May 2016

Essays on structural changes in high dimensional econometric models

Fa Wang

Syracuse University

Abstract

This dissertation consists of three essays on estimating and testing structural changes in high dimensional econometrics models. These essays are based on three working papers joint with Prof. Badi Baltagi and Prof. Chihwa Kao. The first essay considers estimating the date of a single common change in the regression coefficients of a heterogeneous large N and large T panel data model with or without strong cross-sectional dependence. The second essay considers estimating a high dimensional factor model with an unknown number of latent factors and a single common change in the number of factors and/or factor loadings. The third essay considers estimating a high dimensional factor model with an unknown number of latent factors and multiple common changes in the number of factors and/or factor loadings, and also testing procedures to detect the presence and number of structural changes.

The first essay studies the asymptotic properties of the least squares estimator of the common change point in large heterogeneous panel data models under various sets of conditions on the change magnitude and N-T ratio, allowing N and T to go to infinity jointly. Consistency and limiting distribution are established under general conditions. A general Hajek-Renyi inequality is introduced to calculate the order of the expectation of sup-type terms. Both weak and strong cross-sectional dependence are considered. In the former case the least squares estimator is consistent as the number of subjects tends to infinity while in the latter case a two step estimator is proposed and consistency can be recovered once estimated factors are used to control the cross-sectional dependence. The limiting distribution is derived allowing the error process to be serially dependent and

heteroskedastic of unknown form, and inference can be made based on the simulated distribution.

The second essay tackles the identification and estimation of a high dimensional factor model with unknown number of latent factors and a single common break in the number of factors and/or factor loadings. Since the factors are unobservable, the change point estimator is based on the second moments of the estimated pseudo factors. This essay shows that the estimation error of the proposed estimator is bounded in probability as N and T go to infinity jointly. This essay also shows that the proposed estimator has a high degree of robustness to misspecification of the number of pseudo factors. With the estimated change point plugged in, consistency of the estimated number of pre and post-break factors and convergence rate of the estimated pre and post-break factor space are then established under fairly general assumptions. Finite sample performance of the proposed estimators is investigated using Monte Carlo experiments.

The third essay considers high dimensional factor models with multiple common structural changes. Based on the second moments of the estimated pseudo factors, both joint and sequential estimation of the change points are considered. The estimation error of both estimators is bounded in probability as the cross-sectional dimension N and the time dimension T go to infinity jointly. The measurement error contained in the estimated pseudo factors has no effect on the asymptotic properties of the estimated change points as N and T go to infinity jointly, and no N-T ratio condition is needed. The estimated change points are plugged in to estimate the number of factors and the factor space in each regime. Although the estimated change points are inconsistent, using them asymptotically has no effect on subsequent estimation. This essay also proposes (i) tests for the null of no change versus the alternative of l changes and (ii) tests for the null of l changes versus the alternative of l + 1 changes. These tests allow us to make inference on the presence and number of structural changes. Simulation results show good performance of the proposed estimation and testing procedures.

Essays on Structural Changes in High Dimensional Econometric Models

Fa Wang

BA, Xi'an Jiaotong University, 2010

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Graduate School of Syracuse University.

May 2016

Acknowledgements

My deepest gratitude goes to all the people who helped me complete this dissertation and contributed my intellectual progress at Syracuse University.

In particular, I would like to thank my advisors, Professors Chihwa Kao and Badi Baltagi for their encouragement, support and insightful guidance in each step of my progress. I would also like to thank Professors Emil Iantchev, Jerry Kelly, William Horrace, Jan Ondrich, Derek Laing, Devashish Mitra, Lourenzo Paz for helping build my knowledge and techniques for economic research, and thank Professors Dan Coman, Jani Onninen, Philip Griffin, Terry McConnell, Declan Quinn and Graham Leuschke for giving me formal mathematical training. I appreciate the contribution of Professors Promod Varshney, Yoonseok Lee, Alfonso Flores-Lagunes and William Horrace to my dissertation defense. I also benefit a lot from the suggestions and encouragement from Professors Jushan Bai, Lorenzo Trapani, Giovanni Urga, and Liang Chen.

My special thanks go to Professor Yusen Kwoh and Ying Li. I am deeply influenced by their values. I also want to thank Professor Chung Chen for his help. I also benefit a lot from my fellow classmates. They are Ran An, Alexander Falevich, Zaozao He, Stanley Jordan, Bin Peng, Shaofang Qi, Judith Ricks, Zelin Tao, Ian Wright, Tingting Xiong, Jinqi Ye, Pengju Zhang and Hantao Zheng. Moreover, I would like to acknowledge assistance from Mary Santy at Center for Policy Research and Sue Lewis, Laura Sauta and Maureen Eastham in the Economics Department of Syracuse University.

Last, but most important, my lifelong appreciation goes to my parents, my parents in law and especially my wife Nianxia Cao. It is her that gives me a happy PhD life.

Table of Contents

List of Tables and Figures

Essay I: Change Point Estimation in Large Heterogeneous Panels

1 Introduction

Recently, the econometrics literature has witnessed a wave of development in large panel data models (large N and large T), mainly due to its capability of handling cross-sectional dependence. See Pesaran (2006) and Bai (2009), who impose a multifactor error structure, thereby controlling for cross-sectional dependence of the errors and potential correlation between the regressors and the unobservable effects. Meanwhile, the spatial econometrics has also been extended to panel data settings, see for example Yu, De Jong and Lee (2008) and Lee and Yu (2010a, 2010b). Large panels also enable us to test cross-sectional dependence, see Ng (2006), Pesaran (2004, 2012), Pesaran, Ullah and Yamagata (2008) and Baltagi, Feng and Kao (2011, 2012), to mention a few. However, for such panels with a long time span, there is a substantial risk that the underlying data generating process has experienced structural breaks at some unknown time due to various factors. Examples include important economic events such as the European debt crisis, or political events such as the end of the cold war, or gradual but fundamental changes in economic structure due to technological progress, or policy change such as the end of China's one-child policy, to mention a few. If we ignore the parameter changes, standard estimators will be inconsistent and statistical inference will be misleading. Instead, if we explicitly take them into account, the result will be useful for analyzing and evaluating the effect of a policy change, for uncovering the underlying factors that lead to structural change, and for determining whether the response of economic variables are immediate or gradual. This paper therefore studies the parameter change problem in large panel data model with unknown change point.

Change point estimation in linear regression model with single change is analyzed in Bai (1997). Bai and Perron (1998) extend Bai (1997) to the case with multiple changes and also propose tests for the presence of structural change and the number of changes. See also Qu and Perron (2007) for a system of equations, and Bai, Lumsdaine and Stock (1998) for multivariate time series. For other studies on structural change in a finite dimensional setup, see the comprehensive survey by Perron (2006). Bai et al. (1998) find that the number of series is positively related to the accuracy of the change point estimator. To formally analyze this phenomenon, Bai (2010) studies the asymptotic properties of the change point estimator in a panel mean shift setup allowing the number of series N to go to infinity jointly with the sample size T. Based on Bai (2010), Baltagi, Kao and Liu (2014) and Bada, Gualtieri, Kneip and Sickles (2015) study the change point estimation in a homogeneous panel setup, the latter propose a novel Haar wavelet related method. Kim (2011) generalizes Bai (2010) to the case with either mean shift or time trend break or both. Kim (2011) also shows that both cross-sectional and serial dependence of the errors deteriorate the asymptotic behavior of the change point estimator and when the errors have a common factor structure, it reduces to the univariate case. To recover the consistency, Kim (2014) estimates the change point jointly with the factors and factor loadings.

This paper considers least squares estimation of a common change point in a large heterogeneous panel data model, allowing the cross-sectional dependence to be either weak or strong. The heterogeneous framework is general enough to include the most popular panel data models as special cases, so that the results derived here could be applied to these cases with minor adjustment. We first focus on some fundamental difficulties in extending Bai (1997, 2010) to the panel regression setup. The key problem is for random variables $X_{iT} = O_p(1)$ (or $o_p(1)$) as $T \to \infty$, $\frac{1}{N}\sum_{i=1}^{N} X_{iT}$ is not necessarily $O_p(1)$ (or $o_p(1)$ correspondingly) as N and T go to infinity jointly. A simple counterexample is that X_{iT} is identically distributed over T, independent over i, mean zero and variance i^2. This problem is partially solved in Bai (2010) and Kim (2011) by utilizing the specificity of the regressors. In the mean shift setup, $x_{it} = 1$ for all i and t and in the time trend setup, $x_{it} = t$ for all i. However, in the general heterogeneous panel regression setup, it becomes especially troublesome and unavoidable. We solve this problem by introducing a new technique, a general Hajek-Renyi inequality proposed recently in Fazekas and Klesov (2001). An example is given to illustrate how to calculate the order of the expectation of sup-type terms, which in fact is intrinsically related to the uniform law of large numbers. In view of its power, we believe this new tool will also be useful in other places in the econometrics literature.

We then establish the consistency of the estimated common change point under various sets of conditions on the change magnitude and N-T ratio, allowing N and T to go to infinity jointly. As in Kim (2011), we consider both weak and strong cross-sectional dependence of the errors. In the former case, the change point is consistent as the number of series tends to infinity while in the latter case, we propose a two step estimator and show that consistency can be recovered once estimated factors are used to control for cross-sectional dependence. It is also worth noting that because of the powerful tool, our assumptions on the data generating process is fairly general. Rather than assuming specific DGP, e.g., linear process, we only require Doob's maximal inequality to be applicable plus some uniformly bounded moments conditions, see Section 4 for details.

The limiting distribution is derived under the same asymptotic framework as Bai (2010), i.e., shrinking break in the N dimension, but allowing the errors to be cross-sectionally weakly dependent and serially dependent and heteroskedastic of unknown form. The limiting distribution in Bai (2010) is derived assuming the errors are cross-sectionally and serially independent, thus our results generalize those obtained in Bai (2010). This step is nontrivial, see the Appendix for details. Our proof is rigorous and self-contained. Also, our results do not require the DGP to be stationary even within each regime. Based on our results, further parametric assumption can be imposed on the DGP to consistently estimate the parameters in the limiting distribution, and then the distribution can be simulated and inference can be made based on this simulated distribution.

It is worth pointing out the difference and contribution of this paper compared to Baltagi, Feng and Kao (2016), which also study the parameter change problem in large heterogeneous panels. While Baltagi et al. (2016) focus on the asymptotic properties of the estimated regression coefficients and only prove consistency of the change point estimator, this paper studies some fundamental issues in the joint limit asymptotics of change point estimation and the proof of the consistency in Baltagi et al. (2016) is based on solving these issues. Furthermore, this paper derives the limiting distribution of the change point estimator, so that inference regarding the change point can be made. Another difference is how each paper controls for cross-sectional dependence. Baltagi et al. (2016) use cross-sectional averages of the dependent variable and the regressors following Pesaran (2006), while this paper uses estimated factors following Bai (2009).

The rest of the paper is organized as follows. Section 2 introduces the model setup and notation. Section 3 considers least squares estimation of the change point and related fundamental issues. Section 4 studies the asymptotic properties of the least squares estimator when cross-sectional dependence is weak. Section 5 considers estimation of the change point when cross-sectional dependence is strong. Section 6 reports simulation results, while Section 7 concludes. The proofs are given in the Appendix.

2 Model and notation

Consider the following panel data model with a common structural break at k_0:

$$y_{it} = \begin{cases} x_{it}'\beta_i + e_{it}, \; for \; i = 1, ..., N \; and \; t = 1, ..., k_0, \\ x_{it}'\beta_i + z_{it}'\delta_i + e_{it}, \; for \; i = 1, ..., N \; and \; t = k_0 + 1, ..., T, \end{cases} \tag{1}$$

where y_{it} is the dependent variable, x_{it} is a p dimensional vector of regressors, β_i is a p dimensional vector of unknown coefficients, z_{it} is a q dimensional vector of regressors whose coefficients experienced a structural change, δ_i is a q dimensional vector of unknown break magnitude, $z_{it} = R'x_{it}$ and $R = (0_{q\times(p-q)}, I_{q\times q})'$ so that $p > q$ and $p = q$ correspond to partial change and pure change, respectively. e_{it} is the error term allowed to have weak cross-sectional and serial dependence as well as heteroskedasticity. Both N and T are large. In case cross-sectional dependence is strong, a common factor structure is imposed and the model becomes:

$$y_{it} = \begin{cases} x_{it}'\beta_i + F_t^{0\prime}\lambda_i + e_{it}, \; for \; i = 1, ..., N \text{ and } t = 1, ..., k_0, \\ x_{it}'\beta_i + z_{it}'\delta_i + F_t^{0\prime}\lambda_i + e_{it}, \; for \; i = 1, ..., N \text{ and } t = k_0 + 1, ..., T, \end{cases} \tag{2}$$

where F_t^0 is an s dimensional vector of unobservable common factors, λ_i is an s dimensional vector of unobservable factor loadings. In matrix form, the model can be written as

$$Y_i = X_i\beta_i + Z_{0i}\delta_i + e_i, \; for \; i = 1, ..., N, \tag{3}$$

in case the cross-sectional dependence is *weak* and

$$Y_i = X_i\beta_i + Z_{0i}\delta_i + F^0\lambda_i + e_i, \; for \; i = 1, ..., N, \tag{4}$$

in case the cross-sectional dependence is *strong*, where $Z_{0i} = (0_{q\times k_0}, z_{i,k_0+1}, ..., z_{i,T})'$. Also, for any possible change point k, define $Z_{1i} = (z_{i,1}, ..., z_{i,k}, 0_{q\times(T-k)})'$, $Z_{2i} = (0_{q\times k}, z_{i,k+1}, ..., z_{i,T})'$ and $Z_{\Delta i} = (Z_{2i} - Z_{0i})sgn(k_0 - k)$, it follows $Z_{0i} = X_{0i}R$, $Z_{1i} = X_{1i}R$, $Z_{2i} = X_{2i}R$ and $Z_{\Delta i} = X_{\Delta i}R$ once X_{0i}, X_{1i}, X_{2i} and $X_{\Delta i}$ are defined similarly. To study the asymptotic behavior of the change point estimator, the whole set of possible change point, $[1, T]$, is divided into several different regions. Define

$$\begin{aligned} K &= \{k : |k - k_0| \leq T\eta\}, \\ K^c &= \{k : |k - k_0| > T\eta, 1 \leq k \leq T\}, \\ K(k_0) &= \{k : k \neq k_0, |k - k_0| < T\eta)\}, \end{aligned}$$

for some $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, where $\tau_0 = k_0/T$ is the change fraction, and for some $C > 0$,

$$K(C) = \{k : |k - k_0| > C\} \cap K.$$

Throughout the paper, $\|A\| = (trAA')^{\frac{1}{2}}$ denotes the Frobenius norm, $\|A\|_{op}$ denotes the operator norm, $\rho_{\min}(A)$ and $\rho_{\max}(A)$ denote the minimum and maximum eigenvalue of A, $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, c represents a typical constant, $(N,T) \to \infty$ denotes N and T going to infinity jointly.

3 Least squares estimation of the change point

For each possible change point k, the sum of squared residuals is:

$$SSR(k) = \sum\nolimits_{i=1}^{N} SSR_i(k) = \sum\nolimits_{i=1}^{N} Y_i' M_{X_i,Z_{2i}} Y_i, \tag{5}$$

where $M_{X_i,Z_{2i}} = I - P_{X_i,Z_{2i}}$ and $P_{X_i,Z_{2i}}$ is the projection matrix of (X_i, Z_{2i}). The change point estimator is obtained by minimizing the sum of squared residuals:

$$\hat{k} = \arg\min SSR(k).$$

From the identity $Y_i' M_{X_i,Z_{2i}} Y_i = Y_i' M_{X_i} Y_i - \hat{\delta}_i'(k)(Z_{2i}' M_{X_i} Z_{2i})\hat{\delta}_i(k)$, where $(\hat{\beta}_i'(k), \hat{\delta}_i'(k))'$ is the least squares estimator of $(\beta_i', \delta_i')'$ by regressing Y_i on X_i and Z_{2i}, we have

$$SSR(k) = \sum\nolimits_{i=1}^{N} Y_i' M_{X_i} Y_i - \sum\nolimits_{i=1}^{N} \hat{\delta}_i'(k)(Z_{2i}' M_{X_i} Z_{2i})\hat{\delta}_i(k). \tag{6}$$

For simplicity, M_{X_i} is replaced by M_i henceforth. Define $V_i(k) = \hat{\delta}_i'(k)(Z_{2i}' M_i Z_{2i})\hat{\delta}_i(k)$, then $SSR(k) = \sum_{i=1}^{N} Y_i' M_i Y_i - \sum_{i=1}^{N} V_i(k)$ and $SSR(k) - SSR(k_0) = \sum_{i=1}^{N}[V_i(k_0) - V_i(k)]$, it follows that

$$\hat{k} = \arg\min SSR(k) - SSR(k_0) = \arg\max \sum\nolimits_{i=1}^{N}[V_i(k) - V_i(k_0)]. \tag{7}$$

We consider the asymptotic behavior of $\hat{k}$ under different sets of assumptions. Define $\hat{\tau} = \hat{k}/T$ as the estimated break fraction. To show $\hat{\tau} - \tau_0 = o_p(1)$ as $(N,T) \to \infty$, we need to show for any $\epsilon > 0$, $P(\hat{k} \in K^c) < \epsilon$ as $(N,T) \to \infty$. And to show $\hat{k} - k_0 = O_p(1)$, we need to show $P(\hat{k} \in K(C)) < \epsilon$ as $(N,T) \to \infty$ additionally, or $P(\hat{k} \in K(k_0)) < \epsilon$ as $(N,T) \to \infty$, if we want to show $\hat{k}$ is consistent for k_0. Let O represent certain possible region of change point, e.g., K^c. By definition of $\hat{k}$, $\sum_{i=1}^{N}[V_i(\hat{k}) - V_i(k_0)] \geq 0$, hence if $\hat{k} \in O$, then $\sup_{k\in O}\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]\geq 0$. This implies $P(\hat{k}\in O)\leq P(\sup_{k\in O}\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]\geq 0)$, hence to show the former is asymptotically negligible, it suffices to show the latter. In the appendix, we show that the set $\{\omega:\sup_{k\in O}\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]\geq 0\}$ is exactly the same as the set $\{\omega:\sup_{k\in O}\frac{1}{|k-k_0|}\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]\geq 0\}$, hence it suffices to show $P(\sup_{k\in O}\frac{1}{|k-k_0|}\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]\geq 0)<\epsilon$ as $(N,T)\to\infty$.

Remark 1 *The above argument embodies the essence of least squares estimation and appears explicitly, or implicitly in previous change point studies. In fact, the proof of the consistency of β in Bai (2009) is also based on this argument. The difference is that here the supremum is taken with respect to k while in Bai (2009) the supremum is taken with respect to $F'F/T=I$. This argument also can be further generalized and polished to handle other problems featured by the presence of an infinite number of nuisance parameters, by replacing the sum of squared residuals with other criterion function and taking the supremum over their corresponding parameter subspaces. Here we formalize this argument so that it can be easily modified to fit other problems.*

Plug in

$$\begin{aligned}\hat{\delta}_i(k) &= (Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iY_i)=(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_i(X_i\beta_i+Z_{0i}\delta_i+e_i)) \\ &= (Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})\delta_i+(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i),\end{aligned} \tag{8}$$

we have

$$\begin{aligned}\hat{\delta}_i'(k)(Z_{2i}'M_iZ_{2i})\hat{\delta}_i(k) &= \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})\delta_i \\ &\quad +2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) \\ &\quad +(e_i'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i).\end{aligned} \tag{9}$$

$$\begin{aligned}&\sum_{i=1}^{N}V_i(k)-V_i(k_0) \\ =\ &\sum_{i=1}^{N}[\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})\delta_i-\delta_i'Z_{0i}'M_iZ_{0i}\delta_i] \\ &+\sum_{i=1}^{N}[2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i)-2\delta_i'Z_{0i}'M_ie_i] \\ &+\sum_{i=1}^{N}[(e_i'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i)-(e_i'M_iZ_{0i})(Z_{0i}'M_iZ_{0i})^{-1}(Z_{0i}'M_ie_i)].\end{aligned} \tag{10}$$

Define $G_i(k)$ as the first term divided by $-|k_0-k|$ for $k\neq k_0$ and $H_i(k)$ as the last two terms within the summation, then

$$\frac{1}{|k-k_0|}\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]=-\sum_{i=1}^{N}G_i(k)+\frac{1}{|k_0-k|}\sum_{i=1}^{N}H_i(k). \tag{11}$$

Thus $\sup_{k\in O}\frac{1}{|k-k_0|}\sum_{i=1}^N[V_i(k)-V_i(k_0)]\geq 0$ implies $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right|\geq \inf_{k\in O}\sum_{i=1}^N G_i(k)$, and it suffices to show $P(\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right|\geq \inf_{k\in O}\sum_{i=1}^N G_i(k))<\epsilon$ as $(N,T)\to\infty$.

It is worth noting the technical difficulty here. We need to show that the left hand side will be dominated by the right hand side asymptotically. $\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)$ can be written as:

$$2\frac{\sum_{i=1}^N \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i)}{|k-k_0|}-2\frac{\sum_{i=1}^N\delta_i'Z_{0i}'M_ie_i}{|k-k_0|}$$
$$+\frac{\sum_{i=1}^N e_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_ie_i}{|k-k_0|}-\frac{\sum_{i=1}^N e_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_ie_i}{|k-k_0|}, \quad (12)$$

and consider the second term as a representative example. To calculate the stochastic order of $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N\delta_i'Z_{0i}'M_ie_i\right|$, if $N=1$, we are back to Bai (1997) and Hajek-Renyi inequality (Hajek and Renyi (1955)) is applicable. However, if N and T go to infinity jointly, Hajek-Renyi inequality is no longer directly applicable. Noting that $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N\delta_i'Z_{0i}'M_ie_i\right|\leq \sum_{i=1}^N\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i'Z_{0i}'M_ie_i\right|$, we may conclude that $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N\delta_i'Z_{0i}'M_ie_i\right|=O_p(NB_{NT})$, if $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i'Z_{0i}'M_ie_i\right|=O_p(B_{NT})$ for each i, where B_{NT} represents a certain speed. However, this is not necessarily true. We provide three representative counterexamples.

Counterexample 1: X_{iT} is iid over i, $X_{iT}=O_p(1)$, but $E(X_{iT})\to\infty$ as $T\to\infty$.

Suppose $P(X_{iT}=0)=1-\frac{1}{T}$ and $P(X_{iT}=T^2)=\frac{1}{T}$, then $E(X_{iT})=T$, $Var(X_{iT})=T^3-T^2$, $X_{iT}\xrightarrow{p}0$ as $T\to\infty$ for each i and for each T, $\frac{1}{N}\sum_{i=1}^N X_{iT}\xrightarrow{p}\frac{1}{N}\sum_{i=1}^N E(X_{iT})=T$ as $N\to\infty$. This implies that when both N and T are large, $\frac{1}{N}\sum_{i=1}^N X_{iT}$ will be close to a large number with high probability. This contradicts that $\frac{1}{N}\sum_{i=1}^N X_{iT}=O_p(1)$.

Counterexample 2: X_{iT} is independent over i, $X_{iT}=O_p(1)$ and $E(X_{iT})$ is bounded as $T\to\infty$, but $E(X_{iT})$ is not uniformly bounded over i.

Suppose X_{iT} follows $\chi^2(i)$ for all T and is independent over i, then $E(\frac{1}{N}\sum_{i=1}^N X_{iT})=\frac{N+1}{2}$ and $Var(\frac{1}{N}\sum_{i=1}^N X_{iT})=\frac{N+1}{N}$, and it follows that $\frac{1}{N}\sum_{i=1}^N X_{iT}=O_p(N)$.

Counterexample 3: X_{iT} is independent over i, $X_{iT}=O_p(1)$ and $E(X_{iT})$ is uniformly bounded over i and T, but $Var(X_{iT})$ is not uniformly bounded over i.

Suppose X_{iT} follows $N(0,i^2)$ for all T and is independent over i, then $E(X_{iT})=0$ for all i and T, $E(\frac{1}{N}\sum_{i=1}^N X_{iT})=0$ and $Var(\frac{1}{N}\sum_{i=1}^N X_{iT})=\frac{(N+1)(2N+1)}{6N}\approx\frac{N}{3}$, and it follows that $\frac{1}{N}\sum_{i=1}^N X_{iT}=O_p(\sqrt{N})$.

In Bai (2010), Kim (2011) and Kim (2014), this problem is partially solved by utilizing the specificity of the regressors. In the mean shift setup, $x_{it}=1$ for all i and t, and in the time trend setup, $x_{it} = t$ for all i. These two cases do not belong to any of the above counterexamples and for such special regressors, the second term of (12) (as well as the other terms) can be further algebraically simplified, so that calculating the stochastic order is feasible. In the current setup with general regressors, a new method is required. Inspired by the above counterexamples, a feasible solution is to show $E(\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i' Z_{0i}' M_i e_i\right|) \leq MB_{NT}$ for some $M < \infty$ and all i and T. Once this is done, it follows by the Markov inequality that for a large constant C,

$$P(\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^{N}\delta_i' Z_{0i}' M_i e_i\right| > C) \leq P(\sum\nolimits_{i=1}^{N}\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i' Z_{0i}' M_i e_i\right| > C) \leq \frac{NMB_{NT}}{C},$$

so that $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^{N}\delta_i' Z_{0i}' M_i e_i\right| = O_p(NB_{NT})$. Thus to implement this method, the key step is to control the expectation of sup-type terms uniformly over both i and T. For this, we introduce a more powerful tool:

Lemma 1 *General Hajek-Renyi inequality (Theorem 1.1 of Fazekas and Klesov (2001)): Let β_1, β_2,..., β_T be a sequence of nondecreasing positive numbers. Let α_1, α_2,..., α_T be a sequence of nonnegative numbers. Let r be a fixed positive number. Let $\{x_t, t = 1, ...\}$ be a sequence of random variables and $S_l = \sum_{t=1}^{l} x_t$. Assume that for each m with $1 \leq m \leq T$, $E(\sup_{1\leq l\leq m}|S_l|^r) \leq \sum_{l=1}^{m}\alpha_l$, then $E(\sup_{1\leq l\leq T}\left|\frac{S_l}{\beta_l}\right|^r) \leq 4\sum_{l=1}^{T}\frac{\alpha_l}{\beta_l^r}$.*

This lemma permits calculating the order of expectation of sup-type terms, rather than just the stochastic order of sup-type terms. Note that no dependence structure of x_t is assumed. This lemma also permits controlling the expectation uniformly over i if we assume the r-th moment is uniformly bounded over i. Consider the following representative example.

Example 1 *Suppose for each i, $\{x_{it}, t = 1, ...\}$ is a sequence of random variables and $S_{il} = \sum_{t=1}^{l} x_{it}$. If Doob's maximal inequality is applicable, then for each i and each m with $1 \leq m \leq T$, $E(\sup_{1\leq l\leq m}|S_{il}|^r) \leq (\frac{r}{r-1})^r E(|S_{im}|^r)$. Take $r = 2$ and assume $E(S_{im}^2) = O(m)$ uniformly over i, i.e., there exists $M > 0$ such that $E(S_{im}^2) \leq mM$ for all i, we can take $\alpha_{il} = 4M$ so that $E(\sup_{1\leq l\leq m}|S_{il}|^2) \leq \sum_{l=1}^{m}\alpha_{il}$ for each i. If we take $\beta_l = \sqrt{l}$, it follows from this lemma that for each i,*

$$E(\sup_{1\leq l\leq T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^2) \leq 4\sum\nolimits_{l=1}^{T}\frac{\alpha_{il}}{l} = 16M\sum\nolimits_{l=1}^{T}\frac{1}{l} \approx 16M\log T,$$

and for each i and some $\eta > 0$,

$$
\begin{aligned}
E(\sup_{T\eta+1\le l\le T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^2) &\le 16M\sum\nolimits_{l=T\eta+1}^{T}\frac{1}{l} = 16M(\sum\nolimits_{l=1}^{T}\frac{1}{l}-\sum\nolimits_{l=1}^{T\eta}\frac{1}{l}) \\
&= 16M[(\sum\nolimits_{l=1}^{T}\frac{1}{l}-\log T)-(\sum\nolimits_{l=1}^{T\eta}\frac{1}{l}-\log T\eta)+(\log T-\log T\eta)] \\
&\to 16M[\gamma-\gamma+(\log T-\log T\eta)] = 16M\log\frac{1}{\eta},
\end{aligned}
$$

where γ is the Euler-Mascheroni constant, thus $E(\sup_{T\eta+1\le l\le T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^2)$ is uniformly bounded over i. If we take $r = 4$ and $\alpha_{il} = (\frac{4}{3})^4(l^2-(l-1)^2)M$, we can also show that for each i,

$$
E(\sup_{1\le l\le T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^4) \le 4\sum\nolimits_{l=1}^{T}\frac{\alpha_{il}}{l^2} = 4(\frac{4}{3})^4M\sum\nolimits_{l=1}^{T}\frac{2l-1}{l^2} \approx 8(\frac{4}{3})^4M(\log T).
$$

This example illustrates how we calculate the order of expectation of sup-type terms in the Appendix and shows the power of Lemma 1.

4 Asymptotics with weak cross-sectional dependence

This section considers the asymptotic properties of the least squares estimator when cross-sectional dependence is weak. We first present some regularity conditions.

Assumption 1 $\tau_0 = k_0/T \in (0,1)$.

The change point is assumed to be bounded away from 1 and T such that the size of each subsample is a positive fraction of the total sample size. This is a conventional assumption in the change point literature.

Assumption 2 *(1) $E(x_{it}x_{it}') = \Sigma_i^X$ and for all i, $0 < \rho_1 < \rho_{\min}(\Sigma_i^X) < \rho_{\max}(\Sigma_i^X) < \rho_2 < \infty$.*

(2) There exists $\rho_0 > 0$ such that for some $\eta > 0$ and all T and i, $\inf_{k>T(\tau_0-\eta)}\rho_{\min}(\frac{X_{1i}'X_{1i}}{k}) > \rho_0$ and $\inf_{k<T(\tau_0+\eta)}\rho_{\min}(\frac{X_{2i}'X_{2i}}{T-k}) > \rho_0$.

(3) (Doob's maximal inequality) Define $\{R_i(1,k) = \sum_{t=1}^{k}(x_{it}x_{it}'-\Sigma_i^X)\}$, $\{R_i(k,k_0) = \sum_{t=k+1}^{k_0}(x_{it}x_{it}'-\Sigma_i^X)\}$, $\{R_i(k_0+1,k) = \sum_{t=k_0+1}^{k}(x_{it}x_{it}'-\Sigma_i^X)\}$, $\{R_i(k,T) = \sum_{t=k+1}^{T}(x_{it}x_{it}'-\Sigma_i^X)\}$ and $R_{ijm}(1,k)$, $R_{ijm}(k,k_0)$, $R_{ijm}(k_0+1,k)$, $R_{ijm}(k,T)$ as the j-th row and m-th column of $R_i(1,k)$, $R_i(k,k_0)$, $R_i(k_0+1,k)$, $R_i(k,T)$ respectively, then for $1\le j\le p$, $1\le m\le p$ and $1 < r < \infty$,

$E(\sup_{1\le l\le k}|R_{ijm}(1,l)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(1,k)|^r)$ *for all* $1 \le i \le N$ *and* $1 \le k \le T$,

$E(\sup_{k\le l\le k_0-1}|R_{ijm}(l,k_0)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(k,k_0)|^r)$ *for all* $1 \le i \le N$ *and* $0 \le k \le k_0-1$,

$E(\sup_{k_0+1\le l\le k}|R_{ijm}(k_0+1,l)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(k_0,k)|^r)$ *for all* $1 \le i \le N$ *and* $k_0+1 \le k \le T$,

$E(\sup_{k\le l\le T-1}|R_{ijm}(l,T)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(k,T)|^r)$ *for all* $1 \le i \le N$ *and* $0 \le k \le T-1$.

(4) There exists $M > 0$ *such that for* $r = 2, 4$, $1 \le j \le p$ *and* $1 \le m \le p$,

$E(|R_{ijm}(1,k)|^r) < k^{\frac{r}{2}}M$ *for all* $1 \le i \le N$ *and* $1 \le k \le T$,

$E(|R_{ijm}(k,k_0)|^r) < (k_0-k)^{\frac{r}{2}}M$ *for all* $1 \le i \le N$ *and* $0 \le k \le k_0-1$,

$E(|R_{ijm}(k_0+1,k)|^r) < (k-k_0)^{\frac{r}{2}}M$ *for all* $1 \le i \le N$ *and* $k_0+1 \le k \le T$,

$E(|R_{ijm}(k,T)|^r) < (T-k)^{\frac{r}{2}}M$ *for all* $1 \le i \le N$ *and* $0 \le k \le T-1$.

(5) Define $\lambda_N = \sum_{i=1}^N \delta_i'\delta_i$ *and* $\xi = \lim_{N\to\infty}\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i'\Sigma_i^{ZZ}\delta_i$, *for each* t, $\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i' z_{it}z_{it}'\delta_i \xrightarrow{p} \xi$ *as* $N \to \infty$.

Part (1) requires Σ_i^X to be positive definite and bounded uniformly over i. When Σ_i^X is the same for all i, this condition is directly satisfied. Part (2) requires $\frac{X_{1i}'X_{1i}}{k}$ and $\frac{X_{2i}'X_{2i}}{T-k}$ to be uniformly positive definite over i and over $k > T(\tau_0-\eta)$ and $k < T(\tau_0+\eta)$ respectively, so that $\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|$ and $\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|$ are uniformly bounded over i. If a strong law of large numbers is applicable, and together with part (1), part (2) is true almost surely. Part (3) assumes that Doob's maximal inequality is applicable to the process $R_{ijm}(1,k)$, $R_{ijm}(k,k_0)$, $R_{ijm}(k_0+1,k)$ and $R_{ijm}(k,T)$ for $1 \le i \le N$ and $1 \le j,m \le p$. Doob's maximal inequality has proved to be applicable to various processes, including i.i.d. sequences, martingale and submartingale sequences. For economic data, this condition can be easily satisfied. Part (4) further requires the r-th moment of $R_{ijm}(1,k)$, $R_{ijm}(k,k_0)$, $R_{ijm}(k_0+1,k)$ and $R_{ijm}(k,T)$ to be $O(k^{\frac{r}{2}})$, $O((k_0-k)^{\frac{r}{2}})$, $O((k-k_0)^{\frac{r}{2}})$ and $O((T-k)^{\frac{r}{2}})$ uniformly over i, respectively. This will be satisfied if the regressors are weakly dependent over t. Parts (3) and (4) together enable the use of Lemma 1 to calculate the order of sup-type terms. Note that here we do not assume a specific data generating process, thus our assumptions are quite general. Part (5) assumes a weak law of large numbers is applicable to $\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i' z_{it}z_{it}'\delta_i$ for each t.

Assumption 3 *(1)* e_{it} *is independent with* x_{js} *for all* i, t, j, s.

(2) (Doob's maximal inequality) Define $\{S_i(1,k) = \sum_{t=1}^k x_{it}e_{it}\}$, $\{S_i(k,k_0) = \sum_{t=k+1}^{k_0} x_{it}e_{it}\}$, $\{S_i(k_0+1,k) = \sum_{t=k_0+1}^{k} x_{it}e_{it}\}$, $\{S_i(k,T) = \sum_{t=k+1}^{T} x_{it}e_{it}\}$ *and* $S_{ij}(1,k)$, $S_{ij}(k,k_0)$, $S_{ij}(k_0+

$1,k)$, $S_{ij}(k,T)$ as the j-th element of $S_i(1,k)$, $S_i(k,k_0)$, $S_i(k_0+1,k)$, $S_i(k,T)$ respectively, then for $1 \leq j \leq p$ and $1 < r < \infty$,

$E(\sup_{1\leq l\leq k} |S_{ij}(1,l)|^r) \leq (\frac{r}{r-1})^r E(|S_{ij}(1,k)|^r)$ *for all* $1 \leq i \leq N$ *and* $1 \leq k \leq T$,

$E(\sup_{k\leq l\leq k_0-1} |S_{ij}(l,k_0)|^r) \leq (\frac{r}{r-1})^r E(|S_{ij}(k,k_0)|^r)$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq k_0-1$,

$E(\sup_{k_0+1\leq l\leq k} |S_{ij}(k_0+1,l)|^r) \leq (\frac{r}{r-1})^r E(|S_{ij}(k_0,k)|^r)$ *for all* $1 \leq i \leq N$ *and* $k_0+1 \leq k \leq T$,

$E(\sup_{k\leq l\leq T-1} |S_{ij}(l,T)|^r) \leq (\frac{r}{r-1})^r E(|S_{ij}(k,T)|^r)$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq T-1$.

(3) There exists $M > 0$ such that for $r = 2,4$ and for $1 \leq j \leq p$,

$E(|S_{ij}(1,k)|^r) < k^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $1 \leq k \leq T$,

$E(|S_{ij}(k,k_0)|^r) < (k_0-k)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq k_0-1$,

$E(|S_{ij}(k_0+1,k)|^r) < (k-k_0)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $k_0+1 \leq k \leq T$,

$E(|S_{ij}(k,T)|^r) < (T-k)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq T-1$.

(4) Define $\eta_{Nt} = \frac{1}{\sqrt{\lambda_N}} \sum_{i=1}^N \delta_i' z_{it} e_{it}$, there exists $M > 0$ such that

$E(\sup_{k\leq l\leq k_0-1} \left|\sum_{t=l+1}^{k_0} \eta_{Nt}\right|^2) \leq 4E(\left|\sum_{t=k+1}^{k_0} \eta_{Nt}\right|^2) \leq (k_0-k)M$ *for all* N *and* $0 \leq k \leq k_0-1$,

$E(\sup_{k_0+1\leq l\leq k} \left|\sum_{t=k_0+1}^{l} \eta_{Nt}\right|^2) \leq 4E(\left|\sum_{t=k_0+1}^{k} \eta_{Nt}\right|^2) \leq (k-k_0)M$ *for all* N *and* $k_0+1 \leq k \leq T$.

(5) Define $\phi_{st} = \lim_{N\to\infty} E(\frac{1}{\lambda_N} \sum_{i=1}^N \sum_{j=1}^N \delta_i' z_{is} z_{jt}' \delta_j e_{is} e_{jt})$ as the limit of the covariance of η_{Ns} and η_{Nt}. For any fixed $C > 0$, $(\eta_{N,k_0-C}, ..., \eta_{N,k_0+C})' \xrightarrow{d} (Z_{-C}, ..., Z_C)'$ as $N \to \infty$, where $(Z_{-C}, ..., Z_C)'$ follows a multivariate normal distribution with mean zero and covariance $\phi_{st}, k_0 - C \leq s,t \leq k_0 + C$.

Part (1) assumes the error terms are independent of the regressors. Parts (2) and (3) are analogous to parts (3) and (4) of Assumption 2. Part (2) requires Doob's maximal inequality to be applicable to the process $S_{ij}(1,k)$, $S_{ij}(k,k_0)$, $S_{ij}(k_0+1,k)$ and $S_{ij}(k,T)$ for $1 \leq i \leq N$ and $1 \leq j \leq p$. Part (3) requires weak serial dependence of $x_{it}e_{it}$ for each i. Part (4) is a combination of parts (2) and (3), but imposed on the weighted cross-sectional average. Part (5) assumes a central limit theorem is applicable to the fixed dimensional random vector $\{\eta_{Nt}, t = k_0 - C, ..., k_0 + C\}$. Thus cross-sectional dependence of e_{it} can be allowed but need to be weak.

Given the above regularity conditions on the DGP, it is easy to see that asymptotic properties of $\hat{k}$ should depend on the change magnitude, λ_N, and the N-T ratio as $(N,T) \to \infty$. We consider three sets of conditions.

Assumption 4 *Assume* $\max_{1\le i\le N} \delta_i'\delta_i = O(\frac{1}{N})$ *and as* $(N,T)\to\infty$,

(a) $\lambda_N \to \lambda < \infty$ *and* $\frac{N}{\sqrt{T}} \to 0$.

(b) $\lambda_N \to \infty$ *and* $\frac{N}{\sqrt{T}} \to 0$.

(c) $\liminf_{N\to\infty} \frac{\lambda_N}{N} > 0$.

Similar sets of conditions are also considered in Bai (2010). $\max_{1\le i\le N} \delta_i'\delta_i = O(\frac{1}{N})$ is imposed to ensure the change magnitude of each series is of similar order so that no series will be dominant.

Theorem 1 *Under Assumptions 1-3 and 4(a) or 4(b) or 4(c),* $\hat{\tau}$ *is consistent as* $(N,T)\to\infty$.

This result is mainly of theoretical importance. Recall that the least squares estimator is searched in the whole set $[1,T]$, given the consistency of $\hat{\tau}$, the search region can be narrowed down to a local region of k_0. Within this local region, the order of sup-type terms can be established more accurately so that we can move one step further to improve the convergence rate.

Theorem 2 *Under Assumptions 1-3 and 4(a),* $\hat{k} - k_0 = O_p(1)$ *as* $(N,T)\to\infty$.

When $\lambda_N \to \lambda$, the change magnitude is of the same order as that of the univariate case, thus not surprisingly the result is also the same, see Bai (1997). Here the extra condition $\frac{N}{\sqrt{T}} \to 0$[1] is imposed to deal with the nuisance parameters $\beta_i, i = 1,...,N$.[2] With $\hat{\beta}_i$ plugged in the least squares criterion function, for each i, the difference $\hat{\beta}_i - \beta_i$ would result in an extra source of noise. It can be shown that each noise is $O(\frac{1}{\sqrt{T}})$, hence when $\frac{N}{\sqrt{T}} \to 0$, T is large enough to control the total noise resulting from the nuisance parameters. If we let $\lambda_N \to \infty$ while still maintaining $\frac{N}{\sqrt{T}} \to 0$, then we will have consistency of $\hat{k}$.

Theorem 3 *Under Assumptions 1-3 and 4(b) or 4(c),* $\hat{k}$ *is consistent as* $(N,T)\to\infty$.

While consistency under Assumption 4(b) still relies on $\frac{N}{\sqrt{T}} \to 0$, consistency under Assumption 4(c) only requires $T\to\infty$. This is because when $\lambda_N = O(N)$, the change

[1] The condition $\frac{N}{\sqrt{T}} \to 0$ is stricter than that of Bai (2010), $\frac{N\log T}{T} \to 0$, but the spirit is the same. Bai (2010) considers the mean shift setup, in the current setup we do not have the algebraic specificity of the regressors.

[2] In terms of estimating the change point, $\beta_i, i = 1,...,N$ are nuisance parameters.

magnitude is large enough to overwhelm the nuisance parameters problem. Assumption 4(c) is satisfied when the change magnitude of each series is nonnegligible, thus this result confirms Bai (2010) and Kim (2011) in the current regression setup that increasing the number of series helps to identify the change point when cross-sectional dependence of the error terms is weak.

Remark 2 *It's worth pointing out that once $\hat{k}$ is consistent, the convergence rate of $\hat{k}$ is not well defined since $\hat{k}$ has to be an integer. If $\hat{\tau}$ is defined as $\hat{k}/T$, $\hat{\tau}$ has the same problem since $T\hat{\tau}$ has to be an integer. For a sequence of random variables $\{X_n, n = 1, ...\}$ and a sequence of positive numbers $\{C_n, n = 1, ...\}$, $X_n = O_p(C_n)$ is defined in the sense that X_n/C_n is bounded in probability. In most cases, we then derive the limiting distribution of X_n/C_n. However, when X_n is restricted to be integers, this definition is no longer appropriate. Suppose X_n is consistent for some integer θ, i.e., $P(|X_n - \theta| = 0) \to 1$, then for any C_n, $P(|X_n - \theta|/C_n = 0) = P(|X_n - \theta| = 0) \to 1$. This implies that the convergence rate of X_n is arbitrary and the limiting distribution of X_n/C_n is meaningless. Coming back to $\hat{k}$, the convergence rate of $\hat{k}$ will be arbitrary once $\hat{k}$ is consistent, and it is meaningless to derive the limiting distribution of $\hat{k} - k_0$ by multiplying $\hat{k} - k_0$ by some magnifying speed, say, N.*

Except for the above theoretical concern, in practice the change magnitude may be small and some series may not have structural change. Therefore, we will derive the limiting distribution of $\hat{k}$ under Assumption 4(a).

Theorem 4 *Under Assumptions 1-3 and 4(a),*

$$\hat{k} - k_0 \xrightarrow{d} \arg\max W(m),$$

where $W(m)$ is a partial sum process,

$$W(m) = \begin{cases} -|m|\lambda\xi + 2\sqrt{\lambda}\sum_{t=m+1}^{0} Z_t, \ \ for\ m \leq -1, \\ 0, \ \ for\ m = 0, \\ -|m|\lambda\xi - 2\sqrt{\lambda}\sum_{t=1}^{m} Z_t, \ \ for\ m \geq 1, \end{cases} \tag{13}$$

and $\{Z_t, t = -(k_0 - 1), ..., 0, ..., T - k_0\}$ is a discrete time Gaussian process with mean zero and autocovariance $\{\phi_{st}, 1 \leq s, t \leq T\}$.

The key feature of this distribution is that it is free of the underlying DGP so that inference of the change point can be made. Different from the univariate case in which normality comes from applying the functional central limit theorem to the weighted serial average $v_T \sum_{t=k+1}^{k_0} \delta_0' z_t e_t$, where $\delta_T = \delta_0 v_T$ and $v_T \to 0$ as $T \to \infty$, here the normality comes from applying the central limit theorem to the weighted cross-sectional average $\frac{1}{\sqrt{\lambda_N}} \sum_{i=1}^{N} \delta_i' z_{it} e_{it}$, also see Yao (1987), Bai (1997), Bai (2010) and Kim (2011). However, the essence of these two frameworks are the same. A second feature is that this distribution is derived allowing $z_{it}e_{it}$ to be dependent over t, while in Bai (2010) $z_{it}e_{it}$ is assumed to be uncorrelated over t. Thus our result is more general and empirically relevant. This step is nontrivial, see the Appendix for details, our proof is self-contained. Also note that the DGP is not required to be stationary even within each regime. The autocovariance function ϕ_{st} could be of any form, as long as parts (4) and (5) of Assumption 3 are satisfied.

It remains to estimate the parameters in the limiting distribution. λ and ξ can be estimated by $\hat{\lambda}_N = \sum_{i=1}^{N} \hat{\delta}_i' \hat{\delta}_i$ and $\hat{\xi} = \frac{1}{T} \frac{1}{\hat{\lambda}_N} \sum_{t=1}^{T} \sum_{i=1}^{N} \hat{\delta}_i' z_{it} z_{it}' \hat{\delta}_i$, where $\hat{\delta}_i$ and $\hat{e}_{is}$ can be obtained by least squares estimation of each subsample split at $\hat{k}$, and it will not be difficult to show the consistency of $\hat{\lambda}_N$ and $\hat{\xi}$. ϕ_{st} can be estimated by $\hat{\phi}_{st} = \frac{1}{\hat{\lambda}_N} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\delta}_i' z_{is} z_{it}' \hat{\delta}_i \hat{e}_{is} \hat{e}_{it}$ and if we assume that the DGP is independent over i, $\hat{\phi}_{st}$ can be simplified to $\frac{1}{\hat{\lambda}_N} \sum_{i=1}^{N} \hat{\delta}_i' z_{is} z_{it}' \hat{\delta}_i \hat{e}_{is} \hat{e}_{it}$. For each (s,t), it will not be difficult to show the consistency of $\hat{\phi}_{st}$. However, the limiting distribution relies on the consistency of the whole estimated covariance matrix $\{\hat{\phi}_{st}, 1 \leq s,t \leq T\}$. If we impose a further assumption on $z_{it}e_{it}$, e.g., AR(1) or martingale difference, then the consistency of $\{\hat{\phi}_{st}, 1 \leq s,t \leq T\}$ also will not be difficult to show. Once these estimated parameters are available, we can simulate the distribution directly and inference can be made based on this simulated distribution.

5 Estimation with strong cross-sectional dependence

This section considers estimating the change point when cross-sectional dependence is strong due to common factors. When factors are observable and explicitly incorporated into the model, we are back to the case with weak cross-sectional dependence. When factors are unobservable, and we estimate the change point ignoring the factors, the least squares estimator will be inconsistent even under Assumption 4(c). This is because when the cross-sectional dependence is strong, increasing the number of series no longer helps in identifying the change point. Kim (2011) discusses this phenomenon in the time trend break setup. In this case, a feasible way to recover consistency is using estimated factors to control for cross-sectional dependence. A similar method also can be found in Bai (2009) and Kim (2014).

We first present some regularity conditions.

Assumption 5 *(1)* $E\|F_t^0\|^4 < M < \infty$, $E(F_t^0) = 0$, $E(F_t^0 F_t^{0\prime}) = \Sigma_F$ *and* Σ_F *is positive definite.*

(2) (Doob's maximal inequality) Define $\{Q(1,k) = \sum_{t=1}^{k}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$, $\{Q(k,k_0) = \sum_{t=k+1}^{k_0}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$, $\{Q(k_0+1,k) = \sum_{t=k_0+1}^{k}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$, $\{Q(k,T) = \sum_{t=k+1}^{T}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$ *and* $Q_{jm}(1,k)$, $Q_{jm}(k,k_0)$, $Q_{jm}(k_0+1,k)$, $Q_{jm}(k,T)$ *as the j-th row and m-th column of* $Q(1,k)$, $Q(k,k_0)$, $Q(k_0+1,k)$, $Q(k,T)$ *respectively, then for* $1 \le j,m \le s$ *and* $1 < r < \infty$,

$E(\sup_{1\le l\le k}|Q_{jm}(1,l)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(1,k)|^r)$ *for* $1 \le k \le T$,

$E(\sup_{k\le l\le k_0-1}|Q_{jm}(l,k_0)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(k,k_0)|^r)$ *for* $0 \le k \le k_0 - 1$,

$E(\sup_{k_0+1\le l\le k}|Q_{jm}(k_0+1,l)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(k_0,k)|^r)$ *for* $k_0+1 \le k \le T$,

$E(\sup_{k\le l\le T-1}|Q_{jm}(l,T)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(k,T)|^r)$ *for* $0 \le k \le T-1$.

(3) There exists $M > 0$ *such that for* $r = 2,4$ *and* $1 \le j,m \le s$,

$E(|Q_{jm}(1,k)|^r) < k^{\frac{r}{2}} M$ *for* $1 \le k \le T$,

$E(|Q_{jm}(k,k_0)|^r) < (k_0 - k)^{\frac{r}{2}} M$ *for* $0 \le k \le k_0 - 1$,

$E(|Q_{jm}(k_0+1,k)|^r) < (k - k_0)^{\frac{r}{2}} M$ *for* $k_0 + 1 \le k \le T$,

$E(|Q_{jm}(k,T)|^r) < (T - k)^{\frac{r}{2}} M$ *for* $0 \le k \le T-1$.

Part (1) mainly assumes that the factors have a uniformly bounded fourth moment. Parts (2) and (3) are analogous to parts (3) and (4) of Assumption 2. Part (2) requires Doob's maximal inequality to be applicable to the process $Q_{jm}(1,k)$, $Q_{jm}(k,k_0)$, $Q_{jm}(k_0+1,k)$ and $Q_{jm}(k,T)$ for $1 \le j,m \le s$. Part (3) requires the factors to be serially weakly dependent, hence integrated factors are not allowed. Part (3) also implies $\frac{1}{k_0}\sum_{t=1}^{k_0} F_t^0 F_t^{0\prime} \xrightarrow{p} \Sigma_F$ and $\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} F_t^0 F_t^{0\prime} \xrightarrow{p} \Sigma_F$.

Assumption 6 *(1)* x_{it} *is independent of* F_t^0 *for all* i,t.

(2) Define $w_{it} = (x_{it}', F_t^{0\prime})'$, $W_{1i} = (w_{i1},...,w_{ik},0,...,0)'$ *and* $W_{2i} = (0,...,0,w_{i,k+1},...,w_{iT})'$, *there exists* $\rho_0 > 0$ *such that for some* $\eta > 0$ *and all* T *and* i, $\inf_{k>T(\tau_0-\eta)} \rho_{\min}(\frac{W_{1i}'W_{1i}}{k}) > \rho_0$ *and* $\inf_{k<T(\tau_0+\eta)} \rho_{\min}(\frac{W_{2i}'W_{2i}}{T-k}) > \rho_0$.

Part (1) is assumed to simplify the analysis, since our emphasis is the effect of cross-sectional dependence on the asymptotic properties of the change point estimator. If the regressors are correlated with the factors, our estimation procedure is no longer applicable, but in this case the change point can be estimated jointly with the factors by minimizing the sum of squared residuals as in Kim (2014). The results will be the same but the technical proof will be more complex and tedious. Part (2) is analogous to part (2) of Assumption 2 and has similar interpretation.

Assumption 7 $\|\lambda_i\| \leq \bar{\lambda} < \infty$, $\left\|\frac{1}{N}\Lambda'\Lambda - \Sigma_\Lambda\right\| \to 0$ *for some positive definite matrix* Σ_Λ.

Assumption 8 *The eigenvalues of* $\Sigma_F \Sigma_\Lambda$ *are distinct.*

Assumption 9 *(1)* e_{it} *is independent of* F_s^0 *for all* i, t, s.

(2) (Doob's maximal inequality) Define $\{P_i(1,k) = \sum_{t=1}^{k} F_t^0 e_{it}\}$, $\{P_i(k,k_0) = \sum_{t=k+1}^{k_0} F_t^0 e_{it}\}$, $\{P_i(k_0+1,k) = \sum_{t=k_0+1}^{k} F_t^0 e_{it}\}$, $\{P_i(k,T) = \sum_{t=k+1}^{T} F_t^0 e_{it}\}$ *and* $P_{ij}(1,k)$, $P_{ij}(k,k_0)$, $P_{ij}(k_0+1,k)$, $P_{ij}(k,T)$ *as the j-th element of* $P_i(1,k)$, $P_i(k,k_0)$, $P_i(k_0+1,k)$, $P_i(k,T)$ *respectively, then for* $1 \leq j \leq s$ *and* $1 < r < \infty$,

$E(\sup_{1\leq l\leq k} |P_{ij}(1,l)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(1,k)|^r)$ *for all* $1 \leq i \leq N$ *and* $1 \leq k \leq T$,

$E(\sup_{k\leq l\leq k_0-1} |P_{ij}(l,k_0)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(k,k_0)|^r)$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq k_0 - 1$,

$E(\sup_{k_0+1\leq l\leq k} |P_{ij}(k_0+1,l)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(k_0,k)|^r)$ *for all* $1 \leq i \leq N$ *and* $k_0+1 \leq k \leq T$,

$E(\sup_{k\leq l\leq T-1} |P_{ij}(l,T)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(k,T)|^r)$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq T-1$.

(3) There exists $M > 0$ *such that for* $r = 2, 4$ *and for* $1 \leq j \leq s$,

$E(|P_{ij}(1,k)|^r) < k^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $1 \leq k \leq T$,

$E(|P_{ij}(k,k_0)|^r) < (k_0 - k)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq k_0 - 1$,

$E(|P_{ij}(k_0+1,k)|^r) < (k - k_0)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $k_0 + 1 \leq k \leq T$,

$E(|P_{ij}(k,T)|^r) < (T - k)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq T - 1$.

Assumption 10 *There exists a positive constant* $M < \infty$ *such that:*

1 $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$, *for all* $i = 1, ..., N$, *and* $t = 1, ..., T$,

2 $E(\frac{e_s' e_t}{N}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \leq M$ *for* $s = 1, ..., T$, *and for* $t = 1, ..., T$, $\sum_{t=1}^{T} |\gamma_N(s,t)| \leq M$,

3 $E(e_{it}e_{jt}) = \tau_{ij,t}$ *with* $|\tau_{ij,t}| \leq \tau_{ij}$ *for some* τ_{ij} *and* $t = 1, ..., T$, *and for* $i = 1, ..., N$, $\sum_{j=1}^{N} |\tau_{ji}| \leq M$,

4 $E(e_{it}e_{js}) = \tau_{ij,ts}$ *for* $i, j = 1, ..., N$, *and* $t, s = 1, ..., T$, *also*

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |\tau_{ij,ts}| \leq M,$$

5 For every $(t, s = 1, ..., T)$, $E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]\right|^4 \leq M$, *6 For each* $u = 1, ..., T$, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|cov(e_{iu}e_{it}, e_{ju}e_{js})| \le M$ *and for each* $k = 1, ..., N$, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|cov(e_{it}e_{kt}, e_{js}e_{ks})| \le M$.

Assumptions 7 and 8 are standard in the factor literature. Assumption 9 is analogous to parts (1)-(3) of Assumption 3. Assumption 10 requires weak serial and cross-sectional dependence, and heteroskedasticity is allowed. Similar conditions are also assumed in Bai (2009), see the discussion therein for more details.

Assumption 11 *There exists* $M < \infty$ *such that:*

1. *For each* $t = 1, ..., T$, $E(\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{k_0}\sum_{i=1}^{N}F_s^0[e_{is}e_{it} - E(e_{is}e_{it})]\right\|^2) \le M$, *and* $E(\left\|\frac{1}{\sqrt{NT}}\sum_{s=k_0+1}^{T}\sum_{i=1}^{N}F_s^0[e_{is}e_{it} - E(e_{is}e_{it})]\right\|^2) \le M$;
2. $E(\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{k_0}\sum_{i=1}^{N}F_t^0\lambda_i' e_{it}\right\|^2) \le M$ *and* $E(\left\|\frac{1}{\sqrt{NT}}\sum_{t=k_0+1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i' e_{it}\right\|^2) \le M$;
3. *For each* $t = 1, ..., T$, $E(\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right\|^4) \le M$.

Assumption 12 *There exists* $M < \infty$ *such that:*

1. *For every* $s = 1, ..., T$,

$$E(\sup_{k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]\right|^2) \le M,$$

$$E(\sup_{k\le k_0}\frac{1}{k}\sum_{t=1}^{k}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]\right|^2) \le M,$$

$$E(\sup_{k>k_0}\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]\right|^2) \le M,$$

$$E(\sup_{k\ge k_0}\frac{1}{T-k}\sum_{t=k+1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]\right|^2) \le M,$$

2. $E(\sup_{k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right\|^2) \le M$,

$$E(\sup_{k\le k_0}\frac{1}{k}\sum_{t=1}^{k}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right\|^2) \le M,$$

$$E(\sup_{k>k_0}\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right\|^2) \le M,$$

$$E(\sup_{k\ge k_0}\frac{1}{T-k}\sum_{t=k+1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right\|^2) \le M.$$

Assumption 13 *There exists* $M < \infty$ *such that:*

1. $E(\sup_{k<k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{k_0}\sum_{i=1}^{N}F_t^0\lambda_i' e_{it}\right\|^2) \le M$,

$$E(\sup_{k\le k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{k}\sum_{i=1}^{N}F_t^0\lambda_i' e_{it}\right\|^2) \le M,$$

$$E(\sup_{k>k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}F_t^0\lambda_i' e_{it}\right\|^2) \le M,$$

$E(\sup_{k\geq k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i' e_{it}\right\|^2)\leq M.$

2. For each $j=1,...,T$,

$E(\sup_{k<k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{k_0}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M,$

$E(\sup_{k\leq k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{k}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M,$

$E(\sup_{k>k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M,$

$E(\sup_{k\geq k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{T}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M.$

Assumptions 11-13 are not restrictive since the summands are zero mean random variables. If Hajek-Renyi inequality were applicable, these conditions are directly satisfied. If further parametric assumptions are made on the factors, factor loadings and errors, it will not be difficult to verify these conditions. Here we simply lay them out so that these conditions are in their original form.

Assumption 14 *(1) For every* i, *there exists a compact set* B_i *such that* $\beta_i \in B_i$.

(2) For every i, $\frac{X_i' M_{F^0} X_i}{T} \xrightarrow{p} \Sigma_{ii}$ *for some positive definite* Σ_{ii} *as* $T\to\infty$.

(3) There exist $\mu>0$ *such that for every* i, $\inf_{F'F/T=I}\rho_{\min}(\frac{X_i' M_F X_i}{T})\geq\mu$ *as* $T\to\infty$.

(4) $\inf_{F'F/T=I}\rho_{\min}(D)>0$, *where* $D=\frac{1}{N}\sum_{i=1}^{N}D_i$, $D_i=B_i-C_i'A_i^{-1}C_i$, $A_i=\frac{X_i' M_F X_i}{T}$, $B_i=(\lambda_i\lambda_i')\otimes\frac{I_T}{T}$ *and* $C_i=\lambda_i'\otimes\frac{X_i' M_F}{T}$.

(5) For any i,j,t, e_{it} *is independent of* β_j *and* λ_j.

(6) There exist $M<\infty$ *such that for any* i *and* T, $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|E(x_{it}' x_{is}e_{it}e_{is})|\leq M$.

Assumption 14 is mainly borrowed from Assumptions A(iii), B and C in Song (2013) hold[3], see the explanation therein for these conditions. In the proof, we will utilize results in Han and Inoue (2014), Baltagi, Kao and Wang (2015b) and Song (2013) as intermediate steps. It can be verified that the assumptions in these papers are satisfied given all the above assumptions.

To recover consistency, we will use estimated factors as extra regressors to control for cross-sectional dependence. If the true change point k_0 were known, the factors can be estimated globally with the coefficients β_i as in Song (2013). Song (2013) shows that β_i will be $\sqrt{T}$-consistent for each i and the estimated factor space will be consistent. Without knowing k_0, a feasible way is to use $\hat{k}$, the estimated change point ignoring factors.

[3] ϵ_{it} in Song (2013) corresponds to e_{it} here.

Theorem 5 *Under Assumptions 1-3, 4(c), 5 and 6,* $\hat{k} - k_0 = O_p(1)$ *as* $(N, T) \to \infty$.

This result confirms the results in Kim (2011) for the current regression setup, i.e., when cross-sectional dependence is strong, more series do not increase the accuracy of the change point estimator. Nevertheless, $\hat{k} - k_0 = O_p(1)$ is good enough to estimate the factor space. It can be verified that with $O_p(1)$ estimation error, results in Song (2013) remain the same. Once the estimated factors are available and incorporated in the model as extra regressors, consistency of the least squares estimator can be recovered. Define $\tilde{k}$ as the change point estimator in the second step and $\tilde{\tau} = \tilde{k}/T$ as the estimated change fraction, we first show $\tilde{\tau}$ is consistent.

Theorem 6 *Under Assumptions 1-3, 4(c) and 5-14,* $\tilde{\tau} - \tau_0 = o_p(1)$ *as* $(N, T) \to \infty$ *and* $\frac{\sqrt{T}}{N} \to 0$.

Similar to Theorem 1, this result is mainly of theoretical interest and serves as a intermediate step to show the consistency of $\tilde{k}$. The condition $\frac{\sqrt{T}}{N} \to 0$ is required to guarantee the effect of using estimated factors is asymptotically negligible and appears frequently in the factor literature, see for example Bai and Ng (2006).

Theorem 7 *Under Assumption 1-3, 4(c) and 5-14,* $\tilde{k}$ *is consistent as* $(N, T) \to \infty$ *and* $\frac{\sqrt{T}}{N} \to 0$.

Again, $\frac{\sqrt{T}}{N} \to 0$ is required to eliminate the effect of using estimated factors. Note that in Theorem 4, $\frac{N}{\sqrt{T}} \to 0$ is required to eliminate the noise resulting from nuisance parameters, $\beta_i, i = 1, ..., N$. These two conditions are in conflict with each other, and consequently it is infeasible to derive the limiting distribution of $\tilde{k}$.[4] Intuitively speaking, for the factors, T is the dimension and N is the sample size while for β_i, N is the dimension and T is the sample size. If we also regard the factors as nuisance parameters, the effect of these two sets of nuisance parameters will not disappear simultaneously. This is the cost of using heterogeneous coefficients in panel data.

Remark 3 *In Kim (2014), the two N-T conditions can be satisfied simultaneously because Kim (2014) uses estimated factor loadings to control the unobservables and accurate estimation of loadings also requires T to be large relative to N. The reason that Kim (2014) can*

[4] However, if we can relax the condition $\frac{N}{\sqrt{T}} \to 0$ in Theorem 4, then there will be some room for both conditions being satisfied. This is technically quite difficult, but not impossible.

use estimated loadings is because in the time trend setup the regressors are common across different i.

6 Simulations

In this section we evaluate the limiting distribution derived in Section 4 and examine the effect of serial correlation. To simplify the analysis, we assume x_{it} is i.i.d. $N(1,1)$ over both i and t, $e_{it} = \rho e_{i,t-1} + \sigma_i \eta_{it}$ where η_{it} is i.i.d. $N(0,1)$ over both i and t and σ_i^2 is i.i.d. $\chi_2^2/2$ over i, and δ_i is i.i.d. $U(-1,1)$. For this DGP, $\{Z_t, t = -(k_0-1), ..., 0, ..., T-k_0\}$ is a Gaussian process with variance $\phi = \frac{1}{\lambda_N}\sum_{i=1}^N \delta_i' E(z_{it}z_{it}')\delta_i E(e_{it}^2)$ and correlation coefficient $\alpha_{st} = \rho^{|s-t|}\frac{\sum_{i=1}^N \delta_i' E(z_{is}z_{it}')\delta_i\sigma_i^2}{\sum_{i=1}^N \delta_i' E(z_{it}z_{it}')\delta_i\sigma_i^2} = \frac{1}{2}\rho^{|s-t|}$. For given values of N, $\lambda\phi$, $\lambda\xi$ and ρ, we can simulate the distribution of $\arg\max W(m)$ and in the current case $\lambda\phi = \sum_{i=1}^N \delta_i' E(z_{it}z_{it}')\delta_i E(e_{it}^2) \approx 2NE(\delta_i^2)E(\sigma_i^2) = \frac{2}{3}N$ and $\lambda\xi = \sum_{i=1}^N \delta_i' E(z_{it}z_{it}')\delta_i = 2NE(\delta_i^2) = \frac{2}{3}N$. Figures 1-2 are the simulated distributions obtained from 2000 replications with $T = 100$, $k_0 = 50$, $N = 1, 5$, 10 and 20 and $\rho = 0$, 0.4 and 0.8 respectively. When $\rho = 0$, the distribution is well shaped, but when $\rho > 0$, the distribution is no longer bell-shaped and becomes highly nonstandard. The probability of taking both ends and the true change point are high while the probability of taking the other points are approximately the same. Here $(\lambda\xi)^2/\lambda\phi = 2NE(\delta_i^2)$, if $E(\delta_i^2)$ is smaller, the nonstandardness will be more severe. Also note that $\alpha_{st} = \frac{1}{2}\rho^{|s-t|}$, even when $\rho = 0.8$, α_{st} is no more than 0.4. If $\frac{\sum_{i=1}^N \delta_i' E(z_{is}z_{it}')\delta_i\sigma_i^2}{\sum_{i=1}^N \delta_i' E(z_{it}z_{it}')\delta_i\sigma_i^2}$ is larger, the nonstandardness will also be more severe. Furthermore, with $E(\delta_i^2)$ fixed, while large N increases the probability of $\hat{k} = k_0$, it does not make the distribution more bell-shaped.

For such nonstandard distribution, it maybe better to base inference directly on the distribution, rather than on the constructed confidence intervals. Consider the case of $N = 20$ and $\rho = 0.8$ for example. Although the probability of $\hat{k} = k_0$ is already around 0.55, the 90% confidence interval is $[2, 99]$! Therefore, we suggest simulating the distribution directly using the estimated parameters and making inference based on this simulated distribution. For example, in the current setup the parameters λ, ξ, ϕ and α_{st} can be estimated by $\hat{\lambda}_N = \sum_{i=1}^N \hat{\delta}_i'\hat{\delta}_i$, $\hat{\xi} = \frac{1}{T}\frac{1}{\hat{\lambda}_N}\sum_{t=1}^T\sum_{i=1}^N \hat{\delta}_i' z_{it}z_{it}'\hat{\delta}_i$, $\hat{\phi} = \frac{1}{\hat{\lambda}_N}\sum_{i=1}^N \hat{\delta}_i'(\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N z_{it}z_{it}')\hat{\delta}_i(\frac{1}{T}\sum_{t=1}^T \hat{e}_{it}^2)$ and $\hat{\alpha}_{st} = \hat{\rho}^{|s-t|}\frac{\sum_{i=1}^N[\hat{\delta}_i'(\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N z_{it})]^2(\frac{1}{T}\sum_{t=1}^T \hat{e}_{it}^2)}{\sum_{i=1}^N \hat{\delta}_i'(\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N z_{it}z_{it}')\hat{\delta}_i(\frac{1}{T}\sum_{t=1}^T \hat{e}_{it}^2)}$, where $\hat{\delta}_i$ and $\hat{e}_{is}$ can be obtained by least squares estimation of each subsample split at $\hat{k}$ and $\hat{\rho} = \frac{1}{N}\sum_{i=1}^N(\sum_{t=2}^T \hat{e}_{it}\hat{e}_{i,t-1}/\sum_{t=2}^T \hat{e}_{i,t-1}^2)$.

7 Conclusion

This paper studies the joint limit asymptotics of the least squares estimator of a common change point in large heterogeneous panel data models. A general Hajek-Renyi inequality is introduced to solve the fundamental issue that for random variables $X_{iT} = O_p(1)$ (or $o_p(1)$) as $T \to \infty$, $\frac{1}{N}\sum_{i=1}^{N} X_{iT}$ is not necessarily $O_p(1)$ (or $o_p(1)$ correspondingly) as N and T go to infinity jointly. This new technique is quite powerful and we conjecture that it will also be useful in other places. Consistency of the least squares estimator is then established under various sets of conditions on the change magnitude and N-T ratio. Both weak and strong cross-sectional dependence of the errors are considered and in the latter case estimated factors are used to control the cross-sectional dependence. The limiting distribution is derived allowing the errors to be cross-sectionally weakly dependent and serially dependent and heteroskedastic of unknown form, and inference is feasible based on the simulated distribution using estimated parameters.

Acknowledgements

Chapter 1 is based on the working paper Baltagi, Kao and Wang (2015a).

References

[1] Bada, O., Gualtieri, J., Kneip, A., Sickles, R.C., 2015. Panel data models with multiple jump discontinuities in the parameters. Working paper, Rice University.

[2] Bai, J., 1997. Estimation of a change point in multiple regression models. Review of Economics and Statistics 79, 551–563.

[3] Bai, J., Ng, S., 2006. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. Econometrica 74, 1133-1150.

[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 157, 78–92.

[6] Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating breaks in multivariate time series. Review of Economic Studies 65, 395–432.

[7] Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

[8] Baltagi, B.H., Feng, Q., Kao, C., 2011. Testing for sphericity in a fixed effects panel data model. Econometrics Journal 14, 25–47.

[9] Baltagi, B.H., Feng, Q., Kao, C., 2012. A Lagrange Multiplier test for cross-sectional dependence in a fixed effects panel data model. Journal of Econometrics 170, 164-177.

[10] Baltagi, B.H., Feng, Q., Kao, C., 2016. Estimation of heterogeneous panels with structural breaks. Journal of Econometrics 191, 176-195.

[11] Baltagi, B.H., Kao, C., Liu, L., 2014. Estimation and identification of change points in panel models with nonstationary or stationary regressors and error terms. Econometric Reviews, forthcoming.

[12] Baltagi, B.H., Kao, C., Wang, F., 2015a. Change point estimation in large heterogeneous panels. Working Paper, Syracuse University.

[13] Baltagi, B.H., Kao, C., Wang, F., 2015b. Identification and estimation of a large factor model with structural instability. Working paper, Syracuse University.

[14] Fazekas, I., Klesov, O., 2001. A general approach to the strong law of large numbers. Theory of Probability & Its Applications 45, 436-449.

[15] Hajek, J., Renyi, A., 1955. Generalization of an inequality of Kolmogorov. Acta Mathematica Hungarica 6, 281-283.

[16] Han, X., Inoue, A., 2014. Tests for parameter instability in dynamic factor models. Econometric Theory, 1-36.

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

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

[19] Lee, L.F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154, 165-185.

[20] Lee, L.F., Yu, J., 2010. Some recent developments in spatial panel data models. Regional Science and Urban Economics 40, 255-271.

[21] Ng, S., 2006. Testing cross-section correlation in panel data using spacings. Journal of Business and Economic Statistics 24, 12-23.

[22] Perron, P, 2006. Dealing with structural breaks. Palgrave Handbook of Econometrics 1, 278-352.

[23] Pesaran, M.H., 2004. General diagnostic test for cross section dependence in panels. Working Paper, University of Cambridge & USC.

[24] Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with multifactor error structure. Econometrica 74, 967-1012.

[25] Pesaran, M.H., 2012. Testing weak cross-sectional dependence in large panels. IZA Discussion Paper No. 6432, March 2012.

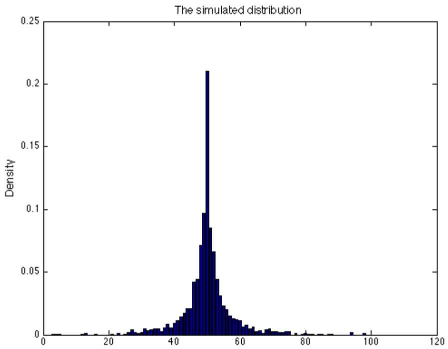
[26] Pesaran, M.H., Ullah, A., Yamagata, T., 2008. A bias-adjusted LM test of error cross section independence. Econometrics Journal 11, 105-127.

[27] Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. Econometrica 75, 459–502.

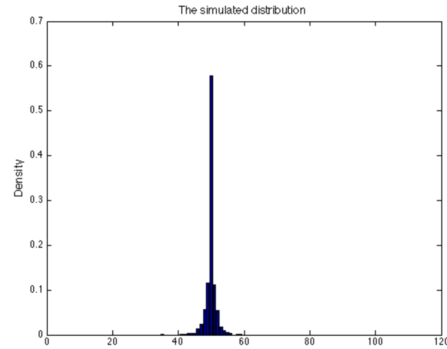
[28] Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Manuscript, Columbia University.

[29] Yao, Y.C., 1987. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. Annals of Statistics 15, 1321–1328.

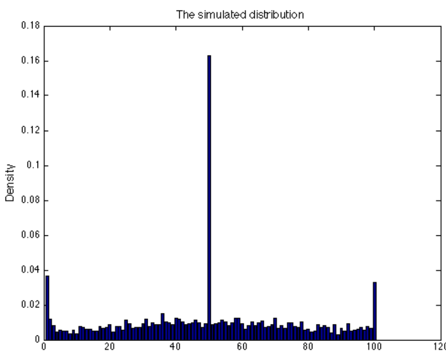
[30] Yu, J., De Jong, R., Lee, L.F. 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146, 118-134.



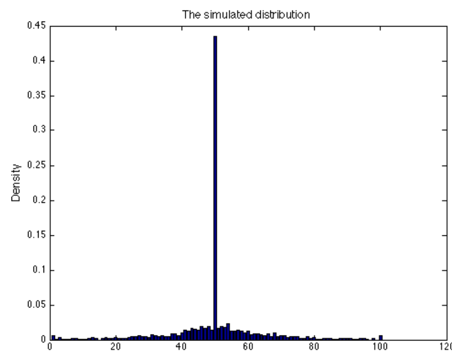
$T = 100,\ k_0 = 50,\ N = 1$ and $\rho = 0$



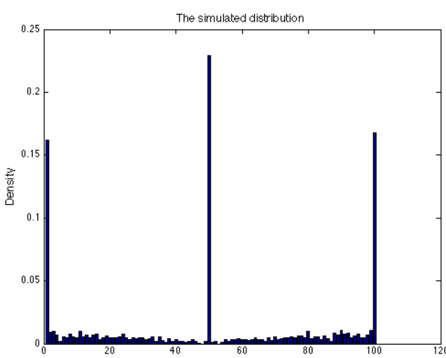
$T = 100,\ k_0 = 50,\ N = 5$ and $\rho = 0$



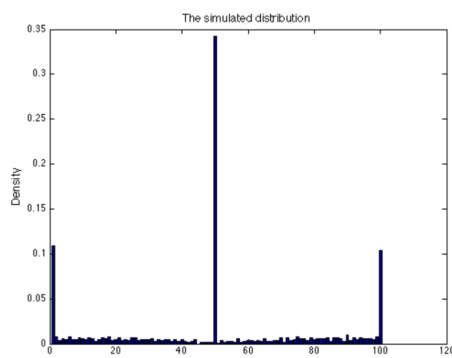
$T = 100,\ k_0 = 50,\ N = 1$ and $\rho = 0.4$



$T = 100,\ k_0 = 50,\ N = 5$ and $\rho = 0.4$

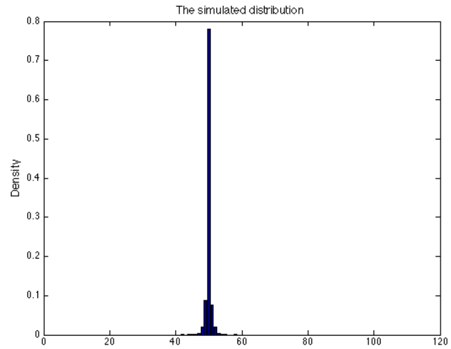


$T = 100,\ k_0 = 50,\ N = 1$ and $\rho = 0.8$

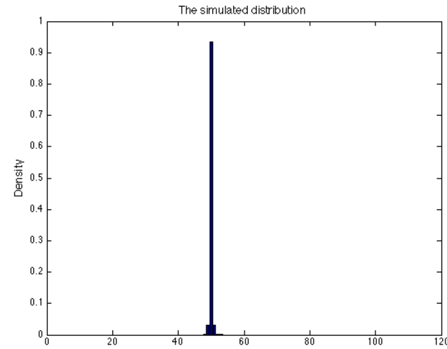


$T = 100,\ k_0 = 50,\ N = 5$ and $\rho = 0.8$

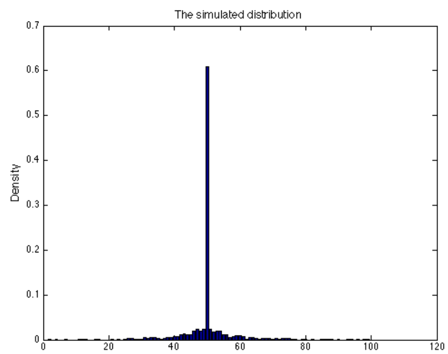
Figure 1: Simulated distribution of $argmaxW(m)$ for $T = 100, k_0 = 50, N = 1$ and 5



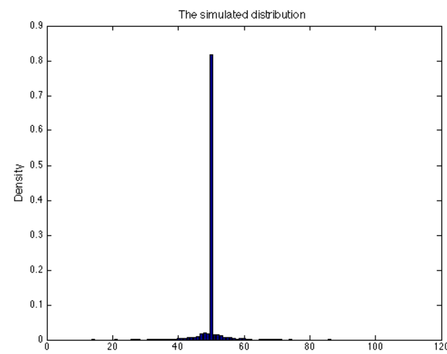
$T = 100,\ k_0 = 50,\ N = 10$ and $\rho = 0$



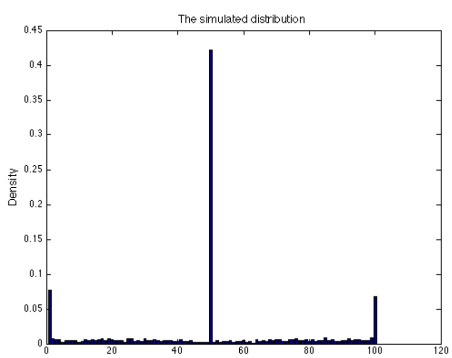
$T = 100,\ k_0 = 50,\ N = 20$ and $\rho = 0$



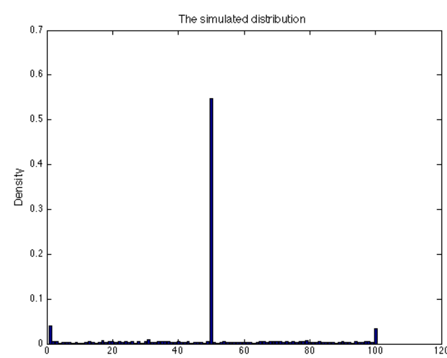
$T = 100,\ k_0 = 50,\ N = 10$ and $\rho = 0.4$



$T = 100,\ k_0 = 50,\ N = 20$ and $\rho = 0.4$



$T = 100,\ k_0 = 50,\ N = 10$ and $\rho = 0.8$



$T = 100,\ k_0 = 50,\ N = 20$ and $\rho = 0.8$

Figure 2: Simulated distribution of $argmaxW(m)$ for $T = 100, k_0 = 50, N = 10$ and 20

APPENDIX

Lemma 2 *For each i and $k < k_0$,*

$$(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i}) \geq R'[(X_{\Delta i}'X_{\Delta i})(X_{2i}'X_{2i})^{-1}(X_{0i}'X_{0i})]R.$$

Proof. See Bai (1997) Lemma A.1. ■

Lemma 3 *Under Assumptions 1-3, there exists $M > 0$ such that for all N and T, for each i,*

(1) $E(\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|^2) \leq M$,

(2) $E(\sup_{k<k_0-C}\left\|\frac{(X_{\Delta i}'X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX}\right\|)^4 \leq \frac{M}{(C+1)^2}$,

(3) $E(\sup_{k<k_0}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1}\right\|)^2 \leq \frac{M}{T}$,

(4) $E(\left\|\frac{X_{0i}'X_{0i}}{T-k_0} - \Sigma_i^{XX}\right\|)^4 \leq \frac{M}{T^2}$,

(5) $E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|^4) \leq M$,

(6) $E(\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4) \leq M$,

(7) $E(\sup_{k<k_0}\left\|\frac{e_i'M_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4) \leq M\log T$,

(8) $E(\sup_{k<k_0}\left\|\frac{e_i'M_iZ_{\Delta i}}{|k-k_0|}\right\|^4) \leq M$,

(9) $E(\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4) \leq M$,

(10) $E(\sup_{k<k_0}\left\|\frac{Z_{2i}'M_ie_i}{\sqrt{T-k}}\right\|^4) \leq M$,

(11) $E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'M_iZ_{2i}}{|k-k_0|}\right\|^4) \leq M$,

(12) $E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2) \leq \frac{M}{T}$,

(13) $E(\left\|(\frac{Z_{0i}'M_iZ_{0i}}{T-k_0})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2 \leq \frac{M}{T}$,

(14) $\sup_{k\in K,k\leq k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\| \leq M$,

(15) $\sup_{k\in K,k\leq k_0}\left\|[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \leq M$,

(16) $\sup_{k\in K^c,k<k_0}\left\|(\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1}\right\| \leq M$,

(17) $E(\sup_{k\in K^c,k<k_0}\left\|\frac{e_i'M_iZ_{1i}}{\sqrt{k}}\right\|^2) \leq M\log T$,

(18) $E(\sup_{k\in K^c,k<k_0}\left\|\frac{Z_{0i}'M_iZ_{1i}}{k}\right\|) \leq M$.

Proof. (1)

$$E(\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|^2) = E(\left\|\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} x_{it}e_{it}\right\|^2) = \sum\nolimits_{j=1}^{p}\frac{1}{T}E[S_{ij}(1,T)]^2 \le pM,$$

where the last inequality follows from part (3) of Assumption 3.

(2) Take $r=4$ in part (3) of Assumption 2, we have for each $1\le j\le p$, $1\le m\le p$, $1\le i\le N$ and $0\le k\le k_0-1$,

$$E(\sup_{k\le t\le k_0-1}|R_{ijm}(t,k_0)|^4) \le (\frac{4}{3})^4 E(|R_{ijm}(k,k_0)|^4) \le (\frac{4}{3})^4(k_0-k)^2M.$$

Next, using Lemma 1 with $r=4$, $S_l = R_{ijm}(k_0-C-l,k_0)$, $\beta_{k_0-k}=k_0-k$ and

$$\alpha_{k_0-k} = \begin{cases} (\frac{4}{3})^4(k_0-k)^2M \;\; for\; k_0-k = C+1 \\ (\frac{4}{3})^4[(k_0-k)^2-(k_0-k-1)^2]M \;\; for\; C+2\le k_0-k\le T(\tau_0-\eta) \end{cases},$$

we have for each $1\le j\le p$, $1\le m\le p$ and $1\le i\le N$,

$$\begin{aligned} E(\sup_{k\in K(C),k<k_0}\left|\frac{1}{k_0-k}R_{ijm}(k,k_0)\right|^4) &\le 4[\frac{(\frac{4}{3})^4(C+1)^2M}{(C+1)^4} + \sum\nolimits_{k_0-k=C+2}^{T\tau_0}\frac{(\frac{4}{3})^4[2(k_0-k)-1]M}{(k_0-k)^4}] \\ &\le 4(\frac{4}{3})^4M[\frac{1}{(C+1)^2} + 2\sum\nolimits_{k_0-k=C+2}^{\infty}\frac{1}{(k_0-k)^3}] \\ &\le \frac{12(\frac{4}{3})^4M}{(C+1)^2}, \end{aligned}$$

where the last inequality follows from

$$\begin{aligned} \sum\nolimits_{i=C+2}^{\infty}\frac{1}{i^3} &< \sum\nolimits_{i=C+2}^{\infty}\frac{1}{i}(\frac{1}{i-1}-\frac{1}{i}) < \frac{1}{C+2}\sum\nolimits_{i=C+2}^{\infty}(\frac{1}{i-1}-\frac{1}{i}) \\ &= \frac{1}{(C+2)(C+1)} \le \frac{1}{(C+1)^2}. \end{aligned}$$

Thus,

$$\begin{aligned} &E(\sup_{k\in K(C),k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}(x_{it}x_{it}'-\Sigma_i^{XX})\right\|^4) \\ \le\;& p^2\sum\nolimits_{j=1}^{p}\sum\nolimits_{m=1}^{p}E(\sup_{k\in K(C),k<k_0}\left|\frac{1}{k_0-k}R_{ijm}(k,k_0)\right|^4) \\ \le\;& \frac{12(\frac{4}{3})^4p^4M}{(C+1)^2}. \end{aligned}$$

(3) Noting that $A^{-1}-B^{-1} = B^{-1}(B-A)A^{-1}$,

$$\begin{aligned} &E(\sup_{k<k_0}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}-(\Sigma_i^{XX})^{-1}\right\|)^2 \\ \le\;& E(\left\|(\Sigma_i^{XX})^{-1}\right\|^2\sup_{k<k_0}\left\|\frac{X_{2i}'X_{2i}}{T-k}-\Sigma_i^{XX}\right\|^2\sup_{k<k_0}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|^2). \end{aligned}$$

By parts (1) and (2) of Assumption 2, the first and third terms are bounded, hence it suffices to show that $E(\sup_{k<k_0}\left\|\frac{X_{2i}'X_{2i}}{T-k}-\Sigma_i^{XX}\right\|)^2=O(\frac{1}{T})$ uniformly over i. Take $r=2$ in part (3) of Assumption 2, we have for each $1\le j\le p$, $1\le m\le p$, $1\le i\le N$ and $0\le k\le T-1$,

$$E(\sup_{k\le t\le T-1}|R_{ijm}(t,T)|^2)\le 4E(|R_{ijm}(k,T)|^2)\le 4(T-k)M,$$

then using Lemma 1 with $r=2$, $S_l=R_{ijm}(k_0-l,T)$, $\beta_{T-k}=T-k$ and

$$\alpha_{T-k}=\begin{cases}4(T-k_0+1)M\ for\ T-k=T-k_0+1\\ 4M\ for\ T-k_0+2\le T-k\le T\end{cases},$$

we have for each $1\le j\le p$, $1\le m\le p$ and $1\le i\le N$,

$$\begin{aligned}E(\sup_{k<k_0}\left|\frac{1}{T-k}R_{ijm}(k,T)\right|^2) &\le 4[\frac{4(T-k_0+1)M}{(T-k_0+1)^2}+\sum\nolimits_{T-k=T-k_0+2}^{T}\frac{4M}{(T-k)^2}]\\ &\le 16M[\frac{1}{T-k_0+1}+\sum\nolimits_{T-k=T-k_0+2}^{T}\frac{1}{(T-k)^2}]\\ &\le \frac{32M}{T-k_0+1}.\end{aligned}$$

Thus,

$$E(\sup_{k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k+1}^{T}(x_{it}x_{it}'-\Sigma_i^{XX})\right\|^2)\le\sum\nolimits_{j=1}^{p}\sum\nolimits_{m=1}^{p}E(\sup_{k<k_0}\left|\frac{1}{T-k}R_{ijm}(k,T)\right|^2)$$
$$\le\frac{32p^2M}{T-k_0+1}.$$

(4)

$$\begin{aligned}E(\left\|\frac{X_{0i}'X_{0i}}{T-k_0}-\Sigma_i^{XX}\right\|)^4 &= E\left\|\frac{1}{T-k_0}\sum\nolimits_{t=k_0+1}^{T}(x_{it}x_{it}'-\Sigma_i^{XX})\right\|^4\\ &= \frac{1}{(T-k_0)^4}E[\sum\nolimits_{j=1}^{p}\sum\nolimits_{m=1}^{p}R_{ijm}^2(k_0,T)]^2\\ &\le \frac{1}{(T-k_0)^4}E[p^2\sum\nolimits_{j=1}^{p}\sum\nolimits_{m=1}^{p}R_{ijm}^4(k_0,T)]\\ &= \frac{p^2}{(T-k_0)^4}\sum\nolimits_{j=1}^{p}\sum\nolimits_{m=1}^{p}E[R_{ijm}^4(k_0,T)]\\ &\le \frac{p^4M}{(T-k_0)^2},\end{aligned}$$

where the last inequality follows from part (4) of Assumption 2.

(5)

$$\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|^2\le\left\|\frac{X_{\Delta i}'X_i}{|k-k_0|}\right\|^2=\left\|\frac{X_{\Delta i}'X_{\Delta i}}{|k-k_0|}\right\|^2\le 2\left\|\frac{X_{\Delta i}'X_{\Delta i}}{|k-k_0|}-\Sigma_i^{XX}\right\|^2+2\left\|\Sigma_i^{XX}\right\|^2,$$

hence

$$\sup_{k<k_0}\left\|\frac{Z'_{\Delta i}X_i}{|k-k_0|}\right\|^4 \le 8\sup_{k<k_0}\left\|\frac{X'_{\Delta i}X_{\Delta i}}{|k-k_0|}-\Sigma_i^{XX}\right\|^4+8\left\|\Sigma_i^{XX}\right\|^4.$$

Take $C=0$ in part (2), the proof is accomplished.

(6) Under part (2) of Assumption 2, the proof follows.

(7)

$$\sup_{k<k_0}\left\|\frac{e'_iM_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4 \le 8\sup_{k<k_0}\left\|\frac{e'_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4+8\sup_{k<k_0}\left\|\frac{e'_iX_i(X'_iX_i)^{-1}X_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4.$$

For the first term, take $r=4$ in part (2) of Assumption 3, we have for each $1\le j\le p$, $1\le i\le N$ and $0\le k\le k_0-1$,

$$E(\sup_{k\le t\le k_0-1}|S_{ij}(t,k_0)|^4)\le(\frac{4}{3})^4E(|S_{ij}(k,k_0)|^4)\le(\frac{4}{3})^4(k_0-k)^2M.$$

Using Lemma 1 with $r=4$, $S_l=S_{ij}(k_0-l,k_0)$, $\beta_{k_0-k}=\sqrt{k_0-k}$ and $\alpha_{k_0-k}=(\frac{4}{3})^4[(k_0-k)^2-(k_0-k-1)^2]M$ for $1\le k_0-k\le T\tau_0$, we have for each $1\le j\le p$ and $1\le i\le N$,

$$\begin{aligned}E(\sup_{k<k_0}\left|\frac{1}{\sqrt{k_0-k}}S_{ij}(k,k_0)\right|^4) &\le 4\sum_{k_0-k=1}^{T\tau_0}\frac{(\frac{4}{3})^4[2(k_0-k)-1]M}{(k_0-k)^2}\\ &\le 8(\frac{4}{3})^4M\sum_{k_0-k=1}^{T\tau_0}\frac{1}{k_0-k}\le 8(\frac{4}{3})^4MO(\log T).\end{aligned}$$

Thus,

$$E(\sup_{k<k_0}\left\|\frac{e'_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4)\le p\sum_{j=1}^{p}E(\sup_{k<k_0}\left|\frac{1}{\sqrt{k_0-k}}S_{ij}(k,k_0)\right|^4)\le 8(\frac{4}{3})^4p^2MO(\log T).$$

For the second term,

$$\begin{aligned}\sup_{k<k_0}\left\|\frac{e'_iX_i(X'_iX_i)^{-1}X'_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4 &\le \left\|\frac{e'_iX_i}{\sqrt{T}}\right\|^4\left\|(\frac{X'_iX_i}{T})^{-1}\right\|^4\sup_{k<k_0}\left\|\frac{X'_iZ_{\Delta i}}{|k-k_0|}\right\|^4\\ &\le M^4\left\|\frac{e'_iX_i}{\sqrt{T}}\right\|^4\sup_{k<k_0}\left\|\frac{X'_iZ_{\Delta i}}{|k-k_0|}\right\|^4,\end{aligned}$$

where the last inequality follows from part (2) of Assumption 2. Hence,

$$E(\sup_{k<k_0}\left\|\frac{e'_iX_i(X'_iX_i)^{-1}X'_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4)\le M^4[E(\left\|\frac{e'_iX_i}{\sqrt{T}}\right\|^8)E(\sup_{k<k_0}\left\|\frac{X'_iZ_{\Delta i}}{|k-k_0|}\right\|^8)]^{\frac{1}{2}}=O(1),$$

in which $E(\left\|\frac{e'_iX_i}{\sqrt{T}}\right\|^8)=O(1)$ and $E(\sup_{k<k_0}\left\|\frac{X'_iZ_{\Delta i}}{|k-k_0|}\right\|^8)=O(1)$ can be proved following the same procedure as part (1) and part (5) respectively.

(8) The proof is similar to part (7). For the first term, the difference is $\beta_{k_0-k} = k_0 - k$ and thus

$$E(\sup_{k<k_0}\left\|\frac{e_i' Z_{\Delta i}}{|k-k_0|}\right\|^4) \leq p\sum_{j=1}^p E(\sup_{k<k_0}\left|\frac{1}{|k-k_0|}S_{ij}(k,k_0)\right|^4)$$
$$\leq p\sum_{j=1}^p 4\sum_{k_0-k=1}^{T\tau_0}\frac{(\frac{4}{3})^4[2(k_0-k)-1]M}{(k_0-k)^4}$$
$$\leq 8(\frac{4}{3})^4 p^2 M\sum_{k_0-k=1}^{T\tau_0}\frac{1}{(k_0-k)^3} = O(1).$$

For the second term, the difference is

$$E(\sup_{k<k_0}\left\|\frac{e_i' X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}}{|k-k_0|}\right\|^4) \leq \frac{1}{T^2}M^4[E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^8)E(\sup_{k<k_0}\left\|\frac{X_i'Z_{\Delta i}}{|k-k_0|}\right\|^8)]^{\frac{1}{2}} = O(\frac{1}{T^2}).$$

(9)

$$E(\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4) \leq 8E(\left\|\frac{e_i'Z_{0i}}{\sqrt{T-k_0}}\right\|^4) + 8E(\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4).$$

Under part (3) of Assumption 3, the first term is $O(1)$. For the second term,

$$E(\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4) \leq E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^4\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4\left\|\frac{X_i'Z_{0i}}{T-k_0}\right\|^4)$$
$$\leq M^4[E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^8)E(\left\|\frac{X_i'Z_{0i}}{T-k_0}\right\|^8)]^{\frac{1}{2}} = O(1).$$

(10) The proof is also similar to part (7).

$$\sup_{k<k_0}\left\|\frac{e_i'M_iZ_{2i}}{\sqrt{T-k}}\right\|^4 \leq 8\sup_{k<k_0}\left\|\frac{e_i'Z_{2i}}{\sqrt{T-k}}\right\|^4 + 8\sup_{k<k_0}\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{2i}}{\sqrt{T-k}}\right\|^4.$$

For the first term, take $r=4$ in part (2) of Assumption 3, we have for each $1 \leq j \leq p$, $1 \leq i \leq N$ and $0 \leq k \leq T-1$,

$$E(\sup_{k\leq t\leq T-1}|S_{ij}(t,T)|^4) \leq (\frac{4}{3})^4 E(|S_{ij}(k,T)|^4) \leq (\frac{4}{3})^4(T-k)^2 M.$$

Using Lemma 1 with $r=4$, $S_l = S_{ij}(k_0-l,T)$, $\beta_{T-k} = \sqrt{T-k}$ and $\alpha_{T-k} = (\frac{4}{3})^4[(T-k)^2 - (T-k-1)^2]M$ for $T-k_0+1 \leq T-k \leq T$, we have for each $1 \leq j \leq p$ and $1 \leq i \leq N$,

$$E(\sup_{k<k_0}\left|\frac{1}{\sqrt{T-k}}S_{ij}(k,T)\right|^4) \leq 4\sum_{T-k=T-k_0+1}^{T}\frac{(\frac{4}{3})^4[2(T-k)-1]M}{(T-k)^2}$$
$$\leq 8(\frac{4}{3})^4 M\sum_{T-k=T-k_0+1}^{T}\frac{1}{T-k}$$
$$\to 8(\frac{4}{3})^4 M\log\frac{1}{1-\tau_0},$$

since

$$\begin{aligned}
& \sum\nolimits_{T-k=T-k_0+1}^{T} \frac{1}{T-k} \\
= & \sum\nolimits_{i=1}^{T} \frac{1}{i} - \sum\nolimits_{i=1}^{T-k_0} \frac{1}{i} \\
= & (\sum\nolimits_{i=1}^{T} \frac{1}{i} - \log T) - (\sum\nolimits_{i=1}^{T-k_0} \frac{1}{i} - \log(T-k_0)) + (\log T - \log(T-k_0)) \\
\to & \gamma - \gamma + \log \frac{1}{1-\tau_0} = \log \frac{1}{1-\tau_0},
\end{aligned}$$

where γ is Euler-Mascheroni constant. Thus,

$$E(\sup_{k<k_0} \left\| \frac{e_i' Z_{2i}}{\sqrt{T-k}} \right\|^4) \leq p \sum\nolimits_{j=1}^{p} E(\sup_{k<k_0} \left| \frac{1}{\sqrt{T-k}} S_{ij}(k,T) \right|^4) \leq 8(\frac{4}{3})^4 p^2 M \log \frac{1}{1-\tau_0} = O(1).$$

For the second term,

$$\begin{aligned}
E(\sup_{k<k_0} \left\| \frac{e_i' X_i (X_i' X_i)^{-1} X_i Z_{2i}}{\sqrt{T-k}} \right\|^4) & \leq E(\left\| \frac{e_i' X_i}{\sqrt{T}} \right\|^4 \left\| (\frac{X_i' X_i}{T})^{-1} \right\|^4 \sup_{k<k_0} \left\| \frac{X_i' Z_{2i}}{T-k} \right\|^4) \\
& \leq M^4 [E(\left\| \frac{e_i' X_i}{\sqrt{T}} \right\|^8) E(\sup_{k<k_0} \left\| \frac{X_i' Z_{2i}}{T-k} \right\|^8)]^{\frac{1}{2}} = O(1),
\end{aligned}$$

in which $E(\sup_{k<k_0} \left\| \frac{X_i' Z_{2i}}{T-k} \right\|^8) = O(1)$ can be proved following the same procedure as part (5).

(11)

$$E(\sup_{k<k_0} \left\| \frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|} \right\|^4) \leq 8E(\sup_{k<k_0} \left\| \frac{Z_{\Delta i}' Z_{2i}}{|k-k_0|} \right\|^4) + 8E(\sup_{k<k_0} \left\| \frac{Z_{\Delta i}' X_i (X_i' X_i)^{-1} X_i' Z_{2i}}{|k-k_0|} \right\|^4).$$

The first term is $O(1)$ based on $\left\| \frac{Z_{\Delta i}' Z_{2i}}{|k-k_0|} \right\| = \left\| \frac{Z_{\Delta i}' Z_{\Delta i}}{|k-k_0|} \right\| \leq \left\| \frac{X_{\Delta i}' X_{\Delta i}}{|k-k_0|} \right\|$ and part (2). For the second term,

$$\begin{aligned}
E(\sup_{k<k_0} \left\| \frac{Z_{\Delta i}' X_i (X_i' X_i)^{-1} X_i' Z_{2i}}{|k-k_0|} \right\|^4) & \leq E(\left\| (\frac{X_i' X_i}{T})^{-1} \right\|^4 \sup_{k<k_0} \left\| \frac{Z_{\Delta i}' X_i}{|k-k_0|} \right\|^4 \sup_{k<k_0} \left\| \frac{X_i' Z_{2i}}{T-k} \right\|^4) \\
& \leq M^4 [E(\sup_{k<k_0} \left\| \frac{Z_{\Delta i}' X_i}{|k-k_0|} \right\|^8) E(\sup_{k<k_0} \left\| \frac{X_i' Z_{2i}}{T-k} \right\|^8)]^{\frac{1}{2}} = O(1).
\end{aligned}$$

(12) Noting that $A^{-1} - B^{-1} = B^{-1}(B-A)A^{-1}$,

$$\begin{aligned}
& E(\sup_{k \in K(k_0), k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\|)^2 \\
\leq & E(\sup_{k \in K(k_0), k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} \right\|^2 \sup_{k \in K(k_0), k<k_0} \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\|^2 \\
& \sup_{k \in K(k_0), k<k_0} \left\| \frac{Z_{2i}' M_i Z_{2i}}{T-k} - [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}] \right\|^2).
\end{aligned}$$

By parts (14) and (15) below, the first and the second terms are bounded. For the third term, we have

$$
\begin{aligned}
& E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}])\right\|^2) \\
= \; & E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'Z_{2i}}{T-k}-\Sigma_i^{ZZ})-(\frac{Z_{2i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{T-k}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ})\right\|^2) \\
\leq \; & 2E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'Z_{2i}}{T-k}-\Sigma_i^{ZZ}\right\|^2) \\
& +2E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{T-k}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2).
\end{aligned}
$$

The first term is $O(\frac{1}{T})$. For the second term,

$$
\begin{aligned}
& E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{T-k}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2) \\
= \; & E(\sup_{k\in K(k_0),k<k_0}(\frac{T-k}{T})^2\left\|\frac{Z_{2i}'X_i}{T-k}(\frac{X_i'X_i}{T})^{-1}\frac{X_i'Z_{2i}}{T-k}-\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2) \\
\leq \; & E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i}{T-k}(\frac{X_i'X_i}{T})^{-1}\frac{X_i'Z_{2i}}{T-k}-\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2)=O(\frac{1}{T}),
\end{aligned}
$$

since $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i}{T-k}-\Sigma_i^{ZX}\right\|^2)=O(\frac{1}{T})$ and $E(\left\|(\frac{X_i'X_i}{T})^{-1}-(\Sigma_i^{XX})^{-1}\right\|^2)=O(\frac{1}{T})$.

(13) Following the same procedure as part (12), the proof is straightforward.

(14)

$$
\begin{aligned}
Z_{2i}'M_iZ_{2i} &= R'[X_{2i}'X_{2i}-X_{2i}'X_i(X_i'X_i)^{-1}X_i'X_{2i}]R \\
&= R'[X_{2i}'X_{2i}-X_{2i}'X_{2i}(X_i'X_i)^{-1}X_{2i}'X_{2i}]R \\
&= R'[X_{2i}'X_{2i}-X_{2i}'X_{2i}(X_i'X_i)^{-1}(X_i'X_i-X_{1i}'X_{1i})]R \\
&= R'[X_{2i}'X_{2i}(X_i'X_i)^{-1}X_{1i}'X_{1i}]R \\
&= R'[(X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}]^{-1}R,
\end{aligned}
$$

where the last equality follows from

$$
\begin{aligned}
[X_{2i}'X_{2i}(X_i'X_i)^{-1}X_{1i}'X_{1i}]^{-1} &= (X_{1i}'X_{1i})^{-1}(X_i'X_i)(X_{2i}'X_{2i})^{-1} \\
&= (X_{1i}'X_{1i})^{-1}(X_{1i}'X_{1i}+X_{2i}'X_{2i})(X_{2i}'X_{2i})^{-1} \\
&= (X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}.
\end{aligned}
$$

It follows that

$$\begin{aligned}
\rho_{\min}(Z_{2i}'M_iZ_{2i}) &= \rho_{\min}(R'[(X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}]^{-1}R) \\
&\geqslant \rho_{\min}([(X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}]^{-1}) \\
&= \frac{1}{\rho_{\max}((X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1})},
\end{aligned}$$

and thus

$$\begin{aligned}
\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\| &\leq \sqrt{q}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\|_{op} = \frac{\sqrt{q}(T-k)}{\rho_{\min}(Z_{2i}'M_iZ_{2i})} \\
&\leq \sqrt{q}(T-k)\rho_{\max}((X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}) \\
&\leq \sqrt{q}(T-k)[\rho_{\max}((X_{1i}'X_{1i})^{-1})+\rho_{\max}((X_{2i}'X_{2i})^{-1})] \\
&= \sqrt{q}(\frac{T-k}{k}\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|+\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|).
\end{aligned}$$

By part (2) of Assumption 2, both $\sup\limits_{k\in K,k\leq k_0}\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|$ and $\sup\limits_{k\in K,k\leq k_0}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|$ are bounded, the proof is finished.

(15) First, noting that $\Sigma_i^{ZX}=(R'\Sigma_i^{XX})$, we have

$$\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}=(R'\Sigma_i^{XX})(\Sigma_i^{XX})^{-1}(\Sigma_i^{XX}R)=R'\Sigma_i^{XX}R=\Sigma_i^{ZZ}.$$

Thus,

$$\begin{aligned}
&\sup_{k\in K,k\leq k_0}\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
=\ &\sup_{k\in K,k\leq k_0}\left\|(\frac{k}{T}\Sigma_i^{ZZ})^{-1}\right\| \leq \frac{\sqrt{q}}{\tau_0-\eta}\left\|(\Sigma_i^{ZZ})^{-1}\right\|_{op} \leq \frac{\sqrt{q}}{\tau_0-\eta}\frac{1}{\rho_{\min}(\Sigma_i^{ZZ})} \\
\leq\ &\frac{\sqrt{q}}{\tau_0-\eta}\frac{1}{\rho_{\min}(\Sigma_i^{XX})} \leq \frac{\sqrt{q}}{\tau_0-\eta}\frac{1}{\rho_1},
\end{aligned}$$

where the second inequality follows from $\rho_{\min}(\Sigma_i^{ZZ})\geqslant\rho_{\min}(\Sigma_i^{XX})$.

(16) Noting that $Z_{2i}'M_iZ_{2i}=Z_{1i}'M_iZ_{1i}$, the proof is the same as part (14), except for

$$\left\|(\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1}\right\| \leq \frac{\sqrt{q}k}{\rho_{\min}(Z_{1i}'M_iZ_{1i})} = \sqrt{q}(\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|+\frac{k}{T-k}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|).$$

(17) The proof is similar to part (7).

(18) The proof is similar to part (11). ■

Lemma 4 *Under Assumptions 1-3 and assume* $\max\limits_{1\leq i\leq N}\frac{\delta_i'\delta_i}{\lambda_N}=O(\frac{1}{N})$*, there exists* $\alpha>0$ *such that for any* $\epsilon>0$*, there exist* $N^*>0$ *and* $T^*>0$ *such that for* $N^*>N$ *and* $T>T^*$*,* $P(\inf\limits_{k<k_0}\sum_{i=1}^N G_i(k)\geq\alpha\lambda_N)>1-\epsilon$*.*

Proof. We will prove by two steps.

Step 1: There exists $\alpha_1 > 0$ such that for any $\epsilon > 0$, there exist $C > 0$ and $T^* > 0$ such that for $T > T^*$, $P(\inf_{k<k_0-C} \sum_{i=1}^N G_i(k) \geq \alpha_1 \lambda_N) > 1 - \epsilon$.

Step 2: There exists $\alpha_2 > 0$ such that for any given $C > 0$ and $\epsilon > 0$, there exist $N^* > 0$ and $T^* > 0$ such that for $N > N^*$ and $T > T^*$, $P(\inf_{k_0-C\leq k<k_0} \sum_{i=1}^N G_i(k) \geq \alpha_2 \lambda_N) > 1 - \epsilon$.

Based on Step 1 and Step 2 and take $\alpha = \min\{\alpha_1, \alpha_2\}$, we have for any $\epsilon > 0$, there exist $N^* > 0$ and $T^* > 0$ such that for $N > N^*$ and $T > T^*$, $P(\inf_{k<k_0} \sum_{i=1}^N G_i(k) < \alpha\lambda_N) \leq P(\inf_{k<k_0-C} \sum_{i=1}^N G_i(k) < \alpha\lambda_N) + P(\inf_{k_0-C\leq k<k_0} \sum_{i=1}^N G_i(k) < \alpha\lambda_N) \leq 2\epsilon$, thus $P(\inf_{k<k_0} \sum_{i=1}^N G_i(k) < \alpha\lambda_N) > 1 - 2\epsilon$.

Proof of Step 1: Define $A_i(k) = \frac{(X_{\Delta i}'X_{\Delta i})(X_{2i}'X_{2i})^{-1}(X_{0i}'X_{0i})}{|k_0-k|}$, then by Lemma 2 we have

$$\begin{aligned}
&\inf_{k<k_0-C} \sum_{i=1}^N G_i(k) \\
= &\inf_{k<k_0-C} \sum_{i=1}^N \frac{\delta_i'[(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})]\delta_i}{|k_0-k|} \\
\geq &\inf_{k<k_0-C} \sum_{i=1}^N \frac{\delta_i'R'[(X_{\Delta i}'X_{\Delta i})(X_{2i}'X_{2i})^{-1}(X_{0i}'X_{0i})]R\delta_i}{|k_0-k|} \\
= &\inf_{k<k_0-C} \sum_{i=1}^N \delta_i'R'A_i(k)R\delta_i \\
\geq &\inf_{k<k_0-C} \sum_{i=1}^N \delta_i'R'(\frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i - \sup_{k<k_0-C} \left|\sum_{i=1}^N \delta_i'R'(A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right| \\
\geq &\sum_{i=1}^N \delta_i'R'\Sigma_i^{XX}R\delta_i - \sup_{k<k_0-C} \left|\sum_{i=1}^N \delta_i'R'(A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right|.
\end{aligned}$$

By Assumption 2,

$$\sum_{i=1}^N \delta_i'R'\Sigma_i^{XX}R\delta_i \geq \sum_{i=1}^N \rho_{\min}(\Sigma_i^{XX})\delta_i'\delta_i \geq \rho\lambda_N,$$

thus it suffices to show for any $\epsilon > 0$ and $\eta > 0$, there exists $C > 0$ and $T^* > 0$ such that for $T > T^*$, $P(\sup_{k<k_0-C} \left|\sum_{i=1}^N \delta_i'R'(A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right| > \eta\lambda_N) < \epsilon$. With assumption $\lim_{N\to\infty} \max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N} < \infty$,

$$\begin{aligned}
&\sup_{k<k_0-C} \left|\sum_{i=1}^N \delta_i'R'(A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right| \\
\leq &\lambda_N \sup_{k<k_0-C} \frac{1}{N} \sum_{i=1}^N \frac{N\delta_i'\delta_i}{\lambda_N} \left\|A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX}\right\| \\
\leq &\lambda_N (\lim_{N\to\infty} \max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N}) \sup_{k<k_0-C} \frac{1}{N} \sum_{i=1}^N \left\|A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX}\right\| \\
\leq &\lambda_N (\lim_{N\to\infty} \max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N}) \frac{1}{N} \sum_{i=1}^N \sup_{k<k_0-C} \left\|A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX}\right\|,
\end{aligned}$$

thus by Markov inequality it suffices to show for any $\epsilon > 0$, there exist $C < \infty$ and $T^* > 0$ such that for $T > T^*$, $E(\sup_{k<k_0-C} \left\| A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX} \right\|) < \epsilon$ for all i. For each i and any given $C > 0$,

$$
\begin{aligned}
& \sup_{k<k_0-C} \left\| A_i(k) - \frac{T-k_0}{T-k}\Sigma_i^{XX} \right\| \\
= & \sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|}\frac{T-k_0}{T-k}(\frac{X'_{2i}X_{2i}}{T-k})^{-1}(\frac{X'_{0i}X_{0i}}{T-k_0}) - \frac{T-k_0}{T-k}\Sigma_i^{XX} \right\| \\
\leq & \sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|}(\frac{X'_{2i}X_{2i}}{T-k})^{-1}(\frac{X'_{0i}X_{0i}}{T-k_0}) - \Sigma_i^{XX} \right\| \\
= & \sup_{k<k_0-C} \left\| \begin{array}{c} (\frac{X'_{\Delta i}X_{\Delta i}}{|k_0-k|} - \Sigma_i^{XX} + \Sigma_i^{XX})[(\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \\ +(\Sigma_i^{XX})^{-1}](\frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} + \Sigma_i^{XX}) - \Sigma_i^{XX} \end{array} \right\| \\
\leq & I + II + III + IV + V + VI + VII.
\end{aligned}
$$

Consider each term one by one. By part (2) and part (3) of Lemma 3, as $C \to \infty$ and $T \to \infty$, for all i

$$
\begin{aligned}
E(I) &= E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\| \sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\| \left\|\Sigma_i^{XX}\right\|) \\
&\leq [E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|)^2 E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|)^2]^{\frac{1}{2}} \left\|\Sigma_i^{XX}\right\| \\
&< \epsilon.
\end{aligned}
$$

By part (2), part (3) and part (4) of Lemma 3, as $C \to \infty$ and $T \to \infty$, for all i,

$$
\begin{aligned}
& E(II) \\
= & E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\| \sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\| \left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|) \\
\leq & [E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|)^2]^{\frac{1}{2}} [E(\\
& \sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|)^4 E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|)^4]^{\frac{1}{4}} \\
< & \epsilon.
\end{aligned}
$$

By part (2) of Lemma 3, as $C \to \infty$, for all i,

$$E(III) = E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|) \left\|(\Sigma_i^{XX})^{-1}\right\| \left\|\Sigma_i^{XX}\right\| < \epsilon.$$

By part (2) and part (4) of Lemma 3, as $C \to \infty$ and $T \to \infty$, for all i,

$$\begin{aligned} E(IV) &= E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\| \left\|(\Sigma_i^{XX})^{-1}\right\| \left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|) \\ &\leq [E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|)^2 E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|)^2]^{\frac{1}{2}} \left\|(\Sigma_i^{XX})^{-1}\right\| \\ &< \epsilon. \end{aligned}$$

By part (3) of Lemma 3, as $T \to \infty$, for all i,

$$E(V) = E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|) \left\|\Sigma_i^{XX}\right\|^2 < \epsilon.$$

By part (3) and part (4) of Lemma 3, as $T \to \infty$, for all i,

$$\begin{aligned} E(VI) &= E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\| \left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\| \left\|\Sigma_i^{XX}\right\|) \\ &\leq [E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|)^2 E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|)^2]^{\frac{1}{2}} \left\|\Sigma_i^{XX}\right\| \\ &< \epsilon. \end{aligned}$$

By part (4) of Lemma 3, as $T \to \infty$, for all i,

$$E(VII) = E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|) < \epsilon.$$

Proof of Step 2: There exists $\alpha_2 > 0$ such that for any given $C > 0$ and $\epsilon > 0$, there exist $N^* > 0$ and $T^* > 0$ such that for $N > N^*$ and $T > T^*$, $P(\inf_{k_0-C \leq k < k_0} \sum_{i=1}^N G_i(k) \geq \alpha_2 \lambda_N) > 1 - \epsilon$.

$$\begin{aligned} \sum_{i=1}^N G_i(k) &= \frac{\sum_{i=1}^N \delta'_i[(Z'_{0i}M_iZ_{0i}) - (Z'_{0i}M_iZ_{2i})(Z'_{2i}M_iZ_{2i})^{-1}(Z'_{2i}M_iZ_{0i})]\delta_i}{|k_0-k|} \\ &= \frac{\sum_{i=1}^N \delta'_i(Z'_{\Delta i}M_iZ_{\Delta i})\delta_i - \sum_{i=1}^N \delta'_i(Z'_{\Delta i}M_iZ_{2i})(Z'_{2i}M_iZ_{2i})^{-1}(Z'_{2i}M_iZ_{\Delta i})\delta_i}{|k_0-k|} \\ &= \frac{\sum_{i=1}^N \delta'_iZ'_{\Delta i}Z_{\Delta i}\delta_i}{|k_0-k|} - \frac{\sum_{i=1}^N \delta'_iZ'_{\Delta i}X_i(X'_iX_i)^{-1}X'_iZ_{\Delta i}\delta_i}{|k_0-k|} \\ &\quad - \frac{\sum_{i=1}^N \delta'_i(Z'_{\Delta i}M_iZ_{2i})(Z'_{2i}M_iZ_{2i})^{-1}(Z'_{2i}M_iZ_{\Delta i})\delta_i}{|k_0-k|}, \end{aligned}$$

thus

$$\begin{aligned} \inf_{k_0-C \leq k < k_0} \sum_{i=1}^N G_i(k) &\geq \inf_{k_0-C \leq k < k_0} \frac{\sum_{i=1}^N \delta'_iZ'_{\Delta i}Z_{\Delta i}\delta_i}{|k_0-k|} \\ &\quad - \sup_{k_0-C \leq k < k_0} \frac{\sum_{i=1}^N \delta'_iZ'_{\Delta i}X_i(X'_iX_i)^{-1}X'_iZ_{\Delta i}\delta_i}{|k_0-k|} \\ &\quad - \sup_{k_0-C \leq k < k_0} \frac{\sum_{i=1}^N \delta'_iZ'_{\Delta i}M_iZ_{2i})(Z'_{2i}M_iZ_{2i})^{-1}(Z'_{2i}M_iZ_{\Delta i})\delta_i}{|k_0-k|}. \end{aligned}$$

Consider the first term. By part (5) of Assumption 2, we have for each t, $\frac{\sum_{i=1}^{N}\delta_i' z_{it} z_{it}' \delta_i}{\lambda_N} \xrightarrow{p} \xi$ as $N \to \infty$. For a given C, $\{\frac{\frac{1}{\lambda_N}\sum_{i=1}^{N}\delta_i' Z_{\Delta i}' Z_{\Delta i}\delta_i}{|k_0-k|}, k_0-C \leq k < k_0\}$ is a finite dimensional random vector, hence $\{\frac{\frac{1}{\lambda_N}\sum_{i=1}^{N}\delta_i' Z_{\Delta i}' Z_{\Delta i}\delta_i}{|k_0-k|}, k_0-C \leq k < k_0\} \xrightarrow{p} (\xi, ..., \xi)'$ as $N \to \infty$. It follows by the continuous mapping theorem that $\inf_{k_0-C\leq k<k_0} \frac{\frac{1}{\lambda_N}\sum_{i=1}^{N}\delta_i' Z_{\Delta i}' Z_{\Delta i}\delta_i}{|k_0-k|} \xrightarrow{p} \xi$ as $N \to \infty$. Next consider the last two terms.

$$
\begin{aligned}
& E(\sup_{k_0-C\leq k<k_0} \frac{\sum_{i=1}^{N}\delta_i' Z_{\Delta i}' X_i (X_i'X_i)^{-1} X_i' Z_{\Delta i}\delta_i}{|k_0-k|}) \\
\leq\ & E(\frac{|k_0-k|}{T} \sup_{k_0-C\leq k<k_0} \sum\nolimits_{i=1}^{N} \left\| \frac{Z_{\Delta i}' X_i}{|k_0-k|} \right\|^2 \left\| (\frac{X_i'X_i}{T})^{-1} \right\| \|\delta_i\|^2) \\
\leq\ & \frac{C\lambda_N}{T} (\lim_{N\to\infty} \max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N}) \frac{1}{N} \sum\nolimits_{i=1}^{N} E(\sup_{k_0-C\leq k<k_0} \left\| \frac{Z_{\Delta i}' X_i}{|k_0-k|} \right\|^2 \left\| (\frac{X_i'X_i}{T})^{-1} \right\|) \\
=\ & O(\frac{\lambda_N}{T}),
\end{aligned}
$$

where the last equality follows from part (2) of Assumption 2 and part (5) of Lemma 3. Similarly,

$$
\begin{aligned}
& E(\sup_{k_0-C\leq k<k_0} \frac{\sum_{i=1}^{N}\delta_i' Z_{\Delta i}' M_i Z_{2i})(Z_{2i}' M_i Z_{2i})^{-1}(Z_{2i}' M_i Z_{\Delta i})\delta_i}{|k_0-k|}) \\
\leq\ & \frac{C\lambda_N}{T-k_0} (\lim_{N\to\infty} \max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N}) \frac{1}{N} \sum\nolimits_{i=1}^{N} E(\sup_{k_0-C\leq k<k_0} \left\| \frac{Z_{\Delta i}' M_i Z_{2i}}{|k_0-k|} \right\|^2 \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} \right\|) \\
=\ & O(\frac{\lambda_N}{T}),
\end{aligned}
$$

where the last equality follows from parts (11) and (14) of Lemma 3. Taken together, the proof is finished. ■

Lemma 5 *Under Assumptions 1-3,*

(1) $\sup_{k\in K(k_0),k<k_0} |A| = \sup_{k\in K(k_0),k<k_0} \left| 2sgn(k_0-k)\frac{\sum_{i=1}^{N}\delta_i' Z_{\Delta i}' e_i}{|k-k_0|} \right| = O_p(\sqrt{\lambda_N})$ *as* $(N,T) \to \infty$*;*

(2) $\sup_{k\in K(k_0),k<k_0} |B| = \sup_{k\in K(k_0),k<k_0} \left| -2sgn(k_0-k)\frac{\sum_{i=1}^{N}\delta_i' Z_{\Delta i}' X_i (X_i'X_i)^{-1} X_i' e_i}{|k-k_0|} \right| = O_p(\frac{\sqrt{N\lambda_N}}{\sqrt{T}})$ *as* $(N,T) \to \infty$*;*

(3) $\sup_{k\in K(k_0),k<k_0} |C| = \sup_{k\in K(k_0),k<k_0} \left| -2sgn(k_0-k)\frac{\sum_{i=1}^{N}\delta_i' (Z_{\Delta i}' M_i Z_{2i})(Z_{2i}' M_i Z_{2i})^{-1}(Z_{2i}' M_i e_i)}{|k-k_0|} \right| = O_p(\frac{\sqrt{N\lambda_N}}{\sqrt{T}})$ *as* $(N,T) \to \infty$*;*

(4) $\sup_{k\in K(k_0),k<k_0} |D| = \sup_{k\in K(k_0),k<k_0} \left| \frac{\sum_{i=1}^{N} e_i' M_i Z_{\Delta i}(Z_{2i}' M_i Z_{2i})^{-1} Z_{\Delta i}' M_i e_i}{|k-k_0|} \right| = O_p(\frac{N\log T}{T})$ *as* $(N,T) \to \infty$*;*

(5) $\sup_{k\in K(k_0),k<k_0} |E| = \sup_{k\in K(k_0),k<k_0} \left| 2sgn(k_0-k)\frac{\sum_{i=1}^{N} e_i' M_i Z_{0i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{\Delta i}'M_ie_i}{|k-k_0|}\right| = O_p(\frac{N}{\sqrt{T}})$ *as* $(N,T)\to\infty$*;*

(6) $\sup_{k\in K(k_0),k<k_0} |F| = \sup_{k\in K(k_0),k<k_0} \left| \frac{\sum_{i=1}^{N} e_i' M_i Z_{0i}[(Z_{2i}'M_iZ_{2i})^{-1}-(Z_{0i}'M_iZ_{0i})^{-1}]Z_{0i}'M_ie_i}{|k-k_0|}\right| = O_p(\frac{N}{\sqrt{T}})$ *as* $(N,T)\to\infty$*.*

Proof. (1) Under part (4) of Assumption 3, there exists $M>0$ such that

$$E(\sup_{k\leq l<k_0}\left|\sum_{t=l+1}^{k_0}\eta_{Nt}\right|^2)\leq 4E(\left|\sum_{t=k+1}^{k_0}\eta_{Nt}\right|^2)\leq (k_0-k)M$$

for all N and $1\leq k<k_0$. Using Lemma 1 and take $r=2$, $\alpha_{k_0-k}=M$, $\beta_{k_0-k}=k_0-k$ for $k_0-k=1,...,T\eta$, we have

$$\begin{aligned}E(\sup_{k\in K(k_0),k<k_0}|A|)^2 &= 4\lambda_N E(\sup_{k\in K(k_0),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\eta_{Nt}\right|^2)\\ &\leq 16\lambda_N M\sum_{k=T(\tau_0-\eta)}^{k_0-1}\frac{1}{(k_0-k)^2}\leq 32\lambda_N M.\end{aligned}$$

(2)

$$\begin{aligned}&\sup_{k\in K(k_0),k<k_0}|B|\\ =\ & \sup_{k\in K(k_0),k<k_0}\left|\frac{2}{|k-k_0|}\sum_{i=1}^{N}\delta_i' Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'e_i\right|\\ \leq\ & \frac{2\sqrt{\lambda_N}}{\sqrt{T}}\sup_{k\in K(k_0),k<k_0}\sum_{i=1}^{N}\left\|\frac{\delta_i}{\sqrt{\lambda_N}}\right\|\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|\left\|(\frac{X_i'X_i}{T})^{-1}\right\|\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|\\ \leq\ & \frac{2\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N}(\lim_{N\to\infty}\max_{1\leq i\leq N}\left\|\frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}}\right\|)\frac{1}{N}\sum_{i=1}^{N}(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|)\left\|(\frac{X_i'X_i}{T})^{-1}\right\|\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|.\end{aligned}$$

Using parts (1), (5) and (6) of Lemma 3,

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}E((\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|)\left\|(\frac{X_i'X_i}{T})^{-1}\right\|\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|)\\ \leq\ & \frac{1}{N}\sum_{i=1}^{N}[E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|^4)E(\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4)]^{\frac{1}{4}}[E(\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|^2)]^{\frac{1}{2}}\\ =\ & O(1),\end{aligned}$$

hence by Markov inequality $\sup_{k\in K(k_0),k<k_0}|B| = O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N})$ as $(N,T)\to\infty$.

(3)

$$
\begin{aligned}
& \sup_{k\in K(k_0),k<k_0} |C| \\
= & \sup_{k\in K(k_0),k<k_0} \left| \frac{2}{\sqrt{T-k}} \sum\nolimits_{i=1}^{N} \delta_i' (\frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|})(\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} (\frac{Z_{2i}' M_i e_i}{\sqrt{T-k}}) \right| \\
\leq & \frac{2\sqrt{\lambda_N}}{\sqrt{T\tau_0}} \sum\nolimits_{i=1}^{N} \left\| \frac{\delta_i}{\sqrt{\lambda_N}} \right\| \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|} \right\| \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{2i}' M_i e_i}{\sqrt{T-k}} \right\| \\
& \sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\| \\
& + \frac{2\sqrt{\lambda_N}}{\sqrt{T\eta}} \sum\nolimits_{i=1}^{N} \left\| \frac{\delta_i}{\sqrt{\lambda_N}} \right\| \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|} \right\| \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{2i}' M_i e_i}{\sqrt{T-k}} \right\| \\
& \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\| \\
\leq & \frac{2\sqrt{N\lambda_N}}{\sqrt{T\tau_0}} (\lim_{N\to\infty} \max_{1\leq i\leq N} \left\| \frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}} \right\|) [\frac{1}{N} \sum\nolimits_{i=1}^{N} \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|} \right\| \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{2i}' M_i e_i}{\sqrt{T-k}} \right\| \\
& \sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\|] \\
& + \frac{2\sqrt{N\lambda_N}}{\sqrt{T\tau_0}} (\lim_{N\to\infty} \max_{1\leq i\leq N} \left\| \frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}} \right\|) [\frac{1}{N} \sum\nolimits_{i=1}^{N} \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|} \right\| \\
& \sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{2i}' M_i e_i}{\sqrt{T-k}} \right\| \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\|] \\
= & \frac{2\sqrt{N\lambda_N}}{\sqrt{T\tau_0}} (\lim_{N\to\infty} \max_{1\leq i\leq N} \left\| \frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}} \right\|)(C_1 + C_2)
\end{aligned}
$$

Using parts (10), (11) and (12) of Lemma 3,

$$
\begin{aligned}
E(C_1) \leq & \frac{1}{N} \sum\nolimits_{i=1}^{N} [E(\sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|} \right\|^4) E(\sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{2i}' M_i e_i}{\sqrt{T-k}} \right\|^4)]^{\frac{1}{4}} \\
& [E(\sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\|^2)]^{\frac{1}{2}} \\
= & O(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

and using parts (10), (11) and (15) of Lemma 3,

$$
\begin{aligned}
E(C_2) \leq & \frac{1}{N} \sum\nolimits_{i=1}^{N} [E(\sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{\Delta i}' M_i Z_{2i}}{|k-k_0|} \right\|^2) E(\sup_{k\in K(k_0),k<k_0} \left\| \frac{Z_{2i}' M_i e_i}{\sqrt{T-k}} \right\|^2)]^{\frac{1}{2}} \\
& \sup_{k\in K(k_0),k<k_0} \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\| \\
= & O(1),
\end{aligned}
$$

thus by Markov inequality $\sup_{k\in K(k_0),k<k_0} |C| = O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N})$ as $(N,T)\to\infty$.

(4)

$$\begin{aligned}
& \sup_{k\in K(k_0),k<k_0} |D| \\
= & \sup_{k\in K(k_0),k<k_0} \left| \frac{N}{T-k}\frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{e_i' M_i Z_{\Delta i}}{\sqrt{|k-k_0|}} (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} \frac{Z_{\Delta i}' M_i e_i}{\sqrt{|k-k_0|}} \right| \\
\leq & \frac{N}{T\tau_0}(\frac{1}{N}\sum\nolimits_{i=1}^{N} \sup_{k\in K(k_0),k<k_0} \left\| \frac{e_i' M_i Z_{\Delta i}}{\sqrt{|k-k_0|}} \right\|^2 \sup_{K(k_0)} \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \right\| \\
& + \frac{1}{N}\sum\nolimits_{i=1}^{N} \sup_{k\in K(k_0),k<k_0} \left\| \frac{e_i' M_i Z_{\Delta i}}{\sqrt{|k-k_0|}} \right\|^2 \\
& \sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \right\|) \\
= & \frac{N}{T\tau_0}(D_1 + D_2),
\end{aligned}$$

Using parts (7) and (15) of Lemma 3,

$$\begin{aligned}
E(D_1) = & \frac{1}{N}\sum\nolimits_{i=1}^{N} E(\sup_{k\in K(k_0),k<k_0} \left\| \frac{e_i' M_i Z_{\Delta i}}{\sqrt{|k-k_0|}} \right\|)^2 \sup_{k\in K(k_0),k<k_0} \\
& \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \right\| \\
= & O(\log T),
\end{aligned}$$

and using parts (7) and (12) of Lemma 3,

$$\begin{aligned}
E(D_2) \leq & \frac{1}{N}\sum\nolimits_{i=1}^{N} [E(\sup_{k\in K(k_0),k<k_0} \left\| \frac{e_i' M_i Z_{\Delta i}}{\sqrt{|k-k_0|}} \right\|^4) \\
& E(\sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \right\|^2)]^{\frac{1}{2}} \\
= & O(\sqrt{\frac{\log T}{T}}),
\end{aligned}$$

thus by Markov inequality $\sup_{k\in K(k_0),k<k_0} |D| = O_p(\frac{N\log T}{T})$ as $(N,T)\to\infty$.

(5)

$$
\begin{aligned}
& \sup_{k\in K(k_0),k<k_0} |E| \\
\leq & \sup_{k\in K(k_0),k<k_0} \left| \frac{2N\sqrt{T-k_0}}{T-k}\frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\frac{Z_{\Delta i}' M_i e_i}{|k-k_0|} \right| \\
& + \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{2N\sqrt{T-k_0}}{T-k}\frac{1}{N}\sum_{i=1}^{N}\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}[(\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} \\ -[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}]\frac{Z_{\Delta i}' M_i e_i}{|k-k_0|} \end{array} \right| \\
\leq & \frac{2}{\sqrt{1-\tau_0}}\frac{N}{\sqrt{T}}(\frac{1}{N}\sum\nolimits_{i=1}^{N}\left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\| \sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}' M_i e_i}{|k-k_0|}\right\| \\
& \sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|) \\
& + \frac{2}{\sqrt{1-\tau_0}}\frac{N}{\sqrt{T}}(\frac{1}{N}\sum\nolimits_{i=1}^{N}\left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\| \sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}' M_i e_i}{|k-k_0|}\right\| \\
& \sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|) \\
= & \frac{2}{\sqrt{1-\tau_0}}\frac{N}{\sqrt{T}}(E_1+E_2).
\end{aligned}
$$

Using parts (8), (9) and (15) of Lemma 3,

$$
\begin{aligned}
E(E_1) \leq & \frac{1}{N}\sum\nolimits_{i=1}^{N}[E(\left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\|^2)E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}' M_i e_i}{|k-k_0|}\right\|^2)]^{\frac{1}{2}} \\
& \sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
= & O(1),
\end{aligned}
$$

and using parts (8), (9) and (12) of Lemma 3,

$$
\begin{aligned}
E(E_2) \leq & \frac{1}{N}\sum\nolimits_{i=1}^{N}[E(\left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\|^4)E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}' M_i e_i}{|k-k_0|}\right\|^4)]^{\frac{1}{4}} \\
& [E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2)]^{\frac{1}{2}} \\
= & O(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

thus by Markov inequality $\sup_{k\in K(k_0),k<k_0} |E| = O_p(\frac{N}{\sqrt{T}})$ as $(N,T)\to\infty$.

(6)

$$
\begin{aligned}
& \sup_{k\in K(k_0),k<k_0} |F| \\
= & \sup_{k\in K(k_0),k<k_0} \left| \frac{1}{|k-k_0|} \sum_{i=1}^{N} e_i' M_i Z_{0i} [\frac{1}{T-k}(\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - \frac{1}{T-k_0}(\frac{Z_{0i}' M_i Z_{0i}}{T-k_0})^{-1}] Z_{0i}' M_i e_i \right| \\
\leq & \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{1}{|k-k_0|} \sum_{i=1}^{N} e_i' M_i Z_{0i} [\frac{1}{T-k}(\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} \\ -\frac{1}{T-k}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}] Z_{0i}' M_i e_i \end{array} \right| \\
& + \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{1}{|k-k_0|} \sum_{i=1}^{N} e_i' M_i Z_{0i} [\frac{1}{T-k}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}] Z_{0i}' M_i e_i \end{array} \right| \\
& + \left| \begin{array}{c} \frac{1}{|k-k_0|} \sum_{i=1}^{N} e_i' M_i Z_{0i} [(\frac{1}{T-k_0}(\frac{Z_{0i}' M_i Z_{0i}}{T-k_0})^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}] Z_{0i}' M_i e_i \end{array} \right| \\
= & \sup_{k\in K(k_0),k<k_0} |F_1| + \sup_{k\in K(k_0),k<k_0} |F_2| + |F_3| .
\end{aligned}
$$

$$
\begin{aligned}
& \sup_{k\in K(k_0),k<k_0} |F_1| \\
= & \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{1}{|k-k_0|} \frac{T-k_0}{T-k} \sum_{i=1}^{N} \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} [(\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} \\ -[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}] \frac{Z_{0i}' M_i e_i}{\sqrt{T-k_0}} \end{array} \right| \\
\leq & \sum_{i=1}^{N} \left\| \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} \right\|^2 \sup_{k\in K(k_0),k<k_0} \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \right\| .
\end{aligned}
$$

Using parts (9) and (12) of Lemma 3,

$$
\begin{aligned}
E(\sup_{k\in K(k_0),k<k_0} |F_1|) \leq & \sum_{i=1}^{N} [E(\left\| \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} \right\|^4) E(\sup_{k\in K(k_0),k<k_0} \\
& \left\| (\frac{Z_{2i}' M_i Z_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \right\|^2)]^{\frac{1}{2}} \\
= & O(\frac{N}{\sqrt{T}}),
\end{aligned}
$$

hence by Markov inequality $\sup_{k\in K(k_0),k<k_0} |F_1| = O_p(\frac{N}{\sqrt{T}})$ as $(N,T) \to \infty$.

$$
\begin{aligned}
& \sup_{k\in K(k_0),k<k_0} |F_2| \\
\leq & \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{1}{|k-k_0|} \sum_{i=1}^{N} e_i' M_i Z_{0i} \{\frac{1}{T-k}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\} Z_{0i}' M_i e_i \end{array} \right| \\
& + \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{1}{|k-k_0|} \sum_{i=1}^{N} e_i' M_i Z_{0i} \{\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\} Z_{0i}' M_i e_i \end{array} \right| \\
= & \sup_{k\in K(k_0),k<k_0} |F_{21}| + \sup_{k\in K(k_0),k<k_0} |F_{22}| .
\end{aligned}
$$

$$
\begin{aligned}
& E(\sup_{k\in K(k_0),k<k_0} |F_{21}|) \\
= & E(\sup_{k\in K(k_0),k<k_0} \left| \frac{1}{T-k} \sum_{i=1}^{N} \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \frac{Z_{0i}' M_i e_i}{\sqrt{T-k_0}} \right|) \\
\leq & \frac{N}{T\tau_0} [\frac{1}{N} \sum_{i=1}^{N} E(\sup_{k\in K(k_0),k<k_0} \left\| \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} \right\|^2) \sup_{k\in K(k_0),k<k_0} \\
& \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\|].
\end{aligned}
$$

Using parts (9) and (15) of Lemma 3 and Markov inequality, $\sup_{k\in K(k_0),k<k_0} |F_{21}| = O_p(\frac{N}{T})$ as $(N,T) \to \infty$.

$$
\begin{aligned}
& E(\sup_{k\in K(k_0),k<k_0} |F_{22}|) \\
\leq & \sum_{i=1}^{N} E(\sup_{k\in K(k_0),k<k_0} \left\| \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} \right\|^2) \sup_{k\in K(k_0),k<k_0} \\
& \frac{1}{|k-k_0|} \left\| \begin{array}{c} [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \\ -[\Sigma_i^{ZZ} - \frac{T-k_0}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \end{array} \right\| .
\end{aligned}
$$

Using part (9) of Lemma 3, the first term is $O(1)$. Noting that $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$, the second term is not larger than

$$
\begin{aligned}
& \sup_{k\in K(k_0),k<k_0} \left\| [\Sigma_i^{ZZ} - \frac{T-k}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\| \left\| \frac{1}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ} \right\| \\
& \left\| [\Sigma_i^{ZZ} - \frac{T-k_0}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\| .
\end{aligned}
$$

Part (15) of Lemma 3 implies this term is $O(\frac{1}{T})$, thus by Markov inequality $\sup_{k\in K(k_0),k<k_0} |F_{22}| = O_p(\frac{N}{T})$ as $(N,T) \to \infty$.

$$
|F_3| \leq \sum_{i=1}^{N} \left\| \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} \right\|^2 \left\| (\frac{Z_{0i}' M_i Z_{0i}}{T-k_0})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k_0}{T} \Sigma_i^{ZX} (\Sigma_i^{XX})^{-1} \Sigma_i^{XZ}]^{-1} \right\|)
$$

Using parts (9) and (13) of Lemma 3,

$$
\begin{aligned}
& E(F_3) \leq \sum_{i=1}^{N} [E(\left\| \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}} \right\|^4) E(\left\| (\frac{Z_{0i}' M_i Z_{0i}}{T-k_0})^{-1} - (\Sigma_i^{ZX} \Sigma_i^{XX} \Sigma_i^{XZ})^{-1} \right\|^2)]^{\frac{1}{2}} \\
= & O(\frac{N}{\sqrt{T}}),
\end{aligned}
$$

thus by Markov inequality $F_3 = O_p(\frac{N}{\sqrt{T}})$ as $(N,T) \to \infty$. Taken together, $\sup_{k\in K(k_0),k<k_0} |F| = O_p(\frac{N}{\sqrt{T}}) + O_p(\frac{N}{T}) + O_p(\frac{N}{\sqrt{T}}) = O_p(\frac{N}{\sqrt{T}})$ as $(N,T) \to \infty$. ■

Lemma 6 *Under Assumptions 5-14, given* $|k - k_0| \leq C$*, there exists* $M > 0$ *such that*

(1) $E(\left\| \frac{1}{T} \sum_{t=1}^{T} (\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \leq \frac{1}{\delta_{NT}^2} M$,

(2) $E(\left\| \frac{1}{T} \sum_{t=1}^{T} (\tilde{F}_t - H'F_t^0) e_{it} \right\|) \leq (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2}) M$ *for each* i,

(3) $E(\left\| \frac{1}{T} \sum_{t=1}^{T} H'F_t^0 (\tilde{F}_t - H'F_t^0)' \right\|) \leq (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2}) M$,

(4) $E(\sup_{k \in K(k_0), k < k_0} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} (\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \leq M$,

(5) $E(\sup_{k \in K(k_0), k < k_0} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} (\tilde{F}_t - H'F_t^0) e_{it} \right\|) \leq M$ *for each* i,

(6) $E(\sup_{k \in K(k_0), k < k_0} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} H'F_t^0 (\tilde{F}_t - H'F_t^0)' \right\|) \leq M$,

(7) $E(\sup_{k \in K(k_0), k < k_0} \left\| \frac{1}{T} \sum_{t=k+1}^{k_0} (\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \leq \frac{1}{\delta_{NT}^2} M$,

(8) $E(\sup_{k \in K(k_0), k < k_0} \left\| \frac{1}{T} \sum_{t=k+1}^{k_0} (\tilde{F}_t - H'F_t^0) e_{it} \right\|) \leq (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2}) M$ *for each* i,

(9) $E(\sup_{k \in K(k_0), k < k_0} \left\| \frac{1}{T} \sum_{t=k+1}^{k_0} H'F_t^0 (\tilde{F}_t - H'F_t^0)' \right\|) \leq (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2}) M$,

(10) $E(\sup_{k \in K(k_0), k \leq k_0} \left\| \frac{1}{T - k} \sum_{t=k+1}^{T} (\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \leq \frac{1}{\delta_{NT}^2} M$,

(11) $E(\sup_{k \in K(k_0), k \leq k_0} \left\| \frac{1}{T - k} \sum_{t=k+1}^{T} (\tilde{F}_t - H'F_t^0) e_{it} \right\|) \leq (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2}) M$ *for each* i,

(12) $E(\sup_{k \in K(k_0), k \leq k_0} \left\| \frac{1}{T - k} \sum_{t=k+1}^{T} H'F_t^0 (\tilde{F}_t - H'F_t^0)' \right\|) \leq (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2}) M$.

Proof. We will show that terms in parentheses have the indicated stochastic order. Given our assumptions on the factor process and the error process and using Holder's inequality, $E \|fg\| \leq (E \|f\|^2)^{\frac{1}{2}} (E \|g\|^2)^{\frac{1}{2}}$ repeatedly, it is easy to show their expectation have the same order.

First note that $[\frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i \hat{\beta}_i(k) - Z_{2i} \hat{\delta}_i(k))(Y_i - X_i \hat{\beta}_i(k) - Z_{2i} \hat{\delta}_i(k))'] \tilde{F} = \tilde{F} V_{NT}$, where V_{NT} is a diagonal matrix consists of the r largest eigenvalues of the matrix in the bracket. Define

$$u_i = X_i(\beta_i - \hat{\beta}_i(k)) + Z_{0i}(\delta_i - \hat{\delta}_i(k)) - (Z_{2i} - Z_{0i}) \hat{\delta}_i(k),$$

then $Y_i - X_i\hat{\beta}_i(k) - Z_{2i}\hat{\delta}_i(k) = u_i + F^0\lambda_i + e_i$. Expanding terms, we have

$$\begin{aligned}
& \tilde{F}V_{NT} - \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_i\lambda_i' F^{0\prime}\tilde{F} \\
= & \frac{1}{NT}\sum_{i=1}^{N} u_i u_i'\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} u_i\lambda_i' F^{0\prime}\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} u_i e_i'\tilde{F} \\
& + \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_i u_i'\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} e_i u_i'\tilde{F} \\
& + \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_i e_i'\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} e_i\lambda_i' F^{0\prime}\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} e_i e_i'\tilde{F} \\
= & I_1 + ... + I_8. \qquad (14)
\end{aligned}$$

Define $H = \frac{1}{NT}\sum_{i=1}^{N}\lambda_i\lambda_i' F^{0\prime}\tilde{F}V_{NT}^{-1}$, then $(\tilde{F} - F^0 H)V_{NT} = I_1 + ... + I_8$.

Parts (1)-(3) correspond to part (ii) of Proposition A.1, part (i) of Lemma A.4 and part (i) of Lemma A.3 respectively in Bai (2009), and can be proved in a similar manner. A key step is to calculate $\left\|\frac{1}{\sqrt{T}}u_i\right\|$. In Bai (2009), $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\left\|\hat{\beta} - \beta\right\|)$ while in the current case $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\left\|\hat{\beta}_i(k) - \beta_i\right\|) + O_p(\left\|\hat{\delta}_i(k) - \delta_i\right\|) + \frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})\hat{\delta}_i(k)\right\|$. If $k = k_0$, Song (2013) shows that $\beta_i - \hat{\beta}_i$ and $\delta_i - \hat{\delta}_i$ are $O_p(\frac{1}{\sqrt{T}})$. It can be verified that this result still holds for $|k - k_0| \le C$. Moreover, given our assumptions on the regressors and factors, this $O_p(\frac{1}{\sqrt{T}})$ is uniform over i. For the last term,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})\hat{\delta}_i(k)\right\| &= \frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})(\hat{\delta}_i(k) - \delta_i)\right\| + \frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})\delta_i\right\| \\
&= O_p(\frac{1}{\sqrt{T}}) + \frac{1}{\sqrt{T}}O_p(1) = O_p(\frac{1}{\sqrt{T}}),
\end{aligned}$$

where the second equality follows from $E\left\|(Z_{2i} - Z_{0i})\delta_i\right\|^2 = |k - k_0|\, E(\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\|z_{it}\|^2) = O(1)$ for $|k - k_0| \le C$. Thus, $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\frac{1}{\sqrt{T}})$.

Next consider parts (4)-(9). Each term in parts (4)-(9) can be decomposed into eight terms according to (14). The proof of the last three terms can be found in the existing literature. For part (4), the last three terms together is $O_p(\frac{1}{\delta_{NT}^2})$, see part (5) of Lemma 5 of Baltagi et al. (2015b). For part (5), the last three terms together is $O_p(\frac{1}{\delta_{NT}})$, see part (4) of Lemma 5 of Baltagi et al. (2015b), replacing F_t^0 by e_{it} does not change the result. For part (6), the last three terms is $O_p(\frac{1}{\delta_{NT}})$, see part (4) of Lemma 5 of Baltagi et al. (2015b). For part (7), the last three terms together is $O_p(\frac{1}{\delta_{NT}^2})$, see part (5) of Lemma 5 of Baltagi et al. (2015b), which is a stronger result. For part (8), the last three terms is $O_p(\frac{1}{\delta_{NT}^2})$, see Lemma 3 of Han and Inoue (2014), replacing $\sum_{t=1}^{\pi T}$ by $\sum_{t=k+1}^{k_0}$ and F_t^0 by e_{it} does not change the result. For part (9), the last three terms is $O_p(\frac{1}{\delta_{NT}^2})$, see Lemma 3 of Han and

Inoue (2014), replacing $\sum_{t=1}^{\pi T}$ by $\sum_{t=k+1}^{k_0}$ does not change the result. The assumptions in Baltagi et al. (2015b) and Han and Inoue (2014) can be verified given Assumptions 5–14. For the first five terms, a key result is $\sup_{k\in K(k_0),k<k_0} \left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} u_{it}^2\right\| = O_p(1)$ for parts (4)-(6) and $\sup_{k\in K(k_0),k<k_0} \left\|\frac{1}{T}\sum_{t=k+1}^{k_0} u_{it}^2\right\| = O_p(\frac{1}{T})$ for parts (7)-(9), which follows directly from $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\frac{1}{\sqrt{T}})$. Based on this, it is easy to see part (4) is $O_p(1) + O_p(\frac{1}{\delta_{NT}^2}) = O_p(1)$, parts (5) and (6) are both $O_p(1)+O_p(\frac{1}{\delta_{NT}}) = O_p(1)$, part (7) is $O_p(\frac{1}{T})+O_p(\frac{1}{\delta_{NT}^2}) = O_p(\frac{1}{\delta_{NT}^2})$, parts (8) and (9) are both $O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\delta_{NT}^2})$.

(10)-(12) can be proved following the same procedure as (7)-(9). ■

Lemma 7 *Under Assumptions 5-14, given $|k-k_0| \leq C$, there exists $M>0$ such that*

(1) $E(\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t e_{it}\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each i,*

(2) $E(\left\|\frac{1}{T}\sum_{t=1}^{T}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each i,*

(3) $E(\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M,$

(4) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\tilde{F}_t e_{it}\right\|) \leq M$ *for each i,*

(5) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq M$ *for each i,*

(6) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq M,$

(7) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{T}\sum_{t=k+1}^{k_0}\tilde{F}_t e_{it}\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each i,*

(8) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{T}\sum_{t=k+1}^{k_0}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each i,*

(9) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{T}\sum_{t=k+1}^{k_0}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M,$

(10) $E(\sup_{k\in K(k_0),k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\tilde{F}_t e_{it}\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each i,*

(11) $E(\sup_{k\in K(k_0),k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each i,*

(12) $E(\sup_{k\in K(k_0),k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M.$

Proof. The proof of parts (2), (5), (8) and (11) are similar to parts (2), (5), (8) and (11) of Lemma 6. Other terms can be easily shown using Lemma 6. ■

A Proof of Theorem 1

Proof. To prove $\hat{\tau} - \tau_0 = o_p(1)$ as $(N,T) \to \infty$, we need to show for any $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\left|\hat{k} - k_0\right| > T\eta) < \epsilon$ as $(N,T) \to \infty$, i.e., we need to show $P(\hat{k} \in$

$K^c) < \epsilon$ as $(N,T) \to \infty$. $\hat{k} = \arg\max \sum_{i=1}^N [V_i(k) - V_i(k_0)]$, hence $\sum_{i=1}^N [V_i(\hat{k}) - V_i(k_0)] \geq 0$. If $\hat{k} \in K^c$, then $\sup_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0$. This implies $P(\hat{k} \in K^c) \leq P(\sup_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0)$, hence it suffices to show for any given $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\sup_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0) < \epsilon$ as $(N,T) \to \infty$. If $\omega \in \{\omega : \sup_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0\}$ and $\arg\max_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] = k^*$, then $\sum_{i=1}^N [V_i(k^*) - V_i(k_0)] \geq 0$. This implies $\frac{\sum_{i=1}^N [V_i(k^*) - V_i(k_0)]}{|k^* - k_0|} \geq 0$ and it follows $\sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} \geq \frac{\sum_{i=1}^N [V_i(k^*) - V_i(k_0)]}{|k^*-k_0|} \geq 0$. This implies $\omega \in \{\omega : \sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} \geq 0\}$, hence $\{\omega : \sup_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0\} \subseteq \{\omega : \sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} \geq 0\}$. Similarly, $\{\omega : \sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} \geq 0\} \subseteq \{\omega : \sup_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0\}$. Thus, $\{\omega : \sup_{k\in K^c} \sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0\} = \{\omega : \sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} \geq 0\}$ and it suffices to show for any given $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} \geq 0) < \epsilon$ as $(N,T) \to \infty$. Note that $\frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} = -\sum_{i=1}^N G_i(k) + \frac{1}{|k_0-k|} \sum_{i=1}^N H_i(k)$ for $k \neq k_0$, thus $\sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k) - V_i(k_0)]}{|k-k_0|} \geq 0$ implies $\sup_{k\in K^c} \left| \frac{1}{|k-k_0|} \sum_{i=1}^N H_i(k) \right| \geq \inf_{k\in K^c} \sum_{i=1}^N G_i(k)$, it suffices to show that for any $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\sup_{k\in K^c} \left| \frac{1}{|k-k_0|} \sum_{i=1}^N H_i(k) \right| \geq \inf_{k\in K^c} \sum_{i=1}^N G_i(k)) < \epsilon$ as $(N,T) \to \infty$. Due to symmetry, it suffices to study the case $k < k_0$.

Consider the left hand side first.

$$\begin{aligned}
& \frac{1}{|k-k_0|} \sum_{i=1}^N H_i(k) \\
= \; & 2\frac{1}{|k-k_0|} \sum_{i=1}^N \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) - 2\frac{1}{|k-k_0|} \sum_{i=1}^N \delta_i' Z_{0i}'M_ie_i \\
& + \frac{1}{|k-k_0|} \sum_{i=1}^N e_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_ie_i \\
& - \frac{1}{|k-k_0|} \sum_{i=1}^N e_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_ie_i
\end{aligned}$$

For the third term, noting that $M_i(Z_{1i} + Z_{2i}) = M_iZ_i = 0$, we have

$$\begin{aligned}
& \sup_{k\in K^c, k<k_0} \frac{1}{|k-k_0|} \sum_{i=1}^N e_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_ie_i \\
= \; & \sup_{k\in K^c, k<k_0} \frac{1}{|k-k_0|} \sum_{i=1}^N e_i'M_iZ_{1i}(Z_{1i}'M_iZ_{1i})^{-1}Z_{1i}'M_ie_i \\
\leq \; & \frac{1}{T\eta} \sum_{i=1}^N \sup_{k\in K^c, k<k_0} \left\| \frac{e_i'M_iZ_{1i}}{\sqrt{k}} \right\|^2 \sup_{k\in K^c, k<k_0} \left\| (\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1} \right\|,
\end{aligned}$$

thus by parts (16) and (17) of Lemma 3 and Markov inequality, this term is $O_p(\frac{N \log T}{T})$. Similarly, the fourth term is not larger than $\frac{1}{T\eta} \sum_{i=1}^N \left\| \frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}} \right\|^2 \left\| (\frac{Z_{0i}'M_iZ_{0i}}{T-k_0})^{-1} \right\|$, and by parts

(9) and (14) of Lemma 3 and Markov inequality, this term is $O_p(\frac{N}{T})$. For the first term, the expectation is not larger than

$$\begin{aligned} &\frac{2\sqrt{N\lambda_N}}{T\eta}(\lim_{N\to\infty}\max_{1\le i\le N}\frac{N\delta_i'\delta_i}{\lambda_N})^{\frac{1}{2}}\frac{1}{N}\sum_{i=1}^N E(\sup_{k\in K^c,k<k_0}\left\|Z_{0i}'M_iZ_{1i}(Z_{1i}'M_iZ_{1i})^{-1}Z_{1i}'M_ie_i\right\|) \\ \le\ & \frac{2\sqrt{N\lambda_N}}{\sqrt{T}\eta}(\lim_{N\to\infty}\max_{1\le i\le N}\frac{N\delta_i'\delta_i}{\lambda_N})^{\frac{1}{2}}\frac{1}{N}\sum_{i=1}^N E(\sup_{k\in K^c,k<k_0}\left\|\frac{Z_{0i}'M_iZ_{1i}}{k}\right\| \\ &\sup_{k\in K^c,k<k_0}\left\|(\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1}\right\|\sup_{k\in K^c,k<k_0}\left\|\frac{Z_{1i}'M_ie_i}{\sqrt{k}}\right\|), \end{aligned}$$

thus by parts (16), (17) and (18) of Lemma 3 and Markov inequality, this term is $O_p(\sqrt{\frac{N\lambda_N\log T}{T}})$. For the second term, using part (9) of Lemma 3, it is easy to see it's $O_p(\sqrt{\frac{N\lambda_N}{T}})$.

Next consider the right hand side. Using Lemma 4, there exists $\alpha>0$ such that for any $\epsilon>0$, $P(\inf_{k\neq k_0}\sum_{i=1}^N G_i(k)\ge\alpha\lambda_N)>1-\epsilon$ as $(N,T)\to\infty$. Noting that $\inf_{k\in K^c,k<k_0}\sum_{i=1}^N G_i(k)\ge\inf_{k\neq k_0}\sum_{i=1}^N G_i(k)$, under Assumption 4(a), or 4(b), or 4(c), $\sup_{k\in K^c,k<k_0}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right|$ will be dominated by $\inf_{k\in K^c,k<k_0}\sum_{i=1}^N G_i(k)$ as $(N,T)\to\infty$. ■

B Proof of Theorem 2

Proof. To prove $\hat{k}-k_0=O_p(1)$ as $(N,T)\to\infty$ and $\frac{N}{\sqrt{T}}\to 0$, we need to show for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\left|\hat{k}-k_0\right|>C)<\epsilon$. Since $P(\left|\hat{k}-k_0\right|>C)<P(\sup_{|k-k_0|>C}\sum_{i=1}^N[V_i(k)-V_i(k_0)]\ge 0)$, it suffices to show for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\sup_{|k-k_0|>C}\sum_{i=1}^N[V_i(k)-V_i(k_0)]\ge 0)<\epsilon$. Since $\hat{\tau}$ is consistent, $P(\hat{k}\in K^c)\to 0$ as $(N,T)\to\infty$. Noting that $K(C)=\{k:|k-k_0|>C\}\cap K$, it suffices to show for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\sup_{k\in K(C)}\sum_{i=1}^N[V_i(k)-V_i(k_0)]\ge 0)<\epsilon$. Since $\sup_{k\in K(C)}\sum_{i=1}^N[V_i(k)-V_i(k_0)]\ge 0$ implies $\sup_{k\in K(C)}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right|\ge\inf_{k\in K(C)}\sum_{i=1}^N G_i(k)$, it suffices to show that for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\sup_{k\in K(C)}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right|\ge\inf_{k\in K(C)}\sum_{i=1}^N G_i(k))<\epsilon$. Again by symmetry, it suffices to study the case $k<k_0$, i.e. $P(\sup_{k\in K(C),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right|\ge\inf_{k\in K(C),k<k_0}\sum_{i=1}^N G_i(k))<\epsilon$. By Lemma 4, there exists $\alpha>0$ such that for any $\epsilon>0$, there exist $N^* > 0$, $T^* > 0$ such that for $N > N^*$, $T > T^*$, $P(\inf_{k\neq k_0} \sum_{i=1}^N G_i(k) \geq \alpha\lambda_N) > 1-\epsilon$. Noting that $\inf_{k\in K(C),k<k_0} \sum_{i=1}^N G_i(k) \geq \inf_{k\neq k_0} \sum_{i=1}^N G_i(k)$, it suffices to show for any $\epsilon > 0$, there exist $C < \infty$, $N^* > 0$, $T^* > 0$ and $\gamma > 0$, such that for $N > N^*$, $T > T^*$ and $\frac{N}{\sqrt{T}} < \gamma$, $P(\sup_{k\in K(C),k<k_0} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \alpha\lambda_N) < \epsilon$. The first two terms of $\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)$ is

$$\begin{aligned}
& 2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) - 2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{0i}'M_ie_i \\
= & [2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{\Delta i}'M_ie_i \\
& -2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i)]sgn(k_0-k) \\
= & 2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{\Delta i}'e_i - 2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'e_i \\
& -2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) \\
= & A+B+C.
\end{aligned}$$

The last two terms of $\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)$ is

$$\begin{aligned}
& \frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_ie_i - \frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_ie_i \\
= & \frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{\Delta i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{\Delta i}'M_ie_i \\
& +2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{0i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{\Delta i}'M_ie_i \\
& +\frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{0i}[(Z_{2i}'M_iZ_{2i})^{-1} - (Z_{0i}'M_iZ_{0i})^{-1}]Z_{0i}'M_ie_i \\
= & D+E+F.
\end{aligned}$$

Thus by Lemma 5,

$$\begin{aligned}
& \sup_{k\in K(C),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \\
\leq & \sup_{k\in K(C),k<k_0}|A| + \sup_{k\in K(C),k<k_0}|B| + \sup_{k\in K(C),k<k_0}|C| \\
& + \sup_{k\in K(C),k<k_0}|D| + \sup_{k\in K(C),k<k_0}|E| + \sup_{k\in K(C),k<k_0}|F| \\
\leq & \sup_{k\in K(C),k<k_0}|A| + \sup_{k\in K(k_0),k<k_0}|B| + \sup_{k\in K(k_0),k<k_0}|C| \\
& + \sup_{k\in K(k_0),k<k_0}|D| + \sup_{k\in K(k_0),k<k_0}|E| + \sup_{k\in K(k_0),k<k_0}|F| \\
= & \sup_{k\in K(C),k<k_0}|A| + O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N}) + O_p(\frac{N\log T}{T}) + O_p(\frac{N}{\sqrt{T}}).
\end{aligned}$$

Under Assumption 4(a), the last three terms are all $o_p(1)$. For the first term, similar to the proof of part (1) of Lemma 5, for all N we have,

$$\begin{aligned} E(\sup_{k\in K(C),k<k_0}|A|)^2 &\leq 4\lambda_N E(\sup_{k\in K(C),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\eta_{Nt}\right|^2) \\ &\leq 16\lambda_N M\sum_{k=T(\tau_0-\eta)}^{k_0-C-1}\frac{1}{(k_0-k)^2}\leq\frac{16\lambda_N M}{C}<\epsilon, \end{aligned}$$

if C is large enough. The proof is finished. ■

C Proof of Theorem 3

Proof. The proof is similar to the proof of Theorem 2. Based on Theorem 1, $\hat{\tau}$ is consistent under Assumption 4(b) or 4(c), i.e., for any $\epsilon > 0$ and $\eta > 0$, $P(\hat{k} \in K^c) < \epsilon$ as $(N,T) \to \infty$, hence it suffices to show for any $\epsilon > 0$ and $\eta > 0$, $P(\hat{k} \in K(k_0)) < \epsilon$ as $(N,T) \to \infty$ under Assumption 4(b) or 4(c). By Lemma 4, there exists $\alpha > 0$ such that for any $\epsilon > 0$, there exist $N^* > 0$, $T^* > 0$ such that for $N^* > N$, $T > T^*$, $P(\inf_{k\in K(k_0)}\sum_{i=1}^N G_i(k) \geq \alpha\lambda_N) > 1-\epsilon$. By Lemma 5,

$$\begin{aligned} \sup_{k\in K(k_0)}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| &\leq \sup_{k\in K(k_0)}|A| + \sup_{k\in K(k_0)}|B| + \sup_{k\in K(k_0)}|C| \\ &\quad + \sup_{k\in K(k_0)}|D| + \sup_{k\in K(k_0)}|E| + \sup_{k\in K(k_0)}|F| \\ &= O_p(\sqrt{\lambda_N}) + O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N}) + O_p(\frac{N\log T}{T}) + O_p(\frac{N}{\sqrt{T}}). \end{aligned}$$

Under Assumption 4(b) or 4(c), all these four terms will be dominated by $\alpha\lambda_N$, the proof is thus finished. ■

D Proof of Theorem 4

Proof. Define $V_{NT}(k) = \sum_{i=1}^N[V_i(k) - V_i(k_0)]$, $U_{NT}(k) = -\sum_{i=1}^N \delta_i' Z_{\Delta i}' Z_{\Delta i}\delta_i + 2sgn(k_0 - k)\sum_{i=1}^N \delta_i' Z_{\Delta i}' e_i$, both $V_{NT}(k)$ and $U_{NT}(k)$ are countable dimensional random vector. For any fixed constant $C < \infty$, define $V_{NT}^C(k) = V_{NT}(k)$ for $|k_0 - k| < C$, $U_{NT}^C(k) = U_{NT}(k)$ for $|k_0 - k| < C$, $W^C(m) = W(m)$ for $|m| < C$. $V_{NT}^C(k)$, $U_{NT}^C(k)$ and $W^C(m)$ are all finite dimensional random vector.

Step 1: Under Assumption 4(a), $V_{NT}^C(k) \xrightarrow{p} U_{NT}^C(k)$ for any fixed $C < \infty$.

Again due to symmetry, it suffices to show the case $k < k_0$.

For $k \neq k_0$, $V_{NT}(k) = -\left|k_0 - k\right| \sum_{i=1}^{N} G_i(k) + \sum_{i=1}^{N} H_i(k)$, where

$$\begin{aligned}
& -\left|k_0 - k\right| \sum_{i=1}^{N} G_i(k) \\
= & -\sum_{i=1}^{N} \delta_i'[(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})]\delta_i \\
= & -\sum_{i=1}^{N} \delta_i'(Z_{\Delta i}'M_iZ_{\Delta i})\delta_i + \sum_{i=1}^{N} \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i \\
= & -\sum_{i=1}^{N} \delta_i'Z_{\Delta i}'Z_{\Delta i}\delta_i + \sum_{i=1}^{N} \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}\delta_i \\
& + \sum_{i=1}^{N} \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i,
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{N} H_i(k) &= \left|k_0 - k\right|(A + B + C + D + E + F) \\
&= 2sgn(k_0 - k)\sum_{i=1}^{N} \delta_i'Z_{\Delta i}'e_i + \left|k_0 - k\right|(B + C + D + E + F).
\end{aligned}$$

Hence for $k \neq k_0$,

$$\begin{aligned}
V_{NT}(k) - U_{NT}(k) = & \sum_{i=1}^{N} \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}\delta_i \\
& + \sum_{i=1}^{N} \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i \\
& + \left|k_0 - k\right|(B + C + D + E + F),
\end{aligned}$$

and for $k = k_0$, $V_{NT}(k) - U_{NT}(k) = 0$. As proved in Step 2 of Lemma 4, the first two terms are both $O_p(\frac{1}{T})$ uniformly over $k_0 - C \leq k < k_0$ as $(N, T) \to \infty$. For the last five terms, using Lemma 5,

$$\begin{aligned}
& \sup_{k_0 - C \leq k < k_0} \left|\left|k_0 - k\right|(B + C + D + E + F)\right| \\
\leq & C(\sup_{k \in K(k_0), k < k_0} |B| + \sup_{k \in K(k_0), k < k_0} |C| + \sup_{k \in K(k_0), k < k_0} |D| + \sup_{k \in K(k_0), k < k_0} |E| + \sup_{k \in K(k_0), k < k_0} |F|) \\
= & o_p(1),
\end{aligned}$$

as $(N, T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$. Taken together, we have $\sup_{k_0 - C \leq k < k_0} |V_{NT}(k) - U_{NT}(k)| \xrightarrow{p} 0$ as $(N, T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$.

Step 2: For any fixed $C < \infty$, as finite dimensional random vectors, $U_{NT}^C(k) \xrightarrow{d} W^C(k - k_0)$ as $N \to \infty$.

Note that

$$U_{NT}(k) = \begin{cases} -\sum_{t=k+1}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i + 2\sum_{t=k+1}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} e_{it}, \ for \ k - k_0 \leq -1, \\ -\sum_{t=k_0+1}^{k} \sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i - 2\sum_{t=k_0+1}^{k} \sum_{i=1}^{N} \delta_i' z_{it} e_{it}, \ for \ k - k_0 \geq 1. \end{cases}$$

Under part (5) of Assumption 3, part (5) of Assumption 2 and Assumption 4(a), for each t $\sum_{i=1}^{N}\delta_i' z_{it} z_{it}' \delta_i \xrightarrow{p} \lambda\xi$ and as a random vector, $(\frac{\sum_{i=1}^{N}\delta_i' z_{i,k_0} e_{i,k_0}}{\sqrt{\lambda_N}}, ..., \frac{\sum_{i=1}^{N}\delta_i' z_{i,k_0-C} e_{i,k_0-C}}{\sqrt{\lambda_N}})' \xrightarrow{d} (Z_0, ..., Z_{-C})'$. Since

$$\begin{aligned}&(\sum_{t=k_0}^{k_0}\sum_{i=1}^{N}\delta_i' z_{it} e_{it}, ..., \sum_{t=k_0-C}^{k_0}\sum_{i=1}^{N}\delta_i' z_{it} e_{it})' \\ = \quad & Q(\sum_{i=1}^{N}\delta_i' z_{i,k_0} e_{i,k_0}, ..., \sum_{i=1}^{N}\delta_i' z_{i,k_0-C} e_{i,k_0-C})',\end{aligned}$$

where Q is a $(C+1)\times(C+1)$ lower triangular matrix with all nonzero element equal to one, we have

$$\begin{aligned}&(\sum_{t=k_0}^{k_0}\sum_{i=1}^{N}\delta_i' z_{it} e_{it}, ..., \sum_{t=k_0-C}^{k_0}\sum_{i=1}^{N}\delta_i' z_{it} e_{it})' \xrightarrow{d} Q(Z_0, ..., Z_{-C})' \\ = \quad & (\sum_{t=0}^{0} Z_t, ..., \sum_{t=-C}^{0} Z_t)'.\end{aligned}$$

Similarly,

$$(\sum_{t=k_0}^{k_0}\sum_{i=1}^{N}\delta_i' z_{it} z_{it}' \delta_i, ..., \sum_{t=k_0-C}^{k_0}\sum_{i=1}^{N}\delta_i' z_{it} z_{it}' \delta_i)' \xrightarrow{p} (\lambda\xi, ..., (C+1)\lambda\xi)'.$$

For the second half of $U_{NT}(k)$, we have similar result. Taken together, we have $U_{NT}^{C}(k) \xrightarrow{d} W^{C}(k-k_0)$ as $N \to \infty$.

Step 3: $V_{NT}^{C}(k) \xrightarrow{d} W^{C}(k-k_0)$ as $(N,T)\to\infty$ and $\frac{N}{\sqrt{T}} \to 0$ for any fixed $C < \infty$.

Based on Step 1 and Step 2 and using Slutsky's Lemma for random vectors, $V_{NT}^{C}(k) \xrightarrow{d} W^{C}(k-k_0)$.

Step 4: $\arg\max V_{NT}^{C}(k) - k_0 \xrightarrow{d} \arg\max W^{C}(m)$ uniformly as $(N,T)\to\infty$ and $\frac{N}{\sqrt{T}} \to 0$ for any fixed $C < \infty$.

Step 4.1: If $W(m)$ does not have a unique maximizer, then these exist $m \neq m'$ such that $W(m) = W(m')$. Consider the case $m' > m \geq 1$, $P(W(m) = W(m')) = P((m'-m)\xi + 2\sqrt{\lambda}\sum_{t=m}^{m'} Z_t = 0) = 0$. Other cases can be proved similarly. Since the number of integer pairs (m, m') is countable and sum of countable zero is still zero, the probability that $W(m)$ does not have a unique maximizer is zero. Therefore, with probability one $\arg\max W(m)$ is unique.

Step 4.2: Based on Step 3 and using continuous mapping theorem, $\arg\max V_{NT}^{C}(k) \xrightarrow{d} \arg\max W^{C}(m)$. Note that for a finite dimensional vector X, $Y = \arg\max X$ is a continuous function. By definition of convergence of distribution, for any $\epsilon > 0$ and any $1 \leq j \leq C$, there exist $N_j^* > 0$, $T_j^* > 0$ and $\gamma_j > 0$ such that if $N > N_j^*$, $T > T_j^*$ and $\frac{N}{\sqrt{T}} < \gamma_j$, then $\left|P(\arg\max V_{NT}^{C}(k) - k_0 = j) - P(\arg\max W^{C}(m) = j)\right| < \epsilon$. Take

$N^* = \max\{N_j^*, 1 \le j \le C\}$, $T^* = \max\{T_j^*, 1 \le j \le C\}$ and $\gamma = \min\{\ \gamma_j, 1 \le j \le C\}$. Since $C < \infty$, we have $N^* < \infty$, $T^* < \infty$ and $\gamma > 0$. For $N > N^*$, $T > T^*$ and $\frac{N}{\sqrt{T}} < \gamma$, $\left|P(\arg\max V_{NT}^C(k) - k_0 = j) - P(\arg\max W^C(m) = j)\right| < \epsilon$ for all $1 \le j \le C$.

Step 5: $\hat{k} - k_0 \xrightarrow{d} \arg\max W(m)$.

Step 5.1:

$\hat{k} - k_0 = O_p(1)$ as $(N,T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$, hence for any $\frac{\epsilon}{3} > 0$, there exist $C_1 < \infty$, $N_1 > 0$, $T_1 > 0$ and $\gamma_1 > 0$, such that for $N > N_1$, $T > T_1$ and $\frac{N}{\sqrt{T}} < \gamma_1$, $P(\left|\hat{k} - k_0\right| > C_1) < \frac{\epsilon}{3}$.

Step 5.2: $\hat{m} = \arg\max W(m) = O_p(1)$.

By the strong law of large numbers, $W(m) \xrightarrow{a.s.} -\infty$ as $|m| \to \infty$. Thus $P(\lim_{C\to\infty} \sup_{|m|>C} W(m) = -\infty) = 1$ and this implies $\lim_{C\to\infty} P(\sup_{|m|>C} W(m) \ge 0) = P(\lim_{C\to\infty} \sup_{|m|>C} W(m) \ge 0) = 0$. Therefore, for any $\frac{\epsilon}{3} > 0$, there exists $C_2 < \infty$ such that $P(\sup_{|m|>C_2} W(m) \ge 0) < \frac{\epsilon}{3}$. Since $W(0) = 0$, $\sup W(m) \ge 0$, and $P(|\hat{m}| > C_2) \le P(\sup_{|m|>C_2} W(m) \ge 0) < \frac{\epsilon}{3}$.

Step 5.3:

Take $C = \max\{C_1, C_2\}$ in Step 4, then for any $\frac{\epsilon}{3} > 0$, there exist $N_2 > 0$, $T_2 > 0$ and $\gamma_2 > 0$, such that for $N > N_2$, $T > T_2$ and $\frac{N}{\sqrt{T}} < \gamma_2$,

$$\left|P(\arg\max V_{NT}^C(k) - k_0 = j) - P(\arg\max W^C(m) = j)\right| < \frac{\epsilon}{3}$$

for all $1 \le j \le C$.

Step 5.4:

Take $N^* = \max\{N_1, N_2\}$, $T^* = \max\{T_1, T_2\}$ and $\gamma = \min\{\ \gamma_1, \gamma_2\}$. For any $N > N^*$, $T > T^*$ and $\frac{N}{\sqrt{T}} < \gamma$, if $|j| > C$,

$$\begin{aligned}
\left|P(\hat{k} - k_0 = j) - P(\hat{m} = j)\right| &< P(\hat{k} - k_0 = j) + P(\hat{m} = j) \\
&< P(\left|\hat{k} - k_0\right| > C) + P(|\hat{m}| > C) \\
&< P(\left|\hat{k} - k_0\right| > C_1) + P(|\hat{m}| > C_2) \\
&< \frac{\epsilon}{3} + \frac{\epsilon}{3} < \epsilon;
\end{aligned}$$

if $|j| \le C$, $\hat{k} - k_0 = j$ implies $\arg\max V_{NT}^C(k) - k_0 = j$, hence

$$P(\hat{k} - k_0 = j) \le P(\arg\max V_{NT}^C(k) - k_0 = j),$$

and $\arg\max V_{NT}^C(k) - k_0 = j$ implies $\hat{k} - k_0 = j$ or $\left|\hat{k} - k_0\right| > C$, hence

$$P(\arg\max V_{NT}^C(k) - k_0 = j) < P(\hat{k} - k_0 = j) + P(\left|\hat{k} - k_0\right| > C).$$

Therefore,

$$\left|P(\hat{k}-k_0=j)-P(\arg\max V_{NT}^C(k)-k_0=j)\right|<P(\left|\hat{k}-k_0\right|>C)<\frac{\epsilon}{3},$$

and similarly

$$\left|P(\hat{m}=j)-P(\arg\max W^C(m)=j)\right|<P(|\hat{m}|>C)<\frac{\epsilon}{3}.$$

It follows that

$$\begin{aligned}\left|P(\hat{k}-k_0=j)-P(\hat{m}=j)\right| &< \left|P(\hat{k}-k_0=j)-P(\arg\max V_{NT}^C(k)-k_0=j)\right| \\ &\quad+\left|P(\arg\max V_{NT}^C(k)-k_0=j)-P(\arg\max W^C(m)=j)\right| \\ &\quad+\left|P(\hat{m}=j)-P(\arg\max W^C(m)=j)\right| \\ &< \frac{\epsilon}{3}+\frac{\epsilon}{3}+\frac{\epsilon}{3}<\epsilon.\end{aligned}$$

We **have** proved that for any $\epsilon>0$, there exist $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $\left|P(\hat{k}-k_0=j)-P(\hat{m}=j)\right|<\epsilon$ for all j. By definition, $\hat{k}-k_0\xrightarrow{d}\arg\max W(m)$. ■

E Proof of Theorem 5

Proof. The proof of Theorem 1 does not rely on weak cross-sectional dependence, hence $\hat{\tau}-\tau_0$ is still consistent when cross-sectional dependence is strong. The rest of the proof follows the same procedure as Theorem 2. The difference is when cross-sectional dependence is strong, $\sup_{k\in K(k_0),k<k_0}|A|$ is $O_p(\sqrt{N\lambda_N})$, which is of the same order as $\inf_{k\in K(k_0)}\sum_{i=1}^N G_i(k)$ given $\lambda_N=O(N)$. And similar to the proof of Theorem 2,

$$\begin{aligned}E(\sup_{k\in K(C),k<k_0}|A|) &\leq \sum_{i=1}^N E(\sup_{k\in K(C),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\delta_i' z_{it}e_{it}\right|) \\ &\leq 16\sqrt{N\lambda_N}M\sum_{k=T(\tau_0-\eta)}^{k_0-C-1}\frac{1}{(k_0-k)^2}\leq\frac{16\sqrt{N\lambda_N}M}{C}<\alpha\lambda_N,\end{aligned}$$

if C is large enough. ■

F Proof of Theorem 6 and Theorem 7

Proof. To prove Theorem 6, what we need to show is for any $\epsilon>0$ and $\eta>0$, there exist $N^*>0$ and $T^*>0$ such that for $N>N^*$ and $T>T^*$, $P(\left|\tilde{k}-k_0\right|>T\eta)<\epsilon$. First note that

$$P(\left|\tilde{k}-k_0\right| > T\eta) = P(\left|\tilde{k}-k_0\right| > T\eta, \left|\hat{k}-k_0\right| > C) + \sum_{j=-C}^{C} P(\left|\tilde{k}-k_0\right| > T\eta, \left|\hat{k}-k_0\right| = j).$$

Since $\left|\hat{k}-k_0\right| = O_p(1)$, there exists $C > 0$ such that $P(\left|\hat{k}-k_0\right| > C) < \frac{\epsilon}{2}$ for large N and large T, it follows that the first term is less than $\frac{\epsilon}{2}$ for large N and large T. Since $P(\left|\tilde{k}-k_0\right| > T\eta, \left|\hat{k}-k_0\right| = j)$ is no larger than $P(\left|\tilde{k}-k_0\right| > T\eta$, given $|k-k_0| = j)$, it suffices to show for each $j = -C, ..., C$, $P(\left|\tilde{k}-k_0\right| > T\eta$, given $|k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T. By Symmetry, it suffices to show for each $j = -C, ..., C$, $P(\tilde{k} \in K^c$ and $\tilde{k} < k_0$,given $|k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T.

Similarly, to prove Theorem 7, it suffices to show that for each $j = -C, ..., C$, $P(\tilde{k} \neq k_0$,given $|k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T. Theorem 6 shows that $\left|\tilde{k}-k_0\right| = o_p(T)$, hence it suffices to show that for each $j = -C, ..., C$, $P(\tilde{k} \in K(k_0)$,given$|k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T. By symmetry, it suffices to show that for each $j = -C, ..., C$, $P(\tilde{k} \in K(k_0)$ *and* $\tilde{k} < k_0$,given $|k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T.

The rest of the proof follows the same procedure as Theorem 1 and Theorem 3 respectively, but in the current case we have extra regressors $\tilde{F}$ and extra error $(F^0H - \tilde{F})H^{-1}\Lambda$. The proof of Theorem 1 relies on Lemma 3 and Lemma 4, which further relies on Lemma 3. The proof of Theorem 3 relies on Lemma 4 and Lemma 5, which further rely on Lemma 3. Thus, to prove Theorem 6 and Theorem 7, it suffices to reestablish Lemma 3 with the presence of the extra regressors $\tilde{F}$ and extra error $(\tilde{F} - F^0H)H^{-1}\Lambda$. Based on Lemma 6, Lemma 7 and our assumptions on the factor process and error process, this can be easily done following the same procedure as proving Lemma 3. Also note that with $\frac{\sqrt{T}}{N} \to 0$, the effect of using estimated factors disappears asymptotically. ■

Essay II: Identification and Estimation of a Large Factor Model with Structural Instability

1 INTRODUCTION

Large factor models where a large number of time series are simultaneously driven by a small number of unobserved factors, provide a powerful framework to analyze high dimensional data. In the past fifteen years, large factor models have been successfully used in business cycle analysis, consumer behavior analysis, asset pricing and economic monitoring and forecasting, see for example Bernanke, Boivin and Eliasz (2005), Lewbel (1991), Ross (1976) and Stock and Watson (2002b), to mention a few. Estimation theory of large factor models also experienced some breakthroughs, see Bai and Ng (2002) and Bai (2003), to mention a few. While most applications implicitly assume that the number of factors and factor loadings are stable, there is broad evidence of structural instability in macroeconomic and financial time series. Stock and Watson (2002a, 2009) argue that given the number of factors, standard principal component estimation of factors is still consistent if the magnitude of the factor loading break is small enough. Bates, Plagborg-Møller, Stock and Watson (2013) further argue that a sufficient condition for consistent estimation of the factor space is that the magnitude of the factor loading break should converge to zero asymptotically. The condition becomes increasingly stringent if one is to ensure the same convergence rate of the estimated factor space derived in Bai and Ng (2002). This plays a crucial role in subsequent forecasting and factor augmented regression models, and in ensuring consistent estimation of the number of factors. However, in many empirical applications, the magnitude of factor loading break could be large and the number of factors may also change over time. Examples include important economic events such as the European debt crisis, or political events such as the end of the cold war, or policy change such as the end of China's one-child policy, to mention a few.

In the presence of a large factor loading break, estimation ignoring this instability leads to serious consequences. First, the estimated number of factors, using any existing method, e.g., Bai and Ng (2002), Onatski (2009, 2010) and Ahn and Horenstein (2013), is no longer consistent and tends to overestimate. This is because a factor model with unstable factor loadings can be represented by an equivalent model with extra pseudo factors but stable factor loadings. Moreover, the inconsistency of the estimated number of factors will be transmitted to the estimated factors. In such cases, it is hard to interpret the estimated factors, and forecasting performance may also deteriorate since adding extra factors in the forecasting equation does not always control the true factor space[1]. Consequently, a series of tests are proposed to test large factor loading break, including Breitung and Eickmeier (2011), Chen, Dolado and Gonzalo (2014), Han and Inoue (2015) and Corradi and Swanson (2014). Once a large factor loading break has been detected, one still has to estimate the change point, determine the number of pre and post-break factors and estimate the factor space.

In fact, identification and estimation of a factor model in the presence of structural instability have inherent difficulties. First, without knowing the change point, it is infeasible to consistently estimate the factors and factor loadings even if the number of pre-break and post-break factors were known. Second, existing change point estimation methods require knowledge of the number of regressors and observability of the regressors, see for example Bai (1994, 1997, 2010). Hence, to estimate the change point along this path, even if the number of pre-break and post-break factors were known, we still need at least a consistent estimator of the factors, which is infeasible without knowing the change point. For example, consider the case where the number of factors is known, constant over time and after a certain time period, the factor loadings are all doubled. This model can be equivalently represented as the model where factor loadings are constant over time, while factors are all doubled after that time period. In this case, estimating the change point directly following Bai (1994, 1997) is not promising. Cheng, Liao and Schorfheide (2015) propose a shrinkage procedure that consistently estimates the number of pre and post-break factors and consistently detects factor loading breaks when the number of factors is constant, without requiring knowledge of the change point. This result is a significant breakthrough. However, it only leads to a consistent estimate of the change fraction and does not lead to consistent estimates of the factors or factor loadings. In addition, Chen (2015) also proposes a consistent estimate of the change fraction.

In contrast with Cheng, Liao and Schorfheide (2015), we first propose a least squares estimator of the change point without requiring knowledge of the number of factors and observability of the factors. Based on the estimated change point, we then split the sample into two subsamples and use each subsample to estimate the number of pre and post-break

[1] Consider the case where all factor loadings are doubled after the change point. Also, the number of factors is imposed a priori as in many empirical studies. In this case, the true factor space would not be controlled for.

factors as well as the factor space. The key observation behind our change point estimator is that the change point of the factor loadings in the original model is the same as the change point of the second moment matrix of the factors in the equivalent model. Estimating the former can therefore be converted to estimating the latter, thereby circumventing the estimation of the original model. This observation was first utilized by Chen et al. (2014) and Han and Inoue (2015) to test the presence of a factor loading break. Here we further exploit this observation to estimate the change point. More specifically, we start by estimating the number of pseudo factors and the pseudo factors themselves ignoring structural change. This leads us to identify the equivalent model. Based on the estimated pseudo factors, we then estimate the pre and post-break second moment matrix of the pseudo factors for all possible sample splits. The change point is estimated by minimizing the sum of squared residuals of this second moment matrix estimation among all possible sample splits.

Under fairly general assumptions, we show that the distance between the estimated and the true change point is $O_p(1)$. Although our change point estimation itself is a two step procedure, a significant advantage is it has some degree of robustness to misspecification of the number of pseudo factors. The underlying mechanism is that if the number of pseudo factors were underestimated, the change point estimator would be based on a subset of its second moment matrix, hence there is still information to identify the change point. While if the number of pseudo factors were overestimated, no information would be lost although extra noise would be brought in by the extra estimated factors. The latter is similar to Moon and Weidner (2015) who show that for panel data with interactive effects, the limiting distribution of the least squares estimator of the regression coefficients is independent of the number of factors as long as it is not underestimated. Estimating the number of pseudo factors therefore can be seen as a procedure selecting the model with the strongest identification strength of the unknown change point. From this perspective, our method shares some similarity with selecting the most relevant instrumental variables (IVs) among a large number of IVs.

Based on the estimated change point, consistency of the estimated pre and post-break number of factors and consistency of the estimated pre and post-break factor space are established. Also, the convergence rate of the estimated factor space is the same as the one in Bai and Ng (2002) for the stable model, which is crucial for eliminating the effect of using estimated factors in factor augmented regressions. Note that these results are based on an inconsistent change point estimator (the first step estimator). This is different from the traditional plug-in procedure, in which even consistency of the first step estimation does not guarantee that its effect on the second step estimation will vanish asymptotically. In general, the effect of the first step error on the second step estimator depends upon the magnitude of the first step error and how the second step estimator is affected by the first step error. In the traditional plug-in procedure, usually the first step error needs to vanish sufficiently fast to eliminate its effect. In the current context, while the first step error does not vanish asymptotically, the second step becomes increasingly less sensitive to the first step error as the time dimension T goes to infinity. That is to say, the robustness of the second step estimators to the first step error relies on large T. Similar robustness has also been established in Bai (1997). In fact, in Bai (1997) it is a direct corollary that the asymptotic property of the estimated regression coefficients is not affected by the inconsistency of the estimated change point. However, in the current factor setup, it is nontrivial to establish this robustness because estimating the number of factors and factor space is totally different from estimating the regression coefficients.

Our assumptions are quite general. We allow for cases with a change in the number of factors, which can be disappearing or emerging factors. We also allow for cases with only partial change in the factor loadings and cases in which a change in the factor loadings do not lead to extra pseudo factors. Our Assumptions 1-7 are either from or slight modification of Assumptions A-G in Bai (2003). These allow for cross-sectional and temporal dependence as well as heteroskedasticity of the idiosyncratic errors. The main extra assumption we impose is that the Hajek-Renyi inequality is applicable to the second moment process of the factors. As discussed in the next section, this assumption is more general than explicitly assuming a specific factor process and can be easily satisfied. It is also worth noting that for a regularly behaved error term, our results do not rely on the relative speed of the number of subjects (N) and the time series length (T).

The rest of the paper is organized as follows. Section 2 introduces the model setup, notation and preliminaries. Section 3 discusses the equivalent representation and assumptions. Section 4 considers estimation of the change point. Section 5 considers estimation of the number of pre and post-break factors. Section 6 considers estimation of the factor space. Section 7 discusses further issues relating to the limiting distribution of the change point estimator. Section 8 reports the simulation results, while Section 9 concludes. All the proofs are given in the Appendix.

2 NOTATION AND PRELIMINARIES

Consider the following large factor model with structural change in the factor loadings:

$$x_{it} = \begin{cases} f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{1,i} + e_{i,t}, \ if \ 1 \le t \le [\tau_0 T] \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{2,i} + e_{i,t}, \ if \ [\tau_0 T]+1 \le t \le T \end{cases} \quad \text{for } i = 1, ..., N \text{ and } t = 1, ..., T, \quad (1)$$

where $f_t = (f'_{0,t}, f'_{1,t})'$. $f_{1,t}$ and $f_{0,t}$ are q and $r-q$ dimensional vectors of factors with and without structural change in their factor loadings, respectively. $\lambda_{0,i}$ is the factor loadings of subject i corresponding to $f_{0,t}$. $\lambda_{1,i}$ and $\lambda_{2,i}$ are factor loadings of subject i corresponding to $f_{1,t}$ before and after the structural change, respectively. It is easy to see that $r-q=0$ and $r-q>0$ correspond to the pure change case and the partial change case respectively. $e_{i,t}$ is the error term allowed to have temporal and cross-sectional dependence as well as heteroskedasticity. $\tau_0 \in (0,1)$ is the change fraction and $k_0 = [\tau_0 T]$ is the change point.

In matrix form, the model can be represented as:

$$X = \begin{bmatrix} F_1^0 \Lambda'_0 + F_1^1 \Lambda'_1 \\ F_2^0 \Lambda'_0 + F_2^1 \Lambda'_2 \end{bmatrix} + E, \quad (2)$$

where $F_1^0 = [f_{0,1}, ..., f_{0,[\tau_0 T]}]'$, $F_2^0 = [f_{0,[\tau_0 T]+1}, ..., f_{0,T}]'$, $F_1^1 = [f_{1,1}, ..., f_{1,[\tau_0 T]}]'$ and $F_2^1 = [f_{1,[\tau_0 T]+1}, ..., f_{1,T}]'$ are of dimensions $[\tau_0 T] \times (r-q)$, $[(1-\tau_0)T] \times (r-q)$, $[\tau_0 T] \times q$ and $[(1-\tau_0)T] \times q$, respectively. $\Lambda_0 = [\lambda_{0,1}, ..., \lambda_{0,N}]'$, $\Lambda_1 = [\lambda_{1,1}, ..., \lambda_{1,N}]'$ and $\Lambda_2 = [\lambda_{2,1}, ..., \lambda_{2,N}]'$ are of dimensions $N \times (r-q)$, $N \times q$ and $N \times q$, respectively, $E = [e_1, ..., e_T]'$ is of dimension $T \times N$. The matrices F_1^0, F_2^0, F_1^1, F_2^1, Λ_0, Λ_1, Λ_2 and E are all unknown. In addition, $\Lambda_{01} = [\Lambda_0, \Lambda_1] = (\lambda_{01,1}, ..., \lambda_{01,N})'$ and $\Lambda_{02} = [\Lambda_0, \Lambda_2] = (\lambda_{02,1}, ..., \lambda_{02,N})'$ are of dimension $N \times r$. Note that in general not only the factor loadings but also the number of factors may have structural change. In our representation, structural change in the number of factors is incorporated as a special case of structural change in factor loadings by allowing either Λ_{01} or Λ_{02} to be degenerate. In case the number of pre-break and post-break factors are r_1 and r_2 respectively, with $r = \max\{r_1, r_2\}$, f_t and λ_i are always r dimensional vectors and both Λ_{01} and Λ_{02} are of dimensions $N \times r$. If $r_1 < r_2$, some columns in Λ_{01} are zeros and the number of such columns is $r_2 - r_1$. In this case, Λ_{01} is degenerate and Λ_{02} is of full rank. Similarly, if $r_1 > r_2$, some columns in Λ_{02} are zeros and Λ_{01} is of full rank. If $r_1 = r_2$, both

Λ_{01} and Λ_{02} are of full rank r. In addition, we want to point out that although cases with either disappearing factors or emerging factors are allowed for, cases with both disappearing factors and emerging factors are not necessarily identifiable within this mathematical setup. A model with s_1 disappearing factors and s_2 emerging factors can be equivalently represented as a model with $s_1 - s_2$ disappearing factors.

Throughout the paper, $\|A\| = (trAA')^{\frac{1}{2}}$ denotes the Frobenius norm, $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, $vec(A)$ denotes the vectorization of matrix A, $r(A)$ denotes the rank of matrix A, $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$, $(N,T) \to \infty$ denotes N and T going to infinity jointly.

3 EQUIVALENT REPRESENTATION AND ASSUMPTIONS

Since at least one of Λ_{01} and Λ_{02} is of full rank, for the moment, suppose that Λ_{01} is of full rank. Due to symmetry, all results can be established similarly in case Λ_{02} is of full rank. When Λ_{01} is of full rank, the rank of the $N \times (r+q)$ matrix $\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_2 \end{bmatrix}$ is between r and $r+q$. Suppose $\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_2 \end{bmatrix}$ is of rank $r+q_1$, where $0 \le q_1 \le q$, then Λ_2 can be decomposed into $\Lambda_2 = \begin{bmatrix} \Lambda_{21} & \Lambda_{22} \end{bmatrix}$, where Λ_{21} is of dimension $N \times q_1$ and contains the columns in Λ_2 that are linearly independent of Λ_{01}. Λ_{22} is of dimension $N \times q_2$ and contains the columns in Λ_2 that are linear combinations of columns in $\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix}$ such that $\Lambda_{22} = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix} Z$ for some $(r+q_1) \times q_2$ matrix Z. Therefore, $\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix}$ is of full rank $(r+q_1)$ and

$$\begin{aligned} \begin{bmatrix} \Lambda_0 & \Lambda_1 \end{bmatrix} &= \begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix} A, \\ \begin{bmatrix} \Lambda_0 & \Lambda_2 \end{bmatrix} &= \begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix} B, \end{aligned}$$

where $A = \begin{bmatrix} I_r \\ 0_{q_1 \times r} \end{bmatrix}$ and $B = \begin{bmatrix} I_{r-q} & 0_{(r-q)\times q_1} & \\ 0_{q\times(r-q)} & 0_{q\times q_1} & Z \\ 0_{q_1\times(r-q)} & I_{q_1} & \end{bmatrix}$. It follows that model (2) has the following equivalent representation with stable factor loadings:

$$
\begin{aligned}
X &= \begin{bmatrix} \begin{bmatrix} F_1^0 & F_1^1 \end{bmatrix} \begin{bmatrix} \Lambda_0 & \Lambda_1 \end{bmatrix}' \\ \begin{bmatrix} F_2^0 & F_2^1 \end{bmatrix} \begin{bmatrix} \Lambda_0 & \Lambda_2 \end{bmatrix}' \end{bmatrix} + E \\
&= \begin{bmatrix} \begin{bmatrix} F_1^0 & F_1^1 \end{bmatrix} (\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix} A)' \\ \begin{bmatrix} F_2^0 & F_2^1 \end{bmatrix} (\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix} B)' \end{bmatrix} + E \\
&= \begin{bmatrix} \begin{bmatrix} F_1^0 & F_1^1 \end{bmatrix} A' \\ \begin{bmatrix} F_2^0 & F_2^1 \end{bmatrix} B' \end{bmatrix} \begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix}' + E. \qquad (3)
\end{aligned}
$$

Next, define $G = (g_1, ..., g_T)' = \begin{bmatrix} \begin{bmatrix} F_1^0 & F_1^1 \end{bmatrix} A' \\ \begin{bmatrix} F_2^0 & F_2^1 \end{bmatrix} B' \end{bmatrix}$ and $\Gamma = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_{21} \end{bmatrix}$, then

$$
\begin{aligned}
X &= G\Gamma' + E, \qquad (4) \\
g_t &= \begin{cases} Af_t, \ if \ 1 \le t \le [\tau_0 T] \\ Bf_t, \ if \ [\tau_0 T] + 1 \le t \le T \end{cases}, \qquad (5)
\end{aligned}
$$

and we call $r + q_1$ the number of pseudo factors. Equivalent representation of model (2) was first formulated by Han and Inoue (2015). Here our representation is unified, generalizes and complements their result. Our representation is fairly general. The big break case discussed in Chen et al. (2014) corresponds to the case $q_1 = q$, while the type 1, type 2 and type 3 breaks discussed in Han and Inoue (2015) correspond to the cases $q_1 = q$, $q_1 = 0$ and $0 < q_1 < q$ respectively. The type 1 and type 2 changes discussed in Cheng et al. (2015) are also special cases of this representation. To ensure this equivalent representation is unique up to a rotation, it remains to show G is asymptotically full rank, i.e., $\frac{1}{T}\sum_{t=1}^{T} g_t g_t' \xrightarrow{p} \Sigma_G$ for some positive definite Σ_G. Define $\Sigma_F = \mathbb{E}(f_t f_t')$, $\Sigma_{G,1} = \mathbb{E}(g_t g_t')$ for $t \le k_0$ and $\Sigma_{G,2} = \mathbb{E}(g_t g_t')$ for $t > k_0$, then

$$
\begin{aligned}
\Sigma_{G,1} &= A\Sigma_F A', \ \Sigma_{G,2} = B\Sigma_F B', \qquad (6) \\
\Sigma_G &= \tau_0 A\Sigma_F A' + (1 - \tau_0) B\Sigma_F B'. \qquad (7)
\end{aligned}
$$

Proposition 1 *If $\tau_0 \in (0,1)$ and Σ_F is positive definite, Σ_G is positive definite.*

For the case where Λ_{02} is of full rank, Λ_1 can be decomposed as $\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \end{bmatrix}$, where $\begin{bmatrix} \Lambda_0 & \Lambda_2 & \Lambda_{11} \end{bmatrix}$ is of full rank and $\Lambda_{12} = \begin{bmatrix} \Lambda_0 & \Lambda_2 & \Lambda_{11} \end{bmatrix} Z$ for some Z. Define $\Theta =$

$\begin{bmatrix} \Lambda_0 & \Lambda_2 & \Lambda_{11} \end{bmatrix}$.

Our assumptions are as follows:

Assumption 1 *(1)* $\mathbb{E}\left\|f_t\right\|^4 < M < \infty$, $\mathbb{E}(f_t f_t') = \Sigma_F$, Σ_F *is positive definite,* $\frac{1}{k_0}\sum_{t=1}^{k_0} f_t f_t' \xrightarrow{p} \Sigma_F$, $\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} f_t f_t' \xrightarrow{p} \Sigma_F$, *(2) there exists* $d > 0$ *such that* $\|A\Sigma_F A' - B\Sigma_F B'\| > d$ *for all* N.

Assumption 2 $\|\lambda_{l,i}\| \leq \bar{\lambda} < \infty$ *for* $l = 0, 1, 2$, $\left\|\frac{1}{N}\Gamma'\Gamma - \Sigma_\Gamma\right\| \to 0$ *for some positive definite matrix* Σ_Γ *or* $\left\|\frac{1}{N}\Theta'\Theta - \Sigma_\Theta\right\| \to 0$ *for some positive definite matrix* Σ_Θ.

Assumption 3 *There exists a positive constant* $M < \infty$ *such that:*

1 $\mathbb{E}(e_{it}) = 0$, $\mathbb{E}\left|e_{it}\right|^8 \leq M$, *for all* $i = 1, ..., N$, *and* $t = 1, ..., T$,

2 $\mathbb{E}(e_{it}e_{js}) = \tau_{ij,ts}$ *for* $i, j = 1, ..., N$, *and* $t, s = 1, ..., T$, *also*

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\tau_{ij,ts}\right| \leq M,$$

3 For every $(t, s = 1, ..., T)$, $\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right|^4 \leq M$.

Assumption 4 *There exists a positive constant* $M < \infty$ *such that:*

$$\begin{aligned}
\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{k_0}}\sum_{t=1}^{k_0} f_t e_{it}\right\|^2) &\leq M, \\
\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T-k_0}}\sum_{t=k_0+1}^{T} f_t e_{it}\right\|^2) &\leq M.
\end{aligned}$$

Assumption 5 *There exists an* $M < \infty$ *such that:*

1 $\mathbb{E}(\frac{e_s' e_t}{N}) = \gamma_N(s,t)$ *and* $\sum_{s=1}^{T}\left|\gamma_N(s,t)\right| \leq M$ *for every* $t \leq T$,

2 $\mathbb{E}(e_{it}e_{jt}) = \tau_{ij,t}$ *with* $|\tau_{ij,t}| \leq \tau_{ij}$ *for some* τ_{ij} *and for all* $t = 1, ..., T$, *and* $\sum_{j=1}^{N}|\tau_{ji}| \leq M$ *for every* $i \leq N$.

Assumption 6 *The largest eigenvalue of* $\frac{1}{NT}EE'$ *is* $O_p(\frac{1}{\delta_{NT}^2})$.

Assumption 7 *The eigenvalues of* $\Sigma_G\Sigma_\Gamma$ *or* $\Sigma_G\Sigma_\Theta$ *are distinct.*

Assumption 8 *Define* $\epsilon_t = vec(f_t f_t' - \Sigma_F)$. *The data generating process of factors is such that the Hajek-Renyi inequality*[2] *applies to the process* $\{\epsilon_t, t = 1, ..., k_0\}$, $\{\epsilon_t, t = k_0, ..., 1\}$, $\{\epsilon_t, t = k_0 + 1, ..., T\}$ *and* $\{\epsilon_t, t = T, ..., k_0 + 1\}$.

[2] See Appendix for an introduction of the Hajek-Renyi inequality.

Assumption 9 $\frac{\log T}{N} \to 0.$

Assumption 10 *There exists* $M < \infty$ *such that:*

1 For every $s = 1, ..., T$, $\mathbb{E}(\sup_{k<k_0} \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right|^2) \leq M,$

$\mathbb{E}(\sup_{k \leq k_0} \frac{1}{k} \sum_{t=1}^{k} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right|^2) \leq M,$

$\mathbb{E}(\sup_{k>k_0} \frac{1}{k-k_0} \sum_{t=k_0+1}^{k} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right|^2) \leq M,$

$\mathbb{E}(\sup_{k \geq k_0} \frac{1}{T-k} \sum_{t=k+1}^{T} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right|^2) \leq M,$

2 $\mathbb{E}(\sup_{k<k_0} \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{it} \right\|^2) \leq M,$

$\mathbb{E}(\sup_{k \leq k_0} \frac{1}{k} \sum_{t=1}^{k} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{it} \right\|^2) \leq M,$

$\mathbb{E}(\sup_{k>k_0} \frac{1}{k-k_0} \sum_{t=k_0+1}^{k} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{it} \right\|^2) \leq M,$

$\mathbb{E}(\sup_{k \geq k_0} \frac{1}{T-k} \sum_{t=k+1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{it} \right\|^2) \leq M.$

Assumptions 1-7 are either from or slight modification of Assumptions A-G in Bai (2003). Assumption 1(1) corresponds to Assumption A in Bai (2003) and should be satisfied within each regime. f_t can be dynamic and contain their lags. Assumption 1(2) enables the identification of the change point and is general enough to cover all patterns of factor loading break likely in practice. It does not matter whether B depends on N or not, as long as the distance between the pre and post-break second moment matrix of g_t is bounded away from zero as $N \to \infty$. If $r(\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_2 \end{bmatrix}) > r(\begin{bmatrix} \Lambda_0 & \Lambda_1 \end{bmatrix})$, then $A\Sigma_F A' \neq B\Sigma_F B'$. If $r(\begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_2 \end{bmatrix}) = r(\begin{bmatrix} \Lambda_0 & \Lambda_1 \end{bmatrix})$, then $A\Sigma_F A' = \Sigma_F$ and $B\Sigma_F B' \neq \Sigma_F$ except for some very unlikely case, for example, some post-break factor loadings are -1 times their pre-break factor loadings. Note that here to simplify analysis, the second moment matrix of the factors is assumed to be stationary over time, since in general how to disentangle structural change in Σ_F from structural change in factor loadings is still unclear. Assumption 2 corresponds to Assumption B in Bai (2003) and implies that $\left\| \frac{1}{N}\Lambda_{01}'\Lambda_{01} - \Sigma_{\Lambda_{01}} \right\| \to 0$ and $\left\| \frac{1}{N}\Lambda_{02}'\Lambda_{02} - \Sigma_{\Lambda_{02}} \right\| \to 0$. Note that one of Λ_{01} and Λ_{02} is allowed to be degenerate. This allows for cases with disappearing or emerging factors. In addition, Λ_0 could contain a small change. Let $\Delta\lambda_{0,i}$ be the change of $\lambda_{0,i}$. As discussed in Bates et al. (2013), if $\Delta\lambda_{0,i} = \frac{1}{\sqrt{NT}}\kappa_i$ and $\|\kappa_i\| \leq \bar{\kappa} < \infty$ for all i, consistency of the estimated number of factors and the factors themselves will not be affected. For simplicity, we assume that Λ_0 is stable. Assumptions

3 and 5 correspond to Assumptions C and E in Bai (2003), which allow for the temporal and cross-sectional dependence as well as heteroskedasticity. Assumption 4 corresponds to Assumption D in Bai (2003) and should be satisfied within each regime. This is implied by Assumptions 1 and 3 if the factors and the errors are independent. Assumption 6 is the key condition for identifying the number of factors and is implicitly assumed in Bai and Ng (2002) and required in almost all existing methods of determining the number of factors or the number of dynamic factors. For example, Onatski (2010) and Ahn and Horenstein (2013) assume $E = A\varepsilon B$, where ε is an i.i.d. $T \times N$ matrix and A and B characterize the temporal and cross-sectional dependence and heteroskedasticity. This is a sufficient but not necessary condition for Assumption 6. In this paper, Assumption 6 can be relaxed to "The largest eigenvalue of $\frac{1}{NT}EE'$ is $o_p(1)$", yet still allows consistent estimation of the number of factors. Assumption 7 corresponds to Assumption G in Bai (2003).

Assumption 8 strengthens Assumption 1(1) and imposes further requirement on the factor process. Instead of assuming a specific data generating process, here we only require that the Hajek-Renyi inequality is applicable to the second moment process of the factors, which incorporates i.i.d., martingale difference, martingale, mixingale and so on as special cases and renders Assumption 8 in its most general form. Assumption 10 imposes further constraints on the idiosyncratic error. Assumption 3(3) and Assumption F3 in Bai (2003) imply that the summands in Assumption 10 are uniformly $O_p(1)$. Assumption 10 strengthens this condition such that the supremum of the average process of these summands is $O_p(1)$. Also note that stationarity is not assumed in Assumption 10. In rare cases, Assumption 10 is not satisfied, but we can still proceed with Assumption 9. Compared to $\frac{\sqrt{T}}{N} \to 0$, which is assumed in Chen et al. (2014), Han and Inoue (2015), Assumption 9 is significantly weaker and much easier to be satisfied since even when T is much larger than N, $\frac{\log T}{N}$ could still be very close to zero.

4 ESTIMATING THE CHANGE POINT

4.1 THE ESTIMATION PROCEDURE

In this subsection, we discuss how to estimate the change point with an unknown number of latent factors. First, we estimate the number of factors ignoring structural change. Define $\tilde{r}$ as the estimated number of factors using the information criteria in Bai and Ng (2002), we will have $\lim_{(N,T)\to\infty} P(\tilde{r} = r + q_1) = 1$, since model (2) can be equivalently represented as model (3). Note that q_1 could be zero, since structural change does not necessarily lead to overestimating the number of factors. Using $\tilde{r}$, we then estimate the factors using the principal component method. This identifies the factors g_t. As noted in (6), the second moment matrix[3] of g_t has a break at the point k_0. Hence, estimating change point of factor loadings can be converted to estimating change point of the second moment matrix of g_t. Although g_t is not directly observable, the principal component estimator $\tilde{g}_t$ is asymptotically close to $J'g_t$ for some rotation matrix J. And $J \xrightarrow{p} J_0 = \Sigma_\Gamma^{\frac{1}{2}}\Phi V^{-\frac{1}{2}}$ as $(N,T) \to \infty$, where V and Φ are the eigenvalue matrix and eigenvector matrix of $\Sigma_\Gamma^{\frac{1}{2}}\Sigma_G\Sigma_\Gamma^{\frac{1}{2}}$ respectively. Hence change point estimation using $\tilde{g}_t$ will be asymptotically equivalent to using $J_0 g_t$. It is easy to see that the second moment matrix of $J_0 g_t$ shares the same change point as that of g_t. Therefore, we proceed to estimate the pre-break and post-break second moment matrix of g_t using the estimated factors $\tilde{g}_t$.

More specifically, following Bai (1994, 1997, 2010), for any $k > 0$ we split the sample into two subsamples and estimate the pre-break and post-break second moment matrix of g_t as

$$\begin{aligned} \tilde{\Sigma}_1 &= \frac{1}{k}\sum\nolimits_{t=1}^{k} \tilde{g}_t\tilde{g}_t', \\ \tilde{\Sigma}_2 &= \frac{1}{T-k}\sum\nolimits_{t=k+1}^{T} \tilde{g}_t\tilde{g}_t', \end{aligned} \tag{8}$$

and define the sum of squared residuals as

$$\tilde{S}(k) = \sum\nolimits_{t=1}^{k}[vec(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_1)]'[vec(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_1)] + \sum\nolimits_{t=k+1}^{T}[vec(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_2)]'[vec(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_2)]. \tag{9}$$

The least squares estimator of the change point[4] is

$$\tilde{k} = \arg\min \tilde{S}(k). \tag{10}$$

Here we use $\tilde{S}(k)$ to emphasize that the sum of squared residuals is based on the estimated factors.

Remark 1 *The change point estimator also can be based on $\hat{g}_t$ instead of $\tilde{g}_t$, where $(\hat{g}_1, ..., \hat{g}_T)' =$*

[3] The first moment of g_t may also help identify the change point, but it requires the true factors f_t to have nonzero mean.

[4] Alternatively, one referee points out that one may consider quasi-maximum likelihood estimation of the change point: $\tilde{k}_{ML} = \arg\max[-k\log\left|\tilde{\Sigma}_1\right| - (T-k)\log\left|\tilde{\Sigma}_2\right|]$.

$\hat{G} = \tilde{G}V_{NT} = (\tilde{g}_1, ..., \tilde{g}_T)'V_{NT}$ *and* V_{NT} *is diagonal and contains the first* $r + q_1$ *largest eigenvalues of* $\frac{1}{NT}XX'$ *in decreasing order.*

4.2 ASYMPTOTIC PROPERTIES OF THE CHANGE POINT ESTIMATOR

In what follows, we shall establish the rate of convergence of the proposed estimator, which allows us to identify the number of pre-break and post-break factors as well as the factor space. Since $\lim_{(N,T)\to\infty} P(\tilde{r} = r + q_1) = 1$, estimation of the change point based on $\tilde{r}$ and the true number of pseudo factors $r + q_1$ is asymptotically equivalent. The proof is similar to footnote 5 in Bai (2003). Therefore, we can treat the number of pseudo factors $r + q_1$ as known in studying the asymptotic properties of our change point estimator.

Define $\tilde{\tau} = \tilde{k}/T$ as the estimated change fraction, we first show that $\tilde{\tau}$ is consistent.

Proposition 2 *Under Assumptions 1-8 and 9 or 10,* $\tilde{\tau} - \tau_0 = o_p(1)$.

This proposition is important for theoretical purposes. In fact, it serves as a first step in proving Theorem 1. Proposition 2 implies that for any $\epsilon > 0$ and $\eta > 0$, $P(\tilde{\tau} \in D) > 1 - \epsilon$ for sufficiently large N and T, where $D = \{k : |k - k_0|/T \leq \eta\}$. Using similar strategy as proving Proposition 2, we can further show that for any $\epsilon > 0$ and $\eta > 0$, there exist an $M > 0$ such that $P(\tilde{k} \in D_M) < \epsilon$ for sufficiently large N and T, where $D_M = \{k : k \in D, |k - k_0| > M\}$. Taken together, we have:

Theorem 1 *Under Assumptions 1-8 and 9 or 10,* $\tilde{k} - k_0 = O_p(1)$.

This theorem implies that the difference between the estimated change point and the true change point is stochastically bounded. This is quite strong since the possible change point is narrowed to a bounded interval no matter how large T is. Although $\tilde{k}$ is still inconsistent, an important observation is that $\tilde{k} - k_0 = O_p(1)$ is already sufficient for consistent estimation of the number of pre-break and post-break factors and consistent estimation of the pre-break and post-break factor space, which will be discussed further in the next three sections.

Theorem 1 differs from existing results in the change point estimation literature. First, in the current setup N goes to infinity jointly with T, thus we should be able to achieve consistency of $\tilde{k}$ as shown in Bai (2010) for the panel mean shift case, because large N will help identify the change point when the change point is common across individuals. Our result is different from Bai (2010) and instead similar to the univariate case, e.g., Bai (1994, 1997), because $\tilde{k}$ is based on $\tilde{g}_t\tilde{g}_t'$ which is a fixed dimensional multivariate time series with mean shift. Second, our result is also different from Bai (1994, 1997) because in the current setup we are using estimated data $\tilde{g}_t\tilde{g}_t'$ rather than the raw data $J_0 g_t g_t' J_0'$ to estimate the change point, i.e., the data $\tilde{g}_t\tilde{g}_t'$ contains measurement error $\tilde{g}_t\tilde{g}_t' - J_0 g_t g_t' J_0'$. Eliminating the effect of this measurement error on estimation of change point relies on large N.

Remark 2 *Proposition 2 and Theorem 1 hold with either Assumption 9 or 10, but we do not need both. Usually Assumption 10 is satisfied. In this case, there is no restriction on the relative speed of N and T going to infinity. Even when Assumption 10 is violated, our results only require $\frac{\log T}{N} \to 0$, which can be easily satisfied.*

Remark 3 *Note that Theorem 1 requires the covariance matrix of the factors to be stationary, and thus is not robust to heteroskedasticity of the factors. This problem is common in the literature, for example, it also appears in Chen et al. (2014), Han and Inoue (2015) and Cheng et al. (2015). It is important to note that Chen (2015)'s change point estimator is robust to heteroskedasticity of the factors.*

4.3 THE EFFECT OF USING ESTIMATED NUMBER OF PSEUDO FACTORS ON ESTIMATION OF THE CHANGE POINT

Since our method for estimating the change point is a two step procedure, a natural question is how will the model selection error in the first step affect the performance of the second step estimation. Although consistent model selection guarantees that asymptotically we can behave as if the true model is known a priori, the finite sample distribution of the post model selection estimator could be dramatically different from its asymptotic limit even when the sample size is very large. This is because the probability of misspecifying the model in the first step may be nonignorable even when the sample size is very large if consistency of the first step model selection is not uniform with respect to the parameter space. The distribution of the post model selection estimator is a weighted average of its distribution given the true model is selected and given some misspecified model is selected, where the weight is given by the probability of selecting that model. When the probability of misspecifying the model is indeed nonignorable and the distributions with the true model selected and with the misspecified model selected are very different, we can imagine that the composite distribution could be far away from its asymptotic limit.

In the current context, the Leeb and Potscher (2005)'s criticism still applies. But, we argue that our change point estimator still has some degree of robustness to the first step estimation error, especially if we only care about the stochastic order of the change point estimation error. This is because if the number of pseudo factors were underestimated, $\tilde{k}$ would be based on a subset of the second moment matrix of $J_0 g_t$. Hence there is still information to identify the change point. While if the number of pseudo factors were overestimated, no information would be lost but extra noise would be brought in by the extra estimated factors. Therefore, estimating the number of pseudo factors can be seen as a procedure selecting the model with the strongest identification strength of the unknown change point. From this perspective, our method shares some similarity with selecting the most relevant instrumental variables (IVs) among a large number of IVs.

In case $\tilde{r}$ is fixed at some positive integer $m < r + q_1$, we have the following result:

Corollary 1 *For any positive integer $m < r + q_1$ and change point estimation based on $\tilde{r} = m$, with J_0 replaced by J_0^m which is of dimension $(r + q_1) \times m$ and contains the first m columns of J_0, and $\|J_0^{m\prime}\Sigma_{G,1}J_0^m - J_0^{m\prime}\Sigma_{G,2}J_0^m\| > d$ for some $d > 0$ and all N, Proposition 2 and Theorem 1 still hold.*

In case $\tilde{r}$ is fixed at some positive integer $m > r + q_1$, we can not prove the robustness of Proposition 2 and Theorem 1. Nonetheless, if the change point estimator were based on $\hat{g}_t$ instead of $\tilde{g}_t$, we can prove:

Corollary 2 *For any positive integer $m > r + q_1$ and change point estimator $\hat{k}$ based on $\hat{g}_t$ and $\tilde{r} = m$, if $\frac{\sqrt{T}}{N} \to 0$, Proposition 2 and Theorem 1 still hold.*

Note that Corollary 1 also applies to $\hat{k}$. Corollary 2 shows that $\hat{k}$ is robust to overestimation of the number of pseudo factors. This result is similar to Moon and Weidner (2015) who show that for panel data with interactive effects, the limiting distribution of the LS estimator is independent of the number of factors used in the estimation, as long as this number is not underestimated.

Remark 4 *If the condition "$\|J_0^{m\prime}\Sigma_{G,1}J_0^m - J_0^{m\prime}\Sigma_{G,2}J_0^m\| > d$ for some $d > 0$ and all N" is not satisfied for all m, estimation errors of the number of the pseudo factors may affect the uniform validity of the estimation procedure. In such case, simply fixing $\tilde{r}$ at the maximum number of pseudo factors may be preferred, especially when this maximum number is small or some prior information is available.*

Remark 5 *As can be seen in the equivalent representation, the pseudo factors induced by structural change are relatively weaker than factors with stable loadings in the original model because a portion of their elements are zeros and the magnitude of those nonzero elements is small if the magnitude of structural change is small. Since underestimation is more harmful*[5] *compared to overestimation, we recommend choosing a less conservative criterion in estimating the number of pseudo factors. We will discuss this further in the simulation section.*

Up to now, we have only touched upon the stochastic order of $\tilde{k} - k_0$. We will postpone the discussion of the imiting distribution and instead put more emphasis on the estimation of the pre and post-break number of factors and factor space. We will show that $\tilde{k} - k_0 = O_p(1)$ is a sufficient condition for the results in subsequent estimation. Thus for the purpose of subsequent estimation, the limiting distribution is not needed.

5 DETERMINING THE NUMBER OF FACTORS

In this section, we study how to consistently estimate the number of factors in the presence of structural instability in the factor loadings or the number of factors themselves. We first relax the sufficient condition proposed by Bates et al. (2013) for the consistent estimation of the number of factors in the presence of structural change using the Bai and Ng (2002) information criteria. The condition they propose is $\frac{1}{N}\|\Delta\|^2 = O(\frac{1}{\delta_{NT}^2})$, where Δ is the matrix of factor loading breaks. In the current setup, $\Delta = \Lambda_2 - \Lambda_1$. We show, in the following proposition, that their condition can be relaxed to $\frac{1}{N}\|\Delta\|^2 = O(\frac{1}{\delta_{NT}^c})$ for some $c > 0$.

Proposition 3 *In the presence of a single common break in factor loadings, the estimator of the number of factors using the Bai and Ng (2002) information criteria is still consistent if* $\frac{1}{N}\|\Delta\|^2 = O(\frac{1}{\delta_{NT}^c})$ *for some* $c > 0$, $g(N,T) \to 0$ *and* $\delta_{NT}^c g(N,T) \to \infty$, *where* $g(N,T)$ *is the penalty function.*

The formal proof is in the Appendix. This proposition complements Theorem 2 below. Note that c can be arbitrarily close to zero, hence our condition is much weaker than that of Bates et al. (2013). The intuition behind our result is that change in factor loadings can be

[5] As discussed above, underestimation will result in loss of useful moment conditions while overestimation will bring in irrelevant moment conditions. In the current setup, loosing useful moment conditions is more harmful.

treated as an extra error term and as long as $c > 0$, the first r largest eigenvalues of XX' are still separated from the rest. By adjusting the speed at which the penalty function goes to zero accordingly, the number of factors can still be consistently determined. Some caveats are the following: When c is less than two, the magnitude of this extra error term becomes large. To outweigh the error term, the speed at which the penalty function $g(N,T)$ goes to zero has to be slower than the speed at which $\frac{1}{N}\|\Delta\|^2$ goes to zero, so that $\frac{g(N,T)}{\frac{1}{N}\|\Delta\|^2} \to \infty$. This may be problematic in real applications, since when c is close to zero, not all factors are necessarily strong enough to outweigh the extra noise brought by the factor loadings breaks. And even if factors are strong enough, we still need to pin down c, which is difficult. In addition, the above result is not applicable for the case where $\frac{1}{N}\|\Delta\|^2 = O(1)$, nor the case where the number of factors also change. In view of these caveats, Proposition 3 is more of theoretical importance and demonstrates how far we can go following Bates et al. (2013).

To estimate the number of pre and post-break factors in the presence of large break, we propose the following procedure: split the sample into two subsamples based on the estimated change point $\tilde{k}$, and then use each subsample to estimate the number of pre and post-break factors. Let $\tilde{r}_1$ and $\tilde{r}_2$ be the estimated number of pre-break and post-break factors using the method in Bai and Ng (2002). We have the following result:

Theorem 2 *Under Assumptions 1-8 and 9 or 10,* $\lim_{(N,T)\to\infty} P(\tilde{r}_1 = r_1) = 1$ *and* $\lim_{(N,T)\to\infty} P(\tilde{r}_2 = r_2) = 1$*, where* r_1 *and* r_2 *are numbers of pre-break and post-break factors, respectively.*

Theorem 2 together with Theorem 1 identifies model (2) and provides the basis for subsequent estimation and inference. Note that $\tilde{k} - k_0 = O_p(1)$ is sufficient for the consistency of $\tilde{r}_1$ and $\tilde{r}_2$, i.e., consistency of the second step estimators $\tilde{r}_1$ and $\tilde{r}_2$ does not require consistency of the first step estimator $\tilde{k}$.[6] This is because $\tilde{k} - k_0 = O_p(1)$ is the exact condition that guarantees the extra noise brought by a change in factor loadings does not affect the speed of eigenvalue separation. In general, the effect of the error in the first step, which could be either estimation or model selection, on the second step estimator depends on the magnitude of the first step error and how the second step estimator is affected by the first step error. In the traditional plug-in procedure, usually the first step error need to vanish sufficiently fast to eliminate its effect. In the current context, although the first step error does not vanish asymptotically, the second step becomes increasingly less sensitive to

[6] When estimating the pre and post-break number of factors and factor space, we consider $\tilde{k}$ as the first step estimator.

the first step error as $T \to \infty$. This can be seen more easily by considering the case in which T is very large while $\left|\tilde{k} - k_0\right|$ is bounded. Since the pre and post-break number of factors and factor space are estimated using each subsample whose size is $O(T)$, misspecifying the change point by a bounded value would affect their behavior very little. In other words, while large T does not help identify the change point, it increases the magnitude of misspecification of change point that can be tolerated.

To better demonstrate the difference between our result and traditional plug-in procedure, we sketch the key steps in proving the consistency of $\tilde{r}_1$. The estimator of the number of pre-break factors $\tilde{r}_1$ is based on the pre-break subsample $t = 1, ..., \tilde{k}$. What we need to show is: for any $\epsilon > 0$, $P(\tilde{r}_1 \neq r_1) < \epsilon$ for large (N, T). Based on $\left|\tilde{k} - k_0\right| = O_p(1)$, we have for any $\epsilon > 0$, there exists $M > 0$ such that $P(\left|\tilde{k} - k_0\right| > M) < \epsilon$ for all (N, T). Based on this M, $P(\tilde{r}_1 \neq r_1)$ can be decomposed as

$$P(\tilde{r}_1 \neq r_1, \left|\tilde{k} - k_0\right| > M) + P(\tilde{r}_1 \neq r_1, k_0 - M \leq \tilde{k} \leq k_0) + P(\tilde{r}_1 \neq r_1, k_0 + 1 \leq \tilde{k} \leq k_0 + M).$$

The first term is less than $P(\left|\tilde{k} - k_0\right| > M)$, hence less than ϵ for all (N, T). The second term can be further decomposed as

$$\sum\nolimits_{k=k_0-M}^{k_0} P(\tilde{r}_1(k) \neq r_1, \tilde{k} = k),$$

where $P(\tilde{r}_1(k) \neq r_1, \tilde{k} = k)$ denotes the joint probability of $\tilde{k} = k$ and $\tilde{r}_1(k) \neq r_1$ and $\tilde{r}_1(k)$ denotes the estimated number of pre-break factors using subsample $t = 1, ..., k$. Obviously, $P(\tilde{r}_1(k) \neq r_1, \tilde{k} = k) \leq P(\tilde{r}_1(k) \neq r_1)$, hence the second term is less than $\sum\nolimits_{k=k_0-M}^{k_0} P(\tilde{r}_1(k) \neq r_1)$. Furthermore, the factor loadings in the pre-break subsample are stable when $k < k_0$ and for $k \in [k_0 - M, k_0]$, $k \to \infty$ at the same speed as k_0, hence we have for each $k \in [k_0 - M, k_0]$, $P(\tilde{r}_1(k) \neq r_1) \leq \frac{\epsilon}{M+1}$ for large (N, T). The second term is therefore less than $\sum\nolimits_{k=k_0-M}^{k_0} \frac{\epsilon}{M+1} = \epsilon$ for large (N, T). The argument for the second term also applies to the third term, except for some modifications. First, the third term can be decomposed similarly as

$$\sum\nolimits_{k=k_0+1}^{k_0+M} P(\tilde{r}_1(k) \neq r_1, \tilde{k} = k) \leq \sum\nolimits_{k=k_0+1}^{k_0+M} P(\tilde{r}_1(k) \neq r_1),$$

hence it remains to show for each $k \in [k_0 + 1, k_0 + M]$, $P(\tilde{r}_1(k) \neq r_1) \leq \frac{\epsilon}{M}$ for large (N, T). Unlike the second term, when $k \in [k_0 + 1, k_0 + M]$ the factor loadings of the pre-break subsample $t = 1, ..., k$ has a break at $t = k_0$, hence results already established for the stable model are not directly applicable. Nevertheless, the number of observations with factor loading break, $k-k_0$, is bounded by M. Hence in estimating the number of factors, these observations will be dominated by the observations $t=1,...k_0$, as $k_0=[\tau_0 T]\to\infty$.

6 ESTIMATING THE FACTOR SPACE

In this section, we discuss the estimation of the pre-break and post-break factor space. As in last section, we split the sample into two subsamples based on the change point estimator $\tilde{k}$, and then use each subsample to estimate the pre-break and post-break factor space. For each possible sample split k, define $X(k)=(x_1,...,x_k)'$, $F_1(k)=(f_1,...,f_k)'$ and $F_2(k)=(f_{k+1},...,f_T)'$. Let u be any prespecified number of pre-break factors, which does not necessarily equal r_1. The principal component estimator of the pre-break factors and factor loadings are obtained by solving $V(u)=\min\frac{1}{Nk}\sum_{t=1}^{k}\sum_{i=1}^{N}(x_{it}-f_t'\lambda_i)^2$. Since the true factors can be identified only up to a rotation, the normalization condition has to be imposed to uniquely determine the solution, and based on different normalization conditions there are two solutions. For the first one, the estimated factors, $\tilde{F}_1^u(k)$, equal $\sqrt{T}$ times the eigenvectors corresponding to the first u largest eigenvalues of $\frac{1}{Nk}X(k)X'(k)$ and $\tilde{\Lambda}_1^u(k)=\frac{1}{k}X'(k)\tilde{F}_1^u(k)$ are the corresponding estimated factor loadings. For the second one, the estimated factor loadings, $\bar{\Lambda}_1^u(k)$, equal $\sqrt{N}$ times the eigenvectors corresponding to the first u largest eigenvalues of $\frac{1}{Nk}X'(k)X(k)$ and $\bar{F}_1^u(k)=\frac{1}{N}X(k)\bar{\Lambda}_1^u(k)$ are the corresponding estimated factors. Following Bai and Ng (2002), we define the rescaled estimator $\hat{F}_1^u(k)=\bar{F}_1^u(k)[\frac{1}{k}\bar{F}_1^{u\prime}(k)\bar{F}_1^u(k)]^{\frac{1}{2}}$. The estimator of the post-break factors $\hat{F}_2^v(k)$ can be obtained similarly based on the post-break subsample, where v is the prespecified number of post-break factors. Next, define $H_1^u(k)=\frac{\Lambda_{01}'\Lambda_{01}}{N}\frac{F_1'(k)\tilde{F}_1^u(k)}{k}$ and $H_2^v(k)=\frac{\Lambda_{02}'\Lambda_{02}}{N}\frac{F_2'(k)\tilde{F}_2^v(k)}{T-k}$. Let $\hat{f}_t^u(\tilde{k})$ and $\hat{f}_t^v(\tilde{k})$ be the estimated factors based on change point estimator $\tilde{k}$ for $t\leq\tilde{k}$ and $t>\tilde{k}$ respectively, we have the following theorem:

Theorem 3 *Under Assumptions 1-8 and 9 or 10,*

$$\begin{aligned}\frac{1}{\tilde{k}}\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k})-H_1^{u\prime}(\tilde{k})f_t\right\|^2 &= O_p(\frac{1}{\delta_{NT}^2}),\\ \frac{1}{T-\tilde{k}}\sum_{t=\tilde{k}+1}^{T}\left\|\hat{f}_t^v(\tilde{k})-H_2^{v\prime}(\tilde{k})f_t\right\|^2 &= O_p(\frac{1}{\delta_{NT}^2}).\end{aligned}$$

Theorem 3 implies that our estimator of the factor space is mean squared consistent within each regime and the convergence rate is the same as that obtained by Bai and Ng

(2002) for the stable model. Consistent estimation of the factor space has proved to be crucial in many cases, including forecasting and factor augmented regressions. Note that the convergence rate $O_p(\frac{1}{\delta_{NT}^2})$ plays a crucial role in eliminating the effect of using estimated factors, for which consistency is not enough. Bates et al. (2013) show that if we ignore the structural change, consistency of the estimated factor space requires $\frac{1}{N}\left\|\Delta\right\|^2 = o(1)$. In contrast, to guarantee the convergence rate $O_p(\frac{1}{\delta_{NT}^2})$ of the estimated factor space, it requires $\frac{1}{N}\left\|\Delta\right\|^2 = O(\frac{1}{\delta_{NT}})$. While reasonable for a small break, these two conditions especially the latter are not suitable for a large break. As discussed in Banerjee, Marcellino and Masten (2008), this is the most likely reason behind the worsening factor-based forecasts. In contrast, our result allows for a large break, and hence improves and complements Bates et al. (2013).

Remark 6 *Note that $\tilde{k}-k_0 = O_p(1)$ is both a necessary and sufficient condition for Theorem 3. If $\left|\tilde{k}-k_0\right|$ is of order larger than $O_p(1)$, the convergence speed in Theorem 3 will be affected.*

Remark 7 *Theorem 3 is based on arbitrarily u and v rather than $\tilde{r}_1$ and $\tilde{r}_2$, the estimated number of pre-break and post-break factors. On the other hand, $\tilde{r}_1$ and $\tilde{r}_2$ are based directly on eigenvalue separation, without using consistency of the estimated pre-break and post-break factor space. Hence, Theorem 3 and Theorem 2 are independent of each other. Alternatively, we can choose $u = \tilde{r}_1$ and $v = \tilde{r}_2$. Since $\tilde{r}_1$ and $\tilde{r}_2$ are consistent, this is asymptotically equivalent to the case in which r_1 and r_2 are known. The same argument was used by Bai (2003) for deriving the limiting distribution of the estimated factors. When r_1 and r_2 are known and under Assumptions 1-8 and 9 or 10, we have $\frac{1}{\tilde{k}}\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t(\tilde{k}) - H_1'(\tilde{k})f_t\right\|^2 = O_p(\frac{1}{\delta_{NT}^2})$ and $\frac{1}{T-\tilde{k}}\sum_{t=\tilde{k}+1}^{T}\left\|\hat{f}_t(\tilde{k}) - H_2'(\tilde{k})f_t\right\|^2 = O_p(\frac{1}{\delta_{NT}^2})$.*

7 FURTHER ISSUES

To make inference about the change point, we seek to derive its limiting distribution. Define

$$\begin{aligned} y_t &= vec(J_0' g_t g_t' J_0 - \Sigma_1) \text{ for } t \leq k_0, \\ y_t &= vec(J_0' g_t g_t' J_0 - \Sigma_2) \text{ for } t > k_0, \end{aligned} \tag{11}$$

where $\Sigma_1 = J_0'\Sigma_{G,1}J_0$ and $\Sigma_2 = J_0'\Sigma_{G,2}J_0$ are the pre-break and post-break means of $J_0' g_t g_t' J_0$. The limiting distribution of $\tilde{k}$ is as follows:

Theorem 4 *Under Assumptions 1-8 and 9 or 10,* $\tilde{k} - k_0 \xrightarrow{d} \arg\min W(l)$*, where*

$$\begin{aligned} W(l) &= -l\left\|\Sigma_2 - \Sigma_1\right\|^2 - 2\sum\nolimits_{t=k_0+l}^{k_0-1}[vec(\Sigma_2 - \Sigma_1)]' y_t \text{ for } l = -1, -2, ..., \\ W(l) &= 0 \text{ for } l = 0, \\ W(l) &= l\left\|\Sigma_2 - \Sigma_1\right\|^2 - 2\sum\nolimits_{t=k_0+1}^{k_0+l}[vec(\Sigma_2 - \Sigma_1)]' y_t \text{ for } l = 1, 2, ... \end{aligned} \quad (12)$$

If y_t is independent over t, then $W(l)$ is a two-sided random walk. Note that y_t is not assumed to be stationary. By definition, if f_t is stationary, then g_t and hence y_t is stationary within each regime. In this case $\sum_{t=k_0+l}^{k_0-1}$ and $\sum_{t=k_0+1}^{k_0+l}$ can be replaced by $\sum_{t=l}^{-1}$ and $\sum_{t=1}^{l}$. The main problem is that this limiting distribution is not free of the underlying DGP, hence constructing a confidence interval is not feasible. In previous change point estimation studies, the shrinking break assumption is required to make the limiting distribution independent of the underlying DGP. However, in the current setup, the break magnitude $\|\Sigma_2 - \Sigma_1\|$ is fixed and it is unreasonable to assume $\|\Sigma_2 - \Sigma_1\| \to 0$ as $T \to \infty$. In fact, feasible inference procedure without the shrinking break assumption is an open question. We conjecture that bootstrap is one possible solution and leave this for future research.

Remark 8 *Bai (2010) also considers a fixed magnitude for the break. The difference between our result and Bai (2010) is that our random walk is not necessarily Gaussian. This is because the dimension of* y_t*,* $(r+q_1)^2$*, is fixed and* y_{jt} *and* y_{kt} *are not independent for* $j \neq k$*. In contrast, in Bai (2010), the dimension of* e_t*,* N*, goes to infinity and* e_{jt} *and* e_{kt} *are independent for* $j \neq k$ *so that the CLT applies to the weighted sum of* e_{it}*.*

Remark 9 *In some special cases, the limiting distribution of* $\tilde{k} - k_0$ *is one-sided, concentrating on* $l \geq 0$*. For example, if* Λ_0*,* Λ_1 *and* $\Lambda_2 - \Lambda_1$ *are orthogonal to each other and the factors are also orthogonal with each other, then* $[vec(\Sigma_2 - \Sigma_1)]' y_t = 0$ *for all* $t < k_0$*. It follows that* $W(l) > W(0)$ *for all* $l < 0$*, hence* $\arg\min W(l) \geq 0$*.*

Remark 10 *As in Proposition 2 and Theorem 1, Theorem 4 holds with either Assumption 9 or 10.*

Remark 11 *As in Remark 1, when change point estimation is based on* $\tilde{r} = m < r + q_1$*, Theorem 4 holds with* J_0 *replaced by* J_0^m*.*

8 SIMULATIONS

In this section, we perform simulations to confirm our theoretical results and examine various elements that may affect the finite sample performance of our estimators.

8.1 DESIGN

Our design roughly follows that of Bates et al. (2013), with the focus switching from small change to large change and from forecasting to estimating the whole model, i.e., estimating the change point, the number of pre-break and post-break factors and the pre-break and post-break factor spaces.

The data is generated as follows:

$$x_{it} = \begin{cases} f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{1,i} + \sqrt{\theta_1}e_{i,t}, \ if \ 1 \le t \le [\tau_0 T] \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{2,i} + \sqrt{\theta_2}e_{i,t}, \ if \ [\tau_0 T]+1 \le t \le T \end{cases} \text{for } i = 1, ..., N \text{ and } t = 1, ..., T.$$

As discussed in Section 2, in case the number of pre-break and post-break factors is r_1 and r_2 respectively, with $r = \max\{r_1, r_2\}$, f_t and λ_i are always r dimensional vectors. If $r_1 < r_2$, the last $r_2 - r_1$ elements of $\lambda_{1,i}$ are zeros while if $r_1 > r_2$, the last $r_1 - r_2$ elements of $\lambda_{2,i}$ are zeros. θ_1 and θ_2 control the magnitude of noise and here we take $\theta_1 = r_1, \theta_2 = r_2$.

The factors are generated as follows:

$$f_{t,p} = \rho f_{t-1,p} + u_{t,p} \text{ for } t = 2, ..., T \text{ and } p = 1, ..., r,$$

where $u_{t,p}$ is i.i.d. $N(0,1)$ for $t = 2, ..., T$ and $p = 1, ..., r$. For $t = 1$, $f_{1,p}$ is i.i.d. $N(0, \frac{1}{1-\rho^2})$ for $p = 1, ..., r$ so that factors have stationary distributions. The scalar ρ captures the serial correlation of factors.

The idiosyncratic errors are generated as follows:

$$e_{i,t} = \alpha e_{i,t-1} + v_{i,t} \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T.$$

The processes $\{u_{t,p}\}$ and $\{v_{i,t}\}$ are mutually independent with $v_t = (v_{1,t}, ..., v_{N,t})'$ being i.i.d. $N(0, \Omega)$ for $t = 2, ..., T$. For $t = 1$, $e_{\cdot,1} = (e_{1,1}, ..., e_{N,1})'$ is $N(0, \frac{1}{1-\alpha^2}\Omega)$ so that the idiosyncratic errors have stationary distributions. The scalar α captures the serial correlation of the idiosyncratic errors. As in Bates et al. (2013), $\Omega_{ij} = \beta^{|i-j|}$ captures the cross-sectional dependence of the idiosyncratic errors.

We consider three different ways of generating factor loadings corresponding to three different representative setups. The first setup allows both change in the number of factors and partial change in the factor loadings, with $(r_1, r_2) = (3, 5)$ and one factor having stable loadings. In this case, $\lambda_{0,i}$ is independent $N(0, x_i(R_i^2))$ across i. Both $\lambda_{1,i}$ and $\lambda_{2,i}$ are four dimensional vectors. The first two elements of $\lambda_{1,i}$ are independent $N(0, x_i(R_i^2)I_2)$ across i and the last two elements of $\lambda_{1,i}$ are zeros. Also, $\lambda_{2,i}$ is independent $N(0, x_i(R_i^2)I_4)$ across i. Hence the number of pseudo factors in the equivalent representation is $r_1 + r_2 - 1 = 7$. The scalar $x_i(R_i^2)$ is determined so that the regression R^2 of series i is equal to R_i^2.[7] The second setup allows only change in the number of factors, with $(r_1, r_2) = (3, 5)$ and three factors having stable loadings. In this case, $\lambda_{0,i}$ is independent $N(0, x_i(R_i^2)I_3)$ across i. Both $\lambda_{1,i}$ and $\lambda_{2,i}$ are two dimensional vectors, $\lambda_{1,i}$ are zeros while $\lambda_{2,i}$ is independent $N(0, x_i(R_i^2)I_2)$ across i. Hence the number of pseudo factors is 5. The third setup allows only partial change in the factor loadings, with $(r_1, r_2) = (3, 3)$ and one factor having stable loadings. In this case, $\lambda_{0,i}$ is independent $N(0, x_i(R_i^2))$ across i. Both $\lambda_{1,i}$ and $\lambda_{2,i}$ are two dimensional vectors, $\lambda_{1,i}$ is independent $N(0, x_i(R_i^2)I_2)$ across i while $\lambda_{2,i} = (1-a)\lambda_{1,i} + \sqrt{2a - a^2}d_i$, where $a \in [0, 1]$ and d_i is independent $N(0, x_i(R_i^2)I_2)$ across i. Hence the number of pseudo factors is 5 except for $a = 0$. The scalar a captures the magnitude of factor loading changes, with the the ratio of mean squared changes in the factor loadings to the pre-break factor loadings being equal to $\frac{4a}{3}$. We consider $a = 0.2$, 0.6 and 1, which correspond to small, medium and large changes, respectively. Finally, all factor loadings are independent of the factors and the idiosyncratic errors.

For each setup, we consider the benchmark DGP with $(\rho, \alpha, \beta) = (0, 0, 0)$ and homogeneous R^2 and the more empirically relevant DGP with $(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$ and heterogeneous R^2. For homogeneous R^2, $R_i^2 = 0.5$ for all i, which is also considered in Bai and Ng (2002), Ahn and Horenstein (2013) (to name a few) as a benchmark case in evaluating estimators of the number of factors. For heterogeneous R^2, R_i^2 is drawn from $U(0.2, 0.8)$ independently. For each DGP, we consider four configurations of data with $T = 100, 200, 400$ and $N = 100, 200$. To see how the position of the structural change affects the performance of our estimators, we consider $\tau_0 = 0.25$ and 0.5.

[7] $x_i(R_i^2) = \frac{1-\rho^2}{1-\alpha^2} \frac{R_i^2}{1-R_i^2}$

8.2 ESTIMATORS AND RESULTS

The number of pseudo factors in the equivalent model is estimated using IC_{p1} in Bai and Ng (2002) for Setups 1 and 2. For Setup 3, it is estimated using IC_{p1} in case $a = 1$ and IC_{p3} in case $a = 0.2$ and 0.6. The maximum number of factors is $rmax = 12$. Estimating the number of pseudo factors is the first step of our estimation procedure, and the performance of $\tilde{r}$ will affect the performance of $\tilde{k}$, which in turn affect the performance of $\tilde{r}_1$, $\tilde{r}_2$ and the estimated pre-break and post-break factor spaces. Therefore, it is worth discussing the choice of criterion in estimating the number of pseudo factors. As can be seen in the equivalent representation, the pseudo factors induced by structural change are not as strong as factors with stable loadings in the original model[8] because a portion of their elements are zeros and the magnitude of those nonzero elements is small if the magnitude of structural change is small. Consequently, estimators of the number of factors which perform well in the normal case tend to underestimate the number of pseudo factors, while estimators which tend to overestimate in the normal case, perform well in estimating the number of pseudo factors. Moreover, the magnitudes of pseudo factors induced by structural change are not only absolutely smaller, but also relatively smaller, especially when the change point is not close to the middle of the sample. This decreases the applicability of the ER and GR estimators in Ahn and Horenstein (2013), whose performance rely on the factors being of similar magnitude. In our current setup, we found that among IC_{p1}, IC_{p2} in Bai and Ng (2002) and ER, GR in Ahn and Horenstein (2013), on the whole IC_{p1} performs best. Compared to IC_{p3}, IC_{p1} is more robust to serial correlation and heteroskedasticity of the errors, but IC_{p3} has an advantage in case the change point is far from middle or the magnitude of change is medium or small[9]. Since IC_{p1} and IC_{p3} are relatively less conservative, these findings are consistent with the above observations. In addition, we also found that underestimation of the number of pseudo factors deteriorates the performance of $\tilde{k}$ significantly more than overestimation. This is because $\tilde{k}$ is based on the second moment matrix of the estimated pseudo factors, hence underestimation will result in loss of information while overestimation will bring in extra noise. As long as the overestimation is not severe, these extra noise have very limited effect on the performance of $\tilde{k}$. In view of these results, we recommend choosing

[8] All factors in the equivalent model are called pseudo factors, but not all pseudo factors are induced by structural change. Factors with stable loadings in the original model are still present in the equivalent model.

[9] Our comparison here is limited by the experiments performed. A more comprehensive comparison in case the change point is far from middle or the magnitude of structural change is medium or small is left for a future study.

a less conservative criterion in estimating the number of pseudo factors.

The change point is estimated as in equation (10). We restrict $\tilde{k}$ to be in $[r_1, T - r_2]$ to avoid the singular matrix in subsequent estimation of the number of pre-break and post-break factors. This will not significantly affect the distribution of $\tilde{k}$ since the probability that $\tilde{k}$ falls out of $[r_1, T - r_2]$ is extremely small. To save space, we only display the distributions of $\tilde{k}$ for $(N, T) = (100, 100)$. Of course, the performance of $\tilde{k}$ improves as (N, T) increases. Figure 1 is the histogram of $\tilde{k}$ of Setup 1 for $(N, T) = (100, 100)$. Figures 2 and 3 are histograms of $\tilde{k}$ of Setup 3 for $(N, T) = (100, 100)$ with $a = 1$ and 0.2, respectively. Each figure contains four subfigures corresponding to $\tau_0 = 0.25$ and 0.5 for $(\rho, \alpha, \beta) = (0, 0, 0)$ with homogeneous R^2 and $(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$ with heterogeneous R^2. Under each subfigure, we also report the average and standard deviation of $\tilde{r}$ used in obtaining $\tilde{k}$. The number of replications is 1,000.

It is easy to see that in each subfigure the mass is concentrated in a small neighborhood of k_0. In most cases, the frequency that $\tilde{k}$ falls into $(k_0 - 5, k_0 + 5)$ is around 90%. This confirms our theoretical result, $\tilde{k} - k_0 = O_p(1)$. In Setup 3, even when a decreases from 1 to 0.2, the performance deteriorates very little. Comparing the left column with the right column of each figure, we can see that the performance of $\tilde{k}$ deteriorates as τ_0 moves from 0.5 to 0.25. This is because when τ_0 is close to the boundary, some pseudo factors in the equivalent model are weak and hence the PC estimator of these factors is noisy. In Setup 3, based on Theorem 4 and the fact that all factors and loadings are generated independently, it is not difficult to see that these weak factors are in $W(l)$ for $l = -1, -2, ...$, hence $\tilde{k} - k_0$ is likely to be negative. This explains the asymmetry of Figures 2 and 3. Comparing the first row with the second row of each figure, we can see that the performance of $\tilde{k}$ deteriorates for $(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$ with heterogeneous R^2. This is consistent with Theorem 4, since y_t is serial correlated when factors are serial correlated and serial correlation increases the variance of $\sum_{t=k_0+l}^{k_0-1}[vec(\Sigma_2 - \Sigma_1)]'y_t$ and $\sum_{t=k_0+1}^{k_0+l}[vec(\Sigma_2 - \Sigma_1)]'y_t$ for each l.

Based on $\tilde{k}$, we then split the sample and estimate the number of pre-break and post-break factors using IC_{p2} in Bai and Ng (2002) and GR in Ahn and Horenstein (2013), with maxima $rmax_1 = 10$ and $rmax_2 = 10$. The performance of ER is similar and will not be reported. Based on $\tilde{k}$, $\tilde{r}_1$ and $\tilde{r}_2$, we then estimate the pre-break and post-break factors using the principal component method. To evaluate the performance, we calculate the R^2 of the multivariate regression of $\hat{F}_1^{\tilde{r}_1}(\tilde{k})$ on $F_1(\tilde{k})$ and $\hat{F}_2^{\tilde{r}_2}(\tilde{k})$ on $F_2(\tilde{k})$, $R^2_{\hat{F},F} =$

$\frac{\left\|P_{F_1(\tilde{k})}\hat{F}_1^{\tilde{r}_1}(\tilde{k})\right\|^2+\left\|P_{F_2(\tilde{k})}\hat{F}_2^{\tilde{r}_2}(\tilde{k})\right\|^2}{\left\|\hat{F}_1^{\tilde{r}_1}(\tilde{k})\right\|^2+\left\|\hat{F}_2^{\tilde{r}_2}(\tilde{k})\right\|^2}$. Theorem 3 states that $R^2_{\hat{F},F}$ should be close to one if N and T are large.

Tables 1-3 report the percentage of underestimation and overestimation of $\tilde{r}_1$, $\tilde{r}_2$ and averages of $R^2_{\hat{F},F}$ over 1,000 replications. x/y denotes that the frequency of underestimation and overestimation is $x\%$ and $y\%$ respectively. On the whole, the performance of IC_{p2} and GR are similar. If we choose the better one in each case, the performance of $\tilde{r}_1$ and $\tilde{r}_2$ behave quite well and in most cases close to the their correspondents based on the true change point k_0. For Setups 1 and 3, $(N, T) = (100, 200)$ is large enough to guarantee good performance in all cases. For the case $\tau_0 = 0.5$, $(N, T) = (100, 100)$ is large enough. Note that for Setup 3, even with a small magnitude of change $a = 0.2$, $\tilde{r}_1$ and $\tilde{r}_2$ still perform well. For Setup 2, $(N, T) = (100, 200)$ is large enough in all cases, except for the case with $\rho = 0.5$. The performance of $R^2_{\tilde{F},F}$ is good for all cases.

Comparing the results of $\tau_0 = 0.5$ with $\tau_0 = 0.25$ and $\rho = 0$ with $\rho = 0.5$ in each table, we can see that the deterioration pattern is in accord with that of $\tilde{k}$. This is not surprising since in the current setup, the estimation error in $\tilde{k}$ is the main cause of misestimating $\tilde{r}_1$ and $\tilde{r}_2$. For $\tilde{r}_1$, underestimation of k_0 decreases the size of the pre-break subsample while overestimation increases the tendency of overestimating r_1. Comparing Tables 2 and 3, we can see that underestimation is less harmful. Finally, it is worth noting that there is still room for improvement of finite sample performance of $\tilde{r}_1$, $\tilde{r}_2$, either through improving the performance of $\tilde{k}$ or through choosing an estimator more robust to misspecification of change point among all estimators of the number of factors in the literature.

9 CONCLUSIONS

This paper studied the identification and estimation of a large dimensional factor model with a single large structural change. Both factor loadings and number of factors are allowed to be unstable. We proposed a least squares estimator of the change point and showed that the distance between this estimator and the true change point is $O_p(1)$. The main appeal of this estimator is that it does not require prior information of the number of factors and observability of the factors and it allows for a change in the number of factors. Based on this change point estimator, we are able to dissect the model into two separate stable models and establish consistency of the estimated pre and post-break number of factors and convergence rate of the estimated pre and post-break factor space. These results provide the foundation for subsequent analysis and applications.

A natural step is to derive the limiting distribution of the estimated factors, factor loadings and common components as in Bai (2003). It will also be rewarding to further improve the finite sample performance of our change point estimator. In addition, following the methods in Bai and Perron (1998), it will be straightforward to extend our results to the case with multiple changes. Many other issues are also on the agenda. For example, what are the asymptotic properties of the estimated change point, estimated number of factors and estimated factors when the factor process is I(1)?

Acknowledgements

Chapter 2 is based on the working paper Baltagi, Kao and Wang (2015).

References

[1] Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.

[2] Amengual, D., Watson, M.W., 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. Journal of Business and Economic Statistics 25, 91–96.

[3] Bai, J., 1994. Least squares estimation of shift in linear processes. Journal of Time Series Analysis 15, 453–472.

[4] Bai, J., 1996. An inequality for vector-valued martingales and its applications. MIT Working Paper No. 96-16.

[5] Bai, J., 1997. Estimation of a change point in multiple regression models. Review of Economics and Statistics 79, 551–563.

[6] Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–173.

[7] Bai, J., 2010. Common breaks in means and variances for panel data. Journal of Econometrics 157, 78–92.

[8] Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

[9] Bai, J., Ng, S., 2006. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. Econometrica 74, 1133-1150.

[10] Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics 25, 52-60.

[11] Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

[12] Baltagi, B.H., Kao, C.,, Wang, F., 2015. Identification and estimation of a large factor model with structural instability. Working Paper, Syracuse University.

[13] Banerjee, A., Marcellino, M., Masten I., 2008. Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. Emerald Group Publishing Limited 3, 149-194.

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

[15] Bernanke B., Boivin, J., Eliasz, P., 2005. Factor augmented vector autoregression and the analysis of monetary policy. Quarterly Journal of Economics 120, 387–422.

[16] Breitung, J., Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. Journal of Econometrics 163, 71–84.

[17] Chen, L., 2015. Estimating the common break date in large factor models. Economics Letters 131, 70-74.

[18] Chen, L., Dolado, J., Gonzalo, J., 2014. Detecting big structural breaks in large factor models. Journal of Econometrics 180, 30–48.

[19] Cheng, X., Liao, Z., Schorfheide, F., 2015. Shrinkage estimation of high-dimensional factor models with structural instabilities. Review of Economic Studies, forthcoming.

[20] Corradi, V., Swanson, N.R., 2014. Testing for structural stability of factor augmented forecasting models. Journal of Econometrics 182, 100–118.

[21] Fazekas, I., Klesov, O., 2001. A general approach to the strong law of large numbers. Theory of Probability & Its Applications 45, 436-449.

[22] Hajek, J., Renyi, A., 1955. Generalization of an inequality of Kolmogorov. Acta Math. Acad. Sci. Hungar 6, 281-283.

[23] Han, X., Inoue, A., 2015. Tests for parameter instability in dynamic factor models. Econometric Theory 31, 1117-1152.

[24] Hallin, M., Liska, R., 2007. Determining the number of factors in the generalized dynamic factor model. Journal of American Statistical Association 102, 603–617.

[25] Leeb, H., Pötscher, B.M., 2005. Model selection and inference: Facts and fiction. Econometric Theory 21, 21-59.

[26] Lewbel, A., 1991. The rank of demand systems: theory and nonparametric Estimation. Econometrica 59, 711–730.

[27] Moon, H.R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83, 1543-1579.

[28] Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447–1479.

[29] Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economics and Statistics 92, 1004–1016.

[30] Ross, S., 1976. The arbitrage theory of capital asset pricing. Journal of Finance 13, 341–360.

[31] Stock, J.H., Watson, M.W., 2002a. Forecasting using principal components from a large number of predictors. Journal of American Statistical Association 97, 1167–1179.

[32] Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20, 147-162.

[33] Stock, J.H., Watson, M.W., 2009. Forecasting in dynamic factor models subject to structural instability. In: Hendry, D.F., Castle, J., Shephard, N. (Eds.), The Methodology and practice of econometrics: A festschrift in honour of David F. Hendry. Oxford University Press, 173–205.

APPENDIX

A HAJEK-RENYI INEQUALITY

Hajek-Renyi inequality is a powerful and almost indispensible tool for calculating the stochastic order of sup-type terms. For a sequence of independent random variables $\{x_t, t = 1, ...\}$ with zero mean and finite variance, Hajek and Renyi (1955) proved that for any integers m and T,

$$P(\sup_{m \le k \le T} c_k \left| \sum_{t=1}^{k} x_t \right| > M) \le \frac{1}{M^2}(c_m^2 \sum_{t=1}^{m} \sigma_t^2 + \sum_{t=m+1}^{T} c_t^2 \sigma_t^2), \tag{A-1}$$

where $\{c_k, k = 1, ...\}$ is a sequence of nonincreasing positive numbers and $\mathbb{E}x_t^2 = \sigma_t^2$. The Hajek-Renyi inequality was extended to various settings, including martingale difference, martingale, mixingale, linear process and vector-valued martingale, see Bai (1996). From expression (A-1), it is easy to see that if σ_t^2 is constant over time,

$$P(\sup_{m \le k \le T} \left| \frac{1}{k} \sum_{t=1}^{k} x_t \right| > M) \le \frac{2\sigma^2}{M^2} \frac{1}{m},$$

hence when $m = 1$, $\sup_{1 \le k \le T} \left| \frac{1}{k} \sum_{t=1}^{k} x_t \right| = O_p(1)$ as $T \to \infty$ and when $m = [T\tau]$ for $\tau \in (0, 1)$, $\sup_{m \le k \le T} \left| \frac{1}{k} \sum_{t=1}^{k} x_t \right| = O_p(\frac{1}{\sqrt{T}})$ as $T \to \infty$; and

$$P(\sup_{m \le k \le T} \left| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} x_t \right| > M) \le \frac{\sigma^2}{M^2}(1 + \sum_{k=m+1}^{T} \frac{1}{k}),$$

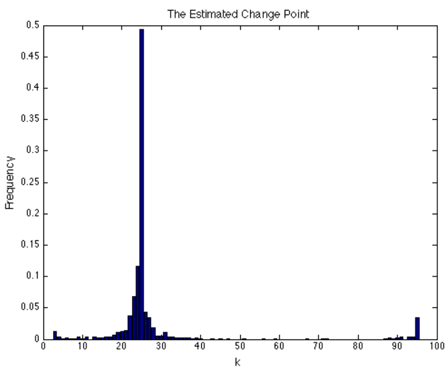
hence when $m = 1$, $\sup_{1 \le k \le T} \left| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} x_t \right| = O_p(\sqrt{\log T})$ as $T \to \infty$ since $\sum_{k=1}^{T} \frac{1}{k} - \log T$ converges to the Euler constant and when $m = [T\tau]$ for $\tau \in (0, 1)$, $\sup_{m \le k \le T} \left| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} x_t \right| = O_p(1)$ as $T \to \infty$ since $\sum_{k=m+1}^{T} \frac{1}{k} = \sum_{k=1}^{T} \frac{1}{k} - \sum_{k=1}^{T\tau} \frac{1}{k} \to \log T - \log T\tau = \log \frac{1}{\tau}$. The last result also can be obtained from the functional central limit theorem.

B SOME NOTATION AND CALCULATION

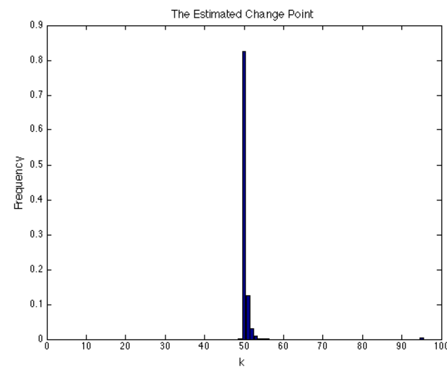
By symmetry, it suffices to study the case $k \le k_0$. To study the asymptotic properties of the change point estimator, we will first decompose the estimation error of pseudo factors and the least squares criterion function $\tilde{S}(k)$.

Define V_{NT} as the diagonal matrix of the first $r + q_1$ largest eigenvalues of $\frac{1}{NT}XX'$ in decreasing order and $\tilde{G}$ as $\sqrt{T}$ times the corresponding eigenvector matrix, V as the

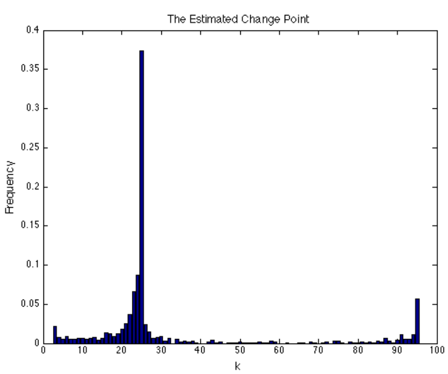
Figure 1: Histogram of $\tilde{k}$ for $(N, T) = (100, 100), (r_1, r_2, r + q_1) = (3, 5, 7)$



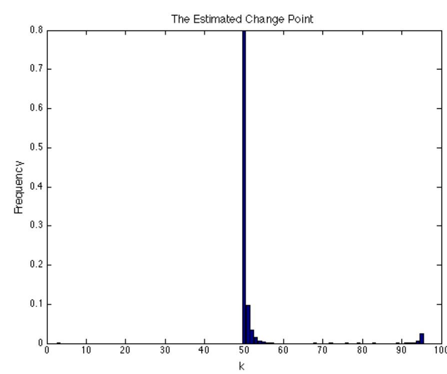
$(\rho, \alpha, \beta) = (0, 0, 0)$, homogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 5.68$, $sd(\tilde{r}) = 0.60$



$(\rho, \alpha, \beta) = (0, 0, 0)$, homogeneous R^2, $\tau_0 = 0.5$, $ave(\tilde{r}) = 6.85$, $sd(\tilde{r}) = 0.38$



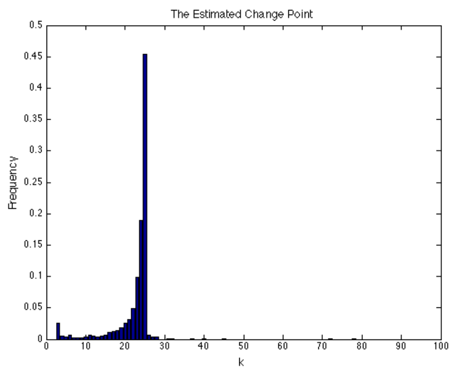
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 5.75$, $sd(\tilde{r}) = 0.58$



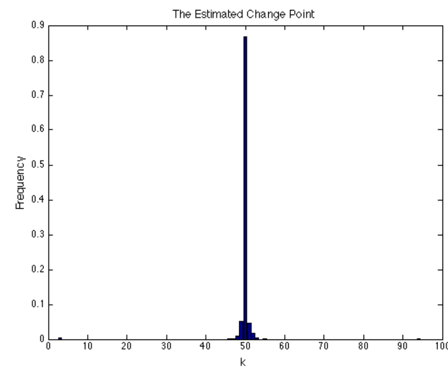
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.5$, $ave(\tilde{r}) = 6.74$, $sd(\tilde{r}) = 0.48$

Notes: ρ, α and β denote factor AR(1) coefficient, error term AR(1) coefficient and error term cross-sectional correlation respectively. $ave(\tilde{r})$ and $sd(\tilde{r})$ denote average and standard deviation of estimated number of pseudo factors that are used to estimate the change point respectively.

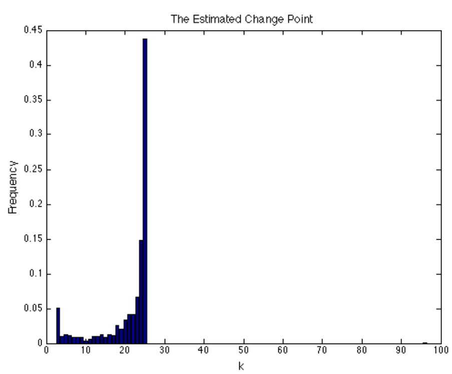
Figure 2: Histogram of $\tilde{k}$ for $(N, T) = (100, 100), (r_1, r_2, r + q_1) = (3, 3, 5), a = 1$



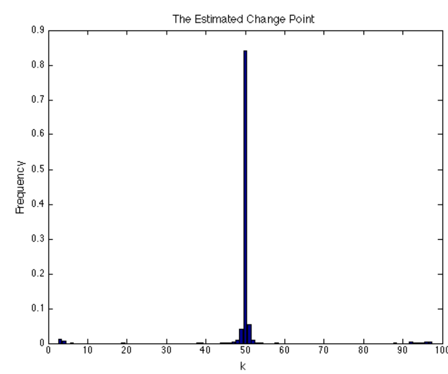
$(\rho, \alpha, \beta) = (0, 0, 0)$, homogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 4.51$, $sd(\tilde{r}) = 0.56$



$(\rho, \alpha, \beta) = (0, 0, 0)$, homogeneous R^2, $\tau_0 = 0.5$, $ave(\tilde{r}) = 5.00$, $sd(\tilde{r}) = 0$



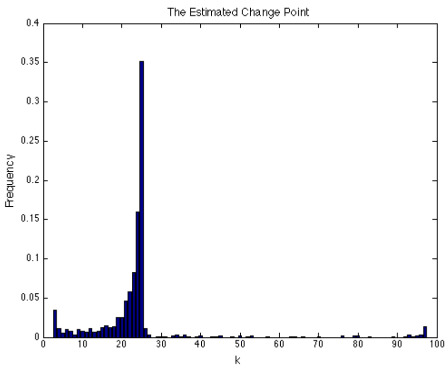
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 4.86$, $sd(\tilde{r}) = 0.35$



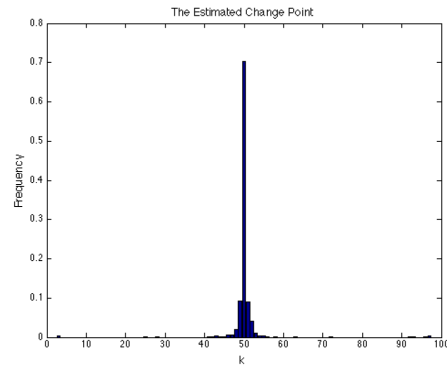
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.5$, $ave(\tilde{r}) = 5.00$, $sd(\tilde{r}) = 0$

Notes: ρ, α and β denote factor AR(1) coefficient, error term AR(1) coefficient and error term cross-sectional correlation respectively. $ave(\tilde{r})$ and $sd(\tilde{r})$ denote average and standard deviation of estimated number of pseudo factors that are used to estimate the change point respectively.

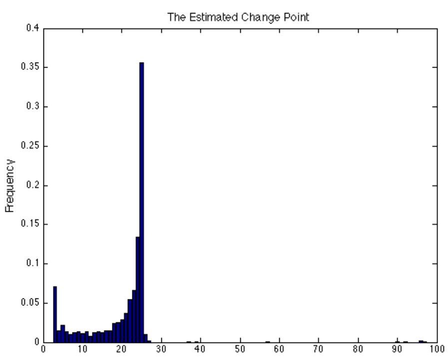
Figure 3: Histogram of $\tilde{k}$ for $(N, T) = (100, 100), (r_1, r_2, r + q_1) = (3, 3, 5), a = 0.2$



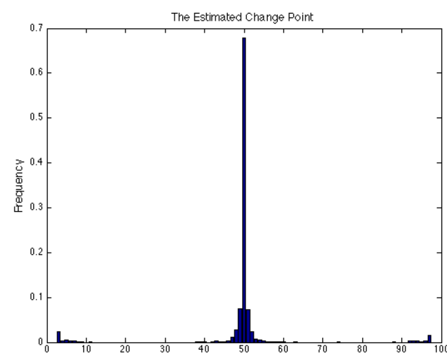
$(\rho, \alpha, \beta) = (0, 0, 0)$, homogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 4.27$, $sd(\tilde{r}) = 0.60$



$\tau_0 = 0.5$, $(\rho, \alpha, \beta) = (0, 0, 0)$, homogeneous R^2, $ave(\tilde{r}) = 4.85$, $sd(\tilde{r}) = 0.36$



$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 5.60$, $sd(\tilde{r}) = 1.17$



$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.5$, $ave(\tilde{r}) = 5.94$, $sd(\tilde{r}) = 1.08$

Notes: ρ, α and β denote factor AR(1) coefficient, error term AR(1) coefficient and error term cross-sectional correlation respectively. $ave(\tilde{r})$ and $sd(\tilde{r})$ denote average and standard deviation of estimated number of pseudo factors that are used to estimate the change point respectively.

Table 1: Estimated number of pre-break and post-break factors and estimated factor space for setup 1 with $r_1 = 3, r_2 = 5, r + q_1 = 7$

N	T	$\tau_0 = 0.25$					$\tau_0 = 0.5$				
		IC_{p2}		GR			IC_{p2}		GR		
		$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_1$	$\tilde{r}_2$	$R^2_{\tilde{F},F}$	$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_1$	$\tilde{r}_2$	$R^2_{\tilde{F},F}$
$\rho = 0, \alpha = 0, \beta = 0$, homogeneous R^2											
100	100	4/8	2/2	11/7	5/1	0.94	0/0	13/0	0/1	2/0	0.96
100	200	0/0	0/0	0/0	0/0	0.95	0/0	0/0	0/0	0/0	0.96
200	200	0/0	0/0	0/0	0/0	0.98	0/0	0/0	0/0	0/0	0.98
200	400	0/0	0/0	0/0	0/0	0.98	0/0	0/0	0/0	0/0	0.98
$\rho = 0.5, \alpha = 0.2, \beta = 0.2$, heterogeneous R^2											
100	100	3/13	2/3	23/4	5/2	0.95	0/4	8/1	1/2	10/0	0.97
100	200	0/2	0/0	2/0	0/1	0.96	0/0	0/0	0/0	0/0	0.97
200	200	0/1	0/3	2/0	0/1	0.98	0/0	0/0	0/0	0/0	0.99
200	400	0/0	0/0	0/0	0/0	0.98	0/0	0/0	0/0	0/0	0.99

Notes: Number of factors in each regime is estimated using IC_{p2} in Bai and Ng (2002) and GR in Ahn and Horenstein (2013). x/y denotes the frequency of underestimation and overestimation is $x\%$ and $y\%$. ρ, α and β denote factor AR(1) coefficient, error term AR(1) coefficient and error term cross-sectional correlation respectively.

Table 2: Estimated number of pre-break and post-break factors and estimated factor space for setup 2 with $r_1 = 3, r_2 = 5, r + q_1 = 5$

N	T	$\tau_0 = 0.25$					$\tau_0 = 0.5$				
		IC_{p2}		GR			IC_{p2}		GR		
		$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_1$	$\tilde{r}_2$	$R^2_{\tilde{F},F}$	$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_1$	$\tilde{r}_2$	$R^2_{\tilde{F},F}$
$\rho = 0, \alpha = 0, \beta = 0$, homogeneous R^2											
100	100	3/41	15/6	9/39	29/0	0.91	0/10	18/2	0/9	12/0	0.96
100	200	0/6	2/1	0/6	5/0	0.95	0/2	1/0	0/1	1/0	0.96
200	200	0/6	2/0	0/5	4/0	0.97	0/1	0/0	0/1	0/0	0.98
200	400	0/1	1/0	0/1	1/0	0.98	0/0	0/0	0/0	0/0	0.98
$\rho = 0.5, \alpha = 0.2, \beta = 0.2$, heterogeneous R^2											
100	100	1/68	20/14	10/59	46/0	0.89	0/26	13/6	1/20	30/0	0.96
100	200	0/27	5/4	2/22	13/0	0.94	0/6	1/2	0/5	4/0	0.97
200	200	0/31	4/5	1/24	14/0	0.95	0/7	1/1	0/6	5/0	0.98
200	400	0/7	1/1	0/5	4/0	0.98	0/2	0/0	0/1	1/0	0.99
$\rho = 0, \alpha = 0.2, \beta = 0.2$, heterogeneous R^2											
100	100	1/43	11/7	9/38	28/0	0.91	0/11	9/2	0/9	12/0	0.96
100	200	0/6	1/1	0/6	4/0	0.96	0/2	0/0	0/1	1/0	0.97
200	200	0/9	1/0	0/5	4/0	0.98	0/1	0/0	0/0	0/0	0.98
200	400	0/1	0/0	0/1	1/0	0.98	0/0	0/0	0/0	0/0	0.98

Notes: Number of factors in each regime is estimated using IC_{p2} in Bai and Ng (2002) and GR in Ahn and Horenstein (2013). x/y denotes the frequency of underestimation and overestimation is $x\%$ and $y\%$. ρ, α and β denote factor AR(1) coefficient, error term AR(1) coefficient and error term cross-sectional correlation respectively.

Table 3: Estimated number of pre-break and post-break factors and estimated factor space for setup 3 with $r_1 = 3, r_2 = 3, r + q_1 = 5$

N	T	$\tau_0 = 0.25$					$\tau_0 = 0.5$				
		IC_{p2}		GR			IC_{p2}		GR		
		$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_1$	$\tilde{r}_2$	$R^2_{\tilde{F},F}$	$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_1$	$\tilde{r}_2$	$R^2_{\tilde{F},F}$
$\rho = 0, \alpha = 0, \beta = 0$, homogeneous R^2, $a = 1$											
100	100	5/4	0/1	14/0	0/1	0.97	0/0	0/0	0/0	0/0	0.97
100	200	0/0	0/0	1/0	0/0	0.97	0/0	0/0	0/0	0/0	0.97
200	200	0/0	0/0	0/0	0/0	0.98	0/0	0/0	0/0	0/0	0.99
200	400	0/0	0/0	0/0	0/0	0.99	0/0	0/0	0/0	0/0	0.99
$\rho = 0.5, \alpha = 0.2, \beta = 0.2$, heterogeneous R^2, $a = 1$											
100	100	3/9	0/8	27/0	0/4	0.97	1/4	0/4	2/1	1/2	0.97
100	200	0/2	0/4	4/0	0/2	0.98	0/1	0/0	0/0	0/0	0.98
200	200	0/1	0/3	2/0	0/2	0.99	0/0	0/0	0/0	0/0	0.99
200	400	0/0	0/1	1/0	0/1	0.99	0/0	0/0	0/0	0/0	0.99
$\rho = 0, \alpha = 0, \beta = 0$, homogeneous R^2, $a = 0.6$											
100	100	4/3	0/1	12/0	0/0	0.97	0/0	0/0	0/0	0/0	0.97
100	200	0/0	0/0	1/0	0/0	0.97	0/0	0/0	0/0	0/0	0.97
200	200	0/0	0/0	0/0	0/0	0.99	0/0	0/0	0/0	0/0	0.99
200	400	0/0	0/0	0/0	0/0	0.99	0/0	0/0	0/0	0/0	0.99
$\rho = 0.5, \alpha = 0.2, \beta = 0.2$, heterogeneous R^2, $a = 0.6$											
100	100	3/9	0/6	26/0	0/3	0.98	1/2	0/3	2/2	2/2	0.98
100	200	0/2	0/3	3/0	0/1	0.98	0/1	0/1	0/0	0/0	0.98
200	200	0/1	0/3	2/0	0/1	0.99	0/0	0/0	0/0	0/0	0.99
200	400	0/0	0/1	1/0	0/1	0.99	0/0	0/0	0/0	0/0	0.99
$\rho = 0, \alpha = 0, \beta = 0$, homogeneous R^2, $a = 0.2$											
100	100	5/8	0/1	18/0	2/0	0.97	0/0	0/0	0/0	1/0	0.97
100	200	2/5	3/7	10/0	16/0	0.97	0/1	1/0	2/0	1/0	0.97
200	200	0/0	0/0	1/0	0/0	0.99	0/0	0/0	0/0	0/0	0.99
200	400	0/0	0/0	0/0	0/0	0.99	0/0	0/0	0/0	0/0	0.99
$\rho = 0.5, \alpha = 0.2, \beta = 0.2$, heterogeneous R^2, $a = 0.2$											
100	100	5/13	0/0	33/0	0/0	0.98	1/2	1/2	3/0	2/0	0.98
100	200	1/3	0/0	7/0	4/0	0.98	0/0	0/0	0/0	1/0	0.98
200	200	0/2	0/0	3/0	0/0	0.99	0/0	0/0	0/0	0/0	0.99
200	400	0/0	0/0	1/0	0/0	0.99	0/0	0/0	0/0	0/0	0.99

Notes: Number of factors in each regime is estimated using IC_{p2} in Bai and Ng (2002) and GR in Ahn and Horenstein (2013). x/y denotes the frequency of underestimation and overestimation is $x\%$ and $y\%$. ρ, α, β and a denote factor AR(1) coefficient, error term AR(1) coefficient and error term cross-sectional correlation and break magnitude respectively.

diagonal matrix of eigenvalues of $\Sigma_\Gamma^{\frac{1}{2}}\Sigma_G\Sigma_\Gamma^{\frac{1}{2}}$ and Φ as the corresponding eigenvector matrix, $J = \frac{\Gamma'\Gamma}{N}\frac{G'\tilde{G}}{T}V_{NT}^{-1}$, $J_0 = \Sigma_\Gamma^{\frac{1}{2}}\Phi V^{-\frac{1}{2}}$. By definition, $\frac{1}{NT}XX'\tilde{G}V_{NT}^{-1} = \tilde{G}$. Plug in $X = G\Gamma' + E$, we have $\tilde{G} - GJ = \frac{1}{NT}(G\Gamma'E'\tilde{G} + E\Gamma G'\tilde{G} + EE'\tilde{G})V_{NT}^{-1}$ and

$$\tilde{g}_t - J'g_t = V_{NT}^{-1}(\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\xi_{st}),$$

where $\zeta_{st} = \frac{e_s'e_t}{N} - \gamma_N(s,t)$, $\eta_{st} = \frac{g_s'\Gamma'e_t}{N}$and $\xi_{st} = \frac{g_t'\Gamma'e_s}{N}$.

Next, define

$$\begin{aligned} z_t &= vec(\tilde{g}_t\tilde{g}_t' - J_0'g_tg_t'J_0) \\ &= vec[(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'] + vec[(\tilde{g}_t - J'g_t)g_t'J] \\ &\quad + vec[J'g_t(\tilde{g}_t - J'g_t)'] + vec[(J' - J_0')g_tg_t'(J' - J_0')'] \\ &\quad + vec[(J' - J_0')g_tg_t'J_0] + vec[J_0'g_tg_t'(J' - J_0')']. \end{aligned} \quad \text{(A-2)}$$

It follows that

$$\begin{aligned} vec(\tilde{g}_t\tilde{g}_t') &= vec(\Sigma_1) + y_t + z_t \text{ for } t \leq k_0, \\ vec(\tilde{g}_t\tilde{g}_t') &= vec(\Sigma_2) + y_t + z_t \text{ for } t > k_0, \end{aligned} \quad \text{(A-3)}$$

where Σ_1, Σ_2 and y_t are defined in Section 7.

For $k \leq k_0$,

$$vec(\tilde{\Sigma}_1) = vec(\Sigma_1) + \frac{1}{k}\sum_{t=1}^{k} y_t + \frac{1}{k}\sum_{t=1}^{k} z_t, \quad \text{(A-4)}$$

$$\begin{aligned} vec(\tilde{\Sigma}_2) &= vec(\Sigma_1) + \frac{T-k_0}{T-k}[vec(\Sigma_2) - vec(\Sigma_1)] \\ &\quad + \frac{1}{T-k}\sum_{t=k+1}^{T} y_t + \frac{1}{T-k}\sum_{t=k+1}^{T} z_t \\ &= \frac{k_0-k}{T-k}[vec(\Sigma_1) - vec(\Sigma_2)] + vec(\Sigma_2) \\ &\quad + \frac{1}{T-k}\sum_{t=k+1}^{T} y_t + \frac{1}{T-k}\sum_{t=k+1}^{T} z_t. \end{aligned} \quad \text{(A-5)}$$

Define

$$a_k = \frac{T-k_0}{T-k}[vec(\Sigma_2) - vec(\Sigma_1)], b_k = \frac{k_0-k}{T-k}[vec(\Sigma_1) - vec(\Sigma_2)], \quad \text{(A-6)}$$

$$\bar{y}_{1k} = \frac{1}{k}\sum\nolimits_{t=1}^{k} y_t,\ \bar{y}_{2k} = \frac{1}{T-k}\sum\nolimits_{t=k+1}^{T} y_t, \tag{A-7}$$

$$\bar{z}_{1k} = \frac{1}{k}\sum\nolimits_{t=1}^{k} z_t,\ \bar{z}_{2k} = \frac{1}{T-k}\sum\nolimits_{t=k+1}^{T} z_t. \tag{A-8}$$

It follows that

$$\begin{aligned} vec(\tilde{\Sigma}_1) &= vec(\Sigma_1) + \bar{y}_{1k} + \bar{z}_{1k}, \\ vec(\tilde{\Sigma}_2) &= vec(\Sigma_1) + a_k + \bar{y}_{2k} + \bar{z}_{2k} = vec(\Sigma_2) + b_k + \bar{y}_{2k} + \bar{z}_{2k}, \end{aligned} \tag{A-9}$$

and for $k < k_0$,

$$\begin{aligned} & \tilde{S}(k) \\ = \ & \sum\nolimits_{t=1}^{k}(y_t + z_t - \bar{y}_{1k} - \bar{z}_{1k})'(y_t + z_t - \bar{y}_{1k} - \bar{z}_{1k}) \\ & + \sum\nolimits_{t=k+1}^{k_0}(y_t + z_t - \bar{y}_{2k} - \bar{z}_{2k} - a_k)'(y_t + z_t - \bar{y}_{2k} - \bar{z}_{2k} - a_k) \\ & + \sum\nolimits_{t=k_0+1}^{T}(y_t + z_t - \bar{y}_{2k} - \bar{z}_{2k} - b_k)'(y_t + z_t - \bar{y}_{2k} - -b_k) \\ = \ & (k_0 - k)a_k' a_k + (T - k_0)b_k' b_k + \sum\nolimits_{t=1}^{T}(y_t + z_t)'(y_t + z_t) \\ & -k(\bar{y}_{1k} + \bar{z}_{1k})'(\bar{y}_{1k} + \bar{z}_{1k}) - (T-k)(\bar{y}_{2k} + \bar{z}_{2k})'(\bar{y}_{2k} + \bar{z}_{2k}) \\ & -2a_k' \sum\nolimits_{t=k+1}^{k_0}(y_t + z_t - \bar{y}_{2k} - \bar{z}_{2k}) \\ & -2b_k' \sum\nolimits_{t=k_0+1}^{T}(y_t + z_t - \bar{y}_{2k} - \bar{z}_{2k}), \end{aligned} \tag{A-10}$$

$$
\begin{aligned}
& \tilde{S}(k) - \tilde{S}(k_0) \\
= \; & (k_0 - k)a_k' a_k \\
& +(T - k_0)b_k' b_k \\
& -\{\frac{1}{k}[\sum\nolimits_{t=1}^{k}(y_t + z_t)]'[\sum\nolimits_{t=1}^{k}(y_t + z_t)] - \frac{1}{k_0}[\sum\nolimits_{t=1}^{k_0}(y_t + z_t)]'[\sum\nolimits_{t=1}^{k_0}(y_t + z_t)]\} \\
& -\{\frac{1}{T-k}[\sum\nolimits_{t=k+1}^{T}(y_t + z_t)]'[\sum\nolimits_{t=k+1}^{T}(y_t + z_t)] \\
& -\frac{1}{T-k_0}[\sum\nolimits_{t=k_0+1}^{T}(y_t + z_t)]'[\sum\nolimits_{t=k_0+1}^{T}(y_t + z_t)]\} \\
& -2a_k' \sum\nolimits_{t=k+1}^{k_0}(y_t + z_t) \\
& -2b_k' \sum\nolimits_{t=k_0+1}^{T}(y_t + z_t) \\
& +2[(k_0 - k)a_k + (T - k_0)b_k]'(\bar{y}_{2k} + \bar{z}_{2k}) \\
= \; & A^* + B^* + C^* + D^* + E^* + F^* + G^*. \qquad \text{(A-11)}
\end{aligned}
$$

C PROOF OF PROPOSITION 1

Proof. In Assumption 1, Σ_F is assumed to be positive definite, hence $A\Sigma_F A'$ and $B\Sigma_F B'$ are both positive semidefinite. For any $r+q_1$ dimensional vector v, if $v'\Sigma_G v = \tau_0 v' A\Sigma_F A' v + (1-\tau_0)v' B\Sigma_F B' v = 0$, it follows that $v' A\Sigma_F A' v = 0$ and $v' B\Sigma_F B' v = 0$. Again because Σ_F is positive definite, this implies $A'v = 0$ and $B'v = 0$. Plug in A, it follows that the first r elements of v are zero. Plug in B, it follows that the last q_1 elements of v are zero. These together imply that $v = 0$ and consequently Σ_G is positive definite. ■

D PROOF OF PROPOSITION 2

Proof. To show $\tilde{\tau} - \tau_0 = o_p(1)$, we need to show for any $\epsilon > 0$ and any $\eta > 0$, $P(|\tilde{\tau} - \tau_0| > \eta) < \epsilon$ as $(N, T) \to \infty$. For the given η, define $D = \{k : (\tau_0 - \eta)T \le k \le (\tau_0 + \eta)T\}$ and D^c as the complement of D, we need to show $P(\tilde{k} \in D^c) < \epsilon$.

$\tilde{k} = \arg\min \tilde{S}(k)$, hence $\tilde{S}(\tilde{k}) - \tilde{S}(k_0) \le 0$. If $\tilde{k} \in D^c$, then $\min\limits_{k\in D^c} \tilde{S}(k) - \tilde{S}(k_0) \le 0$. This implies $P(\tilde{k} \in D^c) \le P(\min\limits_{k\in D^c} \tilde{S}(k) - \tilde{S}(k_0) \le 0)$, hence it suffices to show for any given $\epsilon > 0$ and $\eta > 0$, $P(\min\limits_{k\in D^c} \tilde{S}(k) - \tilde{S}(k_0) \le 0) < \epsilon$ as $(N, T) \to \infty$.

Suppose $\omega \in \{\omega : \min\limits_{k\in D^c} \tilde{S}(k) - \tilde{S}(k_0) \le 0\}$. For any $k^* \in D^c$, if $\arg\min\limits_{k\in D^c} \tilde{S}(k) - \tilde{S}(k_0) = k^*$, then $\tilde{S}(k^*) - \tilde{S}(k_0) \le 0$, and hence $\frac{\tilde{S}(k^*)-\tilde{S}(k_0)}{|k^*-k_0|} \le 0$. Since $k^* \in D^c$, $\min\limits_{k\in D^c} \frac{\tilde{S}(k)-\tilde{S}(k_0)}{|k-k_0|} \le \frac{\tilde{S}(k^*)-\tilde{S}(k_0)}{|k^*-k_0|}$.

Combined together, we have $\min\limits_{k\in D^c}\frac{\tilde{S}(k)-\tilde{S}(k_0)}{|k-k_0|}\leq 0$. In other words, we proved that for any $k^*\in D^c$, $\min\limits_{k\in D^c}\tilde{S}(k)-\tilde{S}(k_0)\leq 0$ together with $\arg\min\limits_{k\in D^c}\tilde{S}(k)-\tilde{S}(k_0)=k^*$ implies $\min\limits_{k\in D^c}\frac{\tilde{S}(k)-\tilde{S}(k_0)}{|k-k_0|}\leq 0$. Thus $\min\limits_{k\in D^c}\tilde{S}(k)-\tilde{S}(k_0)\leq 0$ implies $\min\limits_{k\in D^c}\frac{\tilde{S}(k)-\tilde{S}(k_0)}{|k-k_0|}\leq 0$. Similarly, $\min\limits_{k\in D^c}\frac{\tilde{S}(k)-\tilde{S}(k_0)}{|k-k_0|}\leq 0$ implies $\min\limits_{k\in D^c}\tilde{S}(k)-\tilde{S}(k_0)\leq 0$. Therefore, $\{\omega:\min\limits_{k\in D^c}\tilde{S}(k)-\tilde{S}(k_0)\leq 0\}=\{\omega:\min\limits_{k\in D^c}\frac{\tilde{S}(k)-\tilde{S}(k_0)}{|k-k_0|}\leq 0\}$.

By symmetry, it suffices to study the case $k<k_0$.

$$
\begin{aligned}
P(\min_{k\in D^c,k<k_0}\tilde{S}(k)-\tilde{S}(k_0) &\leq 0)=P(\min_{k\in D^c,k<k_0}\frac{\tilde{S}(k)-\tilde{S}(k_0)}{|k-k_0|}\leq 0)\\
&\leq P(\min_{k\in D^c,k<k_0}\frac{A^*+B^*}{|k-k_0|}\leq\sup_{k\in D^c,k<k_0}\frac{|C^*|}{|k_0-k|}+\sup_{k\in D^c,k<k_0}\frac{|D^*|}{|k_0-k|}\\
&\quad+\sup_{k\in D^c,k<k_0}\frac{|E^*|}{|k_0-k|}+\sup_{k\in D^c,k<k_0}\frac{|F^*|}{|k_0-k|}+\sup_{k\in D^c,k<k_0}\frac{|G^*|}{|k_0-k|}).
\end{aligned}
$$

We will show the right hand side are dominated by the left hand side.

First consider term A^*+B^*,

$$
\begin{aligned}
\min_{k\in D^c,k<k_0}\frac{A^*+B^*}{|k-k_0|} &\geq \min_{k\in D^c,k<k_0}\frac{A^*}{|k_0-k|}=\min_{k\in D^c,k<k_0}a_k'a_k\\
&= \min_{k\in D^c,k<k_0}(\frac{T-k_0}{T-k})^2[vec(\Sigma_2-\Sigma_1)]'[vec(\Sigma_2-\Sigma_1)]\\
&\geq (1-\tau_0)^2\left\|\Sigma_2-\Sigma_1\right\|^2=(1-\tau_0)^2\left\|J_0\right\|^4\left\|\Sigma_{G,2}-\Sigma_{G,1}\right\|^2.
\end{aligned}
$$

Next consider term C^*,

$$
\begin{aligned}
C^* &= -\{\frac{1}{k}[\sum\nolimits_{t=1}^{k}(y_t+z_t)]'[\sum\nolimits_{t=1}^{k}(y_t+z_t)]-\frac{1}{k_0}[\sum\nolimits_{t=1}^{k_0}(y_t+z_t)]'[\sum\nolimits_{t=1}^{k_0}(y_t+z_t)]\}\\
&= -\frac{k_0-k}{k_0}\frac{1}{k}[\sum\nolimits_{t=1}^{k}(y_t+z_t)]'[\sum\nolimits_{t=1}^{k}(y_t+z_t)]\\
&\quad+2\frac{1}{k_0}[\sum\nolimits_{t=1}^{k}(y_t+z_t)]'[\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)]\\
&\quad+\frac{k_0-k}{k_0}\frac{1}{k_0-k}[\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)]'[\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)].
\end{aligned}
$$

Hence,

$$\begin{aligned}\left|\frac{C^*}{k_0-k}\right| &\le \left|\frac{1}{k_0}\frac{1}{k}[\sum_{t=1}^{k}(y_t+z_t)]'[\sum_{t=1}^{k}(y_t+z_t)]\right| \\ &\quad+\left|2\frac{1}{k_0}\frac{1}{k_0-k}[\sum_{t=1}^{k}(y_t+z_t)]'[\sum_{t=k+1}^{k_0}(y_t+z_t)]\right| \\ &\quad+\left|\frac{1}{k_0}\frac{1}{k_0-k}[\sum_{t=k+1}^{k_0}(y_t+z_t)]'[\sum_{t=k+1}^{k_0}(y_t+z_t)]\right| \\ &= C_1^*+C_2^*+C_3^*.\end{aligned}$$

For C_1^*,

$$\begin{aligned}\sup_{k\in D^c,k<k_0} C_1^* &= \sup_{k\in D^c,k<k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}(y_t+z_t)\right\|^2 \\ &\le \sup_{k\in D^c,k<k_0}\frac{1}{k_0}\frac{1}{k}(\left\|\sum_{t=1}^{k}y_t\right\|^2+\left\|\sum_{t=1}^{k}z_t\right\|^2+2\left\|\sum_{t=1}^{k}y_t\right\|\left\|\sum_{t=1}^{k}z_t\right\|) \\ &\le 2\sup_{k\in D^c,k<k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}y_t\right\|^2+2\sup_{k\in D^c,k<k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}z_t\right\|^2.\end{aligned}$$

By part (1) of Lemma 3, $\sup_{k\in D^c,k<k_0}\left\|\frac{1}{\sqrt{k}}\sum_{t=1}^{k}y_t\right\| = O_p(\sqrt{\log T})$, hence the first term is $O_p(\frac{\log T}{T})$. By part (1) of Lemma 7, the second term is $o_p(1)$, hence $\sup_{k\in D^c,k<k_0} C_1^* = o_p(1)$.

For C_2^*,

$$\begin{aligned}\sup_{k\in D^c,k<k_0} C_2^* &\le 2\sup_{k\in D^c,k<k_0}\frac{1}{k_0}\frac{1}{k_0-k}\left\|\sum_{t=1}^{k}(y_t+z_t)\right\|\left\|\sum_{t=k+1}^{k_0}(y_t+z_t)\right\| \\ &\le 2\sup_{k\in D^c,k<k_0}(\left\|\frac{1}{k_0}\sum_{t=1}^{k}y_t\right\|+\left\|\frac{1}{k_0}\sum_{t=1}^{k}z_t\right\|)(\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}y_t\right\| \\ &\quad+\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}z_t\right\|) \\ &\le 2(\sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0}\sum_{t=1}^{k}y_t\right\|+\sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0}\sum_{t=1}^{k}z_t\right\|) \\ &\quad(\sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}y_t\right\|+\sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}z_t\right\|).\end{aligned}$$

By part (1) of Lemma 3, the first term and the third term are $O_p(\frac{1}{\sqrt{T}})$, and by parts (3) and (5) of Lemma 7, the second term and the fourth term are $o_p(1)$, hence $\sup_{k\in D^c,k<k_0} C_2^* = o_p(1)$.

For C_3^*,

$$\begin{aligned}
\sup_{k\in D^c,k<k_0} C_3^* &= \sup_{k\in D^c,k<k_0} \frac{1}{k_0}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)\right\|^2 \\
&\leq \sup_{k\in D^c,k<k_0} \frac{1}{k_0}\frac{1}{k_0-k}(\left\|\sum\nolimits_{t=k+1}^{k_0} y_t\right\| + \left\|\sum\nolimits_{t=k+1}^{k_0} z_t\right\|)^2 \\
&\leq 2\frac{1}{k_0}\sup_{k\in D^c,k<k_0}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} y_t\right\|^2 + 2\frac{1}{k_0}\sup_{k\in D^c,k<k_0}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} z_t\right\|^2.
\end{aligned}$$

By part (1) of Lemma 3, $\sup_{k\in D^c,k<k_0}\left\|\frac{1}{\sqrt{k_0-k}}\sum_{t=k+1}^{k_0} y_t\right\| = O_p(1)$, hence the first term is $O_p(\frac{1}{T})$. By part (7) of Lemma 7, $\sup_{k\in D^c,k<k_0}\frac{1}{k_0}\frac{1}{|k_0-k|}\left\|\sum_{t=k+1}^{k_0} z_t\right\|^2 = o_p(1)$, hence $\sup_{k\in D^c,k<k_0} C_3^* = o_p(1)$.

Therefore, $\sup_{k\in D^c,k<k_0}\left|\frac{C^*}{k_0-k}\right| \leq \sup_{k\in D^c,k<k_0} C_1^* + \sup_{k\in D^c,k<k_0} C_2^* + \sup_{k\in D^c,k<k_0} C_3^* = o_p(1)$.

Similarly,

$$\begin{aligned}
\left|\frac{D^*}{k_0-k}\right| &\leq \left|\frac{1}{T-k_0}\frac{1}{T-k}[\sum\nolimits_{t=k_0+1}^{T}(y_t+z_t)]'[\sum\nolimits_{t=k_0+1}^{T}(y_t+z_t)]\right| \\
&\quad + \left|2\frac{1}{T-k}\frac{1}{k_0-k}[\sum\nolimits_{t=k_0+1}^{T}(y_t+z_t)]'[\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)]\right| \\
&\quad + \left|\frac{1}{T-k}\frac{1}{k_0-k}[\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)]'[\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)]\right| \\
&= D_1^* + D_2^* + D_3^*.
\end{aligned}$$

$$\begin{aligned}
&\sup_{k\in D^c,k<k_0} D_1^* \\
\leq\ & 2\sup_{k\in D^c,k<k_0}\frac{1}{T-k_0}\frac{1}{T-k}\left\|\sum\nolimits_{t=k_0+1}^{T} y_t\right\|^2 + 2\sup_{k\in D^c,k<k_0}\frac{1}{T-k_0}\frac{1}{T-k}\left\|\sum\nolimits_{t=k_0+1}^{T} z_t\right\|^2 \\
=\ & O_p(\frac{1}{T}) + o_p(1) = o_p(1),
\end{aligned}$$

where the equality follows from part (1) of Lemma 3 and part (9) of Lemma 7.

$$\begin{aligned}
\sup_{k\in D^c,k<k_0} D_2^* &\leq 2(\sup_{k\in D^c,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k_0+1}^{T} y_t\right\| + \sup_{k\in D^c,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k_0+1}^{T} z_t\right\|) \\
&\quad (\sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} y_t\right\| + \sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} z_t\right\|) \\
&= (O_p(\frac{1}{\sqrt{T}}) + o_p(1))(O_p(\frac{1}{\sqrt{T}}) + o_p(1)) = o_p(1),
\end{aligned}$$

where the equality follows from part (1) of Lemma 3 and parts (9) and (5) of Lemma 7.

$$\begin{aligned}
&\sup_{k\in D^c,k<k_0} D_3^* \\
\leq\ & 2\sup_{k\in D^c,k<k_0}\frac{1}{T-k}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} y_t\right\|^2 + 2\sup_{k\in D^c,k<k_0}\frac{1}{T-k}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} z_t\right\|^2 \\
=\ & O_p(\frac{1}{T}) + o_p(1) = o_p(1),
\end{aligned}$$

where the equality follows from part (1) of Lemma 3 and part (7) of Lemma 7.

Therefore, $\sup\limits_{k\in D^c,k<k_0}\left|\frac{D^*}{k_0-k}\right| \leq \sup\limits_{k\in D^c,k<k_0} D_1^* + \sup\limits_{k\in D^c,k<k_0} D_2^* + \sup\limits_{k\in D^c,k<k_0} D_3^* = o_p(1)$.

Next consider term E^*.

$$\begin{aligned}
\sup_{k\in D^c,k<k_0}\left|\frac{E^*}{k_0-k}\right| &= 2\sup_{k\in D^c,k<k_0}\frac{1}{k_0-k}\left|a_k'\sum\nolimits_{t=k+1}^{k_0}(y_t+z_t)\right| \\
&\leq 2\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} y_t\right\| \\
&\quad +2\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D^c,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} z_t\right\|.
\end{aligned}$$

By part (1) of Lemma 3, the first term is $O_p(\frac{1}{\sqrt{T}})$. By part (5) of Lemma 7, the second term is $o_p(1)$. Therefore, $\sup\limits_{k\in D^c,k<k_0}\left|\frac{E^*}{k_0-k}\right| = o_p(1)$.

For term F^*,

$$\begin{aligned}
&\sup_{k\in D^c,k<k_0}\left|\frac{F^*}{k_0-k}\right| \\
\leq\ & 2\sup_{k\in D^c,k<k_0}\frac{\|b_k\|\left\|\sum_{t=k_0+1}^{T}(y_t+z_t)\right\|}{|k_0-k|} \\
\leq\ & 2\left\|\Sigma_1-\Sigma_2\right\|\left\|\frac{1}{T-k_0}\sum\nolimits_{t=k_0+1}^{T} y_t\right\| + 2\left\|\Sigma_1-\Sigma_2\right\|\left\|\frac{1}{T-k_0}\sum\nolimits_{t=k_0+1}^{T} z_t\right\|.
\end{aligned}$$

By part (1) of Lemma 3, the first term is $O_p(\frac{1}{\sqrt{T}})$. By part (9) of Lemma 7, $\left\|\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} z_t\right\| \leq \sup\limits_{k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T} z_t\right\| = o_p(1)$. Therefore, $\sup\limits_{k\in D^c,k<k_0}\left|\frac{F^*}{k_0-k}\right| = o_p(1)$.

For term G^*, note that $(k_0-k)a_k=(T-k_0)b_k$,

$$
\begin{aligned}
\sup_{k\in D^c,k<k_0}\left|\frac{G^*}{k_0-k}\right| &= 4\sup_{k\in D^c,k<k_0}\left|a_k'(\bar{y}_{2k}+\bar{z}_{2k})\right| \\
&\leq 4\sup_{k\in D^c,k<k_0}\frac{T-k_0}{T-k}\left\|\Sigma_2-\Sigma_1\right\|\left\|\frac{1}{T-k}\sum\nolimits_{t=k+1}^{T}(y_t+z_t)\right\| \\
&\leq 4\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D^c,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k+1}^{T}y_t\right\| \\
&\quad +4\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D^c,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k+1}^{T}z_t\right\|.
\end{aligned}
$$

The first term is bounded by

$$
\sup_{k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k+1}^{T}y_t\right\|\leq\frac{1}{1-\tau_0}(\sup_{k<k_0}\frac{1}{T}\left\|\sum\nolimits_{t=k+1}^{k_0}y_t\right\|+\sup_{k>k_0}\frac{1}{T}\left\|\sum\nolimits_{t=k_0+1}^{k}y_t\right\|),
$$

and by part (1) of Lemma 3 this term is $O_p(\frac{1}{\sqrt{T}})$. By part (9) of Lemma 7, the second term is $o_p(1)$. Therefore, $\sup_{k\in D^c,k<k_0}\left|\frac{G^*}{k_0-k}\right|=o_p(1)$. ■

E PROOF OF THEOREM 1

Proof. To show $\tilde{k}-k_0=O_p(1)$, we need to show for any $\epsilon>0$ there exist $M>0$ such that $P(\left|\tilde{k}-k_0\right|>M)<\epsilon$ as $(N,T)\to\infty$. By Proposition 2, for any $\epsilon>0$ and $\min\{\tau_0,1-\tau_0\}>\eta>0$, $P(\tilde{k}\in D^c)<\epsilon$ as $(N,T)\to\infty$. For the given η and M, define $D_M=\{k:(\tau_0-\eta)T\leq k\leq(\tau_0+\eta)T,\ |k-k_0|>M\}$, then $P(\left|\tilde{k}-k_0\right|>M)=P(\tilde{k}\in D^c)+P(\tilde{k}\in D_M)$. Hence it suffices to show that for any $\epsilon>0$ and $\eta>0$, there exist $M>0$ such that $P(\tilde{k}\in D_M)<\epsilon$ as $(N,T)\to\infty$. Again by symmetry, it suffices to study the case $k<k_0$. Similar to the proof of Proposition 2, it suffices to show for any given $\epsilon>0$ and $\eta>0$, there exist $M>0$ such that $P(\min_{k\in D_M,k<k_0}\frac{A^*+B^*}{|k_0-k|}\leq \sup_{k\in D_M,k<k_0}\left|\frac{C^*}{k_0-k}\right|+\sup_{k\in D_M,k<k_0}\left|\frac{D^*}{k_0-k}\right|+\sup_{k\in D_M,k<k_0}\left|\frac{E^*}{k_0-k}\right|+\sup_{k\in D_M,k<k_0}\left|\frac{F^*}{k_0-k}\right|+\sup_{k\in D_M,k<k_0}\left|\frac{G^*}{k_0-k}\right|)<\epsilon$ as $(N,T)\to\infty$.

First consider term A^*+B^*,

$$
\begin{aligned}
\min_{k\in D_M,k<k_0}\frac{A^*+B^*}{|k_0-k|} &= \min_{k\in D_M,k<k_0}a_k'a_k=\min_{k\in D_M,k<k_0}(\frac{T-k_0}{T-k})^2[vec(\Sigma_2-\Sigma_1)]'[vec(\Sigma_2-\Sigma_1)] \\
&\geq (1-\tau_0)^2\left\|\Sigma_2-\Sigma_1\right\|^2=(1-\tau_0)^2\left\|J_0\right\|^4\left\|\Sigma_{G,2}-\Sigma_{G,1}\right\|^2.
\end{aligned}
$$

Next consider term C^*. Similar to the proof of Proposition 2,

$$\sup_{k\in D_M,k<k_0}\left|\frac{C^*}{k_0-k}\right| \leq \sup_{k\in D,k<k_0}\left|\frac{C^*}{k_0-k}\right| \leq \sup_{k\in D,k<k_0} C_1^* + \sup_{k\in D,k<k_0} C_2^* + \sup_{k\in D,k<k_0} C_3^*.$$

For C_1^*,

$$\sup_{k\in D,k<k_0} C_1^* \leq 2\sup_{k\in D,k<k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum\nolimits_{t=1}^{k} y_t\right\|^2 + 2\sup_{k\in D,k<k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum\nolimits_{t=1}^{k} z_t\right\|^2.$$

By part (1) of Lemma 3, $\sup_{k\in D,k<k_0}\left\|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} y_t\right\| = O_p(1)$, hence the first term is $O_p(\frac{1}{T})$. By part (2) of Lemma 7, the second term is $o_p(1)$, hence $\sup_{k\in D,k<k_0} C_1^* = o_p(1)$.

For C_2^*,

$$\begin{aligned}\sup_{k\in D,k<k_0} C_2^* \quad\leq\quad & 2\Big(\sup_{k\in D,k<k_0}\left\|\frac{1}{k_0}\sum\nolimits_{t=1}^{k} y_t\right\| + \sup_{k\in D,k<k_0}\left\|\frac{1}{k_0}\sum\nolimits_{t=1}^{k} z_t\right\|\Big)\\ & \Big(\sup_{k\in D,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} y_t\right\| + \sup_{k\in D,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} z_t\right\|\Big).\end{aligned}$$

By part (1) of Lemma 3, the first term is $O_p(\frac{1}{\sqrt{T}})$, the third term is $O_p(1)$ and by parts (4) and (6) of Lemma 7, the second term and the fourth term are $o_p(1)$. Hence $\sup_{k\in D,k<k_0} C_2^* = o_p(1)$.

For C_3^*,

$$\sup_{k\in D,k<k_0} C_3^* \leq 2\frac{1}{k_0}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} y_t\right\|^2 + 2\frac{1}{k_0}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} z_t\right\|^2.$$

By part (1) of Lemma 3, $\sup_{k\in D,k<k_0}\left\|\frac{1}{\sqrt{k_0-k}}\sum_{t=k+1}^{k_0} y_t\right\| = O_p(\sqrt{\log T})$, hence the first term is $O_p(\frac{\log T}{T})$. By part (8) of Lemma 7, $\sup_{k\in D,k<k_0}\frac{1}{k_0}\frac{1}{|k_0-k|}\left\|\sum_{t=k+1}^{k_0} z_t\right\|^2 = o_p(1)$. Hence $\sup_{k\in D,k<k_0} C_3^* = o_p(1)$. Therefore, $\sup_{k\in D_M,k<k_0}\left|\frac{C^*}{k_0-k}\right| = o_p(1)$.

Similarly,

$$\sup_{k\in D_M,k<k_0}\left|\frac{D^*}{k_0-k}\right| \leq \sup_{k\in D,k<k_0}\left|\frac{D^*}{k_0-k}\right| \leq \sup_{k\in D,k<k_0} D_1^* + \sup_{k\in D,k<k_0} D_2^* + \sup_{k\in D,k<k_0} D_3^*.$$

$$\begin{aligned}
&\sup_{k\in D,k<k_0} D_1^* \\
&\leq\ 2\sup_{k\in D,k<k_0}\frac{1}{T-k_0}\frac{1}{T-k}\left\|\sum\nolimits_{t=k_0+1}^{T} y_t\right\|^2+2\sup_{k\in D,k<k_0}\frac{1}{T-k_0}\frac{1}{T-k}\left\|\sum\nolimits_{t=k_0+1}^{T} z_t\right\|^2 \\
&=\ O_p(\frac{1}{T})+o_p(1)=o_p(1),
\end{aligned}$$

where the equality follows from part (1) of Lemma 3 and part (9) of Lemma 7.

$$\begin{aligned}
\sup_{k\in D,k<k_0} D_2^* &\leq 2(\sup_{k\in D,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k_0+1}^{T} y_t\right\|+\sup_{k\in D,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k_0+1}^{T} z_t\right\|) \\
&\quad (\sup_{k\in D,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} y_t\right\|+\sup_{k\in D,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} z_t\right\|) \\
&= (O_p(\frac{1}{\sqrt{T}})+o_p(1))(O_p(1)+o_p(1))=o_p(1),
\end{aligned}$$

where the equality follows from part (1) of Lemma 3 and parts (9) and (6) of Lemma 7.

$$\begin{aligned}
&\sup_{k\in D,k<k_0} D_3^* \\
&\leq\ 2\sup_{k\in D,k<k_0}\frac{1}{T-k}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} y_t\right\|^2+2\sup_{k\in D,k<k_0}\frac{1}{T-k}\frac{1}{k_0-k}\left\|\sum\nolimits_{t=k+1}^{k_0} z_t\right\|^2 \\
&=\ O_p(\frac{\log T}{T})+o_p(1)=o_p(1),
\end{aligned}$$

where the equality follows from part (1) of Lemma 3 and part (8) of Lemma 7.

Therefore, $\sup_{k\in D_M,k<k_0}\left|\frac{D^*}{k_0-k}\right|=o_p(1)$.

Next consider term E^*.

$$\begin{aligned}
\sup_{k\in D_M,k<k_0}\left|\frac{E^*}{k_0-k}\right| &\leq 2\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D_M,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} y_t\right\| \\
&\quad +2\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D_M,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} z_t\right\|.
\end{aligned}$$

For any given $\delta>0$, $P(2\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D_M,k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} y_t\right\|\geq\delta(1-\tau_0)^2\left\|\Sigma_2-\Sigma_1\right\|^2)$ $=P(\sup_{k\in D_M,k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} y_t\right\|\geq\delta\frac{(1-\tau_0)^2}{2}\left\|\Sigma_2-\Sigma_1\right\|)\leq\frac{C}{M\delta^2}\to 0$ as $M\to\infty$, hence the first term is dominated by $\min_{k\in D_M,k<k_0}\frac{A^*+B^*}{|k_0-k|}$. By part (6) of Lemma 7, the second term is $o_p(1)$. Therefore, $\sup_{k\in D_M,k<k_0}\left|\frac{E^*}{k_0-k}\right|$ is dominated by $\min_{k\in D_M,k<k_0}\frac{A^*+B^*}{|k_0-k|}$ as $M\to\infty$.

For term F^*,

$$\begin{aligned}
& \sup_{k\in D_M, k<k_0} \left| \frac{F^*}{k_0-k} \right| \\
\leq & \sup_{k\in D, k<k_0} \left| \frac{F^*}{k_0-k} \right| \leq 2 \sup_{k\in D, k<k_0} \frac{\|b_k\| \left\| \sum_{t=k_0+1}^{T} (y_t+z_t) \right\|}{|k_0-k|} \\
\leq & 2\|\Sigma_1-\Sigma_2\| \left\| \frac{1}{T-k_0} \sum\nolimits_{t=k_0+1}^{T} y_t \right\| + 2\|\Sigma_1-\Sigma_2\| \left\| \frac{1}{T-k_0} \sum\nolimits_{t=k_0+1}^{T} z_t \right\|.
\end{aligned}$$

By part (1) of Lemma 3, the first term is $O_p(\frac{1}{\sqrt{T}})$. By part (9) of Lemma 7, the second term is $o_p(1)$. Therefore, $\sup_{k\in D_M, k<k_0} \left| \frac{F^*}{k_0-k} \right| = o_p(1)$.

For term G^*,

$$\begin{aligned}
\sup_{k\in D_M, k<k_0} \left| \frac{G^*}{k_0-k} \right| \leq & \sup_{k\in D, k<k_0} \left| \frac{G^*}{k_0-k} \right| \leq 4\|\Sigma_2-\Sigma_1\| \sup_{k\in D, k<k_0} \left\| \frac{1}{T-k} \sum\nolimits_{t=k+1}^{T} y_t \right\| \\
& +4\|\Sigma_2-\Sigma_1\| \sup_{k\in D, k<k_0} \left\| \frac{1}{T-k} \sum\nolimits_{t=k+1}^{T} z_t \right\|.
\end{aligned}$$

The first term is bounded by

$$\sup_{k<k_0} \left\| \frac{1}{T-k} \sum\nolimits_{t=k+1}^{T} y_t \right\| \leq \frac{1}{1-\tau_0} (\sup_{k<k_0} \frac{1}{T} \left\| \sum\nolimits_{t=k+1}^{k_0} y_t \right\| + \sup_{k>k_0} \frac{1}{T} \left\| \sum\nolimits_{t=k_0+1}^{k} y_t \right\|),$$

and by part (1) of Lemma 3 this term is $O_p(\frac{1}{\sqrt{T}})$. By part (9) of Lemma 7, the second term is $o_p(1)$. Therefore, $\sup_{k\in D_M, k<k_0} \left| \frac{G^*}{k_0-k} \right| = o_p(1)$. ■

F PROOF OF COROLLARY 1

Proof. The proof is the same as the proof of Proposition 2 and Theorem 1, except for some slight modification. When $m < r + q_1$, V_{NT}, $\tilde{G}$ and J are replaced by V_{NT}^m, $\tilde{G}^m$ and J^m respectively, where V_{NT} is the diagonal matrix of the first m largest eigenvalues of $\frac{1}{NT}XX'$ in decreasing order and $\tilde{G}^m$ is $\sqrt{T}$ times the corresponding eigenvector matrix and $J^m = \frac{\Gamma'\Gamma}{N} \frac{G'\tilde{G}^m}{T} (V_{NT}^m)^{-1}$. $V_{NT}^m \xrightarrow{p} V^m$, where V^m is $m \times m$ diagonal matrix, containing the first m diagonal elements of V. $\frac{G'\tilde{G}^m}{T}$ contains the first m columns of $\frac{G'\tilde{G}}{T}$, hence $\frac{G'\tilde{G}}{T} \xrightarrow{p} \Sigma_\Gamma^{-\frac{1}{2}} \Phi V^{\frac{1}{2}}$ implies $\frac{G'\tilde{G}^m}{T} \xrightarrow{p} D$ where D contains the first m columns of $\Sigma_\Gamma^{-\frac{1}{2}} \Phi V^{\frac{1}{2}}$. Hence $D(V^m)^{-1}$ contains the first m columns of $\Sigma_\Gamma^{-\frac{1}{2}} \Phi V^{-\frac{1}{2}}$ and it follows that $J^m \xrightarrow{p} J_0^m$ where J_0^m contains the first m columns of J_0. ■

G PROOF OF COROLLARY 2

Proof. For any integer $m > r + q_1$, let $\tilde{G}^m$ be the $T \times m$ matrix that contains $\sqrt{T}$ times the eigenvectors corresponding to the first m eigenvalues of $\frac{1}{NT}XX'$ and V_{NT}^m be the $m \times m$ diagonal matrix that contains the first m eigenvalues. Then let $(\hat{g}_1^m, ..., \hat{g}_T^m)' = \hat{G}^m = \tilde{G}^m V_{NT}^m$. When $m = r + q_1$, we simply suppress the superscript m. For any $k > 0$, define $\hat{\Sigma}_1^m = \frac{1}{k}\sum_{t=1}^k \hat{g}_t^m \hat{g}_t^{m\prime}$ and $\hat{\Sigma}_2^m = \frac{1}{T-k}\sum_{t=k+1}^T \hat{g}_t^m \hat{g}_t^{m\prime}$. The sum of squared residuals is

$$\begin{aligned} \hat{S}^m(k) &= \sum\nolimits_{t=1}^{k} [vec(\hat{g}_t^m \hat{g}_t^{m\prime} - \hat{\Sigma}_1^m)]'[vec(\hat{g}_t^m \hat{g}_t^{m\prime} - \hat{\Sigma}_1^m)] \\ &\quad + \sum\nolimits_{t=k+1}^{T} [vec(\hat{g}_t^m \hat{g}_t^{m\prime} - \hat{\Sigma}_2^m)]'[vec(\hat{g}_t^m \hat{g}_t^{m\prime} - \hat{\Sigma}_2^m)], \end{aligned} \tag{A-12}$$

and the least squares estimator of the change point is $\hat{k} = \arg\min \hat{S}^m(k) = \arg\min(\hat{S}^m(k) - \hat{S}^m(k_0))$.

Consider the difference $\hat{S}^m(k) - \hat{S}(k)$. After some calculation, we have

$$\begin{aligned} \hat{S}^m(k) - \hat{S}(k) &= (2\sum\nolimits_{i=1}^{r+q_1}\sum\nolimits_{j=r+q_1+1}^{m} + \sum\nolimits_{i,j=r+q_1+1}^{m}) \\ &\quad [\sum\nolimits_{t=1}^{T}(\hat{g}_{it}^m \hat{g}_{jt}^m)^2 - \frac{1}{k}(\sum\nolimits_{t=1}^{k}\hat{g}_{it}^m \hat{g}_{jt}^m)^2 - \frac{1}{T-k}(\sum\nolimits_{t=k+1}^{T}\hat{g}_{it}^m \hat{g}_{jt}^m)^2]. \end{aligned}$$

It follows that

$$\begin{aligned} & (\hat{S}^m(k) - \hat{S}^m(k_0) - (\hat{S}(k) - \hat{S}(k_0)) \\ = \; & (2\sum\nolimits_{i=1}^{r+q_1}\sum\nolimits_{j=r+q_1+1}^{m} + \sum\nolimits_{i,j=r+q_1+1}^{m})[\frac{1}{k_0}(\sum\nolimits_{t=1}^{k_0}\hat{g}_{it}^m \hat{g}_{jt}^m)^2 + \\ & \frac{1}{T-k_0}(\sum\nolimits_{t=k_0+1}^{T}\hat{g}_{it}^m \hat{g}_{jt}^m)^2 - \frac{1}{k}(\sum\nolimits_{t=1}^{k}\hat{g}_{it}^m \hat{g}_{jt}^m)^2 - \frac{1}{T-k}(\sum\nolimits_{t=k+1}^{T}\hat{g}_{it}^m \hat{g}_{jt}^m)^2] \\ = \; & (2\sum\nolimits_{i=1}^{r+q_1}\sum\nolimits_{j=r+q_1+1}^{m} + \sum\nolimits_{i,j=r+q_1+1}^{m})(L_{ij1} + L_{ij2} - L_{ij3} - L_{ij4}). \end{aligned}$$

Following the same procedure as proving Theorem 1, it is not difficult to show $\arg\min(\hat{S}(k) - \hat{S}(k_0)) - k_0 = O_p(1)$. Thus based on the proof of Proposition 2 and Theorem 1, it suffices to show $\sup_{k \neq k_0}\left|\frac{(\hat{S}^m(k) - \hat{S}^m(k_0) - (\hat{S}(k) - \hat{S}(k_0))}{k-k_0}\right| = o_p(1)$. Consider $\sup_{k \neq k_0}\left|\frac{L_{ij}}{k-k_0}\right|$ for $i \leq r + q_1$ and $j > r + q_1 + 1$ as a representative. By definition, $\frac{1}{T}\sum_{t=1}^T \hat{g}_{lt}^{m2} = V_{NT,l}^2$, where $V_{NT,l}$ is the l-th diagonal element of V_{NT}. Thus $\frac{1}{T}\sum_{t=1}^T \hat{g}_{it}^{m2} = O_p(1)$ and $\frac{1}{T}\sum_{t=1}^T \hat{g}_{jt}^{m2} = O_p(\frac{1}{\delta_{NT}^4})$. It follows that $\sup_{k \neq k_0}\left|\frac{L_{ij1}}{k-k_0}\right| \leq \frac{1}{T\tau_0}\sum_{t=1}^T \hat{g}_{it}^{m2}\sum_{t=1}^T \hat{g}_{jt}^{m2} = O_p(\frac{T}{\delta_{NT}^4})$. Similarly, $\sup_{k \neq k_0}\left|\frac{L_{ij2}}{k-k_0}\right|$, $\sup_{k \neq k_0}\left|\frac{L_{ij3}}{k-k_0}\right|$ and

$\sup_{k\neq k_0}\left|\frac{L_{ij4}}{k-k_0}\right|$ are all $O_p(\frac{T}{\delta_{NT}^4})$. With $\frac{\sqrt{T}}{N}\to 0$, the proof is finished. ■

H PROOF OF THEOREM 2

Proof. Consider the consistency of $\tilde{r}_1$. Due to symmetry, the consistency of $\tilde{r}_2$ can be established similarly. What we need to show is: for any $\epsilon>0$, $P(\tilde{r}_1\neq r_1)<\epsilon$ for large (N,T). Based on $\left|\tilde{k}-k_0\right|=O_p(1)$, we have for any $\epsilon>0$, there exist $M>0$ such that $P(\left|\tilde{k}-k_0\right|>M)<\epsilon$ for all (N,T). Based on this M, $P(\tilde{r}_1\neq r_1)$ can be decomposed as

$$P(\tilde{r}_1\neq r_1,\left|\tilde{k}-k_0\right|>M)+P(\tilde{r}_1\neq r_1,k_0-M\leq\tilde{k}\leq k_0)+P(\tilde{r}_1\neq r_1,k_0+1\leq\tilde{k}\leq k_0+M).$$

The first term is less than $P(\left|\tilde{k}-k_0\right|>M)$, hence less than ϵ for all (N,T). The second term can be further decomposed as

$$\sum\nolimits_{k=k_0-M}^{k_0}P(\tilde{r}_1(k)\neq r_1,\tilde{k}=k),$$

where $P(\tilde{r}_1(k)\neq r_1,\tilde{k}=k)$ denotes the joint probability of $\tilde{k}=k$ and $\tilde{r}_1(k)\neq r_1$ and $\tilde{r}_1(k)$ denotes the estimated number of pre-break factors using subsample $t=1,...,k$. Obviously, $P(\tilde{r}_1(k)\neq r_1,\tilde{k}=k)\leq P(\tilde{r}_1(k)\neq r_1)$, hence the second term is less than $\sum_{k=k_0-M}^{k_0}P(\tilde{r}_1(k)\neq r_1)$. Furthermore, since for each $k\in[k_0-M,k_0]$, the factor loadings in the pre-break subsample are stable, $P(\tilde{r}_1(k)\neq r_1)\leq\frac{\epsilon}{M+1}$ for large (N,T). Therefore, the second term is less than $\sum_{k=k_0-M}^{k_0}\frac{\epsilon}{M+1}=\epsilon$ for large (N,T).

The argument for the second term also applies to the third term, except for some modifications. First, the third can be decomposed similarly as

$$\sum\nolimits_{k=k_0+1}^{k_0+M}P(\tilde{r}_1(k)\neq r_1,\tilde{k}=k)\leq\sum\nolimits_{k=k_0+1}^{k_0+M}P(\tilde{r}_1(k)\neq r_1),$$

hence it remains to show for each $k\in[k_0+1,k_0+M]$, $P(\tilde{r}_1(k)\neq r_1)\leq\frac{\epsilon}{M}$ for large (N,T). Unlike the second term, when $k\in[k_0+1,k_0+M]$ the factor loadings of the pre-break subsample $t=1,...,k$ has a break at $t=k_0$, hence results already established in previous literature for stable model is not directly applicable. To overcome this difficulty, we treat change in factor loadings as an extra error term such that $x_{it}=f_t'\lambda_{02,i}+e_{it}=f_t'\lambda_{01,i}+e_{it}+w_{it}=a_{it}+w_{it}$, where $a_{it}=f_t'\lambda_{01,i}+e_{it}$, $w_{it}=0$ for $1\leq t\leq k_0$ and $w_{it}=f_t'\lambda_{02,i}-f_t'\lambda_{01,i}$ for $t\geq k_0+1$. In other words, when $k\geq k_0+1$ the pre-break subsample $t=1,...,k$

can be regarded as having stable factor loadings and an extra error term in observations $t = k_0 + 1, ...k$. In matrix form, we have $X(k) = A(k) + W(k)$, where $X(k)$, $A(k)$ and $W(k)$ are all $k \times N$ matrix. Define ω_j^k, α_j^k and β_j^k as the j-th largest eigenvalue of $\frac{1}{Nk}X(k)X'(k)$, $\frac{1}{Nk}A(k)A'(k)$ and $\frac{1}{Nk}W(k)W'(k)$ respectively. By Weyl's inequality for singular values, the perturbation effect of the extra error matrix $W(k)$ on the eigenvalues of $A(k)$ is

$$\sqrt{\alpha_j^k} - \sqrt{\beta_1^k} \leq \sqrt{\omega_j^k} \leq \sqrt{\alpha_j^k} + \sqrt{\beta_1^k}, \tag{A-13}$$

hence $(\sqrt{\omega_j^k} - \sqrt{\alpha_j^k})^2 \leq \beta_1^k$. Since the number of nonzero elements in the $k \times N$ matrix $W(k)$ is only $(k - k_0) \times N$ and $k - k_0 \leq M$, simple calculation shows that

$$\begin{aligned} \beta_1^k &\leq tr(\frac{1}{Nk}W(k)W'(k)) = \frac{1}{Nk}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=k_0+1}^{k} w_{it}^2 \\ &\leq 2\frac{1}{Nk_0}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=k_0+1}^{k} \|f_t\|^2 (\|\lambda_{01,i}\|^2 + \|\lambda_{02,i}\|^2) \\ &\leq 8\frac{1}{k_0}\sum\nolimits_{t=k_0+1}^{k_0+M} \|f_t\|^2 \bar{\lambda}^2 = O_p(\frac{1}{T}). \end{aligned} \tag{A-14}$$

In addition, according to Bai and Ng (2002), $\alpha_j^k = \nu_j + o_p(1)$ for $j \leq r_1$, where ν_j is the j-th largest eigenvalue of $\Sigma_F \Sigma_{\Lambda_{01}}$, and $\alpha_j^k = O_p(\frac{1}{\delta_{NT}^2})$ for $j > r_1$. It follows that $\omega_j^k = \alpha_j^k + 2\sqrt{\alpha_j^k} O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{T}) = \nu_j + o_p(1)$ for $j \leq r_1$, and $\omega_j^k = O_p(\frac{1}{\delta_{NT}^2}) + O_p(\frac{1}{\delta_{NT}})O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{T}) = O_p(\frac{1}{\delta_{NT}^2})$ for $j > r_1$. This implies that the estimator of number of factors using Bai and Ng (2002) based on the sample $X(k)$ is still consistent for $k \in [k_0 + 1, k_0 + M]$, hence $P(\tilde{r}_1(k) \neq r_1) \leq \frac{\epsilon}{M}$ for large (N, T). ■

I PROOF OF PROPOSITION 3

Proof. The proof is similar to Theorem 2.

$$\begin{aligned} \beta_1^T &\leq tr(\frac{1}{NT}W(T)W'(T)) = \frac{1}{NT}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=k_0+1}^{T} w_{it}^2 \\ &\leq \frac{1}{NT}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=k_0+1}^{T} \|f_t\|^2 \|\lambda_{02,i} - \lambda_{01,i}\|^2 \\ &= (\frac{1}{T}\sum\nolimits_{t=k_0+1}^{T} \|f_t\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N} \|\lambda_{02,i} - \lambda_{01,i}\|^2) = O_p(\frac{1}{\delta_{NT}^c}). \end{aligned} \tag{A-15}$$

By Weyl's inequality for singular values, $\sqrt{\alpha_j^T} - \sqrt{\beta_1^T} \leq \sqrt{\omega_j^T} \leq \sqrt{\alpha_j^T} + \sqrt{\beta_1^T}$, hence

$(\sqrt{\omega_j^T}-\sqrt{\alpha_j^T})^2 \le \beta_1^T = O_p(\frac{1}{\delta_{NT}^{\frac{c}{2}}})$. It follows that $\omega_j^T = \alpha_j^T + 2\sqrt{\alpha_j^T}O_p(\frac{1}{\delta_{NT}^{\frac{c}{2}}}) + O_p(\frac{1}{\delta_{NT}^{c}}) = \nu_j + o_p(1)$ for $j \le r_1$, and $\omega_j^T = O_p(\frac{1}{\delta_{NT}^2}) + O_p(\frac{1}{\delta_{NT}})O_p(\frac{1}{\delta_{NT}^{\frac{c}{2}}}) + O_p(\frac{1}{\delta_{NT}^{c}}) = O_p(\frac{1}{\delta_{NT}^{c}})$ for $j > r_1$ when $c < 2$. ■

J PROOF OF THEOREM 3

Proof. Again by symmetry, we only need to show the first half.

To show $\frac{1}{\tilde{k}}\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 = O_p(\frac{1}{\delta_{NT}^2})$, we need to show for any $\epsilon > 0$, there exist $C > 0$ such that $P(\delta_{NT}^2\frac{1}{\tilde{k}}\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 > C) < \epsilon$ for all (N,T). First, based on $\left|\tilde{k}-k_0\right| = O_p(1)$ we can choose $M > 0$ such that $P(\left|\tilde{k}-k_0\right| > M) < \frac{\epsilon}{2}$ for the given ϵ. Next,

$$P(\delta_{NT}^2\tfrac{1}{\tilde{k}}\textstyle\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 > C)$$
$$= P(\delta_{NT}^2\tfrac{1}{\tilde{k}}\textstyle\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 > C, \left|\tilde{k}-k_0\right| > M)$$
$$+\textstyle\sum_{k=k_0-M}^{k_0+M} P(\delta_{NT}^2\tfrac{1}{\tilde{k}}\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 > C, \tilde{k}=k).$$
$$\le P(\left|\tilde{k}-k_0\right| > M) + \textstyle\sum_{k=k_0-M}^{k_0+M} P(\delta_{NT}^2\tfrac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2 > C)$$
$$\le \tfrac{\epsilon}{2} + \textstyle\sum_{k=k_0-M}^{k_0+M} P(\delta_{NT}^2\tfrac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2 > C).$$

If we can show $\frac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2 = O_p(\frac{1}{\delta_{NT}^2})$ for each $k \in [k_0-M, k_0+M]$, then for the given ϵ and for each $k \in [k_0-M, k_0+M]$, we can take $C(k) > 0$ such that $P(\delta_{NT}^2\frac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2 > C(k)) < \frac{\epsilon}{2(2M+1)}$ for all (N,T). Take $C = \max\limits_{k\in[k_0-M,k_0+M]} C(k)$, then $P(\delta_{NT}^2\frac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2 > C) \le \frac{\epsilon}{2} + \sum_{k=k_0-M}^{k_0+M}\frac{\epsilon}{2(2M+1)} = \epsilon$ for all (N,T), hence it remains to show for each $k \in [k_0-M, k_0+M]$, $\frac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2$ is $O_p(\frac{1}{\delta_{NT}^2})$.

First consider the case $k_0 - M \le k \le k_0$. In this case, factor loadings are stable and $k_0 - M \le k$ guarantees $k \to \infty$ as $k_0 \to \infty$, hence Theorem 1 of Bai and Ng (2002) is applicable.

Next consider the case $k_0 + 1 \le k \le k_0 + M$. Following the same notation as proof of Theorem 2 and define $E(k) = (e_1, ..., e_k)'$, we have $X(k) = A(k) + W(k) = F_1(k)\Lambda_{01}' + E(k) +$

$W(k)$, thus

$$\begin{aligned} & X(k)X'(k) \\ = \; & F_1(k)\Lambda_{01}'\Lambda_{01}F_1'(k) + F_1(k)\Lambda_{01}'[E(k)+W(k)]' \\ & +[E(k)+W(k)]\Lambda_{01}F_1'(k) + [E(k)+W(k)][E(k)+W(k)]'. \end{aligned} \tag{A-16}$$

It follows that

$$\begin{aligned} \hat{f}_t^u(k) - H_1^{w\prime}(k)f_t \; = \; & \frac{1}{Nk}[\tilde{F}_1^{w\prime}(k)F_1(k)\Lambda_{01}'e_t + \tilde{F}_1^{w\prime}(k)E(k)\Lambda_{01}f_t + \tilde{F}_1^{w\prime}(k)E(k)e_t \\ & +\tilde{F}_1^{w\prime}(k)F_1(k)\Lambda_{01}'w_t + \tilde{F}_1^{w\prime}(k)W(k)\Lambda_{01}f_t + \tilde{F}_1^{w\prime}(k)W(k)w_t \\ & +\tilde{F}_1^{w\prime}(k)E(k)w_t + \tilde{F}_1^{w\prime}(k)W(k)e_t] \\ = \; & Q_{1,t}(k) + Q_{2,t}(k) + Q_{3,t}(k) + Q_{4,t}(k) + Q_{5,t}(k) + Q_{6,t}(k) \\ & +Q_{7,t}(k) + Q_{8,t}(k), \end{aligned} \tag{A-17}$$

and $\frac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{w\prime}(k)f_t\right\|^2 \leq 8\sum_{m=1}^{8}\frac{1}{k}\sum_{t=1}^{k}\|Q_{m,t}(k)\|^2$. Following the same procedure as proof of Theorem 1 in Bai and Ng (2002), it can be shown for $m = 1,2,3$, $\frac{1}{k}\sum_{t=1}^{k}\|Q_{m,t}(k)\|^2 = O_p(\frac{1}{\delta_{NT}^2})$. Next, noting that $w_{it} = 0$ for $1 \leq t \leq k_0$,

$$\begin{aligned} \frac{1}{k}\sum_{t=1}^{k}\|Q_{4,t}(k)\|^2 \; = \; & \frac{1}{k}\sum_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{w\prime}(k)F_1(k)\Lambda_{01}'w_t\right\|^2 \\ \leq \; & \frac{1}{k}\sum_{t=1}^{k}(\frac{1}{k}\sum_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\sum_{s=1}^{k}\|f_s\|^2)\left\|\frac{1}{N}\Lambda_{01}'w_t\right\|^2 \\ \leq \; & (\frac{1}{k}\sum_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\sum_{s=1}^{k}\|f_s\|^2)(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_{01,i}\|^2) \\ & (\frac{1}{k}\sum_{t=1}^{k}\frac{1}{N}\sum_{i=1}^{N}\|w_{it}\|^2) \\ \leq \; & (\frac{1}{k}\sum_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\sum_{s=1}^{k}\|f_s\|^2)(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_{01,i}\|^2) \\ & (\frac{1}{N}\sum_{i=1}^{N}\|\lambda_{01,i} - \lambda_{02,i}\|^2)(\frac{1}{k}\sum_{t=k_0+1}^{k}\|f_t\|^2) \\ = \; & O_p(1)O_p(1)O(1)O(1)O_p(\frac{1}{T}) = O_p(\frac{1}{T}), \end{aligned}$$

$$
\begin{aligned}
\frac{1}{k}\sum\nolimits_{t=1}^{k}\|Q_{5,t}(k)\|^2 &= \frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{u\prime}(k)W(k)\Lambda_{01}f_t\right\|^2 \\
&\leq \frac{1}{k}\sum\nolimits_{t=1}^{k}(\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{N^2}\frac{1}{k}\sum\nolimits_{s=1}^{k}\|w_s'\Lambda_{01}f_t\|^2) \\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}\|^2)(\frac{1}{k}\sum\nolimits_{t=1}^{k}\|f_t\|^2) \\
&\quad (\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}-\lambda_{02,i}\|^2)(\frac{1}{k}\sum\nolimits_{s=k_0+1}^{k}\|f_s\|^2) \\
&= O_p(1)O(1)O_p(1)O(1)O_p(\frac{1}{T}) = O_p(\frac{1}{T}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{k}\sum\nolimits_{t=1}^{k}\|Q_{6,t}(k)\|^2 &= \frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{u\prime}(k)W(k)w_t\right\|^2 \\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)\frac{1}{N^2}(\frac{1}{k}\sum\nolimits_{s=1}^{k}\|w_s\|^2)(\frac{1}{k}\sum\nolimits_{t=1}^{k}\|w_t\|^2) \\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}-\lambda_{02,i}\|^2)^2 \\
&\quad (\frac{1}{k}\sum\nolimits_{s=k_0+1}^{k}\|f_s\|^2)(\frac{1}{k}\sum\nolimits_{t=k_0+1}^{k}\|f_t\|^2) \\
&= O_p(1)O(1)O_p(\frac{1}{T})O_p(\frac{1}{T}) = O_p(\frac{1}{T^2}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{k}\sum\nolimits_{t=1}^{k}\|Q_{7,t}(k)\|^2 &= \frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{u\prime}(k)E(k)w_t\right\|^2 \\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\frac{1}{N}\sum\nolimits_{s=1}^{k}\sum\nolimits_{i=1}^{N}e_{is}^2) \\
&\quad (\frac{1}{k}\sum\nolimits_{t=k_0+1}^{k}\|f_t\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}-\lambda_{02,i}\|^2) \\
&= O_p(1)O_p(1)O_p(\frac{1}{T})O(1) = O_p(\frac{1}{T}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{k}\sum\nolimits_{t=1}^{k}\|Q_{8,t}(k)\|^2 &= \frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{u\prime}(k)W(k)e_t\right\|^2 \\
&\leq \frac{1}{k}\sum\nolimits_{t=1}^{k}(\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)\frac{1}{N^2}(\frac{1}{k}\sum\nolimits_{s=1}^{k}\|w_s'e_t\|^2) \\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\sum\nolimits_{t=k_0+1}^{k}\|f_s\|^2) \\
&\quad (\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}-\lambda_{02,i}\|^2)(\frac{1}{k}\frac{1}{N}\sum\nolimits_{t=1}^{k}\sum\nolimits_{i=1}^{N}e_{it}^2) \\
&= O_p(1)O_p(\frac{1}{T})O(1)O_p(1) = O_p(\frac{1}{T}),
\end{aligned}
$$

hence $\frac{1}{k}\sum_{t=1}^{k}\|Q_{m,t}(k)\|^2 = O_p(\frac{1}{\delta_{NT}^2})$ for $m = 4,5,6,7,8$. ■

K PROOF OF THEOREM 4

Proof. Define $V(k) = \tilde{S}(k) - \tilde{S}(k_0)$, $U(k) = A^* + E^* = (k_0 - k)a_k' a_k - 2a_k' \sum_{t=k+1}^{k_0}(y_t + z_t)$ for $k < k_0$. For any fixed constant $M < \infty$, define $V^M(k) = V(k)$ for $|k_0 - k| < M$, $U^M(k) = U(k)$ for $|k_0 - k| < M$, $W^M(l) = W(l)$ for $|l| < M$. $V^M(k)$, $U^M(k)$ and $W^M(l)$ are all finite dimensional random vector.

Step 1: $V^M(k) \xrightarrow{p} U^M(k)$ as $(N,T) \to \infty$ for any fixed $M < \infty$.

By symmetry we only need to study the case $k < k_0$.

It suffices to show $\sup_{|k_0-k|<M, k<k_0} |V(k) - U(k)| = o_p(1)$.

$$\begin{aligned}\sup_{|k_0-k|<M,k<k_0} |V(k) - U(k)| \leq{}& \sup_{|k_0-k|<M,k<k_0} |B^*| + \sup_{|k_0-k|<M,k<k_0} |C^*| + \\ & \sup_{|k_0-k|<M,k<k_0} |D^*| + \sup_{|k_0-k|<M,k<k_0} |F^*| + \sup_{|k_0-k|<M,k<k_0} |G^*|.\end{aligned}$$

$$\sup_{|k_0-k|<M,k<k_0} |B^*| = \sup_{|k_0-k|<M,k<k_0} (T-k_0)(\frac{k_0-k}{T-k})^2 \|\Sigma_2 - \Sigma_1\|^2 = O(\frac{1}{T}) = o(1).$$

$$\sup_{|k_0-k|<M,k<k_0} |C^*| \leq M \sup_{k\in D,k<k_0} \left|\frac{C^*}{k_0-k}\right| = o_p(1).$$

Similarly, $\sup_{|k_0-k|<M,k<k_0} |D^*|$, $\sup_{|k_0-k|<M,k<k_0} |F^*|$ and $\sup_{|k_0-k|<M,k<k_0} |G^*|$ are all $o_p(1)$.

Step 2: $U^M(k) \xrightarrow{d} W^M(k - k_0)$ as $(N,T) \to \infty$ for any fixed $M < \infty$.

$U^M(k) = (k_0 - k)a_k' a_k - 2a_k' \sum_{t=k+1}^{k_0}(y_t + z_t)$, for $|k_0 - k| < M$ and $k < k_0$.

For $|k_0 - k| < M$,

$$\begin{aligned}(k_0-k)a_k' a_k &= (k_0-k)\|\Sigma_2-\Sigma_1\|^2 + (k_0-k)[(\frac{T-k_0}{T-k})^2 - 1]\|\Sigma_2-\Sigma_1\|^2 \\ &= (k_0-k)\|\Sigma_2-\Sigma_1\|^2 + O(\frac{1}{T}).\end{aligned}$$

By part (6) of Lemma 7,

$$\begin{aligned}\sup_{|k_0-k|<M,k<k_0} \left|-2a_k' \sum_{t=k+1}^{k_0} z_t\right| &\leq 2M\|\Sigma_2-\Sigma_1\| \sup_{|k_0-k|<M,k<k_0} \left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} z_t\right\| \\ &\leq 2M\|\Sigma_2-\Sigma_1\| \sup_{k\in D,k<k_0} \left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} z_t\right\| = o_p(1).\end{aligned}$$

Next,

$$-2a_k' \sum_{t=k+1}^{k_0} y_t = -2[vec(\Sigma_2 - \Sigma_1)]' \sum_{t=k+1}^{k_0} y_t - 2(\frac{T-k_0}{T-k} - 1)[vec(\Sigma_2 - \Sigma_1)]' \sum_{t=k+1}^{k_0} y_t,$$

and

$$\begin{aligned} &\sup_{|k_0-k|<M, k<k_0} \left| -2(\frac{T-k_0}{T-k} - 1)[vec(\Sigma_2 - \Sigma_1)]' \sum_{t=k+1}^{k_0} y_t \right| \\ \leq\ & \frac{2M}{T-k_0} \|\Sigma_2 - \Sigma_1\| \sup_{|k_0-k|<M, k<k_0} \left\| \sum_{t=k+1}^{k_0} y_t \right\| = O_p(\frac{1}{T}) \end{aligned}$$

Taking together, $U^M(k) \xrightarrow{d} (k_0 - k) \|\Sigma_2 - \Sigma_1\|^2 - 2\sum_{t=k+1}^{k_0} [vec(\Sigma_2 - \Sigma_1)]' y_t$ for $|k_0 - k| < M$ and $k < k_0$. Similarly, for $|k_0 - k| < M$ and $k > k_0$, $U^M(k) \xrightarrow{d} (k - k_0) \|\Sigma_2 - \Sigma_1\|^2 - 2\sum_{t=k_0+1}^{k} [vec(\Sigma_2 - \Sigma_1)]' y_t$.

Step 3: $V^M(k) \xrightarrow{d} W^M(k - k_0)$ as $(N, T) \to \infty$ for any fixed $M < \infty$.

Based on step 1 and step 2 and using Slutsky's Lemma, $V^M(k) \xrightarrow{d} W^M(k - k_0)$.

Step 4: $\arg\min V^M(k) - k_0 \xrightarrow{d} \arg\min W^M(l)$ as $(N, T) \to \infty$ for any fixed $M < \infty$.

If $W(l)$ does not have unique maximizer, then these exist $l \neq l'$ such that $W(l) = W(l')$. It's easy to see $P(W(l) = W(l')) = 0$. The number of integer pairs (l, l') is countable and sum of countable zero is zero, hence the probability that $W(l)$ does not have unique maximizer is zero.

Next, for a finite dimensional vector x, $f(x) = \arg\min x$ is a continuous function, hence by continuous mapping theorem we have $\arg\min V^M(k) - k_0 \xrightarrow{d} \arg\min W^M(l)$.

By definition of convergence in distribution, for any $\epsilon > 0$ and any $|j| \leq M$, there exist $N_j^* > 0$ and $T_j^* > 0$ such that for $N > N_j^*$ and $T > T_j^*$,

$$\left| P(\arg\min V^M(k) - k_0 = j) - P(\arg\min W^M(l) = j) \right| < \epsilon.$$

Take $N^* = \max\{N_j^*, |j| \leq M\}$ and $T^* = \max\{T_j^*, |j| \leq M\}$. For $N > N^*$ and $T > T^*$, $\left| P(\arg\min V^M(k) - k_0 = j) - P(\arg\min W^M(l) = j) \right| < \epsilon$ for all $|j| \leq M$.

Step 5: $\tilde{k} - k_0 \xrightarrow{d} \arg\min W(l)$ as $(N, T) \to \infty$.

Step 5.1: By Theorem 1, $\tilde{k} - k_0 = O_p(1)$ as $(N, T) \to \infty$, hence for any $\epsilon > 0$, there exist $M_1 < \infty$, $N_1 > 0$ and $T_1 > 0$, such that for $N > N_1$ and $T > T_1$, $P(\left|\tilde{k} - k_0\right| > M_1) < \frac{\epsilon}{3}$.

Step 5.2: $\tilde{l} = \arg\min W(l) = O_p(1)$ as $(N, T) \to \infty$.

First note that $P(\min_{|l|>M} W(l) \leq 0) \leq P(\min_{l<-M} W_1(l) \leq 0) + P(\min_{l>M} W_2(l) \leq 0)$

$= P(\sup_{l<-M}\{-l\left\|\Sigma_2-\Sigma_1\right\|^2 - 2\sum_{t=k_0+l}^{k_0}[vec(\Sigma_2-\Sigma_1)]'y_t\} \leq 0)$
$+P(\sup_{l>M}\{l\left\|\Sigma_2-\Sigma_1\right\|^2 - 2\sum_{t=k_0+1}^{k_0+l}[vec(\Sigma_2-\Sigma_1)]'y_t\} \leq 0)$
$\leq P(\sup_{l<-M} 2[vec(\Sigma_2-\Sigma_1)]'\frac{1}{l}\sum_{t=k_0+l}^{k_0} y_t \geq \left\|\Sigma_2-\Sigma_1\right\|^2)$
$+P(\sup_{l>M} 2[vec(\Sigma_2-\Sigma_1)]'\frac{1}{l}\sum_{t=k_0+1}^{k_0+l} y_t \geq \left\|\Sigma_2-\Sigma_1\right\|^2)$
$\leq P(\sup_{l<-M}\left\|\frac{1}{-l}\sum_{t=k_0+l}^{k_0} y_t\right\| \geq \frac{\|\Sigma_2-\Sigma_1\|}{2}) + P(\sup_{l>M}\left\|\frac{1}{l}\sum_{t=k_0+1}^{k_0+l} y_t\right\| \geq \frac{\|\Sigma_2-\Sigma_1\|}{2}) = \frac{C}{M}$ by Hajek-Renyi inequality. Hence for any $\epsilon > 0$, there exists $M_2 < \infty$ such that $P(\sup_{|l|>M_2} W(l) \leq 0) < \frac{\epsilon}{3}$. Since $W(0) = 0$, $\min W(l) \leq 0$, therefore $P(\left|\tilde{l}\right| > M_2) \leq P(\min_{|l|>M_2} W(l) \leq 0) < \frac{\epsilon}{3}$.

Step 5.3:

Take $M = \max\{M_1, M_2\}$. Based on step 4, for any $\epsilon > 0$ there exist $N_2 > 0$ and $T_2 > 0$, such that for $N > N_2$ and $T > T_2$, for all $|j| \leq M$,

$$\left|P(\arg\min V^M(k) - k_0 = j) - P(\arg\min W^M(l) = j)\right| < \frac{\epsilon}{3}.$$

Step 5.4:

Take $N^* = \max\{N_1, N_2\}$ and $T^* = \max\{T_1, T_2\}$. For any $N > N^*$ and $T > T^*$,

if $|j| > M$,

$\left|P(\tilde{k} - k_0 = j) - P(\tilde{l} = j)\right| < P(\tilde{k} - k_0 = j) + P(\tilde{l} = j) < P(\left|\tilde{k} - k_0\right| > M) + P(\left|\tilde{l}\right| > M) < \frac{\epsilon}{3} + \frac{\epsilon}{3} < \epsilon$;

if $|j| \leq M$,

$\tilde{k} - k_0 = j$ implies $\arg\min V^M(k) - k_0 = j$, hence $P(\tilde{k} - k_0 = j) \leq P(\arg\min V^M(k) - k_0 = j)$,

$\arg\min V^M(k) - k_0 = j$ implies $\tilde{k} - k_0 = j$ or $\left|\tilde{k} - k_0\right| > M$,

hence $P(\arg\min V^M(k) - k_0 = j) < P(\tilde{k} - k_0 = j) + P(\left|\tilde{k} - k_0\right| > M)$,

therefore $\left|P(\tilde{k} - k_0 = j) - P(\arg\min V^M(k) - k_0 = j)\right| < P(\left|\tilde{k} - k_0\right| > M) < \frac{\epsilon}{3}$.

Similarly $\left|P(\tilde{l} = j) - P(\arg\min W^M(l) = j)\right| < P(\left|\tilde{l}\right| > M) < \frac{\epsilon}{3}$,

therefore $\left|P(\tilde{k} - k_0 = j) - P(\tilde{l} = j)\right| < \left|P(\tilde{k} - k_0 = j) - P(\arg\min V^M(k) - k_0 = j)\right|$
$+\left|P(\arg\min V^M(k) - k_0 = j) - P(\arg\min W^M(l) = j)\right|$
$+\left|P(\hat{l} = j) - P(\arg\min W^M(l) = j)\right|$
$< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} < \epsilon$.

Therefore, we proved that for any $\epsilon > 0$, there exist $N^* > 0$ and $T^* > 0$, such that for $N > N^*$ and $T > T^*$, $\left|P(\tilde{k} - k_0 = j) - P(\tilde{l} = j)\right| < \epsilon$ for all j. By definition, $\tilde{k} - k_0 \xrightarrow{d}$

$\arg\min W(l)$. ■

L PROOF OF LEMMAS

Lemma 1 *Under Assumptions 1-5,* $\frac{1}{T}\sum_{t=1}^{T}\|\tilde{g}_t - J'g_t\|^2 = O_p(\frac{1}{\delta_{NT}^2})$.

Proof. This Lemma is Theorem 1 of Bai and Ng (2002) for the equivalent model, therefore it suffices to verify Assumptions A-D of Bai and Ng (2002).

Assumption A: By Assumption 1,

$$\begin{aligned}
\mathbb{E}\|g_t\|^4 &\leq \max\{\|A\|^4, \|B\|^4\}\mathbb{E}\|f_t\|^4 < M < \infty, \\
\frac{1}{T}\sum_{t=1}^{T} g_t g_t' &= \tau_0 \frac{1}{k_0}\sum_{t=1}^{k_0} A f_t f_t' A' + (1-\tau_0)\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} B f_t f_t' B' \\
&\xrightarrow{p} \tau_0 A \Sigma_F A' + (1-\tau_0) B \Sigma_F B',
\end{aligned}$$

which equals Σ_G and is positive definite.

Assumption B: By Assumption 2,

$$\|\gamma_i\| \leq \left\|(\lambda_{0,i}', \lambda_{1,i}', \lambda_{2,i}')'\right\| = (\|\lambda_{0,i}\|^2 + \|\lambda_{1,i}\|^2 + \|\lambda_{2,i}\|^2)^{\frac{1}{2}} \leq \sqrt{3}\bar{\lambda} < \infty$$

and $\left\|\frac{1}{N}\Gamma'\Gamma - \Sigma_\Gamma\right\| \to 0$ for some positive definite matrix Σ_Γ.

Assumption C: Assumption 3 together with Assumption 5 implies Assumption C.

Assumption D:

$$\begin{aligned}
\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t e_{it}\right\|^2) &\leq 2\|A\|^2 \mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k_0} f_t e_{it}\right\|^2) \\
&\quad +2\|B\|^2 \mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=k_0+1}^{T} f_t e_{it}\right\|^2) \\
&\leq 2\tau_0 M + 2(1-\tau_0)M = 2M.
\end{aligned}$$

■

Lemma 2 *Under Assumptions 1-5 and 7,* $\|J - J_0\| = o_p(1)$.

Proof. This Lemma follows from Proposition 1 of Bai (2003). Assumptions A-D is verified in Lemma 1, Assumption G is identical to Assumption 7. ■

Lemma 3 *Under Assumptions 1-8,*

(1) Hajek-Renyi inequality applies to the process $\{y_t, t = 1, ..., k_0\}$, $\{y_t, t = k_0, ..., 1\}$, $\{y_t, t = k_0+1, ..., T\}$ *and* $\{y_t, t = T, ..., k_0+1\}$,

(2) $\sup\limits_{k\leq k_0} \frac{1}{k}\sum_{t=1}^{k}\|g_t\|^2 = O_p(1)$, $\sup\limits_{k\geq k_0} \frac{1}{T-k}\sum_{t=k+1}^{T}\|g_t\|^2 = O_p(1)$, $\sup\limits_{k<k_0} \frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\|g_t\|^2 = O_p(1)$ *and* $\sup\limits_{k>k_0} \frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\|g_t\|^2 = O_p(1)$.

Proof. (1) $P(\sup\limits_{m\leq k\leq k_0} c_k \left\|\sum_{t=1}^{k} y_t\right\| > M) = P(\sup\limits_{m\leq k\leq k_0} c_k \left\|J_0' A[\sum_{t=1}^{k}(f_t f_t' - \Sigma_F)]A' J_0\right\| > M) \leq P(\|J_0' A\|^2 \sup\limits_{m\leq k\leq k_0} c_k \left\|\sum_{t=1}^{k}\epsilon_t\right\| > M) \leq \frac{C}{M^2}(mc_m^2 + \sum_{k=m+1}^{k_0} c_k^2)$, where the last inequality follows from Hajek-Renyi inequality for process $\{\epsilon_t, t = 1, ..., k_0\}$. Other processes can be proved similarly.

(2) $\sup\limits_{k\leq k_0} \frac{1}{k}\sum_{t=1}^{k}\|g_t\|^2 \leq \|A\|^2 \mathbb{E}\|f_t\|^2 + \|A\|^2 \sup\limits_{k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)$, where $\mathbb{E}\|f_t\|^2 = tr\Sigma_F$. Define $D_k = \frac{1}{k}\sum_{t=1}^{k}(f_t f_t' - \Sigma_F)$, then

$$\left|\frac{1}{k}\sum_{t=1}^{k}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)\right| = |trD_k| \leq \sqrt{r+q_1}(trD_k^2)^{\frac{1}{2}} = \sqrt{r+q_1}\|D_k\|,$$

it follows

$$\left|\sup_{k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)\right| \leq \sup_{k\leq k_0}\left|\frac{1}{k}\sum_{t=1}^{k}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)\right| \leq \sqrt{r+q_1}\sup_{k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}\epsilon_t\right\|,$$

which is $O_p(1)$ by Hajek-Renyi inequality. Thus $\sup\limits_{k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\|g_t\|^2 \leq \|A\|^2 \mathbb{E}\|f_t\|^2 + \|A\|^2 O_p(1) = O_p(1)$. Other terms can be proved similarly. ■

Lemma 4 *General Hajek-Renyi inequality (Theorem 1.1 of Fazekas and Klesov (2001)):*

Let β_1, β_2,..., β_n *be a sequence of nondecreasing positive numbers. Let* α_1, α_2,..., α_n *be a sequence of nonnegative numbers. Let* r *be a fixed positive number. For the partial sum process* $S_l = \sum_{k=1}^{l} X_k$, *assume for each* m *with* $1 \leq m \leq n$, $\mathbb{E}(\sup\limits_{1\leq l\leq m}|S_l|^r) \leq \sum_{l=1}^{m}\alpha_l$, *then* $\mathbb{E}(\sup\limits_{1\leq l\leq n}\left|\frac{S_l}{\beta_l}\right|^r) \leq 4\sum_{l=1}^{n}\frac{\alpha_l}{\beta_l^r}$.

Note that no dependence structure on $\{X_k, k = 1, ...\}$ is assumed.

Lemma 5 *Under Assumptions 1-8 and 10,*

(1) $\sup\limits_{k\in D, k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}(\tilde{g}_t - J' g_t) g_t' J\right\| = O_p(\frac{1}{\delta_{NT}})$,

(2) $\sup\limits_{k\in D^c, k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}(\tilde{g}_t - J' g_t) g_t' J\right\| = O_p(\frac{1}{\delta_{NT}})$,

(3) $\sup_{k\in D^c, k<k_0} \left\| \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} (\tilde{g}_t - J'g_t) g_t' J \right\| = O_p(\frac{1}{\delta_{NT}})$,

(4) $\sup_{k\in D, k<k_0} \left\| \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} (\tilde{g}_t - J'g_t) g_t' J \right\| = O_p(\frac{1}{\delta_{NT}})$,

(5) $\sup_{k\in D, k<k_0} \left\| \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} (\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)' \right\| = O_p(\frac{1}{\delta_{NT}^2})$,

(6) $\sup_{k\in D^c, k\le k_0} \left\| \frac{1}{k} \sum_{t=1}^{k} (\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)' \right\| = O_p(\frac{1}{\delta_{NT}^2})$,

(7) $\sup_{k\le k_0} \left\| \frac{1}{T-k} \sum_{t=k+1}^{T} (\tilde{g}_t - J'g_t) g_t' J \right\| = O_p(\frac{1}{\delta_{NT}})$.

Proof. We will prove parts (2), (5) and (7). Proof of parts (1), (3) and (4) is similar to part (2), proof of part (6) is similar to part (5). First consider part (2).

$$
\begin{aligned}
& \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{k} \sum_{t=1}^{k} (\tilde{g}_t - J'g_t) g_t' J \right\| \\
= & \sup_{k\in D^c, k\le k_0} \left\| \begin{array}{c} \frac{1}{k} \sum_{t=1}^{k} V_{NT}^{-1} \frac{1}{T} (\sum_{s=1}^{T} \tilde{g}_s \gamma_N(s,t) + \sum_{s=1}^{T} \tilde{g}_s \zeta_{st} \\ + \sum_{s=1}^{T} \tilde{g}_s \eta_{st} + \sum_{s=1}^{T} \tilde{g}_s \xi_{st}) g_t' J \end{array} \right\| \\
\le & \Big(\sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} (\tilde{g}_s - J'g_s) g_t' \gamma_N(s,t) \right\| \\
& + \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} J'g_s g_t' \gamma_N(s,t) \right\| \\
& + \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} (\tilde{g}_s - J'g_s) g_t' \zeta_{st} \right\| \\
& + \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} J'g_s g_t' \zeta_{st} \right\| \\
& + \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} (\tilde{g}_s - J'g_s) g_t' \eta_{st} \right\| \\
& + \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} J'g_s g_t' \eta_{st} \right\| \\
& + \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} (\tilde{g}_s - J'g_s) g_t' \xi_{st} \right\| \\
& + \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{T} \frac{1}{k} \sum_{t=1}^{k} \sum_{s=1}^{T} J'g_s g_t' \xi_{st} \right\| \Big) \left\| V_{NT}^{-1} \right\| \left\| J \right\| \\
= & (I + II + III + IV + V + VI + VII + VIII) \left\| V_{NT}^{-1} \right\| \left\| J \right\|.
\end{aligned}
$$

Consider the eight terms one by one.

$$
\begin{aligned}
& I \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\tilde{g}_s - J'g_s\right\|^2)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0} (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k} g_t'\gamma_N(s,t)\right\|^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\tilde{g}_s - J'g_s\right\|^2)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0} [(\frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|g_t\right\|^2)(\frac{1}{T}\frac{1}{k}\sum\nolimits_{s=1}^{T}\sum\nolimits_{t=1}^{k}\left|\gamma_N(s,t)\right|^2)]^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\tilde{g}_s - J'g_s\right\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0}\frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|g_t\right\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0}\frac{1}{T}\frac{1}{k}\sum\nolimits_{s=1}^{T}\sum\nolimits_{t=1}^{k}\left|\gamma_N(s,t)\right|^2)^{\frac{1}{2}} \\
=\ & O_p(\frac{1}{\delta_{NT}})O_p(1)O_p(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

where last equality follows from Lemma 1, Lemma 3 and $\sup_{k\in D^c, k\leq k_0}\frac{1}{T}\frac{1}{k}\sum_{s=1}^{T}\sum_{t=1}^{k}\left|\gamma_N(s,t)\right|^2 \leq \frac{1}{T}\sup_{k\in D^c, k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}(\sum_{s=1}^{T} M\left|\gamma_N(s,t)\right|) \leq \frac{1}{T}M^2$ by part (1) of Assumption 5.

$$
\begin{aligned}
& II \\
\leq\ & \left\|J\right\| \sup_{k\in D^c, k\leq k_0}\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|g_s\right\|\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k} g_t'\gamma_N(s,t)\right\| \\
\leq\ & \left\|J\right\| (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|g_s\right\|^2)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k} g_t'\gamma_N(s,t)\right\|^2)^{\frac{1}{2}} \\
\leq\ & \left\|J\right\| (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|g_s\right\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0}\frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|g_t\right\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0}\frac{1}{T}\frac{1}{k}\sum\nolimits_{s=1}^{T}\sum\nolimits_{t=1}^{k}\left|\gamma_N(s,t)\right|^2)^{\frac{1}{2}} \\
=\ & O_p(1)O_p(1)O_p(1)O_p(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

where the last equality follows from $\sup_{k\in D^c, k\leq k_0}\frac{1}{T}\frac{1}{k}\sum_{s=1}^{T}\sum_{t=1}^{k}\left|\gamma_N(s,t)\right|^2 = O_p(\frac{1}{T})$, Lemma 2, Assumption 1 and Lemma 3.

$$
\begin{aligned}
& III \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\tilde{g}_s - J'g_s\right\|^2)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k} g_t'\frac{1}{N}\sum\nolimits_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right\|^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\tilde{g}_s - J'g_s\right\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0}\frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|g_t\right\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}} \\
& (\frac{1}{T}\sum\nolimits_{s=1}^{T}\sup_{k\in D^c, k\leq k_0}\frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right\|^2)^{\frac{1}{2}} \\
=\ & O_p(\frac{1}{\delta_{NT}})O_p(1)\frac{1}{\sqrt{N}}O_p(1).
\end{aligned}
$$

$$
\begin{aligned}
& IV \\
\leq\ & \|J\| \sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T} g_s g_t' \frac{1}{N}\sum\nolimits_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})] \right\| \\
\leq\ & \|J\| (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \left\| \frac{1}{NT}\sum\nolimits_{s=1}^{T}\sum\nolimits_{i=1}^{N} g_s[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})] \right\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)^{\frac{1}{2}} \\
\leq\ & \|J\| \frac{1}{\sqrt{N}} (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k}\frac{1}{T}\sum\nolimits_{s=1}^{T} \left| \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})] \right|^2)^{\frac{1}{2}} \\
& (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)^{\frac{1}{2}} \\
=\ & O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1)O_p(1) = O_p(\frac{1}{\sqrt{N}}),
\end{aligned}
$$

where the last equalities follow from part (1) of Assumption 10.

$$
\begin{aligned}
& V \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0} (\frac{1}{T}\sum\nolimits_{s=1}^{T} \left\| \frac{1}{k}\sum\nolimits_{t=1}^{k}(\frac{1}{N}\sum\nolimits_{i=1}^{N} g_s'\gamma_i e_{it}) g_t' \right\|^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)^{\frac{1}{2}} (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}} \frac{1}{\sqrt{N}} (\sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{k}\frac{1}{\sqrt{N}}\sum\nolimits_{t=1}^{k}\sum\nolimits_{i=1}^{N}\gamma_i e_{it} g_t' \right\|^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)^{\frac{1}{2}} (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}} \frac{1}{\sqrt{N}} (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)^{\frac{1}{2}} \\
& (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \left\| \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{it} \right\|^2)^{\frac{1}{2}} \\
=\ & O_p(\frac{1}{\delta_{NT}})O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1)
\end{aligned}
$$

$$
\begin{aligned}
& VI \\
\leq\ & \|J\| \sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T} g_s(\frac{1}{N}\sum\nolimits_{i=1}^{N} g_s'\gamma_i e_{it}) g_t' \right\| \\
\leq\ & \|J\| \left\| \frac{1}{T}\sum\nolimits_{s=1}^{T} g_s g_s' \right\| \frac{1}{\sqrt{N}} \sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{k}\frac{1}{\sqrt{N}}\sum\nolimits_{t=1}^{k}\sum\nolimits_{i=1}^{N}\gamma_i e_{it} g_t' \right\| \\
\leq\ & \|J\| \left\| \frac{1}{T}\sum\nolimits_{s=1}^{T} g_s g_s' \right\| \frac{1}{\sqrt{N}} (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)^{\frac{1}{2}} (\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \left\| \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{it} \right\|^2)^{\frac{1}{2}} \\
=\ & O_p(1)O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1),
\end{aligned}
$$

where the last equalities follow from part (2) of Assumption 10.

$$
\begin{aligned}
& VII \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k} g_t'(\frac{1}{N}\sum\nolimits_{i=1}^{N} g_t'\gamma_i e_{is})\right\|^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0}(\frac{1}{T}\sum\nolimits_{s=1}^{T}(\frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2\left\|\frac{1}{N}\sum\nolimits_{i=1}^{N}\gamma_i e_{is}\right\|)^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)^{\frac{1}{2}}(\sup_{k\in D^c, k\leq k_0}\frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)\frac{1}{\sqrt{N}}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{is}\right\|^2)^{\frac{1}{2}} \\
=\ & O_p(\frac{1}{\delta_{NT}})O_p(1)\frac{1}{\sqrt{N}}O_p(1).
\end{aligned}
$$

$$
\begin{aligned}
& VIII \\
\leq\ & \|J\| \sup_{k\in D^c, k\leq k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T} g_s g_t'(\frac{1}{N}\sum\nolimits_{i=1}^{N} g_t'\gamma_i e_{is})\right\| \\
\leq\ & \|J\| \sup_{k\in D^c, k\leq k_0}\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}\|g_s\|\,\|g_t\|^2\left\|\frac{1}{N}\sum\nolimits_{i=1}^{N}\gamma_i e_{is}\right\| \\
\leq\ & \|J\|(\sup_{k\in D^c, k\leq k_0}\frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{is}\right\|^2)^{\frac{1}{2}} \\
=\ & O_p(1)O_p(1)O_p(1)\frac{1}{\sqrt{N}}O_p(1),
\end{aligned}
$$

where the equalities follow from $\mathbb{E}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2) \leq M$, which follows from part (ii) of Lemma 1 in Bai and Ng (2002).

Next consider part (5).

$$
\begin{aligned}
& \sup_{k\in D,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\| \\
\leq\ & \sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\begin{array}{c}\frac{1}{T}(\sum_{s=1}^{T}\tilde{g}_s\gamma_N(s,t)+\sum_{s=1}^{T}\tilde{g}_s\zeta_{st} \\ +\sum_{s=1}^{T}\tilde{g}_s\eta_{st}+\sum_{s=1}^{T}\tilde{g}_s\xi_{st})\end{array}\right\|^2\left\|V_{NT}^{-1}\right\|^2 \\
\leq\ & 4\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}(\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}\tilde{g}_s\gamma_N(s,t)\right\|^2+\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}\tilde{g}_s\zeta_{st}\right\|^2 \\
& +\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}\tilde{g}_s\eta_{st}\right\|^2+\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}\tilde{g}_s\xi_{st}\right\|^2)\left\|V_{NT}^{-1}\right\|^2 \\
\leq\ & 8(\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)\gamma_N(s,t)\right\|^2 \\
& +\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}J'g_s\gamma_N(s,t)\right\|^2 \\
& +\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)\zeta_{st}\right\|^2 \\
& +\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}J'g_s\zeta_{st}\right\|^2 \\
& +\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)\eta_{st}\right\|^2 \\
& +\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}J'g_s\eta_{st}\right\|^2 \\
& +\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)\xi_{st}\right\|^2 \\
& +\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}J'g_s\xi_{st}\right\|^2)\left\|V_{NT}^{-1}\right\|^2 \\
=\ & 8(IX+X+XI+XII+XIII+XIV+XV+XVI)\left\|V_{NT}^{-1}\right\|^2.
\end{aligned}
$$

Consider each term one by one.

$$
\begin{aligned}
IX &\leq (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s-J'g_s\|^2)\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\frac{1}{T}\sum\nolimits_{s=1}^{T}|\gamma_N(s,t)|^2 \\
&= O_p(\frac{1}{\delta_{NT}^2})O_p(\frac{1}{T}).
\end{aligned}
$$

$$
\begin{aligned}
X &\leq \|J\|^2 \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2\right)\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\frac{1}{T}\sum\nolimits_{s=1}^{T}|\gamma_N(s,t)|^2 \\
&= O_p(1)O_p(1)O_p(\frac{1}{T}),
\end{aligned}
$$

where the equalities are explained in proof of term I.

$$
\begin{aligned}
& XI \\
\leq\ & \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)\frac{1}{N}\left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right|^2\right) \\
=\ & O_p(\frac{1}{\delta_{NT}^2})\frac{1}{N}O_p(1).
\end{aligned}
$$

$$
\begin{aligned}
& XII \\
\leq\ & \|J\|^2\left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2\right)\frac{1}{N}\left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right|^2\right) \\
=\ & O_p(1)O_p(1)\frac{1}{N}O_p(1),
\end{aligned}
$$

where the equalities follow from part (1) of Assumption 10.

$$
\begin{aligned}
& XIII \\
\leq\ & \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\frac{1}{T}\sum\nolimits_{s=1}^{T}\left|\frac{1}{N}\sum\nolimits_{i=1}^{N}g_s'\gamma_i e_{it}\right|^2 \\
\leq\ & \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)\left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2\right)\frac{1}{N}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{it}\right\|^2 \\
=\ & O_p(\frac{1}{\delta_{NT}^2})O_p(1)\frac{1}{N}O_p(1).
\end{aligned}
$$

$$
\begin{aligned}
XIV &\leq \|J\|^2\left\|\frac{1}{T}\sum\nolimits_{s=1}^{T}g_s g_s'\right\|^2\frac{1}{N}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{it}\right\|^2 \\
&= O_p(1)O_p(1)\frac{1}{N}O_p(1),
\end{aligned}
$$

where the equalities follow from part (2) of Assumption 10.

$$
\begin{aligned}
& XV \\
\leq\ & \left(\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}g_t'\gamma_i e_{is}\right\|^2 \\
\leq\ & \left(\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)\left(\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\|g_t\|^2\right)\frac{1}{N}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2\right) \\
=\ & O_p\left(\frac{1}{\delta_{NT}^2}\right)O_p(1)\frac{1}{N}O_p(1).
\end{aligned}
$$

$$
\begin{aligned}
& XVI \\
\leq\ & \|J\|^2\left(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2\right)\left(\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\|g_t\|^2\right)\frac{1}{N}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2\right) \\
=\ & O_p(1)O_p(1)O_p(1)\frac{1}{N}O_p(1),
\end{aligned}
$$

where the equalities follow from $\mathbb{E}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2) \leq M$, which follows from part (ii) of Lemma 1 in Bai and Ng (2002).

Finally consider part (7).

$$
\begin{aligned}
\sup_{k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}(\tilde{g}_t - J'g_t)g_t'J\right\| \leq\ & \sup_{k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}(\tilde{g}_t - J'g_t)g_t'J\right\| \\
& + \left\|\frac{1}{T-k_0}\sum_{t=k_0+1}^{T}(\tilde{g}_t - J'g_t)g_t'J\right\|.
\end{aligned}
$$

Based on parts (3) and (4), the first term is $O_p(\frac{1}{\delta_{NT}})$. Following the same procedure as part (2), it can be shown the second term is also $O_p(\frac{1}{\delta_{NT}})$. ■

Lemma 6 *Under Assumptions 1-9, terms (1)-(7) in Lemma 5 are* $o_p(1)$.

Proof. The results can be proved following the same procedure as proving Lemma 5, the differences are stated below. Assumption 10 is used in the proof of III, IV, XI, XII, V, VI, $XIII$, XIV to calculate the stochastic order of

$\sup_{k\in D^c,k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right|^2$,

$\sup_{k\in D,k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right|^2$,

$\sup_{k\in D^c,k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{it}\right\|^2$,

$\sup_{k \in D, k \leq k_0} \frac{1}{k} \sum_{t=1}^{k} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{it} \right\|^2.$

Without Assumption 10, all are no longer necessarily $O_p(1)$. Nevertheless, we can use Lemma 4 to show that all are $O_p(\log T)$ without making any dependence assumption on the error process.

Denote $X_t = \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is} e_{it} - \mathbb{E}(e_{is} e_{it})] \right|^2$, then

$$\frac{1}{k} \sum_{t=1}^{k} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is} e_{it} - \mathbb{E}(e_{is} e_{it})] \right|^2 = \frac{1}{k} \sum_{t=1}^{k} X_t.$$

Taking $r = 1$, $\beta_k = k$ and $\alpha_l = M$, then for each m with $1 \leq m \leq T$,

$$\mathbb{E}(\sup_{1 \leq k \leq m} |S_k|) = \mathbb{E}(S_m) \leq mM \leq \sum_{k=1}^{m} \alpha_k, \tag{A-18}$$

hence by Lemma 4,

$$\mathbb{E}(\sup_{1 \leq k \leq k_0} \left| \frac{S_k}{k} \right|) \leq 4 \sum_{k=1}^{k_0} \frac{M}{k} \leq 4M \log T + 4M\gamma, \tag{A-19}$$

where γ is the Euler-Mascheroni constant. It follows that

$\sup_{k \in D^c, k \leq k_0} \frac{1}{k} \sum_{t=1}^{k} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is} e_{it} - \mathbb{E}(e_{is} e_{it})] \right|^2,$

$\sup_{k \in D, k \leq k_0} \frac{1}{k} \sum_{t=1}^{k} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is} e_{it} - \mathbb{E}(e_{is} e_{it})] \right|^2$ are both $O_p(\log T)$. All other terms can be proved to be $O_p(\log T)$ similarly. Now $III = O_p(\frac{\sqrt{\log T}}{\sqrt{N} \delta_{NT}})$, $IV = O_p(\sqrt{\frac{\log T}{N}})$, $V = O_p(\frac{\sqrt{\log T}}{\sqrt{N} \delta_{NT}})$, $VI = O_p(\sqrt{\frac{\log T}{N}})$, $XI = O_p(\frac{\log T}{N \delta_{NT}^2})$, $XII = O_p(\frac{\log T}{N})$, $XIII = O_p(\frac{\log T}{N \delta_{NT}^2})$ and $XIV = O_p(\frac{\log T}{N})$. With Assumption 9, all terms are $o_p(1)$. ■

Lemma 7 *Under Assumptions 1-8 and 9 or 10,*

(1) $\sup_{k \in D^c, k \leq k_0} \frac{1}{k_0} \frac{1}{k} \left\| \sum_{t=1}^{k} z_t \right\|^2 = o_p(1)$, *(2)* $\sup_{k \in D, k \leq k_0} \frac{1}{k_0} \frac{1}{k} \left\| \sum_{t=1}^{k} z_t \right\|^2 = o_p(1)$,

(3) $\sup_{k \in D^c, k \leq k_0} \frac{1}{k_0} \left\| \sum_{t=1}^{k} z_t \right\| = o_p(1)$, *(4)* $\sup_{k \in D, k \leq k_0} \frac{1}{k_0} \left\| \sum_{t=1}^{k} z_t \right\| = o_p(1)$,

(5) $\sup_{k \in D^c, k < k_0} \frac{1}{k_0 - k} \left\| \sum_{t=k+1}^{k_0} z_t \right\| = o_p(1)$, *(6)* $\sup_{k \in D, k < k_0} \frac{1}{k_0 - k} \left\| \sum_{t=k+1}^{k_0} z_t \right\| = o_p(1)$,

(7) $\sup_{k \in D^c, k < k_0} \frac{1}{k_0} \frac{1}{k_0 - k} \left\| \sum_{t=k+1}^{k_0} z_t \right\|^2 = o_p(1)$, *(8)* $\sup_{k \in D, k < k_0} \frac{1}{k_0} \frac{1}{k_0 - k} \left\| \sum_{t=k+1}^{k_0} z_t \right\|^2 = o_p(1)$,

(9) $\sup_{k \leq k_0} \left\| \frac{1}{T-k} \sum_{t=k+1}^{T} z_t \right\| = o_p(1)$.

Proof. We will prove the results under Assumptions 1-8 and 10 first. Under Assumptions 1-9, the proof follows the same procedure, except for using Lemma 6 instead of Lemma 5.

Recall that $z_t = vec[(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'] + vec[(\tilde{g}_t - J'g_t)g_t'J] + vec[J'g_t(\tilde{g}_t - J'g_t)'] + vec[(J' - J_0')g_tg_t'(J - J_0)] + vec[(J' - J_0')g_tg_t'J_0] + vec[J_0'g_tg_t'(J - J_0)]$.

For parts (1) and (2),

$$\begin{aligned} & \left\|\sum_{t=1}^{k} z_t\right\|^2 \\ \leq\ & (\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| + 2\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\| \\ & + \left\|\sum_{t=1}^{k}(J' - J_0')g_tg_t'(J - J_0)\right\| + 2\left\|\sum_{t=1}^{k}(J' - J_0')g_tg_t'J_0\right\|)^2 \\ \leq\ & 4\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\|^2 + 16\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\|^2 \\ & +4\left\|\sum_{t=1}^{k}(J' - J_0')g_tg_t'(J - J_0)\right\|^2 + 16\left\|\sum_{t=1}^{k}(J' - J_0')g_tg_t'J_0\right\|^2. \end{aligned} \tag{A-20}$$

Consider the four terms one by one.

Using Lemma 1,

$$\begin{aligned} & \sup_{k\in D, k\leq k_0} \frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\|^2 \\ \leq\ & \frac{1}{\tau_0(\tau_0 - \eta)}(\frac{1}{T}\sum_{t=1}^{T}\|\tilde{g}_t - J'g_t\|^2)^2 = O_p(\frac{1}{\delta_{NT}^4}). \end{aligned}$$

Using part (6) of Lemma 5,

$$\begin{aligned} & \sup_{k\in D^c, k\leq k_0} \frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\|^2 \\ \leq\ & \sup_{k\in D^c, k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\|^2 = O_p(\frac{1}{\delta_{NT}^4}). \end{aligned}$$

Using part (1) of Lemma 5,

$$\sup_{k\in D, k\leq k_0} \frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\|^2 \leq \sup_{k\in D, k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\|^2 = O_p(\frac{1}{\delta_{NT}^2}).$$

Using part (2) of Lemma 5,

$$\sup_{k\in D^c, k\leq k_0} \frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\|^2 \leq \sup_{k\in D^c, k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\|^2 = O_p(\frac{1}{\delta_{NT}^2}).$$

Using Lemma 2 and Assumption 3,

$$\sup_{k\le k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\|^2 \le \|J-J_0\|^4\sup_{k\le k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}g_tg_t'\right\|^2 = o_p(1),$$

$$\sup_{k\le k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\|^2 \le \|J-J_0\|^2\|J_0\|^2\sup_{k\le k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}g_tg_t'\right\|^2 = o_p(1).$$

It follows $\sup\limits_{k\in D,k\le k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}z_t\right\|^2$ and $\sup\limits_{k\in D^c,k\le k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}z_t\right\|^2$ are both $o_p(1)$.

For parts (3) and (4),

$$\begin{aligned}&\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}z_t\right\|\\ \le\ &\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|+2\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|\\ &+\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\|+2\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\|.\end{aligned}\qquad\text{(A-21)}$$

Using Lemma 1,

$$\sup_{k\in D,k\le k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\| \le \frac{1}{\tau_0}\frac{1}{T}\sum\nolimits_{t=1}^{T}\|\tilde{g}_t-J'g_t\|^2 = O_p(\frac{1}{\delta_{NT}^2}),$$

$$\sup_{k\in D^c,k\le k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\| \le \frac{1}{\tau_0}\frac{1}{T}\sum\nolimits_{t=1}^{T}\|\tilde{g}_t-J'g_t\|^2 = O_p(\frac{1}{\delta_{NT}^2}).$$

Using part (1) of Lemma 5,

$$\sup_{k\in D,k\le k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\| \le \sup_{k\in D,k\le k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}}).$$

Using part (2) of Lemma 5,

$$\sup_{k\in D^c,k\le k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\| \le \sup_{k\in D^c,k\le k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}}).$$

Using Lemma 2 and Assumption 3,

$$\sup_{k\in D,k\le k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\| \le \|J-J_0\|^2\sup_{k\in D,k\le k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}g_tg_t'\right\| = o_p(1),$$

$$\sup_{k\in D^c,k\le k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\| \le \|J-J_0\|^2\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}g_tg_t'\right\| = o_p(1),$$

$$\sup_{k\in D, k\le k_0} \frac{1}{k_0} \left\| \sum_{t=1}^{k} (J' - J_0') g_t g_t' J_0 \right\| \le \|J - J_0\| \, \|J_0\| \sup_{k\in D, k\le k_0} \left\| \frac{1}{k} \sum_{t=1}^{k} g_t g_t' \right\| = o_p(1),$$

$$\sup_{k\in D^c, k\le k_0} \frac{1}{k_0} \left\| \sum_{t=1}^{k} (J' - J_0') g_t g_t' J_0 \right\| \le \|J - J_0\| \, \|J_0\| \sup_{k\in D^c, k\le k_0} \left\| \frac{1}{k} \sum_{t=1}^{k} g_t g_t' \right\| = o_p(1).$$

It follows that $\sup_{k\in D, k\le k_0} \frac{1}{k_0} \left\| \sum_{t=1}^{k} z_t \right\|$ and $\sup_{k\in D^c, k\le k_0} \frac{1}{k_0} \left\| \sum_{t=1}^{k} z_t \right\|$ are both $o_p(1)$. parts (5), (6), (7), (8) and (9) can be proved following the same procedure. More specifically, part (5) uses Lemma 1, Lemma 2, part (3) of Lemma 5 and Lemma 3; part (6) uses parts (5) and (4) of Lemma 5, Lemma 2 and Lemma 3; parts (7) and (8) follow from (5) and (6) respectively; part (9) uses Lemma 1, Lemma 2, part (7) of Lemma 5 and $\sup_{k\le k_0} \left\| \frac{1}{T-k} \sum_{t=k+1}^{T} g_t g_t' \right\| = O_p(1)$, which is proved below.

$$\begin{aligned}
\sup_{k\le k_0} \left\| \frac{1}{T-k} \sum_{t=k+1}^{T} g_t g_t' \right\| &\le \sup_{k<k_0} \left\| \frac{1}{T-k} \sum_{t=k+1}^{T} g_t g_t' \right\| + \left\| \frac{1}{T-k_0} \sum_{t=k_0+1}^{T} g_t g_t' \right\| \\
&\le \sup_{k<k_0} \left\| \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} g_t g_t' \right\| + 2 \left\| \frac{1}{T-k_0} \sum_{t=k_0+1}^{T} g_t g_t' \right\| \\
&= O_p(1).
\end{aligned}$$

■

Essay III: Estimating and Testing High Dimensional Factor Models with Multiple Structural Changes

1 INTRODUCTION

High dimensional factor models have played a crucial role in business cycle analysis, consumer behavior analysis, asset pricing and macroeconomic forecasting, see for example, Ross (1976), Lewbel (1991), Bernanke, Boivin and Eliasz (2005) and Stock and Watson (2002a, 2002b), to mention a few. This has been enhanced by the increasing availability of big data sets. However, as the time span of the data becomes longer, there is a substantial risk that the underlying data generating process may experience structural changes. Inference ignoring these changes would be misleading. This paper considers multiple changes in the factor loadings of a high dimensional factor model, occurring at dates that are unknown but common to all subjects. We propose a joint estimator of all the change points as well as a sequential estimator of the change points that estimates these change points one by one. Based on the estimated change points, we are able to consistently determine the number of factors and estimate the factor space in each regime. We also propose tests for (i) the null of no change versus the alternative of some fixed number of changes and (ii) tests for the null of l changes versus the alternative of $l+1$ changes. The latter allows us to consistently determine the number of changes. These tests are easy to implement and critical values tabulated in Bai and Perron (1998, 2003) can be used directly to make inference on the presence as well as the number of structural changes. In addition, we also discuss reestimating the change points by direct least squares based on the estimated number of factors in each regime. The reestimated change points tend to be more accurate[1] and could improve the finite sample performance of subsequent estimation and testing procedures.

Bates, Plagborg-Møller, Stock and Watson (2013) argue that as long as the magnitude of the loading breaks converges to zero sufficiently fast, existing estimators ignoring loading breaks are still consistent. However, the conditions required are relatively stringent and the resulting asymptotic properties may not provide a good approximation. Recently, several tests on the stability of the factor loadings in high dimensional factor models have been proposed including Breitung and Eickmeier (2011), Chen, Dolado and Gonzalo (2014), Han and Inoue (2014) and Cheng, Liao and Schorfheide (2014). Recent contributions on estimating high dimensional factor models with loading instability include Baltagi, Kao and Wang

[1] The reestimated change points are applicable only when the number of factors is the same for all regimes.

(2015b), Cheng, Liao and Schorfheide (2014), Massacci (2015) and Shi (2015). However, all of these papers consider the case with a single change.

This paper tackles multiple changes in high dimensional factor models[2]. We start by estimating the number of factors and factor space ignoring structural changes. Since the factor model with changes in the loadings can be equivalently written as another factor model with stable loadings but pseudo factors, this would allow us to identify the equivalent model with stable loadings and give us the estimated pseudo factors. A key observation is that the mean of the second moment matrix of the pseudo factors have changes at exactly the same dates as the loadings. Estimating and testing multiple changes in the latter can be converted to estimating and testing multiple changes in the former. This conversion is crucial because the true factors are unobservable and not estimable without knowing the change points. It is also worth pointing out that after this conversion we are using the estimated pseudo factors, not the pseudo factors themselves. That is to say, the data contains measurement error. We will show that this measurement error has a different effect on testing and estimating structural changes. Once the estimated change points are available, they are plugged in to split the sample and estimate the number of factors and factor space in each regime. The former further enables reestimating the change points by direct least squares while the latter allows us to construct the test for l versus $l+1$ changes.

In the regression setup, influential work on multiple changes include Bai and Perron (1998) and Qu and Perron (2007). This paper differs from these seminal papers in several respects. First, the current paper deals with a high dimensional setup with unobservable regressors, while their papers deal with a fixed dimensional setup with observable regressors. Thus their results are not directly applicable here. Second, in the current setup estimating the number of pseudo factors at the outset plays the role of selecting relevant moment conditions among a large number of candidates while in their setup the moment conditions are known a priori. From this perspective, estimating the number of factors is intrinsically connected to the many instrumental variables literature. Third, after conversion, the data is fixed dimensional with observable regressors[3] and thus conceptually fits into their setup. However, it still relies on high dimension to eliminate the effect of measurement error. Moreover, we

[2] In testing the joint hypothesis of structural stability of both factor loadings and the factor augmented forecasting equation, Corradi and Swanson (2014) also consider the alternative of multiple changes.

[3] The regressors are ones, since the second moment matrix of the pseudo factors is a multivariate time series with mean shifts.

show that to eliminate the effect of measurement error on testing structural changes, we require $\frac{\sqrt{T}}{N} \to 0$ as the dimension N and the sample size T go to infinity jointly, while for estimating change points we only require N and T go to infinity jointly. The latter result is rare in the sense that in the high dimensional econometrics literature very few papers require no N-T (T-N) ratio condition. The latter result is also different from the literature in which the estimated factors are used. For example, Bai and Ng (2006) require $\frac{\sqrt{T}}{N} \to 0$ where estimated factors are used to augment forecasting and vector autoregression. Various N-T ratio conditions are also needed in Bai (2009) where estimated factors are used to control the interactive effects in panel data. The explanation is mainly related to the local nature of the identification of change points, which is also the reason that the estimated change points are inconsistent as $T \to \infty$. For a detailed explanation, see Section 3.1.3. Fourth, the second step in this paper is to estimate the number of factors and factor space in each regime while their second step is to estimate the regression coefficients in each regime. In their setup, it is a direct corollary that the second step is not affected by the inconsistency of the estimated change points, while in the current setup it is not so obvious and requires new analysis. This is because estimating the number of factors and factor space is totally different from estimating the regression coefficients. Fifth, in their setup, due to the fixed dimensionality, the convergence rate of the estimated change points is at best $O_p(1)$, while in the current setup due to the high dimensionality, the reestimated change points could be consistent.

Throughout the paper, $\|A\| = (trAA')^{\frac{1}{2}}$ denotes the Frobenius norm, $\xrightarrow{p}$, $\xrightarrow{d}$ and $\Rightarrow$ denotes convergence in probability, convergence in distribution and weak convergence of stochastic process respectively, $vech(A)$ denotes the half vectorization of matrix A, $\mathbb{E}(\cdot)$ denotes the expectation, $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$ and $(N, T) \to \infty$ denotes N and T going to infinity jointly.

The rest of the paper is organized as follows: Section 2 introduces the model setup, notation and preliminaries. Section 3 considers both joint estimation and sequential estimation of the change points and also the subsequent estimation of the number of factors and factor space in each regime. Section 4 proposes test statistics for multiple changes, derives their asymptotic distributions and discusses how to determine the number of changes. Section 5 discusses reestimating the change points by direct least squares. Section 6 presents simulation results. Section 7 concludes. All the proofs are relegated to the appendix.

2 NOTATION AND PRELIMINARIES

2.1 The Model

Consider the following high dimensional factor model with L changes in the factor loadings:

$$x_{it} = f_{0,t}'\lambda_{0,i} + f_{-0,t}'\lambda_{\kappa,i} + e_{it}, \tag{1}$$

with $k_{\kappa-1,0} + 1 \leq t \leq k_{\kappa,0}$, for $\kappa = 1, ..., L+1, i = 1, ..., N$ and $t = 1, ..., T$, where $f_{0,t}$ and $f_{-0,t}$ are $r-q$ and q dimensional vectors of factors without and with changes in the loadings respectively. Let $f_t = (f_{0,t}', f_{-0,t}')'$. $\lambda_{0,i}$ and $\lambda_{\kappa,i}$ are factor loadings of subject i corresponding to $f_{0,t}$ and $f_{-0,t}'$ in the κ-th regime, respectively. Let $\lambda_{0\kappa,i} = (\lambda_{0,i}', \lambda_{\kappa,i}')'$. e_{it} is the error term allowed to have temporal and cross-sectional dependence as well as heteroskedasticity. $k_{\kappa,0}$ and $\tau_{\kappa,0} = \frac{k_{\kappa,0}}{T}$, $\kappa = 1, ..., L$ are the change points and change fractions respectively, and note that $k_{0,0} = 0$ and $k_{L+1,0} = T$. The goal is to estimate the change points, determine the number of factors and estimate the factors and loadings in each regime.

In matrix form, the model can be expressed as follows:

$$X_{\kappa*} = F_{0\kappa*}\Lambda_0' + F_{-0\kappa*}\Lambda_\kappa' + E_{\kappa*}, \text{ for } \kappa = 1, ..., L+1. \tag{2}$$

$X_{\kappa*} = (x_{k_{\kappa-1,0}+1}, ..., x_{k_{\kappa,0}})'$ and $E_{\kappa*} = (e_{k_{\kappa-1,0}+1}, ..., e_{k_{\kappa,0}})'$ are both of dimension $(k_{\kappa,0} - k_{\kappa-1,0}) \times N$. $F_{0\kappa*} = (f_{0,k_{\kappa-1,0}+1}, ..., f_{0,k_{\kappa,0}})'$ and $F_{-0\kappa*} = (f_{-0,k_{\kappa-1,0}+1}, ..., f_{-0,k_{\kappa,0}})'$ are of dimensions $(k_{\kappa,0} - k_{\kappa-1,0}) \times (r-q)$ and $(k_{\kappa,0} - k_{\kappa-1,0}) \times q$ respectively. Here we use "$\kappa *$" to denote that the sample split is based on the true change points. $\Lambda_0 = (\lambda_{0,1}, ..., \lambda_{0,N})'$ and $\Lambda_\kappa = (\lambda_{\kappa,1}, ..., \lambda_{\kappa,N})'$ are of dimensions $N \times (r-q)$ and $N \times q$ respectively. Also, let $F_{\kappa*} = (F_{0\kappa*}, F_{-0\kappa*}) = (f_{k_{\kappa-1,0}+1}, ..., f_{k_{\kappa,0}})'$ and $\Lambda_{0\kappa} = (\Lambda_0, \Lambda_\kappa) = (\lambda_{0\kappa,1}, ..., \lambda_{0\kappa,N})'$.

Note that in model (1), changes in the number of factors is allowed for, and incorporated as a special case of changes in the loadings by allowing Λ_κ to contain some zero columns for some κ. Let q_κ be the number of nonzero columns in Λ_κ, then $q = \max\{q_\kappa, \kappa = 1, ..., L+1\}$ and the number of factors in the κ-th regime is $r_\kappa = r - q + q_\kappa$. To simplify the analysis, we shall only consider the case where the matrix that contains all the different nonzero columns of Λ_0 and $\Lambda_\kappa, \kappa = 1, ..., L+1$ is full rank.

2.2 Equivalent Representation

Let Λ_{-0} contain all the different nonzero columns in Λ_κ for $\kappa = 1, ..., L+1$ and $\Gamma = (\Lambda_0, \Lambda_{-0})$. It follows that $\Lambda_{0\kappa} = \Gamma R_\kappa$, where R_κ is a selection matrix. Let $G_{\kappa *} = F_{\kappa *} R'_\kappa$, it follows that $g_t = R_\kappa f_t$ if $k_{\kappa-1,0} + 1 \leq t \leq k_{\kappa,0}$ and

$$X_{\kappa *} = F_{\kappa *} \Lambda'_{0\kappa} + E_{\kappa *} = F_{\kappa *} R'_\kappa \Gamma' + E_{\kappa *} = G_{\kappa *} \Gamma' + E_{\kappa *}. \tag{3}$$

Equation (3) is a factor model with stable loadings but pseudo factors g_t, whose number $\bar{r}$ is equal to the column rank of Γ.

We next argue that as long as $\frac{1}{k_{\kappa,0} - k_{\kappa-1,0}} \sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} f_t f'_t - \Sigma_F \xrightarrow{p} 0$ for each κ and $\left\| \frac{1}{N} \Gamma' \Gamma - \Sigma_\Gamma \right\| \to 0$ for some positive definite Σ_F and Σ_Γ[4], then $\frac{1}{T} \sum_{t=1}^{T} g_t g'_t - \Sigma_G \xrightarrow{p} 0$ for some positive definite Σ_G. This ensures the uniqueness (up to a rotation) of the equivalent representation. First, it is not difficult to see that $\Sigma_G = \sum_{\kappa=1}^{L+1} (\tau_{\kappa,0} - \tau_{\kappa-1,0}) \Sigma_{G,\kappa}$, where $\Sigma_{G,\kappa} = R_\kappa \Sigma_F R'_\kappa$ is positive semidefinite for all κ. Thus for any $\bar{r}$ dimensional vector v, $v' \Sigma_G v = 0$ implies $v' \Sigma_{G,\kappa} v = 0$ for all κ, which further implies $v' R_\kappa = 0$ for all κ. Since R_κ is a selection matrix and each element of v is selected by at least one R_κ, each element of v must be zero, and therefore Σ_G is positive definite.

3 ESTIMATING MODELS WITH MULTIPLE CHANGES

In this section, we propose a two step procedure to estimate model (1) when the number of breaks is known. How to determine the number of breaks will be discussed later. The first step is estimating the change points. We propose a joint estimator for all change points as well as a sequential estimator which estimates the change points one by one. For both estimators, we show that the distance between the estimated and the true change points is $O_p(1)$. The second step is plugging in the estimated change points and estimating the number of factors and the factor space in each regime. We show that although the estimated change points are inconsistent, using the estimated change points does not affect the consistency of the estimated number of factors, nor the convergence rate of the estimated factor space.

[4] In case Σ_Γ is degenerate, the factors and loadings can be further transformed to regain positive definiteness. Here we do not consider this case.

3.1 Joint Estimation of the Change Points

We first introduce the estimation procedure, and then impose assumptions to study the asymptotic properties of the proposed estimators.

3.1.1 Estimation Procedure

The estimation procedure is as follows:

1. Estimate the number of factors ignoring structural changes. Let $\tilde{r}$ be the estimated number of factors.

2. Estimate the first $\tilde{r}$ factors using the principal component method. Let $\tilde{g}_t, t = 1, ..., T$ be the estimated factors[5].

3. For any partition $(k_1, ..., k_L)$[6], split the sample into $L + 1$ subsamples, estimate the second moment matrix of g_t in each subsample as $\tilde{\Sigma}_\kappa = \frac{1}{k_\kappa - k_{\kappa-1}} \sum_{t=k_{\kappa-1}+1}^{k_\kappa} \tilde{g}_t \tilde{g}_t'$ and calculate the sum of squared residuals,

$$\tilde{S}(k_1, ..., k_L) = \sum_{\kappa=1}^{L+1} \sum_{t=k_{\kappa-1}+1}^{k_\kappa} [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_\kappa)]'[vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_\kappa)]. \quad (4)$$

Then estimate the change points by minimizing the sum of squared residuals,

$$(\tilde{k}_1, ..., \tilde{k}_L) = \arg\min \tilde{S}(k_1, ..., k_L). \quad (5)$$

The underlying mechanism of the above procedure is as follows:

1. Since model (2) has equivalent representation (3), $\tilde{r}$ is consistent for $\bar{r}$, $\tilde{g}_t$ is asymptotically close to $J' g_t$ for some rotation matrix J , and $J' g_t$ is asymptotically close to $J_0' g_t$, where $J \xrightarrow{p} J_0 = \Sigma_\Gamma^{\frac{1}{2}} \Phi V^{-\frac{1}{2}}$, with V being the diagonal matrix of eigenvalues of $\Sigma_\Gamma^{\frac{1}{2}} \Sigma_G \Sigma_\Gamma^{\frac{1}{2}}$ and Φ the corresponding eigenvector matrix.

2. The second moment matrix of g_t has breaks at the same points as the factor loadings.

3. The second moment matrix of $J_0' g_t$ has breaks at the same points as g_t.

[5] The change points estimator also can be based on $\hat{g}_t$, where $(\hat{g}_1, ..., \hat{g}_T)' = \hat{G} = \tilde{G} V_{NT} = (\tilde{g}_1, ..., \tilde{g}_T)' V_{NT}$ and V_{NT} is a diagonal matrix that contains the first $\bar{r}$ largest eigenvalues of $\frac{1}{NT} X X'$.

[6] $k_0 = 0$ and $k_{L+1} = T$.

More precisely, let $\mathbb{E}(f_t f_t') = \Sigma_F$ for all t. $\Sigma_\kappa = J_0' \Sigma_{G,\kappa} J_0$ is the mean of $J_0' g_t g_t' J_0$ and $y_t = vech(J_0' g_t g_t' J_0 - \Sigma_\kappa)$ for $t = k_{\kappa-1,0}+1, ..., k_{\kappa,0}$ with $\kappa = 1, ..., L+1$, and $z_t = vech(\tilde{g}_t \tilde{g}_t' - J_0' g_t g_t' J_0)$ for $t = 1, ..., T$. It follows that $vech(\tilde{g}_t \tilde{g}_t') = vech(\Sigma_\kappa) + y_t + z_t$ for $t = k_{\kappa-1,0}+1, ..., k_{\kappa,0}$ and $\kappa = 1, ..., L+1$. Since $R_{\kappa-1}$ and R_κ are two different selection matrices, $\Sigma_{\kappa-1} = J_0' R_{\kappa-1} \Sigma_F R_{\kappa-1}' J_0 \neq J_0' R_\kappa \Sigma_F R_\kappa' J_0 = \Sigma_\kappa$. Thus $vech(\tilde{g}_t \tilde{g}_t')$ is a multivariate process with L mean shifts and measurement error z_t. We will show that to asymptotically eliminate the effect of z_t, this requires $(N, T) \to \infty$ and no N-T ratio condition is needed.

Also note that through estimating the number of pseudo factors, we are essentially selecting relevant moment conditions from a large number of candidates. The model with $\tilde{r} = \bar{r}$ has the strongest identification strength for the unknown change points. If $\tilde{r} > \bar{r}$, no information would be lost, but extra noise would be brought in by the extra estimated factors and consequently the identification strength of the change points would be weaker. This is quite similar to Moon and Weidner (2014), who show that for panel data the limiting distribution of the least squares estimator is not affected by overestimation of the number of factors used to control the interactive effects. If $\tilde{r} < \bar{r}$, change point estimation would be based on a subset of $vech(\tilde{g}_t \tilde{g}_t')$. Thus identification of the change points would be weaker or even totally lost. To improve the finite sample performance, we may simply fix $\tilde{r}$ at the maximum of pseudo factors if this maximum is small or some prior information is available. Also, we recommend choosing a less conservative criterion in estimating $\bar{r}$.

3.1.2 Assumptions

The assumptions are as follows:

Assumption 1 $\mathbb{E}\|f_t\|^4 < M < \infty$, $\mathbb{E}(f_t f_t') = \Sigma_F$ *for all* t. Σ_F *is positive definite and* $\frac{1}{k_{\kappa,0} - k_{\kappa-1,0}} \sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} f_t f_t' - \Sigma_F = o_p(1)$ *for* $\kappa = 1, ..., L+1$.

Assumption 2 $\|\lambda_{0\kappa,i}\| \leq \bar{\lambda} < \infty$ *for* $\kappa = 1, ..., L+1$, *and* $\left\|\frac{1}{N}\Gamma'\Gamma - \Sigma_\Gamma\right\| \to 0$ *for some positive definite matrix* Σ_Γ.

Assumption 3 *There exists a positive constant* $M < \infty$ *such that:*

1. $\mathbb{E}(e_{it}) = 0$ *and* $\mathbb{E}|e_{it}|^8 \leq M$ *for all* i *and* t,

2. $\mathbb{E}(e_{it} e_{js}) = \tau_{ij,ts}$ *for all* i, j *and* t, s, *and* $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \leq M$,

3. $\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right|^4 \leq M$ *for all* s,t.

Assumption 4 *There exists an* $M<\infty$ *such that:*

1. $\mathbb{E}(\frac{e_s' e_t}{N})=\gamma_N(s,t)$ *and* $\sum_{s=1}^{T}|\gamma_N(s,t)|\leq M$ *for all* t,

2. $\mathbb{E}(e_{it}e_{jt})=\tau_{ij,t}$ *with* $|\tau_{ij,t}|\leq\tau_{ij}$ *for some* τ_{ij} *and for all* t, *and* $\sum_{j=1}^{N}|\tau_{ji}|\leq M$ *for all* i.

Assumption 5 *The largest eigenvalue of* $\frac{1}{NT}EE'$ *is* $O_p(\frac{1}{\delta_{NT}^2})$.

Assumption 6 *The eigenvalues of* $\Sigma_G\Sigma_\Gamma$ *are distinct.*

Assumption 7 *Define* $\epsilon_t = vech(f_tf_t'-\Sigma_F)$,

1. The data generating process of the factors is such that the Hajek-Renyi inequality[7] *applies to the process* $\{\epsilon_t, t=k_{\kappa-1,0}+1,...,k_{\kappa,0}\}$ *and* $\{\epsilon_t, t=k_{\kappa,0},...,k_{\kappa-1,0}+1\}$ *for* $\kappa=1,...,L+1$,

2. There exist $\delta>0$ *and* $M<\infty$ *such that for* $\kappa=1,...,L+1$ *and for all* $k_{\kappa-1,0}<k<l\leq k_{\kappa,0}$, $\mathbb{E}(\left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l}\epsilon_t\right\|^{4+\delta})<M$.

Assumption 8 *There exists* $M<\infty$ *such that:*

1. $\mathbb{E}(\sup_{0\leq k<l\leq T}\frac{1}{l-k}\sum_{t=k+1}^{l}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right|^2)\leq M$ *for all* s,

2. $\mathbb{E}(\sup_{0\leq k<l\leq T}\frac{1}{l-k}\sum_{t=k+1}^{l}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{it}\right\|^2)\leq M$.

Assumption 1 corresponds to Assumption A in Bai (2003). It requires the law of large number to be applicable to factors within each regime, thus f_t can be dynamic and contain lags. Note that the second moment matrix of the factors is assumed to be stationary over time. Assumption 2 corresponds to Assumption B in Bai (2003). Note that within the statistical setup, only changes in loadings are identifiable, changes in factor identities are not identifiable. And no matter whether the loadings change or not, the factor identities could either change or not. The identities of factors should be determined by other sources

[7] Hajek-Renyi inequality is crucial for pinning down the order of the estimation error in the estimated change points, see the Appendix for more details.

of information. Assumptions 3 and 4 correspond to Assumptions C and E in Bai (2003). Both temporal and cross-sectional dependence as well as heteroskedasticity are allowed for. Assumption 5 is the key condition for determining the number of factors and is required in almost all existing methods. For example, Onatski (2010) and Ahn and Horenstein (2013) assume $E = A\varepsilon B$, where ε is an i.i.d. $T \times N$ matrix and A and B characterize the temporal and cross-sectional dependence and heteroskedasticity. This is a sufficient but not necessary condition for Assumption 5. Also note that once Assumption 5 is imposed, Assumption D in Bai (2003) is not needed. In other words, for the purpose of determining the number of factors, factors could be correlated with the errors. Assumption 6 corresponds to Assumption G in Bai (2003) and ensures uniqueness of the principal component estimator. Assumption 7 imposes a further requirement on the factor process. Instead of assuming a specific data generating process, we require the Hajek-Renyi inequality to be applicable to the second moment process of the factors, so that Assumption 7 is in its most general form. Assumption 8 imposes further constraints on the errors. Assumption 3(3) and Assumption F3 in Bai (2003) imply that the summands are uniformly $O_p(1)$. Assumption 8 strengthens this condition such that the supremum of the average of these summands is $O_p(1)$.

3.1.3 Asymptotic Properties of the Joint Estimator

First note that due to the consistency of $\tilde{r}$ for $\bar{r}$, treating $\bar{r}$ as known will not affect the asymptotic properties of the change point estimator. In what follows we shall show that the distance between the estimated and the true change points is $O_p(1)$. This allows us to identify the number of factors and estimate the factor space in each regime. Define $\tilde{\tau}_\iota = \tilde{k}_\iota / T$ as the estimated change fraction, we first show that $\tilde{\tau}_\iota$ is consistent.

Proposition 1 *Under Assumptions 1-8,* $\tilde{\tau}_\iota - \tau_{\iota 0} = o_p(1)$ *for* $\iota = 1, ..., L$ *as* $(N, T) \to \infty$.

This proposition is important for theoretical purposes. The key observation for its proof and even the whole change point estimation literature is that for any possible region of the change points $\mathcal{O}$, $P((\tilde{k}_1, ..., \tilde{k}_L) \in \mathcal{O})$ is controlled by $P(\min_{(k_1,...,k_L)\in\mathcal{O}} \tilde{S}(k_1, ..., k_L) - \tilde{S}(k_{10}, ..., k_{L0}) \leq 0)$. Based on Proposition 1 and utilizing this observation, we can prove:

Theorem 1 *Under Assumptions 1-8,* $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$ *for* $\iota = 1, ..., L$ *as* $(N, T) \to \infty$.

This theorem implies that no matter how large T is, the possible change points are narrowed to a bounded interval of the true change points. Note that the measurement error

z_t has no effect (asymptotically) on the estimated change points as long as $(N, T) \to \infty$. No N-T ratio condition is needed. This is different from factor-augmented forecasting and factor-augmented vector autoregression (FAVAR), in which $\frac{\sqrt{T}}{N} \to 0$ is required to asymptotically eliminate the effect of using estimated factors. The reason is that identification of the change points relies on observations within a local region of the true change points and consequently the measurement error will not accumulate as $T \to \infty$. In contrast, factor-augmented forecasting and FAVAR relies on all observations and consequently the measurement error will accumulate as $T \to \infty$. Since for each t the measurement error z_t converges to zero as $N \to \infty$, for change points estimation $N \to \infty$ is enough to eliminate the effect of measurement error, while for factor-augmented forecasting and FAVAR, N need to be large relative to T.

Remark 1 *The limiting distribution of $\tilde{k}_\iota - k_{\iota 0}$ has the same form as the single change case. This is because $\tilde{k}_\iota$ also minimizes the sum of squared residuals for the subsample $t = \tilde{k}_{\iota-1} + 1, ..., \tilde{k}_{\iota+1}$. Since $\tilde{k}_{\iota-1} - k_{\iota-1,0}$ and $\tilde{k}_{\iota+1} - k_{\iota+1,0}$ are both $O_p(1)$, $\tilde{k}_\iota$ has the same limiting distribution as the minimizer of the subsample $t = k_{\iota-1,0} + 1, ..., k_{\iota+1,0}$. For more details about the form as well as a proof of the limiting distribution, see Baltagi, Kao and Wang (2015b).*

3.2 Sequential Estimation of the Change Points

This section proposes sequential estimation of the change points one by one, each time treating the model as if there is only one change point. The first two steps are the same as the joint estimation while the third step is slightly adjusted: For any partition k_1, split the sample into two subsamples, estimate the second moment matrix of g_t in each subsample and calculate the sum of squared residuals, $\tilde{S}(k_1) = \sum_{\kappa=1}^{2} \sum_{t=k_{\kappa-1}+1}^{k_\kappa} [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_\kappa)]'[vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_\kappa)]$, then $\hat{k}_1 = \arg\min \tilde{S}(k_1)$. Compared to joint estimation, the main advantage of sequential estimation is that it does not require knowing the number of changes. Instead, together with sequential testing, it allows us to determine the number of changes.

In what follows, we shall show that the distance between the sequentially estimated and the true change points is also $O_p(1)$. First, define $S_0(\tau)$ as the reduction in the sum of squared residuals when $y_t = 0$ and $z_t = 0$ is plugged in to split the sample. If y_t and z_t are indeed zero for all t, the estimated change fraction should be equal to τ among $\tau_{1,0}, ..., \tau_{L,0}$

that leads to the largest reduction in the sum of squared residuals. To simplify the analysis, we require $S_0(\tau_{\iota,0})$ to be different for different ι, and without loss of generality, we assume:

Assumption 9 $S_0(\tau_{1,0}) < ... < S_0(\tau_{L,0})$.

In general, y_t and z_t are not zero for all t, but asymptotically this does not affect the result.

Proposition 2 *Under Assumptions 1-9,* $\hat{\tau}_1 - \tau_{1,0} = o_p(1)$ *as* $(N,T) \to \infty$.

Similar to the joint estimation, the proof is that for any possible region of the change points $\mathcal{O}$, $P(\hat{k}_1 \in \mathcal{O})$ is controlled by $P(\min_{k_1 \in \mathcal{O}} \tilde{S}(k_1) - \tilde{S}(k_{10}) \leq 0)$. Utilizing this strategy, this result can be refined to:

Theorem 2 *Under Assumptions 1-9,* $\hat{k}_1 - k_{1,0} = O_p(1)$ *as* $(N,T) \to \infty$.

Again, no N-T ratio condition is needed to eliminate the effect of the measurement error z_t. Once $\hat{k}_1$ is available, we can plug it in and estimate $k_{2,0}$. Since $\hat{k}_1 - k_{1,0} = O_p(1)$, this is asymptotically equivalent to plugging in $k_{1,0}$, in which case the problem is reduced to estimating the first change point with observations $t = 1, ..., k_{1,0}$ removed[8]. Thus $\hat{k}_2 - k_{2,0}$ will also be $O_p(1)$. Using this argument sequentially, we have

Theorem 3 *Under Assumptions 1-9,* $\hat{k}_\iota - k_{\iota,0} = O_p(1)$ *for* $\iota = 1, ..., L$ *as* $(N,T) \to \infty$.

3.3 Estimating the Number of Factors and the Factor Space

Once the change points estimators are available, we can plug them in and estimate the number of factors and factor space in each regime. Let $\tilde{r}_\kappa$ be the estimated number of factors in the κ-th regime.

Theorem 4 *Under Assumptions 1-2 and 5, with* $\tilde{k}_\kappa - k_{\kappa,0} = O_p(1)$ *and* $\tilde{k}_{\kappa-1} - k_{\kappa-1,0} = O_p(1)$, *we have* $\lim_{(N,T)\to\infty} P(\tilde{r}_\kappa = r_\kappa) = 1$.

[8] In the general case, $\hat{k}_1$ could converge to the change point in the middle of the sampe. Then the problem is reduced to estimating the first change point for subsamples $t = 1, ..., k_{10}$ and $t = k_{10} + 1, ..., T$ and taking $\hat{k}_2$ as the one leading to the largest reduction in sum of squared residuals.

The proof is similar to the single change case, see Baltagi, Kao and Wang (2015b). Note that what we proved is that the speed of eigenvalue separation is not affected by using the estimated change points. Thus, most eigenvalue based estimators are applicable here. For example, Bai and Ng (2002), Ahn and Horenstein (2013), to name a few.

Next, let u_κ be some positive integer, $\tilde{F}_\kappa^{u_\kappa}$ be $\sqrt{T}$ times the eigenvectors corresponding to the first u_κ eigenvalues of $X_\kappa X_\kappa'$, $H_\kappa^{u_\kappa} = \frac{1}{N}\Lambda_{0\kappa}'\Lambda_{0\kappa}\frac{1}{\tilde{k}_\kappa - \tilde{k}_{\kappa-1}}F_\kappa \tilde{F}_\kappa^{u_\kappa}$ and $\hat{F}_\kappa^{u_\kappa} = \tilde{F}_\kappa^{u_\kappa} V_{NT,\kappa}^{u_\kappa}$, where $X_\kappa = (x_{\tilde{k}_{\kappa-1}+1}, ..., x_{\tilde{k}_\kappa})'$, $F_\kappa = (f_{\tilde{k}_{\kappa-1}+1}, ..., f_{\tilde{k}_\kappa})'$and $V_{NT,\kappa}^{u_\kappa}$ is the diagonal matrix that contains the first u_κ eigenvalues of $X_\kappa X_\kappa'$.

Theorem 5 *Under Assumptions 1-4, with* $\tilde{k}_\kappa - k_{\kappa,0} = O_p(1)$ *and* $\tilde{k}_{\kappa-1} - k_{\kappa-1,0} = O_p(1)$*, we have*[9]

$$\frac{1}{\tilde{k}_\kappa - \tilde{k}_{\kappa-1}} \sum\nolimits_{t=\tilde{k}_{\kappa-1}+1}^{\tilde{k}_\kappa} \left\| \hat{f}_t^{u_\kappa} - H_\kappa^{u_\kappa \prime} f_t \right\|^2 = O_p(\frac{1}{\delta_{NT}^2}). \tag{6}$$

The proof is similar to the single change case, see Baltagi, Kao and Wang (2015b). The convergence rate $O_p(\frac{1}{\delta_{NT}^2})$ is crucial to eliminate the effect of using estimated factors in factor-augmented forecasting and FAVAR. In the next section we will use the estimated factors to construct a test for l versus $l+1$ changes, which determines the number of changes sequentially. We shall show that the rate $O_p(\frac{1}{\delta_{NT}^2})$ is again crucial to eliminating the effect of using estimated factors on the limiting distribution of the test statistic.

Note that Theorem 5 and Theorem 4 are based on an inconsistent estimator of the change points (the distance between the estimated and the true change points is $O_p(1)$), i.e., in the current context, inconsistency of the first step estimator has asymptotically no effect on the second step estimator. This is quite different from the traditional plug-in procedure, in which the first step estimation need to be consistent with sufficiently fast speed to eliminate its effect on the second step estimation. The reason for this difference is that in the current context as $T \to \infty$, the second step estimation becomes less and less sensitive to the first step estimation error[10]. Moreover, we also want to point out that the consistency of the estimated change fraction is not enough for Theorem 4 and Theorem 5. The rate $O_p(1)$ is crucial.

[9] Note that for equation (6) to hold, it does not require $\Lambda_{0\kappa}$ to be full column rank.

[10] Although large T does not help identify the change point, it helps absorb the estimation error of the change points.

4 TESTING MULTIPLE CHANGES

In this section we propose two tests for multiple changes. The first one tests no change versus some fixed number of changes. We show that to eliminate the effect of measurement error z_t, this requires $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{\delta_{NT}} \to 0$, where d_T is the bandwidth used in estimating the covariance matrix of the second moments of the estimated factors. The second one tests l versus $l+1$ changes. We show that for this test, we require $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{T^{\frac{1}{4}}} \to 0$ to eliminate the effect of z_t, and using estimated change points does not affect the limiting distribution under the null. We also discuss how to determine the number of changes using this test and the sequential estimation of the change points.

4.1 Testing No Change versus Some Fixed Number of Changes

In this subsection, we discuss how to test $L = 0$ versus $L = l$, where l is some positive integer. This generalizes existing tests in the literature which only consider $L = 0$ versus $L = 1$, for example, Chen, Dolado and Gonzalo (2014), Han and Inoue (2014) and Cheng, Liao and Schorfheide (2014). In case $l = 1$, it can be shown that our test is asymptotically equivalent to the $\sup LM$ test in Han and Inoue (2014). In what follows, we will discuss the construction of the test statistic first, and then modify the assumptions to study its asymptotic properties.

4.1.1 Construction of the Test for $L = 0$ versus $L = l$

First, estimate the number of factors. $\tilde{r}$ is consistent for r and $\bar{r}$ under the null and the alternative respectively. Thus we can behave as if r and $\bar{r}$ were known in studying the asymptotic properties. Note that for testing purposes, estimating the number of factors also plays the role of selecting the relevant moment conditions from a large number of candidates. Next, estimate the factors by the principal component method. Under the null, let $\tilde{f}_t$ be the estimated factors, U_{NT} be the diagonal matrix that contains the r largest eigenvalues of XX', $H = \frac{1}{N}\Lambda'\Lambda\frac{1}{T}F'\tilde{F}U_{NT}^{-1}$ be the rotation matrix, H_0 be the probability limit of H and $z_t^* = vech(\tilde{f}_t\tilde{f}_t' - H_0'f_tf_t'H_0)$. Under the alternative, we follow the same notation as the last section. It follows that under the null $vech(\tilde{f}_t\tilde{f}_t')$ is a multivariate time series $(vech(H_0'f_tf_t'H_0))$ with stable mean $(vech(I_r))$[11] and measurement error z_t^*, while under the

[11]It is not difficult to see that $E(H_0'f_tf_t'H_0) = H_0'\Sigma_F H_0 = I_r$.

alternative $vech(\tilde{g}_t\tilde{g}_t')$ is a multivariate time series with l mean shifts and measurement error z_t. Thus we can base the test on the difference between the restricted and unrestricted sum of squared normalized error.

Let $\Omega = \lim\limits_{T\to\infty} Var(vech(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(H_0'f_tf_t'H_0 - I_r)))$ be the long run covariance matrix of $vech(H_0'f_tf_t'H_0 - I_r)$ and $\tilde{\Omega}(\tilde{F}) = \tilde{\Upsilon}_0(\tilde{F}) + \sum_{j=1}^{T-1} k(\frac{j}{d_T})[\tilde{\Upsilon}_j(\tilde{F}) + \tilde{\Upsilon}_j(\tilde{F})']$ be the HAC estimator of Ω using the estimated factors $\tilde{F}$, where $\tilde{\Upsilon}_j(\tilde{F}) = \frac{1}{T}\sum_{t=j+1}^{T} vech(\tilde{f}_t\tilde{f}_t' - I_{\tilde{r}})vech(\tilde{f}_{t-j}\tilde{f}_{t-j}' - I_{\tilde{r}})'$, $k(\cdot)$ is some kernel function and d_T is the bandwidth. For simplicity, we will suppress $\tilde{\Omega}(\tilde{F})$ as $\tilde{\Omega}$. It follows that the restricted sum of squared normalized error is

$$SSNE_0 = \sum_{t=1}^{T} vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t')'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t'), \qquad (7)$$

and for any partition $(k_1, ..., k_l)$, the unrestricted sum of squared normalized error is

$$\begin{aligned} SSNE(k_1, ..., k_l) \;=\; & \sum_{\iota=1}^{l+1}\sum_{t=k_{\iota-1}+1}^{k_\iota} vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota} \\ & \tilde{f}_t\tilde{f}_t')'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota}\tilde{f}_t\tilde{f}_t'). \qquad (8) \end{aligned}$$

Let $F_{NT}(\tau_1, ..., \tau_l; \frac{\tilde{r}(\tilde{r}+1)}{2}) = \frac{2}{l\tilde{r}(\tilde{r}+1)}[SSNE_0 - SSNE(k_1, ..., k_l)]$ and $\Lambda_\epsilon = \{(\tau_1, ..., \tau_l) : |\tau_{\iota+1} - \tau_\iota| \geq \epsilon, \tau_1 \geq \epsilon, \tau_l \leq 1-\epsilon\}$ for some prespecified $\epsilon > 0$, the test statistic is $\sup\limits_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} F_{NT}(\tau_1, ..., \tau_l; \frac{\tilde{r}(\tilde{r}+1)}{2})$.

4.1.2 Asymptotic Properties of the Test for $L = 0$ versus $L = l$

We first consider the limiting distribution of the proposed test under the null. Since under the null the factor loadings are stable, we use λ_i and Λ to denote the factor loading and the factor loading matrix respectively. The assumptions in the last section are modified as follows:

Assumption 10 *$\mathbb{E}\|f_t\|^4 < M < \infty$, $\mathbb{E}(f_tf_t') = \Sigma_F$ for all t, Σ_F is positive definite and $\frac{1}{T}\sum_{t=1}^{T} f_tf_t' - \Sigma_F = o_p(1)$.*

Assumption 11 *$\|\lambda_i\| \leq \bar{\lambda} < \infty$ and $\left\|\frac{1}{N}\Lambda'\Lambda - \Sigma_\Lambda\right\| = O(\frac{1}{\sqrt{N}})$ for some positive definite matrix Σ_Λ.*

Assumption 12 *The eigenvalues of $\Sigma_F\Sigma_\Lambda$ are distinct.*

Assumption 13 *There exists $M < \infty$ such that:*

1. $\mathbb{E}(\left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{i=1}^{N} f_s[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right\|^2) \leq M$ *for all* t,

2. $\mathbb{E}(\left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} f_t \lambda_i' e_{it} \right\|^2) \leq M$,

3. $\mathbb{E}(\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} \right\|^2) \leq M$ *for all* t.

Assumption 14 *For any* $\epsilon > 0$,

1. $\sup_{T\epsilon \leq k \leq T(1-\epsilon)} \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{k} \sum_{i=1}^{N} f_t \lambda_i' e_{it} \right\| = O_p(1)$,

2. $\sup_{T\epsilon \leq k \leq T(1-\epsilon)} \left\| \frac{1}{\sqrt{NT}} \sum_{t=k+1}^{T} \sum_{i=1}^{N} f_t \lambda_i' e_{it} \right\| = O_p(1)$.

Assumption 15 Ω *is positive definite and* $\frac{1}{\sqrt{T}} \sum_{t=1}^{T\tau} vech[\Omega^{-\frac{1}{2}}(H_0' f_t f_t' H_0 - I_r)] \Rightarrow W_{\frac{r(r+1)}{2}}(\tau)$, *where* $W_{\frac{r(r+1)}{2}}(\cdot)$ *is an* $\frac{r(r+1)}{2}$ *dimensional vector of independent Wiener processes on* $[0,1]$.

Assumption 16 $\tilde{\Omega}(FH_0)$ *is consistent for* Ω.

Assumption 10 only requires the law of large number to be applicable to the factors for the whole sample, and thus it weakens Assumption 1. Assumption 11 specifies the convergence rate of $\frac{1}{N}\Lambda'\Lambda$, and thus strengthens Assumption 2. Assumptions 3-5 are maintained. Assumption 12 ensures the uniqueness of the principal component estimator under the null. Assumption 13 corresponds to and slightly weakens Assumption F in Bai (2003). Assumption 14 requires the term in $\|\cdot\|$ to be uniformly $O_p(1)$. This is not restrictive since all summands have zero means. Assumptions 15 requires the functional central limit theorem to be applicable to $vech(H_0' f_t f_t' H_0 - I_r)$. Assumptions 16 requires the HAC estimator of Ω to be consistent if factors were observable.

Now we are ready to present the limiting distribution:

Theorem 6 *Under Assumptions 3-5, 10-16 and* $L = 0$, *with* $\frac{\sqrt{T}}{N} \to 0$ *and* $\frac{d_T}{\delta_{NT}} \to 0$ *as* $(N,T) \to \infty$,

$$\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} F_{NT}(\tau_1,...,\tau_l; \frac{\tilde{r}(\tilde{r}+1)}{2}) \xrightarrow{d} \sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} F(\tau_1,...,\tau_l; \frac{r(r+1)}{2}),$$

where $F(\tau_1,...,\tau_l; \frac{r(r+1)}{2}) = \frac{2}{lr(r+1)} \sum_{\iota=1}^{l} \frac{\left\| \tau_\iota W_{\frac{r(r+1)}{2}}(\tau_{\iota+1}) - \tau_{\iota+1} W_{\frac{r(r+1)}{2}}(\tau_\iota) \right\|^2}{\tau_\iota \tau_{\iota+1}(\tau_{\iota+1}-\tau_\iota)}$.

Critical values are tabulated in Bai and Perron (1998, 2003). Here the degree of freedom is related to the number of factors. In practical applications the degree of freedom would be $\frac{\tilde{r}(\tilde{r}+1)}{2}$, thus underestimation of the number of factors will not affect the size of the test[12]. Note that $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{\delta_{NT}} \to 0$ are needed to eliminate the effect of the measurement error z_t^*. This is different from the results in the last section but similar to the results in the factor-augmented forecasting and FAVAR. Intuitively, testing for structural changes relies on all the observations and consequently measurement error will accumulate in the test statistic as $T \to \infty$ and $d_T \to \infty$.

We next consider the consistency of the proposed test. Under the alternative, the process $vech(\tilde{g}_t\tilde{g}_t')$ has l mean shifts and measurement error z_t. Thus $vech(\tilde{g}_t\tilde{g}_t')$ is not properly demeaned in calculating the restricted $SSNE$. On the other hand, the test statistic can be written as $\frac{2}{l\tilde{r}(\tilde{r}+1)}[SSNE_0 - \min_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} SSNE(k_1,...,k_l)]$ and by taking the minimum for $(\tau_1,...,\tau_l) \in \Lambda_\epsilon$, it ensures $vech(\tilde{g}_t\tilde{g}_t')$ is properly demeaned. Thus under the alternative, the test statistic will diverge as $(N,T) \to \infty$.

Theorem 7 *Under Assumptions 1-8 and $L = l$, with $\frac{d_T}{T} \to 0$ as $(N,T) \to \infty$,*

$$\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} F_{NT}(\tau_1,...,\tau_l; \frac{\tilde{r}(\tilde{r}+1)}{2}) \xrightarrow{p} \infty.$$

The test discussed above is designed for a given number of changes under the alternative. When the number of changes is misspecified, the test may not be powerful. For example, test for 0 versus 2 changes should be more powerful than the test for 0 versus 1 change when the true DGP contains two changes. Following Bai and Perron (1998), we consider the UDmax and WDmax tests when the number of changes under the alternative is unknown. Given the maximum possible number of changes M and significance level α, the UDmax is simply the maximum of the tests for 0 versus l changes with $l \leq M$ while WDmax is the weighted maximum of the tests for 0 versus l changes with weights $c(\frac{\tilde{r}(\tilde{r}+1)}{2}, \alpha, 1)/c(\frac{\tilde{r}(\tilde{r}+1)}{2}, \alpha, l)$. With Theorem 6, the limiting distributions of both tests have the same form as in Bai and Perron (1998). Comprehensive critical values are tabulated in Bai and Perron (2003).

[12] Of course, it will decrease the power.

4.2 Testing l versus $l+1$ Changes

In this subsection, we discuss how to test $L = l$ versus $L = l+1$ for any prespecified positive integer l. The idea is to estimate l change points first and once they are plugged in, testing $L = l$ versus $L = l+1$ is equivalent to testing no change versus a single change in each regime jointly. The main concern is the effect of using estimated change points and estimated factors on the limiting distribution and consistency of the test statistic. In what follows, we discuss the construction of the test statistic and its asymptotic properties. Also, how to use it to determine the number of changes.

4.2.1 Construction of the Test for $L = l$ versus $L = l+1$

First, we estimate l change points, either jointly or sequentially. Let $\tilde{k}_1, ..., \tilde{k}_l$ be the estimated change points. Then plug $\tilde{k}_1, ..., \tilde{k}_l$ in and estimate the number of factors and factor space in each regime. Let $\tilde{r}_\iota$ be the estimated number of factors in the ι-th regime. Under the null, let $\tilde{F}_\iota = (\tilde{f}_{\iota,\tilde{k}_{\iota-1}+1}, ..., \tilde{f}_{\iota,\tilde{k}_\iota})'$ be the estimated factors, H_ι be the rotation matrix, $H_{\iota 0}$ be the limit of H_ι, $U_{\iota NT}$ be the eigenvalue matrix, U_ι be the limit of $U_{\iota NT}$, $F_\iota = (f_{\iota,\tilde{k}_{\iota-1}+1}, ..., f_{\iota,\tilde{k}_\iota})'$ and $F_{\iota 0} = (f_{\iota,k_{\iota-1,0}+1}, ..., f_{\iota,k_{\iota 0}})'$. Note that $f_{\iota,t}$ is r_ι dimensional and contains the factors that appear in the ι-th regime. Under the alternative, there are $l+1$ changes and the l estimated change points will be close to ($O_p(1)$) the l points that allow the greatest reduction in the sum of squared normalized errors. Without loss of generality, suppose $\tilde{k}_{\iota-1} - k_{\iota-1,0} = O_p(1)$ and $\tilde{k}_\iota - k_{\iota+1,0} = O_p(1)$ for some ι. In this case, the ι-th regime contains an extra change point[13] $k_{\iota,0}$ but can be equivalently represented as having no changes but with pseudo factors $g_{\iota t}$, where $g_{\iota t} = A_{\iota 1} f_t$ for $t \in [\tilde{k}_{\iota-1} + 1, ..., k_{\iota 0}]$ and $g_{\iota t} = A_{\iota 2} f_t$ for $t \in [k_{\iota 0} + 1, ..., \tilde{k}_\iota]$. For this regime, we denote the estimated factors as $\tilde{g}_{\iota t}$ and define $\tilde{G}_\iota$, G_ι, $G_{\iota 0}$, J_ι, $J_{\iota 0}$, $V_{\iota NT}$ and V_ι correspondingly as $\tilde{F}_\iota$, F_ι, $F_{\iota 0}$, H_ι, $H_{\iota 0}$, $U_{\iota NT}$ and U_ι. For the other regimes, we maintain the same notation. It follows that under the null $vech(\tilde{f}_{\iota t}\tilde{f}_{\iota t}')$ is a multivariate time series with stable mean and measurement error $z_{\iota t}^*$ for all ι while under the alternative $vech(\tilde{g}_{\iota t}\tilde{g}_{\iota t}')$ is a multivariate time series with a mean shift and measurement error $z_{\iota t}$ for some ι. Again, the test is based on the difference between the restricted and unrestricted sum of squared normalized error.

[13] When $\tilde{k}_{\iota-1} < k_{\iota-1,0}$ or $\tilde{k}_\iota > k_{\iota+1,0}$, the ι-th regime also contains the change point $k_{\iota-1,0}$ or $k_{\iota+1,0}$, but with $\tilde{k}_{\iota-1} - k_{\iota-1,0} = O_p(1)$ and $\tilde{k}_\iota - k_{\iota+1,0} = O_p(1)$ these two are asymptotically ignorable.

Let $\Omega_\iota = \lim_{T\to\infty} Var(vech(\frac{1}{\sqrt{k_{\iota,0}-k_{\iota-1,0}}} \sum_{t=k_{\iota-1,0}+1}^{k_{\iota,0}} (H'_{\iota 0} f_{\iota t} f'_{\iota t} H_{\iota 0} - I_{r_\iota})))$ be the long run covariance matrix of $vech(H'_{\iota 0} f_{\iota t} f'_{\iota t} H_{\iota 0} - I_{r_\iota})$ and $\tilde{\Omega}_\iota$ be the HAC estimator of Ω_ι using $\tilde{F}_\iota$ and with kernel function $k(\cdot)$ and bandwidth d_T. The test statistic is

$$F_{NT}(l+1\,|l) = SSNE(\tilde{k}_1, ..., \tilde{k}_l) - \min_{1\le\iota\le l+1} \inf_{k\in\Lambda_{\iota,\eta}} SSNE(\tilde{k}_1, ..., \tilde{k}_{\iota-1}, k, \tilde{k}_\iota, ..., \tilde{k}_l), \tag{9}$$

where $SSNE(\tilde{k}_1, ..., \tilde{k}_l)$ is the restricted sum of squared normalized error and equals

$$\begin{aligned}\sum_{\iota=1}^{l+1} SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) &= \sum_{\iota=1}^{l+1} \sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota} vech(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - \frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}} \sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota} \\ &\quad \tilde{f}_{\iota t}\tilde{f}'_{\iota t})'\tilde{\Omega}_\iota^{-1} vech(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - \frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}} \sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota} \tilde{f}_{\iota t}\tilde{f}'_{\iota t}),\end{aligned} \tag{10}$$

$SSNE(\tilde{k}_1, ..., \tilde{k}_{\iota-1}, k, \tilde{k}_\iota, ..., \tilde{k}_l)$ is the unrestricted sum of squared normalized error and equals

$$\sum_{\kappa=1}^{\iota-1} SSNE_\kappa(\tilde{k}_{\kappa-1}, \tilde{k}_\kappa) + SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota) + \sum_{\kappa=\iota+1}^{l+1} SSNE_\kappa(\tilde{k}_{\kappa-1}, \tilde{k}_\kappa), \tag{11}$$

and $\Lambda_{\iota,\eta} = \{k : \tilde{k}_{\iota-1} + (\tilde{k}_\iota - \tilde{k}_{\iota-1})\eta \le k \le \tilde{k}_\iota - (\tilde{k}_\iota - \tilde{k}_{\iota-1})\eta\}$.

4.2.2 Asymptotic Properties of the Test for $L = l$ versus $L = l + 1$

We first consider the limiting distribution. If the true change points were plugged in, Theorem 6 implies that for each regime the effect of using estimated factors can be eliminated if $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{\delta_{NT}} \to 0$. When the estimated change points are plugged in, we will show based on Theorem 4 and Theorem 5 that the result still holds if $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{T^{\frac{1}{4}}} \to 0$.

Since under the null, there are $l+1$ stable regimes, we modify the assumptions in Theorem 6 so that they are satisfied in each regime. More specifically, Assumption 10 is replaced by Assumption 1 while Assumptions 11-16 are modified as follows:

Assumption 17 *$\|\lambda_{0\kappa,i}\| \le \bar{\lambda} < \infty$ for $\kappa = 1, ..., l+1$, and $\left\|\frac{1}{N}\Gamma'\Gamma - \Sigma_\Gamma\right\| = O(\frac{1}{\sqrt{N}})$ for some positive definite matrix Σ_Γ.*

Assumption 18 *Let $\Sigma_{F,\iota}$ be the probability limit of $\frac{1}{k_{\iota,0}-k_{\iota-1,0}} \sum_{t=k_{\iota-1,0}+1}^{k_{\iota,0}} f_{\iota t} f'_{\iota t}$, Λ^ι contain the nonzero columns of Λ_0 and Λ_ι and Σ_{Λ^ι} be the limit of $\frac{1}{N}\Lambda^\iota\Lambda^{\iota\prime}$. The eigenvalues of $\Sigma_{F,\iota}\Sigma_{\Lambda,\iota}$ are distinct for all ι.*

Assumption 19 *There exists $M < \infty$ such that:*

1. $\mathbb{E}(\left\| \frac{1}{\sqrt{NT}} \sum_{s=k_{\iota-1,0}+1}^{k_{\iota,0}} \sum_{i=1}^{N} f_s[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right\|^2) \leq M$ *for all* t *and all* ι,

2. $\mathbb{E}(\left\| \frac{1}{\sqrt{NT}} \sum_{t=k_{\iota-1,0}+1}^{k_{\iota,0}} \sum_{i=1}^{N} f_t \lambda'_{0\iota,i} e_{it} \right\|^2) \leq M$ *for all* ι,

3. $\mathbb{E}(\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{0\iota,i} e_{it} \right\|^2) \leq M$ *for all* $k_{\iota-1,0} < t \leq k_{\iota,0}$ *and all* ι.

Assumption 20 *For any* $\epsilon > 0$ *and all* ι,

1. $\sup_{(k_{\iota,0}-k_{\iota-1,0})\epsilon \leq k-k_{\iota-1,0} \leq (k_{\iota,0}-k_{\iota-1,0})(1-\epsilon)} \left\| \frac{1}{\sqrt{NT}} \sum_{t=k_{\iota-1,0}+1}^{k} \sum_{i=1}^{N} f_t \lambda'_{0\iota,i} e_{it} \right\| = O_p(1)$,

2. $\sup_{(k_{\iota,0}-k_{\iota-1,0})\epsilon \leq k-k_{\iota-1,0} \leq (k_{\iota,0}-k_{\iota-1,0})(1-\epsilon)} \left\| \frac{1}{\sqrt{NT}} \sum_{t=k+1}^{k_{\iota,0}} \sum_{i=1}^{N} f_t \lambda'_{0\iota,i} e_{it} \right\| = O_p(1)$.

Assumption 21 Ω_ι *is positive definite and*

$$\frac{1}{\sqrt{k_{\iota,0} - k_{\iota-1,0}}} \sum_{t=k_{\iota-1,0}+1}^{k_{\iota-1,0}+(k_{\iota,0}-k_{\iota-1,0})\tau} vech[\Omega_\iota^{-\frac{1}{2}}(H'_{\iota 0} f_{\iota t} f'_{\iota t} H_{\iota 0} - I_{r_\iota})] \Rightarrow W_{\frac{r_\iota(r_\iota+1)}{2}}(\tau).$$

Assumption 22 *Let* $\tilde{\Omega}(F_\iota H_{\iota 0})$ *be the HAC estimator of* Ω_ι *using* $F_\iota H_{\iota 0}$, $\tilde{\Omega}(F_\iota H_{\iota 0})$ *is consistent for* Ω_ι.

Now we are ready to present the result:

Theorem 8 *Under Assumptions 1, 3-5, 17-22 and* $L = l$, *with* $\tilde{k}_\iota - k_{\iota,0} = O_p(1)$ *for all* ι, $\frac{\sqrt{T}}{N} \to 0$ *and* $\frac{d_T}{T^{\frac{1}{4}}} \to 0$, *we have* $F_{NT}(l+1\,|\,l) \xrightarrow{d} \sup_{1 \leq \iota \leq l+1} F_\iota$, *where*

$$F_\iota = \sup_{\eta \leq \tau \leq (1-\eta)} \frac{1}{\tau(1-\tau)} \left\| W_{\frac{r_\iota(r_\iota+1)}{2}}(\tau) - \tau W_{\frac{r_\iota(r_\iota+1)}{2}}(1) \right\|^2$$

and F_ι *is independent with each other for different* ι.

Critical values can be obtained via simulations and here they are related to the number of factors in each regime. In case the number of factors is stable, we have:

Corollary 1 *If* $r_\iota = r$ *for all* ι, $\lim_{(N,T)\to\infty} P(F_{NT}(l+1\,|\,l) \leq x) = G_{\frac{r(r+1)}{2},\eta}(x)^{l+1}$, *where* $G_{\frac{r(r+1)}{2},\eta}(x)$ *is the c.d.f. of* $\sup_{\eta \leq \tau \leq (1-\eta)} \frac{1}{\tau(1-\tau)} \left\| W_{\frac{r(r+1)}{2}}(\tau) - \tau W_{\frac{r(r+1)}{2}}(1) \right\|^2$.

Critical values for this case are tabulated in Bai and Perron (1998, 2003). We next show that $F_{NT}(l+1\,|l)$ is also consistent. Let $\Sigma_{G,\iota}$ be the probability limit of $\frac{1}{k_{\iota,0}-k_{\iota-1,0}} g_{\iota t} g_{\iota t}'$, Γ^{ι} contain the nonzero columns of Λ_0, Λ_ι and $\Lambda_{\iota+1}$, and Σ_{Γ^ι} be the limit of $\frac{1}{N}\Gamma^\iota\Gamma^{\iota\prime}$. Assumption 18 is replaced by:

Assumption 23 *The eigenvalues of* $\Sigma_{G,\iota}\Sigma_{\Gamma^\iota}$ *are distinct.*

Since $F_{NT}(l+1\,|l) = \sup_{1\le\kappa\le l+1} \sup_{k\in\Lambda_{\kappa,\eta}} [SSNE_\kappa(\tilde{k}_{\kappa-1},\tilde{k}_\kappa) - SSNE_\kappa(\tilde{k}_{\kappa-1},k,\tilde{k}_\kappa)] \ge SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1},k_{\iota 0},\tilde{k}_\iota)$ and under the alternative $SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota)$ is not properly demeaned, $F_{NT}(l+1\,|l)$ will diverge as $(N,T)\to\infty$.

Theorem 9 *Under Assumptions 1-5, 19-20, 23 and* $L = l+1$, *with* $\left|\tilde{k}_\iota - k_{\iota+1,0}\right| = O_p(1)$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$ *for some* ι *and* $\frac{d_T}{T}\to 0$, *we have* $F_{NT}(l+1\,|l) \xrightarrow{p} \infty$.

4.2.3 Determining the Number of Changes

The sequential test $F_{NT}(l+1\,|l)$ allows us to determine the number of changes. First, estimate l change points, either jointly or sequentially, where l could be suggested by some prior information or just zero. Next, perform the test $F_{NT}(l+1\,|l)$. If rejected[14], estimate $l+1$ change points, either jointly or sequentially, and then perform the test $F_{NT}(l+2\,|l+1)$. Repeat this procedure until the null can not be rejected. Let $\hat{L}$ be the estimated number of changes, it is not difficult to see that $\lim_{(N,T)\to\infty} P(\hat{L}<L) = 0$ and $\lim_{(N,T)\to\infty} P(\hat{L}=L+1) = \alpha$. let $\alpha\to 0$ as $(N,T)\to\infty$, then $\hat{L}$ will be consistent.

5 REESTIMATING THE CHANGE POINTS

Once the estimated number of factors in each regime are available, we can reestimate the change points by minimizing the sum of least squares residuals. More specifically, for any possible change points $(k_1,...,k_L)$, we split the sample into $L+1$ regimes and minimize the sum of squared residuals in each regime. That is to say, choose $f_t^{\tilde{r}_\kappa}$ and $\lambda_{\kappa i}^{\tilde{r}_\kappa}$ to minimize $\sum_{t=k_{\kappa-1}+1}^{k_\kappa}\sum_{i=1}^{N}(x_{it} - f_t^{\tilde{r}_\kappa\prime}\lambda_{\kappa i}^{\tilde{r}_\kappa})^2$. Note that here we use superscript $\tilde{r}_\kappa$ to emphasize that the

[14] It can be shown that the test is also consistent when $L > l+1$.

dimension of the factors and loadings are determined by $\tilde{r}_\kappa$. Denote the estimated factors and loadings in the κ-th regime as $\hat{f}_t^{\tilde{r}_\kappa}$ and $\hat{\lambda}_{\kappa i}^{\tilde{r}_\kappa}$, the total sum of squared residuals is

$$S(k_1, ..., k_L) = \sum_{\kappa=1}^{L+1} \sum_{t=k_{\kappa-1}+1}^{k_\kappa} \sum_{i=1}^{N} (x_{it} - \hat{f}_t^{\tilde{r}_\kappa \prime} \hat{\lambda}_{\kappa i}^{\tilde{r}_\kappa})^2 \tag{12}$$

and the reestimated change points are

$$(\hat{k}_1, ..., \hat{k}_L) = \arg\min S(k_1, ..., k_L). \tag{13}$$

According to Theorem 4, $\tilde{r}_\kappa$ is consistent for r_κ, thus asymptotically we can treat r_κ as known. In case r_κ, $\kappa = 1, ..., L+1$ are all the same, we can prove that

$$(\hat{k}_1, ..., \hat{k}_L) \xrightarrow{p} (k_{1,0}, ..., k_{L,0}) \text{ as } (N, T) \to \infty. \tag{14}$$

When reestimated change points are available, we can recalculate the number of factors and factor space in each regime and also the test of l versus $l+1$ changes. Given consistency of $(\hat{k}_1, ..., \hat{k}_L)$, the theoretical properties of these recalculated estimators and tests remain the same. Their finite sample performance should be better.

Assuming the number of factors in each regime is known and the same, Massacci (2015) uses least squares to estimate the threshold in a high dimensional factor model with a single threshold and proves the estimated threshold is consistent with convergence rate $O_p(1/NT)$. Since structural change is a special case of the threshold model, in the single change case consistency of the estimated change point follows directly from Massacci (2015). Here with multiple changes, our proof is conceptually similar to Massacci (2015) but more difficult. This is because in the single change case the lower boundary of the first regime and the upper boundary of the second regime are known while in multiple changes case, for those regimes in the middle of the sample both boundaries are unknown. Detailed proof of (14) is available upon request from the authors.

Note that consistency of $\tilde{r}_\kappa$ for r_κ is the key for establishing (14). Consider the case with one factor and one structural change at k_0. If we choose $\tilde{r}_1 = \tilde{r}_2 = 1$, the estimated change point $\hat{k}$ in (13) would be consistent. However, if we choose $\tilde{r}_1 = \tilde{r}_2 = 2$, $\hat{k}$ provides no information for k_0. To see this, suppose $k < k_0$ and $(k_0 - k)/T > \eta > 0$. The first subsample does not contain structural change. The second subsample does contain an unaccounted structural change but $\tilde{r}_2 = 2$ still allows the estimated factors to fully capture the true factor space because one factor with one structural change can be represented as two factors with no structural change. Thus with $\tilde{r}_1 = \tilde{r}_2 = 2$, the total sum of squared residuals $S(k)$ would still be close to $S(k_0)$ even if k is far away from k_0.

In general, for the single change case, it is not difficult to establish that the condition to ensure consistency of $\hat{k}$ is

$$\begin{aligned} r_1 &\leq \tilde{r}_1 < \bar{r}, \\ r_2 &\leq \tilde{r}_2 < \bar{r}. \end{aligned} \tag{15}$$

More specifically, $r_1 \leq \tilde{r}_1$ and $\tilde{r}_2 < \bar{r}$[15] ensures $P(\hat{k} < k_0) \to 0$ as $(N, T) \to \infty$ while $\tilde{r}_1 < \bar{r}$ and $r_2 \leq \tilde{r}_2$ ensures $P(\hat{k} > k_0) \to 0$ as $(N, T) \to \infty$. Thus to ensure consistency of $\hat{k}$, $\tilde{r}_1$ and $\tilde{r}_2$ do not have to be the same and are not necessarily unique. Massacci (2015) proposes choosing $\tilde{r}_1 = \tilde{r}_2 = r_{\max}$ in estimating the threshold, but according to (15) this does not necessarily work if $r_{\max} \geq \bar{r}$ or $r_{\max} < r_1$ or $r_{\max} < r_2$. Chen (2015) proposes choosing $\tilde{r}_1 = \tilde{r}_2 = \tilde{r} - 1$ in estimating a single change point. Since $\tilde{r}$ is consistent for $\bar{r}$, this is equivalent to choosing $\bar{r} - 1$. For the single change case with $r_1 = r_2$, Chen (2015)'s method should work because in this case $\bar{r}$ must be larger than r_1 and r_2[16] and thus $\bar{r} - 1$ satisfies condition (15). Strict proof of Chen (2015)'s method is currently unavailable. The main issue of Chen (2015)'s method is that it does not necessarily work in the multiple changes case. Consider the case with one factor and two changes at $k_{1,0}$ and $k_{2,0}$. In this case $\bar{r} = 3$ and thus choosing $\tilde{r} - 1$ is equivalent to choosing $\tilde{r}_1 = \tilde{r}_2 = \tilde{r}_3 = 2$ in calculating $S(k_1, k_2)$ in equation (12). Suppose k_2 is fixed at its true value $k_{2,0}$ and let us compare $S(k_1, k_{2,0})$ to $S(k_{1,0}, k_{2,0})$ with $k_1 < k_{1,0}$ and $(k_{1,0} - k_1)/T > \eta > 0$. First, the sums of squared residuals from the third regime are the same. Thus it reduced to the case with single change point $k_{1,0}$ with sample $t = 1, ..., k_{2,0}$. As discussed above, the first subsample $(t = 1, ..., k_1)$ does not contain structural change. The second subsample $(t = k_1 + 1, ..., k_{2,0})$ does contain an unaccounted structural change but $\tilde{r}_2 = 2$ still allows the estimated factors to fully capture the true factor space in the second regime. Thus $S(k_1, k_{2,0})$ will not be significantly larger than $S(k_{1,0}, k_{2,0})$ even if k_1 is far away from $k_{1,0}$. For our method, since $\tilde{r}_\kappa$ is consistent for r_κ, consistency of the estimated change points is guaranteed in both the single change and multiple changes case. Also, in the single change case, our method is more efficient than

[15] Recall that $\bar{r}$ is the number of pseudo factors for the equivalent model with no structural change

[16] Recall that in this paper we only consider the case where the matrix that contains all different nonzero vectors of factor loadings is full rank.

Chen (2015) because if $\bar{r} - 1 > r_1$ and r_2, the number of factors is overestimated in each regime and it introduces extra error in (12).

Finally, we want to point out that consistency of the reestimated change points is not guaranteed if r_κ are allowed to be different for different κ. That is to say, Massacci (2015)'s method does not necessarily work if r_κ are allowed to be different. Consider the important case where after a single change point one new factor emerges while the loadings of the existing factors does not change. In this case, condition (15) is still required to ensure consistency of $\hat{k}$. But condition (15) can never be satisfied because $r_1 < r_2 = \bar{r}$. The reestimated change point $\hat{k}$ is at best one-sided consistent ($P(\hat{k} > k_0) \to 0$ as $(N, T) \to \infty$). Therefore, in cases where r_κ are different, the first step change points estimator $(\tilde{k}_1, ..., \tilde{k}_L)$ are more robust.

6 MONTE CARLO SIMULATIONS

This section presents simulation results to evaluate the finite sample properties of our proposed estimation and testing procedures. The number of simulations is 1000.

6.1 Data Generating Process

The factors are generated by

$$f_{t,p} = \rho f_{t-1,p} + u_{t,p} \text{ for } t = 2, ..., T \text{ and } p = 1, ..., 3,$$

where $u_t = (u_{t,1}, u_{t,2}, u_{t,3})'$ is i.i.d. $N(0, I_3)$ for $t = 2, ..., T$ and $f_1 = (f_{1,1}, f_{1,2}, f_{1,3})'$ is i.i.d. $N(0, \frac{1}{1-\rho^2} I_3)$ so that the factors are stationary. The idiosyncratic errors are generated by:

$$e_{i,t} = \alpha e_{i,t-1} + v_{i,t} \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T,$$

where $v_t = (v_{1,t}, ..., v_{N,t})'$ is i.i.d. $N(0, \Omega)$ for $t = 2, ..., T$ and $e_1 = (e_{1,1}, ..., e_{N,1})'$ is $N(0, \frac{1}{1-\alpha^2}\Omega)$ so that the idiosyncratic errors are stationary. Ω is generated as $\Omega_{ij} = \beta^{|i-j|}$ so that β captures the degree of cross-sectional dependence of the idiosyncratic errors. In addition, u_t and v_t are mutually independent for all t.

For factor loadings, we consider two different setups. Setup 1 contains no structural change and λ_i is i.i.d. $N(0, \frac{1}{3} I_3)$ across i. Setup 1 will be used to evaluate the size of the tests for multiple changes. Setup 2 contains two structural changes and hence three regimes.

In the first and the second regime, the last element of $\lambda_{1,i}$ and $\lambda_{2,i}$ are zeros for all i while the first two elements of $\lambda_{1,i}$ and $\lambda_{2,i}$ are both i.i.d. $N(0, \frac{1}{2}I_2)$ across i. In the third regime, $\lambda_{3,i}$ is i.i.d. $N(0, \frac{1}{3}I_3)$ across i. Also, $\lambda_{1,i}$, $\lambda_{2,i}$ and $\lambda_{3,i}$ are independent. Thus in Setup 2 the number of factors in the three regimes are 2, 2, 3 respectively and the number of pseudo factors is 7. Setup 2 will be used to evaluate the performance of the estimated change points and the estimated number of factors in each regime. Setup 3 also contains two structural changes while $\lambda_{1,i}$, $\lambda_{2,i}$ and $\lambda_{3,i}$ are all i.i.d. $N(0, \frac{1}{3}I_3)$ across i and independent of each other. Setup 3 will be used to evaluate the power of the tests for multiple changes and the probabilities of selecting the correct number of changes. Once factors, loadings and errors are available, the data is generated as:

$$\text{Setup 1:} \qquad x_{it} = f_t'\lambda_i + e_{it},$$

$$\text{Setup 2 and 3:} \qquad x_{it} = f_t'\lambda_{\kappa,i} + e_{it}, \text{if } [T\tau_{\kappa-1,0}] + 1 \leq t \leq [T\tau_{\kappa,0}] \text{ for } \kappa = 1, 2, 3,$$

where $(\tau_{1,0}, \tau_{2,0}) = (0.3, 0.7)$ are the change fractions. Finally, all factor loadings are independent of the factors and the idiosyncratic errors.

6.2 Estimating the Change Points

We first estimate the number of pseudo factors using IC_{p1} in Bai and Ng (2002) with the maximum number of factors $rmax = 12$. When using other criterion, e.g., IC_{p2}, IC_{p3} in Bai and Ng (2002) and ER, GR in Ahn and Horenstein (2013), the results are similar, and hence omitted. Once estimated pseudo factors are available, the change points are estimated as in equation (5) with minimum sample size of each regime $T \times 0.1$.

Figures 1 and 2 are the histograms of the jointly estimated change points for $(N, T) = (100, 100)$ and $(N, T) = (100, 200)$ respectively. Each figure includes four subfigures corresponding to $(\rho, \alpha, \beta) = (0, 0, 0)$, $(0.7, 0, 0)$, $(0, 0.3, 0)$ and $(0, 0, 0.3)$ respectively. In all subfigures, more than 95 percent of the mass is concentrated within a (-8,8) neighborhood of the true change points. This confirms our theoretical result that $\tilde{k}_\kappa - k_{\kappa,0} = O_p(1)$. Figures 1 and 2 also show that the performance of the estimated change points deteriorates when ρ increases from 0 to 0.7 while serial correlation and cross-sectional dependence of the errors seems to have no effect. This is also in line with the theoretical predictions because the errors only affect estimation of the pseudo factors and does not affect the estimation of change points directly.

6.3 Estimating the Number of Factors in Each Regime

The number of factors in each regime is estimated using IC_{p2} in Bai and Ng (2002) and ER and GR in Ahn and Horenstein (2013), with maximum number of factors 8. We consider various (N,T) combinations and representative (ρ,α,β) combinations. These should be able to cover most empirically relevant cases. The results are shown in Table 1. x/y denotes the frequency of underestimation and overestimation is $x\%$ and $y\%$ respectively. In all cases, the probability of underestimation plus overestimation, $x+y$ is significantly smaller than the probability that the estimated change points differ from the true change points. This implies $O_p(1)$ deviation from the true change points does not significantly affect $\tilde{r}_1$, $\tilde{r}_2$ and $\tilde{r}_3$. Also, when the size of each subsample is large enough, x and y are both zeros, thus the performance of $\tilde{r}_1$, $\tilde{r}_2$ and $\tilde{r}_3$ is as good as the case where change points are known. This further confirms our theoretical result that $\tilde{r}_1$, $\tilde{r}_2$ and $\tilde{r}_3$ are robust to $O_p(1)$ estimation error of the change points.

6.4 Testing Multiple Changes

Now we present the results for the various tests of multiple changes. Table 2 reports size of the test for 0 versus l changes with $l=1,2,3$, size of the UDmax and WDmax tests and the probabilities of selecting changes when the data is generated under Setup 1. We consider two methods of estimating the number of changes. $\hat{L}_1$ and $\hat{L}_2$. $\hat{L}_1$ is obtained by the sequential procedure as discussed in Section 4.2.3 while $\hat{L}_2$ is obtained by using WDmax to test the presence of at least one change first and then performing the sequential procedure starting from 1 versus 2 changes. Table 3 reports the power of the test for 0 versus l changes with $l=1,2,3$, the power of the UDmax and WDmax tests, the power of the test for 1 versus 2 changes, the size of the test for 2 versus 3 changes and the probabilities of selecting changes when the data is generated under Setup 3. For both tables, we consider $(N,T)=(100,100)$ and $(100,200)$ with $\epsilon=0.05, 0.10, 0.15, 0.20$ and 0.25, and $(\rho,\alpha,\beta)=(0,0,0)$, $(0.7,0,0)$ and $(0.7,0.3,0.3)$. We delete the case $T=100$ and $\epsilon=0.05$ to ensure the sample size of each regime is at least 10.

Note that in calculating the HAC estimator of the covariance matrix of the second moments of the estimated factors, Bartlett's kernel is used with bandwidth $T^{1/3}$ for testing 0 versus l changes and $2\times T^{1/5}$ for testing l versus $l+1$ changes. In estimating the number of factors at the very beginning, IC_{p3} [17] is used except for the case $(N, T) = (100, 100)$ and $(\rho, \alpha, \beta) = (0.7, 0.3, 0.3)$. In that case, IC_{p3} overestimates too much, thus we switch to IC_{p1}. The critical values are obtained from Bai and Perron (2003) with nominal size of 5%.

First consider the size properties. Table 2 shows that overall, all tests are slightly undersized. The undersizing phenomenon is quite obvious when $T = 100$ and $\rho = 0$. This is in line with previous findings, see Diebold and Chen (1996). When T increases to 200, the empirical size gets closer to the nominal size 5%. It is also easy to see that when $\rho = 0.7$ and $\epsilon = 0.05$, the tests are significantly oversized. Thus we recommend choosing ϵ at least 0.10 when the factors have serial correlation. Serial and cross-sectional dependence of the errors do not affect the performance too much. Once T is large enough to guarantee the accuracy of the estimated factors, serial and cross-sectional dependence of the errors do not seem to affect the size of the various tests.

Now consider the power properties. Powers of the tests for 0 versus l changes are good in all cases. WDmax has good power except when $T = 100$ and $\epsilon = 0.25$, and is more powerful than UDmax. When $T = 200$, test for 1 versus 2 changes has good power, thus the probabilities of selecting the correct number of changes is always close to 1. However, the power decreases a lot when $T = 100$, and thus $\hat{L}_1$ and $\hat{L}_2$ tend to underestimate the number of changes. This is because when $T = 100$, the sample size of each regime is too small to be robust to the estimation error of the change points. We also conduct simulations gradually increasing T and find that when T increases to 140, the performance is as good as $T = 200$. Of course, the power also depends upon the location of the change points. We suggest that, for each regime, the sample size should be at least 40. Finally, when $T = 100$ serial and cross-sectional dependence of the errors decrease the power. This is again caused by small T. In summary, results in both tables are consistent with our theoretical derivation and show the usefulness of the proposed testing procedure.

7 CONCLUSIONS

This paper studies a high dimensional factor model with multiple changes. The main issues tackled are the estimation of change points, the estimation of the number of factors and the factor space in each regime, tests for the presence of multiple changes and tests for

[17] As discussed in Section 3.1.1, less conservative criterion is recommended in estimating the number of factors in the first step.

determining the number of changes. Our strategy is based on the second moments of the estimated pseudo factors and we show that estimation errors contained in the estimated factors have different effects on estimating and testing structural change. The proposed procedure is easy to implement, computationally efficient and able to take into account the effect of serial correlation. Simulation studies confirm the theoretical results and demonstrate its good performance. A natural next step is to use bootstrap to fix the undersizing issue when T is less than 100, as discussed in Diebold and Chen (1996). It will be also interesting to apply our theoretical results to study the financial market comovement during crises, as discussed in Bekaert, Ehrmann, Fratzscher and Mehl (2014) and Belvisi, Pianeti and Urga (2015).

Acknowledgements

Chapter 3 is based on the working paper Baltagi, Kao and Wang (2015c).

References

[1] AHN, S.C., AND A. R. HORENSTEIN (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227.

[2] BAI, J. (1996): "An Inequality for Vector-Valued Martingales and its Applications," MIT Working Paper No. 96-16.

[3] BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–173.

[4] BAI, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77, 1229-1279.

[5] BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

[6] BAI, J., AND S. NG (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, 1133-1150.

[7] BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.

[8] BAI, J., AND P. PERRON (2003): "Critical values for multiple structural change tests," *Econometrics Journal,* 6, 72-78.

[9] BALTAGI, B. H., C. KAO, AND F. WANG (2015b): "Identification and Estimation of a Large Factor Model with Structural Instability," Working Paper, Syracuse University.

[10] BALTAGI, B. H., C. KAO, AND F. WANG (2015c): "Estimating and Testing High Dimensional Factor Models with Multiple Structural Changes," Working Paper, Syracuse University.

[11] 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.

[12] BEKAERT, G., M. EHRMANN, M. FRATZSCHER, AND A. MEHL (2014): "The Global Crisis and Equity Market Contagion," *Journal of Finance*, 69, 2597-2649.

[13] BELVISI, M., R. PIANETI, AND G. URGA (2015): "Modelling Financial Markets Comovements During Crises: A Dynamic Multi-Factor Approach," Working Paper, Cass Business School

[14] BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Factor Augmented Vector Autoregression and the Analysis of Monetary Policy," *Quarterly Journal of Economics*, 120, 387–422.

[15] BREITUNG, J., AND S. EICKMEIER (2011): "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163, 71–84.

[16] CHEN, L. (2015): "Estimating the Common Break Date in Large Factor Models," *Economics Letters*, 131, 70-74.

[17] CHEN, L., J. DOLADO, AND J. GONZALO (2014): "Detecting Big Structural Breaks in Large Factor Models," *Journal of Econometrics*, 180, 30–48.

[18] CHENG, X., Z. LIAO, AND F. SCHORFHEIDE (2014): "Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities," NBER Working Paper No. 19792.

[19] CORRADI, V., AND N. R. SWANSON (2014): "Testing for Structural Stability of Factor Augmented Forecasting Models," *Journal of Econometrics*, 182, 100–118.

[20] DIEBOLD, F.X., AND C. CHEN (1996): "Testing Structural Stability with Endogenous Breakpoint A Size Comparison of Analytic and Bootstrap Procedures," *Journal of Econometrics*, 70, 221-241.

[21] HAN, X., AND A. INOUE (2014): "Tests for Parameter Instability in Dynamic Factor Models," *Econometric Theory,* 1-36.

[22] LEWBEL, A. (1991): "The Rank of Demand Systems: Theory and Nonparametric Estimation," *Econometrica*, 59, 711–730.

[23] MASSACCI, D. (2015): "Least Squares Estimation of Large Dimensional Threshold Factor Models," Working Paper, Einaudi Institute for Economics and Finance (EIEF).

[24] MOON, H.R., AND M.WEIDNER (2014): "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543-1579.

[25] ONATSKI, A. (2010): "Determining the Number of Factors From Empirical Distribution of Eigenvalues," *Review of Economic and Statistics*, 92, 1004–1016.

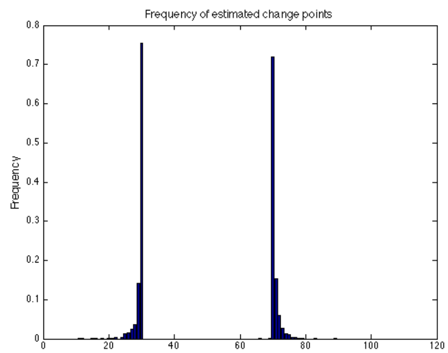
[26] QU, Z., AND P. PERRON (2007): "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica*, 75, 459-502.

[27] ROSS, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Finance*, 13, 341–360.

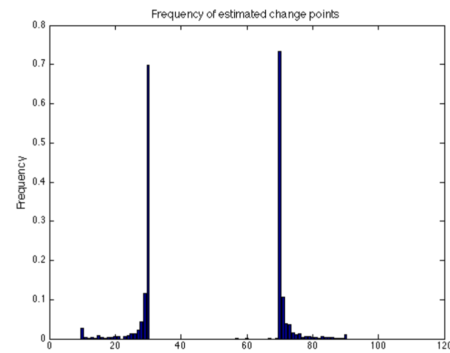
[28] SHI, Y. (2015): "Structural Change in High Dimensional Factor Models," Working Paper, Texas A&M University.

[29] STOCK, J. H., AND M. W. WATSON (2002a): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

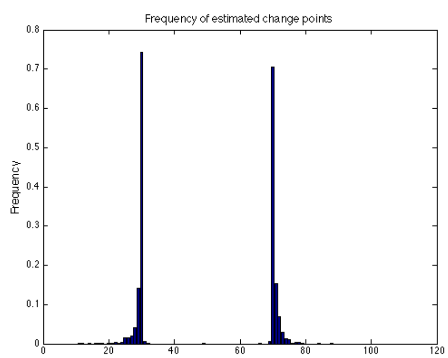
[30] STOCK, J. H., AND M. W. WATSON (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147-162.



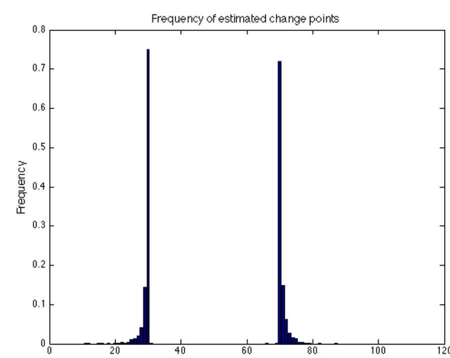
$(\rho, \alpha, \beta) = (0, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0.7, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

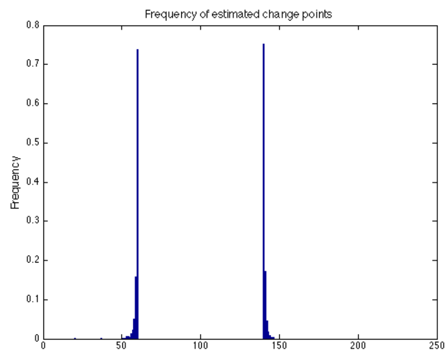


$(\rho, \alpha, \beta) = (0, 0.3, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

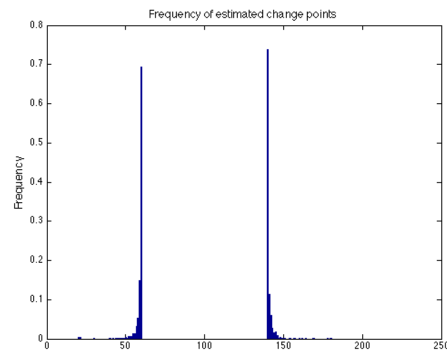


$(\rho, \alpha, \beta) = (0, 0, 0.3)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

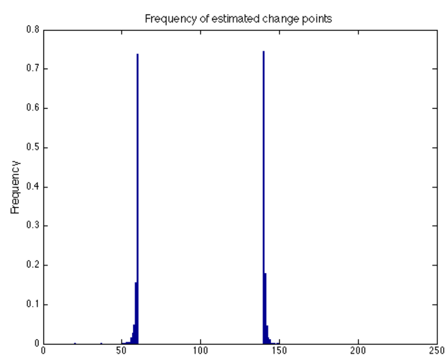
Figure 1: Histogram of estimated change points for $(N, T) = (100, 100), r_1 = 2, r_2 = 2, r_3 = 3, \bar{r} = 7$



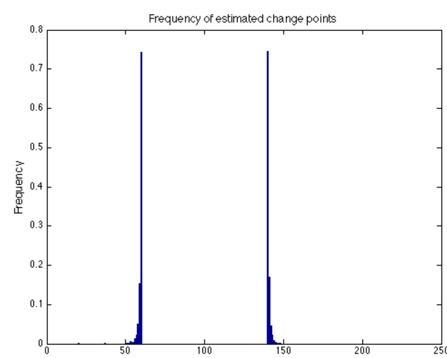
$(\rho, \alpha, \beta) = (0, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0.7, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0, 0.3, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0, 0, 0.3)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

Figure 2: Histogram of estimated change points for $(N, T) = (100, 200), r_1 = 2, r_2 = 2, r_3 = 3, \bar{r} = 7$

Table 1: Estimated number of factors in each regime for $r_1 = 2, r_2 = 2, r_3 = 3, \bar{r} = 7$

N	T	IC_{p2}			GR			ER		
		$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_3$	$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_3$	$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_3$
$\rho = 0, \alpha = 0, \beta = 0$										
100	100	0/0	0/1	1/0	1/0	1/0	5/0	1/0	0/0	3/0
100	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	300	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
$\rho = 0.7, \alpha = 0, \beta = 0$										
100	100	4/4	0/10	1/2	1/2	3/5	12/0	1/0	1/6	6/0
100	200	0/0	0/2	0/0	0/1	0/0	0/0	0/0	0/1	0/0
200	200	0/0	0/3	0/0	0/0	0/1	0/0	0/0	0/1	0/0
200	300	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0
$\rho = 0, \alpha = 0.3, \beta = 0$										
100	100	0/0	0/1	2/0	3/0	1/0	11/0	1/0	1/0	7/0
100	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	300	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
$\rho = 0, \alpha = 0, \beta = 0.3$										
100	100	0/0	0/0	1/0	1/0	1/0	6/0	1/0	0/0	4/0
100	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	300	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Table 2: Size of tests and probabilities of selecting changes

ϵ	$l\|0$			$Dmax$		$\hat{L}_1$			$\hat{L}_2$		
	1	2	3	U	W	0	1	2	0	1	2
$N = 100, T = 100, \rho = 0, \alpha = 0, \beta = 0$											
0.10	0.4	0.2	0.1	0.4	0.2	99.6	0.4	0	99.8	0.2	0
0.15	0.1	0	0	0.1	0.1	99.9	0.1	0	99.9	0.1	0
0.20	0	0	0	0	0	100	0	0	100	0	0
0.25	0.1	0	0	0	0	99.9	0.1	0	100	0	0
$N = 100, T = 200, \rho = 0, \alpha = 0, \beta = 0$											
0.05	1.8	1.8	1.7	1.6	1.4	98.2	1.8	0	98.6	1.4	0
0.10	0.2	0.2	0.5	0.3	0.1	99.8	0.2	0	99.9	0.1	0
0.15	0.6	0.5	0.2	0.7	0.2	99.4	0.6	0	99.8	0.2	0
0.20	0.4	0.3	0.1	0.4	0	99.6	0.4	0	100	0	0
0.25	0.9	0.4	0	0.7	0.2	99.1	0.9	0	99.8	0.2	0
$N = 100, T = 100, \rho = 0.7, \alpha = 0, \beta = 0$											
0.10	2.3	2.6	3.0	2.5	2.2	97.7	2.3	0	97.8	2.2	0
0.15	0.9	1.8	1.0	1.1	1.2	99.1	0.9	0	98.8	1.2	0
0.20	0.9	1.3	0.5	0.9	0.6	99.1	0.9	0	99.4	0.6	0
0.25	0.8	1.3	0	0.7	0.1	99.2	0.8	0	99.9	0.1	0
$N = 100, T = 200, \rho = 0.7, \alpha = 0, \beta = 0$											
0.05	12.7	25.9	23.4	15.9	17.5	87.3	11.8	0.8	82.5	16.1	0.13
0.10	5.3	8.4	8.8	6.4	7.5	94.7	5.1	0.2	92.5	7.2	0.3
0.15	4.5	5.9	4.2	5.1	5.2	95.5	4.5	0	94.8	5.0	0.2
0.20	3.4	4.2	4.0	3.3	3.4	96.6	3.4	0	96.6	3.4	0
0.25	3.6	3.5	0.3	2.8	2.1	96.4	3.6	0	97.9	2.1	0
$N = 100, T = 100, \rho = 0.7, \alpha = 0.3, \beta = 0.3$											
0.10	2.0	2.5	3.1	2.5	2.4	98.0	2.0	0	97.6	2.4	0
0.15	0.8	2.0	1.0	1.0	1.1	99.2	0.8	0	98.9	1.1	0
0.20	1.0	1.4	1.6	1.0	0.7	99.0	1.0	0	99.3	0.7	0
0.25	0.8	1.3	0.1	0.6	0.1	99.2	0.8	0	99.9	0.1	0
$N = 100, T = 200, \rho = 0.7, \alpha = 0.3, \beta = 0.3$											
0.05	12.5	26.8	23.8	16.3	17.8	87.5	11.7	0.7	82.2	16.5	1.2
0.10	5.4	8.0	8.2	6.2	7.3	94.6	5.2	0.2	92.7	7.0	0.3
0.15	4.6	5.6	4.2	5.3	5.3	95.4	4.6	0	94.7	5.2	0.1
0.20	3.7	4.0	1.9	3.6	3.2	96.3	3.7	0	96.8	3.2	0
0.25	3.6	3.5	0.3	2.9	2.1	96.4	3.6	0	97.9	2.0	0.1

Table 3: Power of tests and probabilities of selecting changes for $L = 2$

ϵ	$l\|0$			$Dmax$		$l+1\|l$		$\hat{L}_1$			$\hat{L}_2$		
	1	2	3	U	W	2|1	3|2	0	1	2	0	1	2
$N = 100, T = 100, \rho = 0, \alpha = 0, \beta = 0$													
0.10	100	100	100	98.4	100	23.4	0	0	76.6	23.4	0	76.6	23.4
0.15	100	100	100	23.1	100	12.4	0	0	87.6	12.4	0	87.6	12.4
0.20	100	100	100	4.9	99.9	9.6	0	0	90.4	9.6	0.1	90.3	9.6
0.25	100	100	100	3.6	3.7	11.1	0	0	88.9	11.1	96.3	3.3	0.4
$N = 100, T = 200, \rho = 0, \alpha = 0, \beta = 0$													
0.05	100	100	100	100	100	100	0.5	0	0	99.5	0	0	99.5
0.10	100	100	100	100	100	100	0	0	0	100	0	0	100
0.15	100	100	100	100	100	100	0	0	0	100	0	0	100
0.20	100	100	100	100	100	100	0	0	0	100	0	0	100
0.25	100	100	100	100	100	100	0	0	0	100	0	0	100
$N = 100, T = 100, \rho = 0.7, \alpha = 0, \beta = 0$													
0.10	100	100	100	98.9	100	41.9	0.1	0	58.1	41.8	0	58.1	41.8
0.15	100	100	100	28.7	100	23.3	0	0	76.7	23.3	0	76.7	23.3
0.20	100	100	100	5.9	100	15.8	0	0	84.2	15.8	0	84.2	15.8
0.25	100	100	100	4.3	4.3	15.5	0	0	84.5	15.5	95.7	3.6	0.7
$N = 100, T = 200, \rho = 0.7, \alpha = 0, \beta = 0$													
0.05	100	100	100	100	100	100	3.9	0	0	96.1	0	0	96.1
0.10	100	100	100	100	100	100	0.4	0	0	99.6	0	0	99.6
0.15	100	100	100	100	100	100	0.1	0	0	99.9	0	0	99.9
0.20	100	100	100	100	100	100	0	0	0	100	0	0	100
0.25	100	100	100	100	100	100	0	0	0	100	0	0	100
$N = 100, T = 100, \rho = 0.7, \alpha = 0.3, \beta = 0.3$													
0.10	97.3	98.5	99.9	78.5	97.7	37.0	0.3	2.7	60.6	36.5	2.3	60.9	36.6
0.15	97.5	98.9	100	16.9	96.9	19.6	0	2.5	78.0	19.5	3.1	77.4	19.5
0.20	97.5	99.9	100	1.3	95.1	15.3	0	2.5	82.2	15.3	4.9	80.1	15.0
0.25	97.5	99.9	99.2	0.1	1.4	15.7	0	2.5	81.9	15.6	98.6	1.2	0.2
$N = 100, T = 200, \rho = 0.7, \alpha = 0.3, \beta = 0.3$													
0.05	100	100	100	100	100	100	4.2	0	0	95.8	0	0	95.8
0.10	100	100	100	100	100	100	0.4	0	0	99.6	0	0	99.6
0.15	100	100	100	100	100	100	0.1	0	0	99.9	0	0	99.9
0.20	100	100	100	100	100	100	0	0	0	100	0	0	100
0.25	100	100	100	100	100	100	0	0	0	100	0	0	100

APPENDIX

A HAJEK-RENYI INEQUALITY

For a sequence of independent random variables $\{x_t, t = 1, ...\}$ with $\mathbb{E}x_t = 0$ and $\mathbb{E}x_t^2 = \sigma_t^2$, Hajek and Renyi proved that for any integers m and T,

$$P(\sup_{m\leq k\leq T} c_k \left|\sum_{t=1}^{k} x_t\right| > M) \leq \frac{1}{M^2}(c_m^2 \sum_{t=1}^{m} \sigma_t^2 + \sum_{t=m+1}^{T} c_t^2\sigma_t^2), \tag{16}$$

where $\{c_k, k = 1, ...\}$ is a sequence of nonincreasing positive numbers. It is easy to see that if $\sigma_t^2 \leq \sigma^2$ for all t and $c_k = \frac{1}{k}$, $P(\sup_{m\leq k\leq T} \left|\frac{1}{k}\sum_{t=1}^{k} x_t\right| > M) \leq \frac{2\sigma^2}{M^2}\frac{1}{m}$, thus $\sup_{1\leq k\leq T} \left|\frac{1}{k}\sum_{t=1}^{k} x_t\right| = O_p(1)$ and $\sup_{T\tau\leq k\leq T} \left|\frac{1}{k}\sum_{t=1}^{k} x_t\right| = O_p(\frac{1}{\sqrt{T}})$. If $c_k = \frac{1}{\sqrt{k}}$, $P(\sup_{m\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| > M) \leq \frac{\sigma^2}{M^2}(1 + \sum_{k=m+1}^{T}\frac{1}{k})$, thus $\sup_{1\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| = O_p(\sqrt{\log T})$ since $\sum_{k=1}^{T}\frac{1}{k} - \log T$ converges to the Euler constant and $\sup_{T\tau\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| = O_p(1)$ since $\sum_{k=m+1}^{T}\frac{1}{k} = \sum_{k=1}^{T}\frac{1}{k} - \sum_{k=1}^{T\tau}\frac{1}{k} \to \log T - \log T\tau = \log\frac{1}{\tau}$.

Hajek-Renyi inequality is a more powerful tool than the functional CLT for calculating the stochastic order of sup-type terms. It has been extended to various settings, including martingale difference, martingale, vector-valued martingale, mixingale and linear process, see Bai (1996).

B PROOF OF PROPOSITION 1

Proof. For any $\epsilon > 0$ and $\eta_1 > 0, ..., \eta_L > 0$, define $D = \{(k_1, ..., k_L) : (\tau_{\iota 0} - \eta_\iota)T \leq k_\iota \leq (\tau_{\iota 0} + \eta_\iota)T$ for $\iota = 1, ..., L\}$, we need to show $P((\tilde{k}_1, ..., \tilde{k}_L) \in D^c) < \epsilon$ as $(N, T) \to \infty$. Since $D^c = \cup_{\iota=1}^{L}\{(k_1, ..., k_L)$:for $\kappa = 1, ..., L$, either $k_\kappa < (\tau_{\iota 0} - \eta_\iota)T$ or $k_\kappa > (\tau_{\iota 0} + \eta_\iota)T\} = \cup_{\iota=1}^{L} D_{(\iota)}^c$, it suffices to show $P((\tilde{k}_1, ..., \tilde{k}_L) \in D_{(\iota)}^c) < \epsilon$ as $(N, T) \to \infty$ for all ι. Since $(\tilde{k}_1, ..., \tilde{k}_L) = \arg\min \tilde{S}(k_1, ..., k_L)$, $\tilde{S}(\tilde{k}_1, ..., \tilde{k}_L) \leq \tilde{S}(k_{1,0}, ..., k_{L,0}) \leq \sum_{t=1}^{T}(y_t + z_t)'(y_t + z_t)$. If $(\tilde{k}_1, ..., \tilde{k}_L) \in D_{(\iota)}^c$, then $\min_{(k_1,...,k_L)\in D_{(\iota)}^c} \tilde{S}(k_1, ..., k_L) = \tilde{S}(\tilde{k}_1, ..., \tilde{k}_L)$. Thus $(\tilde{k}_1, ..., \tilde{k}_L) \in D_{(\iota)}^c$ implies $\min_{(k_1,...,k_L)\in D_{(\iota)}^c} \tilde{S}(k_1, ..., k_L) \leq \sum_{t=1}^{T}(y_t + z_t)'(y_t + z_t)$ and it suffices to show $P(\min_{(k_1,...,k_L)\in D_{(\iota)}^c} \tilde{S}(k_1, ..., k_L) - \sum_{t=1}^{T}(y_t + z_t)'(y_t + z_t) \leq 0) < \epsilon$ as $(N, T) \to \infty$.

For any given partition $(k_1, ..., k_L)$, let $\tilde{\Sigma}_\iota = \frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota} \tilde{g}_t\tilde{g}_t'$ and $a_t = vech(\Sigma_\kappa - \tilde{\Sigma}_\iota)$

for $t \in [k_{\iota-1}+1, k_\iota] \cap [k_{\kappa-1,0}+1, ..., k_{\kappa,0}]$, $\iota, \kappa = 1, ..., L+1$. It follows $vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\iota) = a_t + y_t + z_t$ and

$$\begin{aligned} \tilde{S}(k_1, ..., k_L) &= \sum_{\iota=1}^{L+1} \sum_{t=k_{\iota-1}+1}^{k_\iota} [vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\iota)]'[vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\iota)] \\ &= \sum_{t=1}^{T} (y_t + z_t)'(y_t + z_t) + \sum_{t=1}^{T} a_t'a_t + 2\sum_{t=1}^{T} a_t'(y_t + z_t). \end{aligned} \quad (17)$$

Thus it suffices to show $P(\min_{(k_1,...,k_L)\in D^c_{(\iota)}} [\sum_{t=1}^T a_t'a_t + 2\sum_{t=1}^T a_t'(y_t + z_t)] \leq 0) < \epsilon$ as $(N, T) \to \infty$. Since $\min_{(k_1,...,k_L)\in D^c_{(\iota)}} [\sum_{t=1}^T a_t'a_t + 2\sum_{t=1}^T a_t'(y_t + z_t)] \leq 0$ implies $\min_{(k_1,...,k_L)\in D^c_{(\iota)}} \sum_{t=1}^T a_t'a_t \leq 2 \sup_{(k_1,...,k_L)\in D^c_{(\iota)}} \left|\sum_{t=1}^T a_t'(y_t + z_t)\right|$, it suffices to show that the left hand side dominates the right hand side asymptotically.

Consider the left hand side first. For any $(k_1, ..., k_L) \in D^c_{(\iota)}$, there exists κ^* such that $k_{\kappa^*-1} < (\tau_{\iota 0} - \eta_\iota)T$ and $k_{\kappa^*} > (\tau_{\iota 0} + \eta_\iota)T$, thus for $t \in [(\tau_{\iota 0} - \eta_\iota)T, \tau_{\iota 0}T]$, $a_t = vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_\iota)$ and for $t \in [\tau_{\iota 0}T + 1, (\tau_{\iota 0} + \eta_\iota)T]$, $a_t = vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_{\iota+1})$. So for any $(k_1, ..., k_L) \in D^c_{(\iota)}$,

$$\begin{aligned} & \sum_{t=1}^{T} a_t'a_t \\ \geq\ & \sum_{t=(\tau_{\iota 0}-\eta_\iota)T}^{\tau_{\iota 0}T} a_t'a_t + \sum_{t=\tau_{\iota 0}T+1}^{(\tau_{\iota 0}+\eta_\iota)T} a_t'a_t \\ \geq\ & \eta_\iota T[vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_\iota)'vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_\iota) + vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_{\iota+1})'vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_{\iota+1})] \\ \geq\ & \eta_\iota T \frac{vech(\Sigma_\iota - \Sigma_{\iota+1})'vech(\Sigma_\iota - \Sigma_{\iota+1})}{2}, \end{aligned} \quad (18)$$

where the last inequality is due to $(x-a)^2 + (x-b)^2 = 2(x - \frac{a+b}{2})^2 + \frac{(a-b)^2}{2}$ for any x. Thus $\min_{(k_1,...,k_L)\in D^c_{(\iota)}} \sum_{t=1}^T a_t'a_t \geq \eta_\iota T \frac{vech(\Sigma_\iota - \Sigma_{\iota+1})'vech(\Sigma_\iota - \Sigma_{\iota+1})}{2}$. Next, the right hand side is no larger than

$$\left|\sum_{\kappa=1}^{L+1} \sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'(y_t + z_t)\right| \quad (19)$$

$$+ \sup_{(k_1,...,k_L)\in D^c_{(\iota)}} \left|\sum_{\iota=1}^{L+1} \sum_{t=k_{\iota-1}+1}^{k_\iota} vech(\tilde{\Sigma}_\iota)'(y_t + z_t)\right|. \quad (20)$$

For the first term,

$$\begin{aligned}
&\left|\sum_{\kappa=1}^{L+1}\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'y_t\right| \\
\leq\ & \sum_{\kappa=1}^{L+1}\left|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'y_t\right| \leq \sum_{\kappa=1}^{L+1}\|\Sigma_\kappa\|\left\|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} y_t\right\| \\
\leq\ & \sum_{\kappa=1}^{L+1}\|\Sigma_\kappa\|\,\|J_0\|^2\,\|R_\kappa\|^2\left\|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}}(f_tf_t'-\Sigma_F)\right\| = o_p(T),
\end{aligned} \tag{21}$$

where the last equality follows from Assumption 1; and

$$\left|\sum_{\kappa=1}^{L+1}\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'z_t\right| \leq \sum_{\kappa=1}^{L+1}\|\Sigma_\kappa\|\left\|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} z_t\right\| = o_p(T), \tag{22}$$

where the last equality follows from Lemma 5. For the second term, define $b_t = vech(\Sigma_\kappa)$ for $t \in [k_{\kappa-1,0}+1,...,k_{\kappa,0}]$, $\kappa = 1,...,L+1$, then $vech(\tilde{g}_t\tilde{g}_t') = b_t + y_t + z_t$ for all t and $vech(\tilde{\Sigma}_\iota) = \frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota} vech(\tilde{g}_t\tilde{g}_t') = \frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota}(b_t+y_t+z_t)$. It follows that the second term is no larger than

$$\begin{aligned}
&\sup_{(k_1,...,k_L)\in D^c_{(\iota)}}\left|\sum_{\iota=1}^{L+1}\frac{1}{k_\iota-k_{\iota-1}}(\sum_{t=k_{\iota-1}+1}^{k_\iota} b_t)'(\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t))\right| \\
&+\sup_{(k_1,...,k_L)\in D^c_{(\iota)}}\left|\sum_{\iota=1}^{L+1}\frac{1}{k_\iota-k_{\iota-1}}(\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t))'(\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t))\right| \\
\leq\ & (L+1)(\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t)}{\sqrt{k_\iota-k_{\iota-1}}}\right\|^2 + \\
&\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} b_t}{\sqrt{k_\iota-k_{\iota-1}}}\right\|\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t)}{\sqrt{k_\iota-k_{\iota-1}}}\right\|) \\
=\ & (L+1)(B^2+AB).
\end{aligned} \tag{23}$$

For term A, we have $A \leq \sup_{1\leq k_{\iota-1}<k_\iota\leq T}\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|b_t\|}{\sqrt{k_\iota-k_{\iota-1}}} \leq \sup_{1\leq k_{\iota-1}<k_\iota\leq T}\sqrt{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|b_t\|^2} \leq \sqrt{\sum_{t=1}^{T}\|b_t\|^2} = O(\sqrt{T})$. For term B, we have $B^2 \leq 2\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} y_t}{\sqrt{k_\iota-k_{\iota-1}}}\right\|^2 + 2\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} z_t}{\sqrt{k_\iota-k_{\iota-1}}}\right\|^2 = 2B_1^2+2B_2^2$. $B_1 = o_p(\sqrt{T})$, since

$$\begin{aligned}
B_1 \ \leq\ & \sum_{\kappa=1}^{L+1}\sup_{k_{\kappa-1,0}<k<l\leq k_{\kappa,0}}\left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l} y_t\right\| \\
\leq\ & \sum_{\kappa=1}^{L+1}\|J_0\|^2\,\|R_\kappa\|^2\sup_{k_{\kappa-1,0}<k<l\leq k_{\kappa,0}}\left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l}\epsilon_t\right\|,
\end{aligned} \tag{24}$$

and by Assumption 7,

$$\begin{aligned}
&\mathbb{E}(\sup_{k_{\kappa-1,0}<k<l\leq k_{\kappa,0}} \left\| \frac{1}{\sqrt{l-k}} \sum\nolimits_{t=k+1}^{l} \epsilon_t \right\|^{4+\delta}) \\
\leq\ & \sum\nolimits_{k=k_{\kappa-1,0}}^{k_{\kappa,0}-1} \sum\nolimits_{l=k+1}^{k_{\kappa,0}} \mathbb{E}(\left\| \frac{1}{\sqrt{l-k}} \sum\nolimits_{t=k+1}^{l} \epsilon_t \right\|^{4+\delta}) \leq T^2 M. \qquad (25)
\end{aligned}$$

Using Lemma 5, $B_2 = o_p(\sqrt{T})$. Taking together, the right hand side is $o_p(T)$ and thus dominated by the left hand side. ∎

C PROOF OF THEOREM 1

Proof. From Proposition 1, we know that for any $\epsilon > 0$ and $\eta_1 > 0, ..., \eta_L > 0$, $P((\tilde{k}_1, ..., \tilde{k}_L) \in D^c) < \epsilon$ as $(N, T) \to \infty$. Thus to show $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$ for any given $1 \leq \iota \leq L$, we need to show for any $\epsilon > 0$ and $\eta_1 > 0, ..., \eta_L > 0$, there exist $C > 0$ such that $P((\tilde{k}_1, ..., \tilde{k}_L) \in D, \left|\tilde{k}_\iota - k_{\iota 0}\right| > C) < \epsilon$ as $(N, T) \to \infty$. By symmetry, it suffices to show $P((\tilde{k}_1, ..., \tilde{k}_L) \in D, \tilde{k}_\iota < k_{\iota 0} - C) < \epsilon$ as $(N, T) \to \infty$. Define $D(C)_{(\iota)} = D \cap \{k_\iota < k_{\iota 0} - C\}$. Since $(\tilde{k}_1, ..., \tilde{k}_L) = \arg\min \tilde{S}(k_1, ..., k_L)$, $\tilde{S}(\tilde{k}_1, ..., \tilde{k}_L) \leq \tilde{S}(\tilde{k}_1, ..., k_{\iota 0}, ..., \tilde{k}_L)$. Thus if $(\tilde{k}_1, ..., \tilde{k}_L) \in D(C)_{(\iota)}$,

$$\begin{aligned}
&\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} [\tilde{S}(k_1, ..., k_L) - \tilde{S}(k_1, ..., k_{\iota 0}, ..., k_L)] \\
\leq\ & \tilde{S}(\tilde{k}_1, ..., \tilde{k}_L) - \tilde{S}(\tilde{k}_1, ..., k_{\iota 0}, ..., \tilde{k}_L) \leq 0.
\end{aligned}$$

Therefore it suffices to show $P(\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} [\tilde{S}(k_1, ..., k_L) - \tilde{S}(k_1, ..., k_{\iota 0}, ..., k_L)] \leq 0) < \epsilon$ as $(N, T) \to \infty$.

We then show that the event $\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} [\tilde{S}(k_1, ..., k_L) - \tilde{S}(k_1, ..., k_{\iota 0}, ..., k_L)] \leq 0$ is just the event $\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} \frac{\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)}{|k_\iota - k_{\iota 0}|} \leq 0$. Conditioning on the former, for any $(k_1^*, ..., k_L^*) \in D(C)_{(\iota)}$, $\arg\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} [\tilde{S}(k_1, ..., k_L) - \tilde{S}(k_1, ..., k_{\iota 0}, ..., k_L)] = (k_1^*, ..., k_L^*)$ implies $\tilde{S}(k_1^*, ..., k_L^*) - \tilde{S}(k_1^*, ..., k_{\iota 0}, ..., k_L^*) \leq 0$, and this further implies $\frac{\tilde{S}(k_1^*,...,k_L^*)-\tilde{S}(k_1^*,...,k_{\iota 0},...,k_L^*)}{|k_\iota^* - k_{\iota 0}|} \leq 0$. Thus $\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} \frac{\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)}{|k_\iota - k_{\iota 0}|}$, which is not larger than $\frac{\tilde{S}(k_1^*,...,k_L^*)-\tilde{S}(k_1^*,...,k_{\iota 0},...,k_L^*)}{|k_\iota^* - k_{\iota 0}|}$, has to be nonpositive. Note that the above argument holds for any $(k_1^*, ..., k_L^*) \in D(C)_{(\iota)}$, thus the former implies the latter. Similarly, the latter also implies the former. Therefore, it suffices to show $P(\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} \frac{\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)}{|k_\iota - k_{\iota 0}|} \leq 0) < \epsilon$ as $(N,T)\to\infty$.

Next, decompose $\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)$ as

$$[\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_\iota,k_{\iota 0},...,k_L)] \tag{26}$$

$$-[\tilde{S}(k_1,...,k_{\iota 0},...,k_L)-\tilde{S}(k_1,...,k_\iota,k_{\iota 0},...,k_L)]. \tag{27}$$

Term (26) equals

$$\begin{aligned} & \sum\nolimits_{t=k_\iota+1}^{k_{\iota+1}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})] \\ & -\sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^\Delta)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^\Delta)] \\ & -\sum\nolimits_{t=k_{\iota 0}+1}^{k_{\iota+1}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1}^*)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1}^*)] \\ = \ & K_1-K_2-K_3, \end{aligned} \tag{28}$$

and term (27) equals

$$\begin{aligned} & \sum\nolimits_{t=k_{\iota-1}+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^*)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^*)] \\ & -\sum\nolimits_{t=k_{\iota-1}+1}^{k_\iota}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)] \\ & -\sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^\Delta)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^\Delta)] \\ = \ & L_1-L_2-L_3, \end{aligned} \tag{29}$$

where $\tilde{\Sigma}_\iota^\Delta = \frac{\sum_{t=k_\iota+1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}-k_\iota}$, $\tilde{\Sigma}_\iota^* = \frac{\sum_{t=k_{\iota-1}+1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}-k_{\iota-1}}$ and $\tilde{\Sigma}_{\iota+1}^* = \frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1}}\tilde{g}_t\tilde{g}_t'}{k_{\iota+1}-k_{\iota 0}}$. Note that $L_3 = K_2$, thus $(K_1-K_2-K_3)-(L_1-L_2-L_3) = (K_1-K_3)-(L_1-L_2)$. Replacing $\tilde{\Sigma}_{\iota+1}^*$ by $\tilde{\Sigma}_{\iota+1}$, K_3 is magnified, thus $K_1-K_3 \geq \sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]$; and replacing $\tilde{\Sigma}_\iota^*$ by $\tilde{\Sigma}_\iota$, L_1 is magnified, thus $L_1-L_2 \leq \sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]$. Taken together,

$$
\begin{aligned}
& (K_1 - K_3) - (L_1 - L_2) \\
\geq\ & \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} [vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_{\iota+1})]'[vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_{\iota+1})] \\
& - \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} [vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_{\iota})]'[vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_{\iota})] \\
=\ & \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} vech(\Sigma_\iota - \tilde{\Sigma}_{\iota+1})' vech(\Sigma_\iota - \tilde{\Sigma}_{\iota+1}) \\
& - \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} vech(\Sigma_\iota - \tilde{\Sigma}_{\iota})' vech(\Sigma_\iota - \tilde{\Sigma}_{\iota}) \\
& + 2\sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} vech(\Sigma_\iota - \tilde{\Sigma}_{\iota+1})'(y_t + z_t) \\
& - 2\sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} vech(\Sigma_\iota - \tilde{\Sigma}_{\iota})'(y_t + z_t) \\
=\ & K_{\Delta 1} - L_{\Delta 1} + K_{\Delta 2} - L_{\Delta 2}, \qquad (30)
\end{aligned}
$$

thus it suffices to show $P(\min_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}} \frac{K_{\Delta 1} - L_{\Delta 1} + K_{\Delta 2} - L_{\Delta 2}}{|k_\iota - k_{\iota 0}|} \leq 0) < \epsilon$ as $(N, T) \to \infty$.

We consider the case $k_{\iota-1} < k_{\iota-1,0}$ and $k_{\iota+1} > k_{\iota+1,0}$. In case $k_{\iota-1} \geq k_{\iota-1,0}$ or $k_{\iota+1} \leq k_{\iota+1,0}$, the proof is easier and therefore omitted. Plug in $\tilde{\Sigma}_{\iota+1} = \frac{1}{k_{\iota+1}-k_\iota}\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t + z_t) + vech(\frac{1}{k_{\iota+1}-k_\iota}[(k_{\iota 0}-k_\iota)\Sigma_\iota + (k_{\iota+1,0}-k_{\iota 0})\Sigma_{\iota+1} + (k_{\iota+1}-k_{\iota+1,0})\Sigma_{\iota+2}])$ and $\tilde{\Sigma}_\iota = \frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}}^{k_\iota}(y_t + z_t) + vech(\frac{1}{k_\iota-k_{\iota-1}}[(k_{\iota-1,0}-k_{\iota-1})\Sigma_{\iota-1} + (k_\iota - k_{\iota-1,0})\Sigma_\iota])$, and denote $\phi_{k_{\iota-1},k_\iota} = vech(\frac{1}{k_\iota - k_{\iota-1}}(k_{\iota-1,0} - k_{\iota-1})(\Sigma_{\iota-1} - \Sigma_\iota))$ and $\phi_{k_\iota,k_{\iota+1}} = vech(\frac{1}{k_{\iota+1}-k_\iota}[(k_{\iota+1,0} - k_{\iota 0})(\Sigma_{\iota+1} - \Sigma_\iota) + (k_{\iota+1} - k_{\iota+1,0})(\Sigma_{\iota+2} - \Sigma_\iota)])$, we have

$$
\begin{aligned}
\frac{1}{k_{\iota 0} - k_\iota} K_{\Delta 1} &= [\phi_{k_\iota,k_{\iota+1}} + \frac{\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t + z_t)}{k_{\iota+1} - k_\iota}]'[\phi_{k_\iota,k_{\iota+1}} + \frac{\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t + z_t)}{k_{\iota+1} - k_\iota}], \qquad (31) \\
\frac{1}{k_{\iota 0} - k_\iota} L_{\Delta 1} &= [\phi_{k_{\iota-1},k_\iota} + \frac{\sum_{t=k_{\iota-1}}^{k_\iota}(y_t + z_t)}{k_\iota - k_{\iota-1}}]'[\phi_{k_{\iota-1},k_\iota} + \frac{\sum_{t=k_{\iota-1}}^{k_\iota}(y_t + z_t)}{k_\iota - k_{\iota-1}}], \qquad (32) \\
\frac{1}{k_{\iota 0} - k_\iota} K_{\Delta 2} &= -2[\phi_{k_\iota,k_{\iota+1}} + \frac{\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t + z_t)}{k_{\iota+1} - k_\iota}]'\frac{\sum_{t=k_\iota+1}^{k_{\iota 0}}(y_t + z_t)}{k_{\iota 0} - k_\iota}, \qquad (33) \\
\frac{1}{k_{\iota 0} - k_\iota} L_{\Delta 2} &= 2[\phi_{k_{\iota-1},k_\iota} + \frac{\sum_{t=k_{\iota-1}}^{k_\iota}(y_t + z_t)}{k_\iota - k_{\iota-1}}]'\frac{\sum_{t=k_\iota+1}^{k_{\iota 0}}(y_t + z_t)}{k_{\iota 0} - k_\iota}. \qquad (34)
\end{aligned}
$$

For $(k_1, ..., k_L) \in D(C)_{(\iota)}$ and η_ι and $\eta_{\iota+1}$ small enough,

$$\begin{aligned}
& \left\|\phi_{k_\iota, k_{\iota+1}}\right\| \\
\geq\ & \frac{k_{\iota+1,0} - k_{\iota 0}}{k_{\iota+1} - k_\iota} \left\|vech(\Sigma_{\iota+1} - \Sigma_\iota)\right\| - \frac{k_{\iota+1} - k_{\iota+1,0}}{k_{\iota+1} - k_\iota} \left\|vech(\Sigma_{\iota+2} - \Sigma_\iota)\right\| \\
\geq\ & \frac{1}{1 + \frac{\eta_{\iota+1} + \eta_\iota}{\tau_{\iota+1,0} - \tau_{\iota 0}}} \left\|vech(\Sigma_{\iota+1} - \Sigma_\iota)\right\| - \frac{\eta_{\iota+1}}{\eta_{\iota+1} + \tau_{\iota+1,0} - \tau_{\iota 0}} \left\|vech(\Sigma_{\iota+2} - \Sigma_\iota)\right\| \\
\geq\ & \frac{1}{2} \left\|vech(\Sigma_{\iota+1} - \Sigma_\iota)\right\|, \qquad (35)
\end{aligned}$$

and for $\eta_{\iota-1}$ and η_ι small enough,

$$\left\|\phi_{k_{\iota-1}, k_\iota}\right\| = \frac{k_{\iota-1,0} - k_{\iota-1}}{k_\iota - k_{\iota-1}} \left\|vech(\Sigma_{\iota-1} - \Sigma_\iota)\right\| \leq \frac{\eta_{\iota-1}}{\tau_{\iota 0} - \tau_{\iota-1,0} - \eta_\iota} \left\|vech(\Sigma_{\iota-1} - \Sigma_\iota)\right\| \qquad (36)$$

is arbitrarily small.

$$\begin{aligned}
& \sup_{(k_1,...,k_L) \in D(C)_{(\iota)}} \left\| \frac{1}{k_{\iota+1} - k_\iota} \sum\nolimits_{t=k_\iota+1}^{k_{\iota+1}} (y_t + z_t) \right\| \\
\leq\ & \frac{1}{\tau_{\iota+1,0} - \tau_{\iota 0}} \Big(\sup_{(k_1,...,k_L) \in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} (y_t + z_t) \right\| \\
& + \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} (y_t + z_t) \right\| + \sup_{(k_1,...,k_L) \in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota+1,0}+1}^{k_{\iota+1}} (y_t + z_t) \right\| \Big) \\
=\ & o_p(1), \qquad (37)
\end{aligned}$$

where we have used

$$\begin{aligned}
\sup_{(k_1,...,k_L) \in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} y_t \right\| &= o_p(1), \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} y_t \right\| = o_p(1), \\
\sup_{(k_1,...,k_L) \in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota+1,0}+1}^{k_{\iota+1}} y_t \right\| &= o_p(1), \qquad (38) \\
\sup_{(k_1,...,k_L) \in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} z_t \right\| &= o_p(1), \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} z_t \right\| = o_p(1), \\
\sup_{(k_1,...,k_L) \in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota+1,0}+1}^{k_{\iota+1}} z_t \right\| &= o_p(1). \qquad (39)
\end{aligned}$$

The first three terms follow from Hajek-Renyi inequality, which is proved in Lemma 1 to be applicable to y_t within each regime while the last three terms follow from Lemma 5.

Similarly,

$$\sup_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_\iota - k_{\iota-1}} \sum\nolimits_{t=k_{\iota-1}}^{k_\iota} (y_t + z_t) \right\| = o_p(1), \tag{40}$$

using Lemma 5 and

$$\sup_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota-1}+1}^{k_{\iota-1,0}} y_t \right\| = o_p(1), \tag{41}$$

$$\sup_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T} \sum\nolimits_{t=k_{\iota-1,0}+1}^{k_\iota} y_t \right\| = o_p(1). \tag{42}$$

Finally,

$$\begin{aligned} & \sup_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_{\iota 0} - k_\iota} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} (y_t + z_t) \right\| \\ \leq & \sup_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_{\iota 0} - k_\iota} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} y_t \right\| + \sup_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_{\iota 0} - k_\iota} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} z_t \right\| \\ = & O_p(\frac{1}{\sqrt{C}}) + o_p(1), \end{aligned} \tag{43}$$

the first term follows from Hajek-Renyi inequality while the second terms follows from Lemma 5. Taken together and choosing sufficiently large C, the result follows. ∎

D PROOF OF PROPOSITION 2

Proof. To simplify calculation, consider the case with two breaks. For any $\epsilon > 0$ and $\eta > 0$, define $W_\eta = \{k_1 : (\tau_{1,0} - \eta)T \leq k_1 \leq (\tau_{1,0} + \eta)T\}$, we need to show $P((\hat{k}_1 \in W_\eta^c) < \epsilon$ as $(N,T) \to \infty$. Since $\hat{k}_1 = \arg\min \tilde{S}(k_1)$, $\tilde{S}(\hat{k}_1) \leq \tilde{S}(k_{1,0})$. If $\hat{k}_1 \in W_\eta^c$, then $\min_{k_1 \in W_\eta^c} \tilde{S}(k_1) = \tilde{S}(\hat{k}_1)$. Thus $\hat{k}_1 \in W_\eta^c$ implies $\min_{k_1 \in W_\eta^c} \tilde{S}(k_1) \leq \tilde{S}(k_{1,0})$ and it suffices to show $P(\min_{k_1 \in W_\eta^c} \tilde{S}(k_1) - \tilde{S}(k_{1,0}) \leq 0) < \epsilon$ as $(N,T) \to \infty$.

For $k_1 < k_{1,0}$, after some calculation, we have:

$$\tilde{S}(k_1) - \tilde{S}(k_{1,0}) = \Pi_1(k_1) - \Pi(k_{1,0}) + \Psi_1(k_1) - \Psi(k_{1,0}), \tag{44}$$

where

$$\begin{aligned} & \Pi_1(k_1) - \Pi(k_{1,0}) \\ = & \frac{k_{1,0} - k_1}{(T-k_1)(T-k_{1,0})} \left\| vech[(T-k_{1,0})(\Sigma_1 - \Sigma_2) + (T-k_{2,0})(\Sigma_2 - \Sigma_3)] \right\|^2, \end{aligned} \tag{45}$$

$$
\begin{aligned}
&\Psi_1(k_1) \\
= \;& 2\varphi_{k_1}^{1\prime} \sum\nolimits_{\iota=k_1+1}^{k_{1,0}} (y_t + z_t) + 2\varphi_{k_1}^{2\prime} \sum\nolimits_{\iota=k_{1,0}+1}^{k_{2,0}} (y_t + z_t) + 2\varphi_{k_1}^{3\prime} \sum\nolimits_{\iota=k_{2,0}+1}^{T} (y_t + z_t) \\
& -2[(k_{1,0} - k_1)\varphi_{k_1}^1 + (k_{2,0} - k_{1,0})\varphi_{k_1}^2 + (T - k_{2,0})\varphi_{k_1}^3]' \frac{1}{T - k_1} \sum\nolimits_{\iota=k_1+1}^{T} (y_t + z_t) \\
& - \left\| \frac{1}{\sqrt{k_1}} \sum\nolimits_{\iota=1}^{k_1} (y_t + z_t) \right\|^2 - \left\| \frac{1}{\sqrt{T - k_1}} \sum\nolimits_{\iota=k_1+1}^{T} (y_t + z_t) \right\|^2, \qquad (46)
\end{aligned}
$$

$$
\begin{aligned}
\Psi(k_{1,0}) \;=\; & 2\varphi_{k_{1,0}}^{2\prime} \sum\nolimits_{\iota=k_{1,0}+1}^{k_{2,0}} (y_t + z_t) + 2\varphi_{k_{1,0}}^{3\prime} \sum\nolimits_{\iota=k_{2,0}+1}^{T} (y_t + z_t) \\
& -2[(k_{2,0} - k_{1,0})\varphi_{k_{1,0}}^2 + (T - k_{2,0})\varphi_{k_{1,0}}^3]' \frac{1}{T - k_{1,0}} \sum\nolimits_{\iota=k_{1,0}+1}^{T} (y_t + z_t) \\
& - \left\| \frac{1}{\sqrt{k_{1,0}}} \sum\nolimits_{\iota=1}^{k_{1,0}} (y_t + z_t) \right\|^2 - \left\| \frac{1}{\sqrt{T - k_{1,0}}} \sum\nolimits_{\iota=k_{1,0}+1}^{T} (y_t + z_t) \right\|^2, \qquad (47)
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{k_1}^1 &= \frac{1}{T - k_1} vech[(k_{2,0} - k_{1,0})(\Sigma_1 - \Sigma_2) + (T - k_{2,0})(\Sigma_1 - \Sigma_3)], \qquad (48) \\
\varphi_{k_1}^2 &= \frac{1}{T - k_1} vech[(k_{1,0} - k_1)(\Sigma_2 - \Sigma_1) + (T - k_{2,0})(\Sigma_2 - \Sigma_3)], \qquad (49) \\
\varphi_{k_1}^3 &= \frac{1}{T - k_1} vech[(k_{1,0} - k_1)(\Sigma_3 - \Sigma_1) + (k_{2,0} - k_{1,0})(\Sigma_3 - \Sigma_2)]. \qquad (50)
\end{aligned}
$$

Since $\frac{1-\tau_{2,0}}{1-\tau_{1,0}} \|vech(\Sigma_2 - \Sigma_3)\|^2 \leq \frac{\tau_{1,0}}{\tau_{2,0}} \|vech(\Sigma_1 - \Sigma_2)\|^2$, $(1-\tau_{2,0})^2 \|vech(\Sigma_2 - \Sigma_3)\|^2$ is smaller than $(1-\tau_{1,0})^2 \|vech(\Sigma_1 - \Sigma_2)\|^2$, and thus for $k_1 \in W_\eta^c$ and $k_1 < k_{1,0}$, $\Pi_1(k_1) - \Pi(k_{1,0}) \geq cT$ for some c. On the other hand, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \Psi_1(k_1) = o_p(T)$ and $\Psi(k_{1,0}) = o_p(T)$ due to the following:

1. $\|\varphi_{k_1}^1\|$, $\|\varphi_{k_1}^2\|$ and $\|\varphi_{k_1}^3\|$ are uniformly bounded for $k_1 \in W_\eta^c$ and $k_1 < k_{1,0}$.
2. Using Hajek-Renyi inequality, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\sum_{\iota=k_1+1}^{k_{1,0}} y_t\right\|$, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\sum_{\iota=k_1+1}^{T} y_t\right\|$, $\left\|\sum_{\iota=k_{1,0}+1}^{k_{2,0}} y_t\right\|$ and $\left\|\sum_{\iota=k_{2,0}+1}^{T} y_t\right\|$ are all $O_p(\sqrt{T})$, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\frac{1}{\sqrt{k_1}} \sum_{\iota=1}^{k_1} y_t\right\|$ is $O_p(\sqrt{\log T})$ and $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\frac{1}{\sqrt{T-k_1}} \sum_{\iota=k_1+1}^{T} y_t\right\|$ is $O_p(1)$.
3. Using Lemma 5, $\sup_{1 \leq k < l \leq T} \left\|\sum_{\iota=k+1}^{l} z_t\right\|$ and $\sup_{1 \leq k < l \leq T} \left\|\frac{1}{\sqrt{l-k}} \sum_{\iota=k+1}^{l} z_t\right\|^2$ are both $o_p(T)$.

For $k_{1,0}+1 < k_1 \le k_{2,0}$, after some calculation, we have:

$$\tilde{S}(k_1) - \tilde{S}(k_{1,0}) = \Pi_2(k_1) - \Pi(k_{1,0}) + \Psi_2(k_1) - \Psi(k_{1,0}), \tag{51}$$

where

$$\begin{aligned} & \Pi_2(k_1) - \Pi(k_{1,0}) \\ = \; & (k_1 - k_{1,0})[\frac{k_{1,0}}{k_1}\left\|vech(\Sigma_2 - \Sigma_1)\right\|^2 - \frac{(T-k_{2,0})^2}{(T-k_1)(T-k_{1,0})}\left\|vech(\Sigma_3 - \Sigma_2)\right\|^2] \\ \ge \; & (k_1 - k_{1,0})[\frac{k_{1,0}}{k_{2,0}}\left\|vech(\Sigma_2 - \Sigma_1)\right\|^2 - \frac{T-k_{2,0}}{T-k_{1,0}}\left\|vech(\Sigma_3 - \Sigma_2)\right\|^2], \end{aligned} \tag{52}$$

$$\begin{aligned} \Psi_2(k_1) = \; & 2\varphi_{k_1}^{4\prime}\sum_{\iota=1}^{k_{1,0}}(y_t+z_t) + 2\varphi_{k_1}^{5\prime}\sum_{\iota=k_{1,0}+1}^{k_1}(y_t+z_t) + 2\varphi_{k_1}^{6\prime}\sum_{\iota=k_1+1}^{k_{2,0}}(y_t+z_t) \\ & +2\varphi_{k_1}^{7\prime}\sum_{\iota=k_{2,0}+1}^{T}(y_t+z_t) - 2[k_{1,0}\varphi_{k_1}^4 + (k_1-k_{1,0})\varphi_{k_1}^5]'\frac{1}{k_1}\sum_{\iota=1}^{k_1}(y_t+z_t) \\ & -2[(k_{2,0}-k_1)\varphi_{k_1}^6 + (T-k_{2,0})\varphi_{k_1}^7]'\frac{1}{T-k_1}\sum_{\iota=k_1+1}^{T}(y_t+z_t) \\ & -\left\|\frac{1}{\sqrt{k_1}}\sum_{\iota=1}^{k_1}(y_t+z_t)\right\|^2 - \left\|\frac{1}{\sqrt{T-k_1}}\sum_{\iota=k_1+1}^{T}(y_t+z_t)\right\|^2, \end{aligned} \tag{53}$$

$$\varphi_{k_1}^4 = \frac{k_1-k_{1,0}}{k_1}vech(\Sigma_1-\Sigma_2),\;\; \varphi_{k_1}^5 = \frac{k_{1,0}}{k_1}vech(\Sigma_2-\Sigma_1), \tag{54}$$

$$\varphi_{k_1}^6 = \frac{T-k_{2,0}}{T-k_1}vech(\Sigma_2-\Sigma_3),\;\; \varphi_{k_1}^7 = \frac{k_{2,0}-k_1}{T-k_1}vech(\Sigma_3-\Sigma_2). \tag{55}$$

The term in the bracket is positive, thus for $k_1 \in W_\eta^c$ and $k_{1,0}+1 < k_1 \le k_{2,0}$, $\Pi_2(k_1) - \Pi_2(k_{1,0}) \ge Tc$ for some c. Using the same argument as in the previous case, it is easy to show $\sup_{k_1 \in W_\eta^c, k_{1,0}+1<k_1\le k_{2,0}} \Psi_2(k_1) = o_p(T)$.

For $k_{2,0} < k_1 \le T$, after some calculation, we have:

$$\tilde{S}(k_1) - \tilde{S}(k_{1,0}) = \Pi_3(k_1) - \Pi_3(k_{1,0}) + \Psi_3(k_1) - \Psi_3(k_{1,0}). \tag{56}$$

By symmetry, $\Pi_3(k_1) - \Pi_3(k_{2,0})$ has a similar expression as $\Pi_1(k_1) - \Pi_1(k_{1,0})$ and is positive. Thus $\Pi_3(k_1) - \Pi_3(k_{1,0}) \ge \Pi_3(k_{2,0}) - \Pi_3(k_{1,0}) = (k_{2,0} - k_{1,0})[\frac{k_{1,0}}{k_{2,0}}\left\|vech(\Sigma_2 - \Sigma_1)\right\|^2 -$

$\frac{T-k_{2,0}}{T-k_{1,0}}\left\|vech(\Sigma_3-\Sigma_2)\right\|^2]$.

$$\begin{aligned}\Psi_3(k_1) &= 2\varphi_{k_1}^{8\prime}\sum\nolimits_{\iota=1}^{k_{1,0}}(y_t+z_t)+2\varphi_{k_1}^{9\prime}\sum\nolimits_{\iota=k_{1,0}+1}^{k_{2,0}}(y_t+z_t)+2\varphi_{k_1}^{10\prime}\sum\nolimits_{\iota=k_{2,0}+1}^{k_1}(y_t+z_t)\\ &\quad -2[k_{1,0}\varphi_{k_1}^{8}+(k_{2,0}-k_{1,0})\varphi_{k_1}^{9}+(k_1-k_{2,0})\varphi_{k_1}^{10}]'\frac{1}{k_1}\sum\nolimits_{\iota=1}^{k_1}(y_t+z_t)\\ &\quad -\left\|\frac{1}{\sqrt{k_1}}\sum\nolimits_{\iota=1}^{k_1}(y_t+z_t)\right\|^2-\left\|\frac{1}{\sqrt{T-k_1}}\sum\nolimits_{\iota=k_1+1}^{T}(y_t+z_t)\right\|^2, \end{aligned} \tag{57}$$

and similarly $\sup\limits_{k_1\in W_\eta^c, k_1>k_{2,0}}\Psi_3(k_1)=o_p(T)$. ■

E PROOF OF THEOREM 2

Proof. Using similar argument as proving Theorem 1, it suffices to show for any $\epsilon>0$ and $\eta>0$, there exist $C>0$ such that $P(\min\limits_{k_1\in W_\eta, |k_1-k_{1,0}|>C}\frac{\tilde S(k_1)-\tilde S(k_{1,0})}{|k_1-k_{1,0}|}\le 0)<\epsilon$ as $(N,T)\to\infty$.

First consider the case $k_1<k_{1,0}$. Note that

$$\begin{aligned}\tilde S(k_1) &= \sum\nolimits_{t=1}^{k_1}[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1})]'[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1})]\\ &\quad +\sum\nolimits_{t=k_1+1}^{T}[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1}^*)]'[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1}^*)], \end{aligned} \tag{58}$$

where $\tilde\Sigma_{k_1}=\frac{1}{k_1}\sum_{t=1}^{k_1}\tilde g_t\tilde g_t'$ and $\tilde\Sigma_{k_1}^*=\frac{1}{T-k_1}\sum_{t=k_1+1}^{T}\tilde g_t\tilde g_t'$. Replacing $\tilde\Sigma_{k_{1,0}}$ by $\tilde\Sigma_{k_1}$ and $\tilde\Sigma_{k_{1,0}}^*$ by $\tilde\Sigma_{k_1}^*$ in the expression of $\tilde S(k_{1,0})$, $\tilde S(k_{1,0})$ is magnified. Thus

$$\begin{aligned}\frac{\tilde S(k_1)-\tilde S(k_{1,0})}{|k_1-k_{1,0}|} &\ge \frac{1}{|k_1-k_{1,0}|}\{\sum\nolimits_{t=k_1+1}^{k_{1,0}}[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1}^*)]'[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1}^*)]\\ &\quad -\sum\nolimits_{t=k_1+1}^{k_{1,0}}[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1})]'[vech(\tilde g_t\tilde g_t'-\tilde\Sigma_{k_1})]\}. \end{aligned} \tag{59}$$

The right hand side equals

$$\begin{aligned}&vech(\Sigma_1-\tilde\Sigma_{k_1}^*)'vech(\Sigma_1-\tilde\Sigma_{k_1}^*)-vech(\Sigma_1-\tilde\Sigma_{k_1})'vech(\Sigma_1-\tilde\Sigma_{k_1})\\ &+2vech(\Sigma_1-\tilde\Sigma_{k_1}^*)'\frac{\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t)}{k_{1,0}-k_1}-2vech(\Sigma_1-\tilde\Sigma_{k_1})'\frac{\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t)}{k_{1,0}-k_1}\\ =\ &\Xi_1-\Xi_2+\Xi_3-\Xi_4. \end{aligned} \tag{60}$$

Plug in $\tilde{\Sigma}_{k_1}$ and $\tilde{\Sigma}^*_{k_1}$, we have

$$\begin{aligned}
\Xi_1 &= \left\| vech[\frac{k_{2,0}-k_{1,0}}{T-k_1}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_1}(\Sigma_1-\Sigma_3)] \right\|^2 + \left\| \frac{\sum_{t=k_1+1}^{T}(y_t+z_t)}{T-k_1} \right\|^2 \\
&\quad -2vech[\frac{k_{2,0}-k_{1,0}}{T-k_1}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_1}(\Sigma_1-\Sigma_3)]'\frac{\sum_{t=k_1+1}^{T}(y_t+z_t)}{T-k_1} \\
&= \Xi_{11}+\Xi_{12}-\Xi_{13}, \qquad (61)\\
\Xi_2 &= \left\| \frac{1}{k_1}\sum_{t=1}^{k_1}(y_t+z_t) \right\|^2, \qquad (62)\\
\Xi_3 &= 2vech[\frac{k_{2,0}-k_{1,0}}{T-k_1}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_1}(\Sigma_1-\Sigma_3)]'\frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t) \\
&\quad -2[\frac{1}{T-k_1}\sum_{t=k_1+1}^{T}(y_t+z_t)]'\frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t), \qquad (63)\\
\Xi_4 &= -2[\frac{1}{k_1}\sum_{t=1}^{k_1}(y_t+z_t)]'\frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t). \qquad (64)
\end{aligned}$$

If $vech[\frac{k_{2,0}-k_{1,0}}{T-k_{1,0}}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_{1,0}}(\Sigma_1-\Sigma_3)]=0$, then $\Sigma_1-\Sigma_2=\frac{T-k_{2,0}}{T-k_{1,0}}(\Sigma_2-\Sigma_3)$, then $\frac{\tau_{1,0}}{\tau_{2,0}}\|vech(\Sigma_1-\Sigma_2)\|^2=\frac{\tau_{1,0}(1-\tau_{2,0})^2}{\tau_{2,0}(1-\tau_{1,0})^2}\|vech(\Sigma_2-\Sigma_3)\|^2<(\frac{1-\tau_{2,0}}{1-\tau_{1,0}})\|vech(\Sigma_2-\Sigma_3)\|^2$, this contradicts with Assumption 9. Thus $\Xi_{11}>c$ for some c. Using Hajek-Renyi inequality for y_t in each regime and Lemma 5 for z_t,

$$\begin{aligned}
&\sup_{k_1\in W_\eta,k_1<k_{1,0}-C}\left\|\frac{1}{T-k_1}\sum_{t=k_1+1}^{T}(y_t+z_t)\right\|,\\
&\sup_{k_1\in W_\eta,k_1<k_{1,0}-C}\left\|\frac{1}{k_1}\sum_{t=1}^{k_1}(y_t+z_t)\right\|,\\
&\sup_{k_1\in W_\eta,k_1<k_{1,0}-C}\left\|\frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}z_t\right\|,
\end{aligned}$$

are all $o_p(1)$ while $\sup_{k_1\in W_\eta,k_1<k_{1,0}-C}\left\|\frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}y_t\right\|$ is $O_p(\frac{1}{\sqrt{C}})$. Thus for sufficiently large C, all the other terms are dominated by Ξ_{11}.

Next consider the case $k_1>k_{1,0}$. Using the same argument as the case $k_1<k_{1,0}$,

$$\begin{aligned}
\frac{\tilde{S}(k_1)-\tilde{S}(k_{1,0})}{|k_1-k_{1,0}|} &\geq \frac{1}{|k_1-k_{1,0}|}\{\sum_{t=k_{1,0}+1}^{k_1}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{k_1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{k_1})]\\
&\quad -\sum_{t=k_{1,0}+1}^{k_1}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}^*_{k_1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}^*_{k_1})]\}, \qquad (65)
\end{aligned}$$

and the right hand side equals

$$
\begin{aligned}
& vech(\Sigma_2-\tilde{\Sigma}_{k_1})' vech(\Sigma_2-\tilde{\Sigma}_{k_1}) - vech(\Sigma_2-\tilde{\Sigma}^*_{k_1})' vech(\Sigma_2-\tilde{\Sigma}^*_{k_1}) \\
& +2vech(\Sigma_2-\tilde{\Sigma}_{k_1})' \frac{\sum_{t=k_{1,0}+1}^{k_1}(y_t+z_t)}{k_1-k_{1,0}} - 2vech(\Sigma_2-\tilde{\Sigma}^*_{k_1})' \frac{\sum_{t=k_{1,0}+1}^{k_1}(y_t+z_t)}{k_1-k_{1,0}} \\
= \; & \dot{\Xi}_1 - \dot{\Xi}_2 + \dot{\Xi}_3 - \dot{\Xi}_4. \qquad (66)
\end{aligned}
$$

Plug in $\tilde{\Sigma}_{k_1}$ and $\tilde{\Sigma}^*_{k_1}$, we have

$$
\begin{aligned}
\dot{\Xi}_1 &= \left\| \frac{k_{1,0}}{k_1} vech(\Sigma_2-\Sigma_1) \right\|^2 + \left\| \frac{1}{k_1} \sum_{t=1}^{k_1}(y_t+z_t) \right\|^2 \\
& \quad -2vech[\frac{k_{1,0}}{k_1}(\Sigma_2-\Sigma_1)]' \frac{1}{k_1} \sum_{t=1}^{k_1}(y_t+z_t) \\
&= \dot{\Xi}_{11} + \dot{\Xi}_{12} - \dot{\Xi}_{13}, \qquad (67) \\
\dot{\Xi}_2 &= \left\| \frac{T-k_{2,0}}{T-k_1} vech(\Sigma_2-\Sigma_3) \right\|^2 + \left\| \frac{1}{T-k_1} \sum_{t=k_1+1}^{T}(y_t+z_t) \right\|^2 \\
& \quad -2vech[\frac{T-k_{2,0}}{T-k_1}(\Sigma_2-\Sigma_3)]' \frac{1}{T-k_1} \sum_{t=k_1+1}^{T}(y_t+z_t) \\
&= \dot{\Xi}_{21} + \dot{\Xi}_{22} - \dot{\Xi}_{23}, \qquad (68) \\
\dot{\Xi}_3 &= 2vech[\frac{k_{1,0}}{k_1}(\Sigma_2-\Sigma_1)]' \frac{1}{k_1-k_{1,0}} \sum_{t=k_{1,0}+1}^{k_1}(y_t+z_t) \\
& \quad -2[\frac{1}{k_1} \sum_{t=1}^{k_1}(y_t+z_t)]' \frac{1}{k_1-k_{1,0}} \sum_{t=k_{1,0}+1}^{k_1}(y_t+z_t), \qquad (69) \\
\dot{\Xi}_4 &= 2vech[\frac{T-k_{2,0}}{T-k_1}(\Sigma_2-\Sigma_3)]' \frac{1}{k_1-k_{1,0}} \sum_{t=k_{1,0}+1}^{k_1}(y_t+z_t) \\
& \quad -2[\frac{1}{T-k_1} \sum_{t=k_1+1}^{T}(y_t+z_t)]' \frac{1}{k_1-k_{1,0}} \sum_{t=k_{1,0}+1}^{k_1}(y_t+z_t). \qquad (70)
\end{aligned}
$$

For $k_1 \in W_\eta$, $\dot{\Xi}_{11} - \dot{\Xi}_{21} \geq \left\| \frac{\tau_{1,0}}{\tau_{1,0}+\eta} vech(\Sigma_2-\Sigma_1) \right\|^2 - \left\| \frac{1-\tau_{2,0}}{1-\tau_{1,0}-\eta} vech(\Sigma_2-\Sigma_3) \right\|^2$. Thus by Assumption 9, $\dot{\Xi}_{11} - \dot{\Xi}_{21} \geq c$ for some $c > 0$ if η is sufficiently small. Again, using Hajek-Renyi inequality for y_t in each regime and Lemma 5 for z_t, all the other terms are dominated by $\dot{\Xi}_{11} - \dot{\Xi}_{21}$ for sufficiently large C. ■

F PROOF OF THEOREM 6

Proof. It is not difficult to see that

$$\begin{aligned} SSNE_0 &= \sum\nolimits_{t=1}^{T} vech(\tilde{f}_t\tilde{f}_t' - I_r)'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - I_r) \\ &\quad -Tvech(\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{f}_t\tilde{f}_t' - I_r)'\tilde{\Omega}^{-1}vech(\frac{1}{T}\sum\nolimits_{t=1}^{T}\tilde{f}_t\tilde{f}_t' - I_r), \end{aligned} \tag{71}$$

and for any partition $(k_1, ..., k_l)$,

$$\begin{aligned} SSNE(k_1, ..., k_l) &= \sum\nolimits_{t=1}^{T} vech(\tilde{f}_t\tilde{f}_t' - I_r)'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - I_r) - \sum\nolimits_{\iota=1}^{l+1}(k_\iota - k_{\iota-1}) \\ &\quad vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\tilde{f}_t\tilde{f}_t'}{k_\iota - k_{\iota-1}} - I_r)'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\tilde{f}_t\tilde{f}_t'}{k_\iota - k_{\iota-1}} - I_r). \end{aligned} \tag{72}$$

Let $F_{NT}^* = SSNE_0 - SSNE(k_1, ..., k_l)$, it follows that

$$\begin{aligned} F_{NT}^* &= \sum\nolimits_{\iota=1}^{l+1} vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{k_\iota - k_{\iota-1}}})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{k_\iota - k_{\iota-1}}}) \\ &\quad -vech(\frac{\sum_{t=1}^{T}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=1}^{T}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}}) \\ &= \sum\nolimits_{\iota=1}^{l+1} D(k_{\iota-1}+1, k_\iota) - D(1, T) \\ &= \sum\nolimits_{\iota=2}^{l+1}\{D(k_{\iota-1}+1, k_\iota) - [D(1, k_\iota) - D(1, k_{\iota-1})]\} \\ &= \sum\nolimits_{\iota=1}^{l} F_{NT}^*(\iota+1). \end{aligned} \tag{73}$$

After some algebra, we have

$$\begin{aligned} F_{NT}^*(\iota+1) &= \frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1} - k_\iota)} vech[\frac{k_{\iota+1}}{T}\frac{\sum_{t=1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}} - \frac{k_\iota}{T}\frac{\sum_{t=1}^{k_{\iota+1}}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}}]' \\ &\quad \tilde{\Omega}^{-1}vech[\frac{k_{\iota+1}}{T}\frac{\sum_{t=1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}} - \frac{k_\iota}{T}\frac{\sum_{t=1}^{k_{\iota+1}}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}}] \\ &= \frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1} - k_\iota)} B(\tau_\iota, \tau_{\iota+1}; \tilde{F})'\tilde{\Omega}^{-1}B(\tau_\iota, \tau_{\iota+1}; \tilde{F}). \end{aligned} \tag{74}$$

Next, using four facts listed below, we have

$$\begin{aligned}
&\sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon} \sum\nolimits_{\iota=1}^{l} \frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1}-k_\iota)} B(\tau_\iota,\tau_{\iota+1};\tilde{F})'(\tilde{\Omega}^{-1}-\Omega^{-1})B(\tau_\iota,\tau_{\iota+1};\tilde{F}) \\
\leq\ & \frac{1}{\epsilon^3}\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\| \sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon} \sum\nolimits_{\iota=1}^{l}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})\right\|^2 = o_p(1)O_p(1)=o_p(1), \qquad (75)
\end{aligned}$$

$$\begin{aligned}
&\sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon} \sum\nolimits_{\iota=1}^{l} \frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1}-k_\iota)} B(\tau_\iota,\tau_{\iota+1};\tilde{F})'\Omega^{-1}[B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0)] \\
\leq\ & \frac{l\left\|\Omega^{-1}\right\|}{\epsilon^3} \sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})\right\| \sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0)\right\| \\
=\ & O_p(1)o_p(1)=o_p(1). \qquad (76)
\end{aligned}$$

It follows that $F^*_{NT}=\sum_{\iota=1}^{l}\frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1}-k_\iota)}B(\tau_\iota,\tau_{\iota+1};FH_0)'\Omega^{-1}B(\tau_\iota,\tau_{\iota+1};FH_0)+o_p(1)$, where $o_p(1)$ is uniform and by Assumption 15 the first term converges weakly to

$$\sum\nolimits_{\iota=1}^{l}\frac{1}{\tau_\iota\tau_{\iota+1}(\tau_{\iota+1}-\tau_\iota)}\left\|\tau_\iota W_{\frac{r(r+1)}{2}}(\tau_{\iota+1})-\tau_{\iota+1}W_{\frac{r(r+1)}{2}}(\tau_\iota)\right\|^2.$$

1. $\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\|=o_p(1)$ if $\frac{d_T}{\delta_{NT}}\to 0$.
2. $\sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0)\right\|=o_p(1)$ if $\frac{\sqrt{T}}{N}\to 0$.
3. $\sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};FH_0)\right\|=O_p(1)$.
4. $\sup_{(\tau_1,\ldots,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})\right\|=O_p(1)$.

Fact (1) follows from Lemma 8.

Proof of (2): Note that

$$\begin{aligned}
& B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0) \\
=\ & B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH)+B(\tau_\iota,\tau_{\iota+1};FH)-B(\tau_\iota,\tau_{\iota+1};FH_0) \\
=\ & vech[\frac{k_{\iota+1}}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_\iota}(\tilde{f}_t\tilde{f}_t'-H'f_tf_t'H)-\frac{k_\iota}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_{\iota+1}}(\tilde{f}_t\tilde{f}_t'-H'f_tf_t'H)] \\
& +vech[\frac{k_{\iota+1}}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_\iota}(H'(f_tf_t'-\Sigma_F)H-H_0'(f_tf_t'-\Sigma_F)H_0) \\
& -\frac{k_\iota}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_{\iota+1}}(H'(f_tf_t'-\Sigma_F)H-H_0'(f_tf_t'-\Sigma_F)H_0)]. \qquad (77)
\end{aligned}$$

It is not difficult to see

$$\begin{aligned}&\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH)\right\|\\ \leq\ & 2\sup_{T\epsilon\leq k\leq T(1-\epsilon)}\left\|\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k}(\tilde{f}_t\tilde{f}_t'-H'f_tf_t'H)\right\|=O_p(\frac{\sqrt{T}}{\delta_{NT}^2})\end{aligned}\tag{78}$$

by Lemma 7, and

$$\begin{aligned}&\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\|B(\tau_\iota,\tau_{\iota+1};FH)-B(\tau_\iota,\tau_{\iota+1};FH_0)\|\\ \leq\ & 2\sup_{T\epsilon\leq k\leq T(1-\epsilon)}\left\|\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k}(H^{'}(f_tf_t'-\Sigma_F)H-H_0'(f_tf_t'-\Sigma_F)H_0)\right\|=o_p(1)\end{aligned}\tag{79}$$

by part (2) of Lemma 6 and Assumption 15.

Proof of (3): Note that

$$B(\tau_\iota,\tau_{\iota+1};FH_0)=vech[\frac{k_{\iota+1}}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_\iota}(H_0'f_tf_t'H_0-I_r)-\frac{k_\iota}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_{\iota+1}}(H_0'f_tf_t'H_0-I_r)],$$

it is not difficult to see

$$\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\|B(\tau_\iota,\tau_{\iota+1};FH_0)\|\leq 2\sup_{T\epsilon\leq k\leq T(1-\epsilon)}\left\|\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k}(H_0'f_tf_t'H_0-I_r)\right\|,$$

which is $O_p(1)$ by Assumption 15.

Proof of (4): It follows directly from (2) and (3). ■

G PROOF OF THEOREM 7

Proof. Under the alternative, the estimated number of factors converges to the number of pseudo factors and the estimated factors are pseudo factors, g_t. First note that $\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}[SSNE_0-SSNE(k_1,...,k_l)]\geq SSNE_0-SSNE(k_{1,0},...,k_{l0})$, thus it suffices to show the latter goes to infinity in probability.

$$\begin{aligned}
SSNE(k_{1,0},...,k_{l0}) &= \sum_{t=1}^{T} vech(\tilde{g}_t\tilde{g}_t')'\tilde{\Omega}^{-1}vech(\tilde{g}_t\tilde{g}_t') - \sum_{\iota=1}^{l+1}(k_{\iota 0} - \\
&\quad k_{\iota-1,0})vech(\frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota 0}} \tilde{g}_t\tilde{g}_t'}{k_{\iota 0}-k_{\iota-1,0}})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota 0}} \tilde{g}_t\tilde{g}_t'}{k_{\iota 0}-k_{\iota-1,0}}), \quad (80)\\
SSNE_0 &= \sum_{t=1}^{T} vech(\tilde{g}_t\tilde{g}_t')'\tilde{\Omega}^{-1}vech(\tilde{g}_t\tilde{g}_t') \\
&\quad -Tvech(\frac{\sum_{t=1}^{T}\tilde{g}_t\tilde{g}_t'}{T})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=1}^{T}\tilde{g}_t\tilde{g}_t'}{T}). \quad (81)
\end{aligned}$$

Thus similar to (74), $SSNE_0 - SSNE(k_{1,0},...,k_{l0})$ can be written as

$$\begin{aligned}
&\sum_{\iota=1}^{l} \frac{T^3}{k_{\iota 0}k_{\iota+1,0}(k_{\iota+1,0}-k_{\iota 0})} vech(\frac{k_{\iota+1,0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t' - \frac{k_{\iota 0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota+1,0}} \\
&\tilde{g}_t\tilde{g}_t')'\tilde{\Omega}^{-1}vech(\frac{k_{\iota+1,0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t' - \frac{k_{\iota 0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t')' \geq \frac{1}{\rho_{\max}(\tilde{\Omega})} \\
&\sum_{\iota=1}^{l} \frac{T^3}{k_{\iota 0}k_{\iota+1,0}(k_{\iota+1,0}-k_{\iota 0})} \left\| vech(\frac{k_{\iota+1,0}}{T}\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}} - \frac{k_{\iota 0}}{T}\frac{\sum_{t=1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}}) \right\|^2, \quad (82)
\end{aligned}$$

where $\rho_{\max}(\tilde{\Omega})$ is the maximal eigenvalue of $\tilde{\Omega}$. Note that $\frac{k_{\iota+1,0}}{T}\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}} - \frac{k_{\iota 0}}{T}\frac{\sum_{t=1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}} = \frac{(k_{\iota+1,0}-k_{\iota 0})k_{\iota 0}}{T^{\frac{3}{2}}}(\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}} - \frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota+1,0}-k_{\iota 0}})$, thus $SSNE_0 - SSNE(k_{1,0},...,k_{l0})$ is not smaller than

$$\sum_{\iota=1}^{l} \frac{(k_{\iota+1,0}-k_{\iota 0})k_{\iota 0}}{k_{\iota+1,0}\rho_{\max}(\tilde{\Omega})} \left\| vech(\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}} - \frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota+1,0}-k_{\iota 0}}) \right\|^2. \quad (83)$$

Recall that $vech(\tilde{g}_t\tilde{g}_t') = b_t + y_t + z_t$, by Assumption 1, for each ι,

$$\frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} y_t = vech(J_0'R_\iota \frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}(f_tf_t'-\Sigma_F)R_\iota'J_0) = o_p(1),$$

and by Lemma 5, $\frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} z_t = o_p(1)$ for each ι. Thus

$$\frac{1}{k_{\iota 0}}\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t' - \frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t' = \frac{1}{k_{\iota 0}}\sum_{t=1}^{k_{\iota 0}} b_t - \frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} b_t + o_p(1).$$

Recall that $b_t = vech(J_0'R_\iota\Sigma_F R_\iota J_0)$ for $k_{\iota-1,0} < t \leq k_{\iota,0}$ and b_t is different in different regime, thus $\frac{\sum_{t=1}^{k_{\iota 0}} b_t}{k_{\iota 0}} - \frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} b_t}{k_{\iota+1,0}-k_{\iota 0}} \neq 0$ for some ι. It follows that there exists some $c > 0$ such that $SSNE_0 - SSNE(k_{1,0},...,k_{l0}) \geq \frac{Tc}{\rho_{\max}(\tilde{\Omega})}$ with probability approaching one. Next, it is not difficult to see that under the alternative $\rho_{\max}(\tilde{\Omega}) = O_p(d_T)$, since HAC method is used to estimate $\tilde{\Omega}$ while under the alternative $\tilde{g}_t \tilde{g}_t'$ is not properly centered. Noting that $\frac{d_T}{T} \to 0$, the result is proved. ■

H PROOF OF THEOREM 8

Proof. It is easy to see that $F_{NT}(l+1\,|l) = \sup_{1\le\iota\le l+1} \sup_{k\in\Lambda_{\iota,\eta}} [SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota)]$, where $SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota)$ is the sum of squared normalized error of the ι-th regime. Thus testing l versus $l+1$ changes is essentially testing jointly 0 versus 1 change in each regime. In what follows, we reestablish Theorem 6 with $l = 1$ but $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$. Similar to (73), we have

$$\begin{aligned}
& SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota) \\
= \; & vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{k-\tilde{k}_{\iota-1}}})'\tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{k-\tilde{k}_{\iota-1}}}) \\
& + vech(\frac{\sum_{t=k+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - k}})'\tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=k+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - k}}) \\
& - vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}})'\tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}}),
\end{aligned} \tag{84}$$

and similar to (74),

$$\begin{aligned}
& SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota) \\
= \; & \frac{1}{\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\frac{\tilde{k}_\iota-k}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota-\tilde{k}_{\iota-1}}} - \frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota-\tilde{k}_{\iota-1}}})' \\
& \tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota-\tilde{k}_{\iota-1}}} - \frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - I_{r_\iota})}{\sqrt{\tilde{k}_\iota-\tilde{k}_{\iota-1}}}) \\
= \; & \frac{1}{\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\frac{\tilde{k}_\iota-k}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}} C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota)'\tilde{\Omega}_\iota^{-1} C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota).
\end{aligned} \tag{85}$$

Since $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$, asymptotically it suffices to consider the case that $\left|\tilde{k}_\iota - k_{\iota 0}\right| \le C$ for some integer C and all ι. And in such case $\Lambda_{\iota,\eta} \subset (k_{\iota-1,0}, k_{\iota 0}]$ for large T. Next, based on these two properties and using four facts listed below,

$$\sup_{k\in\Lambda_{\iota,\eta}}\left\|\left(\frac{1}{\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}\frac{\tilde{k}_{\iota}-k}{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}-\frac{1}{\frac{k-k_{\iota-1,0}}{k_{\iota 0}-k_{\iota-1,0}}\frac{k_{\iota 0}-k}{k_{\iota 0}-k_{\iota-1,0}}}\right)C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})'\tilde{\Omega}_{\iota}^{-1}C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})\right\|,$$

$$\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})'(\tilde{\Omega}_{\iota}^{-1}-\Omega_{\iota}^{-1})C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})\right\|,$$

$$\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})'\Omega_{\iota}^{-1}(C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0}))\right\|,$$

$$\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})'\Omega_{\iota}^{-1}(C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0}))\right\|$$ are all $o_p(1)$.

Thus

$$\begin{aligned}&\sup_{k\in\Lambda_{\iota,\eta}}[SSNE_{\iota}(\tilde{k}_{\iota-1},\tilde{k}_{\iota})-SSNE_{\iota}(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota})]\\=&\sup_{k\in\Lambda_{\iota,\eta}}\frac{1}{\frac{k-k_{\iota-1,0}}{k_{\iota 0}-k_{\iota-1,0}}\frac{k_{\iota 0}-k}{k_{\iota 0}-k_{\iota-1,0}}}C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})'\Omega_{\iota}^{-1}C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})+o_p(1)\\=&\sup_{k\in\Lambda_{\iota,\eta}}F_{NT,\iota}(k)+o_p(1).\end{aligned}\tag{86}$$

By Assumption 21, with $k=[T\tau]$, $F_{NT,\iota}(k)\Rightarrow\frac{1}{\tau(1-\tau)}\left\|W_{\frac{r_{\iota}(r_{\iota}+1)}{2}}(\tau)-\tau W_{\frac{r_{\iota}(r_{\iota}+1)}{2}}(1)\right\|^2$ for $\tau\in(0,1)$. Furthermore, since Wiener process has independent increments, the limit process of $F_{NT,\iota}(k)$ is independent with each other for different ι. Finally, define $\Lambda_{\iota,\eta}^0=\{k:k_{\iota-1,0}+(k_{\iota 0}-k_{\iota-1,0})\eta\le k\le k_{\iota 0}-(k_{\iota 0}-k_{\iota-1,0})\eta\}$. For any $\eta_1<\eta<\eta_2$, $\Lambda_{\iota,\eta_2}^0\subset\Lambda_{\iota,\eta}\subset\Lambda_{\iota,\eta_1}^0$ for large T, thus $\sup_{k\in\Lambda_{\iota,\eta_2}^0}F_{NT,\iota}(k)\le\sup_{k\in\Lambda_{\iota,\eta}}F_{NT,\iota}(k)\le\sup_{k\in\Lambda_{\iota,\eta_1}^0}F_{NT,\iota}(k)$. Since η_1 and η_2 can be arbitrarily close to η, $\sup_{k\in\Lambda_{\iota,\eta}}F_{NT,\iota}(k)$ has the same distribution as $\sup_{k\in\Lambda_{\iota,\eta}^0}F_{NT,\iota}(k)$. Taking together, we have the desired results.

1. $\left\|\tilde{\Omega}_{\iota}^{-1}-\Omega_{\iota}^{-1}\right\|=o_p(1)$ if $\frac{d_T}{T^{\frac{1}{4}}}\to 0$ and $\frac{d_T}{\sqrt{N}}\to 0$.
2. $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})\right\|=o_p(1)$ if $\frac{\sqrt{T}}{N}\to 0$.
3. $\sup_{k\in\Lambda_{\iota,\eta}}\|C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})\|=O_p(1)$.
4. $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})\right\|=O_p(1)$.

Fact (1) follows from Lemma 11.

Proof of (2): Note that

$$\begin{aligned}
& C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota0}) \\
= & [C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})] \\
& +[C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota0})] \\
= & vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}-\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}) \\
& +vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota0}'(f_tf_t'-\Sigma_F)H_{\iota0})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}} \\
& -\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota0}'(f_tf_t'-\Sigma_F)H_{\iota0})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}). \qquad (87)
\end{aligned}$$

Thus it's not difficult to see $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})\right\|$ is not larger than $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}\right\|+\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}\right\|$, which is $O_p(\frac{\sqrt{T}}{\delta_{NT}^2})$ by Lemma 10. And $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota0})\right\|$ is not larger than

$$\begin{aligned}
& \sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota0}'(f_tf_t'-\Sigma_F)H_{\iota0})}{\sqrt{k-\tilde{k}_{\iota-1}}}\right\| \\
& +\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota0}'(f_tf_t'-\Sigma_F)H_{\iota0})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}\right\|,
\end{aligned}$$

which is $o_p(1)$ by part (2) of Lemma 9 and Assumption 21. Finally, with $\left|\tilde{k}_{\iota}-k_{\iota0}\right|\leq C$ for all ι, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota0})-C(k_{\iota-1,0},k,k_{\iota0};F_{\iota0}H_{\iota0})\right\|=o_p(1)$ is obvious.

Proof of (3): Note that

$$\begin{aligned}
& C(k_{\iota-1,0},k,k_{\iota0};F_{\iota0}H_{\iota0}) \\
= & vech(\frac{\sum_{t=k_{\iota-1,0}+1}^{k}(H_{\iota0}'f_tf_t'H_{\iota0}-I_{r_{\iota}})}{\sqrt{k_{\iota0}-k_{\iota-1,0}}}-\frac{k-k_{\iota-1,0}}{k_{\iota0}-k_{\iota-1,0}}\frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota0}}(H_{\iota0}'f_tf_t'H_{\iota0}-I_{r_{\iota}})}{\sqrt{k_{\iota0}-k_{\iota-1,0}}}),
\end{aligned}$$

for some $\eta_1 < \eta$,

$$\sup_{k\in\Lambda_{\iota,\eta}} \|C(k_{\iota-1,0}, k, k_{\iota 0}; F_{\iota 0}H_{\iota 0})\| \leq \sup_{k\in\Lambda^0_{\iota,\eta_1}} \left\| \frac{\sum_{t=k_{\iota-1,0}+1}^{k}(H'_{\iota 0}f_t f'_t H_{\iota 0} - I_{r_\iota})}{\sqrt{k_{\iota 0} - k_{\iota-1,0}}} \right\| + \left\| \frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota 0}}(H'_{\iota 0}f_t f'_t H_{\iota 0} - I_{r_\iota})}{\sqrt{k_{\iota 0} - k_{\iota-1,0}}} \right\|,$$

which is $O_p(1)$ by Assumption 21.

Proof of (4): It follows directly from (2) and (3). ■

I PROOF OF THEOREM 9

Proof. The calculation of $SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota)$ under the null is still valid under the alternative. Thus following (85) we have

$$\begin{aligned}
& F_{NT}(l+1\,|\,l) \\
\geq\ & \sup_{k\in\Lambda_{\iota,\eta}} [SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota)] \\
\geq\ & SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k_{\iota 0}, \tilde{k}_\iota) \\
\geq\ & \frac{1}{\frac{k_{\iota 0}-\tilde{k}_{\iota-1}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\frac{\tilde{k}_\iota-k_{\iota 0}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}}\frac{1}{\rho_{\max}(\tilde{\Omega}_\iota)} \left\| vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota 0}}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}} - \frac{k_{\iota 0}-\tilde{k}_{\iota-1}}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}}) \right\|^2 \\
=\ & \frac{(k_{\iota 0}-\tilde{k}_{\iota-1})(\tilde{k}_\iota - k_{\iota 0})}{(\tilde{k}_\iota - \tilde{k}_{\iota-1})\rho_{\max}(\tilde{\Omega}_\iota)} \left\| vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota 0}}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{k_{\iota 0} - \tilde{k}_{\iota-1}} - \frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\tilde{k}_\iota - k_{\iota 0}}) \right\|^2. \qquad (88)
\end{aligned}$$

Define $z_{\iota t} = vech(\tilde{g}_{\iota t}\tilde{g}'_{\iota t} - J'_{\iota 0}g_{\iota t}g'_{\iota t}J_{\iota 0})$. By Lemma 13 and Assumption 1,

$$\begin{aligned}
& vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota 0}}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{k_{\iota 0} - \tilde{k}_{\iota-1}}) \\
=\ & \frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota 0}} z_{\iota t}}{k_{\iota 0} - \tilde{k}_{\iota-1}} + vech[J'_{\iota 0}A_{\iota 1}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota 0}}(f_t f'_t - \Sigma_F)}{k_{\iota 0} - \tilde{k}_{\iota-1}}A'_{\iota 1}J_{\iota 0}] + vech(J'_{\iota 0}A_{\iota 1}\Sigma_F A'_{\iota 1}J_{\iota 0}) \\
=\ & vech(J'_{\iota 0}A_{\iota 1}\Sigma_F A'_{\iota 1}J_{\iota 0}) + o_p(1), \qquad (89)
\end{aligned}$$

and similarly $vech(\frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\tilde{k}_\iota - k_{\iota 0}}) = vech(J'_{\iota 0}A_{\iota 2}\Sigma_F A'_{\iota 2}J_{\iota 0}) + o_p(1)$. Since $A_{\iota 1}\Sigma_F A'_{\iota 1} \neq A_{\iota 2}\Sigma_F A'_{\iota 2}$ and $\rho_{\max}(\tilde{\Omega}_\iota) = O_p(d_T)$, there exists some $c > 0$ such that $F_{NT}(l+1\,|\,l) \geq \frac{Tc}{d_T}$ with probability approaching one. ■

J PROOF OF LEMMAS

Lemma 1 *Under Assumption 7(1), Hajek-Renyi inequality applies to the process* $\{y_t, t = k_{\kappa-1,0}+1, ..., k_{\kappa,0}\}$ *and* $\{y_t, t = k_{\kappa,0}, ..., k_{\kappa-1,0}+1\}$*,* $\kappa = 1, ..., L+1$*.*

Proof. Note that $y_t = vech(J_0' R_\kappa (f_t f_t' - \Sigma_F) R_\kappa' J_0)$ for $k_{\kappa-1,0} < k \leq k_{\kappa,0}$.

Thus $P(\sup_{k_{\kappa-1,0}+m \leq k \leq k_{\kappa,0}} c_k \left\| \sum_{t=k_{\kappa-1,0}+1}^{k} y_t \right\| > M)$ is controlled by

$$P(\|J_0' R_\kappa\|^2 \sup_{k_{\kappa-1,0}+m \leq k \leq k_{\kappa,0}} c_k \left\| \sum_{t=k_{\kappa-1,0}+1}^{k} \epsilon_t \right\| > M),$$

which is not larger than $\frac{C}{M^2}(m c_{k_{\kappa-1,0}+m}^2 + \sum_{k=k_{\kappa-1,0}+m+1}^{k_{\kappa,0}} c_k^2)$ by Hajek-Renyi inequality for process $\{\epsilon_t, t = k_{\kappa-1,0}+1, ..., k_{\kappa,0}\}$. Other processes can be proved similarly. ■

Lemma 2 *In case factor loadings have structural changes, under Assumptions 1-6,* $\|J - J_0\| = o_p(1)$ *and* $\|V_{NT} - V\| = o_p(1)$*.*

Proof. The proof follows similar procedure as Proposition 1 in Bai (2003), with J, J_0 and g_t corresponding to H, H_0 and f_t respectively. To avoid repetition, we will only sketch the main steps. In Bai (2003), proof of Proposition 1 relies on $d_{NT} = o_p(1)$ and $V_{NT}^* \xrightarrow{p} V$ (Bai's notation). The former relies on Lemma A.1 and A.3(i)[18] while the latter relies on Lemma A.3(ii). Lemma A.1 relies on Theorem 1 of Bai and Ng (2002) and Lemma A.3(i). Lemma A.3(ii) relies on Lemma A.3(i) and Lemma 1(ii) of Bai and Ng (2002). Thus it suffices to prove Lemma 1(ii) and Theorem 1 of Bai and Ng (2002) and Lemma A.3(i) of Bai (2003). In current context, the first can be proved using Assumption 2 and Assumption 4 (2), the second can be proved using Assumptions 1-4, and the third can be proved using Assumption 5 and Weyl inequality. Finally, Assumption 6 ensures uniqueness of J_0. ■

Lemma 3 *Under Assumptions 1 and 7,*

1. $\sup_{0 \leq k < l \leq T} \frac{1}{\sqrt{T(l-k)}} \sum_{t=k+1}^{l} \|g_t\|^2 = O_p(1)$,

2. $\sup_{k_{\iota-1,0} < l \leq k_{\iota 0}} \frac{1}{l - k_{\iota-1,0}} \sum_{t=k_{\iota-1,0}+1}^{l} \|g_t\|^2 = O_p(1)$,

[18] In Bai (2003), Bai states that it relies on Lemma A.2, but in fact Lemma A.1 and A.3(i) is enough. This is because $d_{NT} = (\frac{\Lambda^{0\prime}\Lambda^0}{N})^{\frac{1}{2}} \frac{F^{0\prime}}{T}(\tilde{F} - F^0 H) V_{NT}$.

3. $\sup_{k_{\iota-1,0}\leq k<k_{\iota 0}} \frac{1}{k_{\iota 0}-k}\sum_{t=k+1}^{k_{\iota 0}}\|g_t\|^2 = O_p(1)$.

Proof. We first prove part (2). Recall that $g_t = R_\iota f_t$ for $k_{\iota-1,0} < t \leq k_{\iota 0}$, thus

$$\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}} \frac{\sum_{t=k_{\iota-1,0}+1}^{l}\|g_t\|^2}{l-k_{\iota-1,0}} \leq \|R_\iota\|^2\,\mathbb{E}\|f_t\|^2 + \|R_\iota\|^2 \sup_{k_{\iota-1,0}<l\leq k_{\iota 0}} \frac{\sum_{t=k_{\iota-1,0}+1}^{l}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)}{l-k_{\iota-1,0}},$$

where $\mathbb{E}\|f_t\|^2 = tr\Sigma_F$. It suffices to show the second term is $O_p(1)$. Let $D_l = \frac{\sum_{t=k_{\iota-1,0}+1}^{l}(f_tf_t'-\Sigma_F)}{l-k_{\iota-1,0}}$, it follows that $\left|\frac{\sum_{t=k_{\iota-1,0}+1}^{l}(\|f_t\|^2-\mathbb{E}\|f_t\|^2)}{l-k_{\iota-1,0}}\right| = |trD_l| \leq \sqrt{r_\iota}(trD_l^2)^{\frac{1}{2}} = \sqrt{r_\iota}\|D_l\|$, thus

$$\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}} \left|\frac{\sum_{t=k_{\iota-1,0}+1}^{l}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)}{l-k_{\iota-1,0}}\right| \leq \sqrt{r_\iota} \sup_{k_{\iota-1,0}<l\leq k_{\iota 0}} \left\|\frac{\sum_{t=k_{\iota-1,0}+1}^{l}\epsilon_t}{l-k_{\iota-1,0}}\right\|,$$

which is $O_p(1)$ by Hajek-Renyi inequality. Proof of part (3) is similar and omitted.

Now we prove part (1). The whole sample $t = 1, ..., T$ is divided into several nonoverlapping segments by the true change points. First consider the case that k and l lie in two different segments. Without loss of generality, suppose k lies in the ι-th segment and l lies in the κ-th segment, then $\sup_{k_{\iota-1,0}<k\leq k_{\iota 0};k_{\kappa-1,0}<l\leq k_{\kappa 0}} \frac{\sum_{t=k+1}^{l}\|g_t\|^2}{\sqrt{T(l-k)}}$ is no larger than $\sup_{k_{\iota-1,0}<k<k_{\iota 0}} \frac{\sum_{t=k+1}^{k_{\iota 0}}\|g_t\|^2}{k_{\iota 0}-k} + \frac{\sum_{t=k_{\iota 0}+1}^{k_{\kappa-1,0}}\|g_t\|^2}{k_{\kappa-1,0}-k_{\iota 0}} + \sup_{k_{\kappa-1,0}<l\leq k_{\kappa 0}} \frac{\sum_{t=k_{\kappa-1,0}+1}^{l}\|g_t\|^2}{l-k_{\iota-1,0}}$(If $\kappa - 1 = \iota$, the second term is zero). By parts (2) and (3), the first term and the third term are $O_p(1)$. The second term is no larger than $\sum_{\upsilon=\iota+1}^{\kappa-1}\|R_\upsilon\|^2(\frac{1}{k_{\upsilon,0}-k_{\upsilon-1,0}}\sum_{t=k_{\upsilon-1,0}+1}^{k_{\upsilon,0}}\|f_t\|^2)$, which is $O_p(1)$. Next consider the case that k and l lie in the same segment. Without loss of generality, suppose they lie in the ι-th segment, then $\sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}} \frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\|g_t\|^2$ is no larger than $\|R_\iota\|^2\,\mathbb{E}\|f_t\|^2 + \|R_\iota\|^2 \sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}} \left|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)\right|$. Similar to part (2), the second term is no larger than $\|R_\iota\|^2\sqrt{r_\iota} \sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}} \left\|\frac{\sum_{t=k+1}^{l}\epsilon_t}{\sqrt{T(l-k)}}\right\|$, which is $o_p(1)$ since by Assumption 7,

$$\begin{aligned} &\mathbb{E}(\sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}} \left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\epsilon_t\right\|^{4+\delta}) \\ = \;& \frac{1}{T^{2+\frac{\delta}{2}}}\sum_{k=k_{\kappa-1,0}}^{k_{\kappa,0}-1}\sum_{l=k+1}^{k_{\kappa,0}}\mathbb{E}(\left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l}\epsilon_t\right\|^{4+\delta}) \leq \frac{M}{T^{\frac{\delta}{2}}}. \end{aligned} \tag{90}$$

Up to now, we have proved the desired result for each possible case. Since the number of cases is finite, the supremum among all $0 \leq k < l \leq T$ will also be $O_p(1)$. ■

Lemma 4 *Under Assumptions 1-8,*

1. $\sup_{0\leq k<l\leq T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|=O_p(\frac{1}{\delta_{NT}^2}),$

2. $\sup_{0\leq k<l\leq T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\tilde{g}_t-J'g_t)g_t'J\right\|=O_p(\frac{1}{\delta_{NT}}),$

3. $\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}}\left\|\frac{1}{l-k_{\iota-1,0}}\sum_{t=k_{\iota-1,0}+1}^{l}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|=O_p(\frac{1}{\delta_{NT}^2})$ *for each* ι,

4. $\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}}\left\|\frac{1}{l-k_{\iota-1,0}}\sum_{t=k_{\iota-1,0}+1}^{l}(\tilde{g}_t-J'g_t)g_t'J\right\|=O_p(\frac{1}{\delta_{NT}})$ *for each* ι,

5. $\sup_{k_{\iota-1,0}\leq k<k_{\iota 0}}\left\|\frac{1}{k_{\iota 0}-k}\sum_{t=k+1}^{k_{\iota 0}}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|=O_p(\frac{1}{\delta_{NT}^2})$ *for each* ι,

6. $\sup_{k_{\iota-1,0}\leq k<k_{\iota 0}}\left\|\frac{1}{k_{\iota 0}-k}\sum_{t=k+1}^{k_{\iota 0}}(\tilde{g}_t-J'g_t)g_t'J\right\|=O_p(\frac{1}{\delta_{NT}})$ *for each* ι.

Proof. Following Bai (2003), we have

$$\tilde{g}_t-J'g_t=V_{NT}^{-1}(\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\gamma_N(s,t)+\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\zeta_{st}+\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\eta_{st}+\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\xi_{st}), \quad (91)$$

where $\zeta_{st}=\frac{e_s'e_t}{N}-\gamma_N(s,t)$, $\eta_{st}=\frac{g_s'\Gamma'e_t}{N}$and $\xi_{st}=\frac{g_t'\Gamma'e_s}{N}$. V_{NT} is the diagonal matrix of the first $\bar{r}$ largest eigenvalues of $\frac{1}{NT}XX'$ in decreasing order, $\tilde{G}$ is $\sqrt{T}$ times the corresponding eigenvector matrix, V is the diagonal matrix of eigenvalues of $\Sigma_\Gamma^{\frac{1}{2}}\Sigma_G\Sigma_\Gamma^{\frac{1}{2}}$ and Φ is the corresponding eigenvector matrix, $J=\frac{\Gamma'\Gamma}{N}\frac{G'\tilde{G}}{T}V_{NT}^{-1}$. First consider part (1).

$$\begin{aligned}&\sup_{0\leq k<l\leq T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|\\ \leq\ & 4\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\gamma_N(s,t)\right\|^2+\left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\zeta_{st}\right\|^2\\ &+\left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\eta_{st}\right\|^2+\left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\xi_{st}\right\|^2)\left\|V_{NT}^{-1}\right\|^2\\ =\ & 4\left\|V_{NT}^{-1}\right\|^2(I+II+III+IV).\end{aligned} \quad (92)$$

By part (1) of Lemma 2, $\left\|V_{NT}^{-1}\right\|\to\|V^{-1}\|$, thus it suffices to consider I, II, III and IV.

By Assumption 4,

$$\begin{aligned} I &\leq \frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s\|^2 \sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\frac{1}{T}\sum\nolimits_{s=1}^{T}\gamma_N(s,t)^2 \\ &\leq \bar{r}\frac{1}{T}\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}(\sum\nolimits_{s=1}^{T}M\left|\gamma_N(s,t)\right|) = O(\frac{1}{T}). \end{aligned} \tag{93}$$

By part (1) of Assumption 8,

$$\begin{aligned} II &\leq \frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s\|^2 \frac{1}{N}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}} \\ &\quad \sum\nolimits_{t=k+1}^{l}\left|\frac{\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]}{\sqrt{N}}\right|^2) \\ &= \bar{r}\frac{1}{N}O_p(1). \end{aligned} \tag{94}$$

By part (2) of Assumption 8,

$$\begin{aligned} III &\leq \frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s\|^2 \sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\frac{1}{T}\sum\nolimits_{s=1}^{T}\left|\frac{1}{N}\sum\nolimits_{i=1}^{N}g_s'\gamma_i e_{it}\right|^2 \\ &\leq \bar{r}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2)\frac{1}{N}\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{it}\right\|^2 \\ &= \bar{r}O_p(1)\frac{1}{N}O_p(1). \end{aligned} \tag{95}$$

By part (1) of Lemma 3 and part (ii) of Lemma 1 in Bai and Ng (2002),

$$\begin{aligned} IV &\leq \frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s\|^2 \sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l} \\ &\quad \|g_t\|^2\frac{1}{N}\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{is}\right\|^2 \\ &= \bar{r}O_p(1)\frac{1}{N}O_p(1). \end{aligned} \tag{96}$$

Next consider part (2).

$$
\begin{aligned}
& \sup_{0\le k<l\le T}\left\| \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} (\tilde{g}_t - J'g_t) g_t' J \right\| \\
\le \; & \left\|V_{NT}^{-1}\right\| \|J\| \sup_{0\le k<l\le T} \left\| \frac{1}{T} \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} \sum\nolimits_{s=1}^{T} \tilde{g}_s g_t' \gamma_N(s,t) \right\| \\
& + \left\|V_{NT}^{-1}\right\| \|J\| \sup_{0\le k<l\le T} \left\| \frac{1}{T} \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} \sum\nolimits_{s=1}^{T} \tilde{g}_s g_t' \zeta_{st} \right\| \\
& + \left\|V_{NT}^{-1}\right\| \|J\| \sup_{0\le k<l\le T} \left\| \frac{1}{T} \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} \sum\nolimits_{s=1}^{T} \tilde{g}_s g_t' \eta_{st} \right\| \\
& + \left\|V_{NT}^{-1}\right\| \|J\| \sup_{0\le k<l\le T} \left\| \frac{1}{T} \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} \sum\nolimits_{s=1}^{T} \tilde{g}_s g_t' \xi_{st} \right\| \\
= \; & \left\|V_{NT}^{-1}\right\| \|J\| (V + VI + VII + VIII). \qquad (97)
\end{aligned}
$$

By Lemma 2, $\left\|V_{NT}^{-1}\right\| \to \|V^{-1}\|$ and $\|J\| \to \|J_0\|$, thus it suffices to consider V, VI, VII and $VIII$. By part (1) of Lemma 3 and Assumption 4,

$$
\begin{aligned}
V \; \le \; & \left(\frac{1}{T} \sum\nolimits_{s=1}^{T} \|\tilde{g}_s\|^2\right)^{\frac{1}{2}} \sup_{0\le k<l\le T} \left(\frac{1}{T} \sum\nolimits_{s=1}^{T} \left\| \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} g_t' \gamma_N(s,t) \right\|^2 \right)^{\frac{1}{2}} \\
\le \; & \bar{r} \left(\sup_{0\le k<l\le T} \frac{\sum_{t=k+1}^{l} \|g_t\|^2}{\sqrt{T(l-k)}} \right)^{\frac{1}{2}} \left(\frac{1}{T} \sup_{0\le k<l\le T} \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} \sum\nolimits_{s=1}^{T} |\gamma_N(s,t)|^2 \right)^{\frac{1}{2}} \\
= \; & O_p(1) O_p\left(\frac{1}{\sqrt{T}}\right). \qquad (98)
\end{aligned}
$$

By part (1) of Lemma 3 and part (1) of Assumption 8,

$$
\begin{aligned}
VI \; \le \; & \left(\frac{1}{T} \sum\nolimits_{s=1}^{T} \|\tilde{g}_s\|^2\right)^{\frac{1}{2}} \sup_{0\le k<l\le T} \left(\frac{1}{T} \sum\nolimits_{s=1}^{T} \left\| \frac{\sum_{t=k+1}^{l} g_t'}{\sqrt{T(l-k)}} \frac{\sum_{i=1}^{N} [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]}{N} \right\|^2 \right)^{\frac{1}{2}} \\
\le \; & \bar{r} \frac{1}{\sqrt{N}} \left(\sup_{0\le k<l\le T} \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} \|g_t\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum\nolimits_{s=1}^{T} \right. \\
& \left. \sup_{0\le k<l\le T} \frac{1}{\sqrt{T(l-k)}} \sum\nolimits_{t=k+1}^{l} \left\| \frac{1}{\sqrt{N}} \sum\nolimits_{i=1}^{N} [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right\|^2 \right)^{\frac{1}{2}} \\
= \; & \frac{1}{\sqrt{N}} O_p(1) O_p(1). \qquad (99)
\end{aligned}
$$

By part (1) of Lemma 3 and part (2) of Assumption 8,

$$\begin{aligned} VII &\leq (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2)^{\frac{1}{2}} \sup_{0\leq k<l\leq T}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{\sum_{t=k+1}^{l}(\frac{1}{N}\sum_{i=1}^{N}g_s'\gamma_i e_{it})g_t'}{\sqrt{T(l-k)}}\right\|^2)^{\frac{1}{2}} \\ &\leq \bar{r}(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}}(\sup_{0\leq k<l\leq T}\left\|\frac{\sum_{t=k+1}^{l}\sum_{i=1}^{N}\gamma_i e_{it}g_t'}{\sqrt{NT(l-k)}}\right\|^2)^{\frac{1}{2}} \\ &\leq \bar{r}(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}}(\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l} \\ &\quad \|g_t\|^2)^{\frac{1}{2}}(\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{it}\right\|^2)^{\frac{1}{2}} \\ &= O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1) \end{aligned} \quad (100)$$

By part (1) of Lemma 3 and part (ii) of Lemma 1 in Bai and Ng (2002),

$$\begin{aligned} VIII &\leq (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2)^{\frac{1}{2}} \sup_{0\leq k<l\leq T}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{\sum_{t=k+1}^{l}g_t'(\frac{1}{N}\sum_{i=1}^{N}g_t'\gamma_i e_{is})}{\sqrt{T(l-k)}}\right\|^2)^{\frac{1}{2}} \\ &\leq \bar{r}(\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\|g_t\|^2)\frac{1}{\sqrt{N}}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2)^{\frac{1}{2}} \\ &= O_p(1)\frac{1}{\sqrt{N}}O_p(1). \end{aligned} \quad (101)$$

For the other parts, proof of parts (3) and (5) are similar to proof of part (1), proof of parts (4) and (6) are similar to proof of part (2). ■

Lemma 5 *Under Assumptions 1-8,*

1. $\sup_{0\leq k<l\leq T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}z_t\right\| = o_p(1)$,
2. $\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}}\left\|\frac{1}{l-k_{\iota-1,0}}\sum_{t=k_{\iota-1,0}+1}^{l}z_t\right\| = o_p(1)$ *for each* ι,
3. $\sup_{k_{\iota-1,0}\leq k<k_{\iota 0}}\left\|\frac{1}{k_{\iota 0}-k}\sum_{t=k+1}^{k_{\iota 0}}z_t\right\| = o_p(1)$ *for each* ι.

Proof. Recall that $z_t = vech[(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'] + vech[(\tilde{g}_t - J'g_t)g_t'J] + vech[J'g_t(\tilde{g}_t - J'g_t)'] + vech[(J'-J_0')g_t g_t'(J-J_0)] + vech[(J'-J_0')g_t g_t'J_0] + vech[J_0'g_t g_t'(J-J_0)]$. From Lemma

2 and part (1) of Lemma 3, we have

$$\begin{aligned} &\sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(J'-J_0')g_tg_t'(J-J_0)\right\| \\ \le\ & \|J-J_0\|^2\sup_{0\le k<l\le T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\|g_t\|^2=o_p(1)O_p(1)=o_p(1), \end{aligned} \tag{102}$$

and similarly $\sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(J'-J_0')g_tg_t'J_0\right\|=o_p(1)$. These together with parts (1) and (2) of Lemma 4 proves part (1). Part (2) can be proved similarly using Lemma 2, part (2) of Lemma 3 and parts (3) and (4) of Lemma 4. Part (3) can be proved similarly using Lemma 2, part (3) of Lemma 3 and parts (5) and (6) of Lemma 4. ■

Lemma 6 *In case factor loadings are stable, under Assumptions 3-5 and 10-12,* $\|H-H_0\|=o_p(1)$ *and* $\|U_{NT}-U\|=o_p(1)$.

Proof. The proof is similar to Lemma 2. ■

Lemma 7 *In case factor loadings are stable, under Assumptions 3-5 and 10-14,*

$$\sup_{T\epsilon\le k\le T(1-\epsilon)}\left\|\frac{1}{T}\sum_{t=1}^{k}(\tilde{f}_t\tilde{f}_t'-H'f_tf_t'H)\right\|=O_p(\frac{1}{\delta_{NT}^2}).$$

Proof. It suffices to show

$$\begin{aligned} \sup_{T\epsilon\le k\le T(1-\epsilon)}\left\|\frac{1}{T}\sum_{t=1}^{k}(\tilde{f}_t-H'f_t)(\tilde{f}_t-H'f_t)'\right\| &= O_p(\frac{1}{\delta_{NT}^2}), \\ \sup_{T\epsilon\le k\le T(1-\epsilon)}\left\|\frac{1}{T}\sum_{t=1}^{k}(\tilde{f}_t-H'f_t)f_t'H\right\| &= O_p(\frac{1}{\delta_{NT}^2}). \end{aligned}$$

The former is not larger than $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_t-H'f_t\right\|^2$, which is $O_p(\frac{1}{\delta_{NT}^2})$ by Lemma A.1 in Bai (2003). The latter is a refinement of part (2) of Lemma 4. For its proof, see Lemma 3 of Han and Inoue (2014), the required conditions (Assumptions 1-8(a) in Han and Inoue (2014)) can be verified. ■

Lemma 8 *In case factor loadings are stable, under Assumptions 3-5, 10-13 and 16, if* $\frac{d_T}{\delta_{NT}}\to 0$ *as* $(N,T)\to\infty$, $\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\|=o_p(1)$.

Proof. First note that $\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\|\le\left\|\tilde{\Omega}^{-1}\right\|\left\|\tilde{\Omega}-\Omega\right\|\|\Omega^{-1}\|$, $\|\Omega^{-1}\|$ is constant, $\left\|\tilde{\Omega}^{-1}\right\|\le\sqrt{\frac{r(r+1)}{2}}\frac{1}{\rho_{\min}(\tilde{\Omega})}$ and $\left|\rho_{\min}(\tilde{\Omega})-\rho_{\min}(\Omega)\right|\le\left\|\tilde{\Omega}-\Omega\right\|$. Thus it remains to show $\left\|\tilde{\Omega}-\Omega\right\|=$

$o_p(1)$. By Assumption 16, $\left\|\tilde{\Omega}(FH_0)-\Omega\right\| = o_p(1)$. By second half of Theorem 2 in Han and Inoue (2014), $\left\|\tilde{\Omega}-\tilde{\Omega}(FH_0)\right\| = o_p(1)$ if $\frac{d_T}{\delta_{NT}} \to 0$. The required conditions in Han and Inoue (2014) can be verified. ■

Lemma 9 *In case factor loadings have structural changes, under Assumptions 1-5 and 18, with* $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$*, we have* $\|H_\iota - H_{\iota 0}\| = o_p(1)$ *and* $\|U_{\iota NT} - U_\iota\| = o_p(1)$.

Proof. First, Assumption 18 ensures uniqueness of $H_{\iota 0}$. The proof of $\|H_\iota - H_{\iota 0}\| = o_p(1)$ follows the same procedure as Proposition 1 in Bai (2003) which, as explained in Lemma 2, relies on Lemma 1(ii), Theorem 1 of Bai and Ng (2002) and Lemma A.3(i) of Bai (2003). Thus it suffices to reestablish these three with $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ and $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$. The first can be proved without adjustment. The second is proved in Theorem 5. The third ($\|U_{\iota NT} - U_\iota\| = o_p(1)$) is proved in Theorem 4. ■

Lemma 10 *Under Assumptions 1-5, 19, 20 and 18, with* $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$, $\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - H_\iota' f_t f_t' H_\iota)}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\right\|$ *and* $\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - H_\iota' f_t f_t' H_\iota)}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\right\|$ *are both* $O_p(\frac{1}{\delta_{NT}^2})$.

Proof. We will only show the first half, proof of the second half is the same. It suffices to prove

$$
\begin{aligned}
\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}} \sum\nolimits_{t=\tilde{k}_{\iota-1}+1}^{k} (\tilde{f}_{\iota t} - H_\iota' f_t)(\tilde{f}_{\iota t} - H_\iota' f_t)'\right\| &= O_p(\frac{1}{\delta_{NT}^2}), \\
\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}} \sum\nolimits_{t=\tilde{k}_{\iota-1}+1}^{k} (\tilde{f}_{\iota t} - H_\iota' f_t) f_t' H_\iota\right\| &= O_p(\frac{1}{\delta_{NT}^2})
\end{aligned}
$$

with $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ and $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$. The former is less than $\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}\left\|\tilde{f}_{\iota t} - H_\iota' f_t\right\|^2}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}$, which is $O_p(\frac{1}{\delta_{NT}^2})$ by Theorem 5 and $\|U_{\iota NT} - U_\iota\| = o_p(1)$ in Lemma 9. To prove the latter, it suffices to show $\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{\sum_{t=k_{\iota-1}+1}^{k}(\tilde{f}_{\iota t} - H_\iota' f_t) f_t' H_\iota}{k_\iota - k_{\iota-1}}\right\| = O_p(\frac{1}{\delta_{NT}^2})$ for each $k_{\iota-1} \in [k_{\iota-1,0} - C, k_{\iota-1,0} + C]$ and $k_\iota \in [k_{\iota,0} - C, k_{\iota,0} + C]$, where C is some positive integer (see Baltagi et al. (2015b) for more details). For the case $k_{\iota-1} \in [k_{\iota-1,0}, k_{\iota-1,0} + C]$ and $k_\iota \in [k_{\iota,0} - C, k_{\iota,0}]$, Lemma 3 of Han and Inoue (2014) is applicable with T replaced by $k_\iota - k_{\iota-1}$. We next prove for the case $k_{\iota-1} \in [k_{\iota-1,0} - C, k_{\iota-1,0} - 1]$ and $k_\iota \in [k_{\iota,0} + 1, k_{\iota,0} + C]$. Proof of the other two cases are the same.

Note that in this case $x_{it} = f_t'\lambda_{0,\iota-1,i} + e_{it}$ for $t \in [k_{\iota-1}+1, k_{\iota-1,0}]$, $x_{it} = f_t'\lambda_{0,\iota,i} + e_{it}$ for $t \in [k_{\iota-1,0}+1, k_{\iota,0}]$ and $x_{it} = f_t'\lambda_{0,\iota+1,i} + e_{it}$ for $t \in [k_{\iota,0}+1, k_\iota]$. Define $w_{it} = f_t'(\lambda_{0,\iota-1,i} - \lambda_{0,\iota,i})$ for $t \in [k_{\iota-1}+1, k_{\iota-1,0}]$, $w_{it} = 0$ for $t \in [k_{\iota-1,0}+1, k_{\iota,0}]$ and $w_{it} = f_t'(\lambda_{0,\iota+1,i} - \lambda_{0,\iota,i})$ for $t \in [k_{\iota,0}+1, k_\iota]$, it follows that $x_{it} = f_t'\lambda_{0,\iota,i} + e_{it} + w_{it}$ for $t \in [k_{\iota-1}+1, k_\iota]$. Define $X_\iota = (x_{k_{\iota-1}+1}, ..., x_{k_\iota})'$, $w_t = (w_{1t}, ..., w_{Nt})'$, $W_\iota = (w_{k_{\iota-1}+1}, ..., w_{k_\iota})'$, $E_\iota = (e_{k_{\iota-1}+1}, ..., e_{k_\iota})'$ and recall $F_\iota = (f_{k_{\iota-1}+1}, ..., f_{k_\iota})'$, it follows that $X_\iota = F_\iota \Lambda_{0\iota}' + E_\iota + W_\iota$. Using the same decomposition as equation A.1 in Bai (2003), we have

$$\begin{aligned} \tilde{f}_{\iota t} - H_\iota' f_t &= U_{\iota NT}^{-1} \frac{1}{N(k_\iota - k_{\iota-1})} [\tilde{F}_\iota' F_\iota \Lambda_{0\iota}' e_t + \tilde{F}_\iota' E_\iota \Lambda_{0\iota} f_t + \tilde{F}_\iota' E_\iota e_t \\ &\quad + \tilde{F}_\iota' F_\iota \Lambda_{0\iota}' w_t + \tilde{F}_\iota' W_\iota \Lambda_{0\iota} f_t + \tilde{F}_\iota' W_\iota w_t + \tilde{F}_\iota' E_\iota w_t + \tilde{F}_\iota' W_\iota e_t] \\ &= U_{\iota NT}^{-1} (Q_{1,t}^\iota + Q_{2,t}^\iota + Q_{3,t}^\iota + Q_{4,t}^\iota + Q_{5,t}^\iota + Q_{6,t}^\iota + Q_{7,t}^\iota + Q_{8,t}^\iota). \end{aligned} \tag{103}$$

By Lemma 9, $\left\|U_{\iota NT}^{-1}\right\|$ and $\|H_\iota\|$ are both $O_p(1)$, thus it suffices to show for $m = 1, ..., 8$, $\sup_{k \in \Lambda_{\iota,\eta}} \left\| \frac{1}{k_\iota - k_{\iota-1}} \sum_{t=k_{\iota-1}+1}^{k} Q_{m,t}^\iota f_t' \right\| = O_p(\frac{1}{\delta_{NT}^2})$.

For $m = 1, 2, 3$, the proof is the same as Lemma 3 of Han and Inoue (2014) except that in current case we use $\frac{1}{k_\iota - k_{\iota-1}} \sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| \tilde{f}_{\iota t} - H_\iota' f_t \right\|^2 = O_p(\frac{1}{\delta_{NT}^2})$ and $\|H_\iota\| = O_p(1)$ for $k_{\iota-1} \in [k_{\iota-1,0} - C, k_{\iota-1,0} - 1]$ and $k_\iota \in [k_{\iota,0}+1, k_{\iota,0}+C]$. These two are proved as intermediate result in Theorem 5 and Lemma 9, respectively. For $m = 4$, $\sup_{k \in \Lambda_{\iota,\eta}} \left\| \frac{\sum_{t=k_{\iota-1}+1}^{k} Q_{4,t}^\iota f_t'}{k_\iota - k_{\iota-1}} \right\|$ is not larger than $\left\| \frac{\tilde{F}_\iota' F_\iota \Lambda_{0\iota}'}{N(k_\iota - k_{\iota-1})} \right\| (\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \|w_t f_t'\|}{k_\iota - k_{\iota-1}})$ and

$$\begin{aligned} \left\| \frac{\tilde{F}_\iota' F_\iota \Lambda_{0\iota}'}{N(k_\iota - k_{\iota-1})} \right\| &\leq (\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota} \left\| \tilde{f}_{\iota s} \right\|^2}{k_\iota - k_{\iota-1}})^{\frac{1}{2}} (\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota} \|f_s\|^2}{k_\iota - k_{\iota-1}})^{\frac{1}{2}} \frac{1}{\sqrt{N}} (\frac{\sum_{i=1}^N \|\lambda_{0,\iota,i}\|^2}{N})^{\frac{1}{2}} \\ &= O_p(\frac{1}{\sqrt{N}}), \end{aligned} \tag{104}$$

$$\begin{aligned} \frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \|w_t f_t'\|}{k_\iota - k_{\iota-1}} &\leq \frac{\sum_{t=k_{\iota-1}+1}^{k_{\iota-1,0}} \|f_t f_t'\|}{k_\iota - k_{\iota-1}} (\sum_{i=1}^N \|\lambda_{0,\iota-1,i} - \lambda_{0,\iota,i}\|^2)^{\frac{1}{2}} \\ &\quad + \frac{\sum_{t=k_{\iota 0}+1}^{k_\iota} \|f_t f_t'\|}{k_\iota - k_{\iota-1}} (\sum_{i=1}^N \|\lambda_{0,\iota+1,i} - \lambda_{0,\iota,i}\|^2)^{\frac{1}{2}} \\ &= O_p(\frac{\sqrt{N}}{T}). \end{aligned} \tag{105}$$

For $m=5$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k} Q_{5,t}^{\iota} f_t'}{k_\iota - k_{\iota-1}}\right\|$ is not larger than $\left\|\frac{\tilde{F}_\iota' W_\iota \Lambda_{0\iota}}{N(k_\iota - k_{\iota-1})}\right\| (\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|f_t f_t'\|}{k_\iota - k_{\iota-1}})$ and

$$\frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota}\|f_t f_t'\| = O_p(1), \tag{106}$$

$$\begin{aligned}
&\left\|\frac{\tilde{F}_\iota' W_\iota \Lambda_{0\iota}}{N(k_\iota - k_{\iota-1})}\right\| \\
\le\ & \left\|\frac{1}{k_\iota - k_{\iota-1}}\tilde{F}_\iota' W_\iota\right\| \frac{1}{\sqrt{N}}(\frac{\sum_{i=1}^{N}\|\lambda_{0,\iota,i}\|^2}{N})^{\frac{1}{2}} \\
\le\ & [\frac{\sum_{s=k_{\iota-1}+1}^{k_{\iota-1,0}}\left\|\tilde{f}_{\iota s} f_s'\right\|}{k_\iota - k_{\iota-1}}(\sum_{i=1}^{N}\|\lambda_{0,\iota-1,i}-\lambda_{0,\iota,i}\|^2)^{\frac{1}{2}} \\
&+\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota}\left\|\tilde{f}_{\iota s} f_s'\right\|}{k_\iota - k_{\iota-1}}(\sum_{i=1}^{N}\|\lambda_{0,\iota+1,i}-\lambda_{0,\iota,i}\|^2)^{\frac{1}{2}}]\frac{1}{\sqrt{N}}(\frac{\sum_{i=1}^{N}\|\lambda_{0,\iota,i}\|^2}{N})^{\frac{1}{2}} \\
=\ & O_p(\frac{1}{T}).
\end{aligned} \tag{107}$$

The last equality is due to $\left\|\tilde{f}_{\iota s} - H_\iota' f_s\right\| = o_p(1)$ for $k_{\iota-1}+1 \le s \le k_\iota$, which can be proved once Lemma A.2 in Bai (2003) is reestablished with $k_{\iota-1}\in[k_{\iota-1,0}-C, k_{\iota-1,0}-1]$ and $k_\iota \in [k_{\iota,0}+1, k_{\iota,0}+C]$. This is not difficult since in Bai (2003) Lemma A.2 is based on Lemma A.1 and Proposition 1, and as explained in the cases $m=1,2,3$, we have reestablished these two with $k_{\iota-1}\in[k_{\iota-1,0}-C, k_{\iota-1,0}-1]$ and $k_\iota\in[k_{\iota,0}+1, k_{\iota,0}+C]$. For $m=6$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k} Q_{6,t}^{\iota} f_t'\right\|$ is not larger than $\frac{1}{N}\left\|\frac{\tilde{F}_\iota' W_\iota}{k_\iota - k_{\iota-1}}\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|w_t f_t'\|}{k_\iota-k_{\iota-1}})$. The second and the third terms are both $O_p(\frac{\sqrt{N}}{T})$, as proved in $m=5$ and $m=4$ respectively. For $m=7$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k} Q_{7,t}^{\iota} f_t'}{k_\iota-k_{\iota-1}}\right\|$ is not larger than $\left\|\frac{1}{N(k_\iota-k_{\iota-1})}\tilde{F}_\iota' E_\iota\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|w_t f_t'\|}{k_\iota-k_{\iota-1}})$. The second term is $O_p(\frac{\sqrt{N}}{T})$, as proved in $m=4$. The first term is not larger than $\frac{1}{\sqrt{N}}(\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota}\|\tilde{f}_{\iota s}\|^2}{k_\iota-k_{\iota-1}})^{\frac{1}{2}}(\frac{\sum_{i=1}^{N}\sum_{s=k_{\iota-1}+1}^{k_\iota} e_{is}^2}{N(k_\iota-k_{\iota-1})})^{\frac{1}{2}}$, which is $O_p(\frac{1}{\sqrt{N}})$. For $m=8$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k} Q_{8,t}^{\iota} f_t'}{k_\iota-k_{\iota-1}}\right\|$ is not larger than $\frac{1}{N}\left\|\frac{1}{k_\iota-k_{\iota-1}}\tilde{F}_\iota' W_\iota\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|e_t f_t'\|}{k_\iota-k_{\iota-1}})$. The second term is $O_p(\frac{\sqrt{N}}{T})$, as proved in $m=5$. The third term is not larger than $(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|f_t\|^2}{k_\iota-k_{\iota-1}})^{\frac{1}{2}}(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\sum_{i=1}^{N} e_{it}^2}{k_\iota-k_{\iota-1}})^{\frac{1}{2}}$, which is $O_p(\sqrt{N})$. Thus $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k} Q_{m,t}^{\iota} f_t'\right\| = O_p(\frac{1}{T})$ for $m=4,...,8$. ■

Lemma 11 *Under Assumptions 1, 3-5, 19, 20, 17, 18 and 22, with* $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ *and*

$\left|\tilde{k}_{\iota-1}-k_{\iota-1,0}\right|=O_p(1)$, $\left\|\tilde{\Omega}_\iota^{-1}-\Omega_\iota^{-1}\right\|=o_p(1)$ *if* $\frac{d_T}{T^{\frac{1}{4}}}\to 0$ *and* $\frac{d_T}{\sqrt{N}}\to 0$ *as* $(N,T)\to\infty$.

Proof. Similar to Lemma 8, it suffices to show $\left\|\tilde{\Omega}_\iota-\tilde{\Omega}(F_\iota H_{\iota 0})\right\|=o_p(1)$, given $\left|\tilde{k}_\iota-k_{\iota 0}\right|=O_p(1)$, $\left|\tilde{k}_{\iota-1}-k_{\iota-1,0}\right|=O_p(1)$, $\frac{d_T}{T^{\frac{1}{4}}}\to 0$ and $\frac{d_T}{\sqrt{N}}\to 0$. This can be proved following the same procedure as Theorem 2 in Han and Inoue (2014). Here we present the adjustment. First, the notation should be replaced correspondingly, for example, in Han and Inoue (2014) the sample is $t=1,...,T$ while here the sample is $t=k_{\iota-1}+1,...,k_\iota$. Next, in Han and Inoue (2014) proof of Theorem 2 relies on their Lemma 7 and Lemma 8, which further relies on their Lemma 5 and Lemma 6 respectively. Once their Lemma 5 and Lemma 6 are reestablished given $\left|\tilde{k}_\iota-k_{\iota 0}\right|=O_p(1)$ and $\left|\tilde{k}_{\iota-1}-k_{\iota-1,0}\right|=O_p(1)$, the proof of Lemma 7, Lemma 8 and Theorem 2 need no adjustment.

We first reestablish parts (i) and (iii) of their Lemma 5. With $\frac{d_T}{T^{\frac{1}{4}}}\to 0$ and $\frac{d_T}{\sqrt{N}}\to 0$, they are enough. From equation (103), we have

$$\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|\tilde{f}_{\iota t}-H_\iota' f_t\right\|^4}{k_\iota-k_{\iota-1}}\leq 8^3\left\|U_{\iota NT}^{-1}\right\|^4\left(\sum_{m=1}^{8}\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|Q_{m,t}^\iota\right\|^4}{k_\iota-k_{\iota-1}}\right). \tag{108}$$

Lemma 5 in Han and Inoue (2014) shows that $\sum_{m=1}^{3}\frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|Q_{m,t}^\iota\right\|^4=O_p(\frac{1}{T})+O_p(\frac{1}{N^2})$, the proof need no adjustment. For $m=4,...,8$, it can be shown that $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|Q_{4,t}^\iota\right\|^4}{k_\iota-k_{\iota-1}}=O_p(\frac{1}{T})$, $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|Q_{5,t}^\iota\right\|^4}{k_\iota-k_{\iota-1}}=O_p(\frac{1}{T})$, $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|Q_{6,t}^\iota\right\|^4}{k_\iota-k_{\iota-1}}=O_p(\frac{1}{T^3})$, $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|Q_{7,t}^\iota\right\|^4}{k_\iota-k_{\iota-1}}=O_p(\frac{1}{T})$ and $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|Q_{8,t}^\iota\right\|^4}{k_\iota-k_{\iota-1}}=O_p(\frac{1}{T^2})$. The proof of Lemma 6 need no adjustment, but note that it utilized $\frac{1}{T}F'(\hat{F}-FH)V_{NT}=O_p(\frac{1}{\delta_{NT}^2})$. Its counterpart in current case is $\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}-H_\iota' f_t)f_t'}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\right\|=O_p(\frac{1}{\delta_{NT}^2})$, which is implicitly proved in Lemma 10. ■

Lemma 12 *Under Assumptions 1-5 and 23, with* $\left|\tilde{k}_\iota-k_{\iota+1,0}\right|=O_p(1)$ *and* $\left|\tilde{k}_{\iota-1}-k_{\iota-1,0}\right|=O_p(1)$*, we have* $\|J_\iota-J_{\iota 0}\|=o_p(1)$ *and* $\|V_{\iota NT}-V_\iota\|=o_p(1)$.

Proof. The proof is similar to Lemma 9. ■

Lemma 13 *Under Assumptions 1-5, 19, 20 and 23, with* $\left|\tilde{k}_\iota-k_{\iota+1,0}\right|=O_p(1)$ *and* $\left|\tilde{k}_{\iota-1}-k_{\iota-1,0}\right|=O_p(1)$, $\frac{1}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota 0}}z_{\iota t}=o_p(1)$ *and* $\frac{1}{\tilde{k}_\iota-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota}z_{\iota t}=o_p(1)$.

Proof. We will show the second half, the first half can be proved similarly. It suffices to show $\left\|\frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota}vech(\tilde{g}_{\iota t}\tilde{g}_{\iota t}'-J_\iota' g_{\iota t}g_{\iota t}'J_\iota)}{\tilde{k}_\iota-k_{\iota 0}}\right\|$ and $\left\|\frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota}vech(J_\iota' g_{\iota t}g_{\iota t}'J_\iota-J_{\iota 0}' g_{\iota t}g_{\iota t}'J_{\iota 0})}{\tilde{k}_\iota-k_{\iota 0}}\right\|$ are both $o_p(1)$.

The first term can be proved similarly as Lemma 10. The second term is not larger than $\left\| \frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota} g_{\iota t} g_{\iota t}'}{\tilde{k}_\iota - k_{\iota 0}} \right\| \|J_\iota - J_{\iota 0}\|^2 + 2 \left\| \frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota} g_{\iota t} g_{\iota t}'}{\tilde{k}_\iota - k_{\iota 0}} \right\| \|J_\iota - J_{\iota 0}\| \|J_{\iota 0}\|$, which is $o_p(1)$ by Lemma 12.
■

Biographical Data

Name of Author: Fa Wang

Date of Birth: March 14, 1988

Place of Birth: Suzhou, Jiangsu, China

Undergraduate Degree: BA in Economics, Xi'an Jiaotong University, China, 2010