Estimation and Inference of Change Points in High-Dimensional Factor Models*

Jushan Bai †
Columbia University

Xu Han ‡
City University of Hong Kong

Yutang Shi §
Texas A&M University

August, 2019

Abstract

In this paper, we consider the estimation of break points in high-dimensional factor models where the unobserved factors are estimated by principal component analysis (PCA). The factor loading matrix is assumed to have a structural break at an unknown time. We establish the conditions under which the least squares (LS) estimator is consistent for the break date. Our consistency result holds for both large and small breaks. We also find the LS estimator's asymptotic distribution. Simulation results confirm that the break date can be accurately estimated by the LS even if the magnitudes of breaks are small. In two empirical applications, we implement the method to estimate break points in the U.S. stock market and U.S. macroeconomy, respectively.

Keywords: structural changes, high-dimensional factor models, break point inference
JEL Classification: C12, C22, C38

*The authors would like to thank co-editor Jianqing Fan, an associate editor, and three anonymous referees for their insightful suggestions. We are also grateful to Mehmet Caner, Jia Li, Liangjun Su, Peng Wang, Jun Yu, Xinghua Zheng, and participants at the 2017 Asian Meeting of the Econometric Society, the First International Conference on Econometrics and Statistics, and the Tenth Annual SoFiE Conference for their helpful comments. Xu Han would like to acknowledge the financial support by the start-up grant from City University of Hong Kong.

†Department of Economics, Columbia University, New York, NY 10027, U.S.A. E-mail: jb3064@columbia.edu
‡Department of Economics and Finance, City University of Hong Kong, Hong Kong S.A.R. 999077, China. E-mail: xuhan25@cityu.edu.hk
§Department of Economics, Texas A&M University, College Station, TX 77843, U.S.A. yshi@tamu.edu

1 Introduction

High-dimensional factor models assume that a small number of factors can capture the common driving forces of a large number of economic variables. This method of dimension reduction is a powerful statistical tool and has been found useful in forecasting (Stock and Watson, 2002), structural factor-augmented VAR analysis (Bernanke, Boivin, and Eliasz, 2005), reducing the number of instruments (Bai and Ng, 2010), and constructing dynamic stochastic general equilibrium (DSGE) models (Boivin and Giannoni, 2006). The method has also been applied to estimate high-dimensional variance-covariance matrices (Fan, Liao and Mincheva, 2011, 2013) and to measure the average treatment effects of policy interventions (Hsiao, Ching and Wan, 2012). While factor models are useful, practitioners have to be cautious about potential structural changes in high-dimensional data sets. This concern is empirically relevant because parameter instability is a pervasive phenomenon in time series data (Stock and Watson, 1996; see also Banerjee and Marcellino, 2008 and Yamamoto, 2016).

Recent studies have developed tests for structural changes in factor models (Stock and Watson, 2008; Breitung and Eickmeier, 2011; Chen et al. , 2014; Corradi and Swanson, 2014; Han and Inoue, 2015; Tanaka and Yamamoto, 2015; and Su and Wang, 2015). Most of these studies focus on testing whether the factor loadings have a structural break at a common date.[1] The rejection of the null hypothesis of no structural breaks naturally leads to the next question of when the break occurred. Estimating the break point in factor models is more challenging than in standard time series regressions because the factors are not observed. Even the consistency of the estimated break fraction (i.e., the break date divided by sample size) was not established until recently. Cheng et al. (2016) develop a shrinkage method that can consistently estimate the break fraction. Chen (2015) considers a least squares (LS) estimator of break points and proves the consistency of estimated break fractions. Baltagi, Kao, and Wang (2017) propose an estimator and show that the distance between the estimated and true break dates is stochastically bounded. For the consistency of the estimated break point, Massacci (2017) studies the LS estimation of structural changes in factor loadings under a threshold model setup. His convergence rate of the estimated break fraction implies the consistency of the estimated break date. In addition, Ma and Su (2018) develop an adaptive fused group Lasso method to consistently estimate all break points under a multi-break setup. Yet, almost all of these papers consider large breaks, and none of them establishes the limiting distribution of the estimated break point.

In an influential study in this area, Bates et al. (2013) show that when the magnitudes of breaks are small, the estimated factors and factor loadings have the same rate of convergence as the case without breaks, even when the breaks are ignored. This result may have the unintended consequence of making researchers think they do not have to explore small changes. However, there are several advantages in studying small breaks. One is that actual changes may indeed be small, or

[1]Su and Wang (2015) do not assume a common break date; instead, they consider smooth changes in the loadings. See Bai and Han (2016) for a detailed survey on these tests.

that only a small portion of series have undergone changes. Second, under large breaks, asymptotic theory implies an exact estimation of the break points as both N and T (the cross-section and time dimensions) go to infinity. In practice, we have only moderate sample sizes, so exact estimation does not hold. Small breaks imply randomness of the estimated breaks even in the limit, providing a way to characterize the uncertainty about the estimated break points.

In this paper, we contribute to the literature mainly in two ways. First, we establish the consistency of the break date estimated by the LS under both large and small breaks. We consider two types of small breaks: (1) the magnitude of the change in each variable's factor loadings is of order $N^{\frac{\alpha-1}{2}}$ for some $0 < \alpha \leq 1$, where N is the total number of variables, and (2) the magnitude of the change is fixed but only $O(N^{\alpha})$ variables have structural changes for some $0 < \alpha \leq 1$. When $\alpha = 1$, these two setups coincide. Almost all existing studies require $\alpha = 1$ to establish the consistency of either the break fraction or the break date. An exception is the work by Massacci (2017), who proves the consistency of the break date estimated by the LS when $O(N^{\alpha})$ many variables have breaks for $0.5 < \alpha \leq 1$. In contrast, we establish the consistency of the estimated break date under both cases (1) and (2) for $0 < \alpha \leq 1$ if $N^{1-\alpha} \log\log T/T \to 0$ and $\log\log(NT)/\min(N,T) \to 0$ as $N, T \to \infty$. This is a much stronger result than those in the literature. It implies that the LS estimator can accurately estimate the break date even when either the break magnitude or the number of variables with breaks is small.

The second major contribution of this paper is that we establish the asymptotic distribution of the break date estimated by LS. We show that the difference between the estimated and true break dates is $O_p(1)$ for $\alpha = 0$ in both cases (1) and (2) addressed in the previous paragraph. In addition, we derive the asymptotic distribution of the estimated break date under case (1) for $\alpha = 0$. The asymptotic distribution would be the same as that obtained by Bai (2010) if the factors were always equal to one for all periods. In general, the distribution of the estimated break date depends on the generating processes of the unobserved factors. Thus, we propose a bootstrap procedure to construct confidence intervals for the estimated break date. The simulation results show that our bootstrap method has good coverage probabilities in finite samples.

To establish the consistency and distribution results, we have to resolve the challenge that factors are estimated by principal components (PCs) rather than observed. Note that Bai (2010) also establishes the consistency and asymptotic distribution of the LS break date estimator with observed regressors. However, the proofs become much more difficult with estimated factors for several reasons. First, when the breaks are large, comparing the sums of squared residuals computed using estimated factors is asymptotically equivalent to comparing those computed based on the true factors. This trick is commonly used in the literature (e.g., Chen, 2015). However, it does not work when the breaks are small. The detailed technical reason is explained in the first two paragraphs of Section 3.2. Second, the factors are estimated by subsample PCs given a potential break date k, so the estimated factors depend on not only time index t but also the split date k. In contrast, the values of the observed regressors in Bai (2010) depend only on time index t regardless of the value of k. Thus, it is more difficult to compare the sums of squared residuals at different split dates with estimated factors. In addition, when the split date is not equal to the true break date, the structural change affects the properties of the factors estimated from the subsample that contains the break point. Lastly, the PC estimation can only identify the factors up to some rotation, which again depends on the split date. In this paper, we resolve all of these challenges and substantially generalize Bai's (2010) results.

In addition to Bai (2010), other researchers study the change point estimation for large panels. Kim (2011) generalizes Bai's (2010) result by allowing structural changes in both deterministic trends and means. Kim (2014) generalizes Kim (2011) by allowing a factor structure in the error terms of the large panel. Baltagi, Kao, and Liu (2017) extend Bai's (2010) work by allowing nonstationary regressors and error terms. Baltagi, Feng, and Kao (2016) show the consistency of the LS break date estimator in Pesaran's (2006) common correlated effects (CCE) model with structural changes on regression coefficients. Baltagi, Kao, and Wang (2015) allow structural breaks in a heterogeneous large panel with interactive fixed effects (Bai, 2009). They obtain the consistency of the estimated break fraction and break date under some conditions. Li et al. (2016) consider the same interactive fixed-effect panel but with multiple breaks on the regression coefficients. To the best of our knowledge, although some studies allow for an unobserved factor structure in large panels, they focus only on structural changes in the coefficients on the observed regressors, and do not consider break point estimation for the factor loadings.

This paper is organized as follows. Section 2 sets up the factor model with a single break on the factor loading matrix and describes the LS estimator for the break date. Section 3 provides the assumptions and establishes the consistency and asymptotic distribution of the break date estimator. A bootstrap procedure for constructing the confidence intervals is also proposed. Section 4 conducts Monte Carlo simulations and shows that the estimator performs well in finite samples. Section 5 presents two empirical applications. Section 6 concludes the paper.

2 High-Dimensional Factor Models with Structural Breaks

Consider the model for $i = 1, ..., N$,

$$x_{it} = \begin{cases} \lambda_{i1}' f_t + e_{it} & \text{for } t = 1, 2, ..., k_0(T) \\ \lambda_{i2}' f_t + e_{it} & \text{for } t = k_0 + 1, k_0(T) + 2, ..., T, \end{cases} \tag{2.1}$$

where f_t is the r-dimensional vector of unobserved common factors, $k_0(T)$ is the unknown break date, λ_{i1} and λ_{i2} are the pre- and post-break factor loadings, respectively, and e_{it} is the idiosyncratic error. Let $x_t = [x_{1t}, ..., x_{Nt}]'$, $e_t = [e_{1t}, ..., e_{Nt}]'$, $\Lambda_1 = [\lambda_{11}, ..., \lambda_{N1}]'$, and $\Lambda_2 = [\lambda_{12}, ..., \lambda_{N2}]'$. The vector representation of (2.1) is

$$x_t = \begin{cases} \Lambda_1 f_t + e_t & \text{for } t = 1, 2, ..., k_0(T) \\ \Lambda_2 f_t + e_t & \text{for } t = k_0 + 1, k_0(T) + 2, ..., T. \end{cases} \tag{2.2}$$

In our asymptotic setup, we set $k_0(T)$ as a sequence depending on T such that

$$k_0(T) = [T\tau_0], \tag{2.3}$$

where $\tau_0 \in (0,1)$ is a fixed constant and $[.]$ denotes the integer part of a real number. Hence, the ratio $k_0(T)/T$ is held constant as $T \to \infty$. The theory in this paper is developed under (2.3), which is commonly used in the structural break literature (e.g., Bai, 2010). The dependence of k_0 on T is suppressed for notational simplicity in the rest of the paper, unless otherwise specified.

For any given $k = 1, ..., T-1$, we define $X_k^{(1)} = [x_1, x_2, ..., x_k]'$ and $X_k^{(2)} = [x_{k+1}, x_{k+2}, ..., x_T]'$. Thus, the subscript k denotes the date to split the sample, and the superscripts (1) and (2) denote the pre- and post-k data, respectively. Similarly, we define

$$\begin{aligned} &F_k^{(1)} = [f_1, ..., f_k]',\ F_k^{(2)} = [f_{k+1}, ..., f_T]', \\ &\mathbf{e}_k^{(1)} = [e_1, ..., e_k]',\ \mathbf{e}_k^{(2)} = [e_{k+1}, ..., e_T]'. \end{aligned} \tag{2.4}$$

To simplify the notation, we use $F^{(1)}$ and $F^{(2)}$ to denote $F_{k_0}^{(1)}$ and $F_{k_0}^{(2)}$, respectively. Hence, (2.1) can be rewritten in the following matrix format

$$\begin{aligned} \begin{bmatrix} X_{k_0}^{(1)} \\ X_{k_0}^{(2)} \end{bmatrix} &= \begin{bmatrix} F^{(1)} & 0_{k_0 \times r} \\ 0_{(T-k_0)\times r} & F^{(2)} \end{bmatrix} \begin{bmatrix} \Lambda_1' \\ \Lambda_2' \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{k_0}^{(1)} \\ \mathbf{e}_{k_0}^{(2)} \end{bmatrix}, \\ &= G\Theta' + \mathbf{e} \end{aligned} \tag{2.5}$$

where

$$G = \begin{bmatrix} F^{(1)} & 0_{k_0 \times r} \\ 0_{(T-k_0)\times r} & F^{(2)} \end{bmatrix},\ \Theta = [\Lambda_1, \Lambda_2],\ \text{and } \mathbf{e} = \begin{bmatrix} \mathbf{e}_{k_0}^{(1)} \\ \mathbf{e}_{k_0}^{(2)} \end{bmatrix}.$$

Thus, (2.5) is an observationally equivalent factor model, where the number of factors is doubled and the factor loadings are time invariant. In this setup, the factor process f_t can be viewed as having a structural change. Baltagi, Kao, and Wang (2017) use this view to estimate the break date.

2.1 Estimation of the break point

Let $\tau_1 > 0$ and $\tau_2 < 1$ be a priori lower and upper bounds for τ_0, so that $0 < \tau_1 \le \tau_0 \le \tau_2 < 1$. The estimation of the break point involves the estimation of unobserved factors. Consider a given date $\tau_1 T \le k \le \tau_2 T$. The factors are estimated via PCs. Let $\tilde{F}_k^{(1)}$ denote $\sqrt{k}$ times the first r eigenvectors of $X_k^{(1)} X_k^{(1)\prime}$ and $\tilde{F}_k^{(2)}$ denote $\sqrt{T-k}$ times the first r eigenvectors of $X_k^{(2)} X_k^{(2)\prime}$, so we have

$$\frac{1}{kN} X_k^{(1)} X_k^{(1)\prime} \tilde{F}_k^{(1)} = \tilde{F}_k^{(1)} \tilde{V}_k^{(1)},$$

$$\frac{1}{(T-k)N} X_k^{(2)} X_k^{(2)\prime} \tilde{F}_k^{(2)} = \tilde{F}_k^{(2)} \tilde{V}_k^{(2)}, \tag{2.6}$$

where $\tilde{V}_k^{(1)}$ and $\tilde{V}_k^{(2)}$ are $r \times r$ diagonal matrices consisting of the first r eigenvalues of $X_k^{(1)} X_k^{(1)\prime}/kN$ and $X_k^{(1)} X_k^{(1)\prime}/(T-k)N$, respectively.

If $k = k_0$, we define

$$\begin{aligned} \hat{V}^{(1)} &\equiv \tilde{V}_{k_0}^{(1)}, \quad \hat{F}^{(1)} \equiv \tilde{F}_{k_0}^{(1)}, \\ \hat{V}^{(2)} &\equiv \tilde{V}_{k_0}^{(2)}, \quad \hat{F}^{(2)} \equiv \tilde{F}_{k_0}^{(2)}, \end{aligned} \tag{2.7}$$

so that the "hat" denotes the case where the break date is correctly specified. Let $\tilde{f}_t$ and $\hat{f}_t$ denote the transpose of the t-th row of $\tilde{F} = \left[\tilde{F}_k^{(1)\prime}, \tilde{F}_k^{(2)\prime}\right]'$ and $\hat{F} = \left[\hat{F}^{(1)\prime}, \hat{F}^{(2)\prime}\right]'$, respectively. Given the estimator $\tilde{f}_t$, the pre- and post-k factor loading matrices can be estimated using OLS, i.e.,

$$\tilde{\lambda}_{i1} = \frac{\tilde{F}_k^{(1)\prime} X_{k,i}^{(1)}}{k}, \quad \tilde{\lambda}_{i2} = \frac{\tilde{F}_k^{(2)\prime} X_{k,i}^{(2)}}{T-k}, \tag{2.8}$$

where $X_{k,i}^{(1)} = [x_{i1}, x_{i2}, ..., x_{ik}]'$, $X_{k,i}^{(2)} = [x_{ik+1}, x_{ik+2}, ..., x_{iT}]'$. It is important to note that $\tilde{f}_t$, $\tilde{\lambda}_{i1}$, and $\tilde{\lambda}_{i2}$ depend on k; however, for notational simplicity, their dependence on k is suppressed. For a given k, define the sum of squared residuals as

$$SSR(k, \tilde{F}) \equiv \sum_{i=1}^{N} \sum_{t=1}^{k} \left(x_{it} - \tilde{f}_t' \tilde{\lambda}_{i1}\right)^2 + \sum_{i=1}^{N} \sum_{t=k+1}^{T} \left(x_{it} - \tilde{f}_t' \tilde{\lambda}_{i2}\right)^2. \tag{2.9}$$

The estimated break date is given by

$$\tilde{k} = \arg \min_{\tau_1 T \le k \le \tau_2 T} SSR(k, \tilde{F}). \tag{2.10}$$

The estimation of $\tilde{k}$ jointly utilizes the information from all N variables. We expect that $\tilde{k}$ benefits from the large N setup and is more accurate than estimators obtained using information from a fixed number of cross-sections.

When $k = k_0$, we use $\hat{\lambda}_{i1}$ and $\hat{\lambda}_{i2}$ to denote the estimated pre- and post-break factor loadings, i.e.,

$$\hat{\lambda}_{i1} = \frac{\hat{F}^{(1)\prime} X_{k_0,i}^{(1)}}{k_0}, \quad \hat{\lambda}_{i2} = \frac{\hat{F}^{(2)\prime} X_{k_0,i}^{(2)}}{T-k_0}, \tag{2.11}$$

so the sum of squared residuals for $k = k_0$ is defined as

$$SSR(k_0, \hat{F}) = \sum_{i=1}^{N} \sum_{t=1}^{k_0} \left(x_{it} - \hat{f}_t' \hat{\lambda}_{i1}\right)^2 + \sum_{i=1}^{N} \sum_{t=k_0+1}^{T} \left(x_{it} - \hat{f}_t' \hat{\lambda}_{i2}\right)^2.$$

These notations are useful for analyzing the asymptotic properties of $\tilde{k}$.

In a strict factor model where the error terms are i.i.d. Gaussian with cross-sectionally homogeneous variances, minimizing the sum of squared residuals follows from the maximum likelihood principle. Under our approximate factor model setup with weakly correlated, heteroskedastic and non-Gaussian errors, $\tilde{k}$ defined in (2.10) can be viewed as a quasi-maximum likelihood (QML) estimator. The factors estimated by PCs are not efficient for approximate factor models, but Bai and Li (2012) provide simulation evidence that the efficiency loss of the PC estimator, relative to the maximum likelihood estimator, is vanishing as N and T diverge. Hence, we expect the LS estimator $\tilde{k}$ to be increasingly accurate as N and T diverge. The large sample theory for $\tilde{k}$ is presented in the next section.

3 Theory

3.1 Assumptions

Let k be a sequence that depends on T, i.e.,

$$k = k(T), \text{ and } T \to \infty. \tag{3.1}$$

For notational simplicity, we suppress its dependence on T and use k to represent the sequence $k(T)$. This notation is maintained throughout the rest of the paper unless otherwise specified.

Assumption 1 *$E(f_t f_t') = \Sigma_F$ is positive definite, $f_t f_t' - \Sigma_F = \sum_{j=0}^{\infty} a_j^{(f)} \nu_{t-j}^{(f)}$, where $\nu_{t-j}^{(f)}$ is i.i.d. over t with zero mean, $E|\nu_t^{(f)}|^{2+\rho} < \infty$ for some $\rho > 0$, and there exists a constant $M < \infty$ such that $\sum_{j=0}^{\infty} j|a_j^{(f)}| \leq M$.*

Assumption 2 *For $\ell = 1, 2$ and $i = 1, ..., N$, $E\|\lambda_{i\ell}\|^4 < \infty$, $\Lambda_\ell' \Lambda_\ell / N - \Sigma_\ell \to_p 0$ for some $r \times r$ positive definite matrix Σ_ℓ and $\Lambda_1' \Lambda_2 / N - \Sigma_{12} \to_p 0$ for some nonsingular matrix Σ_{12}.*

Assumption 3 *There exists $M < \infty$ such that for all T and N,*
(a) $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$ for all $i \leq N$ and $t \leq T$.
(b) $E(e_s' e_t / N) = E(N^{-1} \sum_{i=1}^N e_{is} e_{it}) = \gamma_N(s, t)$, and $\sum_{s=1}^T |\gamma_N(s, t)| \leq M$ for every $t \leq T$.
(c) $E(e_{it} e_{jt}) = \zeta_{ij,t}$ with $|\zeta_{ij,t}| \leq |\zeta_{ij}|$ for some ζ_{ij} and for all t. In addition, $\sum_{j=1}^N |\zeta_{ij}| \leq M$ for every $i \leq N$.
(d) $E(e_{it} e_{js}) = \zeta_{ij,ts}$, and $(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{s=1}^T \sum_{t=1}^T |\zeta_{ij,ts}| \leq M$.
(e) For every (t, s), $E\left|N^{-1/2} \sum_{i=1}^N [e_{is} e_{it} - E(e_{is} e_{it})]\right|^4 \leq M$.

Assumption 4 *(a) The errors $\{e_{it}\}_{i \leq N, t \leq T}$ are independent from the factors $\{f_t\}_{t \leq T}$ and the loadings $\{\lambda_{i1}, \lambda_{i2}\}_{i \leq N}$.*
(b) For each $i = 1, ..., N$, $f_t e_{it} = \sum_{j=0}^{\infty} a_{ij} \nu_{i,t-j}$, where $\nu_{i,t}$ is i.i.d. over t with zero mean, $E|\nu_{i,t}|^{2+\rho} < \infty$ for some $\rho > 0$, and there exists a fixed constant $M < \infty$ such that $\sum_{j=0}^{\infty} j|a_{ij}| \leq M$.

Assumption 5 *There exists $M < \infty$ such that for all T and N,*
(a) For each t, $E\left\|\frac{1}{\sqrt{Nk_0}}\sum_{s=1}^{k_0}\sum_{i=1}^{N} f_s[e_{is}e_{it} - E(e_{is}e_{it})]\right\|^2 \le M$ and $E\|\frac{1}{\sqrt{N(T-k_0)}}\sum_{s=k_0+1}^{T}\sum_{i=1}^{N} f_s[e_{is}e_{it} - E(e_{is}e_{it})]\|^2 \le M$.
(b) For $\ell = 1, 2$, $E\left\|\frac{1}{\sqrt{Nk_0}}\sum_{t=1}^{k_0}\sum_{i=1}^{N} f_t\lambda'_{i\ell}e_{it}\right\|^2 \le M$ and $E\left\|\frac{1}{\sqrt{N(T-k_0)}}\sum_{t=k_0+1}^{T}\sum_{i=1}^{N} f_t\lambda'_{i\ell}e_{it}\right\|^2 \le M$.
(c) For each i and $\ell = 1, 2$, $E\left\|\frac{1}{\sqrt{Nk_0}}\sum_{s=1}^{k_0}\sum_{j=1}^{N} \lambda_{j\ell}[e_{js}e_{is} - E(e_{js}e_{is})]\right\|^2 \le M$ and $E\left\|\frac{1}{\sqrt{N(T-k_0)}}\sum_{s=k_0+1}^{T}\sum_{j=1}^{N} \lambda_{j\ell}[e_{js}e_{is} - E(e_{js}e_{is})]\right\|^2 \le M$.

Assumption 6 *There exists $M < \infty$ such that,*
(a) For each t, $\limsup_{0\le h<k_0,\, N,h\to\infty} \frac{1}{\sqrt{N(h+1)\log\log(NT)}}\|\sum_{s=k_0-h}^{k_0}\sum_{i=1}^{N} f_s[e_{is}e_{it} - E(e_{is}e_{it})]\| \le M$ almost surely (a.s.) and $\limsup_{k_0<h\le T,\, N,h\to\infty} \frac{1}{\sqrt{N(h-k_0)\log\log(NT)}}\|\sum_{s=k_0+1}^{h}\sum_{i=1}^{N} f_s[e_{is}e_{it} - E(e_{is}e_{it})]\| \le M$ a.s.
(b) For $\ell = 1, 2$, $\limsup_{0\le h<k_0,\, N,h\to\infty} \frac{1}{\sqrt{N(h+1)\log\log(NT)}}\|\sum_{t=k_0-h}^{k_0}\sum_{i=1}^{N} f_t\lambda'_{i\ell}e_{it}\| \le M$ a.s. and $\limsup_{k_0<h\le T,\, N,h\to\infty} \frac{1}{\sqrt{N(h-k_0)\log\log(NT)}}\|\sum_{t=k_0+1}^{h}\sum_{i=1}^{N} f_t\lambda'_{i\ell}e_{it}\| \le M$ a.s.
(c) For each i and $\ell = 1, 2$, $\limsup_{0\le h<k_0,\, N,h\to\infty} \frac{1}{\sqrt{N(h+1)\log\log(NT)}}\|\sum_{s=k_0-h}^{k_0}\sum_{j=1}^{N} \lambda_{j\ell}[e_{js}e_{is} - E(e_{js}e_{is})]\| \le M$ a.s. and $\limsup_{k_0<h\le T,\, N,h\to\infty} \frac{1}{\sqrt{N(h-k_0)\log\log(NT)}}\|\sum_{s=k_0+1}^{h}\sum_{j=1}^{N} \lambda_{j\ell}[e_{js}e_{is} - E(e_{js}e_{is})]\| \le M$ a.s.

Assumption 7 *The eigenvalues of $r \times r$ matrix $(\Sigma_\ell\Sigma_F)$ are distinct for $\ell = 1, 2$.*

Assumptions 1–5 are standard in the factor model literature. Assumptions 1 is a strengthened version of Assumption A of Bai (2003). It implies that $E\|f_t\|^4 < \infty$ and f_tf_t' is strictly stationary and ergodic (Proposition 6.1(d) of Hayashi, 2000), so $T^{-1}\sum_{t=1}^{T} f_tf_t' \to_p \Sigma_F$ based on the Ergodic Theorem (Theorem 9.5.5 of Karlin and Taylor, 1975). The linear process setup and summability condition are sufficient to ensure the applicability of the Law of the Iterated Logarithm (LIL) on $f_tf_t' - \Sigma_F$ by Theorem 3.3 of Phillips and Solo (1992).[2] Namely, Assumption 1 implies that $\limsup_{k_0<k\le T,\, k\to\infty} \frac{1}{\sqrt{(k-k_0)\log\log T}}\sum_{t=k_0+1}^{k}(f_tf_t' - \Sigma_F)$ is bounded almost surely. This is applied to provide a uniform bound for the term $(k-k_0)^{-1/2}\sum_{t=k_0+1}^{k}(f_tf_t' - \Sigma_F)$.

Assumption 1 also requires $E(f_tf_t')$ to be constant over time. As both factors and loadings are unobserved, a factor model with a change in the second moment of f_t is observationally equivalent to a model with constant $E(f_tf_t')$, but with a change in the loading matrix (see, for example, the survey by Bai and Han, 2016). Thus, it is common to assume a time-invariant second moment of f_t for identification purposes in the literature of factor models with structural breaks (e.g., Cheng et al., 2016; Han and Inoue, 2015).

[2] The conditions in Assumption 1 are sufficient but not necessary for the LIL. Alternatively, one can assume f_tf_t' to be strong mixing with certain summability condition on the mixing coefficients to ensure the applicability of the LIL. See, for example, Theorem 5 of Oodaira and Yoshihara (1971).

Assumption 2 is similar to Assumption B of Bai (2003). Assumption 3 is a combination of Assumptions C and E of Bai (2003). We show in Theorem 1 that the consistency of $\tilde{k}$ holds when the errors are weakly correlated in both time and cross-sectional dimensions.

Assumption 4(a) is a slightly relaxed version of Assumption D of Bai and Ng (2004). Although the errors are assumed to be independent from factors and loadings, we allow for dependence within each group and dependence between $\{f_t\}_{t\leq T}$ and $\{\lambda_{i1}, \lambda_{i2}\}_{i\leq N}$. Hence, the structural change $\lambda_{i2} - \lambda_{i1}$ does not have to be independent of the factors. Also, e_{it} can be weakly correlated in both cross-sectional and time dimensions. The independence condition in Assumption 4(a) is sufficient but not necessary, and our theory still holds under weaker conditions. Assumption 4(b) implies that the LIL can be applied to bound the term $(k-k_0)^{-1/2}\sum_{t=k_0+1}^{k} f_t e_{it}$ and the Central Limit Theorem (CLT) holds for the term $k_0^{-1/2}\sum_{t=1}^{k_0} f_t e_{it}$ by Theorems 3.3 and 3.4 of Phillips and Solo (1992).

Assumption 5 is a modified version of Assumption F of Bai (2003). Note that all of the summands in Assumption 5 have zero mean. The inequalities in Assumption 5 are implications of the CLT, which holds under general conditions for a wide range of stationary mixing random fields.[3] Also, Jenish and Prucha (2009) establish CLT for mixing random fields without imposing stationarity.

The limsup bounds in Assumption 6 follow from the LIL, which can be established under more primitive assumptions for random fields. For example, Wichura (1973) and Li and Wu (1989) establish the LIL with multiple indices for i.i.d. variables under certain moment conditions. For dependence cases, Theorem 1 of Jiang (1999) shows that the LIL holds for ergodic and strictly stationary martingale differences with a bounded $(2+\rho)$-th moment for some $\rho > 0$.[4] In addition, Schmuland and Sun (2004) establish the LIL for random fields with exponentially decaying correlations. As the conventional LIL (with a single summation) has been well studied for dependent variables (e.g., Oodaira and Yoshihara, 1971; Petrov, 1984; Phillips and Solo, 1992; Zhao and Woodroofe, 2008), the LIL for random fields should continue to hold under quite general weak dependence structures; thus, we make this high-level assumption about LIL.

Assumption 7 is closely related to Assumption G of Bai (2003). It ensures the existence of the probability limit of the rotation matrix introduced by PC estimation.

We make the following assumptions about the magnitude of breaks, which include both large and small breaks.

Assumption 8 *Let $\alpha \in [0,1]$ and $0 < M < \infty$ be a constant that does not depend on N and T. The breaks take either of the following forms:*
(1) For $i = 1, ..., N$,

$$\lambda_{i2} - \lambda_{i1} = \frac{\delta_i}{N^{\frac{1-\alpha}{2}}} \quad \text{for } i = 1, ...N, \tag{3.2}$$

where $M^{-1} \leq E\|\delta_i\|^4 \leq M$ and $E\|N^{-1/2}\sum_{i=1}^{N}\lambda_{i\ell}\delta_i'\| \leq M$ for either $\ell = 1$ or 2.

[3]For stationary random fields, the detailed technical conditions for the CLT can be found in Nahapetian (1980), Bolthausen (1982), and El Machkouri et al. (2013).

[4]For the definitions of ergodicity and martingale differences with two indices, see Jiang (1999) for details.

(2) For $i = 1, ..., N$,

$$\lambda_{i2} - \lambda_{i1} = \begin{cases} \beta_i & \text{for } i = 1, ..., [mN^\alpha], \\ 0 & \text{for } i = [mN^\alpha] + 1, ..., N, \end{cases} \tag{3.3}$$

where $M^{-1} \leq E\|\beta_i\|^4 \leq M$ and m is a fixed constant satisfying $0 < m < \infty$ if $\alpha > 0$ and $1 \leq m < \infty$ if $\alpha = 0$.

Assumption 9 *There exists $M < \infty$ such that for all T and N,*
If the breaks take the form in (3.2), then $E\big(\sup_{k_0+u \leq k \leq k_0+s} |\sum_{t=k_0+u}^{k} \sum_{i=1}^{N} f_t' \delta_i e_{it}|^2\big) \leq MN(s - u + 1)$ for each u, s with $1 \leq u \leq s \leq T - k_0$ and $E\big(\sup_{k_0-s \leq k \leq k_0-u} |\sum_{t=k_0-s+1}^{k+1} \sum_{i=1}^{N} f_t' \delta_i e_{it}|^2\big) \leq MN(s - u + 1)$ for each u, s with $1 \leq u \leq s \leq k_0$;
If the breaks take the form in (3.3), then $E\big(\sup_{k_0+u \leq k \leq k_0+s} |\sum_{t=k_0+u}^{k} \sum_{i=1}^{[mN^\alpha]} f_t' \beta_i e_{it}|^2\big) \leq MN^\alpha(s - u + 1)$ for each u, s with $1 \leq u \leq s \leq T - k_0$ and $E\big(\sup_{k_0-s \leq k \leq k_0-u} |\sum_{t=k_0-s+1}^{k+1} \sum_{i=1}^{[mN^\alpha]} f_t' \beta_i e_{it}|^2\big) \leq MN^\alpha(s - u + 1)$ for each u, s with $1 \leq u \leq s \leq k_0$.

Assumption 10 *Let $\mathbb{A} = (\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2$. There exists a constant $c > 0$ such that*

$$\lim_{N\to\infty} P\left(\frac{1}{N^\alpha}\sum_{i=1}^{N}(\lambda_{i2} - \mathbb{A}'\lambda_{i1})'(\lambda_{i2} - \mathbb{A}'\lambda_{i1}) > c\right) = 1 \tag{3.4}$$

and

$$\lim_{N\to\infty} P\left(f_t'\left[\frac{1}{N^\alpha}(\Lambda_2 - \Lambda_1\mathbb{A})'(\Lambda_2 - \Lambda_1\mathbb{A})\right] f_t > 0\right) = 1 \tag{3.5}$$

for $t = k_0$ and $k_0 + 1$, where α is defined in Assumption 8.

The parameter α in Assumption 8 controls the magnitude of the breaks. When $\alpha = 1$, the setup is the same as in the literature on the large breaks in factor models (e.g., Chen et al., 2014; Han and Inoue, 2015; Cheng et al., 2016). When $\alpha < 1$, the breaks in (3.2) and (3.3) are of different formats: the size of the breaks is shrinking as $N \to \infty$ for all $i = 1, ..., N$ in (3.2), whereas the size of the breaks is fixed for a small portion of variables in (3.3), and the majority of variables do not have breaks. Note that the fraction of variables that have breaks is vanishing as $N \to \infty$ for $\alpha < 1$ in (3.3). In this sense, the setup in (3.3) also captures a kind of small break.

The condition that $E\|N^{-1/2}\sum_{i=1}^{N}\lambda_{i\ell}\delta_i'\| \leq M$ for either $\ell = 1$ or 2 can be satisfied under various circumstances, e.g., δ_i has zero mean and is uncorrelated with either λ_{i1} or λ_{i2}. This condition ensures that $\Theta'(\Lambda_1 - \Lambda_2) = O_p(N^\alpha)$ for $\alpha < 1$ (Lemma 1(a) in the appendix), where Θ is defined in (2.5).[5] This condition is not necessary and can be relaxed for the consistency of $\tilde{k}$ for $0 < \alpha \leq 1$, but it is needed to obtain the stochastic boundedness and asymptotic distribution of $\tilde{k} - k_0$ for $\alpha = 0$ in Theorems 2 and 3, which require a sharper bound for the difference between two rotation matrices caused by the PC estimation (see Lemma 2(c)).

The conditions in Assumption 9 are referred to as the second Kolmogorov type maximal inequality for moments in Fazekas (2014). This assumption ensures the applicability of Hájek-Rényi

[5]Note that $\Theta'(\Lambda_1 - \Lambda_2) = O_p(N)$ automatically holds for $\alpha = 1$ under Assumption 2.

type inequality (see Theorem 2.3 of Fazekas, 2014; see also Lemma 1(d) and its proof in the supplementary appendix). The inequalities in Assumption 9 can be established for weakly dependent random fields under more primitive conditions (see, for example, Lemma 3 of Móricz et al., 2008).

The matrix $\mathbb{A}$ in Assumption 10 is crucial to guaranteeing the identifiability of the break point by (2.10). For example, if $\Lambda_2 = \Lambda_1\mathcal{A}$ for some nonsingular matrix $\mathcal{A}$, then (2.5) can be rewritten as

$$\begin{bmatrix} X_{k_0}^{(1)} \\ X_{k_0}^{(2)} \end{bmatrix} = \begin{bmatrix} F^{(1)} \\ F^{(2)}\mathcal{A} \end{bmatrix} \Lambda_1' + \begin{bmatrix} \mathbf{e}_{k_0}^{(1)} \\ \mathbf{e}_{k_0}^{(2)} \end{bmatrix}, \tag{3.6}$$

which is an observationally equivalent model with constant factor loadings. Consider $k > k_0$ and the pre-k subsample. We can represent $X_k^{(1)}$ as

$$X_k^{(1)} = \begin{bmatrix} F^{(1)} \\ F_{k_0+1:k}\mathcal{A} \end{bmatrix} \Lambda_1' + \mathbf{e}_k^{(1)}.$$

where $F_{k_0+1:k} = [f_{k_0+1}, ..., f_k]'$. Let $\begin{bmatrix} F^{(1)} \\ F_{k_0+1:k}\mathcal{A} \end{bmatrix} \equiv \mathbb{G}_k^{(1)}$. Hence, the PCs extracted from the pre-k subsample are just an estimator for $\mathbb{G}_k^{(1)}$ up to some rotation. The residuals from the regression of $X_k^{(1)}$ on $\mathbb{G}_k^{(1)}$ will be consistent for the pre-k idiosyncratic errors. Also, the post-k subsample does not have a break for $k > k_0$, so the post-k residuals are consistently estimated by the OLS regression of $X_k^{(2)}$(i.e., the post-k data) on the post-k PC estimators. Thus, the break date is not identifiable[6] by minimizing (2.9) for data generated by (3.6). We need Assumption 10 to rule out the breaks of the format in (3.6). This is the price to pay due to the unobservability and multiplicative structure of factors and loadings. Equation (3.4) ensures that the difference between λ_{i1} and λ_{i2} defined in Assumption 8 is genuine in the sense that the sum of squared differences between λ_{i1} and λ_{i2} is of order N^α even if we account for the possible rotation of Λ_2.

Equation (3.4) ensures that the dominant positive term in $SSR(k,\tilde{F}) - SSR(k_0,\hat{F})$ diverges at the rate $(k-k_0)N^\alpha$ if $k-k_0$ is unbounded as $N,T \to \infty$. This condition is applied in the proof of Theorem 1 to show that $\tilde{k}-k_0$ must be stochastically bounded for $\alpha > 0$. Equation (3.5) allows us to show that the dominant positive term in $SSR(k,\tilde{F}) - SSR(k_0,\hat{F})$ is bounded away from zero when $k \neq k_0$ and $k-k_0$ is bounded as $N,T \to \infty$. It is applied to establish the consistency result that $P(\tilde{k}-k_0 \neq 0) \to 0$ in the proof of Theorem 1. The intuition is that it ensures $SSR(k,\tilde{F}) - SSR(k_0,\hat{F})$ to be positive even if k has a small deviation from the true break point k_0, so that $\tilde{k}$ has to equal k_0 to minimize $SSR(k,\tilde{F}) - SSR(k_0,\hat{F})$. Equation (3.5) is not required in Bai (2010) because he considers a constant scalar factor, i.e., $f_t = 1$. We need (3.5) due to the randomness of factors in our setup, but it is flexible enough to allow various data generating processes for f_t. For example, if f_{k_0} and f_{k_0+1} have continuous cumulative distribution functions, then the probability that either f_{k_0} or f_{k_0+1} is in the null space of $N^{-\alpha}(\Lambda_2 - \Lambda_1\mathbb{A})'(\Lambda_2 - \Lambda_1\mathbb{A})$ is

[6]Simulation confirms that $\tilde{k}$ is inconsistent under (3.6) and the results are available upon requests.

zero, given that $N^{-\alpha}(\Lambda_2 - \Lambda_1\mathbb{A})'(\Lambda_2 - \Lambda_1\mathbb{A})$ is positive semi-definite and its rank is no less than one.

3.2 Asymptotic properties of $\tilde{k}$

Before establishing the consistency of $\tilde{k}$ for $\alpha > 0$, we point out several important technical challenges. First, a commonly used trick in the factor model literature is to show the asymptotic equivalence between $SSR(k, \tilde{F})$ and $SSR(k, F)$, where $SSR(k, F)$ is the sum of squared residuals computed using the unobserved (infeasible) f_t. This trick has been applied by Chen (2015) to show the consistency of the estimated break fraction $\hat{\tau}$. However, it cannot be applied in the context of small breaks. Note that the consistency of $\tilde{k}$ for k_0 requires us to show

$$P(\min_{k \neq k_0} SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) > 0) \to 1$$

as $N, T \to \infty$. The conventional method is based on rewriting

$$\begin{aligned} SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) = &\left(SSR(k, \tilde{F}) - SSR(k, F)\right) + (SSR(k, F) - SSR(k_0, F)) \\ &+ \left(SSR(k_0, F) - SSR(k_0, \hat{F})\right). \end{aligned} \tag{3.7}$$

The usual analysis would then proceed to show that the first and third terms on the right-hand side are negligible in comparison with the middle term.

However, for an α close to zero and a k close to k_0, $\tilde{f}_t$ and $\hat{f}_t$ are so close to each other that the left-hand side is much smaller than the first and third terms on the right-hand side of (3.7).[7] That is, the usual argument does not work, and we have to directly compare the objective functions computed based on the estimated factors.

Second, when $k \neq k_0$, say $k > k_0$, the estimator $\tilde{F}_k^{(1)}$ is computed using the pre-k subsample that contains a structural break at k_0. This will affect the properties of PC estimators. We show in Lemma 2 (see the appendix) that $\tilde{f}_t$ is a consistent estimator of f_t subject to different rotation matrices for $t \leq k_0$ and $k_0 < t \leq k$. In addition, the estimated factors $\tilde{f}_t$ depend on the split date k, so $\tilde{f}_t$ and $\hat{f}_t$ are numerically different even for the same t. This causes another layer of technical challenge when we compare $SSR(k, \tilde{F})$ and $SSR(k_0, \hat{F})$. These problems do not arise in the conventional setup (e.g., Bai, 2010), where the regressors f_t are observed. The following proposition establishes the result for the difference between $\tilde{f}_t$ and $\hat{f}_t$.

Proposition 1 *Let* $\delta_{NT}^2 = \min(N, T)$. *Without loss of generality, consider any* $k(T)$ *satisfying* $k_0(T) < k(T) \leq \tau_2 T$ *for all* T, *where* k_0 *and* k *are sequences defined in* (2.3) *and* (3.1), *respectively.*

[7] Under the framework of no structural breaks, Lemma 2 of Bai and Ng (2002) shows that the difference $SSR(\tilde{F}) - SSR(F)$ is $O_p(\delta_{NT}^{-1} NT)$, where $SSR(\tilde{F})$ and $SSR(F)$ denote the sum of squared residuals based on estimated and true factors, respectively, and $\delta_{NT}^2 = \min(N, T)$. Our theoretical derivation shows that $SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) = O_p(N^\alpha (k - k_0))$ which can be much smaller than $O_p(\delta_{NT}^{-1} NT)$ for small α and $k - k_0$.

If Assumptions 1–8 hold, then for $t \leq k_0$ *and* $t > k$

$$R_k \tilde{f}_t - \hat{f}_t = \sqrt{k - k_0} O_p\left(\frac{\sqrt{\log(\log T)}}{T}\right) + (k - k_0) O_p\left(\frac{1}{TN^{1-\alpha}}\right) + O_p\left(\frac{\sqrt{\log\log(NT)}}{\delta_{NT}^2}\right), \quad (3.8)$$

and

$$\frac{1}{k_0}\sum_{t=1}^{k_0} \|R_k \tilde{f}_t - \hat{f}_t\|^2 = (k - k_0) O_p\left(\frac{\log(\log T)}{T^2}\right) + (k - k_0)^2 O_p\left(\frac{1}{T^2 N^{2-2\alpha}}\right) + O_p\left(\frac{\log\log(NT)}{\delta_{NT}^4}\right), \quad (3.9)$$

where $\tilde{f}_t$ *and* $\hat{f}_t$ *denote the PC estimators for* f_t *at a given sample-split date* k *and the true break date* k_0*, respectively;* R_k *is a diagonal matrix consisting of either* $+1$ *or* -1*; and the* O_p *terms in* (3.8) *and* (3.9) *are uniform in* k*.*

The result in Proposition 1 shows that the distance between $R_k \tilde{f}_t$ and $\hat{f}_t$ depends on both the break magnitude (i.e., α) and the extent to which the break date is misspecified (i.e., $|k - k_0|$). Since the rotation matrix does not affect the sum of squared residuals, we assume that the signs of $\tilde{f}_t$ are properly chosen such that $R_k = I_r$ for notational simplicity in the rest of the paper. Suppose that both $\log\log T/N \to 0$ and $\log\log N/T \to 0$, so the third term in (3.8) is $o_p(\delta_{NT}^{-1})$. When $\alpha = 1$ and $|k - k_0| \propto T$, (3.8) implies that $\tilde{f}_t - \hat{f}_t = O_p(1)$, i.e., $\tilde{f}_t$ and $\hat{f}_t$ are generally different even for large N and T. When either $\alpha < 1$ or $|k - k_0|/T \to 0$, the term $\frac{k-k_0}{TN^{1-\alpha}} \to 0$ as $N, T \to \infty$, and $\tilde{f}_t - \hat{f}_t = o_p(1)$. In the case of large break ($\alpha = 1$), for example, we can have $\tilde{f}_t - \hat{f}_t = O_p(T^{-1}\sqrt{\log\log T}) + O_p\left(\frac{\sqrt{\log\log(NT)}}{\delta_{NT}^2}\right)$ if $|k - k_0|$ is a fixed number as $T \to \infty$. It is well known that the distance between the PC estimator $\hat{f}_t$ and f_t, up to some rotation, is $O_p(\delta_{NT}^{-1})$ (see Theorem 1 of Bai, 2003). Thus, Proposition 1 implies that $\tilde{f}_t$ can be in a much closer neighborhood of $\hat{f}_t$ than the true factor f_t (up to some rotation) when the break date is slightly misspecified or the break size is small (i.e., either $|k - k_0|$ or α is small). This again illustrates that one cannot apply the usual trick that replaces $\tilde{f}_t$ and $\hat{f}_t$ with the infeasible f_t to establish the consistency of $\tilde{k}$ for an α close to zero.

Theorem 1 *Suppose that Assumptions 1–10 hold and* $0 < \alpha \leq 1$*. If*

$$\frac{N^{1-\alpha}\log\log T}{T} \to 0, \quad (3.10)$$

$$N^{-1}\log\log T \to 0,\ T^{-1}\log\log N \to 0 \quad (3.11)$$

as $N, T \to \infty$*, then*

$$\lim_{N,T\to\infty} Pr(\tilde{k} = k_0) = 1.$$

We compare the consistency of $\tilde{k}$ for $\alpha > 0$ in Theorem 1 with Bai's (2010) result. Under the condition $T^{-1}N\log\log T \to 0$, Bai (2010) shows the consistency of $\tilde{k}$ in a large panel setup with observed regressors/factors for $\alpha > 0$. Compared with Bai's condition, the rate in (3.10) is less restrictive and allows a larger N/T ratio when $\alpha > 0$. The condition in (3.11) is new compared with that in Bai (2010); it makes the estimation errors of factors negligible uniformly in the split date k. Condition (3.11) is rather flexible and likely to hold in most economic data for factor analysis. It is also remarkable that if $\alpha = 1$, then (3.10) is always satisfied and $\tilde{k}$ is consistent as long as (3.11) holds.[8] Thus, even if N is larger than T in practice, $\tilde{k}$ is still an accurate estimator as long as the break size is large. Theorem 1 is a substantial extension of Bai's (2010) result to the scenario with unobserved factors.

Remark 1: In a conventional time series setup with a fixed N and a large T, it is well known that $\tilde{k}$ is inconsistent and $\tilde{k} - k_0 = O_p(1)$ even if $\alpha = 1$ (see, for example, Bai, 1997a). Baltagi, Kao, and Wang (2017) develop an estimator, denoted as $\tilde{k}_{BKW}$, using the full-sample PC estimator of F. Specifically, they estimate the break point in the factor process F. Due to the fixed cross-sectional dimension of F, $\tilde{k}_{BKW}$ is subject to the same problem as the conventional break point estimator in a small N setup, i.e., $\tilde{k}_{BKW} - k_0$ is only stochastically bounded but does not converge to zero for large breaks (i.e., $\alpha = 1$). In contrast, our estimator is constructed based on the sum of squared residuals of the entire large panel. The consistency of our $\tilde{k}$ benefits from the large N setup, and it holds for both small and large breaks (i.e., $0 < \alpha \leq 1$).

Remark 2: An important contribution of this paper is the consistency under small breaks. The literature shows that the asymptotic properties of the PC estimator of factors depend on whether the breaks are large or small (see the survey by Bai and Han, 2016). Most studies have focused on large breaks in factor models ($\alpha = 1$) (e.g., Breitung and Eickmeier, 2011; Chen et al., 2014; Han and Inoue, 2015; Baltagi, Kao, and Wang, 2017). The large breaks in factor loadings lead to an augmented factor space estimated by PC. For small breaks, Bates et al. (2013) show that under (3.3) with $0 < \alpha \leq 0.5$, the first r PCs converge to the true factors (up to some rotation) at the same standard rate as in Bai (2003), even if the breaks are ignored in the estimation. Thus, the convergence rate does not slow down, and the factor space is not augmented under (3.3) with $0 < \alpha \leq 0.5$. This implies that it is challenging to detect the break date for small breaks. However, the finding of Bates et al. does not imply that the break point cannot be identified for $0 < \alpha \leq 0.5$.[9] Theorem 1 shows that our $\tilde{k}$ is consistent for k when $0 < \alpha \leq 1$, regardless of the asymptotic properties of the PC estimator for F. Hence, our estimator provides a unified framework to consistently estimate the break date for both large and small breaks.

When the breaks are small enough, ignoring the break may lead to better estimation results under some circumstances (e.g., forecasting) due to the bias-variance tradeoff. However, the small break setup has the advantage that it implies randomness of the estimated break point even in the

[8] Bai (2010) shows that $\tilde{k}$ is consistent for $\alpha > 0.5$ in large panel models with observed regressors regardless of the N/T ratio. Theorem 1 imposes a restriction on the relative rate between N and T for $0.5 < \alpha < 1$. This difference is due to the estimation of unobserved factors in our setup.

[9] In a recent study, Massacci (2017) shows the consistency of the LS estimator under small breaks as described by (3.3) with the restriction $\alpha > 0.5$. By the result of Bates et al., such a restriction ensures that enough series experience a regime shift so that ignoring the break is consequential. In contrast, our consistency result under (3.3) only requires $\alpha > 0$ and is thus much stronger.

limit, so that it allows us to analyze the non-degenerate asymptotic distribution of the estimated break point.

Before we establish the asymptotic distribution of $\tilde{k}$, we present the following theorem to show that $\tilde{k} - k_0$ is stochastically bounded when $\alpha = 0$.

Theorem 2 *If Assumptions 1–10 hold, $\alpha = 0$, $T^{-1}(N \log\log T) \to 0$ and $N^{-1} \log\log T \to 0$ as $N, T \to \infty$, then*

$$\tilde{k} - k_0 = O_p(1),$$

where α is defined in Assumption 8.

The stochastic boundedness of $\tilde{k} - k_0$ for $\alpha = 0$ in Theorem 2 is a strong result. Under (3.3), only m variables have structural changes for $\alpha = 0$. Even if m is fixed and $N \to \infty$, $\tilde{k} - k_0$ is still stochastically bounded as long as N and T satisfy the conditions in Theorem 2. To confirm this theoretical result, we conduct a simulation with a two-factor model where only one variable has structural changes with $\lambda_{12} - \lambda_{11} = 1_{2\times 1}$, $N = 100$, $T = 2000$, and $k_0 = T/2$. The factors, the pre-break loadings, and the errors are generated in the same way as in (4.1). Figure 1 presents the distribution of $\tilde{k} - k_0$. The simulation result shows that 95% of the mass of $|\tilde{k} - k_0|$ is less than 20, which is a quite narrow band compared with the sample size $T = 2000$.

Insert Figure 1 about here

Next, we present a result for the limiting distribution for the estimated break point. We impose additional conditions, which are not needed for consistency and the rate of convergence. Also, we only consider the type of small breaks in (3.2).

Theorem 3 *Let $\sum_{i=1}^{N} \delta_i \delta_i'/N \to_p \Sigma_\delta$ and $\Phi_N = N^{-1} \sum_{i=1}^{N} E(\delta_i \delta_i')\sigma_i^2 \to \Phi$ with $\sigma_i^2 = E(e_{it}^2)$. Under Assumptions 1–7, 9, and 10, if f_t is strictly stationary, e_{it} has no serial correlation, the breaks take the form specified in (3.2) with $\alpha = 0$, $N^{-1/2} \sum_{i=1}^{N} \delta_i e_{it} \to_d N(0, \Phi)$, $T^{-1}(N \cdot \log\log T) \to 0$ and $N^{-1} \log\log T \to 0$ as $N, T \to \infty$, then*

$$\tilde{k} - k_0 \to_d \arg\min_{\ell} W^*(\ell),$$

where $W^*(0) = 0$, $W^*(\ell) = W_1(\ell)$ for $\ell > 0$ and $W^*(\ell) = W_2(\ell)$ for $\ell < 0$, with

$$W_1(\ell) = \operatorname{tr}\left(\sum_{t=1}^{\ell} f_t f_t' \cdot \Sigma_\delta\right) + 2\sum_{t=1}^{\ell} f_t' Z_t, \;\; \ell = 1, 2, 3, ...$$

$$W_2(\ell) = \operatorname{tr}\left(\sum_{t=\ell+1}^{0} f_t f_t' \cdot \Sigma_\delta\right) + 2\sum_{t=\ell+1}^{0} f_t' Z_t, \;\; \ell = -1, -2, -3, ... \tag{3.12}$$

and Z_t's are i.i.d. $N(0, \Phi)$ random variables.

Note that the rotation caused by PC estimation does not show up in (3.12). The normality of Z_t is from the CLT over the cross-sections due to the use of panel data. The asymptotic distribution of $\tilde{k}-k_0$ is the same as what we would obtain if f_t were observable. Theorem 3 extends the distribution result of Bai (2010) to a factor model with unobserved factors. Under Bai's (2010) setup, where he considers changes in the mean of a large panel with $f_t = 1$, δ_i reduces to a scalar and the distribution in Theorem 3 reduces to

$$\tilde{k} - k_0 \to_d \arg\min_{\ell}[|\ell|\Sigma_\delta + 2\sqrt{\Phi}U(\ell)], \tag{3.13}$$

where $U(0) = 0$, $U(\ell) = \sum_{t=1}^{\ell} Z_t$ for $\ell = 1, 2, 3, ...$, $U(\ell) = \sum_{t=\ell+1}^{0} Z_t$ for $\ell = -1, -2, -3, ...$, and Z_t's are i.i.d. standard normal random variables. Compared with Bai's (2010) result in (3.13), our result in (3.12) depends on the distribution of f_t as we consider random common factors in general.

Theorem 3 focuses on the breaks as specified in (3.2), which is similar to Bai's (2010) setup for obtaining the asymptotic distribution of the break date estimator. The large N setup yields the result that Z_t follows a normal distribution rather than the distribution of idiosyncratic noises. If e_{it} is serially correlated, then the result in Theorem 3 still holds except that the normal random variables Z_t are serially correlated. If e_{it} has cross-sectional correlation, then we need to include the covariance terms $E(\delta_i\delta_j')E(e_{it}e_{jt})$ in Φ.

This result shows that the estimated break point can have a non-degenerate limiting distribution. It provides theoretical basis for constructing confidence intervals for the true break point.

3.3 Confidence bands of $\tilde{k}$

Using the distribution in (3.12) for inference requires a consistent estimator for Σ_δ. Unlike the conventional panel regressions with observed regressors in Bai (2010), it is challenging to estimate $\delta_i = \sqrt{N}(\lambda_{i1} - \lambda_{i2})$ due to the rotations caused by PC estimation. In general, the pre- and post-break loadings are rotated by different matrices in finite samples even if there is no break. Taking the difference between $\sqrt{N}\tilde{\lambda}_{i1}$ and $\sqrt{N}\tilde{\lambda}_{i2}$ tends to overestimate δ_i due to the different rotations in finite samples. In addition, the asymptotic distribution of $\tilde{k}$ depends on the data-generating process of f_t, which is unknown in practice. To circumvent these issues, we propose the following bootstrap procedure.

Algorithm 1 *(1) For a given data set, we estimate the change point $\tilde{k}$ using the LS method and obtain $\tilde{F}_{\tilde{k}}^{(1)}$, $\tilde{F}_{\tilde{k}}^{(2)}$, $\tilde{\lambda}_{i1}$, $\tilde{\lambda}_{i2}$, and the estimated error $\tilde{e}_i = (\tilde{e}_{i1}, ..., \tilde{e}_{iT})$.*
(2) For $i = 1, ..., N$, draw the T-dimensional vector e_i^* randomly from $(\tilde{e}_1, ..., \tilde{e}_N)$ with replacement. Generate the bootstrap data

$$x_{it}^* = \begin{cases} \tilde{f}_t'\tilde{\lambda}_{i1} + e_{it}^* & \text{for } t = 1, ..., \tilde{k} \\ \tilde{f}_t'\tilde{\lambda}_{i2} + e_{it}^* & \text{for } t = \tilde{k}+1, ..., T \end{cases}$$

for $i = 1, ..., N$. Estimate the break date $\tilde{k}^*$ using the bootstrap data x_{it}^*.

(3) Repeat step (2) B times. Obtain the critical values at a desired level from the sorted values of $\tilde{k}^*$.

The bootstrap procedure in Algorithm 1 requires neither an estimator for Σ_δ nor the distribution of f_t. In addition, the procedure can be applied not only for the case of $\alpha = 0$ but also for the case of $\alpha > 0$. Resampling the entire T-dimensional vector e_i can maintain the serial correlation structure in the error terms, so this procedure should be robust to limited serial correlation in the errors. The limitation of this resampling procedure is the lack of robustness to the cross-sectional correlation in e_{it}. The theoretical validity of this bootstrap procedure is not established and requires a nontrivial amount of work, so we leave it for future research. The finite-sample performance of Algorithm 1 in simulation is reasonably good and presented in the next section.

There are other resampling schemes in the literature. For example, practitioners may consider the wild bootstrap developed by Gonçalves and Perron (2014) to obtain e^*_{it}. A grid bootstrap that imposes the null hypothesis $k_0 = \bar{k}_0$ across values of $\bar{k}_0$ is another potential option. We do not consider the grid bootstrap method due to its high computational costs. The theoretical properties of these resampling schemes are interesting but beyond the scope of this paper. We leave them as topics for future research.

4 Monte Carlo Simulation

In this section, we conduct Monte Carlo simulations to evaluate the performance of the LS method in finite samples. For the consistency results, we consider the following data generating processes (DGPs):

$$x_{it} = \begin{cases} \lambda'_{i1} f_t + e_{it} & \text{for t} = 1, 2, \ldots, k_0 \\ \lambda'_{i2} f_t + e_{it} & \text{for t} = k_0 + 1, k_0 + 2, ..., T \end{cases} \tag{4.1}$$

where each factor follows an AR(1) process with unity variance and the AR coefficient is equal to 0.5, $\lambda_{i1} \sim i.i.d.\ N(0, I_r)$, and $e_{it} \sim i.i.d.\ N(0, r)$.[10] The post-break loadings λ_{i2} are generated using two different setups:

DGP1: $\lambda_{i2} = \lambda_{i1} + \delta_i / N^{(1-\alpha)/2}$, where $\delta_i \sim i.i.d.\ N(0, I_r)$.

DGP2: $\lambda_{i2} = \lambda_{i1} + \beta_i$ for $i = 1, ..., \lceil N^\alpha \rceil$, where $\beta_i \sim i.i.d.\ N(0, I_r)$; $\lambda_{i2} = \lambda_{i1}$ for $i = \lceil N^\alpha \rceil + 1, ..., N$, and $\lceil \rceil$ denotes the ceiling of a real number.

We set $k_0 = T/2$ and $r = 2, 3, 4$. Each experiment is repeated 1000 times. Let $\tilde{k}^{(s)}$ denote the estimated break date for the s-th repetition. We define the root mean square errors of the estimated change point as follows:

$$RMSE = \sqrt{\frac{1}{1000} \sum_{s=1}^{1000} \left(\tilde{k}^{(s)} - k_0 \right)^2}.$$

[10] We also conduct simulations with cross-sectionally correlated error terms. The detailed results are reported in Tables 4A and 4B in the supplementary appendix.

We compare our estimator with three other estimators. First, Baltagi, Kao, and Wang (2017) propose an estimator $\tilde{k}_{BKW}$, which converts the estimation of the change point in factor loadings into the estimation of the change point in the second moments of estimated factors. They show that $\tilde{k}_{BKW} - k_0 = O_p(1)$ for large breaks (i.e., $\alpha = 1$). The number of factors is set equal to $\hat{r}$ in the estimation of $\tilde{k}_{BKW}$, where $\hat{r}$ is estimated by applying Bai and Ng's (2002) IC_{p1} to the entire sample. In addition, to make our test comparable with $\tilde{k}_{BKW}$ under the case where r is unknown, we compute our estimator using $\hat{r} - 1$ factors, following the suggestion of Chen (2014). We denote this estimator as $\tilde{k}_{\hat{r}-1}$.

The third estimator to compare with is computed based on the model selection method proposed by Cheng et al. (2016). Their method can consistently estimate the numbers of pre- and post-break factors and determine the stability of factor loadings if the number of factors does not change. Given the selected model, Cheng et al. (2016) use the same LS break point estimator defined in (2.10) except that r is replaced with $\hat{r}_1$ ($\hat{r}_2$) for the pre-k (post-k) subsamples, where $\hat{r}_1$ and $\hat{r}_2$ are the estimated numbers of pre- and post-break factors by their shrinkage method. We denote this estimator as $\tilde{k}_{CLS}$, which is an alternative to $\tilde{k}_{\hat{r}-1}$ when the number of factors is unknown.

Remark 3: Although Cheng et al. (2016) consider the LS break point estimator, our theoretical work is substantially different from theirs in the following aspects. First, Cheng et al. (2016) mainly focus on the selection consistency of their shrinkage method, and they only show the consistency of the estimated break fraction, given their model selection result. In contrast, we establish the consistency of the LS break point estimator. Second, Cheng et al. (2016) consider the large break setup, i.e., $\alpha = 1$. We establish the consistency of $\tilde{k}$ for both small and large breaks. Note that the penalty terms in Cheng et al. (2016) are not designed for the small break setup. When α is close to zero, their shrinkage method cannot detect the instability, in which case their break point estimator is unavailable. To make a fair comparison, we compute $\tilde{k}_{CLS}$ under the assumption that the existence of break is known. In other words, the method of Cheng et al. provides an alternative way to determine how many factors to use when computing the LS break point estimator. The finite-sample performance of $\tilde{k}_{CLS}$ is studied in our simulation. Its large sample theory for $\alpha < 1$ is beyond the scope of this paper and left for future research.

Table 1A reports the RMSEs of different estimators under DGP1. For large breaks ($\alpha = 1$), the LS estimator $\tilde{k}$ has much smaller RMSEs than $\tilde{k}_{BKW}$ for $N \leq 100$ and $T \leq 200$. The result confirms the consistency of $\tilde{k}$ and that $\tilde{k}_{BKW} - k_0 = O_p(1)$. The RMSEs of the LS estimator based on $\hat{r} - 1$ factors are greater than those of $\tilde{k}$ but still smaller than those of $\tilde{k}_{BKW}$ for small N and T. For smaller breaks with $\alpha \leq 0.75$, the advantage of $\tilde{k}$ over $\tilde{k}_{BKW}$ is more evident. The RMSEs of $\tilde{k}$ shrink as N and T increase.[11] In contrast, $\tilde{k}_{BKW}$ is not consistent and its RMSEs increase with N

[11]The results in Table 1A show a pattern that the RMSE of $\tilde{k}$ tends to substantially decrease with a fixed N and an increasing T for $\alpha < 1$. For example, when $\alpha = 0.25$ and $r = 2$, the RMSE of $\tilde{k}$ decreases from 10.4 to 5.65 as T increases from 100 to 200 and N is fixed at 50. However, this pattern does not imply that we can obtain a consistent estimate by fixing N and increasing T only. Additional simulations show that the RMSE of $\tilde{k}$ approximately stays at

and T for $\alpha \leq 0.5$, which implies that $\tilde{k}_{BKW} - k_0$ is not stochastically bounded. In addition, the RMSEs of $\tilde{k}_{\hat{r}-1}$ are much larger than $\tilde{k}$ but still smaller than $\tilde{k}_{BKW}$. For $\alpha \leq 0.5$, it turns out that Bai and Ng's (2002) estimator $\hat{r}$ is almost always equal to r (i.e., the factor space is not augmented by the small breaks), so using $\hat{r}-1$ factors means that one factor is missing from the regression and treated as part of the error terms. The missing factor contributes to the less accurate estimation by $\tilde{k}_{\hat{r}-1}$. Finally, $\tilde{k}_{CLS}$ performs well in most cases. For $\alpha < 1$ and large samples ($N \geq 100$ and $T \geq 200$), its RMSEs are very close to those of $\tilde{k}$ and much smaller than those of $\tilde{k}_{\hat{k}-1}$. This means that the shrinkage method of Cheng et al. (2016) can provide the correct estimate for r (in large samples) even if the break size is small. For small samples with $N = 50$, $\tilde{k}_{\hat{k}-1}$ sometimes has smaller RMSEs than $\tilde{k}_{CLS}$. For example, the RMSE of $\tilde{k}_{CLS}$ is 28.63 and the RMSE of $\tilde{k}_{\hat{r}-1}$ is 23.50 when $r = 2$, $\alpha = 0.25$, $N = 50$, and $T = 100$. Our simulation shows that Cheng et al's method tends to overfit the number of factors when $N = 50$ and $\alpha < 1$, which may cause the deterioration in performance of $\tilde{k}_{CLS}$. The overall performance of $\tilde{k}_{CLS}$ implies that the shrinkage method of Cheng et al. (2016) provides a good estimate for the number of factors to compute the LS break point estimator in large samples.

Table 1B presents the probabilities of correct estimation of the break date. The results are consistent with those in Table 1A: the LS estimator $\tilde{k}$ can detect the true break date with higher probabilities than the other three estimators regardless of the value of α. The probabilities are increasing with α and the sample size N and T. Even for $\alpha = 0.25$, $\tilde{k}$ can still detect the true break date with a moderate probability. For example, the probability of $\tilde{k} = k_0$ is about 1/4 for $r = 3$, $N = 100$, and $T = 200$. Furthermore, for $\alpha \leq 0.75$, the probability of correction estimation by $\tilde{k}_{BKW}$ is close to zero and consistent with the large RMSEs of $\tilde{k}_{BKW}$ shown in Table 1A. In addition, the performance of $\tilde{k}_{\hat{r}-1}$ is again between $\tilde{k}$ and $\tilde{k}_{BKW}$. The low probabilities of $\tilde{k}_{\hat{r}-1}$ for $\alpha = 0.25$ are due to the same reason discussed in the previous paragraph. Moreover, the probability of correction estimation by $\tilde{k}_{CLS}$ is lower than $\tilde{k}$ but higher than $\tilde{k}_{\hat{r}-1}$. The advantage of $\tilde{k}_{CLS}$ over $\tilde{k}_{\hat{r}-1}$ is more pronounced for $\alpha < 1$. These findings are consistent with the results in Table 1A.

Table 2A reports the RMSEs of the three estimators under DGP2. The pattern of the results is similar to that of Table 1A. For large breaks ($\alpha = 1$), the RMSEs of $\tilde{k}$ are always less than one, even for the case with $(N, T) = (50, 100)$. Under DGP2, Bates et al. (2013) show that if $\alpha \leq 0.5$, the PC estimators for the factors and factor loadings ignoring the breaks achieve the same convergence rate as in Bai (2003). However, such a result does not imply that the break date cannot be identified for $\alpha \leq 0.5$. The consistency of $\tilde{k}$ for $\alpha \leq 0.5$ is confirmed by our simulation results. Note that the number of variables with breaks is in fact quite small for $\alpha = 0.25$ in our experiments. For example, when $r = 2$ and $(N, T) = (100, 500)$, the RMSE of $\tilde{k}$ is equal to 4.84 with only 4 (out of 100) variables[12] with structural breaks, and the probability of correction estimation is 0.39 (see Table 2B). This shows that $\tilde{k}$ can detect the break date with high accuracy even when only a small

the same level and does not converge to zero if we further increase T and keep N fixed. The consistency of $\tilde{k}$ requires both N and T to diverge.

[12]Note that both the ceilings of $100^{0.25}$ and $200^{0.25}$ are equal to 4.

fraction of the variables have undergone structural breaks. The patterns of the results for $\tilde{k}_{BKW}$, $\tilde{k}_{\hat{r}-1}$ and $\tilde{k}_{CLS}$ are similar to those in Table 1A.

Figure 2 shows the distribution of $\tilde{k} - k_0$ under DGP1 and DGP2 for $r = 2$, $N = 100$, and $T = 500$. It is evident that the variation of $|\tilde{k} - k_0|$ decreases as α increases under both DGPs. This pattern is consistent with the prediction of our theory.

Table 2B summarizes the probabilities of correct estimation under DGP2. The probabilities of $\tilde{k}$ are increasing with the sample size and the value of α. For $\alpha = 0.25$, $r = 2$, and $(N, T) = (200, 500)$, the probability of correct estimation by $\tilde{k}$ is 0.34, given that only $4 = \lceil 200^{0.25} \rceil$ variables have breaks. Our simulation also shows that this probability increases to 0.45 (not reported in the table) for $(N, T) = (400, 2000)$, given that $\lceil 400^{0.25} \rceil = 5$ variables have breaks. Thus, this confirms the consistency of $\tilde{k}$ under DGP2 for $\alpha \leq 0.5$.

Remark 4: Our theory for $\tilde{k}$ is based on the assumption that r is known. For unknown r, our simulation results show that $\tilde{k}_{\hat{r}-1}$ outperforms $\tilde{k}_{BKW}$ in terms of both RMSEs and the probabilities of correct estimation, especially for $\alpha \geq 0.5$. The theory of $\tilde{k}_{\hat{r}-1}$ is beyond the scope of this paper and an open question left for future research.

Remark 5: For the case of unknown r, using $\hat{r} - 1$ is not the only choice. If all of the factors undergo large breaks in their loadings, then the number of factors estimated by full-sample PC tends to be doubled, i.e., $P(\hat{r} = 2r) \to 1$ as $N, T \to \infty$ (see Breitung and Eickmeier, 2011). Thus, an alternative choice is to use $\hat{r}/2$ factors in the estimation of $\tilde{k}$. Which one of these two options performs better depends on which is closer to the true value of r. The motivation of $\hat{r}/2$ suggests that it should be a better option when all factors undergo large breaks and $r \geq 2$. If fewer than r factors have breaks in their loadings, then $\hat{r}$ tends to a value less than $2r$ in large samples. Also, if the breaks are small enough (i.e., $\alpha \leq 0.5$), then the number of estimated factors by PC is not inflated, i.e., $P(\hat{r} = r) \to 1$, by the results of Bates et al. (2013). Under these scenarios, $\hat{r} - 1$ can be a better choice than $\hat{r}/2$. We conduct simulations to compare the performance of $\tilde{k}$ using $\hat{r} - 1$ and $\hat{r}/2$ (rounded to the nearest integer) factors. Neither of these two choices dominates the other. Here, $\hat{r}/2$ tends to perform better when $\alpha = 1$ and N and T are large, whereas $\hat{r} - 1$ tends to produce more accurate estimators for $\alpha < 1$ or small N and T. For more details about the simulation results, see Table 5 in the supplementary appendix.

Next, we investigate the coverage probability of the confidence intervals for $\tilde{k}$ obtained from our bootstrap procedure. The data are generated by DGP1 for $\alpha = 0$, 0.25, and 0.5. Each experiment is repeated 500 times, and for each simulated data set the resampling is repeated 1000 times. The effective coverage probabilities are reported in Table 3. For $\alpha = 0$, the length of the confidence intervals tends to be underestimated, but the coverage probabilities approach the nominal levels as N and T increase. For $\alpha = 0.25$ and $\alpha = 0.5$, the coverage probabilities are better than those obtained under $\alpha = 0$. This is because the break date can be estimated more accurately for larger breaks. For a larger sample size and α, the estimated break date tends to have a degenerate distribution, and thus the coverage probabilities are higher than the nominal levels (i.e., the confidence intervals become more conservative), especially for the case of $\alpha = 0.5$.

4.1 Multiple breaks

In this subsection, we investigate the performance of $\tilde{k}$ in the presence of multiple breaks. We consider two setups, both of which may occur in practice. The first setup studies the case where the entire factor loading matrix undergoes two breaks. The data are generated by the following DGP:

$$x_{it} = \begin{cases} \lambda_{i1}' f_t + e_{it} & \text{for } t = 1, 2, \ldots, [T/3] \\ \lambda_{i2}' f_t + e_{it} & \text{for } t = [T/3]+1, ..., [2T/3] \\ \lambda_{i3}' f_t + e_{it} & \text{for } t = [2T/3]+1, ..., T, \end{cases} \tag{4.2}$$

where f_t, λ_{i1}, and e_{it} are generated in the same way as in (4.1), λ_{i2} is generated in the same way as in DGP1, and $\lambda_{i3} = \lambda_{i2} + \delta_{i2}/N^{(1-\alpha)/2}$, where $\delta_{i2} \sim i.i.d.\ N(0, I_r)$. It is expected that $\tilde{k}$ can detect one of the common break dates. Indeed, simulation shows that $\tilde{k}$ tends to equal either $[T/3]$ or $[2T/3]$ as N and T grow. This is similar to Bai's (1997b) result that the LS estimator can be applied to detect one of the break points even if the number of breaks is misspecified in a fixed N and large T linear regression context. Thus, we expect that $P(\tilde{k} = [T/3] \text{ or } [2T/3]) \to 1$ as $N, T \to \infty$ under DGP (4.2). Large sample theory for $\tilde{k}$ in factor models with multiple breaks is challenging and beyond the scope of this paper. We leave it for future research. For simulation results under (4.2), please see Table 6A in the supplementary appendix.

The second setup studies the case where different groups of variables undergo structural breaks in different periods. The data are generated by the following DGP:

$$x_{it} = \begin{cases} \lambda_{i1}' f_t + e_{it} & \text{for } t = 1, 2, \ldots, (k_0 - 3 + b(i)) \\ \lambda_{i2}' f_t + e_{it} & \text{for } t = (k_0 - 2 + b(i)), ..., T, \end{cases} \tag{4.3}$$

where f_t, λ_{i1}, and e_{it} are generated in the same way as in (4.1); λ_{i2} is generated in the same way as in DGP1; and $b(i) = i - 5 \cdot [i/5] + 1$ for $i = 1, ..., N$, so $b(i) \in \{1, 2, 3, 4, 5\}$. Under this setup, N variables are divided into five groups. Each group contains one fifth of the variables, and the $b(i)$-th group undergoes a break at $k_0 - 3 + b(i)$. Hence, the second group undergoes a break one period later than the first group, and so on. This setup is motivated by the structural breaks in some variables potentially causing a structural break in other variables a few periods later. Note that (4.3) can be represented by a factor model with a common break date k_0 by redefining the error terms, i.e., for $i = 1, ..., N$,

$$x_{it} = \begin{cases} \lambda_{i1}' f_t + u_{it} & \text{for } t = 1, 2, \ldots, k_0 \\ \lambda_{i2}' f_t + v_{it} & \text{for } t = k_0 + 1, ..., T, \end{cases} \tag{4.4}$$

where $u_{it} = e_{it} - (\lambda_{i1} - \lambda_{i2})' f_t \cdot \mathbf{1}\{k_0 - 3 + b(i) < t \leq k_0\}$, $v_{it} = e_{it} - (\lambda_{i2} - \lambda_{i1})' f_t \cdot \mathbf{1}\{k_0 < t \leq k_0 - 3 + b(i)\}$ and $\mathbf{1}\{.\}$ denotes the indicator function. In fact, the common break date can be specified as any integer in $[k_0 - 2, k_0 + 2]$ by redefining the error terms. Apparently, the redefined error terms in (4.4) do not satisfy Assumption 3 because $E(u_{it}) \neq 0$ and $\sum_{j=1}^{N} |E(u_{it}u_{jt})|$ is not bounded for some t. Hence, the probability of $\tilde{k} \neq k_0$ does not tend to zero as $N, T \to \infty$. In fact, simulation shows that $\tilde{k}$ could equal any of the break dates from $k_0 - 2$ to $k_0 + 2$, and that the probability $|\tilde{k} - k_0| > 2$ tends to zero as N and T diverge. This implies that $\tilde{k}$ seems able to detect one of the break dates even when multiple break points are close to each other. The theoretical properties of $\tilde{k}$ under (4.3) are left as a future research topic. The simulation results under (4.3) are summarized by Table 6B in the supplementary appendix.

5 Empirical Applications

5.1 Financial asset returns

The first empirical application uses monthly return data for stocks traded on the NYSE, AMEX, and NASDAQ between January 2005 and December 2012. After deleting all of the missing data, the sample size is $T = 96$ and $N = 3716$. Bai and Ng's (2002) IC_{p1} detects $\hat{r} = 2$ for this sample period. To apply our LS method, we set the number of factors equal to $\hat{r} - 1$. The estimated change point is 2009:03. This result is consistent with most major market indexes (such as the SP500, the Dow 30, and the NASDAQ composite) reaching their troughs on March 9, 2009 for the post-2005 sample period. The 99% bootstrap confidence interval is a singleton that consists of the estimated break point. Thus, the estimation uncertainty is considerably small for $\tilde{k}$ in this sample, which implies that the magnitude of the break is large.

5.2 Macroeconomic data

Our second empirical application focuses on estimating the break date of the U.S. macroeconomy during the recent financial crisis. We use the data set adopted by Cheng et al. (2016), which consists of monthly observations of 102 U.S. macroeconomic variables. Since the data are likely to have multiple breaks, we focus on the subsample period between 2001:12 and 2013:01 ($T = 134, N = 102$)[13].

Cheng et al. (2016) find that 2007:12 is the break date, that the pre-break subsample has one factor, and that the post-break subsample has two factors. We follow their estimation result and set the number of factors equal to two in our LS estimation. We search the break date between 20% and 80% of the full sample (i.e., between 2004:01 and 2010:11) and the result is $\tilde{k} = 2008:12$. The bootstrap confidence intervals are $[2008:11,\ 2009:01]$, $[2008:11,\ 2009:01]$, $[2008:09,\ 2009:06]$ at the 90%, 95%, and 99% levels, respectively. Note that the upper bound of the 99% level confidence

[13]Note that 2001:11 is the NBER trough date for the early 2000s recession, so we focus on the data after this trough date.

interval is the same as the NBER trough date of the Great Recession. Our estimation result indicates that the break did not occur immediately after 2007:12 (the NBER beginning date of the Great Recession), yet the break date is close to the end of the Great Recession. Thus, it seems that the factor loading matrix changed to a new regime after the 2008 financial crisis.

6 Conclusion

In this paper, we develop an asymptotic theory for the LS estimator of the break point in high-dimensional factor models where the unobserved factors are estimated by PCA. We establish the conditions under which the LS estimator is consistent for the break point and show that the consistency holds even if the breaks are small. The asymptotic distribution of the estimated break point depends on the data-generating process of the factors. Thus, we propose a bootstrap procedure to construct the confidence intervals for the estimated break date. The simulation results confirm that the break date can be accurately estimated for small breaks. The coverage probabilities of the bootstrap confidence intervals approach the nominal levels as N and T increase. We conduct two empirical applications with U.S. stock return and macroeconomic data. The estimated break date is 2009:04 for the stock market, whereas the estimated break date is 2008:12 for the macroeconomic data.

Appendix

A Lemmas

The detailed proofs for Lemmas 1 to 8 are provided in the supplementary appendix.

Lemma 1 *Under Assumptions 1–4, 8 and 9,*
(a) $\Theta'(\Lambda_2 - \Lambda_1) = O_p(N^{\alpha})$;
(b) There exists $M_1 < \infty$ *such that for all* T *and* N, $E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{i\ell}e_{it}\right\|^2 \leq M_1$ *for* $t = 1, ..., T$ *and* $\ell = 1, 2$;
(c) For all T *and* N *and* $t = 1, ..., T$, $e_t'(\Lambda_2 - \Lambda_1) = O_p(N^{\alpha/2})$;
(d) There exists $0 < C_1 < \infty$ *such that for all* T *and* N, $E\big(\sup_{k\geq k_0+s}|\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}f_t'(\lambda_{i1} - \lambda_{i2})e_{it}|^2\big) \leq s^{-1}C_1MN^{\alpha}$ *and* $E\big(\sup_{k\leq k_0-s}|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\sum_{i=1}^{N}f_t'(\lambda_{i1} - \lambda_{i2})e_{it}|^2\big) \leq s^{-1}C_1MN^{\alpha}$.

A.1 Properties of $\tilde{f}_t$

This subsection provides several useful lemmas about the properties of the estimated factors $\tilde{f}_t$ when $k > k_0$. The case where $k < k_0$ is symmetric and hence omitted.

Consider the pre- and post-k subsamples for $k > k_0$:

$$\begin{aligned} X_k^{(1)} &= G_k^{(1)}\Theta' + \mathbf{e}_k^{(1)}, \\ X_k^{(2)} &= F_k^{(2)}\Lambda_2' + \mathbf{e}_k^{(2)} \end{aligned} \tag{A.1}$$

where $G_k^{(1)}$ denotes the first k rows of G. Expanding (2.6) yields

$$\begin{aligned} \frac{1}{kN}\left(G_k^{(1)}\Theta'\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + G_k^{(1)}\Theta'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)} + \mathbf{e}_k^{(1)}\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + \mathbf{e}_k^{(1)}\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}\right) &= \tilde{F}_k^{(1)}\tilde{V}_k^{(1)} \\ \frac{1}{(T-k)N}\left(F_k^{(2)}\Lambda_2'\Lambda_2 F_k^{(2)'}\tilde{F}_k^{(2)} + F_k^{(2)}\Lambda_2'\mathbf{e}_k^{(2)'}\tilde{F}_k^{(2)} + \mathbf{e}_k^{(2)}\Lambda_2 F_k^{(2)'}\tilde{F}_k^{(2)} + \mathbf{e}_k^{(2)}\mathbf{e}_k^{(2)'}\tilde{F}_k^{(2)}\right) &= \tilde{F}_k^{(2)}\tilde{V}_k^{(2)}. \end{aligned} \tag{A.2}$$

We will show that $\tilde{f}_t$ consistently estimates f_t up to some rotation, which depends on whether $t \leq k_0$, $k_0 < t \leq k$ or $t > k$ (see Lemma 2 below).

For each $1 \leq t \leq k_0$, (A.2) implies

$$\frac{1}{kN}\left(f_t'\Lambda_1'\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + f_t'\Lambda_1'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)} + e_t'\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + e_t'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}\right) = \tilde{f}_t'\tilde{V}_k^{(1)}. \tag{A.3}$$

For each $k_0 + 1 \leq t \leq k$, (A.2) implies

$$\frac{1}{kN}\left(f_t'\Lambda_2'\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + f_t'\Lambda_2'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)} + e_t'\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + e_t'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}\right) = \tilde{f}_t'\tilde{V}_k^{(1)}. \tag{A.4}$$

For each $t \geq k$, (A.2) implies

$$\frac{1}{(T-k)N}\left(f_t'\Lambda_2'\Lambda_2 F_k^{(2)'}\tilde{F}_k^{(2)} + f_t'\Lambda_2'\mathbf{e}_k^{(2)'}\tilde{F}_k^{(2)} + e_t'\Lambda_2 F_k^{(2)'}\tilde{F}_k^{(2)} + e_t'\mathbf{e}_k^{(2)'}\tilde{F}_k^{(2)}\right) = \tilde{f}_t'\tilde{V}_k^{(2)}. \qquad \text{(A.5)}$$

Let

$$\begin{aligned} H_{k,1}^{(1)} &= \frac{1}{kN}\Lambda_1'\Theta G_k^{(1)'}\tilde{F}_k^{(1)}\tilde{V}_k^{(1)^{-1}}, \\ H_{k,2}^{(1)} &= \frac{1}{kN}\Lambda_2'\Theta G_k^{(1)'}\tilde{F}_k^{(1)}\tilde{V}_k^{(1)^{-1}}, \\ H_k^{(2)} &= \frac{1}{(T-k)N}\Lambda_2'\Lambda_2 F_k^{(2)'}\tilde{F}_k^{(2)}\tilde{V}_k^{(2)^{-1}}. \end{aligned} \qquad \text{(A.6)}$$

Lemma 2 *Under Assumptions 1–5 and 8, the following results hold uniformly in $k \in [\tau_1 T, \tau_2 T]$.*
(a) $k_0^{-1}\sum_{t=1}^{k_0}\|\tilde{f}_t - H_{k,1}^{(1)'}f_t\|^2 = O_p(\delta_{NT}^{-2})$ *and* $(T-k)^{-1}\sum_{t=k+1}^{T}\|\tilde{f}_t - H_k^{(2)'}f_t\|^2 = O_p(\delta_{NT}^{-2})$*;*
(b) $(k-k_0)^{-1}\sum_{t=k_0+1}^{k}\|\tilde{f}_t - H_{k,2}^{(1)'}f_t\|^2 = O_p(\delta_{NT}^{-2})$*;*
(c) $H_{k,1}^{(1)} - H_{k,2}^{(1)} = O_p\left(N^{\alpha-1}\right)$.

For notational simplicity, we define

$$L_{2NT} = \log\log(NT). \qquad \text{(A.7)}$$

Lemma 3 *If Assumptions 1–6 hold, then the following results hold uniformly in $k \in [\tau_1 T, \tau_2 T]$ and $k > k_0$.*
(a) $\sum_{i=1}^{N}\sum_{t=1}^{k}\lambda_{i\ell}e_{it}\tilde{f}_t'/kN = O_p(L_{2NT}^{1/2}\delta_{NT}^{-1}N^{-1/2})$ *for $\ell = 1, 2$;*
(b) For each t, $\sum_{s=1}^{k}e_t'e_s\tilde{f}_s'/kN = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$.

For $k > k_0$, let $\tilde{F}_1^{(1)}$ and $\tilde{F}_2^{(1)}$ denote the first k_0 and last $k - k_0$ rows of $\tilde{F}_k^{(1)}$, i.e.,

$$\tilde{F}_k^{(1)} = \begin{bmatrix} \tilde{F}_1^{(1)} \\ \tilde{F}_2^{(1)} \end{bmatrix}. \qquad \text{(A.8)}$$

The following lemma extends Bai's (2003) Lemma B2. Note that our rate in Lemma 4(a) is slowed down by a factor $L_{2NT}^{1/2}$ compared to Bai (2003) because our bound is uniform in $k \in [\tau_1 T, \tau_2 T]$. If k is fixed at k_0, then Lemma 4(a) reduces to Bai's (2003) Lemma B2 for the setup without structural breaks.

Lemma 4 *If Assumptions 1–6 hold, $k > k_0$ and $k \in [\tau_1 T, \tau_2 T]$, then*
(a) $k_0^{-1}F^{(1)'}\left(\tilde{F}_1^{(1)} - F^{(1)}H_{k,1}^{(1)}\right) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$*; and* $(T-k)^{-1}F_k^{(2)'}\left(\tilde{F}_k^{(2)} - F_k^{(2)}H_k^{(2)}\right) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$*;*
(b) $(k-k_0)^{-1}\sum_{t=k_0+1}^{k}f_t(\tilde{f}_t' - f_t'H_{k,2}^{(1)}) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2}) + (k-k_0)^{-1/2}O_p\left(L_{2NT}^{1/2}\delta_{NT}^{-1}\right)$*;*
(c) $(k-k_0)^{-1}\sum_{t=k_0+1}^{k}(\tilde{f}_t\tilde{f}_t' - H_{k,2}^{(1)'}f_tf_t'H_{k,2}^{(1)}) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2}) + (k-k_0)^{-1/2}O_p\left(L_{2NT}^{1/2}\delta_{NT}^{-1}\right)$ *and* $k_0^{-1}(\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)} - H_{k,1}^{(1)'}F^{(1)'}F^{(1)}H_{k,1}^{(1)}) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$*, where the O_p terms in parts (a)-(c) are uniform in k.*

Lemma 5 below extends Bai's (2003) Lemma B1. Note that our rate in Lemma 5(a) is slowed down by a factor $L_{2NT}^{1/2}$ compared to Bai (2003) because our bound is uniform in $k \in [\tau_1 T, \tau_2 T]$. If k is fixed at k_0, then Lemma 5(a) reduces to Bai's (2003) Lemma B1.

Lemma 5 *If Assumptions 1–6 hold, $k > k_0$ and $k \in [\tau_1 T, \tau_2 T]$, then for each i,*
(a) $k_0^{-1} e_{k_0,i}^{(1)\prime} \left(\tilde{F}_1^{(1)} - F^{(1)} H_{k,1}^{(1)}\right) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$; and $(T-k)^{-1} e_{k,i}^{(2)} \left(\tilde{F}_k^{(2)} - F_k^{(2)} H_k^{(2)}\right) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$, where $e_{k_0,i}^{(1)} = [e_{i1}, ..., e_{ik_0}]'$ and $e_{k,i}^{(2)} = [e_{ik+1}, ..., e_{iT}]'$;
(b) $(k-k_0)^{-1} \sum_{t=k_0+1}^{k} e_{it}(\tilde{f}_t' - f_t' H_{k,2}^{(1)}) = O_p(L_{2NT}^{1/2}\delta_{NT}^{-2}) + (k-k_0)^{-1/2} O_p\left(L_{2NT}^{1/2}\delta_{NT}^{-1}\right)$, where the O_p terms in parts (a)-(b) are uniform in k.

In the next lemma, we will show that if the signs of columns in $\tilde{F}^{(1)}$ are properly assigned, then

$$\frac{F^{(1)\prime}(\tilde{F}_1^{(1)} - \hat{F}^{(1)})}{k_0} = \sqrt{k-k_0}\, O_p\left(\frac{\sqrt{\log(\log T)}}{T}\right) + (k-k_0) O_p\left(\frac{1}{TN^{(1-\alpha)}}\right) + O_p\left(\frac{\sqrt{L_{2NT}}}{\delta_{NT}^2}\right), \quad \text{(A.9)}$$

where the $O_p(T^{-1}\sqrt{\log(\log T)})$, $O_p(T^{-1}N^{\alpha-1})$ and $O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$ terms are all uniform in k. The bound in (A.9) is useful to prove Proposition 1. To simplify the notation, we use $O_p(\Pi_{NT,k})$ to denote the bound for $k_0^{-1} F^{(1)\prime}(\tilde{F}_1^{(1)} - \hat{F}^{(1)})$ on the left-hand side of (A.9), i.e.,

$$\Pi_{NT,k} = \frac{\sqrt{(k-k_0)\log(\log T)}}{T} + \frac{k-k_0}{TN^{(1-\alpha)}} + \frac{\sqrt{L_{2NT}}}{\delta_{NT}^2}. \quad \text{(A.10)}$$

Lemma 6 *Under Assumptions 1–8 and for $k > k_0$,*
(a) $\frac{F_k^{(1)\prime}\tilde{F}_k^{(1)}}{k} - \frac{F^{(1)\prime}\tilde{F}_1^{(1)}}{k_0} = O_p(\Pi_{NT,k})$ and $\frac{\Lambda_1'\Theta}{N}\frac{G_k^{(1)\prime}\tilde{F}_k^{(1)}}{k} - \frac{\Lambda_1'\Lambda_1}{N}\frac{F^{(1)\prime}\tilde{F}_1^{(1)}}{k_0} = O_p(\Pi_{NT,k})$;
(b) $\frac{\tilde{F}_1^{(1)\prime}F^{(1)}}{k_0}\frac{\Lambda_1'\Lambda_1}{N}\frac{F^{(1)\prime}\tilde{F}_1^{(1)}}{k_0} - \tilde{V}_k^{(1)} = O_p(\Pi_{NT,k})$;
(c) $\tilde{V}_k^{(1)} - \hat{V}^{(1)} = O_p(\Pi_{NT,k})$;
(d) if the signs of columns in $\tilde{F}^{(1)}$ are properly assigned, then $k_0^{-1} F^{(1)\prime}(\tilde{F}_1^{(1)} - \hat{F}^{(1)}) = O_p(\Pi_{NT,k})$, where $\tilde{F}_1^{(1)}$ denotes the first k_0 rows of $\tilde{F}_k^{(1)}$.

Lemma 6 has some implications on $H_{k,1}^{(1)}$, $H_{k,2}^{(1)}$, and $H_k^{(2)}$. When $\alpha = 1$, Chen (2015) has established the consistency of the estimated break fraction and showed that $\tilde{k}/T - \tau_0 = O_p(\delta_{NT}^{-1})$, i.e., $\tilde{k} - k_0 = O_p(\max[\sqrt{T}, \frac{T}{\sqrt{N}}])$. Hence, for any $m > 0$, $|\tilde{k} - k_0| \le m \cdot \max(T^{3/4}, T/N^{1/4})$ for all large N and T with probability approaching one. In the proofs for consistency, it is sufficient to consider $k \in \mathcal{D}_{NT}$ for the case of $\alpha = 1$, where $\mathcal{D}_{NT} = \{k,\ |k-k_0| \le m_0 \cdot \max(T^{3/4}, T/N^{1/4})\}$ for some $m_0 > 0$. Note that $\sup_{k \in \mathcal{D}_{NT}} \Pi_{NT,k,} \to 0$ as $N, T \to \infty$, so Lemmas 6(a)–6(c) imply that $\frac{\Lambda_1'\Lambda_1}{N}\frac{F^{(1)\prime}\tilde{F}_1^{(1)}}{k_0}$ and $\frac{\Lambda_1'\Theta}{N}\frac{G_k^{(1)\prime}\tilde{F}_k^{(1)}}{k}$ are nonsingular uniformly in $k \in \mathcal{D}_{NT}$ as $N, T \to \infty$ (i.e., their singular values are uniformly bounded away from zero for all $k \in \mathcal{D}_{NT}$ as $N, T \to \infty$). Thus, when $\alpha = 1$, $H_{k,1}^{(1)}$ defined in (A.6) is uniformly nonsingular for all $k \in \mathcal{D}_{NT}$ as $N, T \to \infty$. Similarly, $H_{k,2}^{(1)}$ is also nonsingular because

$$\frac{\Lambda_2'\Theta}{N}\frac{G_k^{(1)\prime}\tilde{F}_k^{(1)}}{k}\tilde{V}_k^{(1)^{-1}} = \frac{\Lambda_2'\Lambda_1}{N}\frac{F^{(1)\prime}\tilde{F}_1^{(1)}}{k}\tilde{V}_k^{(1)^{-1}} + \frac{\Lambda_2'\Lambda_2}{N}\frac{\sum_{t=k_0+1}^{k} f_t\tilde{f}_t'}{k}\tilde{V}_k^{(1)^{-1}},$$

where the second term is $o_p(1)$ uniformly in $k \in \mathcal{D}_{NT}$ and the first term is nonsingular under Assumption 2.

When $0 \le \alpha < 1$, $\Pi_{NT,k,} \to 0$ always holds under (3.11), so $H_{k,1}^{(1)}$ is nonsingular uniformly in $k \in [\tau_1 T, \tau_2 T]$ as $N, T \to \infty$ Lemmas 6(a)–6(c). Also, Lemma 2(c) implies that $H_{k,2}^{(1)}$ is nonsingular for all $k \in [\tau_1 T, \tau_2 T]$ as $N \to \infty$ when $\alpha < 1$.

Lastly, note that the post-k subsample does not undergo a break, $H_k^{(2)}$ is nonsingular uniformly in $k_0 \le k \le \tau_2 T$ by Proposition 1 of Bai (2003).

Lemma 7 *Under Assumptions 1–8,*

$$\frac{1}{k}\sum_{i=1}^{N}\sum_{s=1}^{k_0}(\tilde{f}_s - H_{k,1}^{(1)'} f_s)e_{is}e_{it} = O_p\left(\sqrt{\frac{N}{T}}\right) + O_p\left(\sqrt{\frac{\log\log T}{N}}\right) + O_p\left(\frac{N\sqrt{\log\log N}}{T\sqrt{T}}\right)$$

$$\frac{1}{k}\sum_{i=1}^{N}\sum_{s=k_0+1}^{k}(\tilde{f}_s - H_{k,2}^{(1)'} f_s)e_{is}e_{it} = O_p\left(\frac{\sqrt{L_{2NT}N}}{\sqrt{T}}\right) + +O_p\left(\sqrt{\frac{\log\log T}{N}}\right) + O_p\left(\frac{N\sqrt{\log\log N}}{T\sqrt{T}}\right)$$

uniformly in $k \in [\tau_1 T, \tau_2 T]$ and $k > k_0$.

Proof of Proposition 1:

Let the signs of columns in $\tilde{F}^{(1)}$ be properly assigned so that $R_k = I_r$. For $t \le k_0$, (A.3) implies that

$$\frac{1}{kN}\left(f_t'\Lambda_1'\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + f_t'\Lambda_1'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)} + e_t'\Theta G_k^{(1)'}\tilde{F}_k^{(1)} + e_t'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}\right) = \tilde{f}_t'\tilde{V}_k^{(1)},$$
$$\frac{1}{k_0 N}\left(f_t'\Lambda_1'\Lambda_1 F^{(1)'}\hat{F}^{(1)} + f_t'\Lambda_1'\mathbf{e}_{k_0}^{(1)'}\hat{F}^{(1)} + e_t'\Lambda_1 F^{(1)'}\hat{F}^{(1)} + e_t'\mathbf{e}_{k_0}^{(1)'}\hat{F}^{(1)}\right) = \hat{f}_t'\hat{V}^{(1)}. \tag{A.11}$$

Note that

$$\left\|(\tilde{f}_t' - \hat{f}_t')\tilde{V}_k^{(1)}\right\| = \left\|\tilde{f}_t'\tilde{V}_k^{(1)} - \hat{f}_t'\hat{V}^{(1)} - \hat{f}_t\tilde{V}_k^{(1)} + \hat{f}_t\hat{V}^{(1)}\right\| \le \left\|\tilde{f}_t'\tilde{V}_k^{(1)} - \hat{f}_t'\hat{V}^{(1)}\right\| + \|\hat{f}_t\|\left\|\tilde{V}_k^{(1)} - \hat{V}^{(1)}\right\|.$$

Note that $\tilde{V}_k^{(1)}$ is nonsingular for all $k \in [\tau_1 T, \tau_2 T]$ as $N, T \to \infty$ because the number of nonzero eigenvalues of $X_k^{(1)}X_k^{(1)'}/TN$ as $N, T \to \infty$ is always no less than r by our step. Recall that $\tilde{V}_k^{(1)} - \hat{V}^{(1)} = O_p(\Pi_{NT,k})$ by Lemma 6(c), so it is sufficient to show that $\tilde{f}_t'\tilde{V}_k^{(1)} - \hat{f}_t'\hat{V}^{(1)} = O_p(\Pi_{NT,k})$. To show $k_0^{-1}\sum_{t=1}^{k_0}\|\tilde{f}_t - \hat{f}_t\|^2 = O_p(\Pi_{NT,k}^2)$, it is sufficient to show that $k_0^{-1}\sum_{t=1}^{k_0}\|f_t'(\Lambda_1'\Theta G_k^{(1)'}\tilde{F}_k^{(1)}/kN - \Lambda_1'\Lambda_1 F^{(1)'}\hat{F}^{(1)}/k_0 N)\|^2$, $k_0^{-1}\sum_{t=1}^{k_0}\|f_t'(\Lambda_1'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}/kN - \Lambda_1'\mathbf{e}_{k_0}^{(1)'}\hat{F}^{(1)}/k_0 N)\|^2$, $k_0^{-1}\sum_{t=1}^{k_0}\|e_t'(\Theta G_k^{(1)'}\tilde{F}_k^{(1)}/kN - \Lambda_1 F^{(1)'}\hat{F}^{(1)}/k_0 N)\|^2$, and $k_0^{-1}\sum_{t=1}^{k_0}\|e_t'(\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}/kN - \mathbf{e}_{k_0}^{(1)'}\hat{F}^{(1)}/k_0 N)\|^2$ are all equal to $O_p(\Pi_{NT,k}^2)$.

We consider the terms on the LHS of (A.11). First, by Lemmas 6(a) and 6(d), we have

$$f_t'\left(\frac{\Lambda_1'\Theta}{N}\frac{G_k^{(1)'}\tilde{F}_k^{(1)}}{k} - \frac{\Lambda_1'\Lambda_1}{N}\frac{F^{(1)'}\hat{F}^{(1)}}{k_0}\right) = f_t' O_p(\Pi_{NT,k}), \tag{A.12}$$

where the $O_p(\Pi_{NT,k})$ term does not depend on t.

Second, Lemma 3(a) implies that

$$\frac{1}{kN} f_t' \Lambda_1' \mathbf{e}_k^{(1)'} \tilde{F}_k^{(1)} = f_t' O_p(L_{2NT}^{1/2} \delta_{NT}^{-2}), \quad \frac{1}{k_0 N} f_t' \Lambda_1' \mathbf{e}_{k_0}^{(1)'} \hat{F}^{(1)} = f_t' O_p(\delta_{NT}^{-2}), \tag{A.13}$$

so $f_t'(\Lambda_1' \mathbf{e}_k^{(1)'} \tilde{F}_k^{(1)}/kN - \Lambda_1' \mathbf{e}_{k_0}^{(1)'} \hat{F}^{(1)}/k_0 N) = O_p(L_{2NT}^{1/2} \delta_{NT}^{-2})$ uniformly in k and

$$\begin{aligned} &\frac{1}{k_0} \sum_{t=1}^{k_0} \| f_t' \left(\Lambda_1' \mathbf{e}_k^{(1)'} \tilde{F}_k^{(1)}/KN - \Lambda_1' \mathbf{e}_{k_0}^{(1)'} \hat{F}^{(1)}/k_0 N \right) \|^2 \\ \leq & \frac{2}{k_0} \sum_{t=1}^{k_0} \| f_t \|^2 \left[\| \frac{\Lambda_1' \mathbf{e}_k^{(1)'} \tilde{F}_k^{(1)}}{kN} \|^2 + \| \frac{\Lambda_1' \mathbf{e}_{k_0}^{(1)'} \hat{F}^{(1)}}{k_0 N} \|^2 \right] = O_p(L_{2NT} \delta_{NT}^{-4}). \end{aligned}$$

Third,

$$\begin{aligned} &\frac{e_t' \Theta}{N} \frac{G_k^{(1)'} \tilde{F}_k^{(1)}}{k} - \frac{e_t' \Lambda_1}{N} \frac{F^{(1)'} \hat{F}^{(1)}}{k_0} = \frac{1}{kN} \left(e_t' \Lambda_1 F^{(1)'} \tilde{F}_1^{(1)} + e_t' \Lambda_2 \sum_{s=k_0+1}^{k} f_s \tilde{f}_s' \right) - \frac{1}{k_0 N} e_t' \Lambda_1 F^{(1)'} \hat{F}^{(1)} \\ = & \frac{e_t' \Lambda_1}{N} \left(\frac{F_k^{(1)'} \tilde{F}_k^{(1)}}{k} - \frac{F^{(1)'} \tilde{F}_1^{(1)}}{k_0} + \frac{F^{(1)'} \tilde{F}_1^{(1)}}{k_0} - \frac{F^{(1)'} \hat{F}^{(1)}}{k_0} \right) + \frac{1}{kN} e_t' (\Lambda_2 - \Lambda_1) \sum_{s=k_0+1}^{k} f_s \tilde{f}_s' \\ = & O_p\left(\frac{1}{\sqrt{N}} \right) O_p(\Pi_{NT,k}) + O_p\left(\frac{1}{N^{1-\alpha/2}} \right) O\left(\frac{k-k_0}{T} \right) = o_p(\Pi_{NT,k}) \end{aligned} \tag{A.14}$$

where the last line uses Lemmas 1(c), 6(a), 6(d), and the fact that $O_p\left(\frac{k-k_0}{TN^{1-\alpha/2}}\right)$ is dominated by $O_p\left(\frac{k-k_0}{TN^{1-\alpha}}\right)$. The derivation in (A.14) also implies $k_0^{-1} \sum_{t=1}^{k_0} \| e_t' \Theta G_k^{(1)'} \tilde{F}_k^{(1)}/kN - e_t' \Lambda_1 F^{(1)'} \hat{F}^{(1)}/k_0 N \|^2 = o_p(\Pi_{NT,k}^2)$.

Lastly,

$$\frac{1}{kN} e_t' \mathbf{e}_k^{(1)'} \tilde{F}_k^{(1)} = O_p(L_{2NT}^{1/2} \delta_{NT}^{-2}) \text{ for each } t, \quad \frac{1}{k_0} \sum_{t=1}^{k_0} \| e_t' \mathbf{e}_k^{(1)'} \tilde{F}_k^{(1)}/kN \|^2 = O_p(L_{2NT} \delta_{NT}^{-4}), \tag{A.15}$$

by Lemma 3(b) and similar results hold for $e_t' \mathbf{e}_{k_0}^{(1)'} \hat{F}^{(1)}/k_0 N$. Combining (A.12), (A.13), (A.14), and (A.15), we obtain the desired results.

Q.E.D.

B Proofs for Consistency

We focus on the case $k > k_0$ and the proof for the case $k < k_0$ is the same due to the symmetry. Let $X_{k,i}^{(1)} = [x_{i1}, ..., x_{ik}]'$ and $X_{k,i}^{(2)} = [x_{ik+1}, ..., x_{iT}]'$ for $k > k_0$. We can decompose $SSR(k, \tilde{F}) -$

$SSR(k_0, \hat{F})$ into 3 terms:

$$\begin{aligned}
SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) &= \sum_{i=1}^{N} \sum_{t=1}^{k_0} \left[\left(x_{it} - \tilde{f}_t \tilde{\lambda}_{i1} \right)^2 - \left(x_{it} - \hat{f}_t' \hat{\lambda}_{i1} \right)^2 \right] \\
&+ \sum_{i=1}^{N} \sum_{t=k+1}^{T} \left[\left(x_{it} - \tilde{f}_t' \tilde{\lambda}_{i2} \right)^2 - \left(x_{it} - \hat{f}_t' \hat{\lambda}_{i2} \right)^2 \right] \\
&+ \sum_{i=1}^{N} \sum_{t=k_0+1}^{k} \left[\left(x_{it} - \tilde{f}_t' \tilde{\lambda}_{i1} \right)^2 - \left(x_{it} - \hat{f}_t' \hat{\lambda}_{i2} \right)^2 \right] \\
&= I + II + III. \qquad \text{(B.1)}
\end{aligned}$$

The dependence of terms I, II, and III on k is suppressed in notation. We show that III is positive and dominates the other two terms uniformly in k under certain conditions presented below. The following two lemmas provide useful results: Lemma 8 obtains the rates for terms I and II; Lemma 9 analyzes the rate for term III.

Lemma 8 *Suppose Assumptions 1–8 and* (3.11) *hold.*
(a) Let $A_i(k, k_0) = \left(\tilde{F}_k^{(1)\prime} \tilde{F}_k^{(1)} \right)^{-1} \tilde{F}_k^{(1)\prime} X_{k,i}^{(1)} - \left(\tilde{F}_1^{(1)\prime} \tilde{F}_1^{(1)} \right)^{-1} \tilde{F}_1^{(1)\prime} X_{k_0,i}^{(1)}$. *Then term* I *can be represented as*

$$I = \sum_{i=1}^{N} A_i(k, k_0)' \tilde{F}_1^{(1)\prime} \tilde{F}_1^{(1)} A_i(k, k_0) + (k - k_0) O_p \left(\sqrt{N} + \frac{N}{\delta_{NT}} \right) O_p(\bar{\Pi}_{NT}),$$

where $\sum_{i=1}^{N} A_i(k, k_0)' \tilde{F}_1^{(1)\prime} \tilde{F}_1^{(1)} A_i(k, k_0) = (k - k_0) \left[o_p(1) + O_p \left(N \log(\log T)/T \right) \right] + (k - k_0)^2 O_p(N^{\alpha}/T)$, *the* o_p *and* O_p *terms are uniform in* k*, and*

$$\bar{\Pi}_{NT} \equiv \sqrt{\frac{\log(\log T)}{T}} + \frac{1}{N^{1-\alpha}} + \frac{\sqrt{L_{2NT}}}{\delta_{NT}^2}. \qquad \text{(B.2)}$$

(b) Let $\hat{F}_2^{(2)}$ *denote the last* $T - k$ *rows of* $\hat{F}^{(2)}$ *and let* $B_i(k, k_0) = \left(\hat{F}^{(2)\prime} \hat{F}^{(2)} \right)^{-1} \hat{F}^{(2)\prime} X_{k_0,i}^{(2)} - \left(\hat{F}_2^{(2)\prime} \hat{F}_2^{(2)} \right)^{-1} \hat{F}_2^{(2)\prime} X_{k,i}^{(2)}$. *Then term* II *can be represented as*

$$II = -\sum_{i=1}^{N} B_i(k, k_0)' \hat{F}_2^{(2)\prime} \hat{F}_2^{(2)} B_i(k, k_0) + (k - k_0) O_p \left(\sqrt{N} + \frac{N}{\delta_{NT}} \right) O_p(\bar{\Pi}_{NT}),$$

where $\sum_{i=1}^{N} B_i(k, k_0)' \hat{F}_2^{(2)\prime} \hat{F}_2^{(2)} B_i(k, k_0) = (k - k_0) \left[o_p(1) + O_p \left(N \log(\log T)/T \right) \right]$ *and the* o_p *and* O_p *terms are uniform in* k.

Next, we consider term III in (B.1). Note that by Lemma 2 we have

$$x_{it} = \begin{cases} \tilde{f}_t' H_{k,1}^{(1)^{-1}} \lambda_{i1} + e_{it} + \left(f_t'\lambda_{i1} - \tilde{f}_t' H_{k,1}^{(1)^{-1}} \lambda_{i1} \right) & \text{if } t \leq k_0 \\ \tilde{f}_t' H_{k,2}^{(1)^{-1}} \lambda_{i2} + e_{it} + \left(f_t'\lambda_{i2} - \tilde{f}_t' H_{k,2}^{(1)^{-1}} \lambda_{i2} \right) & \text{if } k_0 \leq t \leq k \\ \hat{f}_t' H_{k_0}^{(2)^{-1}} \lambda_{i2} + e_{it} + \left(f_t'\lambda_{i2} - \hat{f}_t' H_{k_0}^{(2)^{-1}} \lambda_{i2} \right) & \text{if } t \geq k_0, \end{cases} \tag{B.3}$$

where $H_{k_0}^{(2)}$ is the rotation matrix for the post-break subsample when we set $k = k_0$, i.e.,

$$H_{k_0}^{(2)} = \Lambda_2'\Lambda_2 F^{(2)'}\hat{F}^{(2)}\hat{V}^{(2)^{-1}}/N(T-k_0). \tag{B.4}$$

For each i,

$$\begin{aligned} &\sum_{t=k_0+1}^{k} \left[\left(x_{it} - \tilde{f}_t'\tilde{\lambda}_{i1} \right)^2 - \left(x_{it} - \hat{f}_t'\hat{\lambda}_{i2} \right)^2 \right] \\ =& \sum_{t=k_0+1}^{k} \left[f_t'\lambda_{i2} - \tilde{f}_t'\tilde{\lambda}_{i1} + e_{it} \right]^2 - \sum_{t=k_0+1}^{k} \left[f_t'\lambda_{i2} - \hat{f}_t'\hat{\lambda}_{i2} + e_{it} \right]^2 \\ =& z_{1i} - z_{2i} - 2z_{3i} \end{aligned} \tag{B.5}$$

where $z_{1i} = \sum_{t=k_0+1}^{k} \left(\tilde{f}_t'\tilde{\lambda}_{i1} - f_t'\lambda_{i2} \right)^2$, $z_{2i} = \sum_{t=k_0+1}^{k} (\hat{f}_t'\hat{\lambda}_{i2} - f_t'\lambda_{i2})^2$, $z_{3i} = \sum_{t=k_0+1}^{k} e_{it} \left(\tilde{f}_t'\tilde{\lambda}_{i1} - \hat{f}_t'\hat{\lambda}_{i2} \right)$. To prove the consistency of $\tilde{k}$, the following lemma obtains the stochastic bounds of $\sum_{i=1}^{N} z_{1i}$, $\sum_{i=1}^{N} z_{2i}$, and $\sum_{i=1}^{N} z_{3i}$.

Lemma 9 *Under Assumptions 1–10 and* (3.11),
(a) $\sum_{i=1}^{N} z_{1i} \leq (k-k_0) \left[O_p(1) + O_p\left(\frac{\sqrt{N} L_{2NT}^{1/2}}{\delta_{NT}^2} \right) + O_p\left(\frac{\log(\log T)N}{T} \right) \right] + 4\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{2it}^2$, *where the* O_p *terms are uniform in* k, $d_{2it} = \tilde{f}_t'\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}(H_{k,1}^{(1)^{-1}}\lambda_{i1} - H_{k,2}^{(1)^{-1}}\lambda_{i2})/k$ *and* $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{2it}^2$ *diverges at the rate* $N^{\alpha}(k-k_0)$ *if* $k - k_0 \to \infty$ *as* $N, T \to \infty$ *and at the rate* N^{α} *if* $k-k_0$ *is bounded as* $N, T \to \infty$.
(b) $\sum_{i=1}^{N} z_{2i} = (k-k_0) \left[O_p(1) + O_p\left(\frac{N}{T} \right) + O_p\left(\sqrt{\frac{N}{T}} \right) \right]$, *where the* O_p *terms are uniform in* k.
(c) $\sum_{i=1}^{N} z_{3i} = (k-k_0)[O_p(1) + O_p\left(\frac{L_{2NT}^{1/2}\sqrt{N}}{\delta_{NT}^2} \right) + O_p\left(\frac{N^{\alpha}}{N^{1/2}} \right) + O_p\left(\frac{N^{\alpha/2}}{\delta_{NT}} \right) + O_p\left(\sqrt{\frac{NL_{2NT}}{T}} \right) + O_p\left(\frac{N}{T} \right)$ $+ o_p(N^{\alpha})]$, *where the* O_p *and* o_p *terms are uniform in* k.

Proof:
(a) Since $\tilde{\lambda}_{i1} = \tilde{F}_k^{(1)'}X_{k,i}^{(1)}/k$, we can rewrite $\tilde{f}_t'\tilde{\lambda}_{i1} - f_t'\lambda_{i2}$ as

$$\begin{aligned} &\frac{\tilde{f}_t'\left(\tilde{F}_1^{(1)'}F^{(1)}\lambda_{i1} + \sum_{s=k_0+1}^{k}\tilde{f}_s f_s'\lambda_{i2} \right) + \tilde{f}_t'\tilde{F}_k^{(1)'}e_{k,i}^{(1)}}{k} - f_t'\lambda_{i2} \\ =& \frac{\tilde{f}_t'\left(\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1} + \sum_{s=k_0+1}^{k}\tilde{f}_s\tilde{f}_s'H_{k,2}^{(1)^{-1}}\lambda_{i2} \right)}{k} + \frac{\tilde{f}_t'\tilde{F}_k^{(1)'}e_{k,i}^{(1)}}{k} - f_t'\lambda_{i2} \end{aligned}$$

$$+\tilde{f}_t'\underbrace{\frac{\tilde{F}_1^{(1)'}(F^{(1)}-\tilde{F}_1^{(1)}H_{k,1}^{(1)^{-1}})}{k}}_{O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})}\lambda_{i1}+\tilde{f}_t'\underbrace{\frac{\sum_{s=k_0+1}^{k}\tilde{f}_s\left(f_s'-\tilde{f}_s'H_{k,2}^{(1)^{-1}}\right)}{k}}_{O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})}\lambda_{i2}$$

$$=\underbrace{\left(\frac{\tilde{f}_t'\tilde{F}_k^{(1)'}\tilde{F}_k^{(1)}H_{k,2}^{(1)^{-1}}\lambda_{i2}}{k}-f_t'\lambda_{i2}\right)}_{d_{1it}}+\underbrace{\frac{\tilde{f}_t'\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}(H_{k,1}^{(1)^{-1}}\lambda_{i1}-H_{k,2}^{(1)^{-1}}\lambda_{i2})}{k}}_{d_{2it}}+\underbrace{\frac{\tilde{f}_t'\tilde{F}_k^{(1)'}e_{k,i}^{(1)}}{k}}_{d_{3it}}$$

$$+\underbrace{[\tilde{f}_t'O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})\lambda_{i1}+\tilde{f}_t'O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})\lambda_{i2}]}_{d_{4it}},\tag{B.6}$$

where the $O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$ terms are uniform in k by Lemmas 2 and 4 and do not depend on i or t. Note that $\sum_{i=1}^{N} z_{1i}$ is bounded by $4\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}(d_{1it}^2+d_{2it}^2+d_{3it}^2+d_{4it}^2)$, so we will obtain the bounds for $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{jit}^2$ for $j=1,2,3,4$. Our analysis below shows that $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2$ is the dominant term and determines the divergence rate of $\sum_{i=1}^{N}z_{1i}$.

Since $\tilde{F}_k^{(1)'}\tilde{F}_k^{(1)}/k=I_r$, the term d_{1it} in (B.6) reduces to $(\tilde{f}_t'H_{k,2}^{(1)^{-1}}-f_t')\lambda_{i2}$. Recall that

$$\tilde{f}_t'-f_t'H_{k,2}^{(1)}=\frac{1}{kN}e_t'\Theta G_k^{(1)'}\tilde{F}_k^{(1)}\tilde{V}_k^{(1)^{-1}}+\psi_{1t},$$

where $\psi_{1t}=f_t'\Lambda_2'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}\tilde{V}_k^{(1)^{-1}}/kN+e_t'\mathbf{e}_k^{(1)'}\tilde{F}_k^{(1)}\tilde{V}_k^{(1)^{-1}}/kN=O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})$ uniformly in k follows from Lemma 3. Thus, d_{1it} can be represented as

$$\begin{aligned}d_{1it}=(\tilde{f}_t'-f_t'H_{k,2}^{(1)})H_{k,2}^{(1)^{-1}}\lambda_{i2}&=\frac{1}{kN}e_t'\Theta G_k^{(1)'}\tilde{F}_k^{(1)}\tilde{V}_k^{(1)^{-1}}H_{k,2}^{(1)^{-1}}\lambda_{i2}+\psi_{1t}H_{k,2}^{(1)^{-1}}\lambda_{i2}\\&=O_p\left(\frac{1}{\sqrt{N}}\right)\lambda_{i2}+O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})\lambda_{i2}\text{ uniformly in }k.\end{aligned}\tag{B.7}$$

Thus, $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{1it}^2$ can be bounded by

$$\begin{aligned}&\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}\left\{\left[\frac{e_t'\Theta}{N}Z_k\lambda_{i2}\right]^2+2\left\|\frac{e_t'\Theta}{N}\right\|\|Z_k\|\|\psi_{1t}\|\|\lambda_{i2}\|^2\|H_{k,2}^{(1)^{-1}}\|+\|\psi_{1t}\|^2\|\lambda_{i2}\|^2\|H_{k,2}^{(1)^{-1}}\|^2\right\}\\=&(k-k_0)\left[O_p(1)+O_p\left(\frac{\sqrt{N}L_{2NT}^{1/2}}{\delta_{NT}^2}\right)+O_p\left(\frac{NL_{2NT}}{\delta_{NT}^4}\right)\right],\end{aligned}\tag{B.8}$$

where $Z_k\equiv k^{-1}G_k^{(1)'}\tilde{F}_k^{(1)}\tilde{V}_k^{(1)-1}H_{k,2}^{(1)^{-1}}=O_p(1)$ and the O_p terms in the square brackets are uniform in k. We can use the definition of $H_{k,2}^{(1)}$ in (A.6) to further simplify the first term in (B.7) as follows:

$$\frac{e_t'\Theta}{N}\frac{G_k^{(1)'}\tilde{F}_k^{(1)}}{k}\tilde{V}_k^{(1)-1}H_{k,2}^{(1)^{-1}}\lambda_{i2}=\frac{e_t'\Theta G_k^{(1)'}\tilde{F}_k^{(1)}}{Nk}\left(\frac{\Lambda_2'\Theta G_k^{(1)'}\tilde{F}_k^{(1)}}{Nk}\right)^{-1}\lambda_{i2}$$

$$= \left[\underbrace{\frac{e_t' \Lambda_1 F_k^{(1)'} \tilde{F}_k^{(1)}}{Nk}}_{O_p(N^{-1/2})} + \underbrace{\frac{e_t'(\Lambda_2 - \Lambda_1)}{N}}_{O_p(N^{-1+\alpha/2})} \underbrace{\frac{\sum_{s=k_0+1}^k f_s \tilde{f}_s'}{k}}_{O_p(1)} \right] \times \left[\underbrace{\frac{\Lambda_2' \Lambda_1 F_k^{(1)'} \tilde{F}_k^{(1)}}{Nk}}_{O_p(1)} + \underbrace{\frac{\Lambda_2'(\Lambda_2 - \Lambda_1)}{N}}_{O_p(N^{-1+\alpha})} \frac{\sum_{s=k_0+1}^k f_s \tilde{f}_s'}{k} \right]^{-1} \lambda_{i2}$$

$$= \left[\frac{e_t' \Lambda_1}{N} \left(\frac{\Lambda_2' \Lambda_1}{N} \right)^{-1} + O_p\left(\frac{1}{N^{1-\alpha/2}} \right) \right] \lambda_{i2}, \tag{B.9}$$

where the $O_p(N^{-1+\alpha/2})$ term and $O_p(N^{-1+\alpha})$ term in the second equality follow from Lemma 1.

The term d_{3it} can be represented as

$$\begin{aligned}
\frac{\tilde{f}_t' \tilde{F}_k^{(1)'} e_{k,i}^{(1)}}{k} &= \frac{\tilde{f}_t' \left(\tilde{F}_1^{(1)'} e_{k_0,i}^{(1)} + \sum_{s=k_0+1}^k \tilde{f}_s e_{is} \right)}{k} \\
&= \tilde{f}_t' \left(\frac{H_{k,1}^{(1)'} F^{(1)'} e_{k_0,i}^{(1)} + \sum_{s=k_0+1}^k H_{k,2}^{(1)'} f_s e_{is}}{k} \right) \\
&\quad + \tilde{f}_t' \left[\frac{\left(\tilde{F}_1^{(1)} - F^{(1)} H_{k,1}^{(1)} \right)' e_{k_0,i}^{(1)} + \sum_{s=k_0+1}^k (\tilde{f}_s - H_{k,2}^{(1)'} f_s) e_{is}}{k} \right] \\
&= \tilde{f}_t' H_{k,1}^{(1)'} \frac{F_k^{(1)'} e_{k,i}^{(1)}}{k} + \tilde{f}_t' \frac{\sum_{s=k_0+1}^k \left(H_{k,2}^{(1)} - H_{k,1}^{(1)} \right)' f_s e_{is}}{k} + \tilde{f}_t' O_p(L_{2NT}^{1/2} \delta_{NT}^{-2})
\end{aligned} \tag{B.10}$$

where the $O_p(L_{2NT}^{1/2} \delta_{NT}^{-2})$ term follows from Lemma 5 and the fact that

$$\frac{k - k_0}{k} \left(\frac{1}{k - k_0} \sum_{t=k_0+1}^k (\tilde{f}_t - H_{k,2}^{(1)'} f_t) e_{it} \right) = \frac{k - k_0}{k} \left[O_p\left(\frac{L_{2NT}^{1/2}}{\delta_{NT}^2} \right) + \frac{1}{(k - k_0)^{1/2}} O_p\left(\frac{L_{2NT}^{1/2}}{\delta_{NT}} \right) \right] = O_p\left(\frac{L_{2NT}^{1/2}}{\delta_{NT}^2} \right).$$

The uniform bound for term $\tilde{f}_t' H_{k,1}^{(1)'} F_k^{(1)'} e_{k,i}^{(1)}/k$ in (B.10) is $\tilde{f}_t O_p\left(\sqrt{\frac{\log(\log T)}{T}} \right)$ by the LIL under Assumption 4(b). The term $\sum_{s=k_0+1}^k \left(H_{k,2}^{(1)} - H_{k,1}^{(1)} \right)' f_s e_{is}/k = O_p\left(\frac{\sqrt{(k-k_0)\log(\log T)}}{TN^{1-\alpha}} \right) = O_p\left(\frac{\sqrt{\log(\log T)}}{\sqrt{T} N^{1-\alpha}} \right)$ uniformly in k by Lemma 2(c), the LIL and the fact that $|k - k_0| \leq T$. Thus,

$$\sum_{t=k_0+1}^k \sum_{i=1}^N d_{3it}^2 = (k - k_0) \left[O_p\left(\frac{\log(\log T) N}{T} \right) + O_p\left(\frac{\log\log(T)}{TN^{1-2\alpha}} \right) + O_p\left(\frac{L_{2NT} N}{\delta_{NT}^4} \right) \right], \tag{B.11}$$

where the O_p terms in the square brackets are uniform in k.

In addition, $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{4it}^2$ can be bounded by

$$\sum_{i=1}^N \sum_{t=k_0+1}^k \left[\tilde{f}_t' O_p(L_{2NT}^{1/2} \delta_{NT}^{-2}) \lambda_{i1} + \tilde{f}_t' O_p(L_{2NT}^{1/2} \delta_{NT}^{-2}) \lambda_{i2} \right]^2 = (k - k_0) O_p\left(\frac{N L_{2NT}}{\delta_{NT}^4} \right). \tag{B.12}$$

Combining the results in (B.8), (B.11), and (B.12), we obtain

$$\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{1it}^2 + d_{3it}^2 + d_{4it}^2 = (k-k_0)\left[O_p(1) + O_p\left(\frac{\sqrt{N}L_{2NT}^{1/2}}{\delta_{NT}^2}\right) + O_p\left(\frac{\log(\log T)N}{T}\right)\right], \quad \text{(B.13)}$$

because $O_p\left(\frac{\log\log(T)}{TN^{1-2\alpha}}\right)$ is always dominated by $O_p\left(T^{-1}\log(\log T)N\right)$ for $\alpha\in[0,1]$, and $O_p\left(L_{2NT}N\delta_{NT}^{-4}\right)$ is dominated by $O_p(1)$ for $N\le T$ by (3.11) and dominated by $O_p\left(T^{-1}N\log(\log T)\right)$ for $T\le N$ by (3.11).

Next, we analyze the divergence rate of $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{2it}^2$. Rewrite $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{2it}^2$ as

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[\tilde{f}_t'\frac{\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}}{k}H_{k,2}^{(1)^{-1}}(H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1}-\lambda_{i2})\right]^2\\
=&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[[f_t'H_{k,2}^{(1)}+O_p(\delta_{NT}^{-1})]\frac{\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}}{k}H_{k,2}^{(1)^{-1}}(H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1}-\lambda_{i2})\right]^2\\
=&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}(H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1}-\lambda_{i2})'B_k'f_tf_t'B_k(H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1}-\lambda_{i2}) + \text{smaller order},
\end{aligned} \quad \text{(B.14)}$$

where $B_k = H_{k,2}^{(1)}\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}H_{k,2}^{(1)^{-1}}/k$ and there exists $b_1>0$ such that $\lim_{N,T\to\infty}\rho_{min}(B_kB_k')\ge b_1$ for all k considered in the arguments below Lemma 6. Note that the leading term in (B.14) is $O_p((k-k_0)N^\alpha)$ because

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left(f_t'B_kH_{k,2}^{(1)}[H_{k,1}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})+(H_{k,1}^{(1)^{-1}}-H_{k,2}^{(1)^{-1}})\lambda_{i2}]\right)^2\\
\le&2\sum_{t=k_0+1}^{k}\|H_{k,1}^{(1)^{-1'}}H_{k,2}^{(1)'}B_k'f_tf_t'B_kH_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\|\sum_{i=1}^{N}\|\lambda_{i1}-\lambda_{i2}\|^2\\
&+2\sum_{t=k_0+1}^{k}\|H_{k,2}^{(1)'}B_k'f_tf_t'B_kH_{k,2}^{(1)}\|\cdot\|H_{k,1}^{(1)^{-1}}-H_{k,2}^{(1)^{-1}}\|^2\sum_{i=1}^{N}\|\lambda_{i2}\|^2\\
=&(k-k_0)O_p(N^\alpha)+(k-k_0)O_p(N^{-1+2\alpha})
\end{aligned} \quad \text{(B.15)}$$

by Assumption 8 and the fact that $H_{k,1}^{(1)^{-1}} - H_{k,2}^{(1)^{-1}} = O_p(N^{-1+\alpha})$ by Lemma 2(c). For the lower bound, the dominant term in (B.14) reduces to

$$\begin{aligned}
&N^\alpha(k-k_0)\mathrm{tr}\left\{B_k'\left(\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}f_tf_t'\right)B_k\left[\frac{1}{N^\alpha}\sum_{i=1}^{N}(H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1}-\lambda_{i2})(H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1}-\lambda_{i2})'\right]\right\}\\
\ge&N^\alpha(k-k_0)\rho_{min}\left[B_k'\left(\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}f_tf_t'\right)B_k\right]\mathrm{tr}\left\{\left[\frac{1}{N^\alpha}\sum_{i=1}^{N}(\lambda_{i2}-H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1})(\lambda_{i2}-H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1})'\right]\right\}\\
\ge&N^\alpha(k-k_0)\rho_{min}\left(\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}f_tf_t'\right)\rho_{min}(B_kB_k')\left[\frac{1}{N^\alpha}\sum_{i=1}^{N}(\lambda_{i2}-\mathbb{A}'\lambda_{i1})'(\lambda_{i2}-\mathbb{A}'\lambda_{i1})\right],
\end{aligned} \quad \text{(B.16)}$$

where ρ_{min} denotes the smallest eigenvalue, the second line uses the inequality that $\text{tr}(YZ) \geq \rho_{min}(Y)\text{tr}(Z)$ for positive semi-definite matrices Y and Z, and the last line follows from the fact that $\|\Lambda_2 - \Lambda_1 \mathbb{A}\|^2 \leq \|\Lambda_2 - \Lambda_1 \mathcal{A}\|^2 = \sum_{i=1}^N (\lambda_{i2} - \mathcal{A}'\lambda_{i1})'(\lambda_{i2} - \mathcal{A}'\lambda_{i1})$ for any $\mathcal{A}$ by the definition of $\mathbb{A}$.

If k is a sequence such that $k - k_0 \to \infty$ as $T \to \infty$, then $(k-k_0)^{-1}\sum_{t=k_0+1}^k f_t f_t'$ converges to Σ_F a.s. by the Strong Law of Large Numbers (Theorem 3.1, Phillips and Solo, 1992). Thus, for any $\epsilon > 0$, there exists $M_0 > 0$ such that $P\{\sup_{k>k_0+M_0} |\rho_{min}\left((k-k_0)^{-1}\sum_{t=k_0+1}^k f_t f_t'\right) - \rho_{min}(\Sigma_F)| < \frac{1}{2}\rho_{min}(\Sigma_F)\} > 1-\epsilon$, which implies that

$$P\left\{\inf_{k>k_0+M_0} \rho_{min}\left(\frac{1}{k-k_0}\sum_{t=k_0+1}^k f_t f_t'\right) > \frac{1}{2}\rho_{min}(\Sigma_F)\right\} > 1-\epsilon \text{ and}$$

$$P\left(\inf_{k>k_0+M_0}\left[\rho_{min}\left(\frac{\sum_{t=k_0+1}^k f_t f_t'}{k-k_0}\right)\rho_{min}(B_k B_k')\right]\cdot\left[\frac{1}{N^\alpha}\sum_{i=1}^N(\lambda_{i2}-\mathbb{A}'\lambda_{i1})'(\lambda_{i2}-\mathbb{A}'\lambda_{i1})\right] \geq c_1\right) > 1-\epsilon \tag{B.17}$$

for N and T large enough, where $c_1 = \rho_{min}(\Sigma_F)\cdot b_1 c/2$. Hence, there exists $c_1 > 0$ such that the leading term in (B.14) is greater than $N^\alpha(k-k_0)c_1$ for large enough $k - k_0$ with probability approaching one as $N, T \to \infty$.

Next, we consider the case where $k - k_0$ is bounded as $T \to \infty$. Since $\tilde{F}_k^{(1)'}\tilde{F}_k^{(1)}/k = I_r$, it follows that $B_k = I_r + O_p(T^{-1})$ as $k - k_0$ is bounded as $T \to \infty$. The dominant term in (B.14) can be represented as

$$\begin{aligned}
&\sum_{t=k_0+1}^k f_t'\left[\sum_{i=1}^N (H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1} - \lambda_{i2})(H_{k,2}^{(1)}H_{k,1}^{(1)^{-1}}\lambda_{i1} - \lambda_{i2})'\right] f_t \\
\geq & f_{k_0+1}'\left[(\Lambda_2 - \Lambda_1 H_{k,1}^{(1)^{-1'}}H_{k,2}^{(1)'})'(\Lambda_2 - \Lambda_1 H_{k,1}^{(1)^{-1'}}H_{k,2}^{(1)'})\right] f_{k_0+1} \\
\geq & (\Lambda_2 f_{k_0+1} - \Lambda_1 \mathbb{A} f_{k_0+1})'(\Lambda_2 f_{k_0+1} - \Lambda_1 \mathbb{A} f_{k_0+1}) \\
= & N^\alpha f_{k_0+1}'\left[\frac{1}{N^\alpha}(\Lambda_2 - \Lambda_1\mathbb{A})'(\Lambda_2 - \Lambda_1\mathbb{A})\right] f_{k_0+1},
\end{aligned} \tag{B.18}$$

where the second inequality uses the fact that $\|\Lambda_2 f_{k_0+1} - \Lambda_1\mathcal{A}\|^2$ is minimized by setting $\mathcal{A} = (\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2 f_{k_0+1} = \mathbb{A} f_{k_0+1}$. Hence, for a bounded $k - k_0$, (B.18) is of order no smaller than $O_p(N^\alpha)$ by (3.5) as $N, T \to \infty$.

(b) Since $\hat{f}_t$ and $\hat{\lambda}_{i2}$ are standard PC estimators using the post-k_0 data, the analysis of the asymptotics of $\hat{f}_t'\hat{\lambda}_{i2} - f_t'\lambda_{i2}$ is standard. For $t \geq k_0$, we have

$$\hat{f}_t'\hat{\lambda}_{i2} - f_t'\lambda_{i2} = (\hat{f}_t' - f_t' H_{k_0}^{(2)})H_{k_0}^{(2)^{-1}}\lambda_{i2} + \hat{f}_t'(\hat{\lambda}_{i2} - H_{k_0}^{(2)^{-1}}\lambda_{i2}).$$

By (2.6) and Theorem 1 of Bai (2003), we obtain for $t > k_0$

$$\hat{f}_t' - f_t' H_{k_0}^{(2)} = \frac{e_t' \Lambda_2 F^{(2)\prime} \hat{F}^{(2)}}{(T-k_0)N} \hat{V}^{(2)^{-1}} + \psi_{2t}, \tag{B.19}$$

where $\psi_{2t} = \left(f_t' \Lambda_2' \mathbf{e}_{k_0}^{(2)\prime} \hat{F}^{(2)} \hat{V}^{(2)^{-1}} + e_t' \mathbf{e}_{k_0}^{(2)\prime} \hat{F}^{(2)} \hat{V}^{(2)^{-1}} \right) / (T-k_0)N = O_p(\delta_{NT}^{-2})$ uniformly in k by similar arguments in (A.13) and (A.15). (The uniform bound here does not need to be scaled by $L_{2NT}^{1/2}$ because the split date is fixed at k_0). For $\hat{\lambda}_{i2}$, we have

$$\begin{aligned}
\hat{\lambda}_{i2} = \frac{\hat{F}^{(2)\prime} X_{k_0,i}^{(2)}}{T-k_0} &= \frac{\hat{F}^{(2)\prime} \left[\hat{F}^{(2)} H_{k_0}^{(2)^{-1}} \lambda_{i2} + e_{k_0,i}^{(2)} + \left(F^{(2)} - \hat{F}^{(2)} H_{k_0}^{(2)^{-1}} \right) \lambda_{i2} \right]}{T-k_0} \\
&= H_{k_0}^{(2)^{-1}} \lambda_{i2} + \frac{\hat{F}^{(2)\prime} e_{k_0,i}^{(2)}}{T-k_0} + \psi_3 \lambda_{i2},
\end{aligned} \tag{B.20}$$

where $\psi_3 = \hat{F}^{(2)\prime}(F^{(2)} - \hat{F}^{(2)} H_{k_0}^{(2)^{-1}})/(T-k_0) = O_p(\delta_{NT}^{-2})$ uniformly in k by Lemma B3 of Bai (2003). Thus, combining (B.19) and (B.20) yields

$$\hat{f}_t' \hat{\lambda}_{i2} - f_t' \lambda_{i2} = \frac{e_t' \Lambda_2}{N} \left(\frac{\Lambda_2' \Lambda_2}{N} \right)^{-1} \lambda_{i2} + \frac{\hat{f}_t' \hat{F}^{(2)\prime} e_{k_0,i}^{(2)}}{T-k_0} + \psi_{2t} H_{k_0}^{(2)^{-1}} \lambda_{i2} + \hat{f}_t' \psi_3 \lambda_{i2} \tag{B.21}$$

where we use the definition of $H_{k_0}^{(2)}$ in (B.4). Note that $e_t' \Lambda_2 / N = O_p(N^{-1/2})$ by Lemma 1(b) and

$$\frac{\hat{F}^{(2)\prime} e_{k_0,i}^{(2)}}{T-k_0} = \frac{H_{k_0}^{(2)\prime} F^{(2)\prime} e_{k_0,i}^{(2)} + (\hat{F}^{(2)} - F^{(2)} H_{k_0}^{(2)})' e_{k_0,i}^{(2)}}{T-k_0} = \frac{H_{k_0}^{(2)\prime} F^{(2)\prime} e_{k_0,i}^{(2)}}{T-k_0} + O_p(\delta_{NT}^{-2}),$$

where the $O_p(\delta_{NT}^{-2})$ term is uniform in k and follows from Lemma B1 of Bai (2003). Thus,

$$\begin{aligned}
&\sum_{i=1}^{N} \sum_{t=k_0+1}^{k} (\hat{f}_t' \hat{\lambda}_{i2} - f_t' \lambda_{i2})^2 = \sum_{i=1}^{N} \sum_{t=k_0+1}^{k} \left(\frac{e_t' \Lambda_2}{N} \left(\frac{\Lambda_2' \Lambda_2}{N} \right)^{-1} \lambda_{i2} + \frac{\hat{f}_t' \hat{F}^{(2)\prime} e_{k_0,i}^{(2)}}{T-k_0} + (\psi_{2t} H_{k_0}^{(2)^{-1}} + \hat{f}_t' \psi_3) \lambda_{i2} \right)^2 \\
=& \sum_{i=1}^{N} \sum_{t=k_0+1}^{k} \left[\frac{e_t' \Lambda_2}{N} \left(\frac{\Lambda_2' \Lambda_2}{N} \right)^{-1} \lambda_{i2} \right]^2 + \sum_{i=1}^{N} \sum_{t=k_0+1}^{k} \left(\frac{\hat{f}_t' H_{k_0}^{(2)\prime} F^{(2)\prime} e_{k_0,i}^{(2)}}{T-k_0} \right)^2 \\
&+ 2 \sum_{i=1}^{N} \sum_{t=k_0+1}^{k} \left(\frac{e_t' \Lambda_2}{N} \left(\frac{\Lambda_2' \Lambda_2}{N} \right)^{-1} \lambda_{i2} \right) \left(\frac{\hat{f}_t' H_{k_0}^{(2)\prime} F^{(2)\prime} e_{k_0,i}^{(2)}}{T-k_0} \right) + (k-k_0) O_p\left(\frac{N}{\delta_{NT}^3} \right) \\
=& \sum_{i=1}^{N} \sum_{t=k_0+1}^{k} \left[\frac{e_t' \Lambda_2}{N} \left(\frac{\Lambda_2' \Lambda_2}{N} \right)^{-1} \lambda_{i2} \right]^2 + (k-k_0) \left[O_p\left(\frac{N}{T} \right) + O_p\left(\frac{\sqrt{N}}{\sqrt{T}} \right) + O_p\left(\frac{N}{\delta_{NT}^3} \right) \right],
\end{aligned} \tag{B.22}$$

where the first term is $(k-k_0)O_p(1)$ and the $O_p\left(\frac{N}{\delta_{NT}^3}\right)$ term follows from C-S inequality and the facts that $\|N^{-1} e_t' \Lambda_2\| = O_p(N^{-1/2})$, ψ_{2t} in (B.19) is $O_p(\delta_{NT}^{-2})$ for each t, $\|F^{(2)\prime} e_{k_0,i}^{(2)}/(T-k_0)\| = O_p(T^{-1/2})$, $(\hat{F}^{(2)} - F^{(2)} H_{k_0}^{(2)})' e_{k_0,i}^{(2)}/(T-k_0) = O_p(\delta_{NT}^{-2})$ for each i, and ψ_3 defined in (B.20) is $O_p(\delta_{NT}^{-2})$.

(c) Consider term $\sum_{i=1}^N z_{3i}$ in (B.5).

$$\begin{aligned}\sum_{i=1}^N z_{3i} =& \sum_{i=1}^N \sum_{t=k_0+1}^{k} e_{it}\left(\tilde{f}_t' \frac{\tilde{F}_k^{(1)'} X_{k,i}^{(1)}}{k} - \hat{f}_t' \frac{\hat{F}^{(2)'} X_{k_0,i}^{(2)}}{T-k_0}\right)\\ =& \sum_{i=1}^N \sum_{t=k_0+1}^{k} e_{it}\left[\left(\tilde{f}_t' \frac{\tilde{F}_k^{(1)'} X_{k,i}^{(1)}}{k} - f_t'\lambda_{i2}\right) - \left(\hat{f}_t' \frac{\hat{F}^{(2)'} X_{k_0,i}^{(2)}}{T-k_0} - f_t'\lambda_{i2}\right)\right]. \end{aligned} \tag{B.23}$$

The first term in $\sum_{i=1}^N z_{3i}$ can rewritten as

$$\begin{aligned}&\sum_{i=1}^N \sum_{t=k_0+1}^{k} e_{it}\left(\tilde{f}_t' \frac{\tilde{F}_k^{(1)'} X_{k,i}^{(1)}}{k} - f_t'\lambda_{i2}\right)\\ =& \sum_{i=1}^N \sum_{t=k_0+1}^{k} e_{it}\left[\frac{\tilde{f}_t'\left(\tilde{F}_1^{(1)'} F^{(1)}\lambda_{i1} + \sum_{s=k_0+1}^k \tilde{f}_s f_s'\lambda_{i2}\right) + \tilde{f}_t'\tilde{F}_k^{(1)'} e_{k,i}^{(1)}}{k} - f_t'\lambda_{i2}\right]\\ =& \sum_{i=1}^N \sum_{t=k_0+1}^{k} e_{it}\left[\frac{\tilde{f}_t'\left(\tilde{F}_1^{(1)'} \tilde{F}_1^{(1)} H_{k,1}^{(1)^{-1}}\lambda_{i1} + \sum_{s=k_0+1}^k \tilde{f}_s \tilde{f}_s' H_{k,2}^{(1)^{-1}}\lambda_{i2}\right)}{k} + \frac{\tilde{f}_t'\tilde{F}_k^{(1)'} e_{k,i}^{(1)}}{k} - f_t'\lambda_{i2}\right]\\ &+ \sum_{t=k_0+1}^{k}\left[\tilde{f}_t' \underbrace{\frac{\tilde{F}_1^{(1)'}(F^{(1)} - \tilde{F}_1^{(1)} H_{k,1}^{(1)^{-1}})}{k}}_{O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})} \underbrace{\sum_{i=1}^N \lambda_{i1} e_{it}}_{O_p(\sqrt{N})} + \tilde{f}_t' \underbrace{\frac{\sum_{s=k_0+1}^k \tilde{f}_s\left(f_s' - \tilde{f}_s' H_{k,2}^{(1)^{-1}}\right)}{k}}_{O_p(L_{2NT}^{1/2}\delta_{NT}^{-2})} \underbrace{\sum_{i=1}^N \lambda_{i2} e_{it}}_{O_p(\sqrt{N})}\right]\\ =& \sum_{i=1}^N \sum_{t=k_0+1}^{k} e_{it}\left[\left(\frac{\tilde{f}_t'\tilde{F}_k^{(1)'}\tilde{F}_k^{(1)} H_{k,2}^{(1)^{-1}}\lambda_{i2}}{k} - f_t'\lambda_{i2}\right) + \frac{\tilde{f}_t'\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}(H_{k,1}^{(1)^{-1}}\lambda_{i1} - H_{k,2}^{(1)^{-1}}\lambda_{i2})}{k}\right]\\ &+ \sum_{i=1}^N \sum_{t=k_0+1}^{k} e_{it}\frac{\tilde{f}_t'\tilde{F}_k^{(1)'} e_{k,i}^{(1)}}{k} + (k-k_0)O_p\left(\frac{L_{2NT}^{1/2}\sqrt{N}}{\delta_{NT}^2}\right), \end{aligned} \tag{B.24}$$

where the $O_p\left(\frac{L_{2NT}^{1/2}\sqrt{N}}{\delta_{NT}^2}\right)$ term is uniform in k by Lemmas 1(b), 4, and C-S inequality. Since $\tilde{F}_k^{(1)'}\tilde{F}_k^{(1)}/k = I_r$, the first term in (B.24) becomes

$$\begin{aligned}&\sum_{t=k_0+1}^{k}\left[\left(\tilde{f}_t' - f_t' H_{k,2}^{(1)}\right) H_{k,2}^{(1)^{-1}} \sum_{i=1}^N e_{it}\lambda_{i2}\right]\\ =& \sum_{t=k_0+1}^{k} \underbrace{\frac{e_t'\Lambda_1}{N}\left(\frac{\Lambda_2'\Lambda_1}{N}\right)^{-1}\sum_{i=1}^N e_{it}\lambda_{i2}}_{O_p(1)} + (k-k_0)\underbrace{\left[O_p\left(\frac{1}{N^{1/2-\alpha/2}}\right) + O_p\left(\frac{L_{2NT}^{1/2}\sqrt{N}}{\delta_{NT}^2}\right)\right]}_{\text{uniform in } k} \end{aligned} \tag{B.25}$$

by (B.7), (B.9) and Lemma 1(b). The second term in (B.24) is equal to

$$\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[\frac{\tilde{f}_t'\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}(H_{k,1}^{(1)^{-1}}-H_{k,2}^{(1)^{-1}})\lambda_{i1}e_{it}}{k}+\frac{\tilde{f}_t'\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})e_{it}}{k}\right]$$

$$=(k-k_0)O_p\left(\frac{1}{N^{1/2-\alpha}}\right)+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\frac{\left(\tilde{f}_t'-f_t'H_{k,2}^{(1)}\right)\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})e_{it}}{k}$$

$$+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\frac{f_t'H_{k,2}^{(1)}\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})e_{it}}{k}, \tag{B.26}$$

where the $O_p\left(\frac{1}{N^{1/2-\alpha}}\right)$ term is uniform in k and follows from Lemmas 1(b) and 2(c). Note that the second term in (B.26) can be bounded by

$$\sum_{t=k_0+1}^{k}\left(\tilde{f}_t'-f_t'H_{k,2}^{(1)}\right)\frac{\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}}{k}H_{k,2}^{(1)^{-1}}\sum_{i=1}^{N}(\lambda_{i1}-\lambda_{i2})e_{it}=(k-k_0)O_p\left(\frac{N^{\alpha/2}}{\delta_{NT}}\right), \tag{B.27}$$

where the $O_p(\delta_{NT}^{-1}N^{\alpha/2})$ term is uniform in k by Lemmas 1(c), 2, and C-S inequality. For the third term in (B.26), let

$$\xi_t=\sum_{i=1}^{N}f_t'(\lambda_{i1}-\lambda_{i2})e_{it}. \tag{B.28}$$

Lemma 1(d) implies that $E\left(\sup_{k\geq k_0+1}|\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\xi_t|^2\right)\leq C_1MN^{\alpha}$ for some $0<C_1<\infty$. Thus, for any small $\varepsilon>0$

$$P\left(\sup_{k\geq k_0+1}|\frac{\sum_{t=k_0+1}^{k}\xi_t}{k-k_0}|\geq\varepsilon N^{\alpha}\right)=P\left(\sup_{k\geq k_0+1}|\frac{\sum_{t=k_0+1}^{k}\xi_t}{k-k_0}|^2\geq\varepsilon^2N^{2\alpha}\right)$$
$$\leq\frac{C_1MN^{\alpha}}{\varepsilon^2N^{2\alpha}}=O\left(\frac{1}{N^{\alpha}}\right)\to 0 \tag{B.29}$$

for $\alpha>0$ by Markov inequality, so $\sum_{t=k_0+1}^{k}\xi_t$ is uniformly dominated by $\varepsilon(k-k_0)N^{\alpha}$ for arbitrary small $\varepsilon>0$ as $N\to\infty$, which means that $(k-k_0)^{-1}\sum_{t=k_0+1}^{k}\xi_t$ is $o_p(N^{\alpha})$ uniformly in k. Note that $\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}/k$, $H_{k,2}^{(1)}$ and $H_{k,2}^{(1)^{-1}}$ are uniformly $O_p(1)$. Thus, $\mathbb{D}_k$ is also $(k-k_0)o_p(N^{\alpha})$ for $\alpha>0$ and the $o_p(N^{\alpha})$ term is uniform in k, where

$$\mathbb{D}_k=\mathrm{tr}\left(\left[H_{k,2}^{(1)}\left(\frac{\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}}{k}\right)H_{k,2}^{(1)^{-1}}\right]\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}(\lambda_{i1}-\lambda_{i2})e_{it}f_t'\right). \tag{B.30}$$

In addition, the third term of (B.24), $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}e_{it}\tilde{f}_t'\tilde{F}_k^{(1)'}e_{k,i}^{(1)}/k$, can be rewritten as

$$\frac{1}{k}\sum_{t=k_0+1}^{k}\tilde{f}_t'\sum_{i=1}^{N}\sum_{s=1}^{k}H_{k,1}^{(1)'}f_se_{is}e_{it}+\frac{1}{k}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}\tilde{f}_t'\sum_{s=k_0+1}^{k}\left(H_{k,2}^{(1)}-H_{k,1}^{(1)}\right)f_se_{is}e_{it}$$

$$+\frac{1}{k}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}\tilde{f}_t'\sum_{s=1}^{k_0}(\tilde{f}_s-H_{k,1}^{(1)'}f_s)e_{is}e_{it}+\frac{1}{k}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}\tilde{f}_t'\sum_{s=k_0+1}^{k}(\tilde{f}_s-H_{k,2}^{(1)'}f_s)e_{is}e_{it}$$
$$=\omega_1+\omega_2+\omega_3+\omega_4 \tag{B.31}$$

Term ω_1 can be represented as

$$\omega_1=\sum_{t=k_0+1}^{k}\tilde{f}_t'H_{k,1}^{(1)'}\frac{1}{k}\sum_{i=1}^{N}\sum_{s=1}^{k}f_s[e_{is}e_{it}-E(e_{is}e_{it})]+\sum_{t=k_0+1}^{k}\tilde{f}_t'\frac{1}{k}\sum_{i=1}^{N}\sum_{s=1}^{k}H_{k,1}^{(1)'}f_sE(e_{is}e_{it})$$
$$=(k-k_0)\left[O_p\left(\sqrt{\frac{N}{T}}\right)+O_p\left(\frac{N}{T}\right)\right], \tag{B.32}$$

where the $O_p\left(\sqrt{\frac{N}{T}}\right)$ term is uniform in k and follows from Assumption 5(a) and $k/T\in[\tau_1,\tau_2]$, and the $O_p\left(\frac{N}{T}\right)$ term is uniform in k and follows from the fact that $k^{-1}E\|\sum_{i=1}^{N}\sum_{s=1}^{k}f_sE(e_{is}e_{it})\|\leq k^{-1}N(\sum_{s=1}^{k}|\gamma_N(s,t)|)E\|f_s\|=O(N/T)$. Term ω_2 can be represented as

$$\omega_2=\frac{1}{k}\sum_{t=k_0+1}^{k}\tilde{f}_t'\left(H_{k,2}^{(1)}-H_{k,1}^{(1)}\right)\sum_{i=1}^{N}\sum_{s=k_0+1}^{k}\{f_s[e_{is}e_{it}-E(e_{is}e_{it})]+f_sE(e_{is}e_{it})\}.$$

The term $k^{-1}\sum_{i=1}^{N}\sum_{s=k_0+1}^{k}f_s[e_{is}e_{it}-E(e_{is}e_{it})]=O_p\left(\frac{\sqrt{N(k-k_0)L_{2NT}}}{T}\right)$ uniformly in k by Assumption 6(a). The term $k^{-1}\sum_{i=1}^{N}\sum_{s=k_0+1}^{k}f_sE(e_{is}e_{it})=O_p(N/T)$ uniformly in k by the same argument as the second term in (B.32). Since $H_{k,2}^{(1)}-H_{k,1}^{(1)}=O_p(N^{\alpha-1})$ by Lemma 2(c) and $k/T\in[\tau_1,\tau_2]$, we obtain

$$\omega_2=(k-k_0)\left[O_p\left(\frac{\sqrt{NL_{2NT}}}{\sqrt{T}N^{1-\alpha}}\right)+O_p\left(\frac{N}{TN^{1-\alpha}}\right)\right], \tag{B.33}$$

where the O_p terms in (B.33) are uniform in k. Next, term ω_3 can be bounded by

$$\left\|\sum_{t=k_0+1}^{k}\tilde{f}_t'\left(\frac{1}{k}\sum_{i=1}^{N}\sum_{s=1}^{k_0}(\tilde{f}_s-H_{k,1}^{(1)'}f_s)e_{is}e_{it}\right)\right\|$$
$$\leq(k-k_0)\left(\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\|\tilde{f}_t\|^2\right)^{1/2}\left(\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\left\|\frac{1}{k}\sum_{i=1}^{N}\sum_{s=1}^{k_0}(\tilde{f}_s-H_{k,1}^{(1)'}f_s)e_{is}e_{it}\right\|^2\right)^{1/2}$$
$$=(k-k_0)\left[O_p\left(\sqrt{\frac{N}{T}}\right)+O_p\left(\sqrt{\frac{\log\log T}{N}}\right)+O_p\left(\frac{N\sqrt{\log\log N}}{T\sqrt{T}}\right)\right], \tag{B.34}$$

where the O_p terms in (B.34) are uniform in k by Lemma 7. The $O_p\left(\frac{N\sqrt{\log\log N}}{T\sqrt{T}}\right)$ term in (B.34) is dominated by $O_p(N/T)$ in (B.32) under (3.11). By similar arguments to ω_3, it follows that

$$\omega_4 = (k-k_0)\left\{O_p\left(\frac{\sqrt{L_{2NT}N}}{\sqrt{T}}\right)+O_p\left(\sqrt{\frac{\log\log T}{N}}\right)+O_p\left(\frac{N\sqrt{\log\log N}}{T\sqrt{T}}\right)\right\} \tag{B.35}$$

by Lemma 7. The last term in (B.35) is also dominated by (B.32) by (3.11). The $O_p\left(\sqrt{\frac{\log\log T}{N}}\right)$ term in (B.34) and (B.35) is dominated by $O_p\left(\frac{N\sqrt{\log\log N}}{T\sqrt{T}}\right)$ for $T < N$ and has the same rate as the $O_p\left(\frac{\sqrt{N\cdot L_{2NT}}}{\delta_{NT}^2}\right)$ term in (B.24) for $T \geq N$. By (B.32)–(B.35), we can bound (B.31) as $(k-k_0)[O_p(N/T)+O_p(\sqrt{L_{2NT}N/T})+O_p(\delta_{NT}^{-2}\sqrt{N\cdot L_{2NT}})]$.

Lastly, the second term in (B.23) can be rewritten as

$$\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} e_{it}\left(\hat{f}_t'\hat{\lambda}_{i2}-f_t'\lambda_{i2}\right)=\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[(\hat{f}_t'-f_t'H_{k_0}^{(2)})H_{k_0}^{(2)^{-1}}\lambda_{i2}e_{it}+\hat{f}_t'(\hat{\lambda}_{i2}-H_{k_0}^{(2)^{-1}}\lambda_{i2})e_{it}\right] \tag{B.36}$$

The term $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}(\hat{f}_t'-f_t'H_{k_0}^{(2)})H_{k_0}^{(2)^{-1}}\lambda_{i2}e_{it}$ can be represented as

$$\begin{aligned}&\sum_{t=k_0+1}^{k}\left(\left[\frac{e_t'\Lambda_2}{N}\left(\frac{\Lambda_2'\Lambda_2}{N}\right)^{-1}+\psi_{2t}H_{k_0}^{(2)^{-1}}\right]\underbrace{\sum_{i=1}^{N}\lambda_{i2}e_{it}}_{O_p(\sqrt{N})}\right)\\ =&\sum_{t=k_0+1}^{k}\underbrace{\frac{e_t'\Lambda_2}{N}\left(\frac{\Lambda_2'\Lambda_2}{N}\right)^{-1}\sum_{i=1}^{N}\lambda_{i2}e_{it}}_{O_p(1)}+(k-k_0)\underbrace{O_p\left(\frac{\sqrt{N}}{\delta_{NT}^2}\right)}_{\text{uniform in } k},\end{aligned} \tag{B.37}$$

by Lemma 1(b), (B.19) and the definition of $H_{k_0}^{(2)}$ in (B.4). By (B.20), the term $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\hat{f}_t'(\hat{\lambda}_{i2}-H_{k_0}^{(2)^{-1}}\lambda_{i2})e_{it}$ can be rewritten as

$$\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\hat{f}_t'\left[\frac{\hat{F}^{(2)'}e_{k_0,i}^{(2)}}{T-k_0}+\psi_3\lambda_{i2}\right]e_{it}=\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\frac{\hat{f}_t'\hat{F}^{(2)'}e_{k_0,i}^{(2)}e_{it}}{T-k_0}+(k-k_0)\underbrace{O_p\left(\frac{\sqrt{N}}{\delta_{NT}^2}\right)}_{\text{uniform in } k}, \tag{B.38}$$

where $\psi_3 = O_p(\delta_{NT}^{-2})$ uniformly in k is defined in (B.20) and we use Lemma 1(b).

The term $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\hat{f}_t'\hat{F}^{(2)'}e_{k_0,i}^{(2)}e_{it}/(T-k_0)$ in (B.38) can be represented as

$$\begin{aligned}&\frac{1}{T-k_0}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\hat{f}_t'H_{k_0}^{(2)'}F^{(2)'}e_{k_0,i}^{(2)}e_{it}+\frac{1}{T-k_0}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\hat{f}_t'\left(\hat{F}^{(2)}-F^{(2)}H_{k_0}^{(2)}\right)e_{k_0,i}^{(2)}e_{it}\\ =&(k-k_0)\left[O_p\left(\sqrt{\frac{N}{T}}\right)+O_p\left(\frac{N}{T}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)\right]\end{aligned} \tag{B.39}$$

The proof for (B.39) is the same as ω_1 in (B.32) and ω_3 in (B.34). (The uniform bound here does not involve $L_{2NT}^{1/2}$ because the split date is fixed at k_0 for the estimation of $\hat{f}_t$). Thus, combining the results in (B.24), (B.25), (B.26), (B.27), (B.29), (B.31), (B.32), (B.33), (B.34), (B.35), (B.37), (B.38), and (B.39), we obtain the bound for $\sum_{i=1}^N z_{3i}$.
Q.E.D.

Proof of Theorem 1:

By (B.1) and (B.5), we have

$$\frac{1}{k-k_0}[SSR(k,\tilde{F}) - SSR(k_0,\hat{F})] = \frac{1}{k-k_0}\left(I + II + \sum_{i=1}^N z_{1i} - \sum_{i=1}^N z_{2i} - 2\sum_{i=1}^N z_{3i}\right), \tag{B.40}$$

where $\sum_{i=1}^N z_{1i} = \sum_{i=1}^N \sum_{t=k_0+1}^k (d_{1it} + d_{2it} + d_{3it} + d_{4it})^2$ by (B.6). First, we will show that $(k-k_0)^{-1}\sum_{i=1}^N \sum_{t=k_0+1}^k d_{jit}^2$ for $j = 1,3,4$, $(k-k_0)^{-1}\sum_{i=1}^N z_{2i}$, $(k-k_0)^{-1}\sum_{i=1}^N z_{3i}$, and the terms in $(k-k_0)^{-1}I$ and $(k-k_0)^{-1}II$ are $o_p(N^\alpha)$ uniformly in k.

We focus on the case of $k > k_0$ without loss of generality. Note that the term $\sum_{i=1}^N A_i(k,k_0)'$ $\tilde{F}_1^{(1)\prime}\tilde{F}_1^{(1)}A_i(k,k_0)$ defined in part (a) of Lemma 8 is always non-negative, so we only need to show that the remaining terms in I of Lemma 8(a) are $(k-k_0)o_p(N^\alpha)$ and the $o_p(N^\alpha)$ term is uniform in k. The remaining term in I is $(k-k_0)O_p\left(\sqrt{N}\bar{\Pi}_{NT} + \frac{N\bar{\Pi}_{NT}}{\delta_{NT}}\right)$ and the O_p terms are uniform in k by Lemma 8(a). The $O_p\left(\bar{\Pi}_{NT}\sqrt{N}\right)$ term is equal to $O_p\left(\sqrt{\frac{N\log(\log T)}{T}} + \frac{\sqrt{N}}{N^{1-\alpha}} + \frac{\sqrt{NL_{2NT}}}{\delta_{NT}^2}\right)$. Note that $\frac{\sqrt{N\log(\log T)}}{\sqrt{T}N^\alpha} = \frac{N^{(1-\alpha)/2}\sqrt{\log\log T}}{\sqrt{T}}N^{-\alpha/2} \to 0$ by (3.10) and $\frac{\sqrt{N}}{N^{1-\alpha}}/N^\alpha = N^{-1/2} \to 0$. For the term $O_p\left(\frac{\sqrt{NL_{2NT}}}{\delta_{NT}^2}\right)$, it follows that

$$\begin{aligned}
\frac{\sqrt{NL_{2NT}}}{\delta_{NT}^2 N^\alpha} &\leq \frac{N^{1-\alpha}}{T}\sqrt{\frac{\log(2\log N)}{N}} \to 0, \text{ if } N > T \\
\frac{\sqrt{NL_{2NT}}}{\delta_{NT}^2 N^\alpha} &\leq \frac{\sqrt{\log(2\log T)}}{N^{1/2+\alpha}} \to 0, \text{ if } N \leq T
\end{aligned} \tag{B.41}$$

by (3.10) and (3.11). Thus, the $O_p\left(\bar{\Pi}_{NT}\sqrt{N}\right)$ term is uniformly $o_p(N^\alpha)$. The $O_p\left(\frac{\bar{\Pi}_{NT}N}{\delta_{NT}}\right)$ term is the same as the $O_p\left(\bar{\Pi}_{NT}\sqrt{N}\right)$ term for $N \leq T$. For $N > T$, it reduces to

$$\begin{aligned}
&O_p\left(\frac{N\sqrt{\log(\log T)}}{T} + \frac{N}{\sqrt{T}N^{1-\alpha}} + \frac{N\sqrt{L_{2NT}}}{T^{3/2}}\right) \\
\leq &O_p\left(\frac{N\sqrt{\log(\log T)}}{T} + \frac{N^\alpha}{\sqrt{T}} + \frac{N}{T}\sqrt{\frac{\log(2\log N)}{T}}\right),
\end{aligned}$$

which is uniformly $o_p(N^\alpha)$ by (3.10) and (3.11).

The analysis for term II is almost the same as that for term I except that II has a non-positive term $-\sum_{i=1}^N B_i(k,k_0)'\hat{F}_2^{(2)'}\hat{F}_2^{(2)}B_i(k,k_0)$, which is $(k-k_0)\left[o_p(1)+O_p\left(N\log(\log T)/T\right)\right]$ by Lemma 8(b). Since $N\log(\log T)/T$ is dominated by N^α by (3.10) as $N,T\to\infty$, it follows that $-(k-k_0)^{-1}\sum_{i=1}^N B_i(k,k_0)'\hat{F}_2^{(2)'}\hat{F}_2^{(2)}B_i(k,k_0)$ is uniformly $o_p(N^\alpha)$. Thus, the non-positive terms of $(k-k_0)^{-1}I$ and $(k-k_0)^{-1}II$ in Lemma 8 are $o_p(N^\alpha)$ uniformly in k.

Next, the stochastic bounds of $\sum_{i=1}^N\sum_{t=k_0+1}^k d_{jit}^2$ for $j=1,3,4$, $\sum_{i=1}^N z_{2i}$, $\sum_{i=1}^N z_{3i}$ are given in parts (a)–(c) of Lemma 9 for $\alpha>0$. It is sufficient to show that all these bounds in Lemma 9 are $(k-k_0)o_p(N^\alpha)$ and the $o_p(N^\alpha)$ term is uniform in k. By checking the terms in part (a) of Lemma 9, it is straightforward that $(k-k_0)O_p(1)$ is uniformly dominated by $N^\alpha(k-k_0)$ for $\alpha>0$. The $O_p\left(\frac{\sqrt{NL_{2NT}}}{\delta_{NT}^2}\right)$ term is $o_p(N^\alpha)$ uniformly in k by (B.41). Under the condition (3.10), the $O_p\left(\frac{N\log(\log T)}{T}\right)$ term in Lemma 9 is also $o_p(N^\alpha)$.

Also, $\sum_{i=1}^N z_{2i}$ is always dominated by $\sum_{i=1}^N z_{3i}$ by Lemma 9, so the rest of the proofs will only focus on $\sum_{i=1}^N z_{3i}$. The terms $O_p\left(\frac{N^\alpha}{N^{1/2}}\right)$ and $O_p\left(\frac{N^{\alpha/2}}{\delta_{NT}}\right)$ are always $o_p(N^\alpha)$ for $\alpha\in[0,1]$. The $O_p\left(\frac{\sqrt{NL_{2NT}}}{\sqrt{T}}\right)$ term is $o_p(N^\alpha)$ because for $\alpha>0$

$$\frac{\sqrt{NL_{2NT}}}{N^\alpha\sqrt{T}}\leq\begin{cases}\frac{1}{N^{\alpha/2}}\sqrt{\frac{N^{1-\alpha}\log(2\log T)}{T}} & \to 0,\ \text{for } T\geq N;\\ \sqrt{\frac{\log(2\log N)}{N^\alpha}}\sqrt{\frac{N^{1-\alpha}}{T}} & \to 0,\ \text{for } T<N\end{cases}$$

by (3.10). The term $(k-k_0)O_p\left(\frac{N}{T}\right)$ is uniformly $(k-k_0)o_p(N^\alpha)$ by (3.10).

Based on the above results, the term in the right-hand side of (B.40) can be rewritten as

$$\begin{aligned}&I+II+\sum_{i=1}^N z_{1i}-\sum_{i=1}^N z_{2i}-2\sum_{i=1}^N z_{3i}\\ =&\mathbb{Q}_k+\sum_{i=1}^N\sum_{t=k_0+1}^k d_{2it}^2+(I-\mathbb{Q}_k)+II-\sum_{i=1}^N z_{2i}-2\sum_{i=1}^N z_{3i}\\ &+\sum_{i=1}^N\sum_{t=k_0+1}^k(d_{1it}+d_{3it}+d_{4it})^2+2\sum_{i=1}^N\sum_{t=k_0+1}^k d_{2it}(d_{1it}+d_{3it}+d_{4it})\end{aligned}\tag{B.42}$$

where $\mathbb{Q}_k=\sum_{i=1}^N A_i(k,k_0)'\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}A_i(k,k_0)\geq 0$. We have shown that $(k-k_0)^{-1}\sum_{i=1}^N\sum_{t=k_0+1}^k d_{jit}^2$ for $j=1,3,4$, $(k-k_0)^{-1}\sum_{i=1}^N z_{2i}$, $(k-k_0)^{-1}\sum_{i=1}^N z_{3i}$, and $(k-k_0)^{-1}(I-\mathbb{Q}_k)$ and $(k-k_0)^{-1}II$ are $o_p(N^\alpha)$ uniformly in k for $\alpha>0$. In conjunction with (B.15), it follows that the last term in (B.42) is $(k-k_0)o_p(N^\alpha)$ by C-S inequality and the o_p term is uniform in k.

Given the above results, we next prove that $\tilde{k}-k_0=O_p(1)$. We need to show that for any $\epsilon_0>0$, there exists $M>0$, $n>0$ and $S>0$ such that $P(\tilde{k}-k_0>M)<\epsilon_0$, $\forall N\geq n$ and $\forall T\geq S$ (note that $\tilde{k}$ depends on N and T). It is equivalent to show that

$$P\left(\min_{k>k_0+M}SSR(k,\tilde{F})-SSR(k_0,\hat{F})\leq 0\right)$$

$$=P\left(\min_{k>k_0+M}[I+II+\sum_{i=1}^{N}z_{1i}-\sum_{i=1}^{N}z_{2i}-2\sum_{i=1}^{N}z_{3i}]\leq 0\right)<\epsilon_0,\ \forall N\geq n,\ \forall T\geq S \tag{B.43}$$

Recall the results in Eq.(B.16) and (B.17) for the leading term in $(k-k_0)^{-1}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2$. Specifically, there exists $k-k_0=M_\epsilon>0$ and n_0 and $S_0>0$, such that

$$P\left(\inf_{k>k_0+M_\epsilon}\frac{1}{k-k_0}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2<N^\alpha c_1\right)\leq\frac{\epsilon_0}{2},\ \forall N\geq n_0,\ \forall T\geq S_0, \tag{B.44}$$

where c_1 is defined below (B.17).

Also, let $W_{k,NT}=(I-\mathbb{Q}_k)+II-\sum_{i=1}^{N}z_{2i}-2\sum_{i=1}^{N}z_{3i}+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}(d_{1it}+d_{3it}+d_{4it})^2+2\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}(d_{1it}+d_{3it}+d_{4it})$. We have shown that $(k-k_0)^{-1}W_{k,NT}$ is $o_p(N^\alpha)$ uniformly in k. Thus, there exists n_1 and $S_1>0$ such that

$$P\left(\sup_{k_0<k\leq\tau_2 T}\frac{1}{k-k_0}|W_{k,NT}|\geq c_2N^\alpha\right)<\frac{\epsilon_0}{2},\ \forall N\geq n_1,\ \forall T\geq S_1 \tag{B.45}$$

and for arbitrarily small $c_2>0$. It is sufficient to set $c_2=c_1/2$ for our purpose. Let $M=M_0$, $n=\max(n_0,n_1)$, $S=\max(S_0,S_1)$ and

$$\breve{k}=\arg\min_{k>k_0+M}\left(SSR(k,\tilde{F})-SSR(k_0,\hat{F})\right),$$

so $\breve{k}-k_0>M$ by its definition. Hence, for any $N\geq n$ and $T\geq S$, the probability in (B.43) becomes

$$\begin{aligned}
&P\left(\min_{k>k_0+M}[I+II+\sum_{i=1}^{N}z_{1i}-\sum_{i=1}^{N}z_{2i}-2\sum_{i=1}^{N}z_{3i}]\leq 0\right)=P\left\{\min_{k>k_0+M}\left(\mathbb{Q}_k+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2+W_{k,NT}\right)\leq 0\right\}\\
\leq&P\left\{\min_{k>k_0+M}\left(\mathbb{Q}_k+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2+W_{k,NT}\right)\leq 0,\ \frac{1}{\breve{k}-k_0}\sum_{i=1}^{N}\sum_{t=k_0+1}^{\breve{k}}d_{2it}^2\geq N^\alpha c_1\right\}+\frac{\epsilon_0}{2}\\
<&P\left\{\min_{k>k_0+M}\left(\mathbb{Q}_k+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2+W_{k,NT}\right)\leq 0,\ \frac{1}{\breve{k}-k_0}\sum_{i=1}^{N}\sum_{t=k_0+1}^{\breve{k}}d_{2it}^2\geq N^\alpha c_1,\ \sup_{k_0<k\leq\tau_2T}\frac{|W_{k,NT}|}{k-k_0}<c_2N^\alpha\right\}+\epsilon_0\\
<&P\left\{\min_{k>k_0+M}\left(\mathbb{Q}_k+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2+W_{k,NT}\right)\leq 0,\ \frac{1}{\breve{k}-k_0}\sum_{i=1}^{N}\sum_{t=k_0+1}^{\breve{k}}d_{2it}^2\geq N^\alpha c_1,\ \inf_{k_0<k\leq\tau_2T}\frac{W_{k,NT}}{k-k_0}\geq -c_2N^\alpha\right\}+\epsilon_0\\
<&P\left\{\min_{k>k_0+M}\left(\mathbb{Q}_k+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2+W_{k,NT}\right)\leq 0,\ \frac{1}{\breve{k}-k_0}\sum_{i=1}^{N}\sum_{t=k_0+1}^{\breve{k}}d_{2it}^2\geq N^\alpha c_1,\ \frac{1}{\breve{k}-k_0}W_{\breve{k},NT}\geq -c_2N^\alpha\right\}+\epsilon_0\\
\leq&P\left\{\min_{k>k_0+M}\left(\mathbb{Q}_k+\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}d_{2it}^2+W_{k,NT}\right)\leq 0,\ \frac{1}{\breve{k}-k_0}\left(\sum_{i=1}^{N}\sum_{t=k_0+1}^{\breve{k}}d_{2it}^2+W_{\breve{k},NT}\right)\geq\frac{1}{2}N^\alpha c_1\right\}+\epsilon_0\\
=&\epsilon_0,
\end{aligned}$$

where the second line uses (B.44), the third line uses (B.45) with $c_2=c_1/2$, and the last line follows from the fact that the probability of the event in the large brackets is zero by the definition of $\breve{k}$. This completes the proof for $\tilde{k} - k_0 = O_p(1)$.

Lastly, we show that $P(\tilde{k} = k_0) \to 1$ as $N, T \to \infty$, based on the result that $\tilde{k} - k_0 = O_p(1)$. Recall that $SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) = \mathbb{Q}_k + \sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2 + W_{k,NT}$ for $k > k_0$, so

$$P(\tilde{k} > k_0) = P\left(\mathbb{Q}_{\breve{k}} + \sum_{i=1}^{N} \sum_{t=k_0+1}^{\tilde{k}} d_{2it}^2 + W_{\tilde{k},NT} \leq 0\right), \tag{B.46}$$

where $\mathbb{Q}_{\breve{k}}$ is $\mathbb{Q}_k$ evaluated at $\breve{k}$. By similar arguments for (B.45), $W_{\tilde{k},NT}$ is less than $c_3 N^{\alpha}$ for arbitrarily small $c_3 > 0$ as $N, T \to \infty$, given that $\tilde{k} - k_0 = O_p(1)$. Also, $\mathbb{Q}_k \geq 0$ and $N^{-\alpha} \sum_{i=1}^N \sum_{t=k_0+1}^{\tilde{k}} d_{2it}^2$ is greater than $f_{k_0+1}' \left[\lim_{N\to\infty} \frac{1}{N^{\alpha}} (\Lambda_2 - \Lambda_1 \mathbb{A})'(\Lambda_2 - \Lambda_1 \mathbb{A})\right] f_{k_0+1} > 0$ as $N, T \to \infty$ by (3.5) and (B.18). Thus, $\sum_{i=1}^N \sum_{t=k_0+1}^{\tilde{k}} d_{2it}^2$ dominates $W_{\tilde{k},NT}$ and the probability in (B.46) tends to zero as $N, T \to \infty$.

Q.E.D.

C Proofs for the asymptotics of $\tilde{k} - k_0$ when $\alpha = 0$

Proof of Theorem 2:

To show that $\tilde{k} - k_0$ is $O_p(1)$ when $\alpha = 0$, it is sufficient to show that for any $\epsilon > 0$, there exists $S < \infty$ such that

$$P\left(\left[\min_{|k-k_0|>S} SSR(k, \tilde{F}) - SSR(k_0, \hat{F})\right] \leq 0\right) < \epsilon$$

for all large N, T with $T^{-1}(N \log\log T) \to 0$ and $N^{-1} \log\log T \to 0$. We will show that $SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) > 0$ for all $|k - k_0| > S$ and $k \in [\tau_1 T, \tau_2 T]$. Note that the leading term $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2$ in (B.42) is no smaller than $O_p(k - k_0)$ by (B.16) and (B.18) for $\alpha = 0$. Also, the term $\mathbb{Q}_k = \sum_{i=1}^N A_i(k, k_0)' \tilde{F}_1^{(1)'} \tilde{F}_1^{(1)} A_i(k, k_0)$ is always non-negative. The terms in $I - \mathbb{Q}_k$ and II in (B.42) are $(k - k_0) o_p(1)$ for $\alpha = 0$ by the same arguments in the proof of Theorem 1.

We use Lemma 9 to bound $\sum_{i=1}^N z_{ji}$ for $j = 1, 2, 3$ in (B.42). The $O_p\left(\frac{\sqrt{N} L_{2NT}^{1/2}}{\delta_{NT}^2}\right)$, $O_p\left(\frac{\log(\log T) N}{T}\right)$, $O_p\left(\frac{N}{T}\right)$, $O_p\left(\frac{\sqrt{N L_{2NT}}}{\sqrt{T}}\right)$, $O_p\left(\frac{1}{N^{1/2-\alpha}}\right)$, and $O_p\left(\frac{N^{\alpha/2}}{\delta_{NT}}\right)$ terms in Lemma 9 are all $o_p(1)$ for $\alpha = 0$ under the condition that $N \log\log(T)/T \to 0$ (which implies $N < T$) and $\log\log T/N \to 0$. By summarizing the results from (B.6) to (B.39), only the following terms are $(k - k_0) O_p(1)$: $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{1it}^2$ in (B.8), $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{1it} d_{2it}$ in (B.42),[14] the leading term $\sum_{i=1}^N \sum_{t=k_0+1}^k \left[N^{-1} e_t' \Lambda_2 (\Lambda_2' \Lambda_2 / N)^{-1} \lambda_{i2}\right]^2$ in (B.22), $\sum_{t=k_0+1}^k (e_t' \Lambda_1 / N)(\Lambda_2' \Lambda_1 / N)^{-1} \sum_{i=1}^N e_{it} \lambda_{i2}$ in (B.25), $\sum_{t=k_0+1}^k (e_t' \Lambda_2 / N)(\Lambda_2' \Lambda_2 / N)^{-1} \sum_{i=1}^N \lambda_{i2} e_{it}$ in (B.37), and the third term of (B.26).

First, we show that $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{1it} d_{2it}$ can be further bounded by $(k - k_0) o_p(1)$. By the

[14] All other cross-products are $o_p(k - k_0)$ by C-S inequality due to the fact that $\sum_{i=1}^N \sum_{t=k_0+1}^k (d_{3it}^2 + d_{4it}^2) = o_p(k - k_0)$.

definitions of d_{1it} and d_{2it} in (B.6), $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{1it}d_{2it}$ can be represented as

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}[(\tilde{f}_t' H_{k,2}^{(1)^{-1}} - f_t')\lambda_{i2}]\left[(H_{k,1}^{(1)^{-1}}\lambda_{i1} - H_{k,2}^{(1)^{-1}}\lambda_{i2})'\left(\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}/k\right)\tilde{f}_t\right] \\
=&\sum_{t=k_0+1}^{k}\underbrace{(\tilde{f}_t' H_{k,2}^{(1)^{-1}} - f_t')}_{O_p(N^{-1/2})}\underbrace{\left(\sum_{i=1}^{N}\lambda_{i2}\lambda_{i1}'\right)}_{O_p(N)}\underbrace{(H_{k,1}^{(1)^{-1}} - H_{k,2}^{(1)^{-1}})'}_{O_p(N^{-1})}\left(\frac{\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}}{k}\right)\tilde{f}_t \\
&+\sum_{t=k_0+1}^{k}\underbrace{(\tilde{f}_t' H_{k,2}^{(1)^{-1}} - f_t')}_{O_p(N^{-1/2})}\underbrace{\left[\sum_{i=1}^{N}\lambda_{i2}(\lambda_{i1}-\lambda_{i2})'\right]}_{O_p(1)}H_{k,2}^{(1)^{-1'}}\left(\frac{\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}}{k}\right)\tilde{f}_t \\
=&(k-k_0)O_p\left(\frac{1}{\sqrt{N}}\right), \qquad \text{(C.1)}
\end{aligned}$$

where the O_p terms are uniform in k, the $O_p(N^{-1/2})$ term in the second line follows from (B.7), the $O_p(N^{-1})$ term follows from Lemma 2(c) for $\alpha = 0$, and the $O_p(1)$ term follows from Lemma 1(a) for $\alpha = 0$.

Next, note that the $(k-k_0)O_p(1)$ term in $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{1it}^2$ and term $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}[N^{-1}e_t'\Lambda_2 (\Lambda_2'\Lambda_2/N)^{-1}\lambda_{i2}]^2$ in (B.22) are of the opposite signs. We show that the difference between them is $(k-k_0)o_p(1)$. Note that the $(k-k_0)O_p(1)$ term in $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} d_{1it}^2$ is $\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[N^{-1}e_t'\Lambda_1\left(\Lambda_2'\Lambda_1/N\right)^{-1}\lambda_{i2}\right]^2$ by (B.9). Thus, the difference of these two terms is given by

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left(\left[\frac{e_t'\Lambda_1}{N}\left(\frac{\Lambda_2'\Lambda_1}{N}\right)^{-1}\lambda_{i2}\right]^2 - \left[\frac{e_t'\Lambda_2}{N}\left(\frac{\Lambda_2'\Lambda_2}{N}\right)^{-1}\lambda_{i2}\right]^2\right) \\
=&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\underbrace{\left[\left(\frac{e_t'\Lambda_1}{N}\left(\frac{\Lambda_2'\Lambda_1}{N}\right)^{-1} - \frac{e_t'\Lambda_2}{N}\left(\frac{\Lambda_2'\Lambda_2}{N}\right)^{-1}\right)\lambda_{i2}\right]^2}_{O_p(N^{-2})\text{ uniform in } k} + \\
&+2\sum_{t=k_0+1}^{k}\left[\frac{e_t'\Lambda_2}{N}\left(\frac{\Lambda_2'\Lambda_2}{N}\right)^{-1}\sum_{i=1}^{N}\lambda_{i2}\lambda_{i2}'\left(\frac{e_t'\Lambda_1}{N}\left(\frac{\Lambda_2'\Lambda_1}{N}\right)^{-1} - \frac{e_t'\Lambda_2}{N}\left(\frac{\Lambda_2'\Lambda_2}{N}\right)^{-1}\right)'\right] \\
=&(k-k_0)O_p\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}$$

where we use the fact that $e_t'\Lambda_2/N = O_p(N^{1/2})$ by Lemma 1(b) and the fact that

$$\frac{e_t'\Lambda_1}{N}\left(\frac{\Lambda_2'\Lambda_1}{N}\right)^{-1} - \frac{e_t'\Lambda_2}{N}\left(\frac{\Lambda_2'\Lambda_2}{N}\right)^{-1} = O_p(N^{-1}) \qquad \text{(C.2)}$$

because $e_t'(\Lambda_1-\Lambda_2)/N = O_p(N^{-1})$, $\Lambda_2'(\Lambda_1-\Lambda_2)/N = O_p(N^{-1})$ by Lemma 1 for $\alpha = 0$ and $\Lambda_2'\Lambda_2/N$ is asymptotically nonsingular by Assumption 2.

Similarly, the difference between terms $\sum_{t=k_0+1}^{k}(e_t'\Lambda_1/N)(\Lambda_2'\Lambda_1/N)^{-1}\sum_{i=1}^{N}\lambda_{i2}e_{it}$ in (B.25) and $\sum_{t=k_0+1}^{k}(e_t'\Lambda_2/N)(\Lambda_2'\Lambda_2/N)^{-1}\sum_{i=1}^{N}\lambda_{i2}e_{it}$ in (B.37) is also $(k-k_0)o_p(1)$ uniformly in k by (C.2) and

Lemma 1. Hence, except the third term of (B.26), all above terms are $(k-k_0)o_p(1)$ and dominated by $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2$.

Lastly, we show that the third term of (B.26) is also dominated by the leading positive term $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2$ for $k \geq k_0 + S$ and for large S. By (B.44), there exists $S_\epsilon > 0$, N_ϵ and $T_\epsilon > 0$, such that for $\alpha = 0$

$$P\left(\inf_{k\geq k_0+S_\epsilon} \frac{1}{k-k_0}\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2 < c_1\right) \leq \frac{\epsilon}{2},\ \forall N \geq N_\epsilon,\ \forall T \geq T_\epsilon. \tag{C.3}$$

Also, Lemma 1(d) implies $E\big(\sup_{k\geq k_0+S_2} |\frac{1}{k-k_0}\sum_{t=k_0+1}^k \xi_t|^2\big) \leq S_2^{-1}C_1 M$ for $\alpha = 0$ and for some $0 < C_1 < \infty$, where ξ_t is defined in (B.28). Thus, for any small $\varepsilon > 0$

$$\begin{aligned} P\left(\sup_{k\geq k_0+S_2} |\frac{\sum_{t=k_0+1}^k \xi_t}{k-k_0}| \geq \varepsilon c_1\right) &= P\left(\sup_{k\geq k_0+S_2} |\frac{\sum_{t=k_0+1}^k \xi_t}{k-k_0}|^2 \geq \varepsilon^2 c_1^2\right) \\ &\leq \frac{C_1 M}{S_2 \varepsilon^2 c_1^2} = O\left(\frac{1}{S_2}\right) < \frac{\epsilon}{2} \end{aligned} \tag{C.4}$$

as long as S_2 is large enough. Define $S = \max(S_\epsilon, S_2)$, so for any given $\epsilon > 0$, we have

$$\begin{aligned} &P\left(\sup_{k\geq k_0+S} |\frac{\sum_{t=k_0+1}^k \xi_t}{k-k_0}| \geq \inf_{k\geq k_0+S} \frac{\varepsilon}{k-k_0}\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2\right) \\ =&P\left(\sup_{k\geq k_0+S} |\frac{\sum_{t=k_0+1}^k \xi_t}{k-k_0}| \geq \varepsilon c_1,\ \sup_{k\geq k_0+S} |\frac{\sum_{t=k_0+1}^k \xi_t}{k-k_0}| \geq \inf_{k\geq k_0+S} \frac{\varepsilon}{k-k_0}\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2\right) \\ +&P\left(\varepsilon c_1 > \sup_{k\geq k_0+S} |\frac{\sum_{t=k_0+1}^k \xi_t}{k-k_0}| \geq \inf_{k\geq k_0+S} \frac{\varepsilon}{k-k_0}\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2\right) \\ \leq&P\left(\sup_{k\geq k_0+S} |\frac{\sum_{t=k_0+1}^k \xi_t}{k-k_0}| \geq \varepsilon c_1\right) + P\left(\inf_{k\geq k_0+S} \frac{1}{k-k_0}\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2 < c_1\right) \leq \epsilon \end{aligned}$$

where the last line uses (C.3) and (C.4) and $S = \max(S_\epsilon, S_2)$. Thus, $\sum_{i=1}^N \sum_{t=k_0+1}^k f_t'(\lambda_{i1} - \lambda_{i2})e_{it}$ is dominated by $\varepsilon \sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2$ for arbitrary small ε. Since $\tilde{F}_1^{(1)\prime}\tilde{F}_1^{(1)}/k$, $H_{k,2}^{(1)}$ and $H_{k,2}^{(1)^{-1}}$ are all $O_p(1)$ uniformly in k, it follows that $\sum_{i=1}^N \sum_{t=k_0+1}^k f_t' H_{k,2}^{(1)}(\tilde{F}_1^{(1)\prime}\tilde{F}_1^{(1)}/k)H_{k,2}^{(1)^{-1}}(\lambda_{i1} - \lambda_{i2})e_{it}$ in (B.26) is also dominated by $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2$. Since $\sum_{i=1}^N \sum_{t=k_0+1}^k d_{2it}^2$ is positive and dominates all other terms, $SSR(k, \tilde{F}) - SSR(k_0, \hat{F})$ is positive for $k - k_0$ large enough. This completes the proof for Theorem 2.

Q.E.D.

Proof of Theorem 3:

Under (3.2) with $\alpha = 0$, we have $\lambda_{i2} - \lambda_{i1} = \delta_i/\sqrt{N}$. Since we have shown that $\tilde{k} - k_0$ is $O_p(1)$ by Theorem 2, we examine the behavior of $SSR(k, \tilde{F}) - SSR(k_0, \hat{F})$ for $|k - k_0| = O_p(1)$. First, the non-negative term $\sum_{i=1}^{N} A_i(k,k_0)'\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)} A_i(k,k_0)$ in Lemma 8(a) becomes $o_p(1)$ because the $(k-k_0)^2 O_p(N^\alpha/T)$ term reduces to $O_p(T^{-1})$ as $k-k_0 = O_p(1)$ and the $O_p[N\log\log T)/T]$ term is $o_p(1)$ by assumption.

Next, by the proof of Theorem 2, we only need to consider two $O_p(k-k_0)$ terms, the leading term in (B.14) and the third term in (B.26). First, note that the leading term in (B.14) is $O_p(k-k_0)$ when $\alpha = 0$. Since $|k-k_0| = O_p(1)$, it follows that the leading term in (B.14) is also $O_p(1)$ and $\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}/k = I_r + O_p(T^{-1})$. Thus, the leading term in (B.14) reduces to

$$
\begin{aligned}
&= \sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[\tilde{f}_t'\frac{\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}}{k}H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})\right]^2 = \sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[\tilde{f}_t'[I_r+O_p(T^{-1})]H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})\right]^2 \\
&= \sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\left[f_t'(\lambda_{i1}-\lambda_{i2}) + (\tilde{f}_t' - f_t'H_{k,2}^{(1)})H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})\right]^2 + o_p(1) \\
&= \sum_{i=1}^{N}\sum_{t=k_0+1}^{k}[f_t'(\lambda_{i1}-\lambda_{i2})]^2 + 2\mathrm{tr}\left[\underbrace{\sum_{t=k_0+1}^{k} f_t(\tilde{f}_t' - f_t'H_{k,2}^{(1)})}_{o_p(1)} H_{k,2}^{(1)^{-1}}\sum_{i=1}^{N}(\lambda_{i1}-\lambda_{i2})(\lambda_{i1}-\lambda_{i2})'\right] \\
&+ \mathrm{tr}\left[\underbrace{\sum_{t=k_0+1}^{k}\left(H_{k,2}^{(1)^{-1'}}(\tilde{f}_t - H_{k,2}^{(1)'}f_t)(\tilde{f}_t' - f_t'H_{k,2}^{(1)})H_{k,2}^{(1)^{-1}}\right)}_{O_p(\delta_{NT}^{-2})}\sum_{i=1}^{N}(\lambda_{i1}-\lambda_{i2})(\lambda_{i1}-\lambda_{i2})'\right] + o_p(1) \\
&\to_p \mathrm{tr}\left(\sum_{t=k_0+1}^{k} f_t f_t' \cdot \Sigma_\delta\right),
\end{aligned}
$$

where the $o_p(1)$ term in the second line follows from the fact that $2\sum_{t=k_0+1}^{k}\tilde{f}_t' O_p(T^{-1})H_{k,2}^{(1)^{-1}}$ $(N^{-1}\sum_{i=1}^{N}\delta_i\delta_i')H_{k,2}^{(1)^{-1'}}\tilde{f}_t = O_p(T^{-1})$ and $N^{-1}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}[\tilde{f}_t' O_p(T^{-1})H_{k,2}^{(1)^{-1}}\delta_i]^2 = O_p(T^{-2})$ by (3.2) with $\alpha = 0$, and the $o_p(1)$ and $O_p(\delta_{NT}^{-2})$ terms in the second equality follow from Lemmas 4(b) and 2(b), respectively.

Second, the third term in (B.26) is $O_p(k-k_0)$ when $\alpha = 0$. Hence, it is also $O_p(1)$ given that $\tilde{k} - k_0 = O_p(1)$. Since $\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}/k = I_r + O_p(T^{-1})$ and $N/T \to 0$, we have

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k}\frac{f_t'H_{k,2}^{(1)}\tilde{F}_1^{(1)'}\tilde{F}_1^{(1)}H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})e_{it}}{k} \\
=&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} f_t'H_{k,2}^{(1)}[I_r + O_p(T^{-1})]H_{k,2}^{(1)^{-1}}(\lambda_{i1}-\lambda_{i2})e_{it} \\
=&\sum_{i=1}^{N}\sum_{t=k_0+1}^{k} f_t'(\lambda_{i1}-\lambda_{i2})e_{it} + o_p(1)
\end{aligned}
$$

$$= - \sum_{t=k_0+1}^{k} f_t' \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \delta_i e_{it} \right) \to_d \sum_{t=k_0+1}^{k} f_t' Z_t,$$

where $Z_t \sim N(0, \Phi)$ and $\Phi = \lim_{N\to\infty} N^{-1} \sum_{j=1}^{N} E(\delta_j \delta_j') \sigma_j^2$. Therefore, for $k > k_0$

$$SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) \to_d \text{tr} \left(\sum_{t=k_0+1}^{k} f_t f_t' \cdot \Sigma_\delta \right) + 2 \sum_{t=k_0+1}^{k} f_t' Z_t. \tag{C.5}$$

Similarly, for $k < k_0$,

$$SSR(k, \tilde{F}) - SSR(k_0, \hat{F}) \to_d \text{tr} \left(\sum_{t=k+1}^{k_0} f_t f_t' \cdot \Sigma_\delta \right) + 2 \sum_{t=k+1}^{k_0} f_t' Z_t. \tag{C.6}$$

The result in Theorem 3 follows from (C.5), (C.6), and the strict stationarity of f_t.
Q.E.D.

References

[1] Bai, J. (1997a), Estimation of a change point in multiple regression models, *Review of Economics and Statistics*, 79(4): 551–563.

[2] Bai, J. (1997b), Estimating multiple breaks one at a time, *Econometric Theory*, 13, 315–352.

[3] Bai, J. (2003), Inferential theory for factor models of large dimensions, *Econometrica*, 71, 135-172.

[4] Bai, J. (2009), Panel data models with interactive fixed effects, *Econometrica*, 77, 1229-1279.

[5] Bai, J. (2010), Common breaks in means and variances for panel data, *Journal of Econometrics*, 164, 92-115.

[6] Bai, J. and X. Han (2016), Structural changes in high dimensional factor models, *Frontiers of Economics in China*, 11(1), 9-39.

[7] Bai, J. and K. P. Li (2012), Statistical analysis of factor models of high dimension, *Annals of Statistics*, 40(1), 436-465.

[8] Bai, J. and S. Ng (2002), Determining the number of factors in approximate factor models, *Econometrica*, 70(1), 191-221.

[9] Bai, J. and S. Ng (2004), A Panic Attack on Unit Roots and Cointegration, *Econometrica*, 72(4), 1127–1177.

[10] Bai, J. and S. Ng (2010), Instrumental variable estimation in a data rich environment, *Econometric Theory*, 26, 1577-1606.

[11] Banerjee, A. and M. Marcellino (2008), Forecasting Macroeconomic variables using diffusion indexes in short samples with structural change, Emerald Group Publishing Limited, Vol.3, pp. 149-194.

[12] Baltagi, B., Q. Feng, and C. Kao (2016), Estimation of heterogeneous panels with structural breaks, *Journal of Econometrics*, 191(1), 176-195.

[13] Baltagi, B., C. Kao, and L. Liu (2017), Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term, *Econometric Reviews*, 36:1-3, 85-102.

[14] Baltagi, B., C. Kao, and F. Wang (2017), Identification and estimation of a large factor model with structural instability, *Journal of Econometrics*, 197(1), 87-100.

[15] Baltagi, B., C. Kao, and F. Wang (2015), Change point estimation in large heterogeneous panels, manuscript, Syracuse University.

[16] Bates, B., M. Plagborg-Moller, J. H. Stock and M. W. Watson (2013), Consistent factor estimation in dynamic factor models with structural instability, *Journal of Econometrics*, 177, 289-304.

[17] Bernanke, B. S., J. Boivin and P. Eliasz (2005), Measuring the effects of monetary policy: A factor-augmented verctor autoregressive (FAVAR) approach, *Quarterly Journal of Economics*, 120, 387-422.

[18] Boivin, J. and M. Giannoni (2006), DSGE models in a data-rich environment. NBER Working paper, No. 12772.

[19] Bolthausen, E. (1982), One the Central Limit Theorem for Stationary Mixing Random Fields, *Annals of Probability*, 10(4), 1047-1050.

[20] Breitung, J. and S. Eickmeier (2011), Testing for structural breaks in dynamic factor models, *Journal of Econometrics*, 163, 71-74.

[21] Chen, L. (2015) Estimation the common break date in large factor models, *Economics Letters*, 131, 70-74.

[22] Chen, L., Dolado, J. and Gonzalo, J. (2014), Detecting big structural breaks in large factor models, *Journal of Econometrics*, 180(1), 30-48.

[23] Cheng, X., Z. Liao and F. Schorfheide (2016), Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities, *Review of Economic Studies*, 83(4), 1511-1543.

[24] Corradi, V. and N. R. Swanson (2014), Testing for structural stability of factor augmented forecasting models, *Journal of Econometrics*, 182(1), 100-118.

[25] El Machkouri, M., D. Volny, and W. B.Wu (2013), A Central Limit Theorem for Stationary Random Fields, *Stochastic Processes and their Applications*, 123, no.1, 1-14.

[26] Fan, J., Y. Liao and M. Mincheva (2011), High-Dimensional Covariance Matrix Estimation in Approximate Factor Models, *Annals of Statistics*, 39, 3320-3356.

[27] Fan, J., Y. Liao and M. Mincheva (2013), Large Covariance Estimation by Thresholding Principal Orthogonal Complements, *Journal of the Royal Statistical Society: Series B*, 75, 603-680.

[28] Fazekas, I. (2014) On a General Approach to the Strong Laws of Large Numbers, *Journal of Mathematical Sciences*, 200(4), 411–423.

[29] Gonçalves, S. and B. Perron (2014), Bootstrapping factor-augmented regression models, *Journal of Econometrics*, 82, 156–173.

[30] Han, X. and A. Inoue (2015), Tests for Parameter Instability in Dynamic Factor Models, *Econometric Theory*, Volume 31, Issue 05, 1117-1152.

[31] Hayashi, F. (2000), Econometrics, New Jersey: Princeton University Press.

[32] Hsiao, C., H. S. Ching and S. K. Wan (2012), A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China, *Journal of Applied Econometrics*, 27, 705-740.

[33] Jenish, N. and I. R. Prucha (2009), Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields, *Journal of Econometrics*, 150, 86-98.

[34] Jiang, J. (1999) Some Laws of the Iterated Logarithm for Two Parameter Martingales, *Journal of Theoretical Probability*, 12(1), 49-74.

[35] Karlin, S. and H. M. Taylor (1975), A First Course in Stochastic Processes (2d ed.), New York: Academic.

[36] Kim, D. (2011), Estimating a common deterministic time trend break in large panels with cross sectional dependence, *Journal of Econometrics*, 164, 310-330.

[37] Kim, D. (2014), Common breaks in time trends for large panel data with a factor structure, *Econometrics Journal*, 17, 301-337.

[38] Li, D., J. Qian, and L. Su (2016), Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks, *Journal of the American Statistical Association*, 111, 1804-1819.

[39] Li, D. and Z. Wu (1989), The Law of the Iterated Logarithm for B-Valued Random Variables with Multidimensional Indices, *Annals of Probability*, 17(2), 760-774.

[40] Massacci, D. (2017), Least Squares Estimation of Large Dimensional Threshold Factor Models, *Journal of Econometrics*, 197(1), 101-129.

[41] Ma, S. and L. Su (2018), Estimation of Large Dimensional Factor Models with an Unknown Number of Breaks, *Journal of Econometrics*, 207(1), 1-29.

[42] Móricz, F., U. Stadtmüller, and M. Thalmaier (2008), Strong Laws for Blockwise M-Dependent Random Fields, *Journal of Theoretical Probability*, 21: 660–671.

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

[44] Phillips, P. C. B. and V. Solo (1992), Asymptotics for Linear Processes, *Annals of Statistics*, 20(2), 971-1001.

[45] Oodaira H. and K.I. Yoshihara (1971), the Law of the Iterated Logarithm for Stationary Processes Satisfying Mixing Conditions, *Kodai Mathematical Seminar Reports*, 23(3), 311-334.

[46] Petrov, V.V. (1984), Law of the Iterated Logarithm for Sequences of Dependent Random Variables, *Journal of Soviet Mathematics*, 24(5), 611-617.

[47] Schmuland, B. and W. Sun (2004), A Central Limit Theorem and Law of the Iterated Logarithm for a Random Field with Exponential Decay of Correlations, *Canadian Journal of Mathematics*, 56(1), 209-224.

[48] Stock, J. H. and M. W. Watson (1996), Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics*, 14, 11–30.

[49] Stock, J. H. and M. W. Watson (2002), Macroeconomic Forecasting using Diffusion Indexes, *Journal of Business and Economic Statistics*, 20, 147-162.

[50] Stock, J. H. and M. W. Watson (2008), Forecasting in dynamic factor models subject to structural instability. In: Jennifer Castle and Neil Shephard (eds.), The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry, Oxford, UK: Oxford University Press.

[51] Su, L. and X. Wang (2015), On Time-varying Factor Models: Estimation and Inference, manuscript, Singapore Management University.

[52] Wichura, M. (1973), Some Strassen-Type Laws of the Iterated Logarithm for Multiparameter Stochastic Processes with Independent Increments, *Annals of Probability*, 1(2), 272-296.

[53] Yamamoto, Y. (2016), Forecasting with Non-spurious factors in U.S. Macroeconomic time series, *Journal of Business and Economic Statistics*, 34(1), 81-106.

[54] Tanaka, S. and Y. Yamamoto (2015), Testing for Factor Loading Structural Change under Common Breaks, *Journal of Econometrics*, 189(1), 187-206.

[55] Zhao, O and M. Woodroofe (2008), Law of the Iterated Logarithm for Stationary Process, *Annals of Probability*, 36(1), 127-142.

Figure 1: The distribution of $\tilde{k} - k_0$ when only a single variable has a break point with $N = 100$ and $T = 2000$

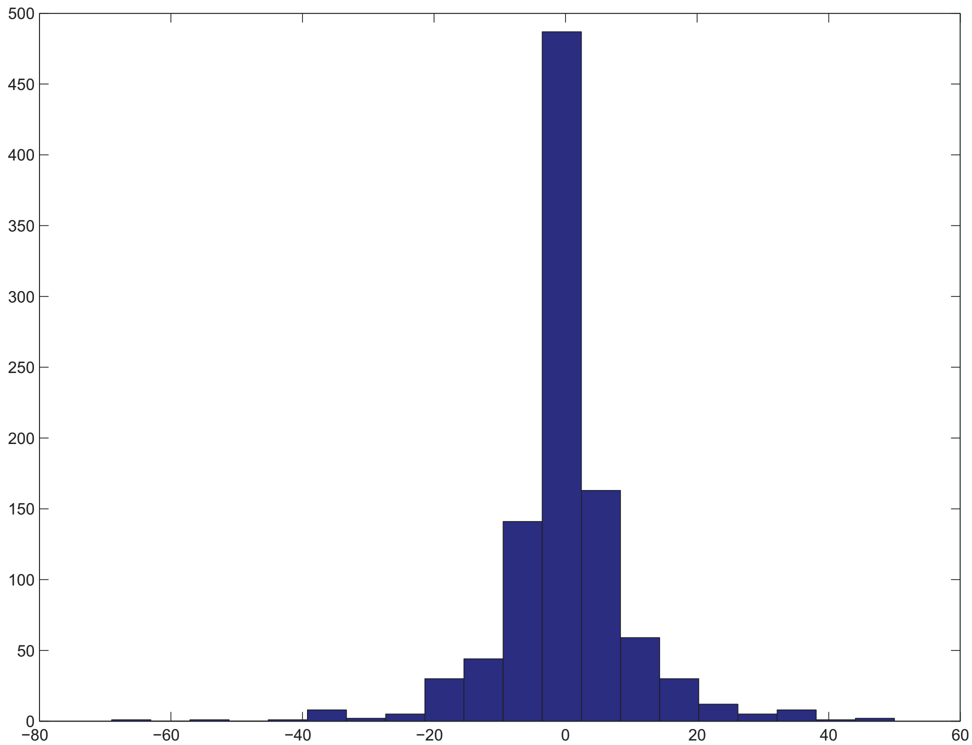

Figure 2: The distribution of $\tilde{k} - k_0$ under DGP1 and DGP2 for $N = 100$ and $T = 500$

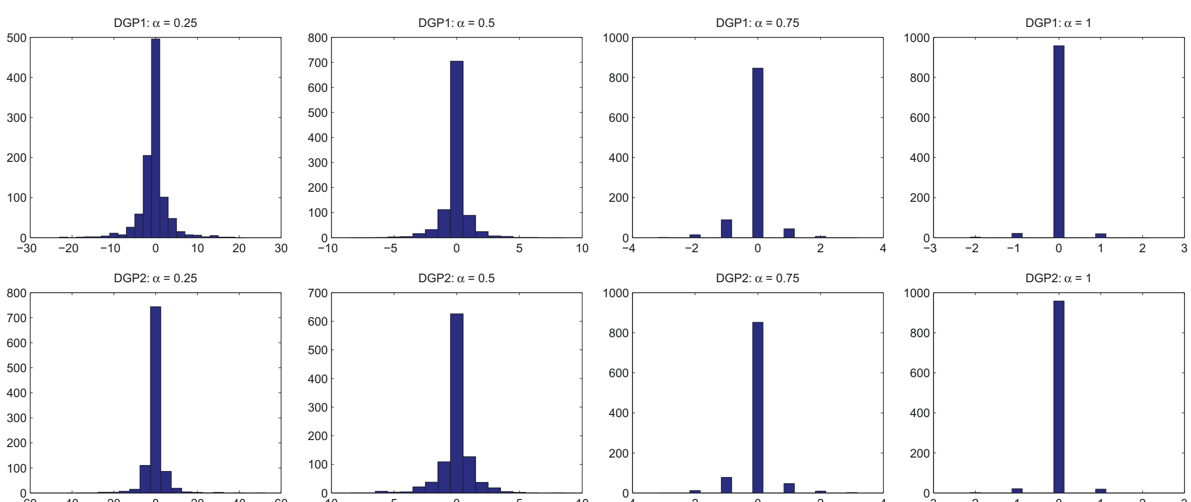

Table 1A: RMSEs under DGP1

$r = 2$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	11.04	27.45	23.62	28.39	2.24	26.66	21.45	22.44	0.87	24.35	16.55	13.05	0.46	7.39	4.22	7.02
50,200	6.21	52.94	46.08	6.21	1.79	52.81	40.17	1.79	0.80	42.50	27.43	0.89	0.47	4.13	3.87	0.50
100,200	6.84	55.11	47.72	6.84	1.30	54.06	42.17	1.30	0.49	42.44	28.39	0.49	0.29	0.47	0.34	0.29
100,500	3.51	136.93	119.99	3.51	1.05	128.49	98.73	1.05	0.48	67.17	41.74	0.52	0.27	0.45	0.30	0.27
200,500	3.26	138.39	118.93	3.26	0.88	135.77	108.75	0.88	0.37	23.99	14.83	0.37	0.22	0.25	0.25	0.22

$r = 3$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	13.28	28.76	21.17	8.62	3.09	28.06	16.91	4.57	0.87	25.40	9.43	1.62	0.41	13.60	3.99	0.82
50,200	8.95	55.92	42.03	8.95	1.86	52.49	30.21	3.86	0.77	43.09	13.20	3.31	0.37	13.84	2.50	2.53
100,200	9.37	56.46	41.77	9.47	1.23	54.38	34.59	1.23	0.42	47.02	16.21	0.59	0.21	1.31	0.27	0.39
100,500	3.41	137.60	106.41	3.41	1.12	126.70	66.47	1.12	0.42	85.01	18.23	0.42	0.21	0.35	0.27	0.23
200,500	3.98	140.69	112.30	3.98	0.71	134.68	88.43	0.71	0.25	96.06	23.10	0.25	0.12	0.14	0.15	0.12

$r = 4$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	15.18	28.73	18.15	9.68	3.85	28.23	12.95	4.65	1.02	25.02	6.07	1.37	0.30	16.48	2.10	0.59
50,200	11.93	56.86	36.58	38.81	2.12	53.33	19.62	12.59	0.73	43.46	5.69	3.82	0.34	19.39	1.35	6.42
100,200	13.19	58.19	38.95	15.31	1.57	56.38	27.76	1.59	0.39	46.91	6.84	2.59	0.14	7.85	0.40	2.96
100,500	4.71	137.23	94.16	4.71	0.95	127.92	45.89	0.96	0.34	80.37	6.58	0.42	0.14	0.77	0.34	0.18
200,500	5.13	140.77	101.45	5.13	0.77	134.83	67.45	0.77	0.24	94.31	11.06	0.26	0.05	0.07	0.08	0.05

Note: $\tilde{k}$ is our LS estimator with a known r, $\tilde{k}_{BKW}$ is Baltagi, Kao, and Wang's (2017) estimator with $\hat{r}$ factors, $\tilde{k}_{\hat{r}-1}$ is the LS estimator using $\hat{r}-1$ factors, where $\hat{r}$ is the full-sample estimator for r using Bai and Ng's (2002) IC_{p1}, and $\tilde{k}_{CLS}$ is the LS estimator using $\hat{r}_1$ pre-break factors and $\hat{r}_2$ post-break factors, where $\hat{r}_1$ and $\hat{r}_2$ are determined by the method of Cheng et al. (2016).

Table 1B: Probability of correct estimation under DGP1

$r = 2$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	0.21	0.02	0.02	0.02	0.48	0.02	0.03	0.10	0.74	0.05	0.13	0.35	0.87	0.66	0.81	0.60
50,200	0.29	0.01	0.02	0.29	0.57	0.01	0.04	0.57	0.76	0.07	0.17	0.76	0.88	0.76	0.84	0.88
100,200	0.27	0.01	0.01	0.27	0.64	0.01	0.02	0.64	0.87	0.12	0.21	0.87	0.94	0.92	0.92	0.94
100,500	0.39	0.00	0.00	0.39	0.67	0.01	0.03	0.67	0.86	0.26	0.40	0.86	0.94	0.92	0.93	0.94
200,500	0.33	0.00	0.00	0.33	0.73	0.00	0.01	0.73	0.92	0.75	0.87	0.92	0.97	0.96	0.96	0.97

$r = 3$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	0.17	0.01	0.03	0.10	0.48	0.01	0.08	0.30	0.75	0.04	0.27	0.60	0.90	0.42	0.74	0.80
50,200	0.24	0.00	0.02	0.24	0.55	0.01	0.08	0.55	0.79	0.05	0.30	0.78	0.91	0.55	0.80	0.86
100,200	0.22	0.00	0.01	0.22	0.66	0.01	0.03	0.66	0.88	0.03	0.23	0.86	0.96	0.88	0.94	0.90
100,500	0.35	0.00	0.01	0.35	0.68	0.01	0.06	0.68	0.90	0.04	0.29	0.90	0.96	0.94	0.94	0.96
200,500	0.34	0.00	0.00	0.34	0.77	0.00	0.02	0.77	0.94	0.06	0.27	0.94	0.99	0.98	0.98	0.99

$r = 4$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	0.10	0.01	0.03	0.10	0.43	0.02	0.13	0.32	0.76	0.06	0.38	0.65	0.94	0.30	0.75	0.85
50,200	0.20	0.01	0.03	0.12	0.51	0.02	0.11	0.43	0.80	0.05	0.44	0.72	0.92	0.36	0.77	0.80
100,200	0.19	0.00	0.01	0.18	0.60	0.01	0.08	0.60	0.90	0.03	0.37	0.86	0.98	0.72	0.95	0.93
100,500	0.32	0.00	0.01	0.32	0.69	0.00	0.10	0.69	0.91	0.04	0.42	0.88	0.98	0.89	0.97	0.98
200,500	0.31	0.00	0.00	0.31	0.77	0.01	0.03	0.77	0.95	0.02	0.33	0.94	1.00	1.00	0.99	1.00

Note: $\tilde{k}$ is our LS estimator with a known r, $\tilde{k}_{BKW}$ is Baltagi, Kao, and Wang's (2017) estimator with $\hat{r}$ factors, $\tilde{k}_{\hat{r}-1}$ is the LS estimator using $\hat{r}-1$ factors, where $\hat{r}$ is the full-sample estimator for r using Bai and Ng's (2002) IC_{p1}, and $\tilde{k}_{CLS}$ is the LS estimator using $\hat{r}_1$ pre-break factors and $\hat{r}_2$ post-break factors, where $\hat{r}_1$ and $\hat{r}_2$ are determined by the method of Cheng et al. (2016).

Table 2A: RMSEs under DGP2

$r=2$

	$\alpha=0.25$				$\alpha=0.5$				$\alpha=0.75$				$\alpha=1$			
N,T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	13.66	27.97	23.50	28.63	3.49	26.61	20.75	21.90	0.86	23.46	16.30	13.94	0.46	7.39	4.23	7.02
50,200	13.57	54.33	45.53	13.56	1.96	49.98	38.42	3.06	0.80	40.27	25.82	0.80	0.47	2.30	3.66	0.48
100,200	12.38	54.86	46.35	12.38	1.66	54.73	43.20	1.66	0.48	38.50	26.99	0.48	0.24	0.43	0.31	0.24
100,500	5.69	136.32	116.69	5.69	1.24	127.95	99.69	1.24	0.49	65.47	39.04	0.49	0.29	0.38	0.38	0.29
200,500	12.13	137.53	122.01	12.13	0.94	132.28	103.18	0.94	0.33	23.29	18.32	0.33	0.18	0.19	0.23	0.18

$r=3$

	$\alpha=0.25$				$\alpha=0.5$				$\alpha=0.75$				$\alpha=1$			
N,T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	14.19	28.33	20.98	8.91	3.33	26.80	15.78	4.48	0.86	24.26	9.18	1.80	0.41	13.78	4.00	0.82
50,200	14.03	55.47	38.09	14.21	2.11	51.11	27.54	3.18	0.70	42.19	12.02	2.67	0.33	11.29	3.31	0.45
100,200	10.71	56.15	40.96	10.71	1.76	55.20	33.70	1.78	0.52	44.45	13.49	0.67	0.17	0.81	0.24	0.19
100,500	6.71	137.57	98.23	6.71	1.14	124.90	66.78	1.17	0.42	85.08	19.62	0.42	0.19	0.30	0.21	0.18
200,500	10.39	138.13	109.74	10.39	0.89	133.20	84.40	0.89	0.24	83.62	21.88	0.24	0.11	0.11	0.15	0.11

$r=4$

	$\alpha=0.25$				$\alpha=0.5$				$\alpha=0.75$				$\alpha=1$			
N,T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	16.22	28.64	18.21	9.92	4.89	27.08	11.97	4.54	0.98	24.00	5.15	1.59	0.34	16.51	2.11	0.59
50,200	17.47	55.71	34.42	37.89	3.28	50.57	18.99	16.51	0.68	40.74	5.64	5.89	0.37	19.58	1.45	5.05
100,200	13.01	56.88	36.60	13.13	1.74	54.15	26.23	1.84	0.43	46.23	7.01	2.59	0.16	5.65	0.41	2.38
100,500	6.16	136.25	88.31	6.16	1.24	124.41	46.62	1.28	0.33	77.71	6.32	0.53	0.11	1.40	0.19	0.15
200,500	7.27	141.48	104.82	7.27	0.81	134.30	61.25	0.81	0.19	93.13	10.15	0.30	0.09	0.09	0.12	0.09

Note: $\tilde{k}$ is our LS estimator with a known r, $\tilde{k}_{BKW}$ is Baltagi, Kao, and Wang's (2017) estimator with $\hat{r}$ factors, $\tilde{k}_{\hat{r}-1}$ is the LS estimator using $\hat{r}-1$ factors, where $\hat{r}$ is the full-sample estimator for r using Bai and Ng's (2002) IC_{p1}, and $\tilde{k}_{CLS}$ is the LS estimator using $\hat{r}_1$ pre-break factors and $\hat{r}_2$ post-break factors, where $\hat{r}_1$ and $\hat{r}_2$ are determined by the method of Cheng et al. (2016).

Table 2B: Probability of correct estimation under DGP2

$r = 2$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	0.21	0.01	0.02	0.03	0.48	0.01	0.05	0.11	0.75	0.08	0.16	0.36	0.87	0.66	0.81	0.61
50,200	0.25	0.01	0.02	0.25	0.55	0.01	0.05	0.55	0.76	0.12	0.22	0.76	0.88	0.77	0.84	0.88
100,200	0.29	0.00	0.01	0.29	0.59	0.01	0.02	0.59	0.85	0.20	0.32	0.85	0.94	0.92	0.92	0.94
100,500	0.35	0.00	0.01	0.35	0.61	0.01	0.03	0.61	0.86	0.38	0.54	0.86	0.93	0.92	0.91	0.93
200,500	0.34	0.00	0.00	0.34	0.73	0.00	0.01	0.73	0.91	0.73	0.85	0.91	0.97	0.97	0.96	0.97

$r = 3$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	0.19	0.02	0.03	0.13	0.46	0.02	0.10	0.29	0.73	0.05	0.25	0.54	0.90	0.41	0.73	0.80
50,200	0.25	0.01	0.03	0.25	0.54	0.03	0.10	0.54	0.80	0.06	0.29	0.79	0.92	0.57	0.82	0.88
100,200	0.29	0.01	0.01	0.29	0.58	0.01	0.06	0.57	0.87	0.06	0.28	0.84	0.98	0.89	0.96	0.97
100,500	0.36	0.00	0.02	0.36	0.68	0.01	0.05	0.68	0.89	0.08	0.32	0.89	0.97	0.94	0.96	0.97
200,500	0.32	0.00	0.01	0.32	0.76	0.00	0.02	0.76	0.95	0.21	0.43	0.95	0.99	0.99	0.98	0.99

$r = 4$

	$\alpha = 0.25$				$\alpha = 0.5$				$\alpha = 0.75$				$\alpha = 1$			
N, T	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$	$\tilde{k}$	$\tilde{k}_{BKW}$	$\tilde{k}_{\hat{r}-1}$	$\tilde{k}_{CLS}$
50,100	0.13	0.01	0.04	0.12	0.38	0.03	0.15	0.31	0.73	0.05	0.39	0.58	0.93	0.30	0.74	0.85
50,200	0.20	0.01	0.04	0.14	0.53	0.03	0.16	0.40	0.78	0.06	0.42	0.65	0.92	0.34	0.74	0.79
100,200	0.24	0.01	0.02	0.24	0.57	0.01	0.09	0.56	0.89	0.04	0.37	0.85	0.98	0.72	0.95	0.92
100,500	0.34	0.00	0.02	0.34	0.66	0.01	0.11	0.66	0.92	0.04	0.44	0.87	0.99	0.88	0.97	0.98
200,500	0.31	0.00	0.01	0.31	0.78	0.00	0.05	0.77	0.97	0.03	0.35	0.96	0.99	0.99	0.99	0.99

Note: $\tilde{k}$ is our LS estimator with a known r, $\tilde{k}_{BKW}$ is Baltagi, Kao, and Wang's (2017) estimator with $\hat{r}$ factors, $\tilde{k}_{\hat{r}-1}$ is the LS estimator using $\hat{r}-1$ factors, where $\hat{r}$ is the full-sample estimator for r using Bai and Ng's (2002) IC_{p1}, and $\tilde{k}_{CLS}$ is the LS estimator using $\hat{r}_1$ pre-break factors and $\hat{r}_2$ post-break factors, where $\hat{r}_1$ and $\hat{r}_2$ are determined by the method of Cheng et al. (2016).

Table 3: Coverage probabilities of the bootstrap confidence intervals

$r = 2$

	$\alpha = 0$			$\alpha = 0.25$			$\alpha = 0.5$		
N, T	90%	95%	99%	90%	95%	99%	90%	95%	99%
50,100	0.856	0.930	0.988	0.898	0.956	0.992	0.924	0.956	0.990
50,200	0.890	0.946	0.994	0.898	0.950	0.996	0.950	0.964	0.996
100,200	0.886	0.956	0.988	0.910	0.966	0.992	0.956	0.980	0.994
100,500	0.882	0.942	0.994	0.928	0.968	0.992	0.940	0.970	0.998
200,500	0.918	0.970	0.992	0.904	0.942	0.988	0.984	0.990	0.996

$r = 3$

	$\alpha = 0$			$\alpha = 0.25$			$\alpha = 0.5$		
N, T	90%	95%	99%	90%	95%	99%	90%	95%	99%
50,100	0.742	0.856	0.968	0.814	0.888	0.978	0.898	0.950	0.992
50,200	0.822	0.924	0.984	0.868	0.934	0.986	0.926	0.958	0.998
100,200	0.840	0.926	0.992	0.850	0.950	0.992	0.950	0.968	0.994
100,500	0.862	0.932	0.986	0.914	0.958	0.996	0.936	0.978	0.994
200,500	0.868	0.936	0.992	0.928	0.962	0.992	0.958	0.984	0.992