Spring 3-2015

Estimation of Heterogeneous Panels with Structural Breaks

Badi Baltagi
bbaltagi@maxwell.syr.edu

THE MAXWELL SCHOOL

CENTER FOR POLICY RESEARCH

WORKING PAPER SERIES

Estimation of Heterogeneous Panels with Structural Breaks

Badi H. Baltagi, Qu Feng, and Chihwa Kao

ISSN: 1525-3066

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114 / email: ctrpol@syr.edu

http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

Paper No. 179
March 2015

CENTER FOR POLICY RESEARCH –Spring 2015

Leonard M. Lopoo, Director
Associate Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

John Yinger
Trustee Professor of Economics and PAIA
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi H. Baltagi ..... Economics
Robert Bifulco ..... PAIA
Thomas Dennison ..... PAIA
Alfonso Flores-Lagunes ..... Economics
Sarah Hamersma ..... PAIA
William C. Horrace ..... Economics
Yilin Hou ..... PAIA
Duke Kao ..... Economics
Sharon Kioko ..... PAIA
Jeffrey Kubik ..... Economics
Yoonseok Lee ..... Economics
Amy Lutz ..... Sociology
Yingyi Ma ..... Sociology
Jerry Miner ..... Economics
Cynthia Morrow ..... PAIA
Jan Ondrich ..... Economics
John Palmer ..... PAIA
David Popp ..... PAIA
Stuart Rosenthal ..... Economics
Ross Rubenstein ..... PAIA
Rebecca Schewe ..... Sociology
Amy Ellen Schwartz ..... PAIA/Economics
Perry Singleton ..... Economics
Abbey Steele ..... PAIA
Michael Wasylenko ..... Economics
Peter Wilcoxen ..... PAIA

GRADUATE ASSOCIATES

Emily Cardon ..... PAIA
Hannah Dalager ..... PAIA
Maidel De La Cruz ..... PAIA
Carlos Diaz ..... Economics
Vantiel Elizabeth Duncan ..... PAIA
Alex Falevich ..... Economics
Lincoln Groves ..... PAIA
Ruby Jennings ..... PAIA
Yusun Kim ..... PAIA
Bridget Lenkiewicz ..... PAIA
Michelle Lofton ..... PAIA
Roberto Martinez ..... PAIA
Qing Miao ..... PAIA
Nuno Abreu Faro E Mota ..... Economics
Judson Murchie ..... PAIA
Sun Jung Oh ..... Social Science
Brian Ohl ..... PAIA
Laura Rodriquez-Ortiz ..... PAIA
Timothy Smilnak ..... PAIA
Kelly Stevens ..... PAIA
Rebecca Wang ..... Sociology
Pengju Zhang ..... Economics
Xirui Zhang ..... Economics

STAFF

Kelly Bogart ..... Administrative Specialist
Karen Cimilluca ..... Office Coordinator
Kathleen Nasto ..... Administrative Assistant
Candi Patterson ..... Computer Consultant
Mary Santy ..... Administrative Assistant
Katrina Wingle ..... Administrative Assistant

Abstract

This paper extends Pesaran's (2006) work on common correlated effects (CCE) estimators for large heterogeneous panels with a general multifactor error structure by allowing for unknown common structural breaks. Structural breaks due to new policy implementation or major technological shocks, are more likely to occur over a longer time span. Consequently, ignoring structural breaks may lead to inconsistent estimation and invalid inference. We propose a general framework that includes heterogeneous panel data models and structural break models as special cases. The least squares method proposed by Bai (1997a, 2010) is applied to estimate the common change points, and the consistency of the estimated change points is established. We find that the CCE estimator has the same asymptotic distribution as if the true change points were known. Additionally, Monte Carlo simulations are used to verify the main results of this paper.

JEL No. C23, C33

Keywords: Heterogeneous Panels, Cross-sectional Dependence, Structural Breaks, Common Correlated Effects

Badi Baltagi-Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244, bbaltagi@maxwell.syr.edu

Qu Feng-Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, HSS-04-48, 14 Nanyang Drive, Singapore 637332, qfeng@ntu.edu.sg

Chihwa Kao-Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244, dkao@maxwell.syr.edu

The authors would like to thank the editor Han Hong, the associate editor and three anonymous referees for their constructive comments and suggestions.

1 Introduction

For panel data models, the presence of cross-sectional dependence due to unobservable common factors or spatial spillover effect is a major concern in estimation and inference. It could lead to invalid inference and inconsistent estimators, see Lee (2002) and Andrews (2005).[1] Several tests for cross-sectional dependence in panel data models have been proposed in the literature. These include Pesaran (2004, 2012), Ng (2006), Pesaran, Ullah and Yamagata (2008), Sarafidis, Yamagata and Robertson (2009), Chen, Gao and Li (2011), Hsiao, Pesaran and Pick (2012), Baltagi, Feng and Kao (2011, 2012), Halunga, Orme and Yamagata (2011), Juhl (2011), and Su and Zhang (2011), to mention a few. To deal with cross-sectional dependence in panels, two general estimation methods have been proposed including spatial estimation methods (Anselin, 1988; Kelejian and Prucha, 1999; Kapoor, Kelejian and Prucha, 2007; Lee, 2007 and Lee and Yu, 2010, to name a few), and factor models (see Pesaran, 2006; and Bai, 2009, to name a few).

In particular, Pesaran (2006) develops common correlated effects (CCE) estimators for large heterogeneous panels with a general multifactor error structure by least squares using augmented data. The common correlated effects (factors) can be aysmptotically partialled out by means of the cross-sectional average of the dependent variable and the individual-specific regressors when the cross-section dimension is large. Kapetanios, Pesaran and Yamagata (2011) show that the CCE estimator can be extended to the case of nonstationary unobserved common factors. Additionally, the CCE approach is also shown to be applicable to situations of spatial and other forms of weak cross-sectional dependent errors (Pesaran and Tosetti, 2011; Chudik, Pesaran and Tosetti, 2011), and heterogenous dynamic panel data models with weakly exogenous regressors (Chudik and Pesaran, 2013).[2]

However, this literature assumes that the slope coefficients are constant over time. This implicit assumption is common in the literature on panel data models with large time dimension, see for example, Kao (1999), Phillips and Moon (1999), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Phillips and Sul (2007), Pesaran and Yamagata (2008), Hayakawa (2009), to name a few. Due to policy implementation or technological shocks,

[1] Baltagi and Pirotte (2010) show that tests of hypotheses based on standard panel data estimators that ignore spatial dependence lead to misleading inference.

[2] In a panel with unobserved common factors, Banerjee and Carrion-i-Silvestre (2011) suggest a test for panel cointegration based on a pooled CCE estimator of the coefficients.

structural breaks are possible especially for panels with a long time span. Consequently, ignoring structural breaks may lead to inconsistent estimation and invalid inference.

This paper extends Pesaran's (2006) heterogeneous panels by allowing for unknown common structural breaks in the slopes. This is useful for example when global technological or financial shocks affect all markets or firms at the *same* time. Since the framework of heterogeneous panels is fairly general and includes popular panel data models as special cases, it allows us to examine the issue of structural breaks in panel data models in a less restrictive way.

By considering both cross-sectional dependence and structural breaks in a general panel data model, this paper also contributes to the change point literature in several ways. First, it extends Bai's (1997a) time series regression model to heterogeneous panels, showing that the consistency of estimated change points can be achieved with the information along the cross-sectional dimension. This result confirms the findings of Bai (2010) and Kim (2011). Second, it also enriches the analysis of common breaks of Bai (2010) and Kim (2011) in a panel mean-shift model and a panel deterministic time trend model by extending them to a regression model using panel data. This makes it possible to allow for structural breaks and cross-sectional dependence in empirical work using panel regressions. In particular, our methods can be applied to regression models using large stationary panel data, such as country-level panels and state/provincial-level panels.[3]

Regarding estimating common breaks in panels, Feng, Kao and Lazarova (2009) and Baltagi, Kao and Liu (2012) also show the consistency of the estimated change point in a simple panel regression model. Hsu and Lin (2012) examine the consistency properties of the change point estimators for nonstationary panels. More recently, Qian and Su (2014) and Li, Qian and Su (2014) study the estimation and inference of common breaks in panel data models with and without interactive fixed effects using Lasso-type methods. In terms of detecting structural breaks in panels, some recent literature includes Horváth and Hušková (2012) in a panel mean shift model with and without cross-sectional dependence, De Wachter and Tzavalis (2012) in dynamic panels, and Pauwels, Chan and Mancini-Griffoli (2012) in heterogeneous panels, to name a few.

The paper is organized as follows. Section 2 introduces heterogeneous panels with a

[3]Some empirical examples include Fleisher, Li and Zhao (2010) using Chinese provincial-level panel data, and Huang (2009) using country-level panels, to mention a few.

common structural break. Section 3 starts with a simple heterogeneous panel model that ignores the unobservable common correlated effects. This model can be regarded as a direct extension of Bai's (1997a) time series regression model to a panel setup. The least squares estimation proposed by Bai (1997a) is applied. With the main results established in Section 3, the discussion of the general model with common correlated effects is presented in Section 4. Section 5 briefly discusses the case of multiple common breaks. In Section 6, Monte Carlo simulations are used to verify the consistency of the estimated common change point for both models considered. Section 7 provides concluding remarks. The mathematical proofs are relegated to the Appendix.

2 Heterogeneous Panels with a Common Structural Break

In a heterogeneous panel data model:

$$y_{it} = x_{it}'\beta_i + e_{it}, \qquad i = 1, ..., N;\ t = 1, ..., T, \tag{1}$$

x_{it} is a $p \times 1$ vector of explanatory variables, and the errors are cross-sectionally correlated, modelled by a multifactor structure

$$e_{it} = \gamma_i' f_t + \varepsilon_{it}, \tag{2}$$

where f_t is an $m \times 1$ vector of unobserved factors and γ_i is the corresponding loading vector. ε_{it} is the idiosyncratic error independent of x_{it}. However, x_{it} could be affected by the unobservable common effects f_t. Projecting x_{it} on f_t, we obtain

$$x_{it} = \Gamma_i' f_t + v_{it},\ i = 1, ..., N;\ t = 1, ..., T, \tag{3}$$

where Γ_i is an $m \times p$ factor loading matrix. v_{it} is a $p \times 1$ vector of disturbances. Due to the correlation between x_{it} and e_{it}, OLS for each individual regression could be inconsistent. Thus, Pesaran (2006) develops the CCE estimator of β_i by least squares using augmented data.[4]

[4]For simplicity, observed common effects like seasonal dummies are not included in (1), but they can be easily handled as in Pesaran (2006).

In this paper, we allow for structural breaks to occur in some or all components of the slopes β_i.[5] Following Bai (2010) and Kim (2011), a structural break at a common unknown date k_0 is assumed. This could be due to a macro policy implementation or a technological shock that affects all markets or firms at the same time. More formally,

$$y_{it} = x'_{it}\beta_i(k_0) + e_{it},\ \ i = 1, ..., N; t = 1, ..., T, \tag{4}$$

where some or all components of $\beta_i(k_0)$ are different before and after the date k_0.[6] Following Bai (1997a), this structural break model can be written as

$$y_{it} = \begin{cases} x'_{it}\beta_i + e_{it}, & t = 1, ..., k_0, \\ x'_{it}\beta_i + z'_{it}\delta_i + e_{it}, & t = k_0 + 1, ..., T, \end{cases} \tag{5}$$

$i = 1, ..., N$, where $z_{it} = R'x_{it}$ denotes a $q \times 1$ subvector of x_{it} with $R' = (0_{q\times(p-q)}, I_q)$. I_q is the $q \times q$ identity matrix with $q \leq p$. The case where $q < p$ denotes a partial change model, while the case where $q = p$ is for a pure change model. Pauwels, Chan and Mancini-Griffoli (2012) propose a testing procedure for k_0 in this setting.

Substituting $z_{it} = R'x_{it}$ in (5), we obtain

$$\beta_i(k_0) = \beta_i + R\delta_i \cdot 1\{t > k_0\} = \begin{cases} \beta_{1i} = \beta_i, & t = 1, ..., k_0, \\ \beta_{2i} = \beta_i + R\delta_i, & t = k_0 + 1, ..., T, \end{cases}$$

so that $\beta_{2i} - \beta_{1i} = R\delta_i$, and δ_i denotes the slope jump for i. When $\delta_i = 0$ there is no structural break in the slope.

The case of multiple break points will be discussed in Section 5. In the next two sections, we consider the simple case of one common break as in model (5). Compared with the heterogeneous panel data model considered in Pesaran (2006), (5) has the extra component $R\delta_i \cdot 1\{t > k_0\}$ in the slope, involving the unknown structural change point k_0. Thus, ignoring the structural break in the slopes may invalidate the CCE estimator proposed by

[5] Pesaran (2004) discusses testing for cross-sectional dependence in a heterogeneous panel model with structural breaks. Kapetanios, Pesaran and Yamagata (2011) examine the performance of the CCE estimator in case of a structural break in the mean of the unobserved factors using Monte Carlo experiments.

[6] As shown in Section 4, to apply the CCE approach, the common break assumption is required. In Kim (2014), the common break assumption is generalized to handle heterogeneous responses to a common shock that follow a common distribution. Liao and Wang (2012) also assume a common distribution, instead of a common break date, and estimate individual-specific structural breaks and their cross-sectional distribution using Bayesian methods. In addition, Li, Chen and Gao (2011) consider a time-varying coefficient panel data model where the slope coefficient is allowed to be different for each time period, e.g., $\beta(t)$, but common for all individuals.

Pesaran (2006).[7] Compared with the simple mean shift panel data model in Bai (2010), our model is enriched by adding a regression structure with $x_{it} \neq 1$ in general, as well as cross-sectional dependence characterized by a multifactor structure in the errors. When there are no unobservable common factors f_t, our model (4) can also be regarded as an extension of Bai (1997a) to a panel data setting. In addition, the model (4) above is similar to Kim (2011), who considered the case of a deterministic time trend with a common break.

Before proceeding to the general model (5), we start with a simple case of heterogeneous panels in the absence of common correlated effects f_t in Section 3, and then extend the main results to the general case in Section 4.

To estimate the common change point k_0, we need the following additional assumptions:

Assumption 1 *$k_0 = [\tau_0 T]$, where $\tau_0 \in (0,1)$ and $[\cdot]$ is the greatest integer function.*

Note that unlike the time series model considered by Bai (1997a), the restriction of $\tau_0 \in (0,1)$ is unnecessary in a panel mean shift setup considered by Bai (2010) as long as $T/N \to 0$. However, this assumption is required in our heterogeneous panels with general regressors. Enough observations are needed to consistently estimate the slopes in each regime.

Define $\phi_N = \sum_{i=1}^{N} \delta_i' \delta_i$. For series i, $\delta_i' \delta_i$ measures the magnitude of the structural break, thus ϕ_N is an indicator of the break magnitude for all N series sharing a common break.

Assumption 2 *$\phi_N \to \infty$ and (i) $\frac{\phi_N}{N}$ is bounded as $N \to \infty$; (ii) $\phi_N \frac{T}{N} \to \infty$ and $\phi_N \sqrt{\frac{T}{N}} \to \infty$ as $(N,T) \to \infty$.*

δ_i could be random with a finite variance across i, with Assumption 2(i) describing this case. When δ_i is considered as random, Assumption 2 means that $\frac{\phi_N}{N}$ is stochastically bounded in part (i), and that $\frac{N}{\phi_N T}$ and $\frac{\sqrt{N}}{\phi_N \sqrt{T}}$ converge in probability to 0 in part (ii). Alternatively, δ_i could denote fixed parameters. Since Assumption 2(i) allows for the case where $\frac{\phi_N}{N} \to 0$ as $N \to \infty$, Assumption 2(ii) implies that it cannot converge to 0 too fast. Consequently, Assumption 2(i) imposes an upper bound on $\frac{\phi_N}{N}$, while Assumption 2(ii) imposes a lower bound on $\frac{\phi_N}{N}$.

[7] As documented in the time series literature, e.g., Pesaran and Timmermann (2002), structural breaks could lead to forecast failure. Stock and Watson (2002, 2009) show that forecasts constructed using dynamic factor models are robust to small structural breaks of factor loadings. However, when the size of the break is big, ignoring structural breaks may yield biased and inconsistent forecasts, as illustrated by Banerjee, Marcellino and Masten (2008), Breitung and Eickmeier (2011), and Chen, Dolado and Gonzalo (2014).

In case T grows at a comparable rate or faster than N, i.e., $T = O(N^{\psi})$ with $\psi \geq 1$, Assumption 2(ii) implies that ϕ_N can diverge at any rate. When ϕ_N increases at a rate less than N, Assumption 2(ii) allows for the possibility of no structural break in some series.[8]

3 Model 1: No Common Correlated Effects

In this section, we assume that there are no unobserved common effects f_t in the errors and regressors. Or the loading vectors γ_i and Γ_i are equal to zero. For $i = 1, ..., N$,

$$y_{it} = \begin{cases} x_{it}'\beta_i + \varepsilon_{it}, & t = 1, ..., k_0, \\ x_{it}'\beta_i + z_{it}'\delta_i + \varepsilon_{it}, & t = k_0 + 1, ..., T. \end{cases} \tag{6}$$

This is the special case of cross-sectionally independent errors, where a common break k_0 occurs in the heterogeneous slopes. This model generalizes Bai (1997a), Bai (2010) and Pesaran and Smith (1995). When $N = 1$, (6) is the time series model considered in Bai (1997a). When $x_{it} = 1$, this model reduces to the one in Bai (2010). In case the lagged dependent variable is included in x_{it} and $\delta_i = 0$, (6) turns out to be the setup in Pesaran and Smith (1995).[9]

Assumption 3 *(i) The disturbances $\varepsilon_{it}, i = 1, ..., N$, are cross-sectionally independent; (ii) For each series i, ε_{it} is independent of x_{it} for all i and t; (iii) ε_{it} is a stationary process with absolute summable autocovariances,*

$$\varepsilon_{it} = \textstyle\sum_{l=0}^{\infty} a_{il}\zeta_{i,t-l}$$

where $\{\zeta_{it}, t = 1, ..., T\}$ are independent and identically distributed (IID) random variables with finite fourth-order cumulants. Assume $0 < Var(\varepsilon_{it}) = \sum_{l=0}^{\infty} a_{il}^2 = \sigma_i^2 < \infty$. Also, for the $T \times 1$ vector $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$, $Var(\varepsilon_i) = \Sigma_{\varepsilon,i}$.

When ε_{it} is serially uncorrelated, lagged dependent variables are predetermined and can be included as regressors in (6).

Assumption 4 *For $i = 1, ..., N$, the matrices $(1/j)\sum_{t=1}^{j} x_{it}x_{it}'$, $(1/j)\sum_{t=T-j+1}^{T} x_{it}x_{it}'$, $(1/j)\sum_{t=k_0-j+1}^{k_0} x_{it}x_{it}'$ and $(1/j)\sum_{t=k_0+1}^{k_0+j} x_{it}x_{it}'$ are stochastically bounded and have minimum*

[8] Assumption $\phi_N \to \infty$ rules out the case where there is no structural break in the slopes in all series.

[9] Heterogeneous dynamic panel data models with weakly exogenous regressors and unobserved common factors are studied by Chudik and Pesaran (2013).

eigenvalues bounded away from zero in probability for all large j. In addition, for each i, $(1/T)\sum_{t=1}^{T} x_{it}x_{it}'$ converges in probability to a nonrandom and positive definite matrix as $T \to \infty$.

This assumption is borrowed from Assumptions A3 and A4 in Bai (1997a).[10] Its counterpart across the cross-sectional dimension is also needed.

Assumption 5 *For any positive finite integer s, the matrices $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=k_0-s+1}^{k_0} x_{it}x_{it}'$ and $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k_0+s} x_{it}x_{it}'$, $i = 1, ..., N$, are stochastically bounded, with minimum eigenvalues bounded away from zero in probability for large N. In addition, for each t, $(1/N)\sum_{i=1}^{N} x_{it}x_{it}'$ is stochastically bounded as $N \to \infty$.*

Assumption 6 *$\{\delta_i, i = 1, ..., N\}$ are drawn independently of $\{x_{it}, i = 1, ..., N\}$.*

Let $b_i = (\beta_i', \delta_i')', i = 1, ..., N$, denote the slope parameters. In the random coefficient model considered by Pesaran and Smith (1995) and Pesaran (2006), we assume:

Assumption 7 *For $i = 1, ..., N$,*

$$b_i = b + v_{b,i}, v_{b,i} \sim IID(0, \Sigma_b), \tag{7}$$

where $b = (\beta', \delta')'$, $v_{b,i} = \begin{pmatrix} v_{\beta,i} \\ v_{\delta,i} \end{pmatrix}$ and $\Sigma_b = \begin{pmatrix} \Sigma_\beta & 0 \\ 0 & \Sigma_\delta \end{pmatrix}$ for $i = 1, 2, ..., N$, where $\|b\| < \infty$, $\|\Sigma_b\| < \infty$, and the random deviations $v_{b,i}$ are independent of x_{it} and ε_{jt} for all i, j and t.

For any matrix or vector A, the norm of A is defined as $\|A\| = \sqrt{tr(AA')}$. This assumption is a simplified version of Assumption 4 of Pesaran (2006). Under Assumption 6, $\{\delta_i, i = 1, ..., N\}$ are not necessarily random. When $\{\delta_i, i = 1, ..., N\}$ are considered as random, as part of Assumption 7, Assumption 6 becomes redundant. Under Assumption 7, $\Sigma_\delta \neq 0$ implies a structural break in the slope.

By (4),

$$y_{it} = x_{it}'\beta_i + x_{it}'R\delta_i 1\{t > k_0\} + \varepsilon_{it},$$

if the structural break is ignored, the term $x_{it}'R\delta_i 1\{t > k_0\}$ is absorbed in the error term $\hat{\varepsilon}_{it} = x_{it}'R\delta_i 1\{t > k_0\} + \varepsilon_{it}$. This leads to inconsistency of OLS for each series due to endogeneity. Thus, estimating k_0 first is essential.

[10] As pointed out by Bai (1997a), Bai and Perron (1998, 2003) trending regressors are allowed in the form of $(t/T)^l$, for $(l > 0)$, or any function of the time trend: $g(t/T)$.

Let $Y_i = (y_{i1}, \cdots, y_{iT})'$, $X_i = (x_{i1}, \cdots, x_{iT})'$ and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$ denote the stacked data and errors for individual $i = 1, ..., N$ over the time periods observed. Similarly, define $Z_{0i} = (0, \cdots, 0, z_{i,k_0+1}, \cdots, z_{iT})'$. (6) can be written in matrix form as

$$Y_i = X_i\beta_i + Z_{0i}\delta_i + \varepsilon_i, \quad i = 1, ..., N. \tag{8}$$

The parameters of interest are β_i, δ_i and the change point k_0. We first estimate k_0 using least squares as proposed by Bai (1997a, 2010). For any possible change point $k = 1, ..., T-1$, define the matrices $X_{2i}(k) = (0, \cdots, 0, x_{i,k+1}, \cdots, x_{iT})'$, and $Z_{2i}(k) = (0, \cdots, 0, z_{i,k+1}, \cdots, z_{iT})'$. When k happens to be the true change point k_0, $Z_{2i}(k_0) = Z_{0i}$. Define $X_{0i} = X_{2i}(k_0)$, thus $Z_{0i} = X_{0i}R$. To make the notation more compact, we let $\mathbb{X}_i(k) = (X_i, Z_{2i}(k))$ and $\mathbb{X}_{0i} = (X_i, Z_{0i})$. Thus, (8) becomes

$$Y_i = X_i\beta_i + Z_{0i}\delta_i + \varepsilon_i = \mathbb{X}_{0i}b_i + \varepsilon_i, \quad i = 1, ..., N. \tag{9}$$

Given any $k = 1, ..., T-1$, one can estimate b_i by least squares,

$$\hat{b}_i(k) = \begin{pmatrix} \hat{\beta}_i(k) \\ \hat{\delta}_i(k) \end{pmatrix} = [\mathbb{X}_i(k)'\mathbb{X}_i(k)]^{-1}\mathbb{X}_i(k)'Y_i, \; i = 1, ..., N. \tag{10}$$

The corresponding sum of squared residuals is given by

$$\begin{aligned} SSR_i(k) &= [Y_i - \mathbb{X}_i(k)\hat{b}_i(k)]'[Y_i - \mathbb{X}_i(k)\hat{b}_i(k)] \\ &= [Y_i - X_i\hat{\beta}_i(k) - Z_{2i}(k)\hat{\delta}_i(k)]'[Y_i - X_i\hat{\beta}_i(k) - Z_{2i}(k)\hat{\delta}_i(k)], \end{aligned}$$

$i = 1, ..., N$. Note that both $\hat{b}_i(k)$ and $SSR_i(k)$ depend on k**.** For each series i, k_0 can be estimated by $\arg\min_{1\le k\le T-1} SSR_i(k)$ as in Bai (1997a). Given that the structural break occurs at a common date for all cross-sectional units in the panel setup, the least squares estimator of k_0 is defined as

$$\hat{k} = \arg \min_{1\le k\le T-1} \sum_{i=1}^{N} \pi_i SSR_i(k). \tag{11}$$

Weights $\{\pi_i \in (0,1), \; i = 1, ..., N, \; \sum_{i=1}^{N} \pi_i = 1\}$ allow for the possibility of different magnitudes, e.g., different variances, across series.

When $N = 1$, $\hat{k}$ defined in (11) boils down to the change-point estimator considered by Bai (1997a) in a time series setting, with $\hat{k} - k_0 = O_p(1)$ for large T. In time series models, only the

break fraction $\tau_0 = k_0/T$, instead of k_0 itself, can be consistently estimated. In a multivariate time series set up, Bai, Lumsdaine and Stock (1998) show that the width of the confidence interval of the estimated change point decreases with the number of time series.[11] This result implies that cross-sectional observations with common breaks improve the accuracy of the estimated change point. In fact, Bai (2010) shows that the least squares estimator of the change point is consistent in a panel mean-shift model, i.e., $\hat{k} - k_0 = o_p(1)$. A similar result is also obtained by Kim (2011) in a panel deterministic time trend model. In our heterogeneous panel regression model, (11) combines the information from each series by summing up $SSR_i(k)$. With a large N, $\hat{k}$ uses more information provided by the multiple time series sharing a common break. Consequently, the panel data estimator $\hat{k}$ is more accurate than the time-series estimator and achieves consistency, i.e., $\hat{k} - k_0 \xrightarrow{p} 0$ as $(N,T) \to \infty$.

Theorem 1 *Under Assumptions 1-6 (or 7),* $\lim_{(N,T)\to\infty} P(\hat{k} = k_0) = 1$.

The proof of Theorem 1 can be found in the Appendix.

Given the estimated change point $\hat{k}$, the corresponding estimator of the slopes is $\hat{b}_i = \hat{b}_i(\hat{k})$, $i = 1, ..., N$. When b_i, $i = 1, ..., N$, are considered as random variables under Assumption 7, the cross-sectional mean b can be consistently estimated by the mean group estimator proposed by Pesaran and Smith (1995) and Pesaran (2006):[12]

$$\hat{b}_{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{b}_i = \frac{1}{N}\sum_{i=1}^{N} [\mathbb{X}_i(\hat{k})_i' \mathbb{X}_i(\hat{k})]_i^{-1} \mathbb{X}_i(\hat{k})' Y_i. \tag{12}$$

4 Model 2: Common Correlated Effects

In this section, we extend Model 1 to the general model with common correlated effects (5): for $i = 1, ..., N$,

$$y_{it} = x_{it}'\beta_i(k_0) + e_{it} = \begin{cases} x_{it}'\beta_i + e_{it}, & t = 1, ..., k_0, \\ x_{it}'\beta_i + z_{it}'\delta_i + e_{it}, & t = k_0 + 1, ..., T. \end{cases}$$

where $e_{it} = \gamma_i' f_t + \varepsilon_{it}$. The regressors x_{it}, $i = 1, ..., N$, are allowed to be correlated with the unobservable factors f_t modelled in (3), $x_{it} = \Gamma_i' f_t + v_{it}$. When $\delta_i = 0$, the model reduces to

[11] Change-point estimators in the multivariate time series literature are discussed in Bai (2000) and Qu and Perron (2006).

[12] Note that the pooled estimator of b considered by Pesaran (2006) can be studied similarly. Since the asymptotic distributions of $\hat{b}_i$ and $\hat{b}_{MG}$ are similar to those derived in Bai (1997a) and Pesaran (2006), they are summarized in the Supplementary Appendix.

the one considered by Pesaran (2006). Kim (2011) considers the special case of $x_{it} = (1, t)'$.[13] In this heterogeneous panel data model with a common break k_0, the parameters of interest are $b_i = (\beta_i', \delta_i')'$, $i = 1, ..., N$, and the change point k_0. The following assumptions are needed.

Assumption 8 *Common factors f_t, $t = 1, ..., T$, are covariance sationary with absolute summable autocovariances, independent of errors ε_{is} and v_{is} for all i, s, t.*

Assumption 9 *Errors ε_{is} and v_{jt} are independent for all i, j, s, t. v_{it}, $i = 1, ..., N$, are linear stationary processes with absolute summable autocovariances, $v_{it} = \sum_{l=0}^{\infty} S_{il} \upsilon_{i,t-l}$, where $(\zeta_{it}, \upsilon_{it}')'$ are $(p+1) \times 1$ vectors of IID random variables with variance-covariance matrix I_{p+1} and finite fourth-order cumulants, and*

$$Var(v_{it}) = \sum_{l=0}^{\infty} S_{il} S_{il}' = \Sigma_{i,v}, \text{ and } 0 < \|\Sigma_{i,v}\| < \infty.$$

Assumption 10 *Factor loadings γ_i and Γ_i are IID across i, and independent of ε_{jt}, v_{jt} and f_t for all i, j, t. Assume $\gamma_i = \gamma + \eta_i$, $\eta_i \sim IID(0, \Omega_\eta)$ and $\Gamma_i = \Gamma + \xi_i$, $\xi_i \sim IID(0, \Omega_\xi)$, $i = 1, ...N$, where the means γ, Γ are non-zero and fixed and the variances Ω_η, Ω_ξ are finite.*

Together with Assumptions 3 and 7, Assumptions 8, 9 and 10 given above are the same as Assumptions 1-3 of Pesaran (2006), with the additional restrictions $\gamma \neq 0$ and $\Gamma \neq 0$.

The correlation between x_{it} and e_{it} due to unobserved common factors f_t renders OLS inconsistent. If f_t were observable, it could be treated as a regressor, and this correlation can be removed using a partitioned regression. Let $F = (f_1, f_2, \cdots, f_T)'$, then the corresponding orthogonal projection matrix is given by $M_f = I_T - F(F'F)^{-1}F'$. In this case, (5) can be written in matrix form as

$$Y_i = X_i \beta_i + Z_{0i} \delta_i + F\gamma_i + \varepsilon_i, \qquad i = 1, ..., N \tag{13}$$

Premultiplying (13) by M_f, we get

$$\breve{Y}_i = \breve{X}_i \beta_i + \breve{Z}_{0i} \delta_i + \breve{\varepsilon}_i, i = 1, ..., N, \tag{14}$$

which is of the same form as equation (8) considered in Section 3, with transformed data $\breve{Y}_i = M_f Y_i$, $\breve{X}_i = M_f X_i = M_f V_i$, $\breve{Z}_{0i}(k_0) = M_f Z_{0i}$ and $\breve{\varepsilon}_i = M_f \varepsilon_i$. For each $i = 1, ..., N$, the

[13] In a similar panel set up without exogenous regressors, Bai and Carrion-i-Silvestre (2009) develop unit root tests applicable to situations of multiple structural breaks and unobserved common dynamic factors.

$T \times p$ vector V_i denotes $(v_{i1}, ..., v_{iT})'$. Conditional on F, $(\breve{X}_i, \breve{Z}_{0i})$ and $\breve{\varepsilon}_i$ are uncorrelated under Assumption 9.

However, f_t, $t = 1, ..., T$, are unobservable. To proceed, we follow Pesaran's (2006) idea of using the cross-sectional averages of y_{it} and x_{it} as proxies for f_t. Combining (5) and (3) yields

$$\underset{(p+1)\times 1}{w_{it}} = \begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \underset{(p+1)\times m}{C_i(k_0)'} \underset{m\times 1}{f_t} + \underset{(p+1)\times 1}{u_{it}(k_0)}, \tag{15}$$

where

$$\underset{m\times(p+1)}{C_i(k_0)} = (\gamma_i, \Gamma_i) \begin{pmatrix} 1 & 0 \\ \beta_i(k_0) & I_p \end{pmatrix} \text{ and } u_{it}(k_0) = \begin{pmatrix} \varepsilon_{it} + v_{it}'\beta_i(k_0) \\ v_{it} \end{pmatrix}.$$

Note that like $\beta_i(k_0)$, the slope $C_i(k_0)$ in (15) also shifts at k_0.

$$C_i(k_0) = \begin{cases} C_{1i} = (\gamma_i + \Gamma_i\beta_{1i},\ \Gamma_i), & t = 1, ..., k_0, \\ C_{2i} = (\gamma_i + \Gamma_i\beta_{2i},\ \Gamma_i), & t = k_0 + 1, ..., T. \end{cases} \tag{16}$$

Common break k_0 splits the data generating process for all individuals into two regimes, and each regime has the same structure as that considered in Pesaran (2006). Consequently, unobserved common factors f_t can be partialled out by using cross-sectional averages in the same spirit.

Let $\bar{w}_t = \sum_{i=1}^N \theta_i w_{it}$ be the cross-sectional averages of w_{it} using weights θ_i, $i = 1, ..., N$. In particular,

$$\bar{w}_t = \bar{C}(k_0)' f_t + \bar{u}_t(k_0) \tag{17}$$

where $\bar{C}(k_0) = \sum_{i=1}^N \theta_i C_i(k_0) = \begin{cases} \bar{C}_1 = \sum_{i=1}^N \theta_i C_{1i}, & t = 1, ..., k_0, \\ \bar{C}_2 = \sum_{i=1}^N \theta_i C_{2i}, & t = k_0 + 1, ..., T. \end{cases}$

The common break assumption is needed, otherwise $\bar{C}(k_0)$ is not well defined. $\bar{u}_t(k_0)$ is defined as

$$\bar{u}_t(k_0) = \sum_{i=1}^N \theta_i u_{it}(k_0) = \begin{cases} \begin{pmatrix} \bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}'\beta_{1i} \\ \bar{v}_t \end{pmatrix}, & t = 1, ..., k_0, \\ \begin{pmatrix} \bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}'\beta_{2i} \\ \bar{v}_t \end{pmatrix}, & t = k_0 + 1, ..., T. \end{cases} \tag{18}$$

As in Pesaran (2006), the weights θ_i, $i = 1, ..., N$, satisfy conditions: $\theta_i = O(\frac{1}{N})$, $\sum_{i=1}^N \theta_i = 1$ and $\sum_{i=1}^N |\theta_i| < \infty$.

Assumption 11 *Rank*$(\bar{C}_1) =$ *Rank*$(\bar{C}_2) = m \leq p + 1$.

We assume that rank condition is satisfied. Pesaran (2006) shows that in the case of deficient rank, it is impossible to obtain consistent estimators of the individual slope coefficients, but their cross-sectional mean can be consistently estimated. When $\bar{C}(k_0)$ is of full rank, f_t can be written as

$$f_t = \left[\bar{C}(k_0)\bar{C}(k_0)'\right]^{-1}\bar{C}(k_0)(\bar{w}_t - \bar{u}_t(k_0)).$$

From (16), the matrix $\bar{C}(k_0)\bar{C}(k_0)'$ has two regimes, shifting at k_0,

$$\bar{C}(k_0)\bar{C}(k_0)' = \begin{cases} \bar{C}_1'\bar{C}_1, & t = 1, ..., k_0, \\ \bar{C}_2'\bar{C}_2, & t = k_0+1, ..., T. \end{cases}$$

Assumption 11 implies that $\bar{C}(k_0)\bar{C}(k_0)'$ is invertible. As shown in Lemma 1 of Pesaran (2006), the cross-sectional average of the errors vanish in both regimes as $N \to \infty$, where $\bar{\varepsilon}_t = \sum_{i=1}^N \theta_i \varepsilon_{it}$, $\bar{v}_t = \sum_{i=1}^N \theta_i v_{it}$, yielding

$$f_t - \left[\bar{C}(k_0)\bar{C}(k_0)'\right]^{-1}\bar{C}(k_0)\bar{w}_t \xrightarrow{p} 0. \tag{19}$$

This suggests that it is asymptotically valid to use $\bar{w}_t$ as observable proxies for f_t. Let $\bar{W} = (\bar{w}_1, \bar{w}_2, \cdots, \bar{w}_T)'$ denote the $T \times (p+1)$ matrix of cross-sectional averages. Denote the $T \times T$ matrix M_w by $M_w = I_T - \bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'$. Thus, similar to the result $M_f F = 0$, by (19) it is expected that the terms involving $M_w F$ are ignorable asymptotically as $N \to \infty$.

Premultiplying (13) by M_w instead of M_f, we obtain

$$M_w Y_i = M_w X_i \beta_i + M_w Z_{0i}\delta_i + M_w F\gamma_i + M_w \varepsilon_i, i = 1, ..., N. \tag{20}$$

Let the $T \times p$ matrix $\tilde{X}_i = M_w X_i = (\tilde{x}_{i1}, \cdots, \tilde{x}_{iT})'$ denote the transformed regressors. Similarly, define $\tilde{Y}_i = M_w Y_i$, $\tilde{Z}_{0i} = M_w Z_{0i}$ and $\tilde{\varepsilon}_i = M_w \varepsilon_i$. Thus, (20) becomes

$$\tilde{Y}_i = \tilde{X}_i\beta_i + \tilde{Z}_{0i}\delta_i + M_w F\gamma_i + \tilde{\varepsilon}_i = \tilde{X}_i\beta_i + \tilde{Z}_{0i}\delta_i + \tilde{\varepsilon}_i^0, i = 1, ..., N, \tag{21}$$

where $\tilde{\varepsilon}_i^0 = M_w F\gamma_i + \tilde{\varepsilon}_i$.

Lemma 6 in the Appendix shows that each element of $M_w F\gamma_i$ is of order $O_p(\frac{1}{\sqrt{N}})$ and vanishes as $(N, T) \to \infty$, implying that $\tilde{\varepsilon}_i^0$ can be treated as $\tilde{\varepsilon}_i$ asymptotically. Based on this intuition, we can follow the procedure proposed in Section 3 to estimate k_0 and $b_i = (\beta_i', \delta_i')'$, using transformed data $\{\tilde{Y}_i, \tilde{X}_i, i = 1, ..., N\}$.

For any possible change point $k = 1, ..., T-1$, define matrices $\tilde{Z}_{2i}(k) = M_w Z_{2i}(k)$, $\tilde{\mathbb{X}}_i(k) = (\tilde{X}_i, \tilde{Z}_{2i}(k))$ and $\tilde{\mathbb{X}}_{0i} = (\tilde{X}_i, \tilde{Z}_{0i})$. With new notation, (21) becomes

$$\tilde{Y}_i = \tilde{\mathbb{X}}_{0i} b_i + \tilde{\varepsilon}_i^0, \ i = 1, ..., N. \tag{22}$$

Given k, slope b_i can be estimated by least squares,

$$\tilde{b}_i(k) = \begin{pmatrix} \tilde{\beta}_i(k) \\ \tilde{\delta}_i(k) \end{pmatrix} = [\tilde{\mathbb{X}}_i(k)' \tilde{\mathbb{X}}_i(k)]^{-1} \tilde{\mathbb{X}}_i(k)' \tilde{Y}_i, \ i = 1, ..., N. \tag{23}$$

The resulting sum of squared residuals is

$$\begin{aligned} \widetilde{SSR}_i(k) &= [\tilde{Y}_i - \tilde{\mathbb{X}}_i(k)\tilde{b}_i(k)]'[\tilde{Y}_i - \tilde{\mathbb{X}}_i(k)\tilde{b}_i(k)] \\ &= [\tilde{Y}_i - \tilde{X}_i\tilde{\beta}_i(k) - \tilde{Z}_{2i}(k)\tilde{\delta}_i(k)]'[\tilde{Y}_i - \tilde{X}_i\tilde{\beta}_i(k) - \tilde{Z}_{2i}(k)\tilde{\delta}_i(k)], \ i = 1, ..., N, \end{aligned}$$

and the estimator of k_0 is defined similarly as

$$\tilde{k} = \arg \min_{1 \le k \le T-1} \textstyle\sum_i \pi_i \widetilde{SSR}_i(k), \tag{24}$$

where π_i are weights, as in (11).

Assumption 12 *For $i = 1, ..., N$, the matrices $\frac{1}{T} X_i' M_w X_i$ and $\frac{1}{T} X_i' M_f X_i$ are nonsingular, and their inverses have finite second-order moments.*

This assumption of identifying b_i and b is adopted from Pesaran (2006).

Let $\tilde{x}_{it}'$ be the t^{th} element of matrix $\tilde{X}_i$, $i = 1, ..., N$. To identify k_0, we need a modified version of Assumptions 4, 5, 6:

Assumption 13 *For $i = 1, ..., N$, the matrices $(1/j)\sum_{t=1}^{j} \tilde{x}_{it}\tilde{x}_{it}'$, $(1/j)\sum_{t=T-j+1}^{T} \tilde{x}_{it}\tilde{x}_{it}'$, $(1/j)\sum_{t=k_0-j+1}^{k_0} \tilde{x}_{it}\tilde{x}_{it}'$ and $(1/j)\sum_{t=k_0+1}^{k_0+j} \tilde{x}_{it}\tilde{x}_{it}'$ are stochastically bounded and have minimum eigenvalues bounded away from zero in probability for all large j. In addition, for each i, $(1/T)\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'$ converges in probability to a nonrandom and positive definite matrix as $T \to \infty$.*

Assumption 14 *For any positive finite integer s, the matrices $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=k_0-s+1}^{k_0} \tilde{x}_{it}\tilde{x}_{it}'$ and $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k_0+s} \tilde{x}_{it}\tilde{x}_{it}'$, $i = 1, ..., N$, are stochastically bounded, with minimum eigenvalues bounded away from zero in probability for large N. In addition, for each t, $(1/N)\sum_{i=1}^{N} \tilde{x}_{it}\tilde{x}_{it}'$ is stochastically bounded as $N \to \infty$.*

Assumption 15 *$\{\delta_i, i = 1, ..., N\}$ are drawn independently of the process of $\{\tilde{x}_{it}, i = 1, ..., N\}$.*

Alternatively, under a random coefficient model, we have a slightly different version of Assumption 7.

Assumption 16 *For $i = 1, ..., N$,*

$$b_i = b + v_{b,i}, v_{b,i} \sim IID(0, \Sigma_b),$$

where $b = (\beta', \delta')'$, $v_{b,i} = \begin{pmatrix} v_{\beta,i} \\ v_{\delta,i} \end{pmatrix}$ and $\Sigma_b = \begin{pmatrix} \Sigma_\beta & 0 \\ 0 & \Sigma_\delta \end{pmatrix}$ for $i = 1, 2, ..., N$, where $\|b\| < \infty$, $\|\Sigma_b\| < \infty$, and the random deviations $v_{b,i}$ are independent of γ_j, Γ_j, ε_{jt}, and v_{jt} for all i, j and t.

Under Assumption 16, b_i is independent of Γ_j, implying that as $N \to \infty$, $\bar{C}_1 = \sum_{i=1}^N \theta_i C_{1i} \xrightarrow{p} E(C_{1i}) = (\gamma + \Gamma\beta, \Gamma)$ and $\bar{C}_2 \xrightarrow{p} E(C_{2i}) = (\gamma + \Gamma(\beta + R\delta), \Gamma)$. In this case, rank condition Assumption 11 requires non-zero means for γ and Γ in Assumption 10 when N is large. Similarly in Model 1, When $\{\delta_i, i = 1, ..., N\}$ are considered as random, as part of Assumption 16, Assumption 15 becomes redundant.

After the transformation (20), it can be shown that the change point estimator $\tilde{k}$ is still consistent in a linear model with a multifactor error structure (5), i.e., $\tilde{k} - k_0 = o_p(1)$.

Theorem 2 *Under Assumptions 1-3, 8-15 (or 16),* $\lim_{(N,T)\to\infty} P(\tilde{k} = k_0) = 1$.

Theorem 2 can be proved similarly to Theorem 1, see the Appendix.

Given the change point estimator $\tilde{k}$, the CCE estimator of the slope coefficients can be written as

$$\tilde{b}_i = \tilde{b}_i(\tilde{k}) = [\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{Y}_i, i = 1, ..., N.$$

Similar to Proposition 3, with the consistency of $\tilde{k}$, the asymptotics of $\tilde{b}_i$ can be established.

Proposition 1 *Under Assumptions 1-3, 8-15, and $\sqrt{T}/N \to 0$ as $(N, T) \to \infty$, for each i,*

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

where

$$\Sigma_{\tilde{\mathbb{X}},i} = plim_{T\to\infty}\frac{1}{T}\tilde{\mathbb{X}}_{0i}'\tilde{\mathbb{X}}_{0i} \text{ and } \Sigma_{\tilde{\mathbb{X}}\tilde{\varepsilon},i} = plim_{T\to\infty}\frac{1}{T}\tilde{\mathbb{X}}_{0i}'\Sigma_{\varepsilon,i}\tilde{\mathbb{X}}_{0i}, \ i = 1, ..., N.$$

An additional condition $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$ is required here, due to the fact that $M_w F\gamma_i$ is included in $\tilde{\varepsilon}_i^0 = M_w F\gamma_i + \tilde{\varepsilon}_i$, the error term of transformed model (21) using cross-sectional averages. This yields an extra term in $\sqrt{T}(\tilde{b}_i - b_i)$ whose order is $O_p(\sqrt{T}/N)+O_p(1/\sqrt{N})$ which is asymptotically ignorable when $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$. See the Supplementary Appendix.

As discussed above, Assumption 2 allows that T can grow faster than N, i.e., $T = O(N^{\psi})$ with $\psi \geq 1$. Here, the relative speed of N and T, $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$ imposes an upper bound on ψ, i.e., $\psi < 2$. Therefore, in the case of $T = O(N^{\psi})$ with $1 \leq \psi < 2$, both Assumption 2 and $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$ required by Proposition 1 are satisfied.

As discussed by Pesaran (2006), a consistent Newey-West type estimator of $\Sigma_{\tilde{\mathbb{X}}\tilde{\varepsilon},i}$ can be obtained using the transformed data,

$$\tilde{\Sigma}_{\tilde{\mathbb{X}}\tilde{\varepsilon},i} = \tilde{\Lambda}_{i0} + \sum_{j=1}^{\omega}(1 - \frac{j}{\omega+1})(\tilde{\Lambda}_{ij} + \tilde{\Lambda}'_{ij}),\ \tilde{\Lambda}_{ij} = \frac{1}{T}\sum_{t=j+1}^{\omega} \tilde{e}_{it}\tilde{e}_{i,t-j}\mathbb{X}_{it}(\hat{k})\mathbb{X}_{it}(\hat{k})',$$

where ω is the window size. $\tilde{e}_{it}$ is the t^{th} element of $\tilde{e}_i = \tilde{Y}_i - \tilde{\mathbb{X}}_i(\tilde{k})\tilde{b}_i$ and $\tilde{\mathbb{X}}_{it}(\tilde{k})$ is the t^{th} row of $\tilde{\mathbb{X}}_i(\tilde{k})$. Since $\Sigma_{\tilde{\mathbb{X}},i}$ can be consistently estimated by $\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})$. Thus, a consistent estimator of $\Sigma_{\tilde{\mathbb{X}},i}^{-1}\Sigma_{\tilde{\mathbb{X}}\tilde{\varepsilon},i}\Sigma_{\tilde{\mathbb{X}},i}^{-1}$ is given by

$$\left[\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})\right]^{-1} \tilde{\Sigma}_{\tilde{\mathbb{X}}\tilde{\varepsilon},i} \left[\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})\right]^{-1}. \tag{25}$$

Since $\tilde{b}_i(\tilde{k})$ has the same limiting distribution as $\tilde{b}_i(k_0)$, parameters b_i, $i = 1, ..., N$, in model (5) can be inferred as if k_0 were known.

The mean group estimator with a common break can be defined similarly:

$$\tilde{b}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\tilde{b}_i = \frac{1}{N}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{Y}_i. \tag{26}$$

Proposition 2 *Under the assumptions 1-3, 8-14, 16,*

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

As in Pesaran (2006), Σ_b can be consistently estimated by

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

For detailed proofs of Propositions 1 and 2, see the Supplementary Appendix. Unlike Pesaran (2006), an additional step is needed, that of estimating k_0. As shown in the propositions above, with the consistency of $\tilde{k}$, the convergence rate of $\tilde{k}$ is not required for deriving the asymptotic distributions of $\tilde{b}_i$, for $i = 1, ..., N$, and $\tilde{b}_{MG}$.

5 Multiple Common Break Points

When multiple common break points, $k_0^{(1)}, ..., k_0^{(B_k)}$, occur in the slopes, there are $B_k + 1$ regimes for each individual:

$$y_{it} = \begin{cases} x_{it}'\beta_i + e_{it}, & t = 1, ..., k_0^{(1)}, \\ x_{it}'\beta_i + z_{it}'\delta_{1i} + e_{it}, & t = k_0^{(1)} + 1, ..., k_0^{(2)}, \\ \vdots & \vdots \\ x_{it}'\beta_i + z_{it}'\delta_{B_k,i} + e_{it}, & t = k_0^{(B_k)} + 1, ..., T, \end{cases} \tag{27}$$

for $i = 1, ..., N$.

Estimation of multiple break points has been discussed by Bai (1997b) and Chong (1995) in a mean-shift model, Bai and Perron (1998) in linear regression models and Bai (2010) in a panel mean-shift model. To deal with this issue in model (27), we can follow the sequential or one at-a-time approach discussed by Bai (1997b, 2010). The number of common breaks, B_k, is assumed *known*.[14] The idea of the sequential approach is to estimate break points one by one. For example, if $B_k = 3$, the estimation of $k_0^{(1)}, k_0^{(2)}$ and $k_0^{(3)}$ can be completed in 3 steps. In the first step, one break point is assumed as in Model 1 (or Model 2) above, and can be estimated by (11) (or (24)), denoted by $\hat{k}^{(1)}$ (or $\tilde{k}^{(1)}$). In the second step, in each of the two sub-panels split by $\hat{k}^{(1)}$ (or $\tilde{k}^{(1)}$), the same procedure (11) (or (24)) is applied. Thus, two single break estimators are obtained in these two sub-panels. $\hat{k}^{(2)}$ (or $\tilde{k}^{(2)}$) is defined as the one associated with a larger reduction in the sum of squared residuals. Similarly, $\hat{k}^{(1)}$ and $\hat{k}^{(2)}$ (or $\tilde{k}^{(1)}$ and $\tilde{k}^{(2)}$) yield 3 sub-panels. In the third step, in each of these 3 sub-panels, one break point can be estimated as in Section 3 (or 4). Among these 3 break estimators, we choose the one associated with the largest reduction of sum of squared residuals, denoted as $\hat{k}^{(3)}$ (or $\tilde{k}^{(3)}$). As suggested by Bai (2010), it can be shown that after rearranging $(\hat{k}^{(1)}, \hat{k}^{(2)}, \hat{k}^{(3)})$ (or

[14] In a time series regression model, a sup $F_T(l+1|l)$ test is proposed by Bai and Perron (1998) to determine the number of structure breaks. Bai and Perron (2003) report the simulation results of this test and compare it with other tests based on information criteria. A panel version of the sup $F_T(l+1|l)$ test can be applied to determine the number of common breaks in our setup.

$(\tilde{k}^{(1)}, \tilde{k}^{(2)}, \tilde{k}^{(3)}))$ in temporal order, $(\hat{k}^{(1)}, \hat{k}^{(2)}, \hat{k}^{(3)})$ (or $(\tilde{k}^{(1)}, \tilde{k}^{(2)}, \tilde{k}^{(3)})$ in Model 2) is consistent for $(k_0^{(1)}, k_0^{(2)}, k_0^{(3)})$ as long as the assumptions listed in Section 3 (or 4) hold in each of the sub-panels.

Once the consistent estimators of $(k_0^{(1)}, ..., k_0^{(B_k)})$ are obtained, the parameters $\beta_i, \delta_{1i}, ..., \delta_{B_k,i}$, $i = 1, ..., N$, can be estimated by least squares as in (10) (or (23)). Thus, their mean-group estimators can be obtained similarly.

6 Monte Carlo Simulations

This section employs Monte Carlo simulations to examine the consistency of the estimated break points $\hat{k}$ and $\tilde{k}$ summarized in Theorems 1 and 2. Since the CCE estimators in Model 2 have the same asymptotic distributions as if the true common breaks were known, their asymptotic properties are not examined here. Two different designs are used for Models 1 and 2, respectively. In Model 1, there are no common correlated effects in the errors and regressors, so least squares can be run for each individual series. While in Model 2, the regressors and errors are correlated due to common correlated effects f_t. A transformation, using cross-sectional averages of the dependent variable and regressors proposed by Pesaran (2006), is needed to remove such effects asymptotically.

In the following experiments, the focus is on the histograms of $\hat{k}$ and $\tilde{k}$ in setups with different combinations of (N, T).

6.1 Model 1: No common correlated effects

The data generating process of Model 1 is modified from that in Pesaran (2004, p.24):

$$\begin{aligned} y_{it} &= \alpha_i + \beta_i(k_0) y_{i,t-1} + e_{it}, i = 1, ..., N; t = 1, ..., T; \\ e_{it} &= \gamma_i f_t + \varepsilon_{it}. \end{aligned}$$

Here we set $\gamma_i = 0$, so there is no cross-sectional dependence in the errors. Instead, in this dynamic heterogeneous panel model, there is a common break $k_0 = 0.5T$ in the slopes β_i, for $i = 1, ..., N$, i.e.,

$$\beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, ..., k_0, \\ \beta_{2i} = \beta_{1i} + \delta_i, & t = k_0 + 1, ..., T, \end{cases}$$

where δ_i is the jump in the slope for each series. We assume $\beta_{1i} \sim iidU(0, 0.8)$ and $\delta_i \sim iidU(0, 0.2)$. We set $\alpha_i = \mu_i(1-\beta_{1i})$, $\mu_i = \varepsilon_{0i}+\eta_i$ where $\varepsilon_{0i} \sim iidN(0,1)$ and $\eta_i \sim iidN(1,2)$. In addition, we assume $y_{i,0} \sim iidN(0,1)$ and $\varepsilon_{it} \sim iidN(0, \sigma_i^2)$, with $\sigma_i^2 \sim \chi_2^2/2$.

In (11), for any possible change point $k = 1, ..., T-1$, the estimated change point $\hat{k}$ is the one that minimizes the sum of N individual sum of squared residuals. 1000 replications are performed to obtain the histogram of $\hat{k}$ for each setup.

Figure 1 reports the histograms of $\hat{k}$ for $T = 20$ and $N = 1, 10, 50, 200$. It shows that the distribution of $\hat{k}$ shrinks with N. The frequency of choosing the true value k_0 increases from 8% to 58% when N increases from 1 to 200. In case $T = 50$, as Figure 2 shows, the frequency of choosing the true value k_0 improves to almost 90% for $N = 200$. This finding supports Theorem 1, confirming that multiple individual series provide additional information on k_0, and that $\hat{k}$ converges to k_0 as the number of series goes to infinity.

To consider the case where there is no structural break in slopes in some series, we set $\delta_i = 0$ in $[N/4]$ series, implying that ϕ_N increases with N at a rate of $O(N^{3/4})$. Figure 3 reports the histograms of $\hat{k}$ for this case with $T = 50$. Similar to Figure 2, the pattern that $\hat{k}$ converges to k_0 as N increases remains. However, the frequency of choosing the true value k_0 is significantly smaller than that in Figure 2. For example, for $N = 50$, the frequency of choosing the true value k_0 drops from 44% in Figure 2 to 34% in Figure 3. This suggests that for the accuracy of the estimated change point, allowing for no break in some series is equivalent to reducing the number of series or the magnitude of break ϕ_N.

6.2 Model 2: Common correlated effects

The data generating process for Model 2 is as follows:

$$\begin{aligned} y_{it} &= \alpha_i + \beta_i(k_0)x_{i,t} + e_{it}, i = 1, ..., N; t = 1, ..., T \\ e_{it} &= \gamma_{1i} f_t + \varepsilon_{it}, \end{aligned}$$

where $\alpha_i \sim iidN(1,1)$ and $\gamma_{1i} \sim iidN(1, 0.2)$. The idiosyncratic errors are generated as $\varepsilon_{it} \sim iidN(0, \sigma_i^2)$ and $\sigma_i^2 \sim iidU(0.5, 1.5)$. There is a common break in the individual slopes:

$$\beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, ..., k_0, \\ \beta_{2i} = \beta_{1i} + \delta_i, & t = k_0 + 1, ..., T, \end{cases} \quad k_0 = 0.5T,$$

where $\beta_{1i} = 1 + \eta_i, \eta_i \sim iidN(0, 0.04)$ and $\delta_i \sim iidN(0, 0.04)$.

Unlike Model 1, the error e_{it} and the regressor x_{it} contain the common correlated effect f_t:

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

where $a_i \sim iidN(0.5, 0.5)$, $\gamma_{2i} \sim iidN(0.5, 0.5)$ and $v_{it} \sim iidN(0, 1 - \rho_{vi}^2)$, with $\rho_{vi} = 0.5$. The factor f_t is generated by the stationary process:

$$\begin{aligned} f_t &= \rho_f f_{t-1} + v_{ft}, t = -49, ..., 0, 1, ..., T; \\ \rho_f &= 0.5, v_{ft} \sim iidN(0, 1 - \rho_f^2), f_{-50} = 0. \end{aligned}$$

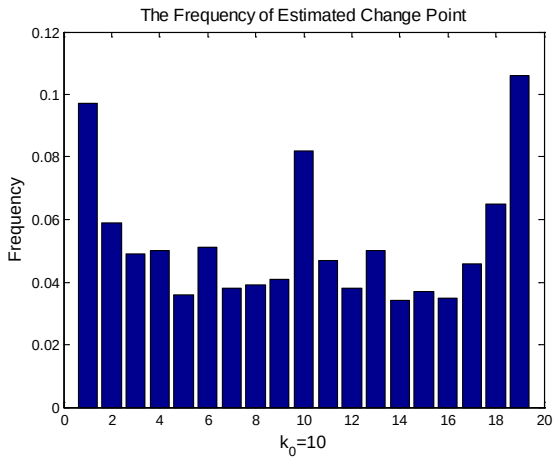
The correlation between x_{it} and e_{it} renders OLS inconsistent in the individual regressions. Thus, transformation (20) using cross-sectional averages of y_{it} and x_{it} is needed to remove f_t before conducting least squares estimation of k_0.

The setup above is a simplified version of the design in Pesaran (2006). First, as in model (4), the observed factors are omitted for simplicity. Second, the number of regressors and unobservable factors are reduced to 1, respectively. Third, the correlation structures in v_{it} and ε_{it} are removed. The only new feature of this model is that there is a common break at k_0, specified as $0.5T$.

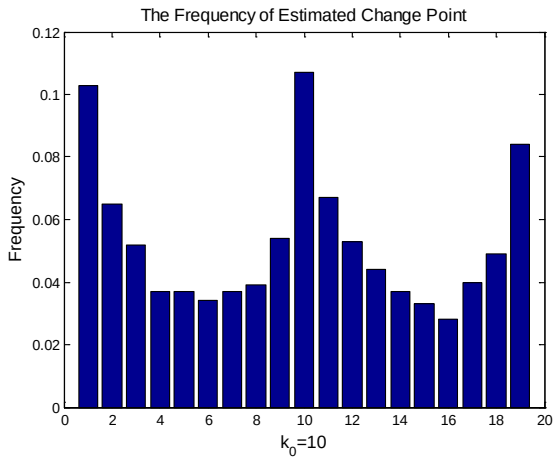
The first row of Figure 4 presents the histograms of the estimated change point $\tilde{k}$ for $T = 20$. It replicates the pattern in Figure 1, showing that after the transformation, the frequency of choosing the true value k_0 increases significantly with N. Figure 4 also reports, in the second row, the histograms of the estimated change point $\hat{k}$ without conducting transformation (20). It indicates that in the presence of common correlated effects, cross-sectional information using multiple series fails to improve the accuracy of the estimated change point.

Figures 5 and 6 report the histograms of $\tilde{k}$ and $\hat{k}$ for $T = 50$ and 200, respectively. The same pattern emerges, suggesting that the distribution of $\tilde{k}$ shrinks to k_0 as $N \to \infty$. Different from Figure 4, the frequency of $\hat{k}$, the estimator without conducting transformation (20), choosing the true break date increases with N in Figure 6 when T is large, although not at a rate as high as that of $\tilde{k}$ using the transformed data. Whether $|\hat{k} - k_0|$ shrinks to 0 or not as $(N, T) \to \infty$ depends upon the correlation between x_{it} and e_{it}. In Figure 7, we increase this correlation by changing the distribution of γ_{1i} from $N(1, 0.2)$ to $N(2, 0.2)$. In this case, the cross-sectional information using multiple series fails to improve the accuracy

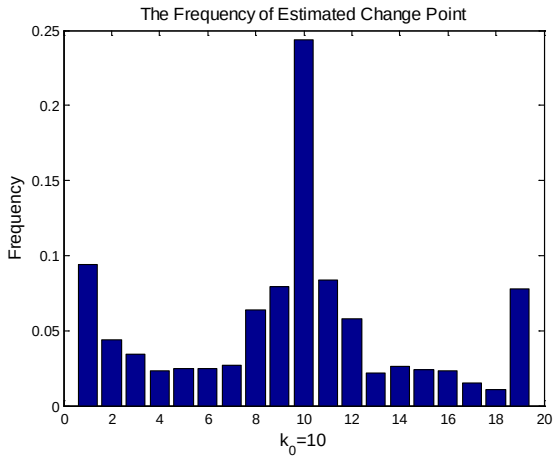
Figure 1: Histograms of $\hat{k}$ in Model 1: T =20



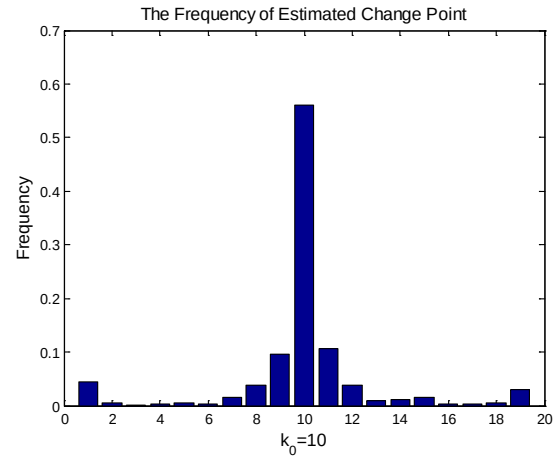
N =1



N =10



N =50



N =200

Note: The DGP is similar to Pesaran (2004, p.24):

$y_{it} = \alpha_i + \beta_i(k_0)y_{i,t-1} + e_{it}, i = 1, \dots, N; t = 1, \dots, T,$

where $\beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, \dots, k_0, \\ \beta_{2i} = \beta_{1i} + \delta_i, & t = k_0 + 1, \dots, T. \end{cases}$

$k_0 = 0.5T = 10,\ e_{it} = \gamma_i f_t + \varepsilon_{it},\ \gamma_i = 0$ and $\alpha_i = \mu_i(1 - \beta_{1i}),\ \mu_i = \varepsilon_{0i} + \eta_i.$

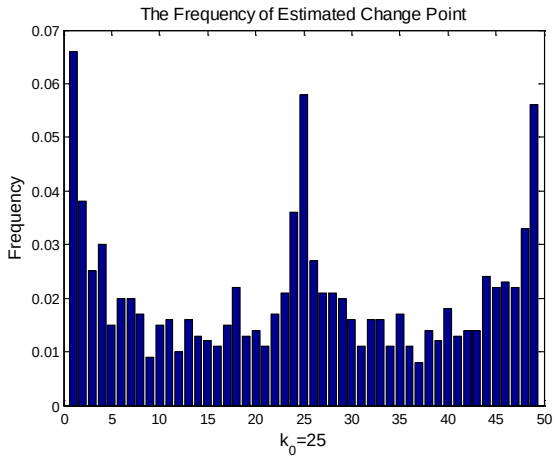
Values used: $\beta_{1i} \sim iidU(0, 0.8),\ \delta_i \sim iidU(0, 0.2),\ \varepsilon_{it} \sim iidN(0, \sigma_i^2),\ \sigma_i^2 \sim \chi_2^2/2,$

$\varepsilon_{0i} \sim iidN(0, 1)$ and $\eta_i \sim iidN(1, 2),\ y_{i,0} \sim iidN(0, 1),\ f_t \sim iidN(0, 1).$

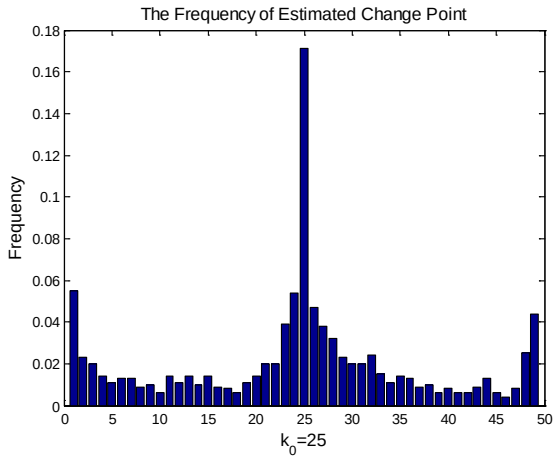
These variables are mutually independent.

The replication number is 1000.

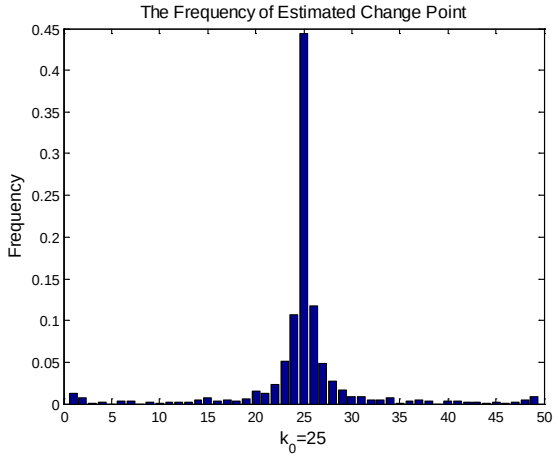
Figure 2: Histograms of $\hat{k}$ in Model 1: T =50



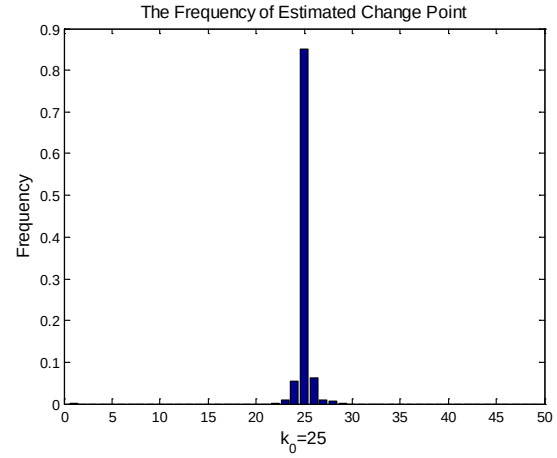
N =1



N =10



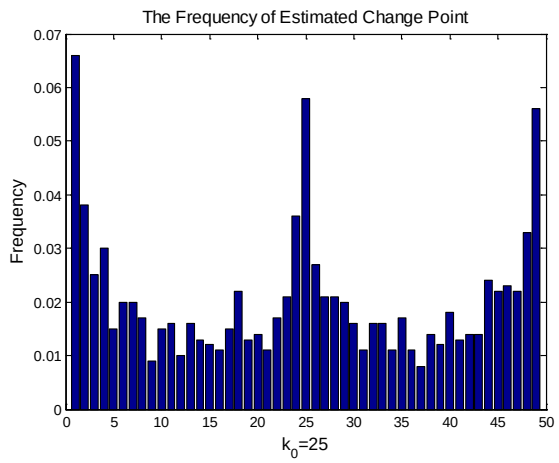
N =50



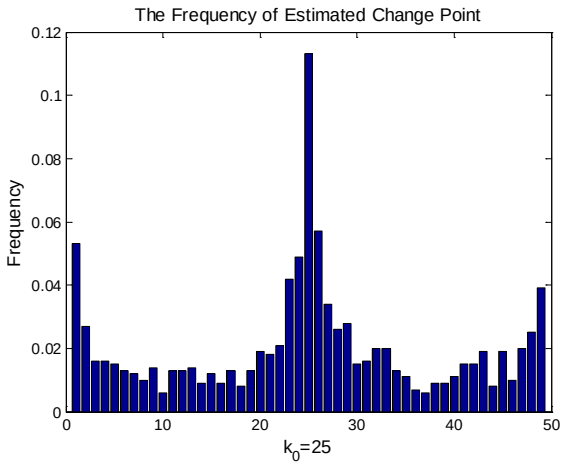
N =200

Note: The DGP is the same as in Figure 1. $k_0 = 0.5T = 25$.

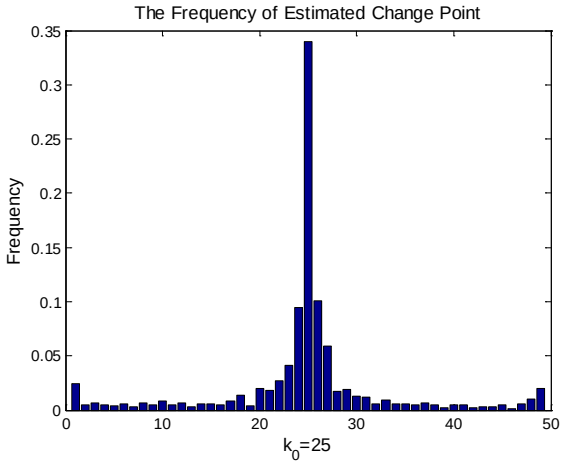
Figure 3: Histograms of $\hat{k}$ in Model 1 with No Break in Some Series: T =50



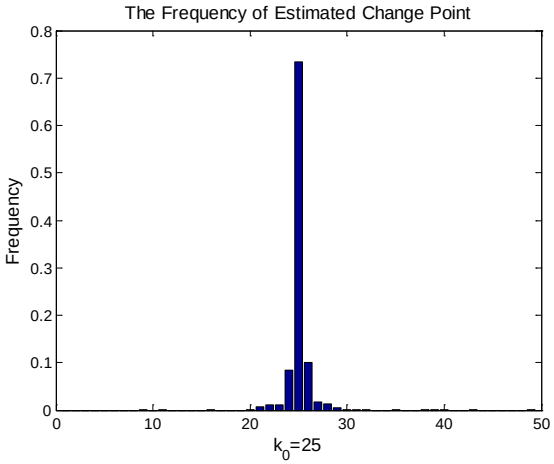
N =1



N =10



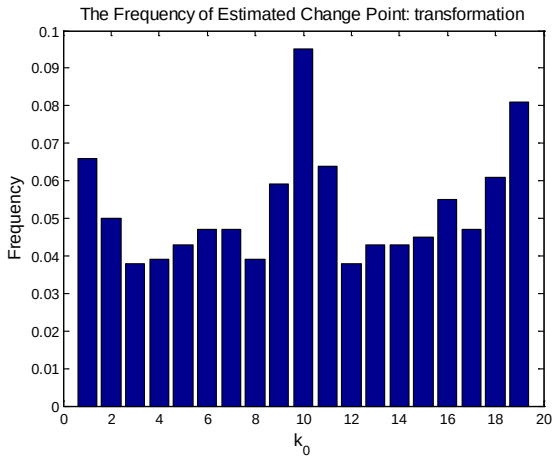
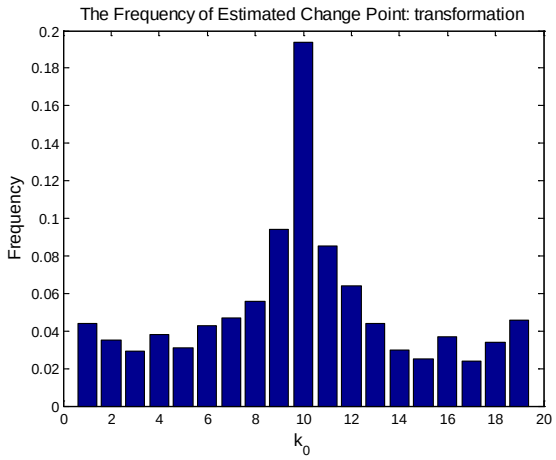
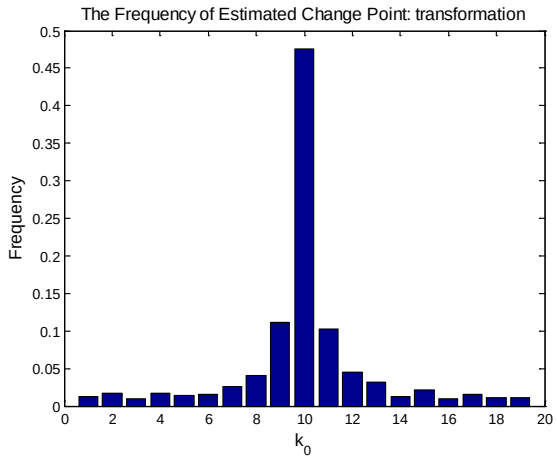
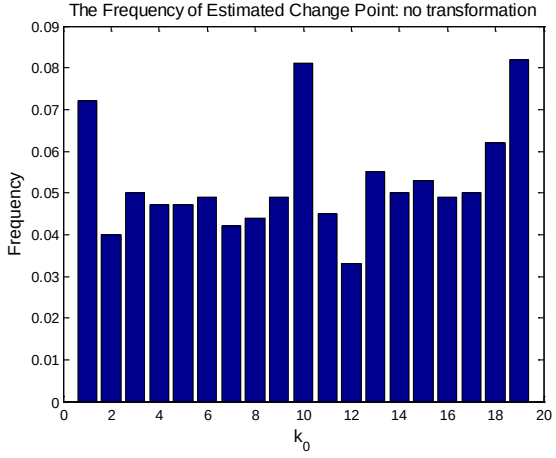
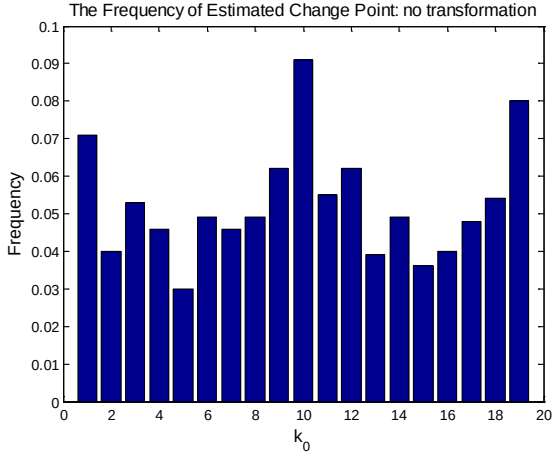
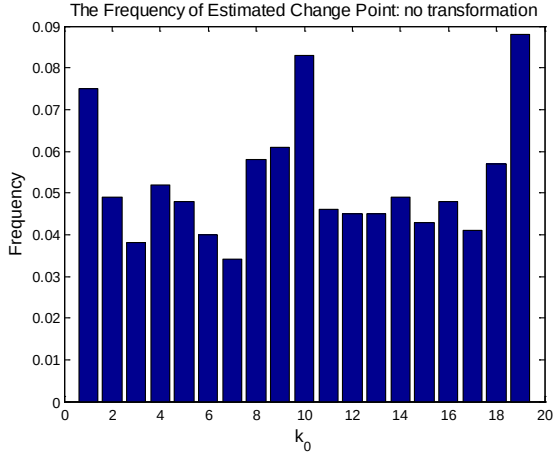
N =50



N =200

Note: The DGP is the same as in Figure 1 except that there is no break in [N/4] series. $k_0 = 0.5T = 25$.

Figure 4: Histograms of $\tilde{k}$ and $\hat{k}$ in Model 2: T =20

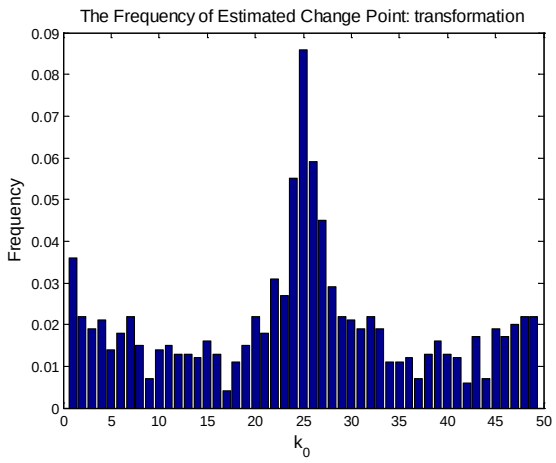
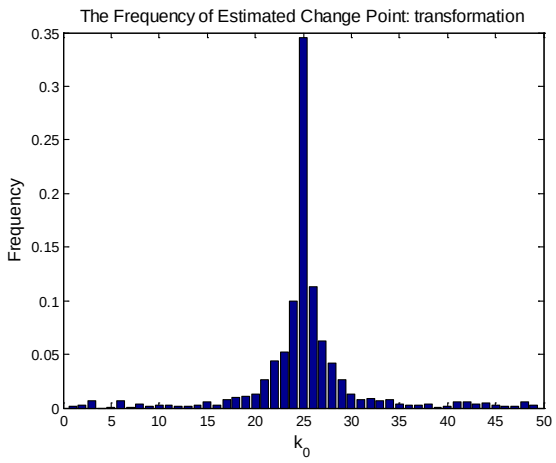
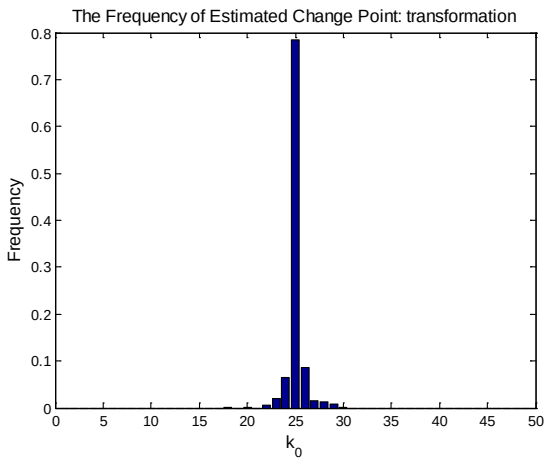
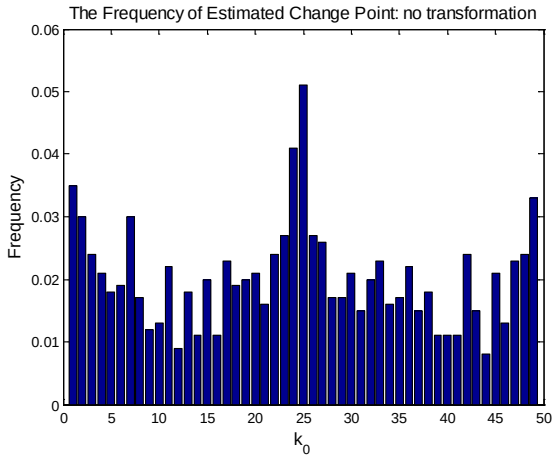
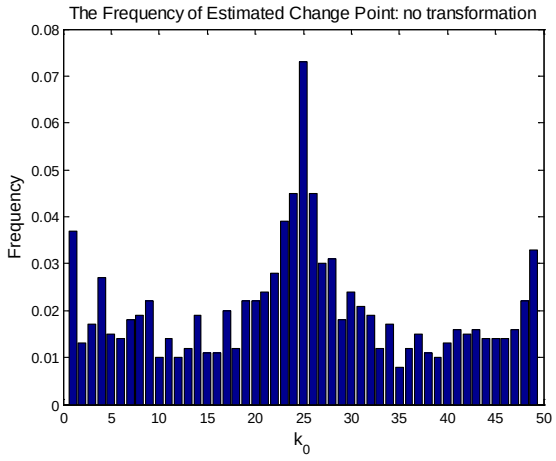
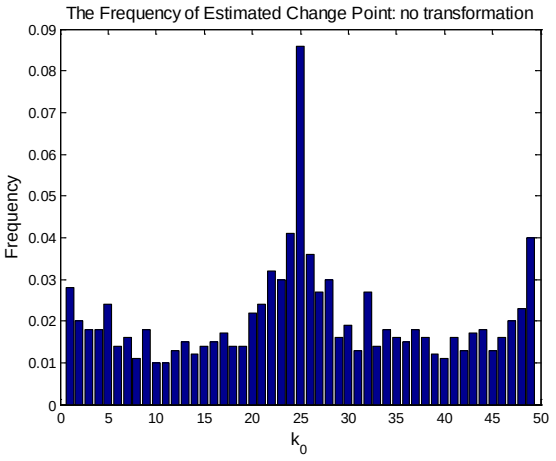
	N =10	N =50	N =200
$\tilde{k}$			
$\hat{k}$			

Note: The DGP is constructed by simplifying the design of Pesaran (2006):

$$y_{it} = \alpha_i + \beta_i(k_0)x_{i,t} + e_{it}, i = 1, \dots, N; t = 1, \dots, T.\ \alpha_i \sim iidN(1,1),\ \beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, \dots, k_0, \\ \beta_{2i} = \beta_{1i} + \delta_i, & t = k_0 + 1, \dots, T. \end{cases}\ k_0 = 0.5T = 10,\ \beta_{1i} \sim iidN(1, 0.04),\ \delta_i \sim iidN(0, 0.04).$$

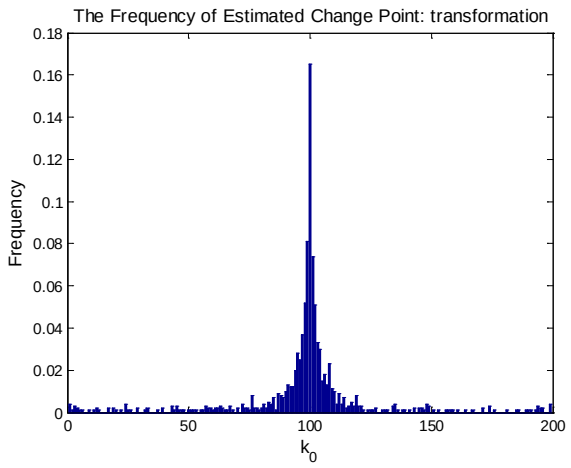
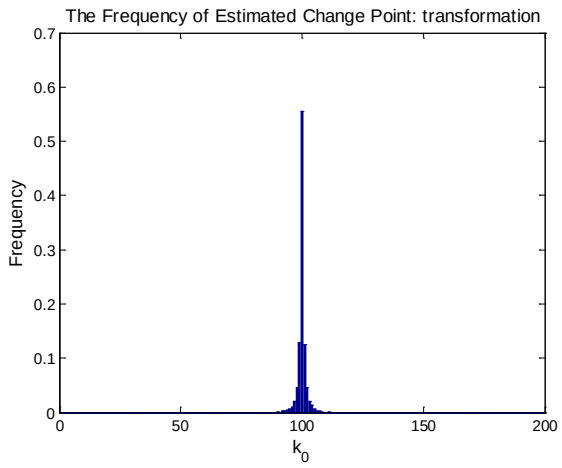
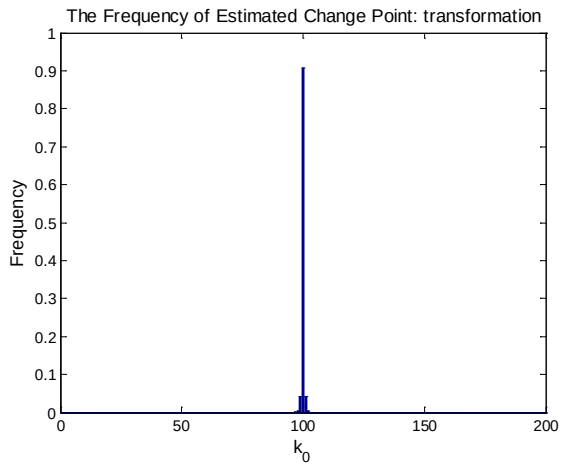
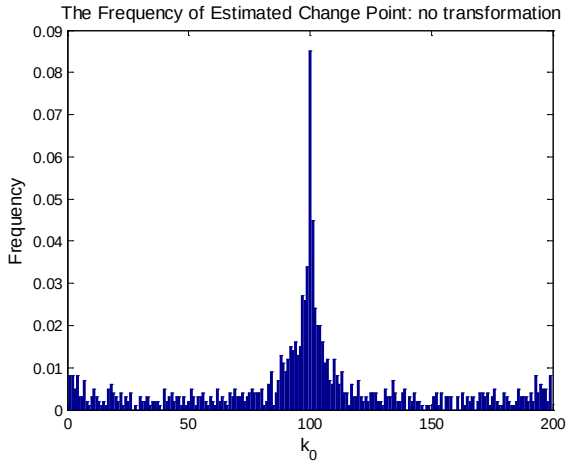
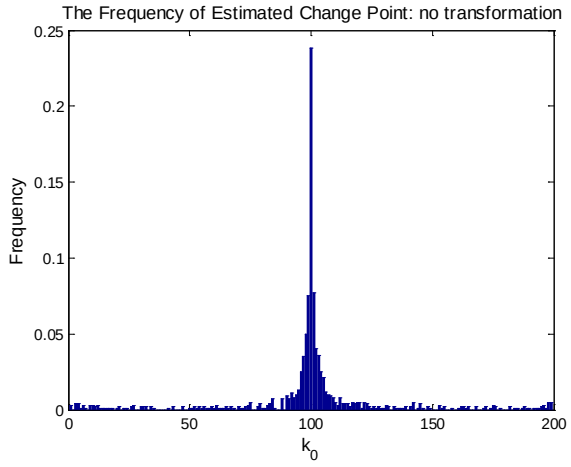
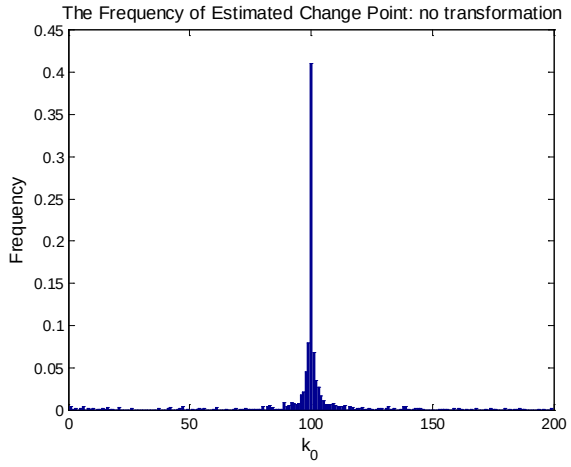
$e_{it} = \gamma_{i1}f_t + \varepsilon_{it},\ x_{i,t} = a_i + \gamma_{i2}f_t + v_{it}; f_t = \rho_f f_{t-1} + v_{ft}, t = -49, \dots, 0, 1, \dots T,\ v_{ft} \sim iidN(0, 1-\rho_f^2),\ \rho_f = 0.5,\ f_{-50} = 0.\ \varepsilon_{it} \sim iidN(0, \sigma_i^2),\ \sigma_i^2 \sim iidU(0.5, 1.5),$
$\gamma_{i1} \sim iidN(1, 0.2),\ \gamma_{i2} \sim iidN(0.5, 0.5),\ a_i \sim iidN(0.5, 0.5),\ v_{it} \sim iidN(0, 1-\rho_{vi}^2),\ \rho_{vi} = 0.5$. These variables are mutually independent. The replication number is 1000.

Figure 5: Histograms of $\tilde{k}$ and $\hat{k}$ in Model 2: T =50

	N =10	N =50	N =200
$\tilde{k}$			
$\hat{k}$			

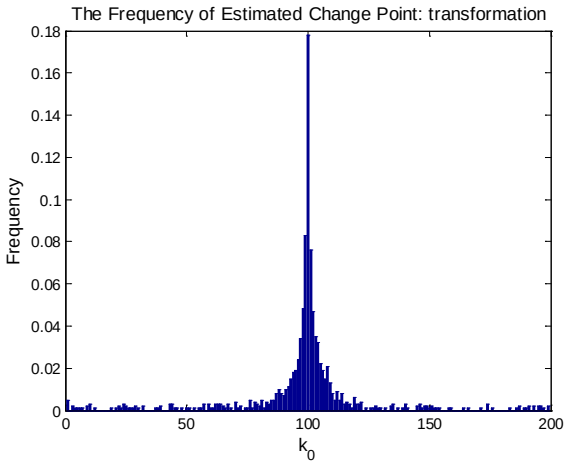
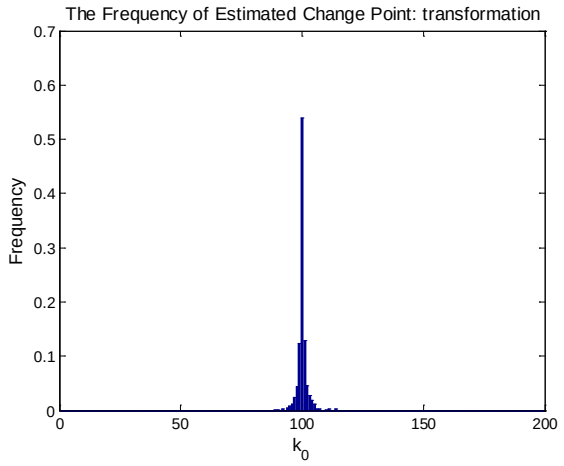
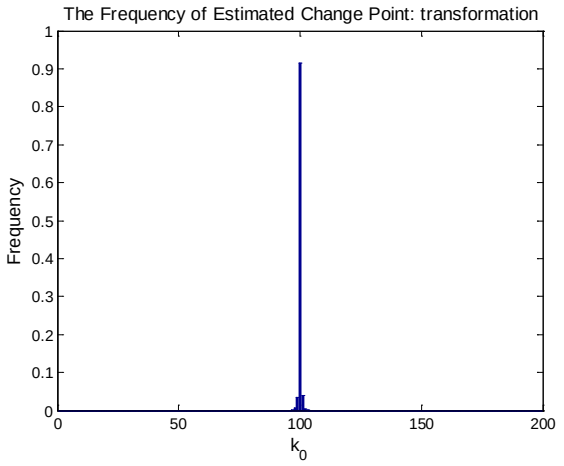
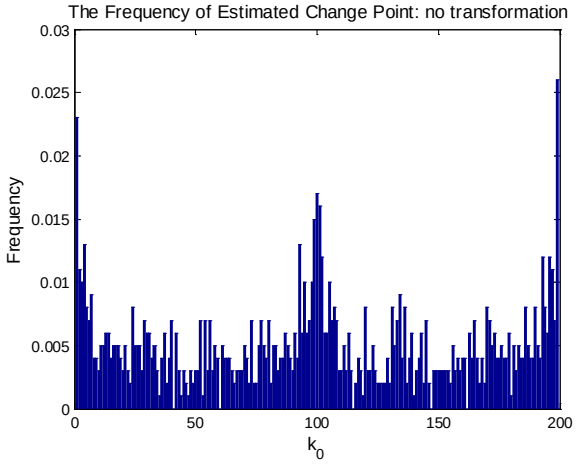
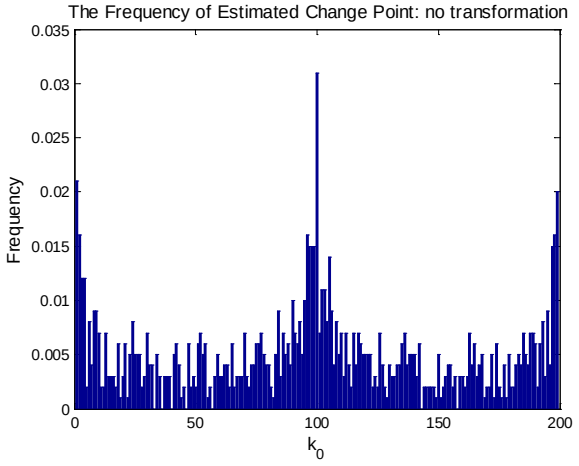
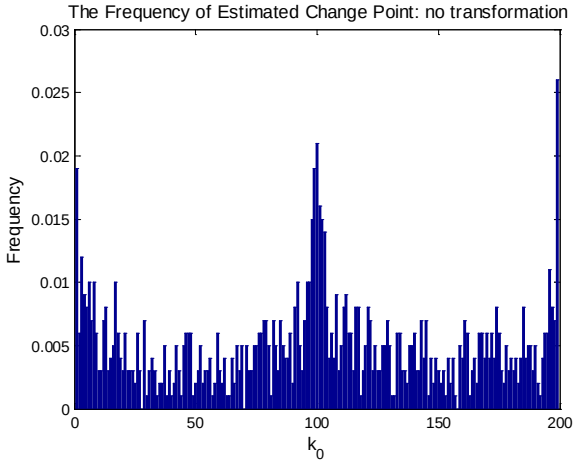
Note: The DGP is the same as in Figure 4. $k_0 = 0.5T = 25$.

Figure 6: Histograms of $\tilde{k}$ and $\hat{k}$ in Model 2: T =200

	N =10	N =50	N =200
$\tilde{k}$			
$\hat{k}$			

Note: The DGP is the same as in Figure 4. $k_0 = 0.5T = 100$.

Figure 7: Histograms of $\tilde{k}$ and $\hat{k}$ in Model 2 (with increased the correlation between x_{it} and e_{it}): T =200

	N =10	N =50	N =200
$\tilde{k}$			
$\hat{k}$			

Note: The DGP is the same as in Figure 4, except that the correlation between x_{it} and e_{it} increases by changing the distribution of γ_{i1} from $iidN(1, 0.2)$ to $iidN(2, 0.2)$.

$k_0 = 0.5T = 100$.

of the estimated change point $\hat{k}$. This is consistent with the findings of Kim (2011).[15]

7 Conclusion

The recent literature on panel data models with large time dimension assumes that the slopes are constant over time. However, due to global policy or technological shocks, slope parameters are likely to fall into different regimes over a longer time span. This paper tackles both structural breaks and cross-sectional dependence by extending Pesaran's (2006) framework of heterogeneous panels to the situation of unknown common breaks in the slopes. The least squares method proposed by Bai (1997a, 2010) is applied to estimate the common change points. This method is married to the CCE estimators proposed by Pesaran (2006) in this setup. Different from the time series change point models, in which the estimated change point is inconsistent, our paper establishes its consistency in a general panel data setup. The properties of the CCE estimators of the slopes at estimated points are also examined. We find that the CCE estimators have the same asymptotic distributions as if the true change points were known.[16]

In this paper, we assume that the rank condition is satisfied. In case of deficient rank, Westerlund and Urbain (2013) show that the CCE estimators could be inconsistent when factor loadings in the error term and in the explanatory variables are correlated. In this case, the iterative principal component approach, proposed by Bai (2009) in homogeneous panels and extended by Song (2012) to dynamic heterogeneous panels, can be an alternative, which allows for correlated and zero-mean factor loadings. However, the issue of structural breaks in this setup is beyond the scope of this paper.

[15] Experiments with different values of parameters and distributions are conducted and similar histograms are obtained. Results are not included here to save space.

[16] Since the convergence rate of the change point estimator is not required to derive the asymptotic distributions of the CCE estimators of slopes and their cross-sectional mean, we leave the derivations of convergence rate and asymptotic distribution of the change point estimator in heterogeneous panels for future research.

Appendix: Mathematical Proofs

Since the panel data model (6) considered here includes the time series model in Bai (1997a) as a special case of $N = 1$, it can be shown similarly that $\hat{k} - k_0 = O_p(1)$. In the proofs that follow, we assume $\hat{k} - k_0$ is stochastically bounded. With more information along the cross-sectional dimension under the common break assumption, we further show that $\hat{k} - k_0 \xrightarrow{p} 0$ as $(N, T) \to \infty$.

For $i = 1, ..., N$, let SSR_i be the sum of squared residuals of regressing Y_i on X_i in case there is no break, i.e., $Z_{2i}(k) = 0_{T\times q}$. Using the identity

$$\begin{aligned} SSR_i - SSR_i(k) &= [Y_i - X_i\hat{\beta}_i(k) - Z_{2i}(k)\hat{\delta}_i(k)]'[Y_i - X_i\hat{\beta}_i(k) - Z_{2i}(k)\hat{\delta}_i(k)] \\ &\quad -[Y_i - X_i\hat{\beta}_i(k)]'[Y_i - X_i\hat{\beta}_i(k)] \\ &= \hat{\delta}_i(k)'[Z_{2i}(k)'M_iZ_{2i}(k)]\hat{\delta}_i(k) \end{aligned}$$

with $M_i = I - X_i(X_i'X_i)^{-1}X_i'$,

$$\hat{k} = \arg\min_{1\le k\le T-1} \sum_{i=1}^{N} \pi_i SSR_i(k) = \arg\max_{1\le k\le T-1} \sum_{i=1}^{N} \pi_i SV_i(k) = \arg\max_{1\le k\le T-1} \sum_{i=1}^{N} \pi_i [SV_i(k) - SV_i(k_0)],$$

where $SV_i(k) = \hat{\delta}_i(k)'[Z_{2i}(k)'M_iZ_{2i}(k)]\hat{\delta}_i(k)$. Note that $SV_i(k_0) = \hat{\delta}_i(k_0)'\left[Z_{0i}'M_iZ_{0i}\right]\hat{\delta}_i(k_0)$ is not a function of k. For simplicity, we assume $\pi_i = 1/N$, $i = 1, ..., N$.[17]

To prove Theorem 1, $\sum_{i=1}^{N}[SV_i(k) - SV_i(k_0)]$ can be decomposed into a deterministic part and a stochastic one. Partitioned regression gives

$$\hat{\delta}_i(k) = \left[Z_{2i}(k)'M_iZ_{2i}(k)\right]^{-1} Z_{2i}(k)'M_iY_i,\ i = 1, ..., N.$$

Substituting $Y_i = X_i\beta_i + Z_{0i}\delta_i + \varepsilon_i$ into the equation above, we obtain

$$\hat{\delta}_i(k) = \left[Z_{2i}(k)'M_iZ_{2i}(k)\right]^{-1} Z_{2i}(k)'M_iZ_{0i}\delta_i + \left[Z_{2i}(k)'M_iZ_{2i}(k)\right]^{-1} Z_{2i}(k)'M_i\varepsilon_i$$

and $\hat{\delta}_i(k_0) = \delta_i + \left(Z_{0i}'M_iZ_{0i}\right)^{-1} Z_{0i}'M_i\varepsilon_i$.

[17] Since the weights π_i, $i = 1, ..., N$, are used to balance different variances across series, we can always employ $\pi_i = 1/N$, $i = 1, ..., N$, by using different notations for X_i and σ_i^2.

To simplify notation, k is suppressed in $\hat{\delta}_i(k)$ and $Z_{2i}(k)$ when no confusion arises. Since

$$SV_i(k) = \hat{\delta}_i'(Z_{2i}'M_iZ_{2i})\hat{\delta}_i = \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})\delta_i$$
$$+ 2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i + \varepsilon_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i,$$

it follows that

$$\begin{aligned} SV_i(k) - SV_i(k_0) &= -\delta_i'[(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})]\delta_i & (28)\\ &\quad +2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - 2\delta_i'Z_{0i}'M_i\varepsilon_i & (29)\\ &\quad +\varepsilon_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - \varepsilon_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_i\varepsilon_i & (30)\end{aligned}$$

The deterministic part is denoted by

$$J_{1i}(k) = \delta_i'[(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})]\delta_i, \quad (31)$$

and the stochastic part is denoted by

$$\begin{aligned} J_{2i}(k) &= 2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - 2\delta_i'Z_{0i}'M_i\varepsilon_i\\ &\quad +\varepsilon_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - \varepsilon_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_i\varepsilon_i.\end{aligned}$$

Thus $SV_i(k) - SV_i(k_0) = -J_{1i}(k) + J_{2i}(k)$ and

$$\hat{k} = \arg\max_{1\le k\le T-1}\sum_{i=1}^{N}[SV_i(k) - SV_i(k_0)] = \arg\max_{1\le k\le T-1}\left[-\sum_{i=1}^{N}J_{1i}(k) + \sum_{i=1}^{N}J_{2i}(k)\right].$$

Define

$$\begin{aligned} X_{\Delta i} &= X_{2i} - X_{0i} = (0,\cdots,0,x_{i,k+1},\cdots,x_{i,k_0},0,\cdots,0)', && \text{for } k < k_0,\\ X_{\Delta i} &= -(X_{2i} - X_{0i}) = (0,\cdots,0,x_{i,k_0+1},\cdots,x_{i,k},0,\cdots,0)', && \text{for } k \ge k_0.\end{aligned}$$

$Z_{\Delta i}$ can be defined similarly.

For a finite large number C_k and arbitrarily small positive number $a < \tau_0$, define the set $K(C_k) = \{k : 1 \le |k - k_0| < C_k, aT < k < (1-a)T\}$. Since $\hat{k} - k_0$ is stochastically bounded, we only consider the values of k that belong to set $K(C_k)$.

Let $\lambda_1(k)$ be the minimum eigenvalue of $\frac{1}{N}\sum_{i=1}^{N}R'(X_{\Delta i}'X_{\Delta i})R$. Define $\lambda_1 = \min_{k\in K(C_k)}\lambda_1(k)$. Under Assumption 5, $\lambda_1(k) > 0$ and $\lambda_1 > 0$.

Lemma 1 *Under Assumptions 1-7, for all large N and T, with probability tending to 1,*

$$\inf_{K(C_k)} \sum_{i=1}^{N} J_{1i}(k) \geq \lambda_1 \phi_N.$$

This lemma is similar to Lemma A.2 in Bai (1997a). The proof can be found in the Supplementary Appendix.

Lemma 2 *Under Assumptions 1-7, uniformly on $K(C_k)$,*

(i) $\sum_{i=1}^{N} \delta_i' Z_{\Delta i}' \varepsilon_i = O_p(\sqrt{\phi_N})$;
(ii) $\frac{1}{\sqrt{T}} \sum_{i=1}^{N} \delta_i' Z_{\Delta i}' X_i (\frac{X_i' X_i}{T})^{-1} \frac{X_i' \varepsilon_i}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{T}})$;
(iii) $\frac{1}{\sqrt{T}} \sum_{i=1}^{N} \delta_i' (Z_{\Delta i}' M_i Z_{2i}) (\frac{Z_{2i}' M_i Z_{2i}}{T})^{-1} \frac{Z_{2i}' M_i \varepsilon_i}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{T}})$;
(iv) $\frac{1}{T} \sum_{i=1}^{N} \varepsilon_i' M_i Z_{\Delta i} (\frac{Z_{2i}' M_i Z_{2i}}{T})^{-1} Z_{\Delta i}' M_i \varepsilon_i = O_p(\frac{N}{T})$;
(v) $\frac{1}{T} \sum_{i=1}^{N} \varepsilon_i' M_i Z_{0i} (\frac{Z_{2i}' M_i Z_{2i}}{T})^{-1} Z_{\Delta i}' M_i \varepsilon_i = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}})$;
(vi) $\sum_{i=1}^{N} \frac{\varepsilon_i' M_i Z_{0i}}{\sqrt{T}} \left[(\frac{Z_{2i}' M_i Z_{2i}}{T})^{-1} - (\frac{Z_{0i}' M_i Z_{0i}}{T})^{-1} \right] \frac{Z_{0i}' M_i \varepsilon_i}{\sqrt{T}} = O_p(\frac{N}{T})$.

Proof of Lemma 2. (i) Under Assumption 3, for large N,

$$Var\left(\sum_{i=1}^{N} \delta_i' Z_{\Delta i}' \varepsilon_i \right) = \sum_{i=1}^{N} \delta_i' Z_{\Delta i}' \Sigma_{\varepsilon,i} Z_{\Delta i} \delta_i.$$

It can be shown equal to $O(\phi_N)$ under Assumptions 4 to 7, similar to the proof of Lemma 1 in the Supplementary Appendix, implying $\sum_{i=1}^{N} \delta_i' Z_{\Delta i}' \varepsilon_i = O_p(\sqrt{\phi_N})$ on $K(C_k)$.

The proofs of Lemma 2(ii)-(vi) are included in the Supplementary Appendix.

With these lemmas, we are ready to prove Theorem 1.

Proof of Theorem 1. To prove $\lim_{(N,T)\to\infty} P(\hat{k} = k_0) = 1$, it is equivalent to showing that for any given $\epsilon > 0$, for both large T and N, $P(|\hat{k} - k_0| \geq 1) < \epsilon$. It is sufficient to show that $P\left(\sup_{K(C_k)} \sum_{i=1}^{N} [SV_i(k) - SV_i(k_0)] \geq 0 \right) < \epsilon$, or

$$P\left(\sup_{K(C_k)} \left| \sum_{i=1}^{N} J_{2i}(k) \right| \geq \inf_{K(C_k)} \sum_{i=1}^{N} J_{1i}(k) \right) < \epsilon.$$

By Lemma 1, it suffices to show $P(\sup_{K(C_k)} \frac{1}{\phi_N} \left| \sum_{i=1}^{N} J_{2i}(k) \right| \geq \lambda_1) < \epsilon$.For any $k \in K(C_k)$,

$$\left|\sum_{i=1}^{N} J_{2i}(k)\right| \leq \left|\sum_{i=1}^{N}\left[2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - 2\delta_i'Z_{0i}'M_i\varepsilon_i\right]\right| + \left|\sum_{i=1}^{N}\left[\varepsilon_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - \varepsilon_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_i\varepsilon_i\right]\right|.$$

Consider the first term, $Z_{2i} = Z_{0i} + Z_{\Delta i}$ for $k < k_0$,

$$\begin{aligned}
&\left|\sum_{i=1}^{N}\left[2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - 2\delta_i'Z_{0i}'M_i\varepsilon_i\right]\right| \\
&= \left|\sum_{i=1}^{N}\left[2\delta_i'Z_{\Delta i}'M_i\varepsilon_i - 2\delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i\right]\right| \\
&\leq 2\left|\sum_{i=1}^{N}\delta_i'Z_{\Delta i}'\varepsilon_i\right| + \frac{2}{\sqrt{T}}\left|\sum_{i=1}^{N}\delta_i'Z_{\Delta i}'X_i(\frac{X_i'X_i}{T})^{-1}\frac{X_i'\varepsilon_i}{\sqrt{T}}\right| \\
&+ \frac{2}{\sqrt{T}}\left|\sum_{i=1}^{N}\delta_i'\left[(Z_{\Delta i}'M_iZ_{2i})(\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1}\frac{Z_{2i}'M_i\varepsilon_i}{\sqrt{T}}\right]\right|.
\end{aligned}$$

By (i), (ii) and (iii) of Lemma 2, the first term

$$\left|\sum_{i=1}^{N}\left[2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - 2\delta_i'Z_{0i}'M_i\varepsilon_i\right]\right| = O_p(\sqrt{\phi_N}). \tag{32}$$

Now consider the second term

$$\begin{aligned}
&\left|\sum_{i=1}^{N}\left[\varepsilon_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - \varepsilon_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_i\varepsilon_i\right]\right| \\
&\leq \frac{1}{T}\left|\sum_{i=1}^{N}\varepsilon_i'M_iZ_{\Delta i}(\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1}Z_{\Delta i}'M_i\varepsilon_i\right| + 2\frac{1}{\sqrt{T}}\left|\sum_{i=1}^{N}\frac{\varepsilon_i'M_iZ_{0i}}{\sqrt{T}}(\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1}Z_{\Delta i}'M_i\varepsilon_i\right| \\
&+ \left|\sum_{i=1}^{N}\frac{\varepsilon_i'M_iZ_{0i}}{\sqrt{T}}\left[(\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} - (\frac{Z_{0i}'M_iZ_{0i}}{T})^{-1}\right]\frac{Z_{0i}'M_i\varepsilon_i}{\sqrt{T}}\right|.
\end{aligned}$$

Similarly, by (iv), (v) and (vi) of Lemma 2, the second term

$$\left|\sum_{i=1}^{N}\left[\varepsilon_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_i\varepsilon_i - \varepsilon_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_i\varepsilon_i\right]\right| = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}}). \tag{33}$$

Combining (32) and (33), we obtain

$$\frac{1}{\phi_N}\left|\sum_{i=1}^{N}J_{2i}(k)\right| = \frac{1}{\phi_N}[O_p(\sqrt{\phi_N}) + O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}})] = O_p(\frac{1}{\sqrt{\phi_N}}) + \frac{1}{\phi_N}[O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}})].$$

Under Assumption 2, $\frac{1}{\phi_N}\left|\sum_{i=1}^{N}J_{2i}(k)\right|$ vanishes for any $k \in K(C_k)$, so does its maximum.

Compared with (8) of Model 1, (21) of Model 2 has the same form using transformed data $\{\tilde{Y}_i, \tilde{X}_i, i = 1, ..., N\}$, except for the additional term $M_w F\gamma_i$. The focus of the proof of Theorem 2 is on showing that $M_w F\gamma_i$ can be ignored asymptotically as $(N, T) \to \infty$.

For $i = 1, ..., N$, let $\widetilde{SSR}_i$ be the sum of squared residuals of regressing $\tilde{Y}_i$ on $\tilde{X}_i$ alone. Using the identity $\widetilde{SSR}_i - \widetilde{SSR}_i(k) = \tilde{\delta}_i(k)'[\tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{Z}_{2i}(k)]\tilde{\delta}_i(k)$ with $\tilde{M}_i = I - \tilde{X}_i(\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'$, we obtain

$$\tilde{k} = \arg \min_{1\le k\le T-1} \sum_{i=1}^{N} \pi_i \widetilde{SSR}_i(k) = \arg \max_{1\le k\le T-1} \sum_{i=1}^{N} \pi_i \widetilde{SV}_i(k) = \arg \max_{1\le k\le T-1} \sum_{i=1}^{N} \pi_i [\widetilde{SV}_i(k) - \widetilde{SV}_i(k_0)]$$

where $\widetilde{SV}_i(k) = \tilde{\delta}_i(k)'[\tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{Z}_{2i}(k)]\tilde{\delta}_i(k)$. Assume $\pi_i = 1/N$, $i = 1, ..., N$ for simplicity. The rest proof proceeds in the same way as that of Theorem 1 using the new notations with "~".

Partitioned regression gives

$$\tilde{\delta}_i(k) = \left[\tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{Z}_{2i}(k)\right]^{-1} \tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{Y}_i.$$

Substituting $\tilde{Y}_i = \tilde{X}_i\beta_i + \tilde{Z}_{0i}\delta_i + \tilde{\varepsilon}_i^0$ into the equation above, we obtain

$$\tilde{\delta}_i(k) = \left[\tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{Z}_{2i}(k)\right]^{-1} \tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{Z}_{0i}\delta_i + \left[\tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{Z}_{2i}(k)\right]^{-1} \tilde{Z}_{2i}(k)'\tilde{M}_i\tilde{\varepsilon}_i^0$$

and $\tilde{\delta}_i(k_0) = \delta_i + \left(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}\right)^{-1} \tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0$. Note that there is an additional term $M_w F\gamma_i$ in $\tilde{\varepsilon}_i^0 = M_w F\gamma_i + \tilde{\varepsilon}_i$.

As in the proof of Theorem 1, we suppress k in $\tilde{\delta}_i(k)$ and $\tilde{Z}_{2i}(k)$ for simplicity. Since

$$\begin{aligned}
\widetilde{SV}_i(k) &= \tilde{\delta}_i'(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})\tilde{\delta}_i \\
&= \delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{0i})\delta_i \\
&\quad + 2\delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 + \tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{2i}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0,
\end{aligned}$$

and $\widetilde{SV}_i(k_0) = \delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})\delta_i + 2\delta_i'\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0 + \tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})^{-1}\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0$, it follows that

$$\begin{aligned}
\widetilde{SV}_i(k) - \widetilde{SV}_i(k_0) &= \delta_i'[(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{0i}) - (\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})]\delta_i \\
&\quad + 2\delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - 2\delta_i'\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0 \\
&\quad + \tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{2i}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - \tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})^{-1}\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0.
\end{aligned}$$

Define

$$\tilde{J}_{1i}(k) = \delta_i'[(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}) - (\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{0i})]\delta_i.$$

In addition, define

$$\begin{aligned}\tilde{J}_{2i}(k) &= 2\delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - 2\delta_i'\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0 \\ &\quad +\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{2i}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - \tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})^{-1}\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0.\end{aligned}$$

Thus, $\widetilde{SV}_i(k) - \widetilde{SV}_i(k_0) = -\tilde{J}_{1i}(k) + \tilde{J}_{2i}(k)$ and

$$\tilde{k} = \arg\max_{1\le k\le T-1}\sum_{i=1}^N[\widetilde{SV}_i(k) - \widetilde{SV}_i(k_0)] = \arg\max_{1\le k\le T-1}\left[-\sum_{i=1}^N\tilde{J}_{1i}(k) + \sum_{i=1}^N\tilde{J}_{2i}(k)\right].$$

For each i, $i = 1, \cdots, N$, define $\tilde{Z}_{\Delta i} = \tilde{Z}_{2i}(k) - \tilde{Z}_{2i}(k_0)$ for $k < k_0$ and $\tilde{Z}_{2i}(k_0) - \tilde{Z}_{2i}(k)$ for $k \ge k_0$. Since $Z_{0i} = X_{0i}R$ and $Z_{2i} = X_{2i}R$, $\tilde{Z}_{\Delta i} = M_w Z_{\Delta i} = M_w X_{\Delta i}R = \tilde{X}_{\Delta i}R$.

As in the proof of Theorem 1, we assume that $\tilde{k} - k_0$ is stochastically bounded, and only consider the values of k belonging to the set $K(C_k) = \{k : 1 \le |k - k_0| < C_k, aT < k < (1-a)T\}$. Let $\tilde{\lambda}_1(k)$ be the minimum eigenvalue of $\frac{1}{N}\sum_{i=1}^N R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R$. Define $\tilde{\lambda}_1 = \min_{k\in K(C_k)}\tilde{\lambda}_1(k)$**.** Under Assumption 14, $\tilde{\lambda}_1(k) > 0$ and $\tilde{\lambda}_1 > 0$**.**

Lemma 3 *Under Assumptions 1, 2, 8-15 (or 16), for all large N and T, with probability tending to 1,*

$$\inf_{K(C_k)}\sum_{i=1}^N\tilde{J}_{1i}(k) \ge \tilde{\lambda}_1\phi_N.$$

Lemma 3 can be shown in the same way as Lemma 1 using the transformed data or the "~" notation. See the Supplementary Appendix.

Different from Model 1, there is an extra term $M_w F\gamma_i$ in the error $\tilde{\varepsilon}_i^0$, thus in $\tilde{J}_{2i}(k)$. To examine the effect of this extra term on the estimated $\tilde{k}$ and $\tilde{b}_i$, we introduce some new matrix notation. Since $x_{it} = \Gamma_i' f_t + v_{it}$ in (3), we write

$$\underset{T\times p}{X_i} = \underset{T\times m}{F}\ \underset{m\times p}{\Gamma_i} + \underset{T\times p}{V_i},$$

where $V_i = (v_{i1,} \cdots, v_{iT})'$. Denote $F_0 = (0, \cdots, 0, f_{k_0+1}, \cdots, f_T)'$ and $V_{0i} = (0, \cdots, 0, v_{i,k_0+1}, \cdots, v_{i,T})'$. Thus,

$$\begin{aligned} X_{0i} &= (0, \cdots, 0, x_{i,k_0+1}, \cdots, x_{i,T})' = (0, \cdots, 0, \Gamma_i' f_{k_0+1} + v_{i,k_0+1}, \cdots, \Gamma_i' f_T + v_{i,T})' \\ &= F_0 \Gamma_i + V_{0i}. \end{aligned}$$

For the error term (18), denote $\bar{u}_t = \begin{pmatrix} \bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}' \beta_i \\ \bar{v}_t \end{pmatrix}$ and

$$\Delta \bar{u}_t(k_0) = \begin{cases} \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & t = 1, ..., k_0, \\ \begin{pmatrix} \sum_{i=1}^N \theta_i v_{it}' R \delta_i \\ 0 \end{pmatrix}, & t = k_0 + 1, ..., T. \end{cases}$$

Thus, $\bar{u}_t(k_0) = \sum_{i=1}^N \theta_i u_{it}(k_0) = \bar{u}_t + \Delta \bar{u}_t(k_0)$. Denote $\bar{U} = (\bar{u}_1, ..., \bar{u}_T)'$ and

$$\Delta \bar{U}(k_0) = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{i=1}^N \theta_i v_{i,k_0+1}' R \delta_i \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} \sum_{i=1}^N \theta_i v_{i,T}' R \delta_i \\ 0 \end{pmatrix} \right)'.$$

Thus, stacking cross-sectional averages $\bar{w}_t = \bar{C}(k_0)' f_t + \bar{u}_t(k_0)$, we obtain

$$\begin{aligned} \underset{T \times (p+1)}{\bar{W}} &= (\bar{w}_1, ..., \bar{w}_{k_0}, \bar{w}_{k_0+1}, \cdots, \bar{w}_T)' \\ &= (\bar{C}_1' f_1 + \bar{u}_1, ..., \bar{C}_1' f_{k_0} + \bar{u}_{k_0}, \bar{C}_2' f_{k_0+1} + \bar{u}_{k_0+1}, \cdots, \bar{C}_2' f_T + \bar{u}_T)' \\ &= F \bar{C}_1 + F_0 (\bar{C}_2 - \bar{C}_1) + \bar{U} + \Delta \bar{U}(k_0). \end{aligned}$$

Denote $\underset{T \times 2m}{\mathbb{F}} = (F, F_0)$, $\underset{2m \times (p+1)}{\overline{\mathbb{C}}} = (\bar{C}_1', (\bar{C}_2 - \bar{C}_1)')'$ and $\underset{T \times (p+1)}{\bar{\mathbb{U}}} = \bar{U} + \Delta \bar{U}(k_0)$. Therefore,

$$\bar{W} = \mathbb{F} \overline{\mathbb{C}} + \bar{\mathbb{U}}. \tag{34}$$

With this notation, we obtain a lemma, which can be proved similarly to Lemmas 1, 2 and 3 in Pesaran (2006).

Lemma 4 *Under Assumptions 1, 2, 8-15, uniformly on $K(C_k)$,*

(i) $\bar{u}_t = O_p(\frac{1}{\sqrt{N}})$, $\Delta \bar{u}_t(k_0) = O_p(\frac{1}{\sqrt{N}})$;

(ii) $\frac{1}{T} \bar{\mathbb{U}}' \bar{\mathbb{U}} = O_p(\frac{1}{N})$; $\frac{1}{T} \mathbb{F}' \bar{\mathbb{U}} = O_p(\frac{1}{\sqrt{NT}})$, $\frac{1}{T} V_i' \mathbb{F} = O_p(\frac{1}{\sqrt{T}})$;

(iii) $\frac{1}{T} V_i' \bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, $\frac{1}{T} \varepsilon_i' \bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, $\frac{1}{T} V_{0i}' \bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$;

(iv) $\frac{1}{T} X_i' \bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$; $\frac{1}{T} X_{0i}' \bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$.

Lemma 5 *Under Assumptions 1, 2, 8-15, uniformly on $K(C_k)$,*

(i) $\frac{1}{T}\mathbb{F}'\mathbb{F} = O_p(1)$; $\frac{1}{T}\mathbb{F}'F = O_p(1)$;

(ii) $\frac{1}{T}X_i'\mathbb{F} = O_p(1)$; $\frac{1}{T}\mathbb{X}_i(k)'\mathbb{F} = O_p(1)$.

Proof of Lemma 5. (i) is obvious by Assumption 8.

(ii) Since $X_i = F\Gamma_i + V_i = (F, F_0)(\Gamma_i', 0)' + V_i$, $\frac{1}{T}X_i'\mathbb{F}$ can be written as $(\Gamma_i', 0)(\frac{1}{T}\mathbb{F}'\mathbb{F}) + \frac{1}{T}V_i'\mathbb{F}$. By (i) and Lemma 4 (iv), $\frac{1}{T}X_i'\mathbb{F} = O_p(1)$. Similarly, $\frac{1}{T}\mathbb{X}_i(k)'\mathbb{F} = O_p(1)$.

With Lemmas 4 and 5, we are ready to establish the property of the $T \times m$ matrix $M_w F\gamma_i$, which will be frequently used in the derivations below. Denote

$$\begin{aligned} \underset{(p+1)\times(p+1)}{E} &= \frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\bar{\mathbb{U}} + \frac{1}{T}\bar{\mathbb{U}}'\mathbb{F}\overline{\mathbb{C}} + \frac{1}{T}\bar{\mathbb{U}}'\bar{\mathbb{U}}; \\ \underset{(p+1)\times(p+1)}{f(E)} &= \sum_{k=1}^{\infty}(-1)^{k+1}[(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1}E]^k(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1}. \end{aligned}$$

By Lemma 5 (v), $E = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, thus $f(E) = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$. In addition, denote

$$\underset{2m\times m}{D_1} = -\overline{\mathbb{C}}f(E)\overline{\mathbb{C}}'\frac{\mathbb{F}'F}{T} + \overline{\mathbb{C}}[(\overline{\mathbb{C}}'\frac{\mathbb{F}'\mathbb{F}}{T}\overline{\mathbb{C}})^{-1} + f(E)]\frac{\bar{\mathbb{U}}'F}{T} \tag{35}$$

and

$$\underset{(p+1)\times m}{D_2} = -[(\overline{\mathbb{C}}'\frac{\mathbb{F}'\mathbb{F}}{T}\overline{\mathbb{C}})^{-1} + f(E)](\overline{\mathbb{C}}'\frac{\mathbb{F}'F}{T} + \frac{\bar{\mathbb{U}}'F}{T}). \tag{36}$$

Since $\overline{\mathbb{C}} = O(1)$, $\frac{\mathbb{F}'F}{T}$ and $\frac{\mathbb{F}'\mathbb{F}}{T}$ are $O_p(1)$, $f(E) = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, and $\frac{\bar{\mathbb{U}}'F}{T} = O_p(\frac{1}{\sqrt{NT}})$,

$$\begin{aligned} D_1 &= O_p(1)[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})]O_p(1) + O_p(1)[O_p(1) + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})]O_p(\frac{1}{\sqrt{NT}}) \\ &= O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}). \end{aligned}$$

Similarly,

$$D_2 = [O_p(1) + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})][O_p(1) + O_p(\frac{1}{\sqrt{NT}})] = O_p(1).$$

Lemma 6 *Under Assumptions 1, 2, 8-15, uniformly on $K(C_k)$,*

$$M_w F\gamma_i = \mathbb{F}D_1\gamma_i + \bar{\mathbb{U}}D_2\gamma_i.$$

By Lemma 4 (i) where each element of $\bar{\mathbb{U}}$ *is* $O_p(\frac{1}{\sqrt{N}})$*, each element of* $M_w F\gamma_i$ *is of order* $O_p(\frac{1}{\sqrt{N}})$.

Proof of Lemma 6. Plugging in (34), we obtain

$$\frac{1}{T}\bar{W}'\bar{W} = \frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}} + \frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\bar{\mathbb{U}} + \frac{1}{T}\bar{\mathbb{U}}'\mathbb{F}\overline{\mathbb{C}} + \frac{1}{T}\bar{\mathbb{U}}'\bar{\mathbb{U}} = \frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}} + E.$$

By Lemma 5 (i), $\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}}$ is $O_p(1)$. Since $E = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, it could be very small when both N and T are large. By Horn and Johnson (1985, p.335)

$$\begin{aligned}
& (\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} - (\frac{1}{T}\bar{W}'\bar{W})^{-1} = (\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} - (\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}} + E)^{-1} \\
= \; & (\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} - [I + (\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1}E]^{-1}(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} \\
= \; & \sum_{k=1}^{\infty}(-1)^{k+1}[(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1}E]^k(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} = f(E).
\end{aligned}$$

This yields

$$(\frac{1}{T}\bar{W}'\bar{W})^{-1} = (\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} + f(E).$$

It follows that

$$\begin{aligned}
M_w F &= [I_T - \bar{W}(\frac{1}{T}\bar{W}'\bar{W})^{-1}\frac{1}{T}\bar{W}']F = [I_T - (\mathbb{F}\overline{\mathbb{C}} + \bar{\mathbb{U}})[(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} + f(E)]\frac{1}{T}(\mathbb{F}\overline{\mathbb{C}} + \bar{\mathbb{U}})']F \\
&= [I_T - (\mathbb{F}\overline{\mathbb{C}})(\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1}(\mathbb{F}\overline{\mathbb{C}})']F - (\mathbb{F}\overline{\mathbb{C}})\{f(E)(\frac{1}{T}\mathbb{F}\overline{\mathbb{C}})' + [(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} + f(E)]\frac{1}{T}\bar{\mathbb{U}}'\}F \\
&\quad -\bar{\mathbb{U}}[(\frac{1}{T}\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1} + f(E)](\frac{1}{T}\mathbb{F}\overline{\mathbb{C}} + \frac{1}{T}\bar{\mathbb{U}})'F,
\end{aligned}$$

As discussed in Pesaran (2006), $M_{\mathbb{F}\overline{\mathbb{C}}} = I_T - (\mathbb{F}\overline{\mathbb{C}})(\overline{\mathbb{C}}'\mathbb{F}'\mathbb{F}\overline{\mathbb{C}})^{-1}(\mathbb{F}\overline{\mathbb{C}})' = I_T - \mathbb{F}(\mathbb{F}'\mathbb{F})^{-1}\mathbb{F}$ under the rank assumption. This implies that the first term is 0. Therefore, plugging in (35) and (36), we obtain

$$M_w F\gamma_i = \mathbb{F}D_1\gamma_i + \bar{\mathbb{U}}D_2\gamma_i. \tag{37}$$

The following lemma collects terms involving $M_w F\gamma_i$.

Lemma 7 *Under Assumptions 1, 2, 8-15 (or 16), uniformly on* $K(C_k)$*,*

(i) $\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'M_wF\gamma_i = O_p(\sqrt{\phi_N})$;

(ii) $\frac{1}{T}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\tilde{X}_i'M_wF\gamma_i = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iii) $\frac{1}{T}\sum_{i=1}^{N}\delta_i'(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{2i}'\tilde{M}_iM_wF\gamma_i = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iv)a. $\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{T})$;

b. $\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{T})$;

(v)a. $\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{\sqrt{T}})$;

b. $\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i = O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}})$;

c. $\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}})$;

(vi)a. $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\tilde{Z}_{0i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$;

b. $\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\tilde{Z}_{0i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{N}) + O_p(\frac{1}{T}) + O_p(\frac{1}{\sqrt{NT}})$.

Proof of Lemma 7.

(i) Consider the term

$$\tilde{Z}_{\Delta i}'M_wF\gamma_i = Z_{\Delta i}'M_wF\gamma_i = Z_{\Delta i}'\mathbb{F}D_1\gamma_i + Z_{\Delta i}'\bar{\mathbb{U}}D_2\gamma_i.$$

Since there are $|k-k_0|$ non-zero elements in $Z_{\Delta i}$, $Z_{\Delta i}'\mathbb{F}$ is equal to the sum of $|k-k_0|$ elements, and is finite on $K(C_k)$. Similarly, $Z_{\Delta i}'\bar{\mathbb{U}}$ is also equal to the sum of $|k-k_0|$ elements, each is $O_p(\frac{1}{\sqrt{N}})$. Due to the fact that $D_1 = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$ and $D_2 = O_p(1)$,

$$Z_{\Delta i}'M_wF\gamma_i = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{\sqrt{N}}) = O_p(\frac{1}{\sqrt{N}}).$$

Therefore,

$$\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'M_wF\gamma_i = O_p(\sqrt{N\phi_N})O_p(\frac{1}{\sqrt{N}}) = O_p(\sqrt{\phi_N}).$$

The rest of this lemma can be proved similarly. See the Supplementary Appendix for a detailed proof.

Lemma 2 in Model 1 can be extended to Model 2 using transformed data as Lemma 8.

Lemma 8 *Under Assumptions 1, 2, 8-15 (or 16), uniformly on $K(C_k)$,*

(i) $\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i = O_p(\sqrt{\phi_N})$;

(ii) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\frac{\tilde{X}_i'\tilde{\varepsilon}_i}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{T}})$;

(iii) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\left[(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i}{\sqrt{T}}\right] = O_p(\sqrt{\frac{\phi_N}{T}})$;

(iv) $\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i = O_p(\frac{N}{T})$;

(v) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}})$;

(vi) $\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i}{\sqrt{T}} = O_p(\frac{N}{T})$.

When the error $\tilde{\varepsilon}_i$ is replaced with $\tilde{\varepsilon}_i^0 = M_wF\gamma_i + \tilde{\varepsilon}_i$, terms involving $M_wF\gamma_i$ in Lemma 7 are used to obtain the following lemma.

Lemma 9 *Under Assumptions 1, 2, 8-15 (or 16), uniformly on $K(C_k)$,*

(i) $\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i^0 = O_p(\sqrt{\phi_N})$;

(ii) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\frac{\tilde{X}_i'\tilde{\varepsilon}_i^0}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iii) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iv) $\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 = O_p(\frac{N}{T})$;

(v) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}}) + O_p(\frac{1}{\sqrt{N}})$;

(vi) $\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0}{\sqrt{T}} = O_p(\frac{N}{T}) + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$.

Proof of Lemma 9. (i) Plugging in $\tilde{\varepsilon}_i^0 = M_wF\gamma_i + \tilde{\varepsilon}_i$, we obtain

$$\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i^0 = \sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'M_wF\gamma_i + \sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i$$

Lemma 7(i) shows that the first term is $O_p(\sqrt{\phi_N})$. By Lemma 8 (i), the second term

$$\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i = O_p(\sqrt{\phi_N})$$

uniformly on $K(C_k)$. It follows that

$$\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i^0 = O_p(\sqrt{\phi_N}) + O_p(\sqrt{\phi_N}) = O_p(\sqrt{\phi_N}).$$

The proofs of Lemma 9(ii)-(vi) are included in the Supplementary Appendix.

With these lemmas, we are ready to prove Theorem 2.

Proof of Theorem 2. To prove $\lim_{(N,T)\to\infty} P(\tilde{k} = k_0) = 1$, it is equivalent to show that for any given $\epsilon > 0$, for both large T and N, $P(|\tilde{k} - k_0| \geq 1) < \epsilon$. It is sufficient to show that $P\left(\sup_{K(C_k)} \sum_{i=1}^{N} \left[\widetilde{SV}_i(k) - \widetilde{SV}_i(k_0)\right] \geq 0\right) < \epsilon$, or

$$P\left(\sup_{K(C_k)} \left|\sum_{i=1}^{N} \tilde{J}_{2i}(k)\right| \geq \inf_{K(C_k)} \sum_{i=1}^{N} \tilde{J}_{1i}(k)\right) < \epsilon.$$

By Lemma 3, it is sufficient to show

$$P(\sup_{K(C_k)} \frac{1}{\phi_N} \left|\sum_{i=1}^{N} \tilde{J}_{2i}(k)\right| \geq \tilde{\lambda}_1) < \epsilon.$$

For any $k \in K(C_k)$,

$$\begin{aligned}\left|\sum_{i=1}^{N} \tilde{J}_{2i}(k)\right| \leq & \left|\sum_{i=1}^{N} \left[2\delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - 2\delta_i'\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0\right]\right| \\ & + \left|\sum_{i=1}^{N} \left[\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{2i}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - \tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})^{-1}\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0\right]\right|.\end{aligned}$$

Consider the first term, with $\tilde{Z}_{2i} = \tilde{Z}_{0i} \pm \tilde{Z}_{\Delta i}$,

$$\begin{aligned} & \left|\sum_{i=1}^{N} \left[2\delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - 2\delta_i'\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0\right]\right| \\ \leq \; & 2\left|\sum_{i=1}^{N} \delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i^0\right| + \frac{2}{\sqrt{T}}\left|\sum_{i=1}^{N} \delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\frac{\tilde{X}_i'\tilde{\varepsilon}_i^0}{\sqrt{T}}\right| \\ & + \frac{2}{\sqrt{T}}\left|\sum_{i=1}^{N} \delta_i'\left[(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0}{\sqrt{T}}\right]\right|. \end{aligned}$$

By (i), (ii) and (iii) of Lemma 9, the first term

$$\left|\sum_{i=1}^{N} \left[2\delta_i'(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0 - 2\delta_i'\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0\right]\right| = O_p(\sqrt{\phi_N}). \tag{38}$$

Now consider the second term

$$\begin{aligned}
& \left|\sum_{i=1}^{N}\left[\tilde{\tilde{\varepsilon}}_i^{0\prime}\tilde{M}_i\tilde{Z}_{2i}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\tilde{\varepsilon}}_i^{0}-\tilde{\tilde{\varepsilon}}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})^{-1}\tilde{Z}_{0i}'\tilde{M}_i\tilde{\tilde{\varepsilon}}_i^{0}\right]\right| \\
\leq\ & \frac{1}{T}\left|\sum_{i=1}^{N}\tilde{\tilde{\varepsilon}}_i^{0\prime}\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\tilde{\varepsilon}}_i^{0}\right| \\
& +2\frac{1}{\sqrt{T}}\left|\sum_{i=1}^{N}\frac{\tilde{\tilde{\varepsilon}}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\tilde{\varepsilon}}_i^{0}\right| \\
& +\left|\sum_{i=1}^{N}\frac{\tilde{\tilde{\varepsilon}}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}-(\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{\tilde{\varepsilon}}_i^{0}}{\sqrt{T}}\right|.
\end{aligned}$$

By (iv), (iv) and (vi) of Lemma 9, the second term

$$\begin{aligned}
& \left|\sum_{i=1}^{N}\left[\tilde{\tilde{\varepsilon}}_i^{0\prime}\tilde{M}_i\tilde{Z}_{2i}(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}\tilde{Z}_{2i}'\tilde{M}_i\tilde{\tilde{\varepsilon}}_i^{0}-\tilde{\tilde{\varepsilon}}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}(\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i})^{-1}\tilde{Z}_{0i}'\tilde{M}_i\tilde{\tilde{\varepsilon}}_i^{0}\right]\right| \\
=\ & O_p(\frac{N}{T})+O_p(\sqrt{\frac{N}{T}})+O_p(\frac{1}{\sqrt{N}}).
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
& \frac{1}{\phi_N}\left|\sum_{i=1}^{N}\tilde{J}_{2i}(k)\right| = \frac{1}{\phi_N}[O_p(\sqrt{\phi_N})+O_p(\frac{N}{T})+O_p(\sqrt{\frac{N}{T}})+O_p(\frac{1}{\sqrt{N}})] \\
=\ & O_p(\frac{1}{\sqrt{\phi_N}})+\frac{1}{\phi_N}[O_p(\frac{N}{T})+O_p(\sqrt{\frac{N}{T}})]+\frac{1}{\phi_N}O_p(\frac{1}{\sqrt{N}}).
\end{aligned}$$

Compared with the proof of Theorem 1, there is an extra term $\frac{1}{\phi_N}O_p(\frac{1}{\sqrt{N}}) = o_p(1)$. But, $\frac{1}{\phi_N}\left|\sum_{i=1}^{N}\tilde{J}_{2i}(k)\right|$ vanishes for any $k \in K(C_k)$ as $(N,T)\to\infty$, so does its maximum.

References

[1] Alvarez, J., and Arellano, M. (2003), The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators, *Econometrica*, 71, 1121-1159.

[2] Anselin, L. (1988), *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

[3] Andrews, D. W. K. (2005), Cross-section Regression with Common Shocks, *Econometrica*, 73, 1551–1585.

[4] Bai, J. (1994), Estimation of Structural Change Based on Wald-Type Statistics, Working paper 94-6, Economics Department, MIT.

[5] Bai, J. (1997a), Estimation of a Change Point in Multiple Regression Models, *Review of Economics and Statistics,* 79, 551-563.

[6] Bai, J. (1997b), Estimating Multiple Breaks One at A Time, *Econometric Theory,* 13, 315-352.

[7] Bai, J. (2000), Vector Autoregressive Models with Structural Change in Regression Coefficients and in Variance-Covariance Matrix, *Annals of Economics and Finance*, Vol. 1, 303-339.

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

[9] Bai, J. (2010), Common Breaks in Means and Variances for Panel Data, *Journal of Econometrics*, 157, 78-92.

[10] Bai, J., Lumsdaine, R., and Stock, J., (1998), Testing For and Dating Common Breaks in Multivariate Time Series, *Review of Economic Studies*, 65, 395-432.

[11] Bai, J., and Carrion-i-Silvestre, J. L. (2009), Structural Changes, Common Stochastic Trends, and Unit Roots in Panel Data, *Review of Economic Studies*, 76, 471-501.

[12] Bai, J., and Perron, P. (1998), Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica*, 66, 47-79.

[13] Bai, J., and Perron, P. (2003), Computation and Analysis of Multiple Structural Change Models, *Journal of Applied Econometrics*, 18, 1-22.

[14] Baltagi, B., and Pirotte, A. (2010), Panel Data Inference Under Spatial Dependence, *Economic Modelling*, 27, 1368–1381.

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

[16] Baltagi, B. H., Feng, Q., and Kao, C. (2012), A Lagrange Multiplier Test for Cross-sectional Dependence in a Fixed Effects Panel Data Model, *Journal of Econometrics*, 170, 164-177.

[17] Baltagi, B. H., Kao, C., and L. Liu (2012), Estimation and Identification of Change Points in Panel Models with Nonstationary or Stationary Regressors and Error Term, *Econometric Reviews*, forthcoming.

[18] Banerjee, A., and Carrion-i-Silvestre, J. L. (2011), Testing for Panel Cointegration Using Common Correlated Effects Estimators, Department of Economics Discussion Paper 11-16, University of Birmingham.

[19] Banerjee, A., Marcellino, M., and Masten, I. (2008), Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change, CEPR Discussion Papers 6706.

[20] Breitung, J., and Eickmeier, S. (2011), Testing for Structural Breaks in Dynamic Factor Models, *Journal of Econometrics*, 163, 71-84.

[21] Chen, L., Dolado, J., and Gonzalo, J. (2014), Detecting Big Structural Breaks in Large Factor Models, *Journal of Econometrics,* 180, 30-48.

[22] Chen, J., Gao, J., and Li, D. (2012), A New Diagnostic Test for Cross-sectional Uncorrelatedness in Nonparametric Panel Data Model, *Econometric Theory,* 28, 1144-1163.

[23] Chong, T. (1995), Partial Parameter Consistency in a Misspecified Structural Change Model, *Economics Letters,* 49, 351-357.

[24] Chudik, A., and Pesaran, M. H. (2013), Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors, Working Paper, USC.

[25] Chudik, A., and Pesaran, M. H. and Tosetti, E. (2011), Weak and Strong Cross-section Dependence and Estimation of Large panels, *Econometrics Journal*, 14, C45-C90.

[26] De Wachter, S., Tzavalis, E., (2012), Detection of Structural Breaks in Linear Dynamic Panel Data Models, *Computational Statistics & Data Analysis*, 56, 3020-3034.

[27] Feng, Q., Kao, C., and Lazarova, S. (2009), Estimation of Change Points in Panels, Working Paper, Syracuse University.

[28] Fleisher, B., Li, H., and Zhao, M. Q. (2010), Human Capital, Economic Growth, and Regional Inequality in China, *Journal of Development Economics,* 92, 215–231.

[29] Hahn, J., and Kuersteiner, G. (2002), Asymptotically Unbiased Inference for a Dynamic Panel Data Model with Fixed Effects, *Econometrica*, 70, 1639-1657.

[30] Halunga, A., Orme, C. D., and Yamagata, T. (2011). A Heteroskedasticity Robust Breusch-Pagan Test for Contemporaneous Correlation in Dynamic Panel Data Models. University of Manchester Economics Discussion Paper Series.

[31] Hayakawa, K. (2009), A Simple Efficient Instrumental Variable Estimator for Panel AR(p) Models When Both N and T are Large, *Econometric Theory*, 25, 873-890.

[32] Horn, R., and Johnson, C. (1985), *Matrix Analysis*, Cambridge University Press.

[33] Hsiao, C., Pesaran, M. H., and Pick, A. (2012), Diagnostic Tests of Cross Section Independence for Nonlinear Panel Data Models, *Oxford Bulletin of Economics and Statistics,* 74, 253-277.

[34] Hsu, C-C., and Lin, C-C. (2012). Change-point Estimation for Nonstationary Panel Data. Working Paper, National Central University.

[35] Horváth, L., and Hušková, M., (2012), Change-point Detection in Panel Data, *Journal of Time Series Analysis,* 33, 631-648.

[36] Huang, Y. (2009), The Political Economy of Financial Reform: Are Abiad and Mody Right? *Journal of Applied Econometrics,* 24, 1207-1213.

[37] Juhl, T., 2011, A Direct Test for Cross-Sectional Correlation in Panel Data Models, Working Paper, University of Kansas.

[38] Kao, C. (1999), Spurious Regression and Residual-based Tests for Cointegration in Panel Data, *Journal of Econometrics*, 90, 1-44.

[39] Kapetanios, G., Pesaran, M. H., and Yamagata, T. (2011), Panels With Nonstationary Multifactor Error Structures, *Journal of Econometrics*, 160, 326-348.

[40] Kapoor, M., Kelejian, H., and Prucha, I. (2007), Panel Data Models with Spatially Correlated Error Components, *Journal of Econometrics,* 140, 97-130

[41] Kelejian, H., and Prucha, I. (1999), A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, *International Economic Review,* 40, 509-533.

[42] Kim, D. (2011), Estimating a Common Deterministic Time Trend Break in Large Panels with Cross Sectional Dependence, *Journal of Econometrics,* 164, 310-330.

[43] Kim, D. (2014), Common Local Breaks in Time Trends for Large Panels, *Econometrics Journal*, 17, 301-337.

[44] Lee, L. (2002), Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models, *Econometric Theory*, 18, 252-277.

[45] Lee, L. (2007), GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, *Journal of Econometrics*, 137, 489-514.

[46] Lee, L., and Yu, J. (2010), Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects, *Journal of Econometrics,* 154, 165-185.

[47] Li, D., Chen, J., and Gao, J. (2011), Nonparametric Time-varying Coefficient Panel Data Models with Fixed Effects, *Econometrics Journal,* 14, 387-408.

[48] Li, D., J. Qian, and Su, L. (2014), Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks, Working Paper, Singapore Management University.

[49] Liao, W., and Wang, P. (2012), Structural Breaks in Panel Data Models: a Common Distribution Method, HKUST Working Paper.

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

[51] Pauwels, L., Chan, F., and Mancini-Griffoli, T. (2012), Testing for Structural Change in Heterogeneous Panels with an Application to the Euro's Trade Effect, *Journal of Time Series Econometrics*, Volume 4, Issue 2, Article 3.

[52] Qu, Z., and Perron, P. (2007), Estimating and Testing Structural Changes in Multivariate Regressions, *Econometrica*, 75, 459-503.

[53] Pesaran, M. H. (2004), General Diagnostic Test for Cross Section Dependence in Panels, Working Paper, University of Cambridge & USC.

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

[55] Pesaran, M. H. (2012), Testing Weak Cross-sectional Dependence in Large Panels. IZA Discussion Paper No. 6432, March 2012.

[56] Pesaran, M. H., and Smith, R. (1995), Estimating Long-run Relationships from Dynamic Heterogeneous Panels, *Journal of Econometrics*, 68, 79-113.

[57] Pesaran, M. H., and Timmermann, A. (2002), Market Timing and Return Prediction Under Model Instability, *Journal of Empirical Finance,* 9, 495-510.

[58] Pesaran, M. H., and Tosetti, E., (2011), Large Panels with Common Factors and Spatial Correlation, *Journal of Econometrics*, 161, 182-202.

[59] Pesaran, M. H., and Yamagata, T. (2008), Testing Slope Homogeneity in Large Panels, *Journal of Econometrics*, 142, 50-93.

[60] Pesaran, M. H., Ullah, A., and Yamagata, T. (2008), A Bias-Adjusted LM Test of Error Cross Section Independence, *Econometrics Journal*, 11, 105-127.

[61] Phillips, P. C. B., and Moon, H. R. (1999), Linear Regression Limit Theory for Non-stationary Panel Data, *Econometrica,* 67, 1057–1111.

[62] Phillips, P. C. B., and Sul, D. (2007), Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence, *Journal of Econometrics*, 137, 162-188.

[63] Qian, J., and Su, L. (2014), Shrinkage Estimation of Common Breaks in Panel Data Models via Adaptive Group Fused Lasso, Working Paper, Singapore Management University.

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

[65] Stock J., and Watson, M. W. (2002), Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association,* 97, 1167-1179.

[66] Stock J., and Watson, M. W. (2009), Forecasting in Dynamic Factor Models Subject to Structural Instability. In: Shephard N, Castle J, *The Methodology and Practice of Econometrics: A Festschrift in Honor of D.F. Hendry.* Oxford University Press, 1-57.

[67] Song, M. (2012), Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-Sectional Dependence and Its Application, Working Paper, Columbia University.

[68] Su, L., and Zhang, Y. (2011), Testing Cross-Sectional Independence in Nonparametric Panel Data Models, Working Paper, Singapore Management University.

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

Supplementary Appendix (not for publication):

The asymptotic distributions of CCE estimators $\tilde{b}_i$ and $\tilde{b}_{MG}$ in model 2 are summarized in Propositions 1 and 2.

Proposition 1. Under Assumptions 1-3, 8-15, and $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$, for each i,

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

where

$$\Sigma_{\tilde{\mathbb{X}},i} = plim_{T\to\infty}\frac{1}{T}\tilde{\mathbb{X}}_{0i}'\tilde{\mathbb{X}}_{0i} \text{ and } \Sigma_{\tilde{\mathbb{X}}\tilde{\varepsilon},i} = plim_{T\to\infty}\frac{1}{T}\tilde{\mathbb{X}}_{0i}'\Sigma_{\varepsilon,i}\tilde{\mathbb{X}}_{0i},\ i = 1, ..., N.$$

Proof of Proposition 1.

By (21), for $i = 1, ..., N$. Since

$$\tilde{Y}_i = \tilde{\mathbb{X}}_{0i}b_i + \tilde{\varepsilon}_i^0 = \tilde{\mathbb{X}}_i(\tilde{k})b_i + \tilde{\varepsilon}_i^*, \tag{39}$$

where $\tilde{\varepsilon}_i^* = \tilde{\varepsilon}_i^0 + (\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i = M_wF\gamma_i + \tilde{\varepsilon}_i + (\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i$, we have

$$\begin{aligned}
\tilde{b}_i &= [\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{Y}_i = b_i + [\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\varepsilon}_i^* \\
&= b_i + [\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'M_wF\gamma_i \\
&\quad +[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\varepsilon}_i + [\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'(\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i
\end{aligned} \tag{40}$$

and

$$\begin{aligned}
&\sqrt{T}\left(\tilde{b}_i - b_i\right) = [\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_i(\tilde{k})'M_wF\gamma_i \\
&+[\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\varepsilon}_i + [\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_i(\tilde{k})'(\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i \\
= &[\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_i(\tilde{k})'M_wF\gamma_i + [\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_{0i}'\varepsilon_i \\
&+[\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i + [\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_i(\tilde{k})'(\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i,
\end{aligned}$$

where the second equality above uses the result $\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\varepsilon}_i = \tilde{\mathbb{X}}_{0i}'\varepsilon_i$.

Due to the results $\tilde{k} - k_0 = o_p(1)$ and $\tilde{\mathbb{X}}_i(\tilde{k}) \xrightarrow{p} \tilde{\mathbb{X}}_{0i}$, for each i,

$$\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k}) \xrightarrow{p} \Sigma_{\tilde{\mathbb{X}},i},$$

where $\Sigma_{\tilde{\mathbb{X}},i} = \text{plim}_{T\to\infty} \frac{1}{T}\tilde{\mathbb{X}}_{0i}'\tilde{\mathbb{X}}_{0i}$. Consider the term

$$\begin{aligned}
& \frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_i(\tilde{k})' M_w F\gamma_i = \frac{1}{\sqrt{T}}\mathbb{X}_i(\tilde{k})' M_w F\gamma_i \\
= & \frac{1}{\sqrt{T}}\mathbb{X}_i(\tilde{k})'\mathbb{F}D_1\gamma_i + \frac{1}{\sqrt{T}}\mathbb{X}_i(\tilde{k})'\bar{\mathbb{U}}D_2\gamma_i \\
= & \sqrt{T}[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})] + \sqrt{T}[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})] \\
= & O_p(\frac{\sqrt{T}}{N}) + O_p(\frac{1}{\sqrt{N}}).
\end{aligned}$$

By Theorem 2, $P(\tilde{k} \neq k_0) = P(|\tilde{k} - k_0| \geq 1) \to 0$. For any $\eta > 0$,

$$\begin{aligned}
& P\left(\left\|\frac{1}{\sqrt{T}}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i\right\| > \eta\right) \\
= & P\left(\left\|\frac{1}{\sqrt{T}}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i\right\| > \eta, \tilde{k} = k_0\right) + P\left(\left\|\frac{1}{\sqrt{T}}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i\right\| > \eta, \tilde{k} \neq k_0\right) \\
= & P(0 > \eta)P(\tilde{k} = k_0) + P\left(\left\|\frac{1}{\sqrt{T}}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i\right\| > \eta \middle| \tilde{k} \neq k_0\right) P\left(\tilde{k} \neq k_0\right) \\
\leq & P(0 > \eta)P(\tilde{k} = k_0) + P\left(\tilde{k} \neq k_0\right) \to 0.
\end{aligned}$$

Thus, $\frac{1}{\sqrt{T}}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i = o_p(1)$. Similarly, $\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_i(\tilde{k})'(\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i = o_p(1)$.

Therefore,

$$\begin{aligned}
& \sqrt{T}\left(\tilde{b}_i - b_i\right) \\
= & O_p(\frac{\sqrt{T}}{N}) + O_p(\frac{1}{\sqrt{N}}) + [\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_{0i}'\varepsilon_i + o_p(1) + o_p(1) \\
= & \Sigma_{\tilde{\mathbb{X}},i}^{-1}\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_{0i}'\varepsilon_i + O_p(\frac{\sqrt{T}}{N}) + o_p(1).
\end{aligned}$$

Since $\frac{1}{\sqrt{T}}\tilde{\mathbb{X}}_{0i}'\varepsilon_i \xrightarrow{d} N\left(0, \Sigma_{\tilde{\mathbb{X}}\varepsilon,i}\right)$ where $\Sigma_{\tilde{\mathbb{X}}\varepsilon,i} = \text{plim}_{T\to\infty} \frac{1}{T}\tilde{\mathbb{X}}_{0i}'\Sigma_{\varepsilon,i}\tilde{\mathbb{X}}_{0i}$, it follows that when $\frac{\sqrt{T}}{N} \to 0$ as $(N,T) \to \infty$,

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

Proposition 2. Under the assumptions 1-3, 8-14, 16,

$$\sqrt{N}(\tilde{b}_{MG} - b) \xrightarrow{d} N(0, \Sigma_b).$$

Proof of Proposition 2.

Under Assumption 16 of a random coefficient model, the asymptotic distribution of mean-group estimator (12) can be derived similarly. Plugging in (39) and (40),

$$\begin{aligned}
\tilde{b}_{MG} &= \frac{1}{N}\sum_{i=1}^{N}\tilde{b}_i = \frac{1}{N}\sum_{i=1}^{N}\left(b_i + [\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\varepsilon}_i^*\right) \\
&= b + \frac{1}{N}\sum_{i=1}^{N} v_{b,i} + \frac{1}{N}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})' M_w F\gamma_i \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'\varepsilon_i + \frac{1}{N}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'(\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i],
\end{aligned}$$

we obtain

$$\begin{aligned}
&\sqrt{N}(\tilde{b}_{MG} - b) \\
= \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} v_{b,i} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})' M_w F\gamma_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_{0i}'\varepsilon_i \\
& + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'(\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i].
\end{aligned}$$

By Assumption 16, the limiting distribution of the first term is $N(0, \Sigma_{b_i})$. Consider the second term

$$\begin{aligned}
& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})' M_w F\gamma_i \\
= \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\frac{\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})}{T}]^{-1}\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\mathbb{F}D_1\gamma_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\frac{\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})}{T}]^{-1}\frac{1}{T}\tilde{\mathbb{X}}_i(\tilde{k})'\bar{\mathbb{U}}D_2\gamma_i \\
= \; & \frac{1}{\sqrt{N}}[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})]O_p(N) + \frac{1}{\sqrt{N}}[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})]O_p(N) \\
= \; & O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}}).
\end{aligned}$$

The third term can be written as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_{0i}'\varepsilon_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})}{T}\right]^{-1}\frac{\tilde{\mathbb{X}}_{0i}'\varepsilon_i}{\sqrt{T}}.$$

Since $\varepsilon_i, i = 1, ..., N$, have zero mean and are distributed independently of $\tilde{\mathbb{X}}_i$, under Assumptions 9, 13, and 14,

$$\begin{aligned} & Var\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})}{T}\right]^{-1}\frac{\tilde{\mathbb{X}}_{0i}'\varepsilon_i}{\sqrt{T}}\right) \\ = & \frac{1}{NT}\sum_{i=1}^{N}\left[\frac{\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})}{T}\right]^{-1}\frac{\tilde{\mathbb{X}}_{0i}'Var(\varepsilon_i)\tilde{\mathbb{X}}_{0i}}{T}\left[\frac{\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})}{T}\right]^{-1} = O(\frac{1}{T}), \end{aligned}$$

implying $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})}{T}\right]^{-1}\frac{\tilde{\mathbb{X}}_{0i}'\varepsilon_i}{\sqrt{T}} = O_p(\frac{1}{\sqrt{T}})$.

As in the proof of Proposition 1, since $P(\tilde{k} \neq k_0) = P(|\tilde{k} - k_0| \geq 1) \to 0$ by Theorem 2, it can be similarly shown that

$$\begin{aligned} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}(\tilde{\mathbb{X}}_i(\tilde{k}) - \tilde{\mathbb{X}}_{0i})'\varepsilon_i &= o_p(1), \\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{\mathbb{X}}_i(\tilde{k})'\tilde{\mathbb{X}}_i(\tilde{k})]^{-1}\tilde{\mathbb{X}}_i(\tilde{k})'(\tilde{Z}_{0i} - \tilde{Z}_{2i}(\tilde{k}))\delta_i] &= o_p(1). \end{aligned}$$

Therefore, as $(N, T) \to \infty$,

$$\sqrt{N}(\tilde{b}_{MG} - b) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} v_{b,i} + O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\sqrt{T}}) + o_p(1) + o_p(1) \xrightarrow{d} N(0, \Sigma_b).$$

The asymptotic distributions of estimators of slopes $\hat{b}_i$ and $\hat{b}_{MG}$ in model 1 are summarized in Propositions 3 and 4.

Proposition 3 *Under Assumptions 1-6, as* $(N,T) \to \infty$, *for each* i,

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

where

$$\Sigma_{\mathbb{X},i} = plim_{T\to\infty}\frac{1}{T}\mathbb{X}_{0i}'\mathbb{X}_{0i} \text{ and } \Sigma_{\mathbb{X}\varepsilon,i} = plim_{T\to\infty}\frac{1}{T}\mathbb{X}_{0i}'\Sigma_{\varepsilon,i}\mathbb{X}_{0i},\ i = 1,...,N.$$

A similar result is obtained in Bai (1997a, Corollary 1). Due to the consistency of $\hat{k}$, $\Sigma_{\mathbb{X},i}$ can be consistently estimated by $\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})$. Also, $\Sigma_{\mathbb{X}\varepsilon,i}$ can be consistently estimated using the Newey and West (1987) approach, as discussed by Bai (1994) and Pesaran (2006), i.e.,

$$\begin{aligned}\hat{\Sigma}_{\mathbb{X}\varepsilon,i} &= \hat{\Lambda}_{i0} + \sum_{j=1}^{\omega}(1 - \frac{j}{\omega+1})(\hat{\Lambda}_{ij} + \hat{\Lambda}_{ij}'),\\ \hat{\Lambda}_{ij} &= \frac{1}{T}\sum_{t=j+1}^{\omega}\hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-j}\mathbb{X}_{it}(\hat{k})\mathbb{X}_{it}(\hat{k})',\end{aligned}$$

where ω is the window size, $\hat{\varepsilon}_{it} = y_{it} - x_{it}'(\hat{\beta}_i(\hat{k}) + R\hat{\delta}_i(\hat{k})\cdot 1\{t > \hat{k}\})$ and $\mathbb{X}_{it}(\hat{k}) = (x_{it}', z_{it}'\cdot 1\{t > \hat{k}\})$. Thus, a consistent estimator of $\Sigma_{\mathbb{X},i}^{-1}\Sigma_{\mathbb{X}\varepsilon,i}\Sigma_{\mathbb{X},i}^{-1}$ is

$$\left[\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})\right]^{-1}\hat{\Sigma}_{\mathbb{X}\varepsilon,i}\left[\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})\right]^{-1}. \tag{41}$$

Proof of Proposition 3.

We follow along the lines of Bai (1997a, Corollary 1) except for using a panel data estimator $\hat{k}$. By (9), for $i = 1,...,N$,

$$Y_i = \mathbb{X}_{0i}b_i + \varepsilon_i = \mathbb{X}_i(\hat{k})b_i + \varepsilon_i^*, \tag{42}$$

where $\varepsilon_i^* = \varepsilon_i + (Z_{0i} - Z_{2i}(\hat{k}))\delta_i$, we have

$$\begin{aligned}\hat{b}_i(\hat{k}) &= [\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'Y_i\\ &= b_i + [\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'\varepsilon_i^*\\ &= b_i + [\mathbb{X}_i(\hat{k})'\mathbb{X}(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'\varepsilon_i + [\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'(Z_{0i} - Z_{2i}(\hat{k}))\delta_i \qquad (43)\end{aligned}$$

and

$$\begin{aligned}
&\sqrt{T}\left(\hat{b}_i(\hat{k}) - b_i\right) \\
= \;& [\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\frac{1}{\sqrt{T}}\mathbb{X}_{0i}'\varepsilon_i + [\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\frac{1}{\sqrt{T}}(\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i})'\varepsilon_i \\
&+[\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\frac{1}{\sqrt{T}}\mathbb{X}_i(\hat{k})'(Z_{0i} - Z_{2i}(\hat{k}))\delta_i.
\end{aligned}$$

By symmetry, assume $\hat{k} \leq k_0$, so

$$\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i} = (0_{T\times p}, (0, ..., 0, z_{i,\hat{k}+1}, ..., z_{i,k_0+1}, 0, ..., 0)') \xrightarrow{p} 0.$$

Thus, due to the consistency of $\hat{k}$, for $i = 1, ..., N$,

$$\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k}) \xrightarrow{p} \Sigma_{\mathbb{X},i},$$

where $\Sigma_{\mathbb{X},i} = \text{plim}_{T\to\infty}\frac{1}{T}\mathbb{X}_{0i}'\mathbb{X}_{0i}$. By Theorem 1, $P(\hat{k} \neq k_0) = P(|\hat{k} - k_0| \geq 1) \to 0$. For any $\eta > 0$,

$$\begin{aligned}
&P\left(\left\|\frac{1}{\sqrt{T}}(\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i})'\varepsilon_i\right\| > \eta\right) \\
= \;& P\left(\left\|\frac{1}{\sqrt{T}}(\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i})'\varepsilon_i\right\| > \eta, \hat{k} = k_0\right) + P\left(\left\|\frac{1}{\sqrt{T}}(\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i})'\varepsilon_i\right\| > \eta, \hat{k} \neq k_0\right) \\
= \;& P(0 > \eta)P(\hat{k} = k_0) + P\left(\left\|\frac{1}{\sqrt{T}}(\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i})'\varepsilon_i\right\| > \eta \middle| \hat{k} \neq k_0\right) P\left(\hat{k} \neq k_0\right) \\
\leq \;& P(0 > \eta)P(\hat{k} = k_0) + P\left(\hat{k} \neq k_0\right) \to 0.
\end{aligned}$$

Thus, $\frac{1}{\sqrt{T}}(\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i})'\varepsilon_i = o_p(1)$. Similarly, $\frac{1}{\sqrt{T}}X_i(\hat{k})'(Z_{0i} - Z_{2i}(\hat{k}))\delta_i = o_p(1)$.

Therefore,

$$\begin{aligned}
&\sqrt{T}\left(\hat{b}_i - b_i\right) \\
= \;& [\frac{1}{T}\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\frac{1}{\sqrt{T}}\mathbb{X}_{0i}'\varepsilon_i + o_p(1) + o_p(1) \\
= \;& \Sigma_{\mathbb{X},i}^{-1}\cdot\frac{1}{\sqrt{T}}\mathbb{X}_{0i}'\varepsilon_i + o_p(1) \xrightarrow{d} N\left(0, \Sigma_{\mathbb{X},i}^{-1}\Sigma_{\mathbb{X}\varepsilon,i}\Sigma_{\mathbb{X},i}^{-1}\right),
\end{aligned}$$

where $\frac{1}{\sqrt{T}}\mathbb{X}_{0i}'\varepsilon_i \xrightarrow{d} N\left(0, \Sigma_{\mathbb{X}\varepsilon,i}\right)$ and $\Sigma_{\mathbb{X}\varepsilon,i} = \text{plim}_{T\to\infty}\frac{1}{T}\mathbb{X}_{0i}'\Sigma_{\varepsilon,i}\mathbb{X}_{0i}$.

Proposition 4 *Under Assumptions 1-5, 7,*

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

As in Pesaran (2006), Σ_b can be consistently estimated by

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

Proof of Proposition 4.

Under Assumption 7 of a random coefficient model, the asymptotic distribution of the mean-group estimator $\hat{b}_{MG}$ defined in (12) can be derived similarly. Plugging in (43),

$$\begin{aligned}\hat{b}_{MG} &= \frac{1}{N}\sum_{i=1}^{N}\hat{b}_i = \frac{1}{N}\sum_{i=1}^{N}\left(b_i + [\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'\varepsilon_i^*\right) \\ &= b + \frac{1}{N}\sum_{i=1}^{N}v_{b,i} + \frac{1}{N}\sum_{i=1}^{N}[\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'\varepsilon_i + \frac{1}{N}\sum_{i=1}^{N}[\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'(Z_{0i} - Z_{2i}(\hat{k}))\delta_i],\end{aligned}$$

we obtain

$$\begin{aligned}\sqrt{N}(\hat{b}_{MG} - b) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}v_{b,i} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_{0i}'\varepsilon_i \\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}(\mathbb{X}_i(\hat{k}) - \mathbb{X}_{0i})'\varepsilon_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'(Z_{0i} - Z_{2i}(\hat{k}))\delta_i].\end{aligned}$$

By Assumption 7, the limiting distribution of the first term is $N(0, \Sigma_b)$.

Consider the second term

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_{0i}'\varepsilon_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})}{T}\right]^{-1}\frac{\mathbb{X}_{0i}'\varepsilon_i}{\sqrt{T}}$$

Since for $i = 1, ..., N$, ε_i has zero mean and is distributed independently of $\mathbb{X}_i$, under Assumptions 3, 4, and 5,

$$\begin{aligned}&Var\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})}{T}\right]^{-1}\frac{\mathbb{X}_{0i}'\varepsilon_i}{\sqrt{T}}\right) \\ = \;& \frac{1}{NT}\sum_{i=1}^{N}\left[\frac{\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})}{T}\right]^{-1}\frac{\mathbb{X}_{0i}'\Sigma_{\varepsilon,i}\mathbb{X}_{0i}}{T}\left[\frac{\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})}{T}\right]^{-1} = O(\frac{1}{T}),\end{aligned}$$

implying $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})}{T}\right]^{-1}\frac{\mathbb{X}_{0i}'\varepsilon_i}{\sqrt{T}} = O_p(\frac{1}{\sqrt{T}})$.

As in the proof of Proposition 3, since $P(\hat{k} \neq k_0) = P(|\hat{k} - k_0| \geq 1) \to 0$ by Theorem 1, it can be similarly shown that the third and fourth terms above are $o_p(1)$. Hence, as $(N,T) \to \infty$,

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

Complete Proofs of Lemmas 1, 2, 3, 7 and 9.

Lemma 1. Under Assumptions 1-7, for all large N and T, with probability tending to 1,

$$\inf_{K(C_k)} \sum_{i=1}^{N} J_{1i}(k) \geq \lambda_1 \phi_N.$$

Proof of Lemma 1. By symmetry, we assume $k < k_0$. From (31) and Lemma A.1 in Bai (1997a),

$$\begin{aligned} J_{1i}(k) &= \delta_i'[(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})]\delta_i \\ &\geq \delta_i'R'(X_{\Delta i}'X_{\Delta i})(X_{2i}'X_{2i})^{-1}(X_{0i}'X_{0i})R\delta_i. \end{aligned}$$

Since $X_{2i}'X_{2i} = X_{0i}'X_{0i} + X_{\Delta i}'X_{\Delta i}$ and $\frac{1}{T}X_{\Delta i}'X_{\Delta i} = O_p(\frac{1}{T})$ on $K(C_k)$ for large T under Assumption 4, we obtain

$$\begin{aligned} &\sum_{i=1}^{N} \delta_i'R'(X_{\Delta i}'X_{\Delta i})(X_{2i}'X_{2i})^{-1}(X_{0i}'X_{0i})R\delta_i \\ = &\sum_{i=1}^{N} \delta_i'R'(X_{\Delta i}'X_{\Delta i})(\frac{X_{0i}'X_{0i}}{T} + \frac{X_{\Delta i}'X_{\Delta i}}{T})^{-1}(\frac{X_{0i}'X_{0i}}{T})R\delta_i \\ = &(1 + O_p(\frac{1}{T}))\sum_{i=1}^{N} \delta_i'R'(X_{\Delta i}'X_{\Delta i})R\delta_i. \end{aligned}$$

Thus, the term above is dominated by $\sum_{i=1}^{N} \delta_i'R'(X_{\Delta i}'X_{\Delta i})R\delta_i$. Denote $a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R)$ the $(j,l)^{th}$ element of matrix $R'(X_{\Delta i}'X_{\Delta i})R$.

$$\begin{aligned} &\sum_{i=1}^{N} \delta_i'R'(X_{\Delta i}'X_{\Delta i})R\delta_i = \sum_{i=1}^{N}\left[\sum_{j=1}^{p}\sum_{l=1}^{p} \delta_{i,j}\delta_{i,l} \cdot a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R)\right] \\ = &N\sum_{j=1}^{p}\sum_{l=1}^{p}\left[\frac{1}{N}\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l} \cdot a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R)\right]. \end{aligned}$$

By Assumption 6 or 7 that δ_i is independent of X_i, $\delta_{i,j}\delta_{i,l}$ is independent of $a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R)$. Thus,

$$E[(\delta_{i,j}\delta_{i,l}) \cdot a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R)] = E[(\delta_{i,j}\delta_{i,l})]E[a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R)],$$

implying that for large N, with probability tending to 1,

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l} \cdot a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R) = \left(\frac{1}{N}\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l}\right)\left(\frac{1}{N}\sum_{i=1}^{N} a_{j,l}(R'(X_{\Delta i}'X_{\Delta i})R)\right) \\ = &\left(\frac{1}{N}\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l}\right) a_{j,l}\left(\frac{1}{N}\sum_{i=1}^{N} R'(X_{\Delta i}'X_{\Delta i})R\right). \end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{i=1}^{N} \delta_i' R'(X_{\Delta i}' X_{\Delta i}) R \delta_i = N \sum_{j=1}^{p} \sum_{l=1}^{p} \left[\left(\frac{1}{N} \sum_{i=1}^{N} \delta_{i,j} \delta_{i,l} \right) a_{j,l} \left(\frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R \right) \right] \\
= &\sum_{j=1}^{p} \sum_{l=1}^{p} \sum_{i=1}^{N} \left[\delta_{i,j} \delta_{i,l} \cdot a_{j,l} \left(\frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R \right) \right] \\
= &\sum_{i=1}^{N} \delta_i' \left(\frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R \right) \delta_i
\end{aligned}$$

with probability tending to 1 for large N.

Denote $A = \frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R$. Under Assumptions 4, 5, A is positive definite, with a nonzero minimum eigenvalue, i.e., $\lambda_1(k) > 0$. In addition, there exists an orthogonal matrix S such that $A = S' \Lambda S$, where Λ is a diagonal matrix comprising of the eigenvalues of matrix A. Hence,

$$\sum_{i=1}^{N} \delta_i' \left(\frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R \right) \delta_i = \sum_{i=1}^{N} \delta_i' S' \Lambda S \delta_i = \sum_{i=1}^{N} \tilde{\delta}_i' \Lambda \tilde{\delta}_i \geq \sum_{i=1}^{N} \lambda_1(k) \tilde{\delta}_i' \tilde{\delta}_i = \lambda_1(k) \sum_{i=1}^{N} \tilde{\delta}_i' \tilde{\delta}_i,$$

where $\tilde{\delta}_i = S \delta_i$. Due to the fact $\tilde{\delta}_i' \tilde{\delta}_i = \delta_i' S' S \delta_i = \delta_i' \delta_i$, for large T and N, with probability tending to 1, we obtain

$$\sum_{i=1}^{N} J_{1i}(k) \geq \sum_{i=1}^{N} \delta_i' \left(\frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R \right) \delta_i \geq \lambda_1(k) \sum_{i=1}^{N} \delta_i' \delta_i = \lambda_1(k) \phi_N.$$

Thus, $\inf_{K(C_k)} \sum_{i=1}^{N} J_{1i}(k) \geq \lambda_1 \phi_N$.

Remark: Since $\delta_i' R'(X_{\Delta i}' X_{\Delta i})(X_{2i}' X_{2i})^{-1}(X_{0i}' X_{0i}) R \delta_i$ is positive and bounded, its sum over i seems of order $O_p(N)$. However, since the largest eigenvalue of $\frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R$ is bounded under Assumptions 4, 5, as in the proof of Lemma 1 above, it can be shown that $\sum_{i=1}^{N} \delta_i' \left(\frac{1}{N} \sum_{i=1}^{N} R'(X_{\Delta i}' X_{\Delta i}) R \right) \delta_i = O_p(\phi_N)$. Since Assumption 2 implies that ϕ_N could grow at a rate less than N, this allows for the possibility of no breaks in some series or small magnitude of breaks in slopes, as in Bai (2010).

Lemma 2 Under Assumptions 1-7, uniformly on $K(C_k)$,

(i) $\sum_{i=1}^N \delta_i' Z_{\Delta i}' \varepsilon_i = O_p(\sqrt{\phi_N})$;

(ii) $\frac{1}{\sqrt{T}} \sum_{i=1}^N \delta_i' Z_{\Delta i}' X_i (\frac{X_i'X_i}{T})^{-1} \frac{X_i'\varepsilon_i}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{T}})$;

(iii) $\frac{1}{\sqrt{T}} \sum_{i=1}^N \delta_i' (Z_{\Delta i}' M_i Z_{2i}) (\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} \frac{Z_{2i}'M_i\varepsilon_i}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{T}})$;

(iv) $\frac{1}{T} \sum_{i=1}^N \varepsilon_i' M_i Z_{\Delta i} (\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} Z_{\Delta i}' M_i \varepsilon_i = O_p(\frac{N}{T})$;

(v) $\frac{1}{T} \sum_{i=1}^N \varepsilon_i' M_i Z_{0i} (\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} Z_{\Delta i}' M_i \varepsilon_i = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}})$;

(vi) $\sum_{i=1}^N \frac{\varepsilon_i' M_i Z_{0i}}{\sqrt{T}} \left[(\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} - (\frac{Z_{0i}'M_iZ_{0i}}{T})^{-1} \right] \frac{Z_{0i}'M_i\varepsilon_i}{\sqrt{T}} = O_p(\frac{N}{T})$.

Proof of Lemma 2.(ii) and (iii) can be proved similarly as in part (i).

(iv) Since $M_i Z_{\Delta i} (Z_{2i}' M_i Z_{2i})^{-1} Z_{\Delta i}' M_i$ is positive semidefinite, $\varepsilon_i' M_i Z_{\Delta i} (Z_{2i}' M_i Z_{2i})^{-1} Z_{\Delta i}' M_i \varepsilon_i \geq 0$. For bounded $k - k_0$ on $K(C_k)$,

$$\frac{1}{N} \sum_{i=1}^N \left[\varepsilon_i' M_i Z_{\Delta i} (\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} Z_{\Delta i}' M_i \varepsilon_i \right] = O_p(1).$$

Thus, $\frac{1}{T} \sum_{i=1}^N \varepsilon_i' M_i Z_{\Delta i} (\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} Z_{\Delta i}' M_i \varepsilon_i = O_p(\frac{N}{T})$.

(v) For finite q, the $q \times q$ matrix $\frac{Z_{2i}'M_iZ_{2i}}{T} = O_p(1)$ for large T, thus the order of $\frac{1}{T} \sum_{i=1}^N \varepsilon_i' M_i Z_{0i} (\frac{Z_{2i}'M_iZ_{2i}}{T})^{-1} Z_{\Delta i}' M_i \varepsilon_i$ is same as that of $\frac{1}{T} \sum_{i=1}^N \varepsilon_i' M_i Z_{0i} Z_{\Delta i}' M_i \varepsilon_i$. Substituting $M_i = I - X_i (X_i'X_i)^{-1} X_i'$, we obtain

$$\begin{aligned}
\frac{1}{T} \sum_{i=1}^N \varepsilon_i' M_i Z_{0i} Z_{\Delta i}' M_i \varepsilon_i &= \frac{1}{T} \sum_{i=1}^N \varepsilon_i' Z_{0i} Z_{\Delta i}' \varepsilon_i - \frac{1}{T} \sum_{i=1}^N \varepsilon_i' X_i (X_i'X_i)^{-1} X_i' Z_{0i} Z_{\Delta i}' \varepsilon_i \\
&\quad - \frac{1}{T} \sum_{i=1}^N \varepsilon_i' Z_{0i} Z_{\Delta i}' X_i (X_i'X_i)^{-1} X_i' \varepsilon_i \\
&\quad + \frac{1}{T} \sum_{i=1}^N \varepsilon_i' X_i (X_i'X_i)^{-1} X_i' Z_{0i} Z_{\Delta i}' X_i (X_i'X_i)^{-1} X_i' \varepsilon_i.
\end{aligned}$$

Consider the first term $\frac{1}{T} \sum_{i=1}^N \varepsilon_i' Z_{0i} Z_{\Delta i}' \varepsilon_i$. If $k < k_0$, for bounded $k_0 - k$

$$\begin{aligned}
\frac{1}{T} \sum_{i=1}^N \varepsilon_i' Z_{0i} Z_{\Delta i}' \varepsilon_i &= \frac{1}{T} \sum_{i=1}^N (\sum_{t=k_0+1}^T \varepsilon_{it} z_{it}') (\sum_{\tau=k+1}^{k_0} \varepsilon_{i\tau} z_{i\tau}) \\
&= \frac{1}{T} \sum_{\tau=k+1}^{k_0} \sum_{i=1}^N \sum_{t=k_0+1}^T \varepsilon_{it} \varepsilon_{i\tau} z_{it}' z_{i\tau} = \frac{1}{T} O_p(\sqrt{NT}) = O_p(\sqrt{\frac{N}{T}}),
\end{aligned}$$

on $K(C_k)$. If $k > k_0$,

$$\begin{aligned}\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i' Z_{0i} Z_{\Delta i}'\varepsilon_i &= \frac{1}{T}\sum_{i=1}^{N}(\sum_{t=k_0+1}^{T}\varepsilon_{it} z_{it}')(\sum_{\tau=k_0+1}^{k}\varepsilon_{i\tau} z_{i\tau}) = \frac{1}{T}\sum_{i=1}^{N}\sum_{\tau=k_0+1}^{k}\sum_{t=k_0+1}^{T}\varepsilon_{it}\varepsilon_{i\tau} z_{it}' z_{i\tau} \\ &= \frac{1}{T}\sum_{\tau=k_0+1}^{k}\sum_{i=1}^{N}\varepsilon_{i\tau}^2 z_{i\tau}' z_{i\tau} + \frac{1}{T}\sum_{\tau=k_0+1}^{k}\sum_{t\neq\tau, t=k_0+1}^{T}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{i\tau} z_{it}' z_{i\tau} \\ &= O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}}),\end{aligned}$$

on $K(C_k)$.

Similarly, it can be shown that the second term $\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i' X_i (X_i'X_i)^{-1} X_i' Z_{0i} Z_{\Delta i}'\varepsilon_i = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}})$.

For bounded $|k-k_0|$ on $K(C_k)$, $Z_{\Delta i}' X_i = \sum_{t=k+1}^{k_0} z_{it} x_{it}'$ is bounded. Thus, As above, it is easy to show the third term

$$\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i' Z_{0i} Z_{\Delta i}' X_i (X_i'X_i)^{-1} X_i'\varepsilon_i = \frac{1}{T}\sum_{i=1}^{N}\frac{\varepsilon_i' Z_{0i}}{\sqrt{T}} Z_{\Delta i}' X_i (\frac{X_i'X_i}{T})^{-1}\frac{X_i'\varepsilon_i}{\sqrt{T}} = O_p(\frac{N}{T})$$

and the fourth term

$$\begin{aligned}&\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i' X_i (X_i'X_i)^{-1} X_i' Z_{0i} Z_{\Delta i}' X_i (X_i'X_i)^{-1} X_i'\varepsilon_i \\ =\ &\frac{1}{T}\sum_{i=1}^{N}\frac{\varepsilon_i' X_i}{\sqrt{T}}(\frac{X_i'X_i}{T})^{-1}\frac{X_i' Z_{0i}}{T} Z_{\Delta i}' X_i (\frac{X_i'X_i}{T})^{-1}\frac{X_i'\varepsilon_i}{\sqrt{T}} = O_p(\frac{N}{T}),\end{aligned}$$

on $K(C_k)$.

Therefore,

$$\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i' M_i Z_{0i} Z_{\Delta i}' M_i\varepsilon_i = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}}).$$

(vi) Since $|k-k_0|$ is bounded on $K(C_k)$,

$$\begin{aligned}&(\frac{Z_{2i}' M_i Z_{2i}}{T})^{-1} - (\frac{Z_{0i}' M_i Z_{0i}}{T})^{-1} \\ =\ &(\frac{Z_{0i}' M_i Z_{0i}}{T} + \frac{Z_{\Delta i}' M_i Z_{0i}}{T} + \frac{Z_{0i}' M_i Z_{\Delta i}}{T} + \frac{Z_{\Delta i}' M_i Z_{\Delta i}}{T})^{-1} - (\frac{Z_{0i}' M_i Z_{0i}}{T})^{-1} \\ =\ &(\frac{Z_{0i}' M_i Z_{0i}}{T} + O_p(\frac{1}{T}))^{-1} - (\frac{Z_{0i}' M_i Z_{0i}}{T})^{-1} \\ =\ &(\frac{Z_{0i}' M_i Z_{0i}}{T})^{-1}(1 + O_p(\frac{1}{T})) - (\frac{Z_{0i}' M_i Z_{0i}}{T})^{-1} = O_p(\frac{1}{T}),\end{aligned}$$

we obtain

$$\sum_{i=1}^{N} \frac{\varepsilon_i' M_i Z_{0i}}{\sqrt{T}} \left[(\frac{Z_{2i}' M_i Z_{2i}}{T})^{-1} - (\frac{Z_{0i}' M_i Z_{0i}}{T})^{-1} \right] \frac{Z_{0i}' M_i \varepsilon_i}{\sqrt{T}} = O_p(\frac{N}{T}).$$

Lemma 3 Under Assumptions 1, 2, 8-15 (or 16), for large N and T, with probability tending to 1,

$$\inf_{K(C_k)} \sum_{i=1}^{N} \tilde{J}_{1i}(k) \geq \tilde{\lambda}_1 \phi_N.$$

Proof of Lemma 3. By symmetry, we assume $k < k_0$. From Lemma A.1 of Bai (1997a),

$$\begin{aligned} \tilde{J}_{1i}(k) &= \delta_i'[(\tilde{Z}_{0i}{}'\tilde{M}_i\tilde{Z}_{0i}) - (\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{2i})(\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i})^{-1}(\tilde{Z}_{2i}{}'\tilde{M}_i\tilde{Z}_{0i})]\delta_i \\ &\geq \delta_i' R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})(\tilde{X}_{2i}'\tilde{X}_{2i})^{-1}(\tilde{X}_{0i}'\tilde{X}_{0i})R\delta_i. \end{aligned}$$

Plugging in $\tilde{X}_{2i}'\tilde{X}_{2i} = \tilde{X}_{0i}'\tilde{X}_{0i} + \tilde{X}_{\Delta i}'\tilde{X}_{\Delta i}$ gives

$$\begin{aligned} & \sum_{i=1}^{N} \delta_i' R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})(\tilde{X}_{2i}'\tilde{X}_{2i})^{-1}(\tilde{X}_{0i}'\tilde{X}_{0i})R\delta_i \\ = & \sum_{i=1}^{N} \delta_i' R'(X_{\Delta i}'X_{\Delta i})(\frac{\tilde{X}_{0i}'\tilde{X}_{0i}}{T} + \frac{\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i}}{T})^{-1}(\frac{\tilde{X}_{0i}'\tilde{X}_{0i}}{T})R\delta_i \\ = & \sum_{i=1}^{N} \delta_i' R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\delta_i + o_p(1). \end{aligned}$$

Given that $\tilde{k} - k_0$ is stochastically bounded on $K(C_k)$, $\frac{1}{T}\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i} = o_p(1)$ for a large T on $K(C_k)$. Thus, the term above is dominated by $\sum_{i=1}^{N} \delta_i' R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\delta_i$. Denote $a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R)$ the $(j,l)^{th}$ element of matrix $R'(X_{\Delta i}'X_{\Delta i})R$.

$$\begin{aligned} \sum_{i=1}^{N} \delta_i' R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\delta_i &= \sum_{i=1}^{N} \left[\sum_{j=1}^{p}\sum_{l=1}^{p} \delta_{i,j}\delta_{i,l} \cdot a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R) \right] \\ &= \sum_{j=1}^{p}\sum_{l=1}^{p} \left[\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l} \cdot a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R) \right]. \end{aligned}$$

By Assumption 15 that δ_i is independent of X_i, $\delta_{i,j}\delta_{i,l}$ is independent of $a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R)$. Thus,

$$E[(\delta_{i,j}\delta_{i,l}) \cdot a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R)] = E[(\delta_{i,j}\delta_{i,l})]E[a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R)],$$

implying that for a large N, with probability tending to 1,

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l} \cdot a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R) &= \left(\frac{1}{N}\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l} \right) \left(\frac{1}{N}\sum_{i=1}^{N} a_{j,l}(R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R) \right) \\ &= \left(\frac{1}{N}\sum_{i=1}^{N} \delta_{i,j}\delta_{i,l} \right) a_{j,l} \left(\frac{1}{N}\sum_{i=1}^{N} R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R \right). \end{aligned}$$

It follows that

$$\begin{aligned}\sum_{i=1}^{N}\delta_i' R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\delta_i &= N\sum_{j=1}^{p}\sum_{l=1}^{p}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{i,j}\delta_{i,l}\right)a_{j,l}\left(\frac{1}{N}\sum_{i=1}^{N}R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\right)\right]\\ &= \sum_{i=1}^{N}\sum_{j=1}^{p}\sum_{l=1}^{p}\left[\delta_{i,j}\delta_{i,l}\cdot a_{j,l}\left(\frac{1}{N}\sum_{i=1}^{N}R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\right)\right]\\ &= \sum_{i=1}^{N}\delta_i'\left(\frac{1}{N}\sum_{i=1}^{N}R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\right)\delta_i\end{aligned}$$

with probability tending to 1 for large N. Therefore, similar to the proof of Lemma 1,

$$\sum_{i=1}^{N}\tilde{J}_{1i}(k) \geq \sum_{i=1}^{N}\delta_i'\left(\frac{1}{N}\sum_{i=1}^{N}R'(\tilde{X}_{\Delta i}'\tilde{X}_{\Delta i})R\right)\delta_i \geq \sum_{i=1}^{N}\tilde{\lambda}_1(k)\delta_i'\delta_i = \tilde{\lambda}_1(k)\phi_N,$$

thus, $\inf_{K(C_k)}\sum_{i=1}^{N}\tilde{J}_{1i}(k) \geq \tilde{\lambda}_1\phi_N$.

Lemma 7 Under Assumptions 1, 2, 8-15 (or 16), uniformly on $K(C_k)$,

(i) $\sum_{i=1}^N \delta_i' \tilde{Z}_{\Delta i}' M_w F \gamma_i = O_p(\sqrt{\phi_N})$;

(ii) $\frac{1}{T}\sum_{i=1}^N \delta_i' \tilde{Z}_{\Delta i}' \tilde{X}_i (\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1} \tilde{X}_i' M_w F \gamma_i = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iii) $\frac{1}{T}\sum_{i=1}^N \delta_i' (\tilde{Z}_{\Delta i}' \tilde{M}_i \tilde{Z}_{2i}) (\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} \tilde{Z}_{2i}' \tilde{M}_i M_w F \gamma_i = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iv)a. $\frac{1}{T}\sum_{i=1}^N \tilde{\varepsilon}_i' \tilde{M}_i \tilde{Z}_{\Delta i} (\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} \tilde{Z}_{\Delta i}' \tilde{M}_i M_w F \gamma_i = O_p(\frac{1}{T})$;

b. $\frac{1}{T}\sum_{i=1}^N \gamma_i' F' M_w' \tilde{M}_i \tilde{Z}_{\Delta i} (\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} \tilde{Z}_{\Delta i}' \tilde{M}_i M_w F \gamma_i = O_p(\frac{1}{T})$;

(v)a. $\frac{1}{T}\sum_{i=1}^N \tilde{\varepsilon}_i' \tilde{M}_i \tilde{Z}_{0i} (\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} \tilde{Z}_{\Delta i}' \tilde{M}_i M_w F \gamma_i = O_p(\frac{1}{\sqrt{T}})$;

b. $\frac{1}{T}\sum_{i=1}^N \gamma_i' F' M_w \tilde{M}_i \tilde{Z}_{0i} (\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} \tilde{Z}_{\Delta i}' \tilde{M}_i \tilde{\varepsilon}_i = O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}})$;

c. $\frac{1}{T}\sum_{i=1}^N \gamma_i' F' M_w \tilde{M}_i \tilde{Z}_{0i} (\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} \tilde{Z}_{\Delta i}' \tilde{M}_i M_w F \gamma_i = O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}})$;

(vi)a. $\frac{1}{\sqrt{T}}\sum_{i=1}^N \frac{\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}} \left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1} \right] \tilde{Z}_{0i}' \tilde{M}_i M_w F \gamma_i = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$;

b. $\frac{1}{T}\sum_{i=1}^N \gamma_i' F' M_w \tilde{M}_i \tilde{Z}_{0i} \left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1} \right] \tilde{Z}_{0i}' \tilde{M}_i M_w F \gamma_i = O_p(\frac{1}{N}) + O_p(\frac{1}{T}) + O_p(\frac{1}{\sqrt{NT}})$.

Proof of Lemma 7.

(ii) By Lemmas 5 (ii) and 4 (vii),

$$\begin{aligned}
\frac{1}{T}\tilde{X}_i' M_w F \gamma_i &= \frac{1}{T} X_i' M_w F \gamma_i = \frac{X_i'\mathbb{F}}{T} D_1 \gamma_i + \frac{X_i'\bar{\mathbb{U}}}{T} D_2 \gamma_i \\
&= O_p(1)[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})] + [O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})] O_p(1) \\
&= O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}).
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{1}{T}\sum_{i=1}^N \delta_i' \tilde{Z}_{\Delta i}' \tilde{X}_i (\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1} \tilde{X}_i' M_w F \gamma_i &= \sum_{i=1}^N \delta_i' O_p(1)[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})] \\
&= O_p(\sqrt{\phi_N}) O_p(\sqrt{N})[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})] \\
&= O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}}).
\end{aligned}$$

(iii) can be proved in the same way as (i) and (ii).

(iv) Since

$$\tilde{Z}_{\Delta i}' \tilde{M}_i M_w F \gamma_i = \tilde{Z}_{\Delta i}' \tilde{M}_i \mathbb{F} D_1 \gamma_i + \tilde{Z}_{\Delta i}' \tilde{M}_i \bar{\mathbb{U}} D_2 \gamma_i = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{\sqrt{N}}) = O_p(\frac{1}{\sqrt{N}}),$$

we obtain ((iv)a)

$$\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = \frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i}O_p(1)O_p(\frac{1}{\sqrt{N}})$$
$$= \frac{1}{T}O_p(\frac{1}{\sqrt{N}})O_p(\sqrt{N}) = O_p(\frac{1}{T}).$$

$\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i}$ is the sum of finite elements on $K(C_k)$, so $\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i} = O_p(\sqrt{N})$. In addition, ((iv)c)

$$\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = \frac{1}{T}\sum_{i=1}^{N}O_p(\frac{1}{\sqrt{N}})O_p(1)O_p(\frac{1}{\sqrt{N}}) = O_p(\frac{1}{T}).$$

(v) Since $\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{\sqrt{N}})$ and $\frac{1}{T}\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i} = O_p(1)$, we obtain ((v)a)

$$\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = \frac{1}{T}O_p(\sqrt{NT})O_p(\frac{1}{\sqrt{N}}) = O_p(\frac{1}{\sqrt{T}}).$$

Similarly,

$$\tilde{Z}_{0i}'\tilde{M}_iM_wF\gamma_i = \tilde{Z}_{0i}'\tilde{M}_i\mathbb{F}D_1\gamma_i + \tilde{Z}_{0i}'\tilde{M}_i\bar{\mathbb{U}}D_2\gamma_i = T\frac{\tilde{Z}_{0i}'\tilde{M}_i\mathbb{F}}{T}D_1\gamma_i + T\frac{\tilde{Z}_{0i}'\tilde{M}_i\bar{\mathbb{U}}}{T}D_2\gamma_i$$
$$= T\left(O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\right) + T\left(O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\right)$$
$$= O_p(\frac{T}{N}) + O_p(\sqrt{\frac{T}{N}}).$$

It follows that ((v)b)

$$\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i = \frac{1}{T}(O_p(\frac{T}{N}) + O_p(\sqrt{\frac{T}{N}}))O_p(\sqrt{N})$$
$$= O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}})$$

and ((v)c),

$$\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i = O_p(\frac{1}{\sqrt{N}})\frac{1}{T}\left(O_p(\frac{T}{N}) + O_p(\sqrt{\frac{T}{N}})\right)O_p(N)$$
$$= O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}}).$$

(vi) Since $\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right] = O_p(\frac{1}{T})$ and $Z_{0i}'\tilde{M}_i M_w F\gamma_i = O_p(\frac{T}{N}) + O_p(\sqrt{\frac{T}{N}})$, we obtain ((vi)a)

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\tilde{Z}_{0i}'\tilde{M}_i M_w F\gamma_i \\
= \quad & \frac{1}{\sqrt{T}}O_p(\frac{1}{T})[O_p(\frac{T}{N}) + O_p(\sqrt{\frac{T}{N}})]O_p(\sqrt{N}) = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}).
\end{aligned}$$

Similarly, (vi)b

$$\begin{aligned}
&\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\tilde{Z}_{0i}'\tilde{M}_i M_w F\gamma_i \\
= \quad & \frac{1}{T}O_p(\frac{1}{T})[O_p(\frac{T}{N}) + O_p(\sqrt{\frac{T}{N}})][O_p(\frac{T}{N}) + O_p(\sqrt{\frac{T}{N}})]O_p(N) \\
= \quad & O_p(\frac{1}{N}) + O_p(\frac{1}{T}) + O_p(\frac{1}{\sqrt{NT}}).
\end{aligned}$$

Lemma 9. Under Assumptions 1, 2, 8-15 (or 16), uniformly on $K(C_k)$,

(i) $\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{\varepsilon}_i^0 = O_p(\sqrt{\phi_N})$;

(ii) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\frac{\tilde{X}_i'\tilde{\varepsilon}_i^0}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iii) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0}{\sqrt{T}} = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$;

(iv) $\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 = O_p(\frac{N}{T})$;

(v) $\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 = O_p(\frac{N}{T}) + O_p(\sqrt{\frac{N}{T}}) + O_p(\frac{1}{\sqrt{N}})$;

(vi) $\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}\left[(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{Z}_{0i}}{T})^{-1}\right]\frac{\tilde{Z}_{0i}'\tilde{M}_i\tilde{\varepsilon}_i^0}{\sqrt{T}} = O_p(\frac{N}{T}) + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$.

Proof of Lemma 9.

(ii) Since

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\frac{\tilde{X}_i'\tilde{\varepsilon}_i^0}{\sqrt{T}} \\
= &\frac{1}{T}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\tilde{X}_i'M_wF\gamma_i + \frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\frac{\tilde{X}_i'\tilde{\varepsilon}_i}{\sqrt{T}},
\end{aligned}$$

combining Lemma 7 (ii) and Lemma 8 (ii), we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\tilde{Z}_{\Delta i}'\tilde{X}_i(\frac{\tilde{X}_i'\tilde{X}_i}{T})^{-1}\frac{\tilde{X}_i'\tilde{\varepsilon}_i^0}{\sqrt{T}} \\
= &O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}}) + O_p(\sqrt{\frac{\phi_N}{T}}) = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}}).
\end{aligned}$$

(iii)

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\left[(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0}{\sqrt{T}}\right] \\
= &\frac{1}{T}\sum_{i=1}^{N}\delta_i'\left[(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{2i}'\tilde{M}_iM_wF\gamma_i\right] \\
&+\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\left[(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i}{\sqrt{T}}\right]
\end{aligned}$$

Lemma 7(iii) shows that the order of first term is $O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}})$. By Lemma 8(iii), the second term is $O_p(\sqrt{\frac{\phi_N}{T}})$. Thus,

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\delta_i'\left[(\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{Z}_{2i})(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{\varepsilon}_i^0}{\sqrt{T}}\right] = O_p(\sqrt{\frac{\phi_N}{N}}) + O_p(\sqrt{\frac{\phi_N}{T}}).$$

(iv) Plugging in $\tilde{\varepsilon}_i^0 = M_w F\gamma_i + \tilde{\varepsilon}_i$ yields

$$\begin{aligned}
&\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 \\
= &\frac{1}{T}\sum_{i=1}^{N}(M_wF\gamma_i+\tilde{\varepsilon}_i)'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i(M_wF\gamma_i+\tilde{\varepsilon}_i) \\
= &\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i+\frac{2}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i \\
&+\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w'\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i
\end{aligned}$$

Lemma 8(iv) shows that the first term is $O_p(\frac{N}{T})$. In addition, Lemma 7(iv) shows that the second term and the third term are $O_p(\frac{1}{T})$. Therefore,

$$\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{\Delta i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 = O_p(\frac{N}{T})+O_p(\frac{1}{T})+O_p(\frac{1}{T}) = O_p(\frac{N}{T}).$$

(v) can be proved similarly. Plug in $\tilde{\varepsilon}_i^0 = M_wF\gamma_i+\tilde{\varepsilon}_i$,

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 \\
= &\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i+\frac{1}{T}\sum_{i=1}^{N}\tilde{\varepsilon}_i'\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i \\
&+\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i+\frac{1}{T}\sum_{i=1}^{N}\gamma_i'F'M_w\tilde{M}_i\tilde{Z}_{0i}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_iM_wF\gamma_i.
\end{aligned}$$

By Lemma 8(v) and Lemma 7(v), we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\frac{\tilde{\varepsilon}_i^{0\prime}\tilde{M}_i\tilde{Z}_{0i}}{\sqrt{T}}(\frac{\tilde{Z}_{2i}'\tilde{M}_i\tilde{Z}_{2i}}{T})^{-1}\tilde{Z}_{\Delta i}'\tilde{M}_i\tilde{\varepsilon}_i^0 \\
= &O_p(\frac{N}{T})+O_p(\sqrt{\frac{N}{T}})+O_p(\frac{1}{\sqrt{T}})+O_p(\frac{1}{\sqrt{N}})+O_p(\frac{1}{\sqrt{T}})+O_p(\frac{1}{\sqrt{N}})+O_p(\frac{1}{\sqrt{T}}) \\
= &O_p(\frac{N}{T})+O_p(\sqrt{\frac{N}{T}})+O_p(\frac{1}{\sqrt{N}}).
\end{aligned}$$

(vi) Similarly,

$$
\begin{aligned}
& \sum_{i=1}^{N} \frac{\tilde{\varepsilon}_i^{0\prime} \tilde{M}_i \tilde{Z}_{0i}}{\sqrt{T}} \left[(\frac{\tilde{Z}_{2i}' \tilde{M}_i \tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{Z}_{0i}}{T})^{-1} \right] \frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{\varepsilon}_i^0}{\sqrt{T}} \\
= & \sum_{i=1}^{N} \frac{(M_w F \gamma_i + \tilde{\varepsilon}_i)' \tilde{M}_i \tilde{Z}_{0i}}{\sqrt{T}} \left[(\frac{\tilde{Z}_{2i}' \tilde{M}_i \tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{Z}_{0i}}{T})^{-1} \right] \frac{\tilde{Z}_{0i}' \tilde{M}_i (M_w F \gamma_i + \tilde{\varepsilon}_i)}{\sqrt{T}} \\
= & \sum_{i=1}^{N} \frac{\tilde{\varepsilon}_i' \tilde{M}_i \tilde{Z}_{0i}}{\sqrt{T}} \left[(\frac{\tilde{Z}_{2i}' \tilde{M}_i \tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{Z}_{0i}}{T})^{-1} \right] \frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{\varepsilon}_i}{\sqrt{T}} \\
& + \frac{2}{\sqrt{T}} \sum_{i=1}^{N} \frac{\tilde{\varepsilon}_i' \tilde{M}_i \tilde{Z}_{0i}}{\sqrt{T}} \left[(\frac{\tilde{Z}_{2i}' \tilde{M}_i \tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{Z}_{0i}}{T})^{-1} \right] \tilde{Z}_{0i}' \tilde{M}_i M_w F \gamma_i \\
& + \frac{1}{T} \sum_{i=1}^{N} \gamma_i' F' M_w \tilde{M}_i \tilde{Z}_{0i} \left[(\frac{\tilde{Z}_{2i}' \tilde{M}_i \tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{Z}_{0i}}{T})^{-1} \right] \tilde{Z}_{0i}' \tilde{M}_i M_w F \gamma_i .
\end{aligned}
$$

Lemma 8 (vi) and Lemma 7 (vi) show that these three terms are $O_p(\frac{N}{T})$, $O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$ and $O_p(\frac{1}{N}) + O_p(\frac{1}{T}) + O_p(\frac{1}{\sqrt{NT}})$, respectively. Therefore,

$$
\sum_{i=1}^{N} \frac{\tilde{\varepsilon}_i^{0\prime} \tilde{M}_i \tilde{Z}_{0i}}{\sqrt{T}} \left[(\frac{\tilde{Z}_{2i}' \tilde{M}_i \tilde{Z}_{2i}}{T})^{-1} - (\frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{Z}_{0i}}{T})^{-1} \right] \frac{\tilde{Z}_{0i}' \tilde{M}_i \tilde{\varepsilon}_i^0}{\sqrt{T}} = O_p(\frac{N}{T}) + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}).
$$