assumed. Baltagi, Kao and Liu (2017) show the convergence rates of break estimator in homogeneous cointegrated panels and stationary panel regression are  $O_p(1/NT)$  and  $O_p(1/N)$ , respectively, suggesting the benefit of using observations in the cross-sectional dimension under the common break assumption in panels. In our model, similar insights can be carried over. However, when the slopes are heterogeneous, the derivation of convergence rate and limiting distribution of the break point estimators is technically nontrivial. In addition, as shown in the following proposition, the convergence rate of  $\hat{k}_j$  is not required for the asymptotic distribution of the slope estimators, so we leave it for future research.

Given the consistency of estimated structural breaks  $\hat{\mathcal{K}}$  above, we can obtain consistent estimators of the slope parameters.

**Proposition 1** Under Assumptions 1-11, as  $(N,T) \to \infty$ , and  $\frac{\sqrt{T}}{N} \to 0$ , for each  $i = \{1, \dots, N\},\$ 

$$\sqrt{T}\left(\hat{b}_i - b_i\right) \xrightarrow{d} N\left(0, \Sigma_{X,i}^{-1} \Sigma_{X\varepsilon,i} \Sigma_{X,i}^{-1}\right),$$

where  $\Sigma_{X,i} = plim_{T \to \infty} \frac{1}{T} \underline{X}_i(\mathcal{K}_0)' M_{\bar{X}} \underline{X}_i(\mathcal{K}_0) \text{ and } \Sigma_{X\varepsilon,i} = plim_{T \to \infty} \frac{1}{T} \underline{X}_i(\mathcal{K}_0) M_{\bar{X}} \Sigma_{\varepsilon,i} M_{\bar{X}} \underline{X}_i(\mathcal{K}_0)'.$ 

As in Pesaran (2006), KPY and BFK, a consistent Newey-West type estimator of  $\Sigma_{X\varepsilon,i}$  can be obtained as

$$\widehat{\Sigma}_{X\varepsilon,i} = \widehat{\Lambda}_{i0} + \sum_{j=1}^{\omega} \left( 1 - \frac{j}{\omega+1} \right) \left( \widehat{\Lambda}_{ij} + \widehat{\Lambda}'_{ij} \right), \ \widehat{\Lambda}_{ij} = \frac{1}{T} \sum_{t=j+1}^{\omega} e_{it} e_{i,t-j} \underline{X}_{it}(\widehat{\mathcal{K}}_0) \underline{X}_{it}(\widehat{\mathcal{K}}_0)', \ (17)$$

where  $\omega$  is the window size,  $e_{it}$  is the  $t^{th}$  element of  $e_i = M_{\bar{X}}Y_i - M_{\bar{X}}\underline{X}_i(\hat{\mathcal{K}}_0)\hat{b}_i$  and  $\underline{X}_{it}(\hat{\mathcal{K}}_0)$  is the  $t^{th}$  row of  $M_{\bar{X}}\underline{X}_i(\hat{\mathcal{K}}_0)$ . Thus, a consistent Newey-West type estimator of  $\Sigma_{X,i}^{-1}\Sigma_{X\varepsilon,i}\Sigma_{X,i}^{-1}$  is given by

$$\left[\frac{1}{T}\underline{X}_{i}(\hat{\mathcal{K}}_{0})'M_{\bar{X}}\underline{X}_{i}(\hat{\mathcal{K}}_{0})\right]^{-1}\widehat{\Sigma}_{X\varepsilon,i}\left[\frac{1}{T}\underline{X}_{i}(\hat{\mathcal{K}}_{0})'M_{\bar{X}}\underline{X}_{i}(\hat{\mathcal{K}}_{0})\right]^{-1}.$$
(18)

**Proposition 2** Under Assumptions 1-11, and  $(N,T) \rightarrow \infty$ ,

$$\sqrt{N}\left(\hat{b}_{MG}-b\right) \xrightarrow{d} N\left(0,\Sigma_b\right),$$

where  $\Sigma_b$  can be consistently estimated by

$$\frac{1}{N-1}\sum_{i=1}^{N} \left(\hat{b}_i - \hat{b}_{MG}\right) \left(\hat{b}_i - \hat{b}_{MG}\right)'.$$

### 5 Nonstationary Regressors

In this section, our analysis of nonstationary panels is extended to the case of both nonstationary  $f_t$  and  $v_{it}$ . Idiosyncratic errors  $\varepsilon_{it}$  remain I(0). Compared with Section 5 of Phillips and Moon (1999), our model acommodates additional features of an error factor structure and multiple breaks in slopes. In equation (3)  $x_{it} = \Gamma'_i f_t + v_{it}$ , errors  $v_{it}$ follow I(1) processes:

$$v_{it} = v_{i,t-1} + \varsigma_{it}, \ i = 1, ..., N,$$
(19)

where  $\varsigma_{it}$  follows the assumption below:

Assumption 12  $\varsigma_{it}$ , i = 1, ..., N, are cross-sectionally independent. For each i,  $(i) \varsigma_{it} = \Psi_i(L)\epsilon_{it}$  with  $\epsilon_{it}$  is IID random variables with zero mean and has a finite fourth-order moments; (ii)  $Var(\epsilon_{it}) = \Sigma_{\epsilon,i} = P_i P'_i$ , and  $\Psi_i(L) = \sum_{j=0}^{\infty} \Psi_{ij} L^j$  with  $\sum_{j=0}^{\infty} j ||\Psi_{ij}|| < \infty$ , and  $\Psi_i(1) = \sum_{j=0}^{\infty} \Psi_{ij}$ .

Different from Case 1 of stationary  $v_{it}$  considered in Section 4, in Case 2 of  $I(1) v_{it}$ , the CCE transformed regressors in the partitioned regression (15) remain nonstationary. We will show that  $\hat{\mathcal{K}}$  defined in equation (13) above are still consistent and  $\hat{b}_i$  is *T*-consistent. In addition, different from Case 1, the restriction on the relative diverging rate between *T* and *N* in Assumption 11 is not required here. **Theorem 2** Under Assumptions 1-10, 12, as  $(N,T) \to \infty$ ,  $\lim_{(N,T)\to\infty} P\left(\hat{k}_j = k_j^0\right) = 1, j = \{1, \dots, m\}.$ 

With an additional Assumption 13 on  $\varphi_t$  below, we obtain the following Proposition 3. In line with equation (5.8) in Phillips and Moon (1999) in nonstationary heterogeneous panels without structural breaks and error factors, for each i = 1, ..., N,  $\hat{b}_i$  is also super consistent in our model.

Assumption 13  $\varphi_t$  is linear stationary process, (i)  $\varphi_t = \Pi(L)u_t$  with  $\mu_t$ , t = 1, ..., T, have a finite fourth-order moments; (ii)  $Var(u_t) = \Sigma_u = QQ'$ , and  $\Pi(L) = \sum_{j=0}^{\infty} \Pi_j L^j$ with  $\sum_{j=0}^{\infty} j \|\Pi_j\| < \infty$ , and  $\Pi(1) = \sum_{j=0}^{\infty} \Pi_j$ .

**Proposition 3** Under Assumptions 1-8, 11, 12 and 13, for each i,  $T(\hat{b}_i - b_i)$  converges weakly to a non-degenerate distribution, as  $(N,T) \to \infty$ .

Intercept estimator is not included in  $\hat{b}_i$  above, and its convergence rate is  $\sqrt{T}$  as in a cointegration model (Hamilton, 1994, p.588). The intercept can be wiped out by adding vector of ones to  $\bar{X}$  in the  $M_{\bar{X}}$ .

For the mean group estimator of b,

$$\sqrt{N}\left(\hat{b}_{MG}-b\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} v_{b,i} + \frac{1}{\sqrt{N}T}\sum_{i=1}^{N} \left[ \left(\frac{1}{T^2}\underline{X}_i(\mathcal{K}_0)'M_{\bar{X}}\underline{X}_i(\mathcal{K}_0)\right)^{-1} \frac{1}{T}\underline{X}_i(\mathcal{K}_0)'M_{\bar{X}}\varepsilon_i \right] + o_p\left(1\right)$$

$$\tag{20}$$

The second term is dominated by the first term in the above equation. Thus, we can obtain a similar result to Proposition 2 in Case 1:  $\sqrt{N} \left( \hat{b}_{MG} - b \right) \xrightarrow{d} N(0, \Sigma_b)$  as  $(N, T) \to \infty$ .

In a special case of homogeneous slopes  $b_i = b$  with  $v_{b,i} = 0$ , the first term in equation (20) disappears. Thus, equation (20) reduces to

$$\sqrt{N}T\left(\hat{b}_{MG}-b\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \left(\frac{1}{T^2} \underline{X}_i(\mathcal{K}_0)' M_{\bar{X}} \underline{X}_i(\mathcal{K}_0)\right)^{-1} \frac{1}{T} \underline{X}_i(\mathcal{K}_0)' M_{\bar{X}} \varepsilon_i \right] + o_p\left(1\right).$$
(21)

The convergence rate of  $\hat{b}_{MG}$  in a homogeneous panel becomes  $\sqrt{NT}$ , same as in Bai, Kao and Ng (2009) and Huang et al. (2020).

With an additional Assumption 14 below, we obtain the following Proposition 4.

**Assumption 14**  $\frac{1}{T^2} \underline{X}_i(\hat{\mathcal{K}}_0)' M_{\overline{X}} \underline{X}_i(\hat{\mathcal{K}}_0)$  is nonsingular, and its inverse has a finite second-order moment.

**Proposition 4** Under Assumptions 1-8, 11 and 12-14, in a homogeneous panel with  $b_i = b$ , as  $(N,T) \rightarrow \infty$ ,

$$\sqrt{N}T\left(\hat{b}_{MG}-b\right) \xrightarrow{d} N\left(0,\Sigma_{MG}\right),$$

where  $\Sigma_{MG}$  is defined in the Appendix A.4.

For simplicity, asymptotic bias mentioned in Theorem 8 of Phillips and Moon (1999) and Proposition 1 of Bai, Ng and Kao (2009) disappears here under the assumptions of no serial/ cross-sectional correlation and heteroskedasticity. In addition, we leave the expression of  $\Sigma_{MG}$  in the appendix since its form is too complicated to be used in practice.

### 6 Monte Carlo Simulations

In this section, Monte Carlo experiments are conducted to examine the finite sample properties of the break estimators. We consider the case of three breaks, i.e., m = 3, including two common breaks in slopes  $(k_1^0, k_2^0)$  and a third one in error factor loadings  $k_3^0$ in various scenarios. We find supporting results to the main findings in Theorems 1 and 2. This is done by looking at the frequency of choosing true breaks using the proposed break estimators. For nonstationary panels, nonstationarity could come from either  $f_t$  or  $v_{it}$  or both under the common factor assumption (3). Thus, we consider six different scenarios: i) Case 1 with I(1) factors  $f_t$  and  $I(0) v_{it}$ ; ii) Case 1 under rank deficiency; iii) Case 2 of a panel cointegration model with I(1)  $f_t$  and  $I(1) v_{it}$ ; iv) Case 2 with I(0)  $f_t$  and I(1) $v_{it}$ ; v) Case 2 with I(1) errors  $\varepsilon_{it}$ ; vi) Case 1 with mixed stationary and nonstationary regressors and factors.

### 6.1 Data Generating Process

Our basic design is similar to the one used in KPY but now with multiple breaks:

$$y_{it} = \alpha_i + \beta_i \left( k_1^0, k_2^0 \right) x_{it} + \gamma_{1,i} \left( k_3^0 \right) f_t + \varepsilon_{it}, \ i = 1, ..., N; \ t = 1, ..., T,$$
(22)

where  $\alpha_i \sim iidN(1,1)$ . The scalar regressor  $x_{it}$  is affected by the common correlated effect  $f_t$ :

$$x_{it} = a_i + \gamma_{2,i} f_t + v_{it}, (23)$$

with  $a_i \sim iidN(0.5, 0.5)$  and  $\gamma_{2,i} \sim iidN(0.5, 0.5)$ . The scalar factor  $f_t$  follows an I(1) process:

$$f_t = f_{t-1} + v_{ft}, t = -49, \dots, 0, 1, \dots, T;$$

where  $f_{-50} = 0$ ,  $v_{ft} \sim iidN(0, 1)$ .

Two common breaks  $k_1^0, k_2^0$  in slopes are assumed at [0.3T] and [0.5T] of the time span:

$$\beta_i(k_1^0, k_2^0) = \begin{cases} \beta_i, & t = 1, ..., k_1^0, \\ \beta_i + \Delta \beta_i, & t = k_1^0 + 1, ..., T, \\ \beta_i + 2\Delta \beta_i, & t = k_2^0 + 1, ..., T \end{cases}$$

where  $\beta_i \sim iidN(1, 0.04)$  and  $\Delta\beta_i \sim iidN(0, 0.5)$ . A third break  $k_3^0 = [0.7T]$  occurs in the error factor loadings:

$$\gamma_{1,i}(k_3^0) = \begin{cases} \gamma_{1,i}, & t = 1, \dots, k_3^0, \\ \gamma_{1,i} + \Delta \gamma_i, & t = k_3^0 + 1, \dots, T, \end{cases}$$
(24)

where  $\gamma_{1,i} \sim iidN(1,0.2)$  and  $\Delta \gamma_i \sim iidN(0.5,0.5)$ .

In scenario (i) of Case 1, as in KPY, both  $\varepsilon_{it}$  and  $v_{it}$  are stationary.  $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i \left(1 - \rho_{i\varepsilon}^2\right)^{0.5} \omega_{it}$ , for i = 1, 2, ..., [N/2] and  $\varepsilon_{it} = \sigma_i \left(1 + \theta_{i\varepsilon}^2\right)^{-0.5} (\omega_{it} + \theta_{i\varepsilon}\omega_{i,t-1})$ , for i = [N/2] + 1, ..., N, with  $\omega_{it} \sim iidN(0, 1), \sigma_i^2 \sim iidU[0.5, 1.5], \rho_{i\varepsilon} = iidU[0.05, 0.95]$  and  $\theta_{i\varepsilon} \sim iidU[0, 1]$ . Similarly,  $v_{it} = \rho_{vi}v_{i,t-1} + \psi_{it}, \psi_{it} \sim iidN(0, 1 - \rho_{vi}^2)$ , with  $v_{i,-49} = 0$ , and  $\rho_{vi} \sim iidU[0.05, 0.95]$ .<sup>7</sup>

In scenario (ii), we consider the importance of rank deficiency in finite samples. The DGP here is the same as above, except that the means of  $a_i$  and  $\gamma_{2,i}$  change to zero, i.e.,  $a_i \sim iidN(0, 0.5)$  and  $\gamma_{2,i} \sim iidN(0, 0.5)$  in equation (23). In the current design, the rank condition is not satisfied asymptotically.

In scenario (iii) of Case 2 of panel cointegration, both  $v_{it}$  and  $\varepsilon_{it}$  follow I(1) processes,

$$v_{it} = v_{i,t-1} + \psi_{it}, \ \psi_{it} \sim iidN(0,1), \ t = -49, ..., 0, 1, ..., T.$$

We also allow for I(0)  $f_t$  in the design above in scenario (iv). In addition, in scenario (v), we examine the impact of nonstationary errors on break point estimators, we also consider Case 2 with nonstationary errors, i.e.,  $I(1) \varepsilon_{it}$ ,  $\varepsilon_{it} = \varepsilon_{i,t-1} + \vartheta_{it}$ ,  $\vartheta_{it} \sim iidN(0,1)$ , t = -49, ..., 0, 1, ..., T.

<sup>&</sup>lt;sup>7</sup>In this design, the signal-to-noise ratio is about 1.5.

Finally, in scenario (vi), we also consider the case of mixed stationary and nonstationary regressors and factors. To allow for a stationary regression, we add an additional regressor and factor in the regression (22) above. More specifically,

$$y_{it} = \alpha_i + \beta_{1,i} \left( k_1^0 \right) x_{1,it} + \beta_{2,i} \left( k_2^0 \right) x_{2,it} + \gamma_{11,i} \left( k_3^0 \right) f_{1,t} + \gamma_{12,i} \left( k_3^0 \right) f_{2,t} + \varepsilon_{it},$$

where both regressors are generated by

$$x_{1,it} = a_i + \gamma_{21,i} f_{1,t} + \gamma_{22,i} f_{2,t} + v_{1,it},$$
  
$$x_{2,it} = a_i + 0 \cdot f_{1,t} + \gamma_{23,i} f_{2,t} + v_{2,it}.$$

We assume that both  $v_{1,it}$  and  $v_{2,it}$  are I(0) as  $v_{it}$  in Case 1 above. Two factors  $f_{1,t}$  and  $f_{2,t}$  are generated as I(1) and I(0) processes, respectively, as follows:

$$f_{1,t} = f_{1,t-1} + v_{1,ft}$$
, and  $f_{2,t} = 0.5f_{2,t-1} + v_{2,ft}$ 

Thus,  $x_{1,it}$  is I(1) and  $x_{2,it}$  is I(0). Same as  $\gamma_{2i}$ , loadings  $\gamma_{21,i}, \gamma_{22,i}, \gamma_{23,i} \sim iidN(0.5, 0.5)$ . The break points  $k_1^0 = [0.3T], k_2^0 = [0.5T]$  appear in the slopes:

$$\begin{split} \beta_{1,i} \left( k_{1}^{0} \right) &= \begin{cases} \beta_{11,i}, & t = 1, \dots, k_{1}^{0}, \\ \beta_{11,i} + \Delta \beta_{1,i}, & t = k_{1}^{0} + 1, \dots, T, \end{cases} \\ \beta_{2,i} \left( k_{2}^{0} \right) &= \begin{cases} \beta_{21,i}, & t = 1, \dots, k_{2}^{0}, \\ \beta_{21,i} + \Delta \beta_{2,i}, & t = k_{2}^{0} + 1, \dots, T, \end{cases}$$

where  $\Delta\beta_{1,i}, \Delta\beta_{2,i} \sim iidN(0, 0.16)$ . Here  $\gamma_{11,i}(k_3^0)$  and  $\gamma_{12,i}(k_3^0)$  have the same design as  $\gamma_{1,i}(k_3^0)$  in (24) but the variance of  $\Delta\gamma_i$  changes from 0.5 to 0.16.

Different combinations of T = 20, 50, 100 and N = 10, 50, 200 are considered in the Monte Carlo experiments with 1,000 replications. Due to limited space, only the results with T = 50 are reported in the paper.

### 6.2 Results

Figure 1 presents the histograms of estimators  $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$  in Case 1 with nonstationary factors for T = 50. The true values of the break points are  $k_1^0 = 15$ ,  $k_2^0 = 25$ ,  $k_3^0 = 35$ . In each replication, a dynamic programming algorithm proposed by Bai and Perron (2003) is applied to obtain  $\hat{k}_1$ ,  $\hat{k}_2$ ,  $\hat{k}_3$  simultaneously. The upper, middle and lower panels represent the empirical distributions of  $\hat{k}_1$ ,  $\hat{k}_2$  and  $\hat{k}_3$ , respectively. Figure 1 shows that the frequencies of choosing  $(k_1^0, k_2^0, k_3^0)$  increase substantially as N increases from 10 to 200. For example, the probability of choosing  $k_1^0$  increases from 36% for N = 10 to 69% for N = 200. This finding supports the results in Theorem 1.

Figure 2 reports the histograms of  $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$  in Case 1 for T = 50 under rank deficiency. The rank condition is required for the validity of the CCE approach to deal with unobserved common factors. We examine the finite sample properties of these break estimators when the rank condition is not satisfied asymptotically. Although the probabilities of choosing the true break points are smaller than those in Figure 1, they still increase substantially with N, showing that under rank deficiency, the estimators  $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$  are still very informative about choosing  $(k_1^0, k_2^0, k_3^0)$  when N is large.

In Figure 3, we consider Case 2 of panel cointegration with nonstationary regressors and both  $f_t$  and  $v_{it}$  are nonstationary in  $x_{it}$ . Similar patterns as in Figure 1 are observed. The probabilities of choosing true break dates increase with N, e.g., nearly 100% for choosing  $k_1^0$  by  $\hat{k}_1$  for N = 200 and T = 50. This finding supports the consistency of the break estimators in Theorem 2. In Figure 4, we also consider a scenario of an I(0)  $f_t$  and I(1)  $v_{it}$  in Case 2, where  $f_t = 0.5f_{t-1} + v_{ft}$  and  $v_{ft} \sim iidN(0, 0.75)$ . As expected, as long as  $x_{it}$  is still I(1),  $\hat{k}_1$ ,  $\hat{k}_2$ ,  $\hat{k}_3$  are consistent. Little impact is spotted from changing  $f_t$  from I(1) to I(0) in Figure 4.

In Figure 5, we consider the scenario of nonstationary errors  $\varepsilon_{it}$  in the design of Case 2 above. Under the current design,  $f_t, v_{it}$  and  $\varepsilon_{it}$  follow I(1) processes. Different from Case 2,  $I(1) \varepsilon_{it}$  could lead to a spurious regression and thus, the least squares estimators of slopes could be inconsistent. In addition, nonstationary  $\varepsilon_{it}$  could lead to a smaller signal-to-noise ratio in the DGP of Figure 5 than that of Figure 3 with  $I(0) \varepsilon_{it}$ . Thus, we observe smaller probabilities of choosing  $(k_1^0, k_2^0, k_3^0)$  here, even though the same pattern remains. That is, big N helps to date the break points.

Lastly, we examine the scenario of mixed stationary and nonstationary regressors in Figure 6, as in Bai, Kao and Ng (2009), Huang, Jin, Phillips, Su (2021). Slightly different from the designs used in Figures 1-4, an additional regressor and factor are added to the design (22). In our modified design, given an I(1)  $f_{1,t}$  and an I(0)  $f_{2,t}$ ,  $x_{1,it}$ ,  $x_{2,it}$  are I(1)and I(0), respectively. We consider I(0)  $v_{it}$  in this scenario to avoid potential spurious regression after  $f_{1,t}$  and  $f_{2,t}$  are partialled out from the regressors and  $y_{it}$ . As expected, the frequency of choosing  $k_2^0$ , the break point in the stationary regressor, is smaller than that of choosing  $k_1^0$  under the same design parameters for a same N. After scaling up the magnitude of the break in  $\beta_{2,i}(k_2^0)$ , we find a similar pattern as in Figure 1, still observing increasing probabilities of dating the true break points with N in the histograms of  $\hat{k}_1$ ,  $\hat{k}_2$ ,  $\hat{k}_3$ .

Moreover, we also conduct additional robustness checks, including using  $(\bar{y}_{\cdot t}, \bar{x}_{\cdot t})$  instead of  $\bar{x}_{i\cdot}$ , to proxy  $f_t$ , boundary breaks, fixed effects model, different magnitude of breaks in slopes and factor loadings, adding a time trend etc. These results can be found in Figures A1-A6 in the supplementary Appendix B. The results with T = 20 and 100 are in line with those with T = 50 reported above, and are available upon request from the authors.

[Insert Figures 1-6 Here]

## 7 Conclusion

This paper proposes the estimation of unknown multiple structural breaks both in slopes and factor loadings in nonstationary panels with common factors. Based on KPY's approach for dealing with nonstationary factors in panels, we extend Bai and Perron's least squares estimator for multiple breaks in time series regression to nonstationary heterogeneous panels with unobserved factors in errors. We show that the proposed estimators, including the estimated structural breaks and slopes, are consistent in both cases of nonstationary factors and nonstationary regressors. These main findings are supported by the Monte Carlo simulations.

There are potentially two important issues to explore in the current framework. One is testing for multiple structural changes in nonstationary panels. In this paper, we only assume multiple breaks in slopes and factor loadings and estimate these break points. It would be meaningful to test the existence of the breaks in many empirical studies before applying our estimation methods. A candidate is to extend Bai and Perron's (1998)  $\sup F$  or double maximum tests into nonstationary panels. Another important issue is related to sequential estimation of the break points. In this paper, we estimate multiple breaks simultaneously. In the case of mixed stationary and nonstationary factors and regressors as considered in Figures 4 and 5, it would matter a lot whether breaks are estimated simultaneously or sequentially. It would be interesting to explore the asymptotic properties of sequential estimation of multiple breaks as in Bai and Perron (1998) and Pang, Du and Chong (2021). We leave these research questions for future research.

### References

- [1] Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4), 551-563.
- [2] Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229-1279.
- [3] Bai, J. (2010). Common breaks in means and variances for panel data. Journal of Econometrics, 157(1): 78-92.
- [4] Bai, J., & Kao, C. (2006). On the estimation and inference of a panel cointegration model with cross-sectional dependence. *Contributions to Economic Analysis*, 274, 3-30.
- [5] Bai J, Kao, C, and Ng, S. (2009). Panel cointegration with global stochastic trends. Journal of Econometrics, 149(1): 82-99.
- [6] Bai, J., Han, X., & Shi, Y. (2020). Estimation and inference of change points in high-dimensional factor models. *Journal of Econometrics*, 219(1), 66-100.
- [7] Bai, J., & Li, K. (2014). Theory and methods of panel data models with interactive effects. Annals of Statistics, 42(1), 142-170.
- [8] Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1), 47-78.
- [9] Baltagi, B. H., Feng, Q., & Kao, C. (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, 191(1), 176-195.
- [10] Baltagi, B. H., Feng, Q., & Kao, C. (2019). Structural changes in heterogeneous panels with endogenous regressors. *Journal of Applied Econometrics*, 34(6), 883-892.
- [11] Baltagi, B. H., Kao, C., & Liu, L. (2017). Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term. *Econometric Reviews*, 36(1-3), 85-102.
- [12] Baltagi, B. H., Kao, C., & Wang, F. (2015). Change point estimation in large heterogeneous panels. Manuscript, Syracuse University.
- [13] Baltagi, B. H., Kao, C., & Wang, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics*, 197(1), 87-100.
- [14] Baltagi, B. H., Kao, C., & Wang, F. (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics*, 220(2), 349-365.
- [15] Breitung, J., & Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1), 71-84.
- [16] Boldea, O., Drepper, B., & Gan, Z. (2020). Change point estimation in panel data with time-varying individual effects. *Journal of Applied Econometrics*, 35(6), 712-727.
- [17] Chen, L., Dolado, J. J., & Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, 180(1), 30-48.

- [18] Cheng, X., Liao, Z., & Schorfheide, F. (2016). Shrinkage estimation of highdimensional factor models with structural instabilities. *Review of Economic Studies*, 83(4), 1511-1543.
- [19] Dong, C., Gao, J., & Peng, B. (2021). Varying-coefficient panel data models with nonstationarity and partially observed factor structure. *Journal of Business & Economic Statistics*, 39(3), 700-711.
- [20] Ditzen, J., Karavias, Y., & Westerlund, J. (2022). Multiple structural breaks in interactive effects panel data and the impact of quantitative easing on bank lending. arXiv:2211.06707.
- [21] Duan, J., Bai, J., & Han, X. (2023). Quasi-maximum likelihood estimation of break point in high-dimensional factor models. *Journal of Econometrics*, 233(1), 209-236.
- [22] Feng, Q., and C. Kao (2020), Large-Dimensional Panel Data Econometrics: Testing, Estimation and Structural Changes, World Scientific.
- [23] Hamilton, J. (1994), *Time Series Analysis*, Princeton University Press.
- [24] Holly, S., Pesaran, M. H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, 158(1), 160-173.
- [25] Hsiao, C., Shi, Z., & Zhou, Q. (2021). Transformed Estimation for Panel Interactive Effects Models. *Journal of Business & Economic Statistics*, 1-18.
- [26] Hsiao, C. (2018). Panel models with interactive effects. Journal of Econometrics, 206(2), 645-673.
- [27] Huang, W., Jin, S., Phillips, P. C. B., & Su, L. (2021). Nonstationary panel models with latent group structures and cross-section dependence. *Journal of Econometrics*, 221(1), 198-222.
- [28] Huang, W., Jin, S., & Su, L. (2020). Identifying latent grouped patterns in cointegrated panels. *Econometric Theory*, 36(3), 410.
- [29] Juodis, A., Karabiyik, H., & Westerlund, J. (2021). On the robustness of the pooled CCE estimator. *Journal of Econometrics*, 220(2), 325-348.
- [30] Juodis, A., Reese S., (2022). The Incidental Parameters Problem in Testing for Remaining Cross-Section Correlation, *Journal of Business & Economic Statistics*, 40(3), 1191-1203.
- [31] Kaddoura, Y., & Westerlund, J. (2023). Estimation of panel data models with random interactive effects and multiple structural breaks when T is fixed. *Journal of Business & Economic Statistics*, 41(3), 778-790.
- [32] Kapetanios, G., Pesaran, M. H., & Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics*, 160(2), 326-348.
- [33] Karavias, Y., Narayan, P. K., & Westerlund, J. (2023). Structural breaks in interactive effects panels and the stock market reaction to COVID-19. *Journal of Business* & *Economic Statistics*, 41(3), 653-666.

- [34] Karabiyik, H., Reese, S., & Westerlund, J. (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics*, 197(1), 60-64.
- [35] Karabiyik, H., Urbain, J. P., & Westerlund, J. (2019). CCE estimation of factoraugmented regression models with more factors than observables. *Journal of Applied Econometrics*, 34(2), 268-284.
- [36] Kim, D. (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics*, 164(2), 310-330.
- [37] Kim, D. (2014). Common breaks in time trends for large panel data with a factor structure. *Econometrics Journal*, 17(3), 301-337.
- [38] Li, D, Qian, J, Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516): 1804-1819.
- [39] Lumsdaine, R. L., Okui, R., & Wang, W. (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, 233(1), 45-65.
- [40] Ma, C., & Tu, Y. (2023). Group fused Lasso for large factor models with multiple structural breaks. *Journal of Econometrics*, 233(1), 132-154.
- [41] Pang, T., Du, L and Chong, T. (2021), Estimating multiple breaks in nonstationary autoregressive models, *Journal of Econometrics*, 221 (1), 2021, 277-311.
- [42] Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4), 967-1012.
- [43] Phillips, P. C. B., and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5), 1057-1111.
- [44] Stock, J. H., & Watson, M. (2009). Forecasting in dynamic factor models subject to structural instability. The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry, 173, 205.
- [45] Stock, J. H., & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Handbook of macroeconomics (Vol. 2, pp. 415-525). Elsevier.
- [46] Yamamoto, Y., & Tanaka, S. (2015). Testing for factor loading structural change under common breaks. *Journal of Econometrics*, 189(1), 187-206.

## Appendixes: Proofs of Theorems and Propositions

The section includes detailed proofs of the main results in the text. To further simplify notation, in this section we consider the case of three breaks, m = 3, including two in slopes,  $(k_1^0, k_2^0)$ , and one in error factor loadings,  $k_3^0$ . The proofs of the general case in model (10) can be presented at the cost of additional notation.

Specifically, Appendix A includes detailed proofs of Theorems 1 and 2, Propositions 1-4. Subsection A.1 provides necessary Lemmas and detailed proof of Theorem 1. Similarly, subsection A.2 provides necessary lemmas and proof of Theorem 2. Lastly, Subsection A.3 provides proofs of Propositions 1- 2, and A.4 provides proofs of Propositions 3-4 respectively. Detailed proofs of lemmas are collected in the supplementary Appendix B.

# Appendix A: Proofs of Theorems and Propositions

#### A.1 Proof of Theorem 1

#### Proof of Theorem 1.

Following Bai and Perron (1998), we decompose the analysis of multiple breaks into several problems involving a single structural change in each. Without loss of generality, we only provide the proof of  $\lim_{(N,T)\to\infty} P(\hat{k}_1 = k_1^0) = 1$ . The proof of  $\lim_{(N,T)\to\infty} P(\hat{k}_j = k_j^0) = 1$ , j = 2, 3, can be shown similarly and is omitted.

To show  $\hat{k}_1 - k_1^0 \xrightarrow{p} 0$ , it is equivalent to show that for any given  $\epsilon > 0$ , for both large T and N,  $P(|\hat{k}_1 - k_1^0| \ge 1) < \epsilon$ . As in BFK (2016), we assume that  $\hat{k}_1 - k_1^0$ ,  $\hat{k}_2 - k_2^0$  and  $\hat{k}_3 - k_3^0$  are bounded here for simplicity. Under Assumptions 1 and that the estimators of break fractions are consistent, we consider the set  $K(C_k) = \{(k_1, k_2, k_3) :$  $1 \le |k_1 - k_1^0|, |k_j - k_j^0| \le C_k, aT \le k_j \le (1 - a)T, j = 1, 2, 3\}$  for a finite constant  $C_k$ and a > 0. By definition,  $S(k_1, k_2, k_3) = \sum_{i=1}^N SSR_i(k_1, k_2, k_3)$  is minimized globally at  $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$ , i.e.,  $S(\hat{k}_1, \hat{k}_2, \hat{k}_3) \le S(k_1^0, \hat{k}_2, \hat{k}_3)$  with probability 1.

Therefore, we examine the behavior of  $S(k_1, k_2, k_3)$  on the set  $K(C_k)$ . It is sufficient to show that for each  $\epsilon > 0$ , for both large T and N,  $P(\min_{K(C_k)}[S(k_1, k_2, k_3) - S(k_1^0, k_2, k_3)] \le$  $0) < \epsilon$ . Without loss of generality, assume  $k_1 < k_1^0 < k_2$ ,

$$S(k_{1}, k_{2}, k_{3}) - S(k_{1}^{0}, k_{2}, k_{3})$$

$$= [S(k_{1}, k_{2}, k_{3}) - S(k_{1}, k_{0}^{0}, k_{2}, k_{3})] - [S(k_{1}^{0}, k_{2}, k_{3}) - S(k_{1}, k_{1}^{0}, k_{2}, k_{3})].$$

$$= \sum_{i=1}^{N} [SSR_{i}(k_{1}, k_{2}, k_{3}) - SSR_{i}(k_{1}, k_{0}^{0}, k_{2}, k_{3})]$$

$$- \sum_{i=1}^{N} [SSR_{i}(k_{1}^{0}, k_{2}, k_{3}) - SSR_{i}(k_{1}, k_{1}^{0}, k_{2}, k_{3})],$$
(25)

where,  $SSR_i(k_1, k_1^0, k_2, k_3)$  is the sum of squared residuals in the regression with four breaks at  $(k_1, k_1^0, k_2, k_3)$  for series *i* and  $S(k_1, k_1^0, k_2, k_3) = \sum_{i=1}^N SSR_i(k_1, k_1^0, k_2, k_3)$ . Thus, the analysis of a three-break or multiple break problem can be decomposed into two problems involving a single break. The first term  $SSR_i(k_1, k_2, k_3) - SSR_i(k_1, k_1^0, k_2, k_3)$  allows an additional fourth break  $k_1^0$  between  $k_1$  and  $k_2$ , and the second term  $SSR_i(k_1^0, k_2, k_3) - SSR_i(k_1, k_1^0, k_2, k_3)$  allows an additional fourth break  $k_1^0$  between  $k_1$  and  $k_2$ , and the second term  $SSR_i(k_1^0, k_2, k_3) - SSR_i(k_1, k_1^0, k_2, k_3)$  adds an additional fourth break at  $k_1$  between 1 and  $k_1^0$ . Thus, it is convenient to derive each part above as a single common break issue in panel data as in BFK (2016).

Following Bai and Perron (1998), we denote  $\hat{\delta}_i(\hat{k}_1, \hat{k}_2, \hat{k}_3) = (\hat{\delta}'_{i1}, \hat{\delta}'_{i2}, \hat{\delta}'_{i3}, \hat{\delta}'_{i4})'$  the estimator of  $(\delta_{i1}, \delta_{i2}, \delta_{i3}, \delta_{i4})$  in the regression with three breaks  $k_1, k_2$  and  $k_3$ , and  $(\hat{\delta}^*_{i1}, \hat{\delta}_{i\Delta}, \hat{\delta}^*_{i2}, \hat{\delta}^*_{i3}, \hat{\delta}^*_{i4})$  the estimator of  $(\delta_{i1}, \delta_{i1}, \delta_{i2}, \delta_{i3}, \delta_{i4})$  based on the partition  $(k_1, k_1^0, k_2, k_3)$ . In particular,  $\hat{\delta}^*_{i1}$  is an estimate of  $\delta_{i1}$  associated with regressor  $(z_{i1}, ..., z_{i,k_1}, 0, ..., 0)', \hat{\delta}_{i\Delta}$  is the estimate of  $\delta_{i1}$  associated with regressor  $Z_{i\Delta} = (0, ..., 0, z_{i,k_1+1}, ..., z_{i,k_1}^0, 0, ..., 0)', \hat{\delta}^*_{i3}, \hat{\delta}^*_{i4}$  can be estimate of  $\delta_{i2}$  associated with regressor  $(0, ..., 0, z_{i,k_1^0+1}, ..., z_{i,k_2}, 0, ..., 0)'$ .  $\hat{\delta}^*_{i3}, \hat{\delta}^*_{i4}$  can be defined similarly.

By definition,

$$SSR_{i}(k_{1}, k_{2}, k_{3}) = \sum_{t=1}^{k_{1}} \left( y_{it} - z'_{it} \hat{\delta}_{i1} \right)^{2} + \sum_{t=k_{1}+1}^{k_{2}} \left( y_{it} - z'_{it} \hat{\delta}_{i2} \right)^{2} + \sum_{t=k_{2}+1}^{k_{3}} \left( y_{it} - z'_{it} \hat{\delta}_{i3} \right)^{2} + \sum_{t=k_{3}+1}^{T} \left( y_{it} - z'_{it} \hat{\delta}_{i4} \right)^{2},$$

and

$$SSR_{i}(k_{1},k_{1}^{0},k_{2},k_{3}) = \sum_{t=1}^{k_{1}} \left( y_{it} - z_{it}'\hat{\delta}_{i1}^{*} \right)^{2} + \sum_{t=k_{1}+1}^{k_{1}^{0}} \left( y_{it} - z_{it}'\hat{\delta}_{i\Delta} \right)^{2} + \sum_{t=k_{1}+1}^{k_{2}} \left( y_{it} - z_{it}'\hat{\delta}_{i2}^{*} \right)^{2} + \sum_{t=k_{2}+1}^{k_{3}} \left( y_{it} - z_{it}'\hat{\delta}_{i3}^{*} \right)^{2} + \sum_{t=k_{3}+1}^{T} \left( y_{it} - z_{it}'\hat{\delta}_{i4}^{*} \right)^{2}.$$

It's worth noting that  $\hat{\delta}_{i1}$  and  $\hat{\delta}_{i1}^*$  are the estimators associated with same regressor  $(z_{i1}, ..., z_{i,k_1}, 0, ..., 0)'$ , thus,  $\hat{\delta}_{i1} = \hat{\delta}_{i1}^*$ . Similarly,  $\hat{\delta}_{i3} = \hat{\delta}_{i3}^*$ ,  $\hat{\delta}_{i4} = \hat{\delta}_{i4}^*$ . Thus,

$$SSR_{i}(k_{1}, k_{2}, k_{3}) - SSR_{i}(k_{1}, k_{1}^{0}, k_{2}, k_{3})$$

$$= \sum_{t=k_{1}+1}^{k_{2}} \left( y_{it} - z'_{it}\hat{\delta}_{i2} \right)^{2} - \left[ \sum_{t=k_{1}+1}^{k_{1}^{0}} \left( y_{it} - z'_{it}\hat{\delta}_{i\Delta} \right)^{2} + \sum_{t=k_{1}^{0}+1}^{k_{2}} \left( y_{it} - z'_{it}\hat{\delta}_{i2}^{*} \right)^{2} \right].$$

$$(26)$$

Since the term  $SSR_i(k_1, k_2, k_3) - SSR_i(k_1, k_1^0, k_2, k_3)$  involves a regression with a break  $k_1^0$  between  $k_1$  and  $k_2$ , we focus on the interval  $[k_1 + 1, k_2]$ .  $k_1^0$  splits  $[k_1 + 1, k_2]$  into

two parts  $[k_1 + 1, k_1^0]$  and  $[k_1^0 + 1, k_2]$ . These three intervals are referred to as  $\bigstar$ ,  $\Delta$ and  $\bigstar - \Delta$ , respectively, i.e.,  $\bigstar = [\Delta, \bigstar - \Delta]$ . Under the current assumptions, the number of observations on interval  $\Delta$  is finite, different from that on  $\bigstar$  or  $\bigstar - \Delta$ . Define  $Y_{i\bigstar} = (y_{i,k_1+1}, ..., y_{i,k_2})'$ ,  $Y_{i\Delta} = (y_{i,k_1+1}, ..., y_{i,k_1^0}, 0, ..., 0)'$  and  $Y_{i(\bigstar -\Delta)} = Y_{i\bigstar} - Y_{i\Delta} =$  $(0, ..., 0, y_{i,k_1^0+1}, ..., y_{i,k_2})'$ .  $Z_{i\bigstar}, \varepsilon_{i\bigstar}^*, Z_{i\Delta}, Z_{i(\bigstar -\Delta)}$  can be defined in the same fashion. By construction,  $Y'_{i\Delta}Y_{i(\bigstar -\Delta)} = 0$  and  $Z'_{i\Delta}Z_{i(\bigstar -\Delta)} = 0$ .

Recall that the OLS estimators of  $(\delta_{i1}, \delta_{i2})$  on intervals of  $[k_1 + 1, k_1^0]$  and  $[k_1^0 + 1, k_2]$ are  $\hat{\delta}_{i\Delta}, \hat{\delta}_{i2}^*$ , respectively. Without considering a break in slopes on the interval  $[k_1 + 1, k_2]$ , the OLS estimator for  $\delta_{i2}$  is  $\hat{\delta}_{i2}$ . The first term in (26),  $\sum_{t=k_1+1}^{k_2} (y_{it} - z'_{it}\hat{\delta}_{i2})^2 = [Y_{i\star} - Z_{i\star}\hat{\delta}_{i2}]'[Y_{i\star} - Z_{i\star}\hat{\delta}_{i2}]$  is the sum of squared residuals in the regression of y on z for series i using time series sample on the interval  $[k_1 + 1, k_2]$ . The second term in equation (26)

$$\sum_{t=k_{1}+1}^{k_{1}^{0}} (y_{it} - z'_{it}\hat{\delta}_{i\Delta})^{2} + \sum_{t=k_{1}^{0}+1}^{k_{2}} (y_{it} - z'_{it}\hat{\delta}_{i2}^{*})^{2}$$

$$= \sum_{t=k_{1}+1}^{k_{1}^{0}} (y_{it} - z'_{it}\hat{\delta}_{i\Delta})^{2} + \sum_{t=k_{1}^{0}+1}^{k_{2}} (y_{it} - z'_{it}\hat{\delta}_{i\Delta} + z'_{it}(\hat{\delta}_{i\Delta} - \hat{\delta}_{i2}^{*}))^{2}$$

$$= [Y_{i\star} - Z_{i\star}\hat{\delta}_{i\Delta} - Z_{i(\star-\Delta)}(\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta})]'[Y_{i\star} - Z_{i\star}\hat{\delta}_{i\Delta} - Z_{i(\star-\Delta)}(\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta})]$$

is the sum of squared residuals in the regression of y on z for series i with a break  $k_1^0$  on the interval  $[k_1 + 1, k_2]$ . Thus, according to Amemiya (1985, p. 31),

$$SSR_{i}(k_{1}, k_{2}, k_{3}) - SSR_{i}(k_{1}, k_{1}^{0}, k_{2}, k_{3}) = (\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta})' Z_{i(\bigstar - \Delta)}' M_{Z_{i\bigstar}} Z_{i(\bigstar - \Delta)}(\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta})$$
$$= (\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta})' Z_{i\Delta}' M_{Z_{i\bigstar}} Z_{i\Delta}(\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta}).$$

The second equality above is due to the facts of  $Z_{i(\bigstar - \Delta)} = Z_{i\bigstar} - Z_{i\Delta}$  and

$$Z'_{i(\bigstar-\Delta)}M_{Z_{i\bigstar}}Z_{i(\bigstar-\Delta)}=Z'_{i\Delta}M_{Z_{i\bigstar}}Z_{i\Delta},$$

where  $M_{Z_{i\star}} = I_{k_2-k_1+1} - Z_{i\star} \left( Z'_{i\star} Z_{i\star} \right) Z'_{i\star}$  and  $I_{(k_2-k_1+1)}$  is the  $(k_2-k_1+1) \times (k_2-k_1+1)$ identity matrix. Next, following BFK (2016) we derive the expression of  $SSR_i(k_1, k_2, k_3) - SSR_i(k_1, k_1^0, k_2, k_3)$ .

For  $t \in [k_1+1, k_1^0]$ ,  $\hat{\delta}_{i\Delta} = (Z'_{i\Delta}Z_{i\Delta})^{-1}Z'_{i\Delta}Y_{i\Delta}$  and  $\hat{\delta}^*_{i2} = (Z'_{i(\bigstar -\Delta)}Z_{i(\bigstar -\Delta)})^{-1}Z'_{i(\bigstar -\Delta)}Y_{i(\bigstar -\Delta)}$ for  $t \in [k_1^0 + 1, k_2]$ . Partitioned regression gives

$$\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta} = (Z_{i(\bigstar - \Delta)}' M_{Z_{i\bigstar}} Z_{i(\bigstar - \Delta)})^{-1} Z_{i(\bigstar - \Delta)}' M_{Z_{i\bigstar}} Y_{i\bigstar}$$
$$= -(Z_{i\Delta}' M_{Z_{i\bigstar}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\bigstar}} Y_{i\bigstar}.$$

Plugging  $Y_{i\bigstar} = Z_{i\bigstar}\delta_{i1} + Z_{i(\bigstar-\Delta)}(\delta_{i2} - \delta_{i1}) + \varepsilon^*_{i\bigstar}$  into the equation above gives,

$$\hat{\delta}_{i2}^{*} - \hat{\delta}_{i\Delta} = (\delta_{i2} - \delta_{i1}) + (Z'_{i(\bigstar - \Delta)} M_{Z_{i\bigstar}} Z_{i(\bigstar - \Delta)})^{-1} Z'_{i(\bigstar - \Delta)} M_{Z_{i\bigstar}} \varepsilon^{*}_{i\bigstar}$$

$$= (\delta_{i2} - \delta_{i1}) - (Z'_{i\Delta} M_{Z_{i\bigstar}} Z_{i\Delta})^{-1} Z'_{i\Delta} M_{Z_{i\bigstar}} \varepsilon^{*}_{i\bigstar}.$$

$$(27)$$

Thus, we can get

$$SSR_{i}(k_{0},k_{1}) - SSR_{i}(k_{0},k_{0}^{0},k_{1}) = (\delta_{i2} - \delta_{i1})' Z_{i\Delta}' M_{Z_{i\star}} Z_{i\Delta} (\delta_{i2} - \delta_{i1})$$

$$-2 (\delta_{i2} - \delta_{i1})' Z_{i\Delta}' M_{Z_{i\star}} \varepsilon_{i\star}^{*}$$

$$+ \varepsilon_{i\star}^{*\prime} M_{Z_{i\star}} Z_{i\Delta} (Z_{i\Delta}' M_{Z_{i\star}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\star}} \varepsilon_{i\star}^{*}.$$

$$(28)$$

Similarly, the second term  $SSR_i(k_1^0, k_2, k_3) - SSR_i(k_1, k_1^0, k_2, k_3)$  in (25) involves a regression with a break at  $k_1$  between 1 and  $k_1^0$ . Denote the interval  $[1, k_1^0]$  by  $\Diamond$ .  $k_1$  splits  $[1, k_1^0]$  into two parts  $[1, k_1]$  and  $[k_1 + 1, k_1^0]$ . Note that the latter interval has been denoted as  $\Delta$  above. Similarly, define  $Y_{i\Diamond} = (y_{i,1}, \dots, y_{i,k_1^0})'$ ,  $Z_{i\Diamond}$  and  $\varepsilon_{i\Diamond}^*$  on the interval  $\Diamond$ . The number of observations on the interval  $\Diamond$  is unbounded under Assumption 1 as  $T \to \infty$ . Note that there is no true break in slopes on the interval  $[1, k_1^0]$  and the corresponding true slope parameter is  $\delta_{i1}$ . The OLS estimators of  $(\delta_{i1}, \delta_{i1})$  on intervals of  $[1, k_1]$  and  $[k_1 + 1, k_1^0]$  are  $\hat{\delta}_{i1}^*$ ,  $\hat{\delta}_{i\Delta}$ , respectively. As in equation (25), we can obtain

$$SSR_{i}(k_{1}^{0}, k_{2}, k_{3}) - SSR_{i}(k_{1}, k_{1}^{0}, k_{2}, k_{3}) = (\hat{\delta}_{i\Delta} - \hat{\delta}_{i1}^{*})' Z_{i\Delta}' M_{Z_{i\Diamond}} Z_{i\Delta} (\hat{\delta}_{i\Delta} - \hat{\delta}_{i1}^{*})$$

Partitioned regression gives  $\hat{\delta}_{i\Delta} - \hat{\delta}_{i1}^* = (Z'_{i\Delta}M_{Z_{i\Diamond}}Z_{i\Delta})^{-1}Z'_{i\Delta}M_{Z_{i\Diamond}}Y_{i\Diamond}$ , where  $M_{Z_{i\Diamond}} = I_{k_1^0} - Z_{i\Diamond}(Z'_{i\Diamond}Z_{i\Diamond})^{-1}Z'_{i\Diamond}$ . Plugging  $Y_{i\Diamond} = Z_{i\Diamond}\delta_{i1} + \varepsilon^*_{i\Diamond}$  into the equation above gives

$$\hat{\delta}_{i\Delta} - \hat{\delta}_{1i}^* = (Z_{i\Delta}' M_{Z_{i\Diamond}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\Diamond}} \varepsilon_{i\Diamond}^*.$$
<sup>(29)</sup>

Since there is no break in slopes on the interval  $[1, k_1^0]$ , no slope shift term appears in (29), which is different from (27). Thus, we can get

$$SSR_{i}(k_{1}^{0}, k_{2}, k_{3}) - SSR_{i}(k_{1}, k_{1}^{0}, k_{2}, k_{3}) = \varepsilon_{i\Diamond}^{*'} M_{Z_{i\Diamond}} Z_{i\Delta} (Z_{i\Delta}' M_{Z_{i\Diamond}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\Diamond}} \varepsilon_{i\Diamond}^{*}.$$
 (30)

Combining equations (28) and (30), we obtain,

$$S(k_{1}, k_{2}, k_{3}) - S(k_{1}^{0}, k_{2}, k_{3})$$

$$= \sum_{i=1}^{N} \left[ S_{i}(k_{1}, k_{2}, k_{3}) - S_{i}(k_{1}, k_{1}^{0}, k_{2}, k_{3}) \right] - \sum_{i=1}^{N} \left[ S_{i}(k_{1}^{0}, k_{2}, k_{3}) - S_{i}(k_{1}, k_{1}^{0}, k_{2}, k_{3}) \right]$$

$$= \sum_{i=1}^{N} \left( \delta_{i2} - \delta_{i1} \right)' Z'_{i\Delta} M_{Z_{i\star}} Z_{i\Delta} \left( \delta_{i2} - \delta_{i1} \right) - 2 \sum_{i=1}^{N} \left( \delta_{i2} - \delta_{i1} \right)' Z'_{i\Delta} M_{Z_{i\star}} \varepsilon_{i\star}^{*}$$

$$+ \sum_{i=1}^{N} \varepsilon_{i\star}^{*'} M_{Z_{i\star}} Z_{i\Delta} \left( Z'_{i\Delta} M_{Z_{i\star}} Z_{i\Delta} \right)^{-1} Z'_{i\Delta} M_{Z_{i\star}} \varepsilon_{i\star}^{*}$$

$$- \sum_{i=1}^{N} \varepsilon_{i\diamond}^{*'} M_{Z_{i\diamond}} Z_{i\Delta} \left( Z'_{i\Delta} M_{Z_{i\diamond}} Z_{i\Delta} \right)^{-1} Z'_{i\Delta} M_{Z_{i\diamond}} \varepsilon_{i\diamond}^{*}.$$

Like in Bai (1997) and BFK (2016), here  $S(k_1, k_2, k_3) - S(k_1^0, k_2, k_3)$  can be expressed as the sum of a deterministic part  $\sum_{i=1}^{N} J_{1i}(k_1, k_2, k_3)$  and a stochastic term  $-\sum_{i=1}^{N} J_{2i}(k_1, k_2, k_3)$ , where  $J_{1i}(k_1, k_2, k_3) = (\delta_{i2} - \delta_{i1})' Z'_{i\Delta} M_{Z_{i\star}} Z_{i\Delta}(\delta_{i2} - \delta_{i1})$ ,  $J_{2i}(k_1, k_2, k_3) = [2(\delta_{i2} - \delta_{i1})' Z'_{i\Delta} M_{Z_{i\star}} \varepsilon^*_{i\star}] - [\varepsilon^{*'}_{i\star} M_{Z_{i\star}} Z_{i\Delta}(Z'_{i\Delta} M_{Z_{i\star}} Z_{i\Delta})^{-1} Z'_{i\Delta} M_{Z_{i\star}} \varepsilon^*_{i\star}]$   $+ [\varepsilon^{*'}_{i\Diamond} M_{Z_{i\Diamond}} Z_{i\Delta}(Z'_{i\Delta} M_{Z_{i\Diamond}} Z_{i\Delta})^{-1} Z'_{i\Delta} M_{Z_{i\Diamond}} \varepsilon^*_{i\Diamond}].$  Thus,  $S(k_1, k_2, k_3) - S(k_1^0, k_2, k_3) = \sum_{i=1}^N J_{1i}(k_1, k_2, k_3) - \sum_{i=1}^N J_{2i}(k_1, k_2, k_3).$ 

To prove Theorem 1 and the statement  $P(\min_{K(C_k)}[S(k_1, k_2, k_3) - S(k_1^0, k_2, k_3)] \leq 0) < \epsilon$  for both large T and N, it suffices to show

$$P(\sup_{K(C_k)} \left| \frac{1}{T} \sum_{i=1}^{N} J_{2i}(k_1, k_2, k_3) \right| \ge \inf_{K(C_k)} \frac{1}{T} \sum_{i=1}^{N} J_{1i}(k_1, k_2, k_3)) < \epsilon.$$
(31)

Consider the term  $\frac{1}{T}Z'_{i\Delta}M_{Z_{i\star}}Z_{i\Delta}$  in  $J_{1i}(k_1,k_2,k_3)$ . Since  $Z_{i\star} = Z_{i\Delta} + Z_{i(\star-\Delta)}$  and  $Z_{i\Delta}/Z_{i(\star-\Delta)} = 0$ ,

$$T^{-1}Z'_{i\Delta}M_{Z_{i\star}}Z_{i\Delta} = T^{-1}Z'_{i\Delta}Z_{i\Delta} - T^{-1}Z'_{i\Delta}Z_{i\star}(Z'_{i\star}Z_{i\star})^{-1}Z'_{i\star}Z_{i\Delta}$$
$$= T^{-1}Z'_{i\Delta}Z_{i\Delta} - T^{-2}Z'_{i\Delta}Z_{i\Delta}(T^{-2}Z'_{i\star}Z_{i\star})^{-1}T^{-1}Z'_{i\Delta}Z_{i\Delta}.$$

Note that the numbers of observations on the intervals of  $\bigstar$  and  $\Delta$  are  $k_2 - k_1$  and  $k_1^0 - k_1$ . On the set  $K(C_k)$ ,  $k_1^0 - k_1$  is finite, while  $k_2 - k_1$  is unbounded as  $T \to \infty$ . By Lemma 1(i),  $\frac{1}{T}Z'_{i\Delta}Z_{i\Delta} = O_p(1)$  and  $\frac{1}{T^2}Z'_{i\Delta}Z_{i\Delta}(\frac{1}{T^2}Z'_{i\bigstar}Z_{i\bigstar})^{-1}\frac{1}{T}Z'_{i\Delta}Z_{i\Delta} = o_p(1)$  on  $K(C_k)$ , thus,  $T^{-1}Z'_{i\Delta}M_{Z_{i\bigstar}}Z_{i\Delta} = T^{-1}Z'_{i\Delta}Z_{i\Delta} + o_p(1)$ . Last,

$$\inf_{K(C_k)} \frac{1}{T} \sum_{i=1}^N J_{1i}(k_1, k_2, k_3) = \inf_{K(C_k)} \sum_{i=1}^N \left(\delta_{i2} - \delta_{i1}\right)' \left(\frac{1}{T} Z'_{i\Delta} Z_{i\Delta}\right) \left(\delta_{i2} - \delta_{i1}\right) + o_p(1).$$

Under Assumption 7, let a finite  $\rho_{min} > 0$  be the minimum eigenvalue of  $\frac{1}{N} \sum_{i=1}^{N} (\frac{1}{T} Z'_{i\Delta} Z_{i\Delta})$ uniformly on  $K(C_k)$ . Following the proof of Lemma 1 in BFK's (2016) appendix, we obtain

$$\inf_{K(C_k)} \frac{1}{T} \sum_{i=1}^{N} J_{1i}(k_1, k_2, k_3) \ge \varrho_{min} \phi_{N,1}$$

with probability tending to 1 and  $\phi_{N,1} = \sum_{i=1}^{N} (\delta_{i2} - \delta_{i1})' (\delta_{i2} - \delta_{i1})$ . Thus, from equation (31), to prove Theorem 1, it is sufficient to show

$$P(\sup_{K(C_k)} \frac{1}{T\phi_{N,1}} |\sum_{i=1}^N J_{2i}(k_1, k_2, k_3)| \ge \rho_{min}) < \epsilon.$$
(32)

By Lemma 2,

$$\begin{aligned} |\sum_{i=1}^{N} J_{2i}(k_{1}, k_{2}, k_{3})| &\leq |\sum_{i=1}^{N} [2(\delta_{i2} - \delta_{i1})' Z_{i\Delta}' M_{Z_{i\star}} \varepsilon_{i\star}^{*}]| \\ &+ |\sum_{i=1}^{N} [\varepsilon_{i\star}^{*\prime} M_{Z_{i\star}} Z_{i\Delta} (Z_{i\Delta}' M_{Z_{i\star}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\star}} \varepsilon_{i\star}^{*}]| \\ &+ |\sum_{i=1}^{N} [\varepsilon_{i\diamond}^{*\prime} M_{Z_{i\diamond}} Z_{i\Delta} (Z_{i\Delta}' M_{Z_{i\diamond}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\diamond}} \varepsilon_{i\diamond}^{*}]| \\ &= O_{p} (T^{1/2} \phi_{N,1}^{1/2}) + O_{p}(N). \end{aligned}$$

Thus,  $\frac{1}{T\phi_{N,1}} \left| \sum_{i=1}^{N} J_{2i}(k_1, k_2, k_3) \right| = O_p(\frac{1}{\sqrt{T\phi_{N,1}}}) + O_p(\frac{N}{T\phi_{N,1}})$ . Under Assumption 11 that  $\frac{N}{T\phi_{N,1}} \to 0$ , as  $(N, T) \to \infty$ , the term  $\frac{1}{T\phi_{N,1}} |J_2(k_1, k_2, k_3)|$  vanishes for any  $(k_1, k_2, k_3) \in K(C_k)$ . Therefore, (32) and then Theorem 1 are established.

The following Lemmas 1 and 2 are needed to prove Theorem 1.

**Lemma 1** Under Assumptions 1-5, 8,9, and uniformly over  $K(C_k)$ , as  $(N,T) \to \infty$ , for i = 1, ..., N,

Lemma 2 Under Assumptions 1-9, uniformly on  $K(C_k)$ , (i)  $\sum_{i=1}^{N} (\delta_{i2} - \delta_{i1})' Z'_{i\Delta} M_{Z_{i\star}} \varepsilon^*_{i\star} = O_p(\sqrt{T\phi_{N,1}});$ (ii)  $\sum_{i=1}^{N} \varepsilon^{**}_{i\star} M_{Z_{i\star}} Z_{i\Delta} (Z'_{i\Delta} M_{Z_{i\star}} Z_{i\Delta})^{-1} Z'_{i\Delta} M_{Z_{i\star}} \varepsilon^*_{i\star} = O_p(N);$ (iii)  $\sum_{i=1}^{N} \varepsilon^{**}_{i\Diamond} M_{Z_{i\Diamond}} Z_{i\Delta} (Z'_{i\Delta} M_{Z_{i\Diamond}} Z_{i\Delta})^{-1} Z'_{i\Delta} M_{Z_{i\Diamond}} \varepsilon^*_{i\Diamond} = O_p(N).$ 

The proofs of Lemmas 1 and 2 can be found in the supplementary Appendix B.

#### A.2 Proof of Theorem 2

The proof of Theorem 2 is similar to that of Theorem 1. To obtain the inequality (31) in Case 2 of I(1)  $v_{it}$ , Lemmas 3 and 4 are needed.

**Lemma 3** Under Assumptions 1-10 and 12, uniformly on  $K(C_k)$  and for each i = 1, ..., N, as  $(N,T) \to \infty$ , (i)  ${}^{1}Z' Z = O(1) {}^{1}Z' Z = O(1)$ ;

$$\begin{aligned} (i) \quad &\frac{1}{T} Z_{i\Delta} Z_{i\Delta} = O_p(1), \quad &\frac{1}{T^2} Z_{i\star} Z_{i\star} = O_p(1); \\ (ii) \quad &\frac{1}{\sqrt{T}} Z_{i\Delta}' \varepsilon_{i\star} = \frac{1}{\sqrt{T}} Z_{i\Delta}' \varepsilon_{i\Delta} = O_p(1), \quad &\frac{1}{T} Z_{i\star}' \varepsilon_{i\star} = O_p(1); \\ (iii) \quad &\frac{1}{\sqrt{T}} Z_{i\Delta}' \varepsilon_{i\diamond} = \frac{1}{\sqrt{T}} Z_{i\Delta}' \varepsilon_{i\Delta} = O_p(1), \quad &\frac{1}{T} Z_{i\diamond}' \varepsilon_{i\diamond} = O_p(1); \\ (iv) \quad &\frac{1}{T^2} \bar{V}_{\star}' \bar{V}_{\star} = O_p\left(\frac{1}{N}\right), \quad &\frac{1}{T} Z_{i\Delta}' \bar{V}_{\star} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad &\frac{1}{T\sqrt{T}} Z_{i\star}' \bar{V}_{\star} = O_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned}$$

Lemma 4 Under Assumptions 1-10 and 12, uniformly on 
$$K(C_k)$$
,  
(i)  $\sum_{i=1}^{N} (\delta_{i2} - \delta_{i1})' Z'_{i\Delta} M_{Z_{i\star}} \varepsilon^*_{i\star} = O_p \left(\sqrt{T\phi_{N,1}}\right) + O_p \left(T\sqrt{\frac{\phi_{N,1}}{N}}\right);$   
(ii)  $\sum_{i=1}^{N} \varepsilon^*_{i\star} M_{Z_{i\star}} Z_{i\Delta} \left(Z'_{i\Delta} M_{Z_{i\star}} Z_{i\Delta}\right)^{-1} Z'_{i\Delta} M_{Z_{i\star}} \varepsilon^*_{i\star} = O_p (N) + O_p (T);$   
(iii)  $\sum_{i=1}^{N} \varepsilon^*_{i\diamond} M_{Z_{i\diamond}} Z_{i\Delta} \left(Z'_{i\Delta} M_{Z_{i\diamond}} Z_{i\Delta}\right)^{-1} Z'_{i\Delta} M_{Z_{i\diamond}} \varepsilon^*_{i\diamond} = O_p (N) + O_p (T).$ 

**Proof of Theorem 2.** As in the proof of Theorem 1, it is suffices to show for any  $\epsilon > 0$ , for large N and T,

$$P(\sup_{K(C_k)} \left| \frac{1}{T} \sum_{i=1}^N J_{2i}(k_1, k_2, k_3) \right| \ge \inf_{K(C_k)} \frac{1}{T} \sum_{i=1}^N J_{1i}(k_1, k_2, k_3)) < \epsilon.$$

In Case 2, the only difference lies in that  $v_{it}$  changes from I(0) to I(1). Since  $x_{it} = \Gamma'_i f_t + v_{it}$ and  $\bar{x}_t = \bar{\Gamma}' f_t + \bar{v}_t$ ,  $z_{it} = (x'_{it}, \bar{x}'_{\cdot t})'$  remains I(1) for I(1) f<sub>t</sub>. Thus, with Lemma 3, the following result remains unchanged,  $\inf_{K(C_k)} \frac{1}{T} \sum_{i=1}^N J_{1i}(k_1, k_2, k_3) \geq \rho_{\min} \phi_{N,1}$  with probability tending to 1. As in the proof of Theorem 1, we need to show

$$P(\sup_{K(C_k)} \frac{1}{T\phi_{N,1}} |\sum_{i=1}^N J_{2i}(k_1, k_2, k_3)| \ge \varrho_{min}) < \epsilon.$$
(33)

By Lemma 4,

$$\begin{aligned} |\sum_{i=1}^{N} J_{2i}(k_{1}, k_{2}, k_{3})| &\leq |\sum_{i=1}^{N} [(\delta_{i2} - \delta_{i1})' Z_{i\Delta}' M_{Z_{i\star}} \varepsilon_{i\star}^{*}]| \\ &+ |\sum_{i=1}^{N} [\varepsilon_{i\star}^{*'} M_{Z_{i\star}} Z_{i\Delta} (Z_{i\Delta}' M_{Z_{i\star}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\star}} \varepsilon_{i\star}^{*}]| \\ &+ |\sum_{i=1}^{N} [\varepsilon_{i\Diamond}^{*'} M_{Z_{i\Diamond}} Z_{i\Delta} (Z_{i\Delta}' M_{Z_{i\Diamond}} Z_{i\Delta})^{-1} Z_{i\Delta}' M_{Z_{i\Diamond}} \varepsilon_{i\Diamond}^{*}]| \\ &= O_{p} (\sqrt{T\phi_{N,1}}) + O_{p} (T\phi_{N,1}^{1/2} N^{-1}) + O_{p} (N) + O_{p} (T) .\end{aligned}$$

Thus,

$$\frac{1}{T\phi_{N,1}} \left| J_2(k_1, k_2, k_3) \right| = O_p(T^{-1/2}\phi_{N,1}^{-1/2}) + O_p(N^{-1/2}\phi_{N,1}^{-1/2}) + O_p(NT^{-1}\phi_{N,1}^{-1}) + O_p(\frac{1}{\phi_{N,1}}).$$

Under Assumption 11,  $\phi_{N,1} \to \infty$  and  $\frac{N}{T\phi_{N,1}} \to 0$ , as as  $(N,T) \to \infty$ ,  $\frac{1}{T\phi_{N,1}} |J_2(k_1, k_2, k_3)|$  vanishes for any  $(k_1, k_2, k_3) \in K(C_k)$ . Therefore, (33) is established, and Theorem 2 is proved.

#### A.3 Proofs of Propositions 1 and 2

In this subsection, we also assume m = 3, including two breaks  $k_1^0, k_2^0$  in slopes and a third one  $k_3^0$  in error factor loadings. In order to prove Propositions 1 and 2, we first give the following lemma.

**Lemma 5** Under Assumptions 1-5, 8, 9, and uniformly over  $K(C_k)$  and for each i = 1, ..., N, as  $(N, T) \to \infty$ ,

$$(i) \ \frac{1}{T} \bar{V}' M_{\bar{X}} \bar{V} = O_p(N^{-1}), \ \frac{1}{T} V_i' M_{\bar{X}} V_i = O_p(1)$$
  

$$(ii) \frac{1}{T} F' M_{\bar{X}} F = O_p(N^{-1}), \ \frac{1}{T} V_i' M_{\bar{X}} F = O_p(N^{-1/2});$$
  

$$(iii) \ \frac{1}{T} X_i' M_{\bar{X}} X_i = O_p(1), \ \frac{1}{T} \underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \underline{X}_i(\hat{k}_1, \hat{k}_2) = O_p(1)$$

**Proof of Proposition** 1. For individual series i = 1, ..., N, equation (14) can be written as

$$Y_{i} = \underline{X}_{i}(k_{1}^{0}, k_{2}^{0})b_{i} + \bar{X}\gamma_{i}^{*}(k_{3}^{0}) + \varepsilon_{i}^{*}$$

$$= \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})b_{i} + [\underline{X}_{i}(k_{1}^{0}, k_{2}^{0}) - \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})]b_{i} + \bar{X}\gamma_{i}^{*}(k_{3}^{0}) + \varepsilon_{i} - \bar{V}\bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1}\gamma_{i}(k_{3}^{0}).$$
(34)

Plugging equation (34) above into the expression of  $\hat{b}_i$  gives,

$$\hat{b}_{i} = \hat{b}_{i}(\hat{k}_{1}, \hat{k}_{2}) = [\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})]^{-1} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} Y_{i}$$

$$= b_{i} + [\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})]^{-1} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} [\underline{X}_{i}(k_{1}^{0}, k_{2}^{0}) - \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})] b_{i}$$

$$+ [\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})]^{-1} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \varepsilon_{i}$$

$$- [\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})]^{-1} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \overline{V} \overline{\Gamma}'(\overline{\Gamma} \overline{\Gamma}')^{-1} \gamma_{i}(k_{3}^{0}).$$
(35)

Thus, we decompose  $\sqrt{T}(\hat{b}_i - b_i)$  into five terms,

$$\begin{aligned}
\sqrt{T}(\hat{b}_{i}-b_{i}) &= \left[\frac{1}{T}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]^{-1}\frac{1}{\sqrt{T}}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})'M_{\bar{X}}\varepsilon_{i} \\
&- \left[\frac{1}{T}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]^{-1}\frac{1}{\sqrt{T}}\left[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]M_{\bar{X}}\varepsilon_{i} \\
&- \left[\frac{1}{T}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]^{-1}\frac{1}{\sqrt{T}}\underline{X}_{i}\left(k_{1}^{0},k_{2}^{0}\right)'M_{\bar{X}}\bar{V}\bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1}\gamma_{i}(k_{3}^{0}) \\
&+ \left[\frac{1}{T}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]^{-1}\frac{1}{\sqrt{T}}\left[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]M_{\bar{X}}\bar{V}\bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1}\gamma_{i}(k_{3}^{0}) \\
&+ \left[\frac{1}{T}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]^{-1}\frac{1}{\sqrt{T}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\left[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})\right]b_{i}.
\end{aligned}$$
(36)

As in KPY, in the model considered in Case 1, after the transformation using  $M_{\bar{X}}$ ,  $M_{\bar{X}}\underline{X}_i(\hat{k}_1, \hat{k}_2)$  becomes stationary since I(1)  $f_t$  is removed asymptotically in regressors  $x_{it}$ . Thus,  $\frac{1}{T}\underline{X}_i(\hat{k}_1, \hat{k}_2)'M_{\bar{X}}\underline{X}_i(\hat{k}_1, \hat{k}_2) = O_p(1)$  for large N and T. For the second term, by Theorem 1,  $P\left(\hat{k}_1 \neq k_1^0, \hat{k}_2 \neq k_2^0\right) = P\left(|\hat{k}_1 \neq k_1^0| \geq 1, |\hat{k}_2 \neq k_2^0| \geq 1\right) \to 0$ . For any  $\eta > 0$ ,

$$\begin{split} &P(||T^{-1/2}[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]M_{\bar{X}}\varepsilon_{i}|| > \eta) \\ &= P(||T^{-1/2}[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]M_{\bar{X}}\varepsilon_{i}|| > \eta, \hat{k}_{1} = k_{1}^{0}, \hat{k}_{2} = k_{2}^{0}) \\ &+ P(||T^{-1/2}[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]M_{\bar{X}}\varepsilon_{i}|| > \eta, \hat{k}_{1} \neq k_{1}^{0}, \hat{k}_{2} \neq k_{2}^{0}) \\ &= P(0 > \eta)P(\hat{k}_{1} = k_{1}^{0}, \hat{k}_{2} = k_{2}^{0}) \\ &+ P(||T^{-1/2}[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]M_{\bar{X}}\varepsilon_{i}|| > \eta \left| \hat{k}_{1} \neq k_{1}^{0}, \hat{k}_{2} \neq k_{2}^{0} \right)P(\hat{k}_{1} \neq k_{1}^{0}, \hat{k}_{2} \neq k_{2}^{0}) \\ &\leq P(0 > \eta)(\hat{k}_{1} = k_{1}^{0}, \hat{k}_{2} = k_{2}^{0}) + P(\hat{k}_{1} \neq k_{1}^{0}, \hat{k}_{2} \neq k_{2}^{0}) \to 0. \end{split}$$

$$\begin{split} \text{Thus,} \ T^{-1/2}[\underline{X}_i(k_1^0,k_2^0) - \underline{X}_i(\hat{k}_1,\hat{k}_2)] M_{\bar{X}} \varepsilon_i &= o_p(1). \text{ Similarly, } T^{-1/2} \underline{X}_i(\hat{k}_1,\hat{k}_2)' M_{\bar{X}}[\underline{X}_i(k_1^0,k_2^0) - \underline{X}_i(\hat{k}_1,\hat{k}_2)] b_i &= o_p(1) \text{ and } T^{-1/2}[\underline{X}_i(k_1^0,k_2^0) - \underline{X}_i(\hat{k}_1,\hat{k}_2)] M_{\bar{X}} \bar{V} \bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1} \gamma_i(k_3^0) &= o_p(1). \end{split}$$

According to Lemma 5(ii),  $\frac{1}{T} \underline{X}_i(k_1^0, k_2^0)' M_{\bar{X}} \bar{V} = O_p(N^{-1}), T^{-1/2} \underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \bar{V} = O_p(T^{1/2}N^{-1})$ , thus, the third term

$$[T^{-1}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}T^{-1/2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\bar{V}\bar{\Gamma}'(\bar{\Gamma}\bar{\Gamma}')^{-1}\gamma_{i}(k_{3}^{0}) = O_{p}(T^{1/2}N^{-1}).$$

Combining all these terms, we obtain

$$\sqrt{T}(\hat{b}_i - b_i) = \left[\frac{1}{T}\underline{X}_i(\hat{k}_1, \hat{k}_2)'M_{\bar{X}}\underline{X}_i(\hat{k}_1, \hat{k}_2)\right]^{-1}T^{-1/2}\underline{X}_i(\hat{k}_1, \hat{k}_2)'M_{\bar{X}}\varepsilon_i + O_p\left(T^{1/2}N^{-1}\right) + o_p(1).$$

Now we consider the asymptotic distribution of

$$[T^{-1}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}T^{-1/2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\varepsilon_{i}.$$

Under Theorem 1,  $\hat{k}_1 - k_1^0 = o_p(1)$  and  $\hat{k}_2 - k_2^0 = o_p(1)$ , for each  $i, \underline{X}_i(\hat{k}_1, \hat{k}_2) \xrightarrow{p} \underline{X}_i(k_1^0, k_2^0)$ . Thus, under Assumption 6 (ii),

$$\frac{1}{T}\underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}}\underline{X}_i(\hat{k}_1, \hat{k}_2) - \frac{1}{T}\underline{X}_i(k_1^0, k_2^0)' M_{\bar{X}}\underline{X}_i(k_1^0, k_2^0) \xrightarrow{p} 0.$$

Let  $\Sigma_{X,i} = plim_{T \to \infty} \frac{1}{T} \underline{X}_i(k_0^0)' M_{\bar{X}} \underline{X}_i(k_0^0)$  and  $\Sigma_{X\varepsilon,i} = plim_{T \to \infty} \frac{1}{T} \underline{X}_i(k_1^0, k_2^0)' M_{\bar{X}} \Sigma_{\varepsilon,i} M_{\bar{X}} \underline{X}_i(k_1^0, k_2^0)$ , as  $T^{1/2} N^{-1} \to 0$ , we obtain  $\sqrt{T}(\hat{b}_i - b_i) \xrightarrow{d} N(0, \Sigma_{X,i}^{-1} \Sigma_{X\varepsilon,i} \Sigma_{X,i}^{-1})$ .

**Proof of Proposition** 2. Under Assumption 4, the asymptotic distribution of meangroup estimator can be derived similarly. Thus, we obtain

$$\begin{split} \sqrt{N} \left( \hat{b}_{MG} - b \right) &= N^{-1/2} \sum_{i=1}^{N} v_{b,i} \\ &+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \frac{1}{T} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}) \right]^{-1} \frac{1}{T} \underline{X}_{i}(k_{1}^{0}, k_{2}^{0})' M_{\bar{X}} \varepsilon_{i} \\ &+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \frac{1}{T} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}) \right]^{-1} \frac{1}{T} \left[ \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}) - \underline{X}_{i}(k_{1}^{0}, k_{2}^{0}) \right]' M_{\bar{X}} \varepsilon_{i} \\ &+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \frac{1}{T} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}) \right]^{-1} \frac{1}{T} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \left[ \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}) - \underline{X}_{i}(k_{1}^{0}, k_{2}^{0}) \right] b_{i} \\ &+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \frac{1}{T} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}) \right]^{-1} \frac{1}{T} \underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})' M_{\bar{X}} \overline{V} \overline{\Gamma}' (\overline{\Gamma} \overline{\Gamma}')^{-1} \gamma_{i}(k_{3}^{0}). \end{split}$$

By Assumption 4, the limiting distribution of the first term is  $N(0, \Sigma_b)$ . For the second term,

$$\begin{aligned} &Var(N^{-1/2}\sum_{i=1}^{N}[T^{-1}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}T^{-1}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})'M_{\bar{X}}\varepsilon_{i}) \\ &= \frac{1}{NT}\sum_{i=1}^{N}(T^{-1}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2}))^{-1}(T^{-1}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})'M_{\bar{X}}Var(\varepsilon_{i})M_{\bar{X}}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})) \\ &\times (T^{-1}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2}))^{-1} = O_{p}(T^{-1}). \end{aligned}$$

Thus,  $N^{-1/2} \sum_{i=1}^{N} [T^{-1} \underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \underline{X}_i(\hat{k}_1, \hat{k}_2)]^{-1} T^{-1} \underline{X}_i(k_1^0, k_2^0)' M_{\bar{X}} \varepsilon_i = O_p(T^{-1/2}).$  Similarly, the last term is  $O_p(N^{-1/2}T^{-1}).$ 

As in the proof of Proposition 1, the second and third terms are also  $o_p(1)$ . Therefore, as  $(N,T) \to \infty$ ,

$$\sqrt{N}(\hat{b}_{MG} - b) = N^{-1/2} \sum_{i=1}^{N} v_{b,i} + o_p(1) \xrightarrow{d} N(0, \Sigma_b).$$

#### A.4 Proofs of Propositions 3, 4

**Proof of Proposition 3**. We will show that the convergence rate of  $\hat{b}_i$  is T. From equation (34),  $T(\hat{b}_i - b_i)$  can be decomposed into five terms,

$$\begin{split} T(\hat{b}_{i}-b_{i}) &= [T^{-2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}\frac{1}{T}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})'M_{\bar{X}}\varepsilon_{i} \\ &- [T^{-2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}\frac{1}{T}[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]M_{\bar{X}}\varepsilon_{i} \\ &- [T^{-2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}\frac{1}{T}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\overline{V}\overline{\Gamma}'(\overline{\Gamma}\overline{\Gamma}')^{-1}\gamma_{i}(k_{3}^{0}) \\ &+ [T^{-2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}\frac{1}{T}[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]'M_{\bar{X}}\overline{V}\overline{\Gamma}'(\overline{\Gamma}\overline{\Gamma}')^{-1}\gamma_{i}(k_{3}^{0}) \\ &+ [T^{-2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]^{-1}\frac{1}{T}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}[\underline{X}_{i}(k_{1}^{0},k_{2}^{0})-\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})]b_{i}. \end{split}$$

Under Theorem 2,  $\hat{k}_1 - k_1^0 = o_p(1)$  and  $\hat{k}_2 - k_2^0 = o_p(1)$ , for each i,  $\underline{X}_i(\hat{k}_1, \hat{k}_2) - \underline{X}_i(k_1^0, k_2^0) \xrightarrow{p} 0$ . Thus, similar to equation (37) in the proof of Proposition 1, except the first term, the other four terms above are  $o_p(1)$ , i.e.,

$$T(\hat{b}_i - b_i) = [T^{-2}\underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \underline{X}_i(\hat{k}_1, \hat{k}_2)]^{-1} T^{-1} \underline{X}_i(k_1^0, k_2^0)' M_{\bar{X}} \varepsilon_i + o_p(1).$$

Thus, to prove Proposition 3, we need to show that the first term above converges weakly to a non-degenerate distribution. Given that

$$T^{-2}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1},\hat{k}_{2}) - T^{-2}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})'M_{\bar{X}}\underline{X}_{i}(k_{1}^{0},k_{2}^{0}) \xrightarrow{p} 0,$$

it is equivalent to show that  $\left[\frac{1}{T^2}\underline{X}_i(k_1^0,k_2^0)'M_{\bar{X}}\underline{X}_i(k_1^0,k_2^0)\right]^{-1}\frac{1}{T}\underline{X}_i(k_1^0,k_2^0)'M_{\bar{X}}\varepsilon_i$  converges weakly to a non-degenerate distribution.

Following Phillips and Moon (1999), we will show that as  $T \to \infty$ ,

$$\frac{1}{T^2} \underline{X}_i(k_1^0, k_2^0)' M_{\bar{X}} \underline{X}_i(k_1^0, k_2^0) \Rightarrow G_i, 
\frac{1}{T} \underline{X}_i(k_1^0, k_2^0)' M_{\bar{X}} \varepsilon_i \Rightarrow H_i,$$

where  $G_i$  and  $H_i$  are two non-degenerate distributions, respectively, which will be specified below. Therefore, as  $T \to \infty$ ,  $T(\hat{b}_i - b_i) \Rightarrow G_i^{-1}H_i$ .

 $\begin{array}{l} \text{Consider the term } \frac{1}{T^2}\underline{X}_i(k_1^0,k_2^0)'M_{\bar{X}}\underline{X}_i(k_1^0,k_2^0) \text{ first. Denote } \underline{X}_i(k_1^0,k_2^0) = diag\left(X_{i1},X_{i2},X_{i3}\right) \\ \text{with } \underset{(k_1^0\times p)}{X_{i1}(k_1^0)} = (x_{i1},...,x_{i,k_1^0})', \underset{(k_2^0-k_1^0)\times p}{X_{i2}(k_1^0,k_2^0)} = (x_{i,k_1^0+1},...,x_{ik_2^0})', \underset{(T-k_2^0)\times p}{X_{i3}(k_2^0)} = (x_{i,k_2^0+1},...,x_{iT})'. \\ F_1 = (f_1,...,f_{k_1^0})', F_2 = (f_{k_1^0+1},...,f_{k_2^0})', \text{ and } F_3 = (f_{k_2^0+1},...,f_T)', \text{ and } V_{i1}, V_{i2}, V_{i3}, \varepsilon_{1i}, \varepsilon_{2i}, \\ \varepsilon_{3i} \text{ are similarly defined. Thus, } \underbrace{X_i(k_1^0,k_2^0) = diag(F_1\Gamma_i + V_{i1},F_2\Gamma_i + V_{i2},F_3\Gamma_i + V_{i3}). \end{array}$ 

When the rank condition is satisfied and  $\bar{X} = F\bar{\Gamma} + \bar{V}$ ,  $M_{\bar{X}}\underline{X}_i(k_1^0, k_2^0) = M_F\underline{X}_i(k_1^0, k_2^0) + o_p(1)$ , as  $(N, T) \to \infty$ . Thus,

$$T^{-2}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})'M_{F}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})$$

$$= T^{-2}diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})'\times diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})$$

$$- [T^{-2}diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})'F](T^{-2}F'F)^{-1}$$

$$\times [T^{-2}F'diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})].$$
(37)

According to Phillips and Moon (1999, P.1062), under Assumption 12, for any  $0 \leq \tau_1 \leq \tau_2 \leq 1,$ 

$$T^{-2} \sum_{[\tau_1 T]}^{[\tau_2 T]} v_{it} v_{it}' \Rightarrow \Psi_i(1) P_i(\int_{\tau_1}^{\tau_2} W_{\varsigma,i} W_{\varsigma,i}') P_i' \Psi_i(1)' = \int_{\tau_1}^{\tau_2} B_{\varsigma,i} B_{\varsigma,i}',$$
(38)

where  $B_{\varsigma,i}$  is a Brownian motion with covariance  $\Psi_i(1)P_iP'_i\Psi_i(1)'$ . Similarly, under Assumptions 5, 12 and 13,

$$T^{-2} \sum_{[\tau_1 T]}^{[\tau_2 T]} f_t f'_t \Rightarrow \Pi(1) Q(\int_{\tau_1}^{\tau_2} W_{\varphi} W'_{\varphi}) Q' \Pi(1)' = \int_{\tau_1}^{\tau_2} B_{\varphi} B'_{\varphi}, \tag{39}$$

$$T^{-2} \sum_{[\tau_1 T]}^{[\tau_2 T]} f_t v'_{it} \Rightarrow \Pi(1) Q(\int_{\tau_1}^{\tau_2} W_{\varphi} W'_{\varsigma,i}) P'_i \Psi_i(1)' = \int_{\tau_1}^{\tau_2} B_{\varphi} B'_{\varsigma,i}.$$
 (40)

In addition, under Assumptions 5, 8, and Lemma 8 of Phillips and Moon (1999),

$$T^{-1}\sum_{[cT]}^{[dT]} f_t \varepsilon_{it} \Rightarrow \Pi(1)Q(\int_c^d W_{\varphi}d(W_{\varepsilon,i}))\sigma_i + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E(\varphi_t \varepsilon_{i,t+s})$$
$$= \int_c^d B_{\varphi}d(B_{\varepsilon,i}) + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E(\varphi_t \varepsilon_{i,t+s}).$$
(41)

Moreover, under Assumptions 5, 12 and 13,

$$T^{-1}\sum_{[cT]}^{[dT]} v_{it}\varepsilon_{it} \Rightarrow \Psi_i(1)P_i(\int_c^d W_{\varsigma,i}d(W_{\varepsilon,i}))\sigma_i + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E(\varsigma_{it}\varepsilon_{i,t+s})$$
$$= \int_c^d B_{\varsigma,i}d(B_{\varepsilon,i}) + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E(\varsigma_{it}\varepsilon_{i,t+s}).$$
(42)

Consider the first term in equation (37) above,

 $T^{-2}diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})' \times diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})$ = $T^{-2}diag((F_{1}\Gamma_{i}+V_{i1})'(F_{1}\Gamma_{i}+V_{i1}),(F_{2}\Gamma_{i}+V_{i2})'(F_{2}\Gamma_{i}+V_{i2}),(F_{3}\Gamma_{i}+V_{i3})'(F_{3}\Gamma_{i}+V_{i3})).$  According to equations (38)-(40),

$$T^{-2}(F_j\Gamma_i+V_{j1})'(F_j\Gamma_i+V_{ij}) \Rightarrow \Gamma'_i \int_{\lambda_{j-1}^0}^{\lambda_j^0} B_{\varphi}B'_{\varphi}\Gamma_i + (\int_{\lambda_{j-1}^0}^{\lambda_j^0} B_{\varsigma,i}B'_{\varphi})\Gamma_i + \Gamma'_i (\int_{\lambda_{j-1}^0}^{\lambda_j^0} B_{\varphi}B'_{\varsigma,i}) + \int_{\lambda_{j-1}^0}^{\lambda_j^0} B_{\varsigma,i}B'_{\varphi,i}$$

for  $j = \{1, 2, 3\}$  with  $\lambda_0^0 = 0$  and  $\lambda_3^0 = 1$ . Thus,

$$T^{-2}diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})' \cdot diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})$$

$$\Rightarrow diag(\Gamma_{i}'\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i}+(\int_{0}^{\lambda_{1}^{0}}B_{\varsigma,i}B_{\varphi}')\Gamma_{i}+\Gamma_{i}'(\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varsigma,i}')+\int_{0}^{\lambda_{1}^{0}}B_{\varsigma,i}B_{\varsigma,i}',$$

$$\Gamma_{i}'\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i}+(\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varsigma,i}B_{\varphi}')\Gamma_{i}+\Gamma_{i}'(\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}')+\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varsigma,i}B_{\varsigma,i}',$$

$$\Gamma_{i}'\int_{\lambda_{2}^{0}}^{1}B_{\varphi}B_{\varphi}'\Gamma_{i}+(\int_{\lambda_{2}^{0}}^{1}B_{\varsigma,i}B_{\varphi}')\Gamma_{i}+\Gamma_{i}'(\int_{\lambda_{2}^{0}}^{1}B_{\varphi}B_{\varsigma,i}')+\int_{\lambda_{2}^{0}}^{1}B_{\varsigma,i}B_{\varsigma,i}').$$

Similarly, according to equations (39) and (40), the second term in equation (37)

$$T^{-2}diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})'F$$

$$=T^{-2}(F_{1}'F_{1}\Gamma_{i}+F_{1}'V_{i1},F_{2}'F_{2}\Gamma_{i}+F_{2}'V_{i2},F_{3}'F_{3}\Gamma_{i}+F_{3}'V_{i3})'$$

$$\Rightarrow(\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i}+\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varsigma,i}',\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i}+\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}',\int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i}+\int_{\lambda_{2}^{0}}^{1}B_{\varphi}B_{\varphi}'\Gamma_{i}+\int_{\lambda_{2}^{0}}^{1}B_{\varphi}B_{\varsigma,i}')',$$

and  $\frac{1}{T^2}F'F \Rightarrow \int_0^1 B_{\varphi}B'_{\varphi}$ . Thus, we obtain

$$T^{-2}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})'M_{F}\underline{X}_{i}(k_{1}^{0},k_{2}^{0})$$
(43)  

$$\Rightarrow diag(\Gamma_{i}'\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + (\int_{0}^{\lambda_{1}^{0}}B_{\varsigma,i}B_{\varphi}')\Gamma_{i} + \Gamma_{i}'(\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varsigma,i}') + \int_{0}^{\lambda_{1}^{0}}B_{\varsigma,i}B_{\varsigma,i}',$$
$$\Gamma_{i}'\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + (\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varsigma,i}B_{\varphi}')\Gamma_{i} + \Gamma_{i}'(\int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}') + \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varsigma,i}B_{\varsigma,i}',$$
$$\Gamma_{i}'\int_{\lambda_{2}^{0}}^{\lambda_{1}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + (\int_{\lambda_{2}^{0}}^{\lambda_{1}^{0}}B_{\varphi}B_{\varphi}')\Gamma_{i} + \Gamma_{i}'(\int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}') + \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varsigma,i}B_{\varsigma,i}',$$
$$(\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + \int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varsigma,i}', \int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + \int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}', \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}')'(\int_{0}^{1}B_{\varphi}B_{\varphi}')^{-1}$$
$$\times (\int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + \int_{0}^{\lambda_{1}^{0}}B_{\varphi}B_{\varsigma,i}', \int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + \int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}', \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}'\Gamma_{i} + \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}') \equiv G_{i}.$$

Likewise,

$$T^{-1}\underline{X}_{i}\left(k_{1}^{0},k_{2}^{0}\right)'M_{\bar{X}}\varepsilon_{i}$$

$$= T^{-1}diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})'\varepsilon_{i}$$

$$- T^{-1}diag(F_{1}\Gamma_{i}+V_{i1},F_{2}\Gamma_{i}+V_{i2},F_{3}\Gamma_{i}+V_{i3})'F(F'F)^{-1}F\varepsilon_{i}$$

$$= \begin{bmatrix} T^{-1}\Gamma_{i}'F_{1}'\varepsilon_{1i}+T^{-1}V_{i1}'\varepsilon_{1i}\\ T^{-1}\Gamma_{i}'F_{2}'\varepsilon_{2i}+T^{-1}V_{i2}'\varepsilon_{2i}\\ T^{-1}\Gamma_{i}'F_{3}'\varepsilon_{3i}+T^{-1}V_{i3}'\varepsilon_{3i} \end{bmatrix} - \begin{bmatrix} T^{-2}\Gamma_{i}'F_{1}'F_{1}+T^{-2}V_{i1}'F_{1}\\ T^{-2}\Gamma_{i}'F_{2}'F_{2}+T^{-2}V_{i2}'F_{2}\\ T^{-2}\Gamma_{i}'F_{2}'F_{2}+T^{-2}V_{3i}'F_{3} \end{bmatrix} \left(T^{-2}F'F\right)^{-1}\left(\frac{1}{T}F'\varepsilon_{i}\right).$$

According to equations (39), (41), and (42),

$$T^{-1}\underline{X}_{i}\left(k_{1}^{0},k_{2}^{0}\right)'M_{\bar{X}}\varepsilon_{i} \tag{44}$$

$$\Rightarrow \begin{bmatrix} \Gamma_{i}'\int_{0}^{\lambda_{1}^{0}}B_{\varphi}d(B_{\varepsilon,i}) + \Gamma_{i}'\sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E\left(\varphi_{t}\varepsilon_{i,t+s}\right) + \int_{0}^{\lambda_{1}^{0}}B_{\varsigma,i}d(B_{\varepsilon,i}) + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E\left(\varsigma_{it},\varepsilon_{i,t+s}\right) \\ \Gamma_{i}'\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}d(B_{\varepsilon,i}) + \Gamma_{i}'\sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E\left(\varphi_{t}\varepsilon_{i,t+s}\right) + \int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varsigma,i}d(B_{\varepsilon,i}) + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E\left(\varsigma_{it},\varepsilon_{i,t+s}\right) \\ \Gamma_{i}'\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}d(B_{\varepsilon,i}) + \Gamma_{i}'\sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E\left(\varphi_{t}\varepsilon_{i,t+s}\right) + \int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varsigma,i}d(B_{\varepsilon,i}) + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E\left(\varsigma_{it},\varepsilon_{i,t+s}\right) \end{bmatrix} \\ - \begin{bmatrix} \Gamma_{i}'\int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}' + \int_{\lambda_{1}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}' \\ \Gamma_{i}'\int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}' + \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}' \\ \Gamma_{i}'\int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varphi}' + \int_{\lambda_{2}^{0}}^{\lambda_{2}^{0}}B_{\varphi}B_{\varsigma,i}' \\ \end{bmatrix} \begin{pmatrix} \int_{0}^{1}B_{\varphi}B_{\varphi}' \end{pmatrix}^{-1} \begin{bmatrix} \int_{0}^{1}B_{\varphi}d(B_{\varepsilon,i}) + \sum_{t=0}^{\infty}\sum_{s=0}^{\infty}E\left(\varphi_{t}\varepsilon_{i,t+s}\right) \end{bmatrix} = H_{i}$$

**Proof of Proposition 4**. By the same argument in the proof of Proposition 2, we can obtain equation (20),

$$\sqrt{N}\left(\hat{b}_{MG}-b\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{b,i} + \frac{1}{\sqrt{N}T} \sum_{i=1}^{N} \left[ \left(\frac{1}{T^2} \underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \underline{X}_i(\hat{k}_1, \hat{k}_2) \right)^{-1} \frac{1}{T} \underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \varepsilon_i \right] + o_p(1) .$$

In a special case of homogeneous slopes  $b_i = b$  with  $v_{b,i} = 0$ , we have,

$$\sqrt{N}T\left(\hat{b}_{MG}-b\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \left(\frac{1}{T^2} \underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \underline{X}_i(\hat{k}_1, \hat{k}_2)\right)^{-1} \frac{1}{T} \underline{X}_i(\hat{k}_1, \hat{k}_2)' M_{\bar{X}} \varepsilon_i \right] + o_p(1) \, .$$

As in the proof of Proposition 3 above,  $\left(\frac{1}{T^2}\underline{X}_i(\hat{k}_1,\hat{k}_2)'M_{\bar{X}}\underline{X}_i(\hat{k}_1,\hat{k}_2)\right)^{-1}\frac{1}{T}\underline{X}_i(\hat{k}_1,\hat{k}_2)'M_{\bar{X}}\varepsilon_i$ weakly converges a non-degenerate distribution  $G_i^{-1}H_i$ .

Under the assumptions that  $\varepsilon_{it}, \varphi_s, \varsigma_{jt'}$  are independent for all (i, j) and (t, s, t'), and  $E(\varepsilon_{it}) = 0$ ,  $E\left[\frac{1}{T}\underline{X}_i(\hat{k}_1, \hat{k}_2)'M_{\bar{X}}\varepsilon_i\right] = 0$ . Thus,  $\sqrt{NT}\left(\hat{b}_{MG} - b\right)$  is consistent, as  $(N,T) \to \infty$ . In addition, under the assumption cross-sectional independence of  $\varepsilon_{it}, G_i^{-1}H_i$ are independent across *i*. Thus, by the Central Limit Theory, the limiting distribution of  $\sqrt{NT}\left(\hat{b}_{MG} - b\right)$  is multivariate normal, i.e., as  $(N,T) \to \infty, \sqrt{NT}\left(\hat{b}_{MG} - b\right) \xrightarrow{d} N(0, \Sigma_{MG})$ . Next, we derive the expression of  $\Sigma_{MG}$ . For simplicity, asymptotic bias mentioned in Theorem 8 of Phillips and Moon (1999) and Proposition 1 of Bai, Ng and Kao (2009) disappears here under the assumptions of no serial/ cross-sectional correlation and heteroskedasticity.

Let  $w_{it} = (\varepsilon_{it}, \varphi'_t, \varsigma'_{it})'$ . Denote the long-run covariance matrix of  $w_{it}$ , partitioned conformably for  $w_{it}$ , by

$$\Omega_{i} = \sum_{j=-\infty}^{\infty} E(w_{i0}w_{ij}') = \begin{bmatrix} \Omega_{\varepsilon.i} & \Omega_{\varepsilon\varphi i} & \Omega_{\varepsilon\varsigma i} \\ \Omega_{\varphi\varepsilon i} & \Omega_{\varphi} & \Omega_{\varphi\varsigma i} \\ \Omega_{\varsigma\varepsilon i} & \Omega_{\varsigma\varphi i} & \Omega_{\varsigma.i} \end{bmatrix}.$$

Denote  $L_1 \sim N(0, I_r)$  and  $L_2 \sim N(0, I_p)$ . thus, as  $T \to \infty$ ,  $\frac{1}{T}F'\varepsilon_i \Rightarrow \int_0^1 B_{\varphi}d(B_{\varepsilon,i}) \equiv \xi_{i1} \sim \Omega_{\varepsilon,i}^{1/2}\Omega_{\varphi}^{1/2} \times L_1$ ,  $\frac{1}{T}V'_i\varepsilon_i \Rightarrow \int_0^1 B_{\varsigma,i}d(B_{\varepsilon,i}) \equiv \xi_{i2} \sim \Omega_{\varepsilon,i}^{1/2}\Omega_{\varsigma,i}^{1/2} \times L_2$ , where  $\xi_{i1}$  and  $\xi_{i2}$  are Gaussian processes, independent across *i*. Similarly, as  $T \to \infty$ ,

$$\frac{1}{T^2}V_i'F_1 \Rightarrow \int_0^1 B_{\varphi}B_{\varsigma,i}' \equiv \xi_{i3}, \quad \frac{1}{T^2}F'F \Rightarrow \int_0^1 B_{\varphi}B_{\varphi}' \equiv \xi_4, \quad \frac{1}{T^2}V_i'V_i \Rightarrow \int_0^1 B_{\varsigma,i}B_{\varsigma,i}' \equiv \xi_{i5} = \xi_{i5}$$

where  $\xi_{i3}, \xi_{i4}$  and  $\xi_{i5}$  are Gaussian processes. The proof of Proposition 3 above shows,  $\frac{1}{T^2} \underline{X}_i(k_1^0, k_2^0)' M_F \underline{X}_i(k_1^0, k_2^0) \Rightarrow G_i$ . According to the definitions of  $\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_4$ , and let  $\lambda = (\lambda_1^0, \lambda_2^0 - \lambda_1^0, 1 - \lambda_2^0)'$  we obtain

$$\begin{split} G_{i} &= diag(\lambda_{1}^{0}, \lambda_{2}^{0} - \lambda_{1}^{0}, 1 - \lambda_{2}^{0}) \otimes (\Gamma_{i}'\xi_{4}\Gamma_{i} + \xi_{i3}\Gamma_{i} + \Gamma_{i}'\xi_{i3}' + \xi_{i5}) \\ &- \begin{bmatrix} \lambda_{1}^{0}(\Gamma_{i}'\xi_{4} + \xi_{i3}) \\ (\lambda_{2}^{0} - \lambda_{1}^{0})(\Gamma_{i}'\xi_{4} + \xi_{i3}) \\ (1 - \lambda_{2}^{0})(\Gamma_{i}'\xi_{4} + \xi_{i3}) \end{bmatrix} \xi_{4}^{-1} [\lambda_{1}^{0}(\xi_{4}\Gamma_{i} + \xi_{i3}'), (\lambda_{2}^{0} - \lambda_{1}^{0})(\xi_{4}\Gamma_{i} + \xi_{i3}'), (1 - \lambda_{2}^{0})(\xi_{4}\Gamma_{i} + \xi_{i3}')] \\ &= diag(\lambda_{1}^{0}, \lambda_{2}^{0} - \lambda_{1}^{0}, 1 - \lambda_{2}^{0}) \otimes (\Gamma_{i}'\xi_{4}\Gamma_{i} + \xi_{i3}\Gamma_{i} + \Gamma_{i}'\xi_{i3}' + \xi_{i5}) \\ &- [\lambda \otimes (\Gamma_{i}'\xi_{4} + \xi_{i3})] \xi_{4}^{-1} [\lambda \otimes (\Gamma_{i}'\xi_{4} + \xi_{i3})]' \,. \end{split}$$

Similarly, since  $\frac{1}{T} \underline{X}_i (k_1^0, k_2^0)' M_{\bar{X}} \varepsilon_i \Rightarrow H_i$  and

$$H_{i} = \begin{bmatrix} \lambda_{1}^{0} \Gamma'_{i} \xi_{i1} + \lambda_{1}^{0} \xi_{i2} \\ (\lambda_{2}^{0} - \lambda_{1}^{0}) \Gamma'_{i} \xi_{i1} + (\lambda_{2}^{0} - \lambda_{1}^{0}) \xi_{i2} \\ (1 - \lambda_{2}^{0}) \Gamma'_{i} \xi_{i1} + (1 - \lambda_{2}^{0}) \xi_{i2} \end{bmatrix} - \begin{bmatrix} \lambda_{1}^{0} \Gamma'_{i} \xi_{4} + \lambda_{1}^{0} \xi_{i3} \\ (\lambda_{2}^{0} - \lambda_{1}^{0}) \Gamma'_{i} \xi_{4} + (\lambda_{2}^{0} - \lambda_{1}^{0}) \xi_{i3} \\ (1 - \lambda_{2}^{0}) \Gamma'_{i} \xi_{4} + (1 - \lambda_{2}^{0}) \xi_{i3} \end{bmatrix} \xi_{4}^{-1} \xi_{i1}$$
$$= \begin{bmatrix} \lambda_{1}^{0} (\xi_{i2} - \xi_{i3} \xi_{4}^{-1} \xi_{i1}) \\ (\lambda_{2}^{0} - \lambda_{1}^{0}) (\xi_{i2} - \xi_{i3} \xi_{4}^{-1} \xi_{i1}) \\ (1 - \lambda_{2}^{0}) (\xi_{i2} - \xi_{i3} \xi_{4}^{-1} \xi_{i1}) \end{bmatrix} = \lambda \otimes (\xi_{i2} - \xi_{i3} \xi_{4}^{-1} \xi_{i1}).$$

Therefore,

$$\Sigma_{MG} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E[(T^{-2}\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}))^{-1}(T^{-1}\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})'M_{\bar{X}}\varepsilon_{i}) \\ \times (T^{-1}\varepsilon_{i}'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})')(T^{-2}\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2})'M_{\bar{X}}\underline{X}_{i}(\hat{k}_{1}, \hat{k}_{2}))^{-1}] \\ = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E\{G_{i}^{-1} \left[\lambda \otimes (\xi_{i2} - \xi_{i3}\xi_{4}^{-1}\xi_{i1})\right] \left[\lambda \otimes (\xi_{i2} - \xi_{i3}\xi_{4}^{-1}\xi_{i1})\right]'G_{i}^{-1}\} \\ = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E\{G_{i}^{-1} \left[(\lambda\lambda') \otimes \left((\xi_{i2} - \xi_{i3}\xi_{4}^{-1}\xi_{i1})(\xi_{i2} - \xi_{i3}\xi_{4}^{-1}\xi_{i1})'\right)\right]G_{i}^{-1}\}.$$



Figure 1: Histograms of Break Point Estimators in Case 1 with Nonstationary Factors (T = 50)

where  $a_i \sim iidN(0.5, 0.5)$ ,  $\gamma_{2,i} \sim iidN(0.5, 0.5)$  and  $v_{it} = \rho_{vi}v_{i,t-1} + \psi_{it}, \psi_{it} \sim iidN(0, 1 - \rho_{vi}^2), v_{i,-50} = 0$ ,  $\rho_{vi} \sim iidU[0.05, 0.95]$ .  $f_t = f_{t-1} + v_{ft}, t = -50, ..., T$ ,  $v_{ft} \sim iidN(0, 1 - \rho_f^2), r_{t-50} = 0$ ,  $\rho_{vi} \sim iidU[0.05, 0.95]$ .  $f_t = f_{t-1} + v_{ft}, t = -50, ..., T$ ,  $v_{ft} \sim iidN(0, 1 - \rho_f^2), r_{t-50} = 0$ ,  $\rho_{vi} \sim iidU[0.05, 0.95]$ .  $f_t = f_{t-1} + v_{ft}, t = -50, ..., T$ ,  $v_{ft} \sim iidN(0, 1 - \rho_f^2), r_{t-50} = 0$ ,  $\rho_{vi} \sim iidU[0.05, 0.95]$ .  $f_t = f_{t-1} + v_{ft}, t = -50, ..., T$ ,  $v_{ft} \sim iidN(0, 1 - \rho_f^2), r_{t-50} = 0$ .  $\gamma_{1,i}(k_3^0) = \begin{cases} \gamma_{1,i}, & t = 1, ..., k_3^0, \\ \gamma_{1,i} + \Delta\gamma_{i}, t = k_3^0 + 1, ..., T, \end{cases}$ ,  $\gamma_{1,i} \sim iidN(1, 0.2), \Delta\gamma_i \sim iidN(0.5, 0.5).$   $\varepsilon_{it} = 0$ .  $\begin{cases} \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\omega_{it}, & i = 1, ..., [N/2], \\ \sigma_i(1 + \theta_{i\varepsilon}^2)^{-0.5}(\omega_{it} + \theta_{i\varepsilon}\omega_{i,t-1}), & i = [N/2] + 1, ..., T, \end{cases}$ ,  $\sigma_i^2 \sim iidU(0.5, 1.5), \omega_{1i} \sim iidN(0, 1), \rho_{i\varepsilon} \sim iidU[0.05, 0.95], \theta_{i\varepsilon} \sim iidU[0, 1]$ . These variables are mutually independent. The replication number is 1,000.  $T = 50, k_1^0 = [0.3T] = 15, k_2^0 = [0.5T] = 25, k_3^0 = [0.7T] = 35.$   $\hat{k}_j$ : The OLS estimator of the change point  $k_j^0, j = 1, 2, 3$ .



Figure 2: Histograms of Break Point Estimators in Case 1 Under Rank Deficiency (T = 50)

Note: The DGP is the same as that in Figure 1, except that the means of  $a_i$  and  $\gamma_{2,i}$  change to zero, i.e.,  $a_i \sim iidN(0, 0.5)$ ,  $\gamma_{2i} \sim iidN(0, 0.5)$ . In the current design, the rank condition is not satisfied asymptotically.



Figure 3: Histograms of Break Point Estimators in Case 2 (Panel Cointegration) (T = 50)

Note: The DGP is the same as that of Figure 1, except nonstationary  $v_{it} = v_{i,t-1} + \psi_{it}, \psi_{it} \sim iidN(0, 1 - \rho_{vi}^2), v_{i,-50} = 0.$ 



Figure 4: Histograms of Break Point Estimators in Case 2 with Stationary Factors (T = 50)

Note: The DGP is the same as that of Figure 3, except stationary factors,  $f_t = 0.5f_{t-1} + v_{ft}$ , t = -49, ..., 0,1, ... T,  $v_{ft} \sim iidN(0, 1 - \rho_f^2)$  with  $\rho_f = 0.5$ ,  $f_{-50} = 0$ .



Figure 5: Histograms of Break Point Estimators in Case 2 with Nonstationary Error (T = 50)

Note: The DGP is the same as that of Figure 3, except nonstationary errors,  $\varepsilon_{it} = \varepsilon_{i,t-1} + \vartheta_{it}$ , t = -49, ..., 0, 1, ..., T,  $\vartheta_{it} \sim iidN(0, 1 - \rho_{\varepsilon i}^2)$ ,  $\varepsilon_{i,-50} = 0$ .



Figure 6: Histograms of Break Point Estimators in Case 1 with Mixed Stationary and Nonstationary Regressors (T = 50)

Note: An additional regressor and factor are added in the DGP used in Figure 1 to allow for mixed stationary and nonstationary regressors.  $y_{it} = \alpha_i + \beta_{1,i}(k_1^0)x_{1,it} + \beta_{2,i}(k_2^0)x_{2,it} + \gamma_{11,i}(k_3^0)f_{1,t} + \gamma_{12,i}(k_3^0)f_{2,t} + \varepsilon_{it}$ , where  $x_{1,it} = a_i + \gamma_{21,i}f_{1,t} + \gamma_{22,i}f_{2,t} + v_{1,it}$ ,  $x_{2,it} = a_i + \gamma_{23,i}f_{2,t} + v_{2,it}$ .  $\gamma_{21,i}, \gamma_{22,i}, \gamma_{23,i} \sim iidN(0.5, 0.5)$ . Two factors  $f_{1,t} = f_{1,t-1} + v_{1,ft}$ ,  $f_{2,t} = 0.5f_{2,t-1} + v_{2,ft}$ ,  $f_{1,-50} = f_{2,-50} = 0$ .  $k_1^0 = 0.3T$ ,  $k_2^0 = 0.5T$ ,  $k_3^0 = 0.7T$ .  $\gamma_{11,i}(k_3^0)$ ,  $\gamma_{12,i}(k_3^0)$  have the same design as  $\gamma_{1,i}(k_3^0)$  in Figure 1 except the variance of the shift term changes from 0.5 to 0.16.  $\beta_{1,i}(k_1^0) = \begin{cases} \beta_{11,i}, & t = 1, ..., k_1^0 \\ \beta_{11,i} + \Delta\beta_{1,i}, & t = k_1^0 + 1, ..., T \end{cases}$  and  $\beta_{2,i}(k_2^0) = \begin{cases} \beta_{21,i}, & t = 1, ..., k_2^0 \\ \beta_{21,i} + \Delta\beta_{2,i}Af = k_2^0 + 1, ..., T \end{cases}$  with  $\Delta\beta_{1,i} \sim iidN(0, 0.16)$ .  $\Delta\beta_{2,i} \sim iidN(0, 0.16)$ .