University of Connecticut

Department of Economics Working Paper Series

Identification and Estimation of a Large Factor Model with Structural Instability

by

Badi H. Baltagi
Syracuse University

Chihwa Kao
University of Connecticut

Fa Wang
Shanghai University of Finance and Economics

Working Paper 2016-34
October 2016

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

IDENTIFICATION AND ESTIMATION OF A LARGE FACTOR MODEL WITH STRUCTURAL INSTABILITY*

Badi H. Baltagi[†] Chihwa Kao[‡] Fa Wang[§]

October 12, 2016

Abstract

This paper tackles the identification and estimation of a high dimensional factor model with unknown number of latent factors and a single break in the number of factors and/or factor loadings occurring at unknown common date. First, we propose a least squares estimator of the change point based on the second moments of estimated pseudo factors and show that the estimation error of the proposed estimator is $O_p(1)$. We also show that the proposed estimator has some 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. The finite sample performance of our estimators is investigated using Monte Carlo experiments.

Keywords: high dimensional factor model, structural change, rate of convergence, number of factors, model selection, factor space, panel data

JEL Classification: C13; C33.

*We would like to thank the editor Jianqing Fan, the associate editor and three referees for their valuable comments and suggestions. We would also like to thank Jushan Bai, Xu Cheng, Yoonseok Lee, Lorenzo Trapani, Giovanni Urga and participants of 2015 NY Camp Econometrics, 2015 International Panel Data Conference and the 2015 World Congress of the Econometric Society for helpful comments and suggestions.

[†]Department of Economics and Center for Policy Research, Syracuse University, Syracuse, New York, U.S.A. Also, Department of Economics, Leicester University, Leicester LE176EE, UK.

[‡]Department of Economics, University of Connecticut

[§]School of Economics, Shanghai University of Finance and Economics

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. 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 the 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 the 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,

[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.

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 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 that 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:

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

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:

$$X = \begin{bmatrix} F_1^0 \Lambda_0' + F_1^1 \Lambda_1' \\ F_2^0 \Lambda_0' + F_2^1 \Lambda_2' \end{bmatrix} + E, \tag{2}$$

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 $0 \leq q_1 \leq q$, 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:

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

Next, define $G=(g_1,...,g_T)'=\left[\begin{array}{c}\left[\begin{array}{cc}F_1^0 & F_1^1\end{array}\right]A' \\ \left[\begin{array}{cc}F_2^0 & F_2^1\end{array}\right]B'\end{array}\right]$ and $\Gamma=\left[\begin{array}{ccc}\Lambda_0 & \Lambda_1 & \Lambda_{21}\end{array}\right]$, then

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

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 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 $t\leq k_0$ and $\Sigma_{G,2}=\mathbb{E}(g_t g_t')$ for $t>k_0$, then

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

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 $\left[\begin{array}{cc}\Lambda_{11} & \Lambda_{12}\end{array}\right]$, where $\left[\begin{array}{ccc}\Lambda_0 & \Lambda_2 & \Lambda_{11}\end{array}\right]$ is of full rank and $\Lambda_{12}=\left[\begin{array}{ccc}\Lambda_0 & \Lambda_2 & \Lambda_{11}\end{array}\right]Z$ for some Z. Define $\Theta=\left[\begin{array}{ccc}\Lambda_0 & \Lambda_2 & \Lambda_{11}\end{array}\right]$.

Our assumptions are as follows:

Assumption 1 *(1) $\mathbb{E}\|f_t\|^4<M<\infty$, $\mathbb{E}(f_t f_t')=\Sigma_F$, 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:*

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

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* $\left|\tau_{ij,t}\right| \leq \tau_{ij}$ *for some* τ_{ij} *and for all* $t = 1, ..., T$, *and* $\sum_{j=1}^{N}\left|\tau_{ji}\right| \leq M$ *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* $\{\epsilon_t, t = T, ..., k_0 + 1\}$.

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

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

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

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

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

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

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

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. 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

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

[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: $\tilde{k}_{ML} = \arg\max[-k \log \left|\tilde{\Sigma}_1\right| - (T-k) \log \left|\tilde{\Sigma}_2\right|]$.

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 $(\hat{g}_1, ..., \hat{g}_T)' = \hat{G} = \tilde{G}V_{NT} = (\tilde{g}_1, ..., \tilde{g}_T)'V_{NT}$ 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 show in the appendix that $\tilde{\tau}$ is consistent for τ_0. This implies 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 stategy, 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 *Theorem 1 holds 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 also appears in other recent papers in the literature, for example, Chen et al. (2014), Han and Inoue (2015) and Cheng et al. (2015).*[5]

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

[5] A main drawback of this approach is that the proposed procedure is not able to distinguish breaks in the factor loadings (an important issue in empirical practice) and breaks in the factor variance (an issue of minor importance since the estimator remains consistent in this case). We would like to thank an anonymous referee for emphasizing this fact.

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 and misspecified models 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, Theorem 1 still holds.*

In case $\tilde{r}$ is fixed at some positive integer $m > r + q_1$, we can not prove the robustness of 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 $\frac{\sqrt{T}}{N} \to 0$, Theorem 1 still holds.*

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[6] 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.

[6] 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.

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 $\frac{1}{N}\left\|\Delta\right\|^2 = O(\frac{1}{\delta_{NT}^2})$, 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 $\frac{1}{N}\left\|\Delta\right\|^2 = O(\frac{1}{\delta_{NT}^c})$ for some $c > 0$.

Proposition 2 *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* $\frac{1}{N}\left\|\Delta\right\|^2 = O(\frac{1}{\delta_{NT}^c})$ *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 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 $\frac{1}{N}\left\|\Delta\right\|^2$ goes to zero, so that $\frac{g(N,T)}{\frac{1}{N}\left\|\Delta\right\|^2} \to \infty$. 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 $\frac{1}{N}\left\|\Delta\right\|^2 = O(1)$, nor the case where the number of factors also change. In view of these caveats, Proposition 2 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 $\tilde{k}$.[7] 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 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.

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

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

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

and 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, the 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. Therefore, 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,*

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

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*

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

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 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:

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

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.[8] 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

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

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.

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[9] because a portion of their elements are zeros and the magnitude

[9] 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

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[10]. 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 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 $(N, T) =$

the equivalent model.

[10]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.

$(100, 100)$ 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 $\tilde{r}_1$,

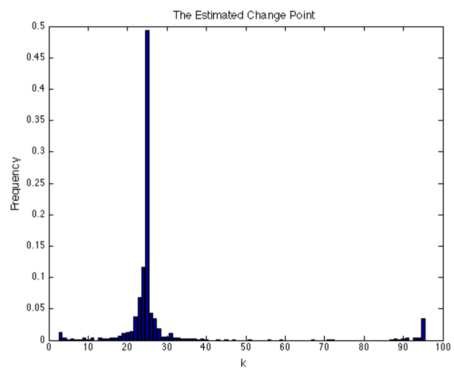
$\tilde{r}_2$ and averages of $R^2_{\hat{F},F}$ over 1,000 replications. x/y denotes that the frequency of underestimation and overestimation is $x\%$ and $y\%$ respectively. On the whole, the performance of IC_{p2} and GR are similar. If we choose the better one in each case, the performance of $\tilde{r}_1$ and $\tilde{r}_2$ behave quite well and in most cases close to 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 the 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 the change point among all estimators of the number of factors in the literature.

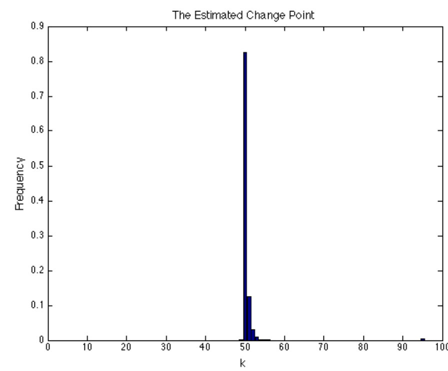
9 CONCLUSIONS

This paper studied the identification and estimation of a large dimensional factor model with a single large structural change. Both the factor loadings and the 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

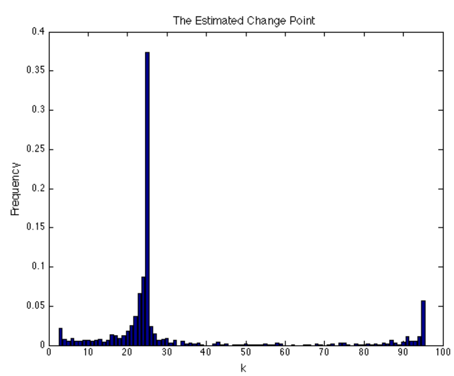
Figure 1: Histogram of $\tilde{k}$ for $(N, T) = (100, 100), (r_1, r_2, r + q_1) = (3, 5, 7)$



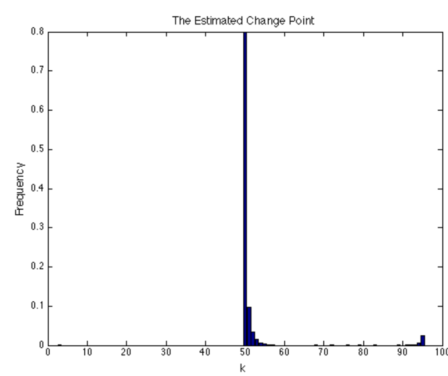
$(\rho, \alpha, \beta) = (0, 0, 0)$, homogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 5.68$, $sd(\tilde{r}) = 0.60$



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



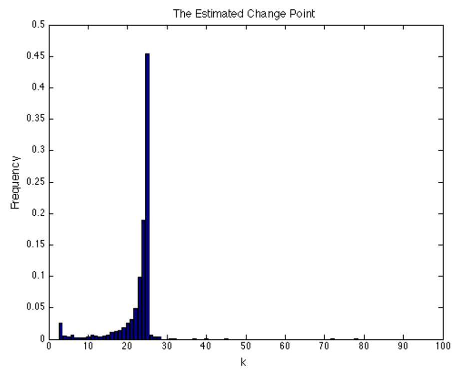
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 5.75$, $sd(\tilde{r}) = 0.58$



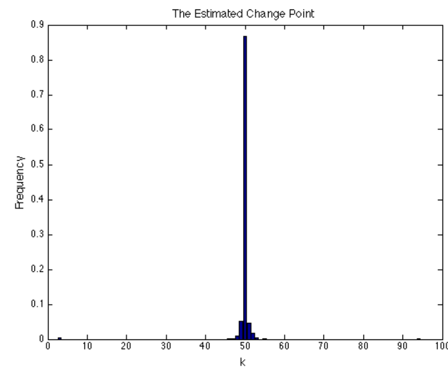
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous R^2, $\tau_0 = 0.5$, $ave(\tilde{r}) = 6.74$, $sd(\tilde{r}) = 0.48$

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

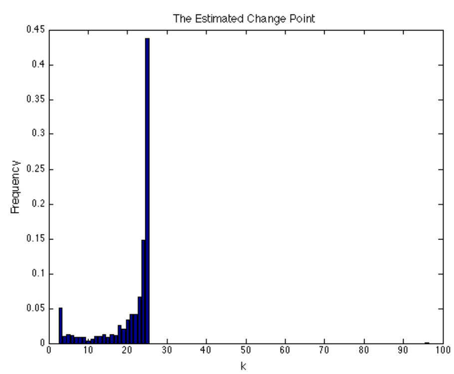
Figure 2: Histogram of $\tilde{k}$ for $(N,T) = (100,100), (r_1, r_2, r+q_1) = (3,3,5), a = 1$



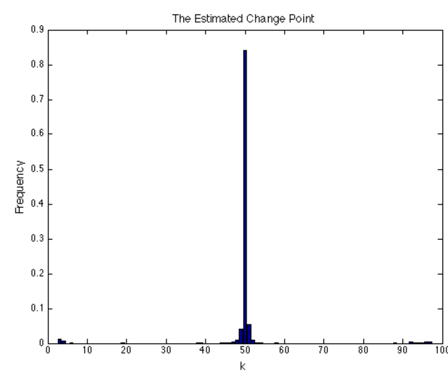
$(\rho, \alpha, \beta) = (0,0,0)$, homogeneous R^2,
$\tau_0 = 0.25$, $ave(\tilde{r}) = 4.51$, $sd(\tilde{r}) = 0.56$



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



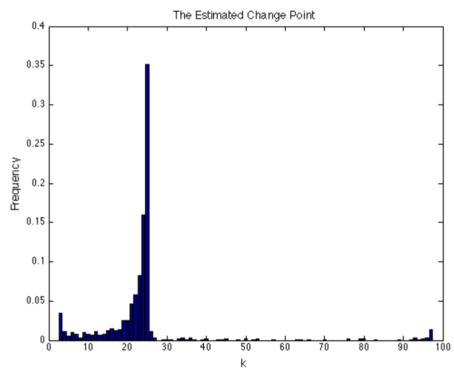
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous
R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 4.86$, $sd(\tilde{r}) = 0.35$



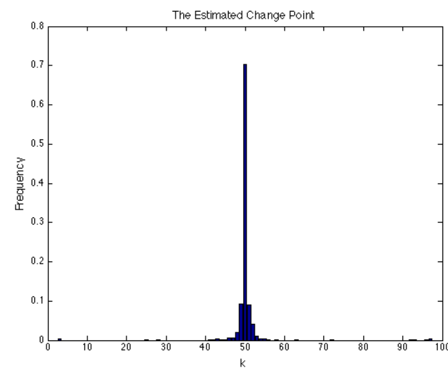
$(\rho, \alpha, \beta) = (0.5, 0.2, 0.2)$, heterogeneous
R^2, $\tau_0 = 0.5$, $ave(\tilde{r}) = 5.00$, $sd(\tilde{r}) = 0$

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

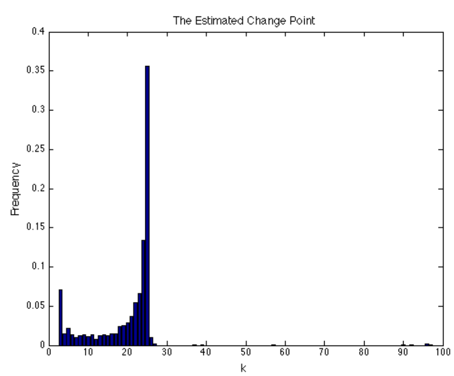
Figure 3: Histogram of $\tilde{k}$ for $(N,T) = (100,100), (r_1, r_2, r+q_1) = (3,3,5), a = 0.2$



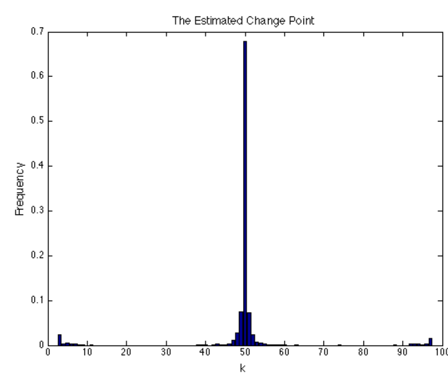
$(\rho,\alpha,\beta) = (0,0,0)$, homogeneous R^2, $\tau_0 = 0.25$, $ave(\tilde{r}) = 4.27$, $sd(\tilde{r}) = 0.60$



$\tau_0 = 0.5$, $(\rho,\alpha,\beta) = (0,0,0)$, 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 the factor AR(1) coefficient, the error term AR(1) coefficient and the error term cross-sectional correlation respectively. $ave(\tilde{r})$ and $sd(\tilde{r})$ denote the average and the standard deviation of the 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 the factor AR(1) coefficient, the error term AR(1) coefficient and the 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 the factor AR(1) coefficient, the error term AR(1) coefficient and the 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 the factor AR(1) coefficient, the error term AR(1) coefficient, the error term cross-sectional correlation, and the break magnitude respectively.

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)?

References

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

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

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

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

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

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

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

[8] 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.

[9] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

[23] 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.

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

[25] 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.

IDENTIFICATION AND ESTIMATION OF A LARGE FACTOR MODEL WITH STRUCTURAL INSTABILITY
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,

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

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,

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

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

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

hence when $m = 1$, $\sup_{1\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| = O_p(\sqrt{\log T})$ 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 $\tau \in (0, 1)$, $\sup_{m\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| = O_p(1)$ 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 $k \leq k_0$. 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 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

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

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

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

It follows that

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

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

For $k \leq k_0$,

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

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

Define

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

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

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

It follows that

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

and for $k < k_0$,

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

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

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 CONSISTENCY OF $\tilde{\tau}$

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 $D = \{k : (\tau_0-\eta)T \leq k \leq (\tau_0+\eta)T\}$ 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 $\tilde{S}(\tilde{k})-\tilde{S}(k_0)\leq 0$. If $\tilde{k}\in D^c$, then $\min_{k\in D^c}\tilde{S}(k)-\tilde{S}(k_0)\leq 0$. This implies $P(\tilde{k}\in D^c)\leq P(\min_{k\in D^c}\tilde{S}(k)-\tilde{S}(k_0)\leq 0)$, hence it suffices to show for any given $\epsilon > 0$ and $\eta > 0$, $P(\min_{k\in D^c}\tilde{S}(k) - \tilde{S}(k_0) \leq 0) < \epsilon$ as $(N,T) \to \infty$.

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

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,

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

For C_1^*,

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

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^*,

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

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^*,

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

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,

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

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

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

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, $\sup_{k\in D^c,k<k_0}\left|\frac{E^*}{k_0-k}\right| = o_p(1)$.

For term F^*,

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

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

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

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

The first term is bounded by

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

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 consistency of $\tilde{\tau}$, 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 consistency of $\tilde{\tau}$, 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 consistency of $\tilde{\tau}$,

$$\sup_{k \in D_M, k<k_0} \left|\frac{C^*}{k_0 - k}\right| \leq \sup_{k \in D, k<k_0} \left|\frac{C^*}{k_0 - k}\right| \leq \sup_{k \in D, k<k_0} C_1^* + \sup_{k \in D, k<k_0} C_2^* + \sup_{k \in D, k<k_0} C_3^*.$$

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^*,

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

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^*,

$$\sup_{k\in D,k<k_0} C_3^* \le 2\frac{1}{k_0}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\left\|\sum_{t=k+1}^{k_0} y_t\right\|^2 + 2\frac{1}{k_0}\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\left\|\sum_{t=k+1}^{k_0} z_t\right\|^2.$$

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,

$$\sup_{k\in D_M,k<k_0}\left|\frac{D^*}{k_0-k}\right| \le \sup_{k\in D,k<k_0}\left|\frac{D^*}{k_0-k}\right| \le \sup_{k\in D,k<k_0} D_1^* + \sup_{k\in D,k<k_0} D_2^* + \sup_{k\in D,k<k_0} D_3^*.$$

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

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

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

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^*.

$$
\begin{aligned}
\sup_{k\in D_M,k<k_0}\left|\frac{E^*}{k_0-k}\right| \leq\ & 2\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D_M,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} y_t\right\| \\
& +2\left\|\Sigma_2-\Sigma_1\right\|\sup_{k\in D_M,k<k_0}\left\|\frac{1}{k_0-k}\sum\nolimits_{t=k+1}^{k_0} z_t\right\|.
\end{aligned}
$$

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^*,

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

By part (1) of Lemma 3, the first term is $O_p(\frac{1}{\sqrt{T}})$. By part (9) of Lemma 7, 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^*,

$$\sup_{k\in D_M,k<k_0}\left|\frac{G^*}{k_0-k}\right| \leq \sup_{k\in D,k<k_0}\left|\frac{G^*}{k_0-k}\right| \leq 4\left\|\Sigma_2-\Sigma_1\right\| \sup_{k\in D,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k+1}^{T} y_t\right\|$$
$$+4\left\|\Sigma_2-\Sigma_1\right\| \sup_{k\in D,k<k_0}\left\|\frac{1}{T-k}\sum\nolimits_{t=k+1}^{T} z_t\right\|.$$

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, $\sup\limits_{k\in D_M,k<k_0}\left|\frac{G^*}{k_0-k}\right| = o_p(1)$. ■

F PROOF OF COROLLARY 1

Proof. The proof is the same as the proof of 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

$$\begin{aligned}\hat{S}^m(k) &= \sum\nolimits_{t=1}^{k}[vec(\hat{g}_t^m\hat{g}_t^{m\prime}-\hat{\Sigma}_1^m)]'[vec(\hat{g}_t^m\hat{g}_t^{m\prime}-\hat{\Sigma}_1^m)] \\ &\quad+\sum\nolimits_{t=k+1}^{T}[vec(\hat{g}_t^m\hat{g}_t^{m\prime}-\hat{\Sigma}_2^m)]'[vec(\hat{g}_t^m\hat{g}_t^{m\prime}-\hat{\Sigma}_2^m)], \qquad \text{(A-12)}\end{aligned}$$

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 consistency of $\tilde{\tau}$ and Theorem 1, it suffices to show $\sup\limits_{k\neq k_0}\left|\frac{(\hat{S}^m(k)-\hat{S}^m(k_0)-(\hat{S}(k)-\hat{S}(k_0))}{k-k_0}\right|=o_p(1)$. Consider $\sup\limits_{k\neq k_0}\left|\frac{L_{ij}}{k-k_0}\right|$ 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 $\sup\limits_{k\neq k_0}\left|\frac{L_{ij1}}{k-k_0}\right|\leq\frac{1}{T\tau_0}\sum_{t=1}^{T}\hat{g}_{it}^{m2}\sum_{t=1}^{T}\hat{g}_{jt}^{m2}=O_p(\frac{T}{\delta_{NT}^4})$. Similarly, $\sup\limits_{k\neq k_0}\left|\frac{L_{ij2}}{k-k_0}\right|$, $\sup\limits_{k\neq k_0}\left|\frac{L_{ij3}}{k-k_0}\right|$ and $\sup\limits_{k\neq k_0}\left|\frac{L_{ij4}}{k-k_0}\right|$ 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

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

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

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

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 2

Proof. The proof is similar to Theorem 2.

$$\begin{aligned}\beta_1^T &\leq tr(\frac{1}{NT}W(T)W'(T)) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=k_0+1}^{T} w_{it}^2 \\ &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=k_0+1}^{T} \|f_t\|^2 \|\lambda_{02,i} - \lambda_{01,i}\|^2 \\ &= (\frac{1}{T}\sum_{t=k_0+1}^{T}\|f_t\|^2)(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_{02,i}-\lambda_{01,i}\|^2) = O_p(\frac{1}{\delta_{NT}^c}). \qquad \text{(A-15)}\end{aligned}$$

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 $(\sqrt{\omega_j^T} - \sqrt{\alpha_j^T})^2 \leq \beta_1^T = O_p(\frac{1}{\delta_{NT}^{\frac{c}{2}}})$. 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 $j \leq r_1$, 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,

$$\begin{aligned}&P(\delta_{NT}^2\tfrac{1}{\tilde{k}}\textstyle\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 > C) \\ &= P(\delta_{NT}^2\tfrac{1}{\tilde{k}}\textstyle\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 > C, \left|\tilde{k}-k_0\right| > M) \\ &+ \textstyle\sum_{k=k_0-M}^{k_0+M} P(\delta_{NT}^2\tfrac{1}{\tilde{k}}\sum_{t=1}^{\tilde{k}}\left\|\hat{f}_t^u(\tilde{k}) - H_1^{u\prime}(\tilde{k})f_t\right\|^2 > C, \tilde{k} = k). \\ &\leq P(\left|\tilde{k}-k_0\right| > M) + \textstyle\sum_{k=k_0-M}^{k_0+M} P(\delta_{NT}^2\tfrac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2 > C) \\ &\leq \tfrac{\epsilon}{2} + \textstyle\sum_{k=k_0-M}^{k_0+M} P(\delta_{NT}^2\tfrac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{u\prime}(k)f_t\right\|^2 > C).\end{aligned}$$

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 $C = \max_{k\in[k_0-M,k_0+M]} C(k)$, then $P(\delta_{NT}^2 \frac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{w\prime}(k)f_t\right\|^2 > C) \leq \frac{\epsilon}{2} + \sum_{k=k_0-M}^{k_0+M}\frac{\epsilon}{2(2M+1)} = \epsilon$ for all (N,T), hence it remains to show for each $k \in [k_0 - M, k_0 + M]$, $\frac{1}{k}\sum_{t=1}^{k}\left\|\hat{f}_t^u(k) - H_1^{w\prime}(k)f_t\right\|^2 = O_p(\frac{1}{\delta_{NT}^2})$.

First consider the case $k_0 - M \leq k \leq k_0$. In this case, factor loadings are stable and $k_0 - M \leq k$ guarantees $k \to \infty$ as $k_0 \to \infty$, hence Theorem 1 of Bai and Ng (2002) is applicable.

Next consider the case $k_0 + 1 \leq k \leq k_0 + M$. 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

$$\begin{aligned} & X(k)X'(k) \\ = \; & F_1(k)\Lambda_{01}'\Lambda_{01}F_1'(k) + F_1(k)\Lambda_{01}'[E(k)+W(k)]' \\ & +[E(k)+W(k)]\Lambda_{01}F_1'(k) + [E(k)+W(k)][E(k)+W(k)]'. \qquad \text{(A-16)} \end{aligned}$$

It follows that

$$\begin{aligned} \hat{f}_t^u(k) - H_1^{w\prime}(k)f_t \;=\; & \frac{1}{Nk}[\tilde{F}_1^{w\prime}(k)F_1(k)\Lambda_{01}'e_t + \tilde{F}_1^{w\prime}(k)E(k)\Lambda_{01}f_t + \tilde{F}_1^{w\prime}(k)E(k)e_t \\ & +\tilde{F}_1^{w\prime}(k)F_1(k)\Lambda_{01}'w_t + \tilde{F}_1^{w\prime}(k)W(k)\Lambda_{01}f_t + \tilde{F}_1^{w\prime}(k)W(k)w_t \\ & +\tilde{F}_1^{w\prime}(k)E(k)w_t + \tilde{F}_1^{w\prime}(k)W(k)e_t] \\ =\; & Q_{1,t}(k) + Q_{2,t}(k) + Q_{3,t}(k) + Q_{4,t}(k) + Q_{5,t}(k) + Q_{6,t}(k) \\ & +Q_{7,t}(k) + Q_{8,t}(k), \qquad \text{(A-17)} \end{aligned}$$

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 $1\leq t\leq k_0$,

$$\begin{aligned}
\frac{1}{k}\sum\nolimits_{t=1}^{k}\|Q_{4,t}(k)\|^2 &= \frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{u\prime}(k)F_1(k)\Lambda_{01}'w_t\right\|^2\\
&\leq \frac{1}{k}\sum\nolimits_{t=1}^{k}(\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\sum\nolimits_{s=1}^{k}\|f_s\|^2)\left\|\frac{1}{N}\Lambda_{01}'w_t\right\|^2\\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\sum\nolimits_{s=1}^{k}\|f_s\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}\|^2)\\
&\quad (\frac{1}{k}\sum\nolimits_{t=1}^{k}\frac{1}{N}\sum\nolimits_{i=1}^{N}\|w_{it}\|^2)\\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{k}\sum\nolimits_{s=1}^{k}\|f_s\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}\|^2)\\
&\quad (\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}-\lambda_{02,i}\|^2)(\frac{1}{k}\sum\nolimits_{t=k_0+1}^{k}\|f_t\|^2)\\
&= O_p(1)O_p(1)O(1)O(1)O_p(\frac{1}{T}) = O_p(\frac{1}{T}),
\end{aligned}$$

$$\begin{aligned}
\frac{1}{k}\sum\nolimits_{t=1}^{k}\|Q_{5,t}(k)\|^2 &= \frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{u\prime}(k)W(k)\Lambda_{01}f_t\right\|^2\\
&\leq \frac{1}{k}\sum\nolimits_{t=1}^{k}(\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{N^2}\frac{1}{k}\sum\nolimits_{s=1}^{k}\|w_s'\Lambda_{01}f_t\|^2)\\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}\|^2)(\frac{1}{k}\sum\nolimits_{t=1}^{k}\|f_t\|^2)\\
&\quad (\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}-\lambda_{02,i}\|^2)(\frac{1}{k}\sum\nolimits_{s=k_0+1}^{k}\|f_s\|^2)\\
&= O_p(1)O(1)O_p(1)O(1)O_p(\frac{1}{T}) = O_p(\frac{1}{T}),
\end{aligned}$$

$$\begin{aligned}
\frac{1}{k}\sum\nolimits_{t=1}^{k}\|Q_{6,t}(k)\|^2 &= \frac{1}{k}\sum\nolimits_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{u\prime}(k)W(k)w_t\right\|^2\\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)\frac{1}{N^2}(\frac{1}{k}\sum\nolimits_{s=1}^{k}\|w_s\|^2)(\frac{1}{k}\sum\nolimits_{t=1}^{k}\|w_t\|^2)\\
&\leq (\frac{1}{k}\sum\nolimits_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2)(\frac{1}{N}\sum\nolimits_{i=1}^{N}\|\lambda_{01,i}-\lambda_{02,i}\|^2)^2\\
&\quad (\frac{1}{k}\sum\nolimits_{s=k_0+1}^{k}\|f_s\|^2)(\frac{1}{k}\sum\nolimits_{t=k_0+1}^{k}\|f_t\|^2)\\
&= O_p(1)O(1)O_p(\frac{1}{T})O_p(\frac{1}{T}) = O_p(\frac{1}{T^2}),
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{k}\sum_{t=1}^{k}\left\|Q_{7,t}(k)\right\|^2 &= \frac{1}{k}\sum_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{w\prime}(k)E(k)w_t\right\|^2 \\
&\leq \left(\frac{1}{k}\sum_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2\right)\left(\frac{1}{k}\frac{1}{N}\sum_{s=1}^{k}\sum_{i=1}^{N}e_{is}^2\right) \\
&\quad \left(\frac{1}{k}\sum_{t=k_0+1}^{k}\left\|f_t\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_{01,i}-\lambda_{02,i}\right\|^2\right) \\
&= O_p(1)O_p(1)O_p\left(\frac{1}{T}\right)O(1) = O_p\left(\frac{1}{T}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{k}\sum_{t=1}^{k}\left\|Q_{8,t}(k)\right\|^2 &= \frac{1}{k}\sum_{t=1}^{k}\left\|\frac{1}{Nk}\tilde{F}_1^{w\prime}(k)W(k)e_t\right\|^2 \\
&\leq \frac{1}{k}\sum_{t=1}^{k}\left(\frac{1}{k}\sum_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2\right)\frac{1}{N^2}\left(\frac{1}{k}\sum_{s=1}^{k}\left\|w_s'e_t\right\|^2\right) \\
&\leq \left(\frac{1}{k}\sum_{s=1}^{k}\left\|\tilde{f}_s^u(k)\right\|^2\right)\left(\frac{1}{k}\sum_{t=k_0+1}^{k}\left\|f_s\right\|^2\right) \\
&\quad \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_{01,i}-\lambda_{02,i}\right\|^2\right)\left(\frac{1}{k}\frac{1}{N}\sum_{t=1}^{k}\sum_{i=1}^{N}e_{it}^2\right) \\
&= O_p(1)O_p\left(\frac{1}{T}\right)O(1)O_p(1) = O_p\left(\frac{1}{T}\right),
\end{aligned}
$$

hence $\frac{1}{k}\sum_{t=1}^{k}\left\|Q_{m,t}(k)\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ 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 $P(\min_{|l|>M} W(l) \leq 0) \leq P(\min_{l<-M} W_1(l) \leq 0) + P(\min_{l>M} W_2(l) \leq 0)$
$= P(\sup_{l<-M} \{-l \|\Sigma_2 - \Sigma_1\|^2 - 2\sum_{t=k_0+l}^{k_0} [vec(\Sigma_2 - \Sigma_1)]' y_t\} \leq 0)$
$+P(\sup_{l>M} \{l \|\Sigma_2 - \Sigma_1\|^2 - 2\sum_{t=k_0+1}^{k_0+l} [vec(\Sigma_2 - \Sigma_1)]' y_t\} \leq 0)$
$\leq P(\sup_{l<-M} 2[vec(\Sigma_2 - \Sigma_1)]' \frac{1}{l} \sum_{t=k_0+l}^{k_0} y_t \geq \|\Sigma_2 - \Sigma_1\|^2)$
$+P(\sup_{l>M} 2[vec(\Sigma_2 - \Sigma_1)]' \frac{1}{l} \sum_{t=k_0+1}^{k_0+l} y_t \geq \|\Sigma_2 - \Sigma_1\|^2)$
$\leq P(\sup_{l<-M} \left\|\frac{1}{-l} \sum_{t=k_0+l}^{k_0} y_t\right\| \geq \frac{\|\Sigma_2 - \Sigma_1\|}{2}) + P(\sup_{l>M} \left\|\frac{1}{l} \sum_{t=k_0+1}^{k_0+l} y_t\right\| \geq \frac{\|\Sigma_2 - \Sigma_1\|}{2}) = \frac{C}{M}$
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, $\mathbb{E}\left\|g_t\right\|^4 \leq \max\{\left\|A\right\|^4, \left\|B\right\|^4\}\mathbb{E}\left\|f_t\right\|^4 < M < \infty$, $\frac{1}{T}\sum_{t=1}^{T} g_t g_t' = \tau_0 \frac{1}{k_0}\sum_{t=1}^{k_0} A f_t f_t' A' + (1-\tau_0)\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} B f_t f_t' B' \xrightarrow{p} \tau_0 A\Sigma_F A' + (1-\tau_0)B\Sigma_F B' = \Sigma_G$ and Σ_G is positive definite.

Assumption B:

By Assumption 2, $\left\|\gamma_i\right\| \leq \left\|(\lambda_{0,i}', \lambda_{1,i}', \lambda_{2,i}')'\right\| = (\left\|\lambda_{0,i}\right\|^2 + \left\|\lambda_{1,i}\right\|^2 + \left\|\lambda_{2,i}\right\|^2)^{\frac{1}{2}} \leq \sqrt{3}\bar{\lambda} < \infty$ 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:

$$\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t e_{it}\right\|^2\right) \leq 2\left\|A\right\|^2 \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k_0} f_t e_{it}\right\|^2\right)$$
$$+2\left\|B\right\|^2 \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=k_0+1}^{T} f_t e_{it}\right\|^2\right) \leq 2\tau_0 M + 2(1-\tau_0)M = 2M. \quad \blacksquare$$

Lemma 2 *Under Assumptions 1-5 and 7,* $\left\|J - J_0\right\| = o_p(1)$.

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) $\sup\limits_{k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\left\|g_t\right\|^2 = O_p(1)$, $\sup\limits_{k\geq k_0}\frac{1}{T-k}\sum_{t=k+1}^{T}\left\|g_t\right\|^2 = O_p(1)$, $\sup\limits_{k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\left\|g_t\right\|^2 = O_p(1)$ *and* $\sup\limits_{k>k_0}\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\left\|g_t\right\|^2 = O_p(1)$.

Proof. (1) $P(\sup\limits_{m\leq k\leq k_0} c_k\left\|\sum_{t=1}^{k} y_t\right\| > M) = P(\sup\limits_{m\leq k\leq k_0} c_k\left\|J_0' A[\sum_{t=1}^{k}(f_t f_t' - \Sigma_F)]A' J_0\right\| > M) \leq P(\left\|J_0' A\right\|^2 \sup\limits_{m\leq k\leq k_0} c_k\left\|\sum_{t=1}^{k}\epsilon_t\right\| > M) \leq \frac{C}{M^2}(m c_m^2 + \sum_{k=m+1}^{k_0} c_k^2)$, 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 $\left|\sup\limits_{k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}(\|f_t\|^2 - \mathbb{E}\,\|f_t\|^2)\right| \leq \sup\limits_{k\leq k_0}\left|\frac{1}{k}\sum_{t=1}^{k}(\|f_t\|^2 - \mathbb{E}\,\|f_t\|^2)\right| \leq \sqrt{r+q_1}\sup\limits_{k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}\epsilon_t\right\|$, 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) $\sup\limits_{k\in D^c,k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}(\tilde{g}_t - J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}})$,

(4) $\sup\limits_{k\in D,k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}(\tilde{g}_t - J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}})$,

(5) $\sup\limits_{k\in D,k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| = O_p(\frac{1}{\delta_{NT}^2})$,

(6) $\sup\limits_{k\in D^c,k\leq k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| = O_p(\frac{1}{\delta_{NT}^2})$,

(7) $\sup\limits_{k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}(\tilde{g}_t - J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}})$.

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).

$$
\begin{aligned}
& \sup_{k\in D^c,k\le k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\| \\
= & \sup_{k\in D^c,k\le k_0}\left\|\begin{array}{c}\frac{1}{k}\sum_{t=1}^{k}V_{NT}^{-1}\frac{1}{T}(\sum_{s=1}^{T}\tilde{g}_s\gamma_N(s,t)+\sum_{s=1}^{T}\tilde{g}_s\zeta_{st} \\ +\sum_{s=1}^{T}\tilde{g}_s\eta_{st}+\sum_{s=1}^{T}\tilde{g}_s\xi_{st})g_t'J\end{array}\right\| \\
\le & (\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)g_t'\gamma_N(s,t)\right\| \\
& +\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}J'g_sg_t'\gamma_N(s,t)\right\| \\
& +\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)g_t'\zeta_{st}\right\| \\
& +\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}J'g_sg_t'\zeta_{st}\right\| \\
& +\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)g_t'\eta_{st}\right\| \\
& +\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}J'g_sg_t'\eta_{st}\right\| \\
& +\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}(\tilde{g}_s-J'g_s)g_t'\xi_{st}\right\| \\
& +\sup_{k\in D^c,k\le k_0}\left\|\frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T}J'g_sg_t'\xi_{st}\right\|)\left\|V_{NT}^{-1}\right\|\|J\| \\
= & (I+II+III+IV+V+VI+VII+VIII)\left\|V_{NT}^{-1}\right\|\|J\|.
\end{aligned}
$$

Consider the eight terms one by one.

$$
\begin{aligned}
& I \\
\le & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s-J'g_s\|^2)^{\frac{1}{2}}\sup_{k\in D^c,k\le k_0}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}g_t'\gamma_N(s,t)\right\|^2)^{\frac{1}{2}} \\
\le & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s-J'g_s\|^2)^{\frac{1}{2}}\sup_{k\in D^c,k\le k_0}[(\frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)(\frac{1}{T}\frac{1}{k}\sum\nolimits_{s=1}^{T}\sum\nolimits_{t=1}^{k}|\gamma_N(s,t)|^2)]^{\frac{1}{2}} \\
\le & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s-J'g_s\|^2)^{\frac{1}{2}}(\sup_{k\in D^c,k\le k_0}\frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2)^{\frac{1}{2}}(\sup_{k\in D^c,k\le k_0}\frac{1}{T}\frac{1}{k}\sum\nolimits_{s=1}^{T}\sum\nolimits_{t=1}^{k}|\gamma_N(s,t)|^2)^{\frac{1}{2}} \\
= & O_p(\frac{1}{\delta_{NT}})O_p(1)O_p(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

where last equality follows from Lemma 1, Lemma 3 and $\sup_{k\in D^c,k\leq k_0}\frac{1}{T}\frac{1}{k}\sum_{s=1}^{T}\sum_{t=1}^{k}|\gamma_N(s,t)|^2\leq \frac{1}{T}\sup_{k\in D^c,k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}(\sum_{s=1}^{T}M\,|\gamma_N(s,t)|)\leq\frac{1}{T}M^2$ by part (1) of Assumption 5.

$$
\begin{aligned}
& II \\
\leq\ & \|J\|\sup_{k\in D^c,k\leq k_0}\frac{1}{T}\sum_{s=1}^{T}\|g_s\|\left\|\frac{1}{k}\sum_{t=1}^{k}g_t'\gamma_N(s,t)\right\| \\
\leq\ & \|J\|\,(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}}\sup_{k\in D^c,k\leq k_0}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{k}\sum_{t=1}^{k}g_t'\gamma_N(s,t)\right\|^2)^{\frac{1}{2}} \\
\leq\ & \|J\|\,(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}}(\sup_{k\in D^c,k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\|g_t\|^2)^{\frac{1}{2}}(\sup_{k\in D^c,k\leq k_0}\frac{1}{T}\frac{1}{k}\sum_{s=1}^{T}\sum_{t=1}^{k}|\gamma_N(s,t)|^2)^{\frac{1}{2}} \\
=\ & O_p(1)O_p(1)O_p(1)O_p(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

where the last equality follows from Lemma 2, Assumption 1, Lemma 3 and $\sup_{k\in D^c,k\leq k_0}\frac{1}{T}\frac{1}{k}\sum_{s=1}^{T}\sum_{t=1}^{k}|\gamma_N(s,t)|^2=O_p(\frac{1}{T})$ as explained above.

$$
\begin{aligned}
& III \\
\leq\ & (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s-J'g_s\|^2)^{\frac{1}{2}}\sup_{k\in D^c,k\leq k_0}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{k}\sum_{t=1}^{k}g_t'\frac{1}{N}\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right\|^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s-J'g_s\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}}(\frac{1}{T}\sum_{s=1}^{T}\sup_{k\in D^c,k\leq k_0}\left\|\frac{1}{k}\frac{1}{\sqrt{N}}\sum_{t=1}^{k}\sum_{i=1}^{N}g_t'[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right\|^2)^{\frac{1}{2}} \\
\leq\ & (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s-J'g_s\|^2)^{\frac{1}{2}}(\sup_{k\in D^c,k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\|g_t\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}} \\
& (\frac{1}{T}\sum_{s=1}^{T}\sup_{k\in D^c,k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right\|^2)^{\frac{1}{2}} \\
=\ & O_p(\frac{1}{\delta_{NT}})O_p(1)\frac{1}{\sqrt{N}}O_p(1).
\end{aligned}
$$

$$
\begin{aligned}
& IV \\
\leq\; & \|J\| \sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T} g_s g_t' \frac{1}{N}\sum\nolimits_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right\| \\
\leq\; & \|J\| \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \left\| \frac{1}{NT}\sum\nolimits_{s=1}^{T}\sum\nolimits_{i=1}^{N} g_s[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right\|^2 \right)^{\frac{1}{2}} \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \|g_t\|^2 \right)^{\frac{1}{2}} \\
\leq\; & \|J\| \frac{1}{\sqrt{N}} \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2\right)^{\frac{1}{2}} \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \frac{1}{T}\sum\nolimits_{s=1}^{T} \left| \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right|^2 \right)^{\frac{1}{2}} \\
& \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \|g_t\|^2 \right)^{\frac{1}{2}} \\
=\; & O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1)O_p(1) = O_p\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}
$$

where the last equalities follow from part (1) of Assumption 10.

$$
\begin{aligned}
& V \\
\leq\; & \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)^{\frac{1}{2}} \sup_{k\in D^c, k\leq k_0} \left(\frac{1}{T}\sum\nolimits_{s=1}^{T} \left\| \frac{1}{k}\sum\nolimits_{t=1}^{k} \left(\frac{1}{N}\sum\nolimits_{i=1}^{N} g_s'\gamma_i e_{it}\right) g_t' \right\|^2 \right)^{\frac{1}{2}} \\
\leq\; & \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)^{\frac{1}{2}} \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2\right)^{\frac{1}{2}} \frac{1}{\sqrt{N}} \left(\sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{k}\frac{1}{\sqrt{N}}\sum\nolimits_{t=1}^{k}\sum\nolimits_{i=1}^{N} \gamma_i e_{it} g_t' \right\|^2 \right)^{\frac{1}{2}} \\
\leq\; & \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2\right)^{\frac{1}{2}} \left(\frac{1}{T}\sum\nolimits_{s=1}^{T}\|g_s\|^2\right)^{\frac{1}{2}} \frac{1}{\sqrt{N}} \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2 \right)^{\frac{1}{2}} \\
& \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \left\| \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{it} \right\|^2 \right)^{\frac{1}{2}} \\
=\; & O_p\left(\frac{1}{\delta_{NT}}\right)O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1)
\end{aligned}
$$

$$
\begin{aligned}
& VI \\
\leq\; & \|J\| \sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{T}\frac{1}{k}\sum\nolimits_{t=1}^{k}\sum\nolimits_{s=1}^{T} g_s\left(\frac{1}{N}\sum\nolimits_{i=1}^{N} g_s'\gamma_i e_{it}\right) g_t' \right\| \\
\leq\; & \|J\| \left\| \frac{1}{T}\sum\nolimits_{s=1}^{T} g_s g_s' \right\| \frac{1}{\sqrt{N}} \sup_{k\in D^c, k\leq k_0} \left\| \frac{1}{k}\frac{1}{\sqrt{N}}\sum\nolimits_{t=1}^{k}\sum\nolimits_{i=1}^{N}\gamma_i e_{it} g_t' \right\| \\
\leq\; & \|J\| \left\| \frac{1}{T}\sum\nolimits_{s=1}^{T} g_s g_s' \right\| \frac{1}{\sqrt{N}} \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k}\|g_t\|^2 \right)^{\frac{1}{2}} \left(\sup_{k\in D^c, k\leq k_0} \frac{1}{k}\sum\nolimits_{t=1}^{k} \left\| \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}\gamma_i e_{it} \right\|^2 \right)^{\frac{1}{2}} \\
=\; & O_p(1)O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1),
\end{aligned}
$$

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\nolimits_{s=1}^{T}\tilde{g}_s\gamma_N(s,t)+\sum\nolimits_{s=1}^{T}\tilde{g}_s\zeta_{st}\\ +\sum\nolimits_{s=1}^{T}\tilde{g}_s\eta_{st}+\sum\nolimits_{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.

$$
\begin{aligned}
& XV \\
\leq\ & (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}g_t'\gamma_i e_{is}\right\|^2 \\
\leq\ & (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s - J'g_s\|^2)(\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\|g_t\|^2)\frac{1}{N}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2) \\
=\ & O_p(\frac{1}{\delta_{NT}^2})O_p(1)\frac{1}{N}O_p(1).
\end{aligned}
$$

$$
\begin{aligned}
& XVI \\
\leq\ & \|J\|^2(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)(\sup_{k\in D,k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\|g_t\|^2)\frac{1}{N}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2) \\
=\ & O_p(1)O_p(1)O_p(1)\frac{1}{N}O_p(1),
\end{aligned}
$$

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),

$$\begin{aligned}
&\left\|\sum_{t=1}^{k} z_t\right\|^2 \\
\leq\ & (\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| + 2\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\| \\
&+\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\| + 2\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\|)^2 \\
\leq\ & 4\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\|^2 + 16\left\|\sum_{t=1}^{k}(\tilde{g}_t - J'g_t)g_t'J\right\|^2 \\
&+4\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\|^2 + 16\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\|^2. \quad \text{(A-20)}
\end{aligned}$$

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,

$$\sup_{k\in D^c,k\leq k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|^2\leq\sup_{k\in D^c,k\leq k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|^2=O_p(\frac{1}{\delta_{NT}^2}).$$

Using Lemma 2 and Assumption 3,

$$\sup_{k\leq k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\|^2\leq\|J-J_0\|^4\sup_{k\leq k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}g_tg_t'\right\|^2=o_p(1),$$

$$\sup_{k\leq k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\|^2\leq\|J-J_0\|^2\|J_0\|^2\sup_{k\leq k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}g_tg_t'\right\|^2=o_p(1).$$

It follows $\sup_{k\in D,k\leq k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}z_t\right\|^2$ and $\sup_{k\in D^c,k\leq k_0}\frac{1}{k_0}\frac{1}{k}\left\|\sum_{t=1}^{k}z_t\right\|^2$ are both $o_p(1)$.

For parts (3) and (4),

$$\begin{aligned}&\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}z_t\right\|\\ \leq\;&\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|+2\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|\\ &+\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\|+2\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\|.\quad\text{(A-21)}\end{aligned}$$

Using Lemma 1,

$$\sup_{k\in D,k\leq k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|\leq\frac{1}{\tau_0}\frac{1}{T}\sum\nolimits_{t=1}^{T}\|\tilde{g}_t-J'g_t\|^2=O_p(\frac{1}{\delta_{NT}^2}),$$

$$\sup_{k\in D^c,k\leq k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)(\tilde{g}_t-J'g_t)'\right\|\leq\frac{1}{\tau_0}\frac{1}{T}\sum\nolimits_{t=1}^{T}\|\tilde{g}_t-J'g_t\|^2=O_p(\frac{1}{\delta_{NT}^2}).$$

Using part (1) of Lemma 5,

$$\sup_{k\in D,k\leq k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|\leq\sup_{k\in D,k\leq k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|=O_p(\frac{1}{\delta_{NT}}).$$

Using part (2) of Lemma 5,

$$\sup_{k\in D^c,k\leq k_0}\frac{1}{k_0}\left\|\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|\leq\sup_{k\in D^c,k\leq k_0}\left\|\frac{1}{k}\sum\nolimits_{t=1}^{k}(\tilde{g}_t-J'g_t)g_t'J\right\|=O_p(\frac{1}{\delta_{NT}}).$$

Using Lemma 2 and Assumption 3,

$$\sup_{k\in D,k\le k_0}\frac{1}{k_0}\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\| \le \|J-J_0\|^2 \sup_{k\in D,k\le k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}g_tg_t'\right\| = o_p(1),$$

$$\sup_{k\in D^c,k\le k_0}\frac{1}{k_0}\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'(J-J_0)\right\| \le \|J-J_0\|^2 \sup_{k\in D^c,k\le k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}g_tg_t'\right\| = o_p(1),$$

$$\sup_{k\in D,k\le k_0}\frac{1}{k_0}\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\| \le \|J-J_0\|\,\|J_0\| \sup_{k\in D,k\le k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}g_tg_t'\right\| = o_p(1),$$

$$\sup_{k\in D^c,k\le k_0}\frac{1}{k_0}\left\|\sum_{t=1}^{k}(J'-J_0')g_tg_t'J_0\right\| \le \|J-J_0\|\,\|J_0\| \sup_{k\in D^c,k\le k_0}\left\|\frac{1}{k}\sum_{t=1}^{k}g_tg_t'\right\| = o_p(1).$$

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 $\sup_{k\le k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}g_tg_t'\right\| = O_p(1)$, which is proved below.

$$\begin{aligned}
\sup_{k\le k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}g_tg_t'\right\| &\le \sup_{k<k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}g_tg_t'\right\| + \left\|\frac{1}{T-k_0}\sum_{t=k_0+1}^{T}g_tg_t'\right\| \\
&\le \sup_{k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}g_tg_t'\right\| + 2\left\|\frac{1}{T-k_0}\sum_{t=k_0+1}^{T}g_tg_t'\right\| \\
&= O_p(1).
\end{aligned}$$

■

References

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

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

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

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

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