Estimating and testing high dimensional factor models with multiple structural changes

Baltagi, Badi H. and Kao, Chihwa and Wang, Fa

Syracuse University, University of Connecticut, Cass Business School

1 March 2016

ESTIMATING AND TESTING HIGH DIMENSIONAL FACTOR MODELS WITH MULTIPLE STRUCTURAL CHANGES

Badi H. Baltagi* Chihwa Kao† Fa Wang‡

July 26, 2019

Abstract

This paper considers multiple changes in the factor loadings of a high dimensional factor model occurring at dates that are unknown but common to all subjects. Since the factors are unobservable, the problem is converted to estimating and testing structural changes in the second moments of the pseudo factors. We consider both joint and sequential estimation of the change points and show that the distance between the estimated and the true change points is $O_p(1)$. We find that the estimation error contained in the estimated pseudo factors has no effect on the asymptotic properties of the estimated change points as the cross-sectional dimension N and the time dimension T go to infinity jointly. No N-T ratio condition is needed. We also propose (i) tests for the null of no change versus the alternative of l changes (ii) tests for the null of l changes versus the alternative of $l+1$ changes, and show that using estimated factors asymptotically has no effect on their limit distributions if $\sqrt{T}/N \to 0$. These tests allow us to make inference on the presence and number of structural changes. Simulation results show good performance of the proposed procedure. In an application to US quarterly macroeconomic data we detect two possible breaks.

Keywords: factor model, multiple changes, model selection, panel data

JEL Classification: C13; C33.

*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

‡Faculty of Finance, Cass Business School, London, EC1Y8TZ, UK

1 INTRODUCTION

High dimensional factor models have played a crucial role in business cycle analysis, consumer behavior analysis, asset pricing and macroeconomic forecasting, see for example, Ross (1976), Lewbel (1991), Bernanke, Boivin and Eliasz (2005) and Stock and Watson (2002a, 2002b), to mention a few. This has been enhanced by the increasing availability of big data sets. However, as the time span of the data becomes longer, there is a substantial risk that the underlying data generating process may experience structural changes. Inference ignoring these changes would be misleading. This paper considers multiple changes in the factor loadings of a high dimensional factor model, occurring at dates that are unknown but common to all subjects. We propose a joint estimator of all the change points as well as a sequential estimator of the change points that estimates these change points one by one. Based on the estimated change points, we are able to consistently determine the number of factors and estimate the factor space in each regime. We also propose tests for (i) the null of no change versus the alternative of some fixed number of changes and (ii) tests for the null of l changes versus the alternative of $l+1$ changes. The latter allows us to consistently determine the number of changes. These tests are easy to implement and critical values tabulated in Bai and Perron (1998, 2003) can be used directly to make inference on the presence as well as the number of structural changes.

Stock and Watson (2009) and Bates, Plagborg-Møller, Stock and Watson (2013) argue that as long as the magnitude of the loading breaks converges to zero sufficiently fast, existing estimators ignoring loading breaks are still consistent. Recently, several tests on the stability of the factor loadings in high dimensional factor models have been proposed, including Breitung and Eickmeier (2011), Chen, Dolado and Gonzalo (2014), Han and Inoue (2015) and Cheng, Liao and Schorfheide (2016). Recent contributions on estimating high dimensional factor models with loading instability include Baltagi, Kao and Wang (2017), Cheng et al. (2016), Massacci (2017) and Bai, Han and Shi (2016). All of these papers consider the case with a single change. The number of factors is explicitly allowed to change in the former two papers. The change point estimator of Bai et al. (2016) is consistent (hence more accurate than those of

the former two papers), but it does not allow the number of factors to change.

This paper tackles multiple changes in high dimensional factor models[1]. We start by estimating the number of factors and factor space ignoring structural changes. Since the factor model with changes in the loadings can be equivalently written as another factor model with stable loadings but pseudo factors, this would allow us to identify the equivalent model with stable loadings and give us the estimated pseudo factors. A key observation is that the mean of the second moment matrix of the pseudo factors have changes at exactly the same dates as the loadings. Estimating and testing multiple changes in the latter can be converted to estimating and testing multiple changes in the former. This conversion is crucial because the true factors are unobservable and not estimable without knowing the change points. It is also worth pointing out that after this conversion we are using the estimated pseudo factors, not the pseudo factors themselves. That is to say, the data contains estimation error. We will show that this estimation error has a different effect on testing and estimating structural changes. Once the estimated change points are available, they are plugged in to split the sample and estimate the number of factors and factor space in each regime, which are further used to construct the test for l versus $l+1$ changes.

In the regression setup, influential work on multiple changes include Bai and Perron (1998) and Qu and Perron (2007). This paper differs from these seminal papers in several respects. First, to estimate and test structural changes, this paper utilizes estimated pseudo factors rather than the raw data. Second, the estimated pseudo factors have a multivariate time series setup, while Bai and Perron (1998) have a regression setup. Third, the estimated pseudo factors contain estimation error and we show that to eliminate the effect of estimation error, for testing structural changes we need $\frac{\sqrt{T}}{N} \to 0$ as N and T go to infinity jointly, but no N-T (T-N) ratio condition is needed for estimating change points. The latter is rare in the high dimensional econometrics literature since very few papers require no N-T (T-N) ratio condition[2].

[1] In testing the joint hypothesis of stability of both factor loadings and the factor augmented forecasting equation, Corradi and Swanson (2014) also consider the alternative of multiple changes.

[2] For example, Bai and Ng (2006) require $\frac{\sqrt{T}}{N} \to 0$ where estimated factors are used to augment forecasting and vector autoregression. Various N-T ratio conditions are also needed in Bai (2009) where estimated factors are used to control the interactive effects in panel data.

This paper is also closely related to Han and Inoue (2015), Baltagi et al. (2017) and Ma and Su (2018). Han and Inoue (2015) propose Wald and LM type tests for single change in a factor model. These tests can not be extended to cases with multiple changes directly since they are based on the difference of the second moments of factors between two subsamples. This paper solves this issue by considering likelihood ratio type tests. This paper also contains results on estimating change points, which is not covered in Han and Inoue (2015). Baltagi et al. (2017) propose an estimator for a single change point in a factor model and prove that the distance between the estimated and the true change point is $O_p(1)$, and this $O_p(1)$ error asymptotically has no effect on the estimated number of factors and factor space in each regime. This paper differs from Baltagi et al. (2017) in two respects. First, for the multiple changes case, although the distance between the estimated and the true change points are still $O_p(1)$, the proof is different. This is because when analyzing the location of one change point, the locations of the previous and the next change point are unknown[3]. Second, this paper also studies the testing procedure to determine the presence and number of structural changes, which is not covered in Baltagi et al. (2017). Ma and Su (2018) propose an adaptive fused group Lasso method to estimate and test multiple structural changes in factor models. Their method follows from the Lasso literature while our method follows from converting the original high dimensional setup to a fixed dimensional setup. Their estimator of the changes points is consistent, but their method requires the number of factors to be stable. Our method allows (i) the number of factors to change as well as (ii) more general type of changes. Allowing the number of factors to change is important and has been considered by Stock and Watson (2012) and Cheng et al. (2016). The latter found that given macroeconomic and financial indicators have a factor representation, one new factor that captures financial comovement emerges at the beginning of the Great Recession.

Throughout the paper, $\|A\| = (trAA')^{\frac{1}{2}}$ denotes the Frobenius norm, $\xrightarrow{p}$, $\xrightarrow{d}$ and $\Rightarrow$ denotes convergence in probability, convergence in distribution and weak convergence of stochastic process respectively, $vech(A)$ denotes the half vectorization of matrix A, $E(\cdot)$ denotes the expectation, $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$ and $(N,T) \to \infty$ de-

[3]Obviously, the single change case does not have this issue.

notes N and T going to infinity jointly. The rest of the paper is organized as follows: Section 2 introduces the model setup, notation and preliminaries. Section 3 considers both joint estimation and sequential estimation of the change points and also the subsequent estimation of the number of factors and factor space in each regime. Section 4 proposes test statistics for multiple changes, derives their asymptotic distributions and discusses how to determine the number of changes. Section 5 presents simulation results. Section 6 provides an empirical application to US macroeconomic data. Section 7 concludes. All the proofs are relegated to the appendix.

2 NOTATION AND PRELIMINARIES

2.1 The Model

Consider the following high dimensional factor model with L changes in the factor loadings:

$$x_{it} = f_{0,t}'\lambda_{0,i} + f_{-0,t}'\lambda_{\kappa,i} + e_{it}, \tag{1}$$

with $k_{\kappa-1,0} + 1 \leq t \leq k_{\kappa,0}$, for $\kappa = 1, ..., L+1, i = 1, ..., N$ and $t = 1, ..., T$, where $f_{0,t}$ and $f_{-0,t}$ are $r-q$ and q dimensional vectors of factors without and with changes in the loadings respectively. Let $f_t = (f_{0,t}', f_{-0,t}')'$. $\lambda_{0,i}$ and $\lambda_{\kappa,i}$ are factor loadings of subject i corresponding to $f_{0,t}$ and $f_{-0,t}'$ in the κ-th regime, respectively, and let $\lambda_{0\kappa,i} = (\lambda_{0,i}', \lambda_{\kappa,i}')'$. e_{it} is the error term allowed to have temporal and cross-sectional dependence as well as heteroskedasticity. For $\kappa = 1, ..., L$, $k_{\kappa,0}$ are change points ($k_{0,0} = 0$ and $k_{L+1,0} = T$), $\tau_{\kappa,0} = \frac{k_{\kappa,0}}{T}$ are change fractions and considered fixed in the asymptotic analysis. When there is no change in all factor loadings, let λ_i and Λ denote the factor loading and the factor loading matrix.

In matrix form, the model can be expressed as follows:

$$X_{\kappa*} = F_{0\kappa*}\Lambda_0' + F_{-0\kappa*}\Lambda_\kappa' + E_{\kappa*}, \text{ for } \kappa = 1, ..., L+1. \tag{2}$$

$X_{\kappa*} = (x_{k_{\kappa-1,0}+1}, ..., x_{k_{\kappa,0}})'$ and $E_{\kappa*} = (e_{k_{\kappa-1,0}+1}, ..., e_{k_{\kappa,0}})'$ are both of dimension $(k_{\kappa,0} - k_{\kappa-1,0}) \times N$. $F_{0\kappa*} = (f_{0,k_{\kappa-1,0}+1}, ..., f_{0,k_{\kappa,0}})'$ and $F_{-0\kappa*} = (f_{-0,k_{\kappa-1,0}+1}, ..., f_{-0,k_{\kappa,0}})'$ are of dimensions $(k_{\kappa,0} - k_{\kappa-1,0}) \times (r-q)$ and $(k_{\kappa,0} - k_{\kappa-1,0}) \times q$ respectively. Here we use "$\kappa*$"

to denote that the sample split is based on the true change points. $\Lambda_0 = (\lambda_{0,1}, ..., \lambda_{0,N})'$ and $\Lambda_\kappa = (\lambda_{\kappa,1}, ..., \lambda_{\kappa,N})'$ are of dimensions $N \times (r-q)$ and $N \times q$ respectively. Also, let $F_{\kappa *} = (F_{0\kappa *}, F_{-0\kappa *}) = (f_{k_{\kappa-1,0}+1}, ..., f_{k_{\kappa,0}})'$ and $\Lambda_{0\kappa} = (\Lambda_0, \Lambda_\kappa) = (\lambda_{0\kappa,1}, ..., \lambda_{0\kappa,N})'$.

2.2 Equivalent Representation

First note that in model (1), changes in the number of factors are allowed for, and incorporated as a special case of changes in the loadings by allowing Λ_κ to contain some zero columns for some κ. Second, for each factor considered in model (1), its loadings are nonzero for at least one κ, otherwise it would be totally irrelevant. Third, zero columns are allowed to appear at different locations of Λ_κ for different κ. This means that both emerging as well as disappearing factors are possible. Note that for this case we can still identify the break point, although we may not be able to identify whether the true model has both emerging and disappearing factors or the same factors with different loadings in two regimes.

To derive the equivalent representation, define Λ_{-0} as follows: Starting from the first column of Λ_1, if it is nonzero and linearly independent with Λ_0, put it in Λ_{-0}. If the second column of Λ_1 is nonzero and linearly independent with Λ_0 and the first column, put it in Λ_{-0}. In general, if the j-th column of Λ_κ is nonzero and linearly independent with Λ_0 and those columns are already in Λ_{-0}, put it in Λ_{-0}. Repeat this procedure for all κ and j.

Let $\Gamma = (\Lambda_0, \Lambda_{-0})$ and $\bar{r}$ denote the number of columns in Γ. From the definition of Λ_{-0} it is easy to see that (1) Γ is full column rank, (2) the j-th column of Λ_κ is allowed to be the same as the j-th column of $\Lambda_{\kappa-s}$ for some positive integer s, i.e., the factor loadings are allowed to switch back to their previous values after one or more breaks, (3) $\Lambda_{0\kappa} = \Gamma R_\kappa$ for some $\bar{r} \times r$ dimensional R_κ because $\Lambda_{0\kappa} = (\Lambda_0, \Lambda_\kappa)$, columns of Λ_0 are all included in Γ, and columns of Λ_κ are either included in Γ or linear combinations of columns in Γ.[4]

Let $G_{\kappa *} = (g_{k_{\kappa-1,0}+1}, ..., g_{k_{\kappa,0}})' = F_{\kappa *} R_\kappa'$. It follows that $g_t = R_\kappa f_t$ if $k_{\kappa-1,0} + 1 \leq$

[4] Zero columns of Λ_κ are also linear combinations of columns in Γ.

$t \leq k_{\kappa,0}$, and

$$X_{\kappa*} = F_{\kappa*}\Lambda_{0\kappa}' + E_{\kappa*} = F_{\kappa*}R_\kappa'\Gamma' + E_{\kappa*} = G_{\kappa*}\Gamma' + E_{\kappa*}, \tag{3}$$

which is a factor model with stable loadings Γ and $\bar{r}$ dimensional pseudo factors g_t. Equation (3) generalizes the equivalent representation in Baltagi et al. (2017) to cases with multiple changes.

Remark 1 *The identification condition for the κ-th change point is $\Sigma_{G,\kappa} \neq \Sigma_{G,\kappa+1}$. This is satisfied since $\Sigma_{G,\kappa} = R_\kappa \Sigma_F R_\kappa'$ and $R_\kappa \neq R_{\kappa+1}$.*

Remark 2 *To ensure the uniqueness (up to a rotation) of the equivalent representation, here we show that as long as $\frac{1}{k_{\kappa,0}-k_{\kappa-1,0}} \sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} f_t f_t' - \Sigma_F \xrightarrow{p} 0$ for each κ and $\left\| \frac{1}{N}\Gamma'\Gamma - \Sigma_\Gamma \right\| \to 0$ for some positive definite Σ_F and Σ_Γ, $\frac{1}{T}\sum_{t=1}^{T} g_t g_t' - \Sigma_G \xrightarrow{p} 0$ for some positive definite Σ_G. First, it is not difficult to see that $\Sigma_G = \sum_{\kappa=1}^{L+1}(\tau_{\kappa,0} - \tau_{\kappa-1,0})\Sigma_{G,\kappa}$, where $\Sigma_{G,\kappa} = R_\kappa \Sigma_F R_\kappa'$ is positive semidefinite for all κ. Thus for any $\bar{r}$ dimensional vector v, $v'\Sigma_G v = 0$ implies $v'\Sigma_{G,\kappa} v = 0$ for all κ, which further implies $v'R_\kappa = 0$ for all κ. Since the $\bar{r} \times r(L+1)$ matrix $(R_1, ..., R_{L+1})$ has rank $\bar{r}$, v has to be zero, and therefore Σ_G is positive definite.*

Remark 3 *Break in the intercept of x_{it} is absorbed into a break in the loadings. Break in the variance of x_{it} could be due to a break in the variance-covariance of factors, or a break in the factor loadings, or a break in the error variance. Our method cannot distinguish between a break in the factor loadings and a break in the factor variance, but can distinguish between these two possibilities and a break in the error variance, because our method only detects the breaks in the second moment matrix of the pseudo factors.*

3 ESTIMATING MODELS WITH MULTIPLE CHANGES

In this section, we propose a joint estimator for all change points as well as a sequential estimator which estimates the change points one by one, assuming the number of breaks is known. How to determine the number of breaks will be discussed in the next

section. For both estimators, we show that the distance between the estimated and the true change points is $O_p(1)$. In economic studies, the estimated change points may provide guidance for uncovering the underlying factors or mechanism of the structural change, or analyzing the effect of economic policy. The estimated change points also have important implications for factor-augmented forecasting, which will be discussed at the end of this section.

Based on the estimated change points, we can split the sample and estimate the number of factors and the factor space in each regime. As discussed extensively in the literature, consistently estimated factors can be helpful for business cycle analysis, asset pricing and other issues. In this paper, the estimated factors will be used to construct a test for l versus $l+1$ breaks.

3.1 Joint Estimation of the Change Points

We first introduce the estimation procedure, and then impose assumptions to study the asymptotic properties of the proposed estimators.

3.1.1 Estimation Procedure

The estimation procedure is as follows:

1. Using any consistent estimator, e.g., Bai and Ng (2002), Ahn and Horenstein (2013) to estimate the number of factors ignoring structural changes, i.e., to estimate the number of pseudo factors. Denote this estimator by $\tilde{r}$.

2. Estimate the first $\tilde{r}$ factors using the principal component method. Let $\tilde{g}_t, t = 1, ..., T$ be the estimated factors[5].

3. For any partition $(k_1, ..., k_L)$, split the sample into $L+1$ subsamples, estimate the second moment matrix of g_t in each subsample as $\tilde{\Sigma}_\kappa = \frac{1}{k_\kappa - k_{\kappa-1}} \sum_{t=k_{\kappa-1}+1}^{k_\kappa} \tilde{g}_t \tilde{g}_t'$ and calculate the sum of squared residuals,

$$\tilde{S}(k_1, ..., k_L) = \sum\nolimits_{\kappa=1}^{L+1} \sum\nolimits_{t=k_{\kappa-1}+1}^{k_\kappa} [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_\kappa)]' [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_\kappa)]. \quad (4)$$

[5]The change points estimator also can be based on $\hat{g}_t$, where $(\hat{g}_1, ..., \hat{g}_T)' = \hat{G} = \tilde{G} V_{NT} = (\tilde{g}_1, ..., \tilde{g}_T)' V_{NT}$ and V_{NT} is a diagonal matrix that contains the first $\bar{r}$ largest eigenvalues of $\frac{1}{NT} XX'$.

Then estimate the change points by minimizing the sum of squared residuals,

$$(\tilde{k}_1, ..., \tilde{k}_L) = \arg\min \tilde{S}(k_1, ..., k_L). \tag{5}$$

The underlying mechanism is as follows:

1. Since model (2) has an equivalent representation (3), $\tilde{r}$ is consistent for $\bar{r}$, $\tilde{g}_t$ is asymptotically close to $J'g_t$ for some rotation matrix J , and $J'g_t$ is asymptotically close to $J_0'g_t$, where $J \xrightarrow{p} J_0 = \Sigma_\Gamma^{\frac{1}{2}} \Phi V^{-\frac{1}{2}}$, with V being the diagonal matrix of eigenvalues of $\Sigma_\Gamma^{\frac{1}{2}} \Sigma_G \Sigma_\Gamma^{\frac{1}{2}}$ and Φ the corresponding eigenvector matrix.

2. The second moment matrix of g_t has breaks at the same points as the factor loadings.

3. The second moment matrix of $J_0'g_t$ has breaks at the same points as g_t.

More precisely, let $\mathbb{E}(f_t f_t') = \Sigma_F$ for all t, then $\Sigma_\kappa = J_0' \Sigma_{G,\kappa} J_0$ is the mean of $J_0' g_t g_t' J_0$. Let $y_t = vech(J_0' g_t g_t' J_0 - \Sigma_\kappa)$ for $t = k_{\kappa-1,0}+1, ..., k_{\kappa,0}$ with $\kappa = 1, ..., L+1$ and $z_t = vech(\tilde{g}_t \tilde{g}_t' - J_0' g_t g_t' J_0)$ for $t = 1, ..., T$, it follows that $vech(\tilde{g}_t \tilde{g}_t') = vech(\Sigma_\kappa) + y_t + z_t$ for $t = k_{\kappa-1,0} + 1, ..., k_{\kappa,0}$ and $\kappa = 1, ..., L+1$. Since $\Sigma_{G,\kappa} \neq \Sigma_{G,\kappa+1}$, $\Sigma_\kappa = J_0' \Sigma_{G,\kappa} J_0 \neq J_0' \Sigma_{G,\kappa+1} J_0 \neq \Sigma_{\kappa+1}$. Thus $vech(\tilde{g}_t \tilde{g}_t')$ is a multivariate process with L mean shifts and extra error z_t. We will show that to asymptotically eliminate the effect of z_t, this requires $(N, T) \to \infty$ and no N-T ratio condition is needed.

Remark 4 *Through estimating the number of pseudo factors, we are essentially selecting relevant moment conditions from a large number of candidates. The model with $\tilde{r} = \bar{r}$ has the strongest identification strength for the unknown change points. If $\tilde{r} > \bar{r}$, no information would be lost, but extra noise would be brought in by the extra estimated factors. If $\tilde{r} < \bar{r}$, change point estimation would be based on a subset of $vech(\tilde{g}_t \tilde{g}_t')$, thus identification of the change points would be weaker or even totally lost.*

3.1.2 Assumptions

The assumptions are as follows:

Assumption 1 $\mathbb{E}\|f_t\|^4 < M < \infty$, $\mathbb{E}(f_t f_t') = \Sigma_F$ *for all* t. Σ_F *is positive definite and* $\frac{1}{k_{\kappa,0}-k_{\kappa-1,0}} \sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} f_t f_t' - \Sigma_F = o_p(1)$ *for* $\kappa = 1, ..., L+1$. *Note that when there is no break,* $L = 0$, $k_{0,0} = 0$ *and* $k_{1,0} = T$.

Assumption 2 $\|\lambda_{0\kappa,i}\| \leq \bar{\lambda} < \infty$ *for* $\kappa = 1, ..., L+1$, *and* $\left\|\frac{1}{N}\Gamma'\Gamma - \Sigma_\Gamma\right\| = O(\frac{1}{\sqrt{N}})$ *for some positive definite matrix* Σ_Γ. *When there is no break,* $\|\lambda_i\| \leq \bar{\lambda} < \infty$ *and* $\left\|\frac{1}{N}\Lambda'\Lambda - \Sigma_\Lambda\right\| = O(\frac{1}{\sqrt{N}})$ *for some positive definite matrix* Σ_Λ.

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

1. $\mathbb{E}(e_{it}) = 0$ *and* $\mathbb{E}|e_{it}|^8 \leq M$ *for all* i *and* t,
2. $\mathbb{E}(e_{it}e_{js}) = \tau_{ij,ts}$ *for all* i, j *and* t, s, *and* $\frac{1}{NT}\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\tau_{ij,ts}| \leq M$,
3. $\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right|^4 \leq M$ *for all* s, t.

Assumption 4 *There exists an* $M < \infty$ *such that:*

1. $\mathbb{E}(\frac{e_s'e_t}{N}) = \gamma_N(s,t)$ *and* $\sum_{s=1}^T |\gamma_N(s,t)| \leq M$ *for all* t,
2. $\mathbb{E}(e_{it}e_{jt}) = \tau_{ij,t}$ *with* $|\tau_{ij,t}| \leq \tau_{ij}$ *for some* τ_{ij} *and for all* t, *and* $\sum_{j=1}^N |\tau_{ji}| \leq M$ *for all* i.

Assumption 5 *The largest eigenvalue of* $\frac{1}{NT}EE'$ *is* $O_p(\frac{1}{\delta_{NT}^2})$.

Assumption 6 *When there is no break, the eigenvalues of* $\Sigma_F\Sigma_\Lambda$ *are distinct. When there are breaks, the eigenvalues of* $\Sigma_G\Sigma_\Gamma$ *are distinct.*

Assumption 7 *Define* $\epsilon_t = vech(f_t f_t' - \Sigma_F)$,

1. *The data generating process of the factors is such that the Hajek-Renyi inequality*[6] *applies to the process* $\{\epsilon_t, t = k_{\kappa-1,0}+1, ..., k_{\kappa,0}\}$ *and* $\{\epsilon_t, t = k_{\kappa,0}, ..., k_{\kappa-1,0}+1\}$ *for* $\kappa = 1, ..., L+1$,

[6]Hajek-Renyi inequality is crucial for pinning down the order of the estimation error in the estimated change points, see the Appendix A for more details.

2. *There exist* $\delta > 0$ *and* $M < \infty$ *such that for* $\kappa = 1, ..., L+1$ *and for all* $k_{\kappa-1,0} < k < l \leq k_{\kappa,0}$, $\mathbb{E}(\left\| \frac{1}{\sqrt{l-k}} \sum_{t=k+1}^{l} \epsilon_t \right\|^{4+\delta}) < M$.

Assumption 8 *There exists* $M < \infty$ *such that:*

1. $\mathbb{E}(\sup_{0\leq k<l\leq T} \frac{1}{l-k} \sum_{t=k+1}^{l} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})] \right|^2) \leq M$ *for all* s,

2. $\mathbb{E}(\sup_{0\leq k<l\leq T} \frac{1}{l-k} \sum_{t=k+1}^{l} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{it} \right\|^2) \leq M$.

Assumption 1 requires the law of large numbers to be applicable to factors within each regime, thus f_t can be dynamic and contain lags. Note that the second moment matrix of the factors is assumed to be stationary over time. Assumption 2 requires the factor loadings to be uniformly bounded and $\frac{1}{N}\Gamma'\Gamma$ (or $\frac{1}{N}\Lambda'\Lambda$) converges to its limit at the speed $O(\frac{1}{\sqrt{N}})$. Assumptions 3 and 4 allow for both temporal and cross-sectional dependence as well as heteroskedasticity. Assumption 5 is the key condition for determining the number of factors and is required in almost all existing methods. For example, Onatski (2010) and Ahn and Horenstein (2013) assume $E = A\varepsilon B$, where ε is an i.i.d. $T \times N$ matrix and A and B characterize the temporal and cross-sectional dependence and heteroskedasticity. This is a sufficient but not necessary condition for Assumption 5. Also note that once Assumption 5 is imposed, Assumption D in Bai (2003) is not needed. In other words, for the purpose of determining the number of factors, factors could be correlated with the errors. Assumption 6 ensures uniqueness of the principal component estimator in large samples.

Assumption 7 imposes a further requirement on the factor process. Instead of assuming a specific data generating process, we require the Hajek-Renyi inequality to be applicable to the second moment process of the factors, so that Assumption 7 is in its most general form. Processes that satisfy Assumption 7 include martingale difference, mixing process and linear process, see Bai (1996). Hajek-Renyi inequality is a more powerful tool than the functional CLT for calculating the stochastic order of sup-type terms. It allows us to calculate the order of $\sup_{m\leq k\leq T} \left| c_k \sum_{t=1}^{k} x_t \right|$ while FCLT only allows us to calculate the order of $\sup_{\tau\leq k/T\leq 1} \left| T^{-\frac{1}{2}} \sum_{t=1}^{k} x_t \right|$, i.e., for Hajek-Renyi

inequality the supremum is taken with respect to all k while for FCLT the supremum is taken with respect to the fraction.

Assumption 8 imposes further constraints on the errors. Assumption 3(3) and Assumption F3 in Bai (2003) imply that the summands are uniformly $O_p(1)$. Assumption 8 strengthens this condition such that the supremum of the average of these summands is $O_p(1)$.

Remark 5 *The assumptions above are the same as or similar to the assumptions in Bai (2003). Assumptions 1, 2, 3, 4 and 6 correspond to Assumptions A, B, C, E and G in Bai (2003), respectively. Assumption 5 replaces Assumption D, and Assumption 8 strengthens Assumption F3 in Bai (2003).*

3.1.3 Asymptotic Properties of the Joint Estimator

First note that due to the consistency of $\tilde{r}$ for $\bar{r}$, treating $\bar{r}$ as known does not affect the asymptotic properties of the change point estimator[7]. Define $\tilde{\tau}_\iota = \tilde{k}_\iota/T$ as the estimated change fraction, we first show that $\tilde{\tau}_\iota$ is consistent.

Proposition 1 *Under Assumptions 1-8,* $\tilde{\tau}_\iota - \tau_{\iota 0} = o_p(1)$ *for* $\iota = 1, ..., L$ *as* $(N, T) \to \infty$.

Remark 6 *For change points estimation, a key observation is that for any possible region of the change points* O, $P((\tilde{k}_1, ..., \tilde{k}_L) \in O)$ *is controlled by* $P(\min_{(k_1,...,k_L)\in\mathcal{O}} \tilde{S}(k_1, ..., k_L) - \tilde{S}(k_{10}, ..., k_{L0}) \leq 0)$. *The proof of Proposition 1 utilizes this observation.*

Proposition 1 establishes the consistency of the estimated change fraction, and serves as an intermediate step for the following theorem.

Theorem 1 *Under Assumptions 1-8,* $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$ *for* $\iota = 1, ..., L$ *as* $(N, T) \to \infty$.

Theorem 1 implies that no matter how large T is, the possible change points are narrowed to a bounded interval of the true change points. Note that the extra error z_t has no effect (asymptotically) on our estimator of the change points as

[7] The proof of the consistency of $\tilde{r}$ for $\bar{r}$ is omitted since under Assumptions 1-5, assumptions in Bai and Ng (2002) are satisfied.

long as $(N,T) \to \infty$. No N-T ratio condition is needed. This is different from factor-augmented forecasting and factor-augmented vector autoregression (FAVAR), in which $\frac{\sqrt{T}}{N} \to 0$ is required to asymptotically eliminate the effect of using estimated factors.

Remark 7 *Identification of the change points relies on observations within a local region of the true change points and consequently the extra error z_t will not accumulate as $T \to \infty$. In contrast, factor-augmented forecasting and FAVAR relies on all observations and consequently the extra error z_t will accumulate as $T \to \infty$. This is why z_t asymptotically has no effect on the estimated change points and no N-T ratio condition is needed.*

Remark 8 *The limiting distribution of $\tilde{k}_\iota - k_{\iota 0}$ has the same form as the single change case. This is because $\tilde{k}_\iota$ also minimizes the sum of squared residuals for the subsample $t = \tilde{k}_{\iota-1}+1, ..., \tilde{k}_{\iota+1}$. Since $\tilde{k}_{\iota-1} - k_{\iota-1,0}$ and $\tilde{k}_{\iota+1} - k_{\iota+1,0}$ are both $O_p(1)$, $\tilde{k}_\iota$ has the same limiting distribution as the minimizer of the subsample $t = k_{\iota-1,0}+1, ..., k_{\iota+1,0}$.*

3.2 Sequential Estimation of the Change Points

This section considers sequential estimation of the change points one by one, each time treating the model as if there is only one change point. The first two steps are the same as the joint estimation while the third step is slightly adjusted: For any partition k_1, split the sample into two subsamples, estimate the second moment matrix of g_t in each subsample and calculate the sum of squared residuals, $\tilde{S}(k_1) = \sum_{\kappa=1}^{2} \sum_{t=k_{\kappa-1}+1}^{k_\kappa} [vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\kappa)]'[vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\kappa)]$, then $\hat{k}_1 = \arg\min \tilde{S}(k_1)$. Compared to joint estimation, the main advantage of sequential estimation is that it does not require knowing the number of changes[8]. Instead, together with sequential testing, it allows us to determine the number of changes.

In what follows, we show that the distance between the sequentially estimated and the true change points is also $O_p(1)$. First, define $S_0(\tau)$ as the reduction in the sum of squared residuals when $y_t = 0$ and $z_t = 0$ is plugged in. If y_t and z_t

[8] Sequential estimation is also computationally simpler.

are indeed zero for all t, the estimated change fraction should be equal to τ among $\tau_{1,0}, ..., \tau_{L,0}$ that leads to the largest reduction in the sum of squared residuals. To simplify the analysis, we require $S_0(\tau_{\iota,0})$ to be different for different ι, and without loss of generality, we assume:

Assumption 9 $S_0(\tau_{1,0}) < ... < S_0(\tau_{L,0})$.

In general, y_t and z_t are not zero for all t, but asymptotically this does not affect the result.

Proposition 2 *Under Assumptions 1-9,* $\hat{\tau}_1 - \tau_{1,0} = o_p(1)$ *as* $(N, T) \to \infty$.

Similar to the joint estimation, Proposition 2 can be refined to:

Theorem 2 *Under Assumptions 1-9,* $\hat{k}_1 - k_{1,0} = O_p(1)$ *as* $(N, T) \to \infty$.

Again, no N-T ratio condition is needed to eliminate the effect of the extra error z_t. Once $\hat{k}_1$ is available, we can plug it in and estimate $k_{2,0}$. Since $\hat{k}_1 - k_{1,0} = O_p(1)$, it can be shown that this is asymptotically equivalent to plugging in $k_{1,0}$, in which case the problem is reduced to estimating the first change point with observations $t = 1, ..., k_{1,0}$ removed[9]. Thus $\hat{k}_2 - k_{2,0}$ will also be $O_p(1)$. Using this argument sequentially, we have

Theorem 3 *Under Assumptions 1-9,* $\hat{k}_\iota - k_{\iota,0} = O_p(1)$ *for* $\iota = 1, ..., L$ *as* $(N, T) \to \infty$.

Note that Theorems 1-3 require the change fractions $\tau_{\kappa,0}$ to be positive and different. Theorems 1-3 no longer hold if $k_{\kappa,0} - k_{\kappa-1,0} = o(T)$.

[9]In the general case, $\hat{k}_1$ could converge to the change point in the middle of the sample. Then the problem is reduced to estimating the first change point for subsamples $t = 1, ..., k_{1,0}$ and $t = k_{1,0}+1, ..., T$ and taking $\hat{k}_2$ as the one leading to the largest reduction in the sum of squared residuals.

3.3 Estimating the Number of Factors and the Factor Space

Once the change points estimators are available, we plug them in and estimate the number of factors and factor space in each regime. If true change points are plugged in, consistency of the estimated number of factors and convergence rate of the estimated factor space are well established. Thus the main concern is the effect of using estimated change points. We show that although the estimated change points are inconsistent, this effect is asymptotically negligible.

Let $\tilde{r}_\kappa$ and r_κ be the estimated (using the method in Bai and Ng (2002) or Ahn and Horenstein (2013)) and the true number of factors in the κ-th regime.

Theorem 4 *Under Assumptions 1-2 and 5, with* $\tilde{k}_\kappa - k_{\kappa,0} = O_p(1)$ *and* $\tilde{k}_{\kappa-1} - k_{\kappa-1,0} = O_p(1)$*, we have* $\lim_{(N,T)\to\infty} P(\tilde{r}_\kappa = r_\kappa) = 1$.

Next, let u_κ be some positive integer, $\tilde{F}_\kappa^{u_\kappa}$ be $\sqrt{T}$ times the eigenvectors corresponding to the first u_κ eigenvalues of $X_\kappa X_\kappa'$, $H_\kappa^{u_\kappa} = \frac{1}{N}\Lambda_{0\kappa}'\Lambda_{0\kappa}\frac{1}{\tilde{k}_\kappa - \tilde{k}_{\kappa-1}}F_\kappa\tilde{F}_\kappa^{u_\kappa}$ and $\hat{F}_\kappa^{u_\kappa} = \tilde{F}_\kappa^{u_\kappa}V_{NT,\kappa}^{u_\kappa}$, where $X_\kappa = (x_{\tilde{k}_{\kappa-1}+1}, ..., x_{\tilde{k}_\kappa})'$, $F_\kappa = (f_{\tilde{k}_{\kappa-1}+1}, ..., f_{\tilde{k}_\kappa})'$and $V_{NT,\kappa}^{u_\kappa}$ is the diagonal matrix that contains the first u_κ eigenvalues of $X_\kappa X_\kappa'$.

Theorem 5 *Under Assumptions 1-4, with* $\tilde{k}_\kappa - k_{\kappa,0} = O_p(1)$ *and* $\tilde{k}_{\kappa-1} - k_{\kappa-1,0} = O_p(1)$*, we have*

$$\frac{1}{\tilde{k}_\kappa - \tilde{k}_{\kappa-1}}\sum\nolimits_{t=\tilde{k}_{\kappa-1}+1}^{\tilde{k}_\kappa}\left\|\hat{f}_t^{u_\kappa} - H_\kappa^{u_\kappa\prime}f_t\right\|^2 = O_p(\frac{1}{\delta_{NT}^2}). \tag{6}$$

The convergence rate $O_p(\frac{1}{\delta_{NT}^2})$ is crucial to eliminate the effect of using estimated factors in factor-augmented forecasting and FAVAR. In the next section we will use the estimated factors to construct a test for l versus $l+1$ changes. We show that the rate $O_p(\frac{1}{\delta_{NT}^2})$ is also crucial in eliminating the effect of using estimated factors on the limiting distribution of the test statistic.

Remark 9 *The proof for Theorem 4 and Theorem 5 are similar to the single change case, see Baltagi et al. (2017).*

Remark 10 *Theorem 4 and Theorem 5 rely on* $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$*. Consistency of the estimated change fractions is not enough.*

3.4 Implications for Factor-augmented Forecasting

Consider the following factor-augmented regression model:

$$y_{t+h} = \alpha' f_t + \beta' W_t + \epsilon_{t+h}, \tag{7}$$

where W_t contains some observable regressors relevant for forecasting, f_t contains unobservable factors and will be estimated from x_{it}, and h is the lead time between the dependent variable and information available. When there is no break, f_t will be replaced by the estimated factors to estimate α and β, and forecasts can be constructed based on the estimated factors, W_T, $\hat{\alpha}$ and $\hat{\beta}$. When there are breaks in factor loadings, we have two choices to handle the breaks.

The first choice is to plug in the estimated change points and estimate the factor space in each regime. Let u_κ in Theorem 5 be the true number of factors. Theorem 5 implies that using $\hat{f}_t$ is equivalent to using $H_\kappa' f_t$ in the κ-th regime. Since $y_{t+h} = \alpha'(H_\kappa^{-1})' H_\kappa' f_t + \beta' W_t + \epsilon_{t+h}$ in the κ-th regime and H_κ is different for different κ, we need to allow α to have breaks at the estimated break points $(\tilde{k}_1, ..., \tilde{k}_L)$ in estimating the forecasting model.

The second choice is to ignore the breaks and obtain the estimated pseudo factors $\tilde{g}_t$, and then use $\tilde{g}_t$ to construct forecasts. Since $\tilde{g}_t - J' g_t$ is asymptotically negligible and $g_t = R_\kappa f_t$ for $k_{\kappa-1,0} + 1 \leq t \leq k_{\kappa,0}$, this is equivalent to using $J' R_\kappa f_t$ for $k_{\kappa-1,0} + 1 \leq t \leq k_{\kappa,0}$.

If there is no zero column in Λ_κ and columns in Λ_0 and all Λ_κ are linearly independent, then there exists an $\bar{r}$ dimensional θ such that $\theta' J' R_\kappa = \alpha'$ for all κ. For example, if $\Lambda_0 = 0$ and there are two breaks, we have $\theta = (\alpha', \alpha', \alpha')'$. Thus for this case, equation (7) can be written as $y_{t+h} = \theta' J' g_t + \beta' W_t + \epsilon_{t+h}$, and there is no need to consider structural breaks for the forecasting model.

If there is no zero column in Λ_κ but columns in Λ_0 and Λ_κ are linearly dependent, then using $\tilde{g}_t$ will induce breaks in the forecasting model. For example, if $\Lambda_0 = 0$, there is only one break and $\Lambda_2 = 2\Lambda_1$, then $g_t = f_t$ and $g_t = 2f_t$ in the first and the second regime respectively. It follows that $y_{t+h} = \alpha' g_t + \beta' W_t + \epsilon_{t+h}$ in the first regime and $y_{t+h} = \frac{1}{2}\alpha' g_t + \beta' W_t + \epsilon_{t+h}$ in the second regime. Thus for this case we need to

allow α to have breaks at $(\tilde{k}_1, ..., \tilde{k}_L)$.

If there are zero columns in Λ_κ, then α and β are not identifiable even if we know the change points, because when a certain column of Λ_κ is zero, the corresponding factor can not be estimated. However, it is also high likely that this factor does not appear in the forecasting model either. If this is true, then we need to allow α to have breaks when we estimate the forecasting model[10].

In summary, since we do not know the specific form of structural breaks in the factor loadings, we need to consider breaks in α for the forecasting model. The number and locations of breaks of α are the same as those of the factor loadings. We can also use $\tilde{g}_t$ for f_t and apply Bai and Perron's test directly to (7) to detect and estimate the change points of α. The change points estimated in this way also have bounded errors. Also, it's worth pointing out that using $\tilde{g}_t$ (or $\hat{f}_t$) and the full sample to construct forecasts is better than simply using the last subsample because the full sample estimator of β is more accurate.

4 TESTING MULTIPLE CHANGES

In this section we propose two tests for multiple changes. The first one tests no change ($L = 0$) versus some fixed number of changes ($L = l$). The second one tests l versus $l + 1$ changes, and together with sequential estimation of the change points, can be used to determine the number of changes.

Our testing procedure is the same as Bai and Perron (1998), but the construction of the test statistics are slightly different, because our tests are based on a vector (multivariate) process. The main concern is the effect of using estimated factors on the asymptotic and finite sample performance of the test statistics.

4.1 Construction of the Test for $L = 0$ versus $L = l$

First, estimate the number of factors and then estimate the factor space by principal components. Under the null, let $\tilde{f}_t$ be the estimated factors, U_{NT} be the diagonal matrix that contains the r largest eigenvalues of XX', $H = \frac{1}{N}\Lambda'\Lambda\frac{1}{T}F'\tilde{F}U_{NT}^{-1}$ be the

[10] This also applies when we use $\hat{f}_t$ to estimate the forecasting model.

rotation matrix, H_0 be the probability limit of H and $z_t^* = vech(\tilde{f}_t\tilde{f}_t' - H_0'f_tf_t'H_0)$. Under the alternative, we follow the same notation as the last section. It follows that under the null $vech(\tilde{f}_t\tilde{f}_t')$ is a multivariate time series ($vech(H_0'f_tf_t'H_0)$) with stable mean ($vech(I_r)$)[11] and extra error z_t^*, while under the alternative $vech(\tilde{g}_t\tilde{g}_t')$ is a multivariate time series with l mean shifts and extra error z_t. Thus we can base the test on the difference between the restricted and unrestricted sum of squared normalized error.

Let $\Omega = \lim_{T\to\infty} Var(vech(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(H_0'f_tf_t'H_0 - I_r)))$ be the long run covariance matrix of $vech(H_0'f_tf_t'H_0 - I_r)$ and $\tilde{\Omega}(\tilde{F}) = \tilde{\Upsilon}_0(\tilde{F}) + \sum_{j=1}^{T-1}\ker(\frac{j}{d_T})[\tilde{\Upsilon}_j(\tilde{F}) + \tilde{\Upsilon}_j(\tilde{F})']$ be the HAC estimator of Ω using the estimated factors $\tilde{F}$, where $\tilde{\Upsilon}_j(\tilde{F}) = \frac{1}{T}\sum_{t=j+1}^{T} vech(\tilde{f}_t\tilde{f}_t' - I_{\tilde{r}})vech(\tilde{f}_{t-j}\tilde{f}_{t-j}' - I_{\tilde{r}})'$, $\ker(\cdot)$ is some kernel function and d_T is the bandwidth. For simplicity, we will suppress $\tilde{\Omega}(\tilde{F})$ as $\tilde{\Omega}$. It follows that the restricted sum of squared normalized error is

$$SSNE_0 = \sum_{t=1}^{T} vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t')'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t'), \qquad (8)$$

and for any partition $(k_1, ..., k_l)$, the unrestricted sum of squared normalized error is

$$\begin{aligned} SSNE(k_1, ..., k_l) \;=\; & \sum_{\iota=1}^{l+1}\sum_{t=k_{\iota-1}+1}^{k_\iota} vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota} \\ & \tilde{f}_t\tilde{f}_t')'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - \frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota}\tilde{f}_t\tilde{f}_t'). \qquad (9) \end{aligned}$$

Let $F_{NT}(\tau_1, ..., \tau_l; \frac{\tilde{r}(\tilde{r}+1)}{2}) = \frac{2}{l\tilde{r}(\tilde{r}+1)}[SSNE_0 - SSNE(k_1, ..., k_l)]$ and $\Lambda_\epsilon = \{(\tau_1, ..., \tau_l) : |\tau_{\iota+1} - \tau_\iota| \geq \epsilon,\ \tau_1 \geq \epsilon,\ \tau_l \leq 1 - \epsilon\}$ for some prespecified $\epsilon > 0$, the test statistic is

$$\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} F_{NT}(\tau_1, ..., \tau_l; \frac{\tilde{r}(\tilde{r}+1)}{2}).$$

For the kernel function $\ker(\cdot)$ and bandwidth d_T, we consider three popular choices:

1. Bartlett kernel with $d_T = O(T^{\frac{1}{3}})$.

2. Parzen kernel with $d_T = O(T^{\frac{1}{5}})$.

[11] It is not difficult to see that $E(H_0'f_tf_t'H_0) = H_0'\Sigma_F H_0 = I_r$.

3. Quadratic spectral kernel with $K_1 T^{\frac{1}{5}} \leq d_T \leq K_2 T^{\frac{1}{5}}$ for some $K_1, K_2 > 0$.

4.2 Construction of the Test for $L = l$ versus $L = l + 1$

First, we estimate l change points and plug them in to estimate the number of factors and factor space in each regime. Then testing $L = l$ versus $L = l + 1$ is equivalent to testing no change versus a single change in each regime jointly. The main concern is the effect of using estimated change points and estimated factors on the limiting distribution and consistency of the test statistic.

Let $\tilde{k}_1, ..., \tilde{k}_l$ be the estimated change points and $\tilde{r}_\iota$ be the estimated number of factors in the ι-th regime. Under the null, let $\tilde{F}_\iota = (\tilde{f}_{\iota,\tilde{k}_{\iota-1}+1}, ..., \tilde{f}_{\iota,\tilde{k}_\iota})'$ be the estimated factors, H_ι be the rotation matrix, $H_{\iota 0}$ be the limit of H_ι, $U_{\iota NT}$ be the eigenvalue matrix, U_ι be the limit of $U_{\iota NT}$, $F_\iota = (f_{\iota,\tilde{k}_{\iota-1}+1}, ..., f_{\iota,\tilde{k}_\iota})'$ and $F_{\iota 0} = (f_{\iota,k_{\iota-1,0}+1}, ..., f_{\iota,k_{\iota 0}})'$. Note that $f_{\iota,t}$ is r_ι dimensional and contains the factors that appear in the ι-th regime. Under the alternative, there are $l + 1$ changes and the l estimated change points will be close to $(O_p(1))$ the l points that allow the greatest reduction in the sum of squared normalized errors. Without loss of generality, suppose $\tilde{k}_{\iota-1} - k_{\iota-1,0} = O_p(1)$ and $\tilde{k}_\iota - k_{\iota+1,0} = O_p(1)$ for some ι. In this case, the ι-th regime contains an extra change point[12] $k_{\iota,0}$ but can be equivalently represented as having no changes but with pseudo factors $g_{\iota t}$, where $g_{\iota t} = A_{\iota 1} f_t$ for $t \in [\tilde{k}_{\iota-1} + 1, ..., k_{\iota 0}]$ and $g_{\iota t} = A_{\iota 2} f_t$ for $t \in [k_{\iota 0} + 1, ..., \tilde{k}_\iota]$. For this regime, we denote the estimated factors as $\tilde{g}_{\iota t}$ and define $\tilde{G}_\iota$, G_ι, $G_{\iota 0}$, J_ι, $J_{\iota 0}$, $V_{\iota NT}$ and V_ι correspondingly as $\tilde{F}_\iota$, F_ι, $F_{\iota 0}$, H_ι, $H_{\iota 0}$, $U_{\iota NT}$ and U_ι. For the other regimes, we maintain the same notation. It follows that under the null $vech(\tilde{f}_{\iota t}\tilde{f}'_{\iota t})$ is a multivariate time series with stable mean and extra error $z^*_{\iota t}$ for all ι while under the alternative $vech(\tilde{g}_{\iota t}\tilde{g}'_{\iota t})$ is a multivariate time series with a mean shift and extra error $z_{\iota t}$ for some ι. Again, the test is based on the difference between the restricted and unrestricted sum of squared normalized error.

Let $\Omega_\iota = \lim_{T\to\infty} Var(vech(\frac{1}{\sqrt{k_{\iota,0}-k_{\iota-1,0}}} \sum_{t=k_{\iota-1,0}+1}^{k_{\iota,0}} (H'_{\iota 0} f_{\iota t} f'_{\iota t} H_{\iota 0} - I_{r_\iota})))$ be the long run covariance matrix of $vech(H'_{\iota 0} f_{\iota t} f'_{\iota t} H_{\iota 0} - I_{r_\iota})$ and $\tilde{\Omega}_\iota$ be the HAC estimator of Ω_ι

[12] When $\tilde{k}_{\iota-1} < k_{\iota-1,0}$ or $\tilde{k}_\iota > k_{\iota+1,0}$, the ι-th regime also contains the change point $k_{\iota-1,0}$ or $k_{\iota+1,0}$, but with $\tilde{k}_{\iota-1} - k_{\iota-1,0} = O_p(1)$ and $\tilde{k}_\iota - k_{\iota+1,0} = O_p(1)$ these two are asymptotically ignorable.

using $\tilde{F}_\iota$ and with kernel function $\ker(\cdot)$ and bandwidth d_T. The test statistic is

$$F_{NT}(l+1\,|l) = SSNE(\tilde{k}_1,...,\tilde{k}_l) - \min_{1\le\iota\le l+1}\inf_{k\in\Lambda_{\iota,\eta}} SSNE(\tilde{k}_1,...,\tilde{k}_{\iota-1},k,\tilde{k}_\iota,...,\tilde{k}_l), \quad (10)$$

where $SSNE(\tilde{k}_1,...,\tilde{k}_l)$ is the restricted sum of squared normalized error and equals

$$\begin{aligned}\sum_{\iota=1}^{l+1} SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota) &= \sum_{\iota=1}^{l+1}\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota} vech(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - \frac{1}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota} \\ &\quad \tilde{f}_{\iota t}\tilde{f}_{\iota t}')'\tilde{\Omega}_\iota^{-1} vech(\tilde{f}_{\iota t}\tilde{f}_{\iota t}' - \frac{1}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}\tilde{f}_{\iota t}\tilde{f}_{\iota t}'), \quad (11)\end{aligned}$$

$SSNE(\tilde{k}_1,...,\tilde{k}_{\iota-1},k,\tilde{k}_\iota,...,\tilde{k}_l)$ is the unrestricted sum of squared normalized error and equals

$$\sum_{\kappa=1}^{\iota-1} SSNE_\kappa(\tilde{k}_{\kappa-1},\tilde{k}_\kappa) + SSNE_\iota(\tilde{k}_{\iota-1},k,\tilde{k}_\iota) + \sum_{\kappa=\iota+1}^{l+1} SSNE_\kappa(\tilde{k}_{\kappa-1},\tilde{k}_\kappa), \quad (12)$$

with $\Lambda_{\iota,\eta} = \{k : \tilde{k}_{\iota-1} + (\tilde{k}_\iota - \tilde{k}_{\iota-1})\eta \le k \le \tilde{k}_\iota - (\tilde{k}_\iota - \tilde{k}_{\iota-1})\eta\}$.

4.3 Asymptotic Properties of the Test Statistics

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

1. $\mathbb{E}(\left\| \frac{1}{\sqrt{NT}}\sum_{s=k_{\iota-1,0}+1}^{k_{\iota,0}}\sum_{i=1}^{N} f_s[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right\|^2) \le M$ *for all* t *and* $\iota = 1,...,L+1$,
2. $\mathbb{E}(\left\| \frac{1}{\sqrt{NT}}\sum_{t=k_{\iota-1,0}+1}^{k_{\iota,0}}\sum_{i=1}^{N} f_t\lambda_{0\iota,i}'e_{it}\right\|^2) \le M$ *for* $\iota = 1,...,L+1$,
3. $\mathbb{E}(\left\| \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_{0\iota,i}e_{it}\right\|^2) \le M$ *for all* $k_{\iota-1,0} < t \le k_{\iota,0}$ *and* $\iota = 1,...,L+1$.

Assumption 11 *For any* $\epsilon > 0$ *and* $\iota = 1,...,L+1$,

1. $\sup_{(k_{\iota,0}-k_{\iota-1,0})\epsilon\le k-k_{\iota-1,0}\le(k_{\iota,0}-k_{\iota-1,0})(1-\epsilon)} \left\| \frac{1}{\sqrt{NT}}\sum_{t=k_{\iota-1,0}+1}^{k}\sum_{i=1}^{N} f_t\lambda_{0\iota,i}'e_{it}\right\| = O_p(1)$,
2. $\sup_{(k_{\iota,0}-k_{\iota-1,0})\epsilon\le k-k_{\iota-1,0}\le(k_{\iota,0}-k_{\iota-1,0})(1-\epsilon)} \left\| \frac{1}{\sqrt{NT}}\sum_{t=k+1}^{k_{\iota,0}}\sum_{i=1}^{N} f_t\lambda_{0\iota,i}'e_{it}\right\| = O_p(1)$.

Assumption 12 *For $\iota = 1, ..., L+1$, Ω_ι is positive definite and*

$$\frac{1}{\sqrt{k_{\iota,0} - k_{\iota-1,0}}} \sum_{t=k_{\iota-1,0}+1}^{k_{\iota-1,0}+(k_{\iota,0}-k_{\iota-1,0})\tau} vech[\Omega_\iota^{-\frac{1}{2}}(H_{\iota 0}' f_{\iota t} f_{\iota t}' H_{\iota 0} - I_{r_\iota})] \Rightarrow W_{\frac{r_\iota(r_\iota+1)}{2}}(\tau),$$

where $W_{\frac{r_\iota(r_\iota+1)}{2}}(\cdot)$ is an $\frac{r_\iota(r_\iota+1)}{2}$ dimensional vector of independent Wiener processes on $[0,1]$.

Assumption 13 *For $\iota = 1, ..., L+1$, let $\tilde{\Omega}(F_\iota H_{\iota 0})$ be the HAC estimator of Ω_ι using $F_\iota H_{\iota 0}$, $\tilde{\Omega}(F_\iota H_{\iota 0})$ is consistent for Ω_ι. When $L = 0$, $\tilde{\Omega}(F H_0)$ is consistent for Ω.*

Assumption 14 *For $\iota = 1, ..., L+1$, the eigenvalues of $\Sigma_{F,\iota}\Sigma_{\Lambda,\iota}$ are distinct, the eigenvalues of $\Sigma_{G,\iota}\Sigma_{\Gamma^\iota}$ are also distinct. ($\Sigma_{F,\iota} = plim \frac{1}{k_{\iota,0}-k_{\iota-1,0}} \sum_{t=k_{\iota-1,0}+1}^{k_{\iota,0}} f_{\iota t} f_{\iota t}'$, $\Sigma_{G,\iota} = plim \frac{1}{k_{\iota,0}-k_{\iota-1,0}} g_{\iota t} g_{\iota t}'$, $\Sigma_{\Lambda^\iota} = plim \frac{1}{N} \Lambda^\iota \Lambda^{\iota\prime}$, $\Sigma_{\Gamma^\iota} = plim \frac{1}{N} \Gamma^\iota \Gamma^{\iota\prime}$, Λ^ι contains the nonzero factor loadings of the ι-th regime, Γ^ι contains the linearly independent vectors of factor loadings of the ι-th and $(\iota+1)$-th regime.)*

Note that when $L = 0$, $k_{0,0} = 0$ and $k_{1,0} = T$, $\lambda_{0\iota,i}$ is replaced by λ_i in Assumptions 10 and 11, and in Assumption 12 $f_{\iota t}$, Ω_ι and r_ι is replaced by f_t, Ω and r respectively.

Assumption 10 corresponds to and slightly weakens Assumption F in Bai (2003). Assumption 11 requires the term in $\|\cdot\|$ to be uniformly $O_p(1)$. This is not restrictive since all summands have zero means. Assumptions 10 and 11 are satisfied by various mixing processes. Assumptions 12 requires the functional central limit theorem to be applicable to $vech(H_{\iota 0}' f_{\iota t} f_{\iota t}' H_{\iota 0} - I_{r_\iota})$ in each regime. Assumptions 13 requires the HAC estimator of Ω_ι to be consistent if factors were observable. Assumptions 14 ensures that for each regime, no matter whether there is break or not, the principal component estimator is unique in large samples.

4.3.1 Asymptotic Properties of the Test for $L = 0$ versus $L = l$

Now we are ready to present the limiting distribution:

Theorem 6 *Under Assumptions 1-6, 10-13 and $L = 0$, with $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{\delta_{NT}} \to 0$ as $(N,T) \to \infty$,*

$$\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} F_{NT}(\tau_1, ..., \tau_l; \frac{\tilde{r}(\tilde{r}+1)}{2}) \xrightarrow{d} \sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon} F(\tau_1, ..., \tau_l; \frac{r(r+1)}{2}),$$

where $F(\tau_1,...,\tau_l;\frac{r(r+1)}{2}) = \frac{2}{lr(r+1)}\sum_{\iota=1}^{l}\frac{\left\|\tau_\iota W_{\frac{r(r+1)}{2}}(\tau_{\iota+1})-\tau_{\iota+1}W_{\frac{r(r+1)}{2}}(\tau_\iota)\right\|^2}{\tau_\iota\tau_{\iota+1}(\tau_{\iota+1}-\tau_\iota)}$.

Note that $\frac{\sqrt{T}}{N}\to 0$ and $\frac{d_T}{\delta_{NT}}\to 0$ are needed to eliminate the effect of the extra error z_t^*. This is different from the results in the last section but similar to the results in the factor-augmented forecasting and FAVAR. Intuitively, testing for structural changes relies on all the observations and consequently z_t^* will accumulate in the test statistic as $T\to\infty$ and $d_T\to\infty$.

We next consider the consistency of the proposed test. Under the alternative, the process $vech(\tilde{g}_t\tilde{g}_t')$ has l mean shifts and extra error z_t. Thus $vech(\tilde{g}_t\tilde{g}_t')$ is not properly demeaned in calculating the restricted $SSNE$. On the other hand, the test statistic can be written as $\frac{2}{l\tilde{r}(\tilde{r}+1)}[SSNE_0-\min_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}SSNE(k_1,...,k_l)]$ and by taking the minimum for $(\tau_1,...,\tau_l)\in\Lambda_\epsilon$, it ensures $vech(\tilde{g}_t\tilde{g}_t')$ is properly demeaned. Thus under the alternative the test statistic will diverge as $(N,T)\to\infty$.

Theorem 7 *Under Assumptions 1-8 and $L=l$, with $\frac{d_T}{T}\to 0$ as $(N,T)\to\infty$,* $\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}F_{NT}(\tau_1,...,\tau_l;\frac{\tilde{r}(\tilde{r}+1)}{2})\xrightarrow{p}\infty$.

The test discussed above is designed for a given number of changes under the alternative. When the number of changes is misspecified, the test may not be powerful. For example, test for 0 versus 2 changes should be more powerful than the test for 0 versus 1 change when the true DGP contains two changes. Following Bai and Perron (1998), we consider the UDmax and WDmax tests when the number of changes under the alternative is unknown. Let $c(q,\alpha,l)$ be the asymptotic critical value of the test for 0 versus l changes with degree of freedom q and significance level α. Given the maximum possible number of changes M, UDmax is simply the maximum of the tests for 0 versus l changes with $l\le M$ while WDmax is the weighted maximum of the tests for 0 versus l changes with weights $c(\frac{\tilde{r}(\tilde{r}+1)}{2},\alpha,1)/c(\frac{\tilde{r}(\tilde{r}+1)}{2},\alpha,l)$. With Theorem 6, the limiting distributions of both tests have the same form as in Bai and Perron (1998).

Remark 11 *Comprehensive critical values for Theorem 6 and the UDmax and WDmax tests are tabulated in Bai and Perron (2003).*

Remark 12 *Since $\tilde{r}$ is consistent for r and $\bar{r}$ under the null and the alternative respectively, in the asymptotic analysis we can treat r and $\bar{r}$ as known.*

Remark 13 *Now consider the finite sample effect of $\tilde{r} \neq r$ (or $\bar{r}$) on the performance of the test. Underestimation of the number of factors will not affect the size of the test but will decrease the power, because under the null the degrees of freedom $\frac{\tilde{r}(\tilde{r}+1)}{2}$ (and consequently the critical value) adjust automatically with the estimated number of factors $\tilde{r}$, while under the alternative important second moment conditions of the pseudo factors will be lost. Overestimation of the number of factors will not significantly affect the power because all second moment conditions are utilized. However, overestimation will make the test undersized if the errors are stationary, because under the null it will magnify the degrees of freedom but will not magnify the test statistic. If the errors are heteroscedastic, overestimation may introduce breaks from the errors.*

4.3.2 Asymptotic Properties of the Test for $L = l$ versus $L = l + 1$

If the true change points were plugged in, Theorem 6 implies that for each regime the effect of using estimated factors can be eliminated if $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{\delta_{NT}} \to 0$. When the estimated change points are plugged in, we show based on Theorem 4 and Theorem 5 that the result still holds if $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{T^{\frac{1}{4}}} \to 0$.

Theorem 8 *Under Assumptions 1-6, 10-14 and $L = l$, with $\tilde{k}_\iota - k_{\iota,0} = O_p(1)$ for all ι, $\frac{\sqrt{T}}{N} \to 0$ and $\frac{d_T}{T^{\frac{1}{4}}} \to 0$, we have $F_{NT}(l+1\,|l) \xrightarrow{d} \sup_{1\leq\iota\leq l+1} F_\iota$, where $F_\iota = \sup_{\eta\leq\tau\leq(1-\eta)} \frac{1}{\tau(1-\tau)} \left\| W_{\frac{r_\iota(r_\iota+1)}{2}}(\tau) - \tau W_{\frac{r_\iota(r_\iota+1)}{2}}(1) \right\|^2$ and F_ι is independent with each other for different ι.*

Critical values can be obtained via simulations and here they are related to the number of factors in each regime. In case the number of factors is stable, we have:

Corollary 1 *If $r_\iota = r$ for all ι, $\lim_{(N,T)\to\infty} P(F_{NT}(l+1\,|l) \leq x) = G_{\frac{r(r+1)}{2},\eta}(x)^{l+1}$, where $G_{\frac{r(r+1)}{2},\eta}(x)$ is the c.d.f. of $\sup_{\eta\leq\tau\leq(1-\eta)} \frac{1}{\tau(1-\tau)} \left\| W_{\frac{r(r+1)}{2}}(\tau) - \tau W_{\frac{r(r+1)}{2}}(1) \right\|^2$.*

We next consider the consistency of $F_{NT}(l+1\,|l)$. Since

$$\begin{aligned} F_{NT}(l+1\,|l) &= \sup_{1\leq\kappa\leq l+1}\sup_{k\in\Lambda_{\kappa,\eta}}[SSNE_\kappa(\tilde{k}_{\kappa-1},\tilde{k}_\kappa)-SSNE_\kappa(\tilde{k}_{\kappa-1},k,\tilde{k}_\kappa)] \\ &\geq SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota)-SSNE_\iota(\tilde{k}_{\iota-1},k_{\iota 0},\tilde{k}_\iota) \end{aligned}$$

and under the alternative $SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota)$ is not properly demeaned, $F_{NT}(l+1\,|l)$ will diverge as $(N,T)\to\infty$.

Theorem 9 *Under Assumptions 1-5, 10-11, 14 and $L=l+1$, with $\left|\tilde{k}_\iota-k_{\iota+1,0}\right|=O_p(1)$ and $\left|\tilde{k}_{\iota-1}-k_{\iota-1,0}\right|=O_p(1)$ for some ι and $\frac{d_T}{T}\to 0$, we have $F_{NT}(l+1\,|l)\xrightarrow{p}\infty$.*

Remark 14 *Since $\tilde{r}_\iota$ is consistent for r_ι under the null, in the asymptotic analysis we can treat r_ι as known.*

Remark 15 *For the finite sample effect of $\tilde{r}_\iota$ on $F_{NT}(l+1\,|l)$, the discussion in Remark 13 also applies here.*

4.4 Determining the Number of Changes

The sequential test $F_{NT}(l+1\,|l)$ allows us to determine the number of changes. First, estimate l change points, either jointly or sequentially, where l could be suggested by some prior information or just zero. Next, perform the test $F_{NT}(l+1\,|l)$. If rejected[13], estimate $l+1$ change points, either jointly or sequentially, and then perform the test $F_{NT}(l+2\,|l+1)$. Repeat this procedure until the null can not be rejected. Let $\hat{L}$ be the estimated number of changes, it is not difficult to see that $\lim_{(N,T)\to\infty}P(\hat{L}<L)=0$ and $\lim_{(N,T)\to\infty}P(\hat{L}\geq L+1)=\alpha$. let $\alpha\to 0$ as $(N,T)\to\infty$, then $\hat{L}$ will be consistent.

Remark 16 *For the error accumulation issue of our multi-step testing procedure, note that the estimator of the number of factors and the factor space is robust to bounded error of $\tilde{k}_\iota-k_{\iota,0}$, and as explained in Remarks 13 and 15, our test is to some degree robust to $\tilde{r}_\iota\neq r_\iota$.*

[13] It can be shown that the test is also consistent when $L>l+1$.

5 MONTE CARLO SIMULATIONS

This section presents simulation results to evaluate the finite sample properties of our proposed estimation and testing procedures. The number of simulations is 1000.

5.1 Data Generating Process

The factors are generated by

$$f_{t,p} = \rho f_{t-1,p} + u_{t,p} \text{ for } t = 2, ..., T \text{ and } p = 1, ..., 3,$$

where $u_t = (u_{t,1}, u_{t,2}, u_{t,3})'$ is i.i.d. $N(0, I_3)$ for $t = 2, ..., T$ and $f_1 = (f_{1,1}, f_{1,2}, f_{1,3})'$ is i.i.d. $N(0, \frac{1}{1-\rho^2} I_3)$ so that the factors are stationary. The idiosyncratic errors are generated by:

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

where $v_t = (v_{1,t}, ..., v_{N,t})'$ is i.i.d. $N(0, \Omega)$ for $t = 2, ..., T$ and $e_1 = (e_{1,1}, ..., e_{N,1})'$ is $N(0, \frac{1}{1-\alpha^2}\Omega)$ so that the idiosyncratic errors are stationary. Ω is generated as $\Omega_{ij} = \beta^{|i-j|}$ so that β captures the degree of cross-sectional dependence of the idiosyncratic errors. In addition, u_t and v_t are mutually independent for all t.

For factor loadings, we consider two different setups. Setup 1 contains no structural change and λ_i is i.i.d. $N(0, \frac{1}{3} I_3)$ across i. Setup 1 will be used to evaluate the size of the tests for multiple changes. Setup 2 contains two structural changes and hence three regimes. In the first and the second regime, the last element of $\lambda_{1,i}$ and $\lambda_{2,i}$ are zeros for all i while the first two elements of $\lambda_{1,i}$ and $\lambda_{2,i}$ are both i.i.d. $N(0, \frac{1}{2} I_2)$ across i. In the third regime, $\lambda_{3,i}$ is i.i.d. $N(0, \frac{1}{3} I_3)$ across i. Also, $\lambda_{1,i}$, $\lambda_{2,i}$ and $\lambda_{3,i}$ are independent. Thus in Setup 2 the number of factors in the three regimes are 2, 2, 3 respectively and the number of pseudo factors is 7. Setup 2 will be used to evaluate the performance of the estimated change points and the estimated number of factors in each regime. Setup 3 also contains two structural changes while $\lambda_{1,i}$, $\lambda_{2,i}$ and $\lambda_{3,i}$ are all i.i.d. $N(0, \frac{1}{3} I_3)$ across i and independent of each other. Setup 3 will be used to evaluate the power of the tests for multiple changes and the probabilities

of selecting the correct number of changes. Once factors, loadings and errors are available, the data is generated as:

$$\begin{aligned} \text{Setup 1:} \quad & x_{it} = f_t'\lambda_i + e_{it}, \\ \text{Setup 2 and 3:} \quad & x_{it} = f_t'\lambda_{\kappa,i} + e_{it}, \text{if } [T\tau_{\kappa-1,0}] + 1 \leq t \leq [T\tau_{\kappa,0}] \text{ for } \kappa = 1, 2, 3, \end{aligned}$$

where $(\tau_{1,0}, \tau_{2,0}) = (0.3, 0.7)$ are the change fractions. Finally, all factor loadings are independent of the factors and the idiosyncratic errors.

5.2 Estimating the Change Points

We first estimate the number of pseudo factors using IC_{p1} in Bai and Ng (2002) with the maximum number of factors $rmax = 12$. When using other criterion, e.g., IC_{p2}, IC_{p3} in Bai and Ng (2002) and ER, GR in Ahn and Horenstein (2013), the results are similar, and hence omitted. Once estimated pseudo factors are available, the change points are estimated as in equation (5) with minimum sample size of each regime $T \times 0.1$.

Figures 1 and 2 are the histograms of the jointly estimated change points for $(N, T) = (100, 100)$ and $(N, T) = (100, 200)$ respectively. Each figure includes four subfigures corresponding to $(\rho, \alpha, \beta) = (0, 0, 0)$, $(0.7, 0, 0)$, $(0, 0.3, 0)$ and $(0, 0, 0.3)$ respectively. In all subfigures, more than 95 percent of the mass is concentrated within a (-8,8) neighborhood of the true change points. This confirms our theoretical result that $\tilde{k}_\kappa - k_{\kappa,0} = O_p(1)$. Figures 1 and 2 also show that the performance of the estimated change points deteriorates when ρ increases from 0 to 0.7 while serial correlation and cross-sectional dependence of the errors seems to have no effect. This is also in line with the theoretical predictions because the errors affect the estimation of the pseudo factors and not the estimation of the change points directly.

5.3 Estimating the Number of Factors in Each Regime[14]

The number of factors in each regime is estimated using IC_{p2} in Bai and Ng (2002) and ER and GR in Ahn and Horenstein (2013), with maximum number of factors 8. We consider various (N, T) combinations and representative (ρ, α, β) combinations. These should cover the most empirically relevant cases. The results are shown in Table 1. x/y denotes the frequency of underestimation and overestimation is $x\%$ and $y\%$ respectively. In all cases, the probability of underestimation plus overestimation, $x+y$ is significantly smaller than the probability that the estimated change points differ from the true change points. This implies $O_p(1)$ deviation from the true change points does not significantly affect $\tilde{r}_1$, $\tilde{r}_2$ and $\tilde{r}_3$. Also, when the size of each subsample is large enough, x and y are both zeros. This further confirms our theoretical result that $\tilde{r}_1$, $\tilde{r}_2$ and $\tilde{r}_3$ are robust to $O_p(1)$ estimation error of the change points.

5.4 Testing Multiple Changes

Now we present the results for the various tests of multiple changes. Table 2 reports size of the test for 0 versus l changes with $l = 1, 2, 3$, size of the UDmax and WDmax tests and the probabilities of selecting changes when the data is generated under Setup 1. We consider two methods of estimating the number of changes, $\hat{L}_1$ and $\hat{L}_2$. $\hat{L}_1$ is obtained by the sequential procedure as discussed in Section 4.4 while $\hat{L}_2$ is obtained by using WDmax to test the presence of at least one change first and then performing the sequential procedure starting from 1 versus 2 changes. Table 3 reports the power of the test for 0 versus l changes with $l = 1, 2, 3$, the power of the UDmax and WDmax tests, the power of the test for 1 versus 2 changes, the size of the test for 2 versus 3 changes and the probabilities of selecting changes when the data is generated under Setup 3. For both tables, we consider $(N, T) = (100, 100)$ and $(100, 200)$ with $\epsilon = 0.05, 0.10, 0.15, 0.20$ and 0.25, and $(\rho, \alpha, \beta) = (0, 0, 0)$, $(0.7, 0, 0)$ and $(0.7, 0.3, 0.3)$. We delete the case $T = 100$ and $\epsilon = 0.05$ to ensure the sample size of each regime is at least 10.

[14]The finite sample performance of the estimated factor space should be similar to that of the single change case, which are evaluated in Baltagi et al. (2016).

Note that in calculating the HAC estimator of the covariance matrix of the second moments of the estimated factors, Bartlett kernel is used with bandwidth $T^{1/3}$ for testing 0 versus l changes and $2\times T^{1/5}$ for testing l versus $l+1$ changes[15]. In estimating the number of factors at the very beginning, IC_{p3} [16] is used except for the case $(N,T) = (100,100)$ and $(\rho,\alpha,\beta) = (0.7,0.3,0.3)$. In that case, IC_{p3} overestimates too much, thus we switch to IC_{p1}. The critical values are obtained from Bai and Perron (2003) with nominal size of 5%.

First consider the size properties. Table 2 shows that overall, all tests are slightly undersized. The undersizing phenomenon is quite obvious when $T = 100$ and $\rho = 0$. This is in line with previous findings, see Diebold and Chen (1996). When T increases to 200, the empirical size gets closer to the nominal size 5%. It is also easy to see that when $\rho = 0.7$ and $\epsilon = 0.05$, the tests are significantly oversized. Thus we recommend choosing ϵ at least 0.10 when the factors have serial correlation. Once T is large enough to guarantee the accuracy of the estimated factors, serial and cross-sectional dependence of the errors do not seem to affect the size of the various tests.

Now consider the power properties. Powers of the tests for 0 versus l changes are good in all cases. WDmax has good power except when $T = 100$ and $\epsilon = 0.25$, and is more powerful than UDmax. When $T = 200$, test for 1 versus 2 changes has good power, thus the probabilities of selecting the correct number of changes is always close to 1. However, the power decreases a lot when $T = 100$, and thus $\hat{L}_1$ and $\hat{L}_2$ tend to underestimate the number of changes. This is because when $T = 100$, the sample size of each regime is too small to be robust to the estimation error of the change points. We also conduct simulations gradually increasing T and find that when T increases to 140, the performance is as good as $T = 200$. Of course, the power also depends upon the location of the change points. We suggest that, for each regime, the sample size should be at least 40. Finally, when $T = 100$ serial and cross-sectional dependence of the errors decrease the power. This is again caused by small T. In summary, results

[15] For Bartlett kernel, the condition on the bandwidth is $d_T = O(T^{1/3})$. We simply choose $d_T = T^{1/3}$. For testing l versus $l+1$ changes, since Theorem 8 requires $\frac{d_T}{T^{\frac{1}{4}}} \to 0$, we choose $d_T = 2\times T^{1/5}$. For space limitations, finite sample performances of different bandwidth choices are not carried out.

[16] As discussed in Section 3.1.1, less conservative criterion is recommended in estimating the number of factors in the first step.

in both tables are consistent with our theoretical derivation and show the usefulness of the proposed testing procedure.

6 APPLICATION

In this section we apply the proposed method to detect breaks in Stock and Watson (2009)'s US macroeconomic data set. The original data set contains 108 monthly and 79 quarterly time series of US nominal and real variables, including prices, interest rates, money and credit aggregates, stock prices, exchange rates, etc, ranging from 1959:Q1 to 2006:Q4. The transformed data is a balanced panel of standardized variables with $N = 109$ and $T = 190$, ranging from 1959:Q3 to 2006:Q4, see Stock and Watson (2009) for the details of data description and transformation.

We use WDmax to detect the presence of at least one break. The trimming parameter ϵ equals 0.1. Using Bai and Ng (2002)'s IC_1, IC_2 and IC_3, the estimated number of pseudo factors $\tilde{r}$ equals 4, 2 and 10 respectively. Using Ahn and Horenstein (2013)'s ER or GR estimator, $\tilde{r}$ equals 1. At significance level 5%, WDmax fails to reject the null when $\tilde{r} = 4$, 2, and 1 and reject the null when $\tilde{r} = 10$. UDmax also rejects the null when $\tilde{r} = 10$. To check the robustness, we set the trimming parameter ϵ to be 0.05 and 0.15. The results are the same. We also set $\tilde{r}$ manually with maximum 12. We find that WDmax always fails to reject the null when $\tilde{r} \leq 5$, and always rejects the null when $\tilde{r} \geq 6$. As discussed in Remark 13, under the null a larger $\tilde{r}$ (i.e., overestimating the number of factors) will make the test even less likely to reject the null, while under the alternative, a smaller $\tilde{r}$ (i.e., underestimating the number of factors) may make the test fail to reject the null. Therefore, we conclude there exists at least one break[17].

We then use the sequential test $F_{NT}(l+1\,|l\,)$ to determine the number of breaks, starting from 1 versus 2 breaks. We find that $F_{NT}(2\,|1\,)$ rejects the null but $F_{NT}(3\,|2\,)$ fails to reject the null, thus we conclude there are two breaks. The estimated break points are 1979:Q1 and 1983:Q4. The first break could be due to the impact of the

[17]The detected breaks may also come from the idiosyncratic noises if the number of factors is overestimated. Since this section is mainly for illustration, further empirical evidence to support the presence of breaks in the factor loadings is out of scope.

Iranian revolution on the oil price and US inflation, which at least partially motivated the Federal Reserve to tighten monetary policy. This break is also detected by Chen et al. (2014) and Ma and Su (2018). The second break could be due to the great moderation, and is also considered by Stock and Watson (2009) and Ma and Su (2018). The estimated number of factors in three regimes is 3, 3, 4 respectively. Decomposing the breaks into breaks in loadings of old factors and emergence of new factors and identifying the extra factor is beyond the scope of this paper.

7 CONCLUSIONS

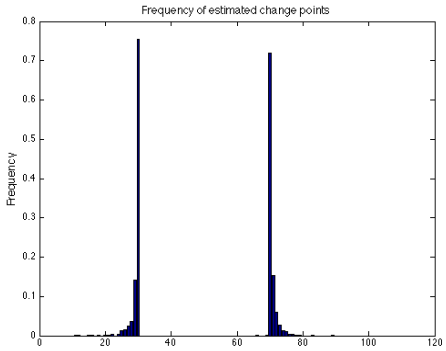
This paper studies a high dimensional factor model with multiple changes. The main issues tackled are the estimation of change points, tests for the presence of multiple changes and tests for determining the number of changes. Our strategy is based on the second moments of the estimated pseudo factors and we show that estimation errors contained in the estimated factors have different effects on estimating and testing structural changes. Simulation studies confirm the theoretical results and demonstrate its good performance. An application to U.S. macroeconomic dataset illustrates our procedure for testing and estimating structural breaks. A natural next step is to use bootstrap to fix the undersizing issue when T is less than 100, as discussed in Diebold and Chen (1996). It will be also interesting to apply our theoretical results to study the financial market comovement during crises, as discussed in Bekaert, Ehrmann, Fratzscher and Mehl (2014) and Belvisi, Pianeti and Urga (2015).

References

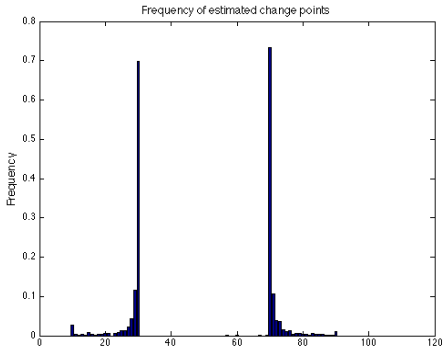
[1] Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.

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

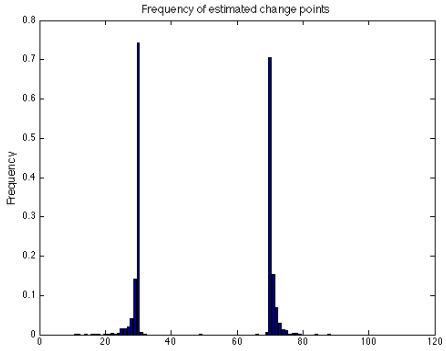
[3] Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–173.



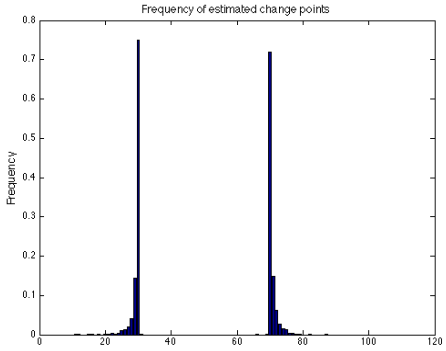
$(\rho, \alpha, \beta) = (0, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0.7, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

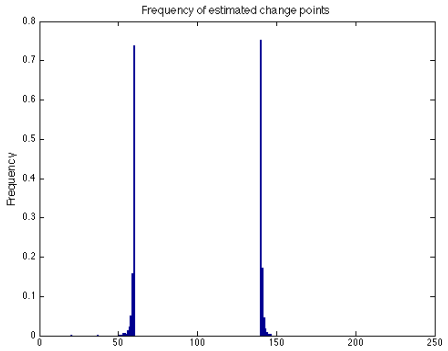


$(\rho, \alpha, \beta) = (0, 0.3, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

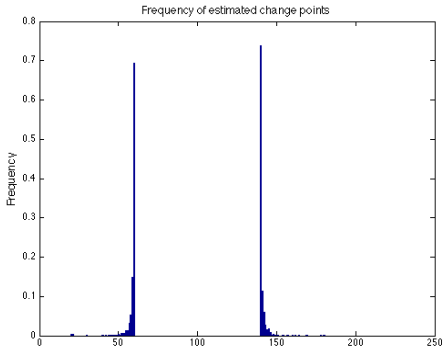


$(\rho, \alpha, \beta) = (0, 0, 0.3)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

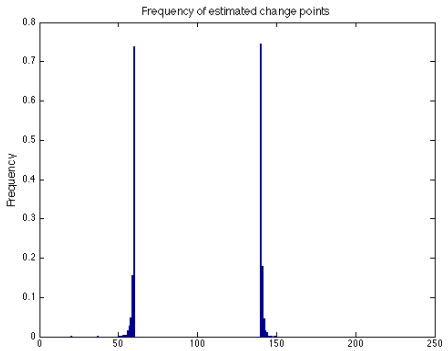
Figure 1: Histogram of estimated change points for $(N, T) = (100, 100), r_1 = 2, r_2 = 2, r_3 = 3, \bar{r} = 7$



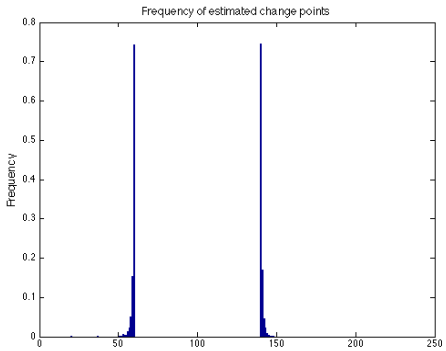
$(\rho, \alpha, \beta) = (0, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0.7, 0, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0, 0.3, 0)$, $(\tau_1, \tau_2) = (0.3, 0.7)$



$(\rho, \alpha, \beta) = (0, 0, 0.3)$, $(\tau_1, \tau_2) = (0.3, 0.7)$

Figure 2: Histogram of estimated change points for $(N, T) = (100, 200), r_1 = 2, r_2 = 2, r_3 = 3, \bar{r} = 7$

Table 1: Estimated number of factors in each regime for $r_1 = 2, r_2 = 2, r_3 = 3, \bar{r} = 7$

N	T	IC_{p2}			GR			ER		
		$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_3$	$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_3$	$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_3$
$\rho = 0, \alpha = 0, \beta = 0$										
100	100	0/0	0/1	1/0	1/0	1/0	5/0	1/0	0/0	3/0
100	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	300	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
$\rho = 0.7, \alpha = 0, \beta = 0$										
100	100	4/4	0/10	1/2	1/2	3/5	12/0	1/0	1/6	6/0
100	200	0/0	0/2	0/0	0/1	0/0	0/0	0/0	0/1	0/0
200	200	0/0	0/3	0/0	0/0	0/1	0/0	0/0	0/1	0/0
200	300	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0
$\rho = 0, \alpha = 0.3, \beta = 0$										
100	100	0/0	0/1	2/0	3/0	1/0	11/0	1/0	1/0	7/0
100	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	300	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
$\rho = 0, \alpha = 0, \beta = 0.3$										
100	100	0/0	0/0	1/0	1/0	1/0	6/0	1/0	0/0	4/0
100	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	200	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
200	300	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Table 2: Size of tests and probabilities of selecting changes

ϵ	$l\|0$			$Dmax$		$\hat{L}_1$			$\hat{L}_2$		
	1	2	3	U	W	0	1	2	0	1	2
$N = 100, T = 100, \rho = 0, \alpha = 0, \beta = 0$											
0.10	0.4	0.2	0.1	0.4	0.2	99.6	0.4	0	99.8	0.2	0
0.15	0.1	0	0	0.1	0.1	99.9	0.1	0	99.9	0.1	0
0.20	0	0	0	0	0	100	0	0	100	0	0
0.25	0.1	0	0	0	0	99.9	0.1	0	100	0	0
$N = 100, T = 200, \rho = 0, \alpha = 0, \beta = 0$											
0.05	1.8	1.8	1.7	1.6	1.4	98.2	1.8	0	98.6	1.4	0
0.10	0.2	0.2	0.5	0.3	0.1	99.8	0.2	0	99.9	0.1	0
0.15	0.6	0.5	0.2	0.7	0.2	99.4	0.6	0	99.8	0.2	0
0.20	0.4	0.3	0.1	0.4	0	99.6	0.4	0	100	0	0
0.25	0.9	0.4	0	0.7	0.2	99.1	0.9	0	99.8	0.2	0
$N = 100, T = 100, \rho = 0.7, \alpha = 0, \beta = 0$											
0.10	2.3	2.6	3.0	2.5	2.2	97.7	2.3	0	97.8	2.2	0
0.15	0.9	1.8	1.0	1.1	1.2	99.1	0.9	0	98.8	1.2	0
0.20	0.9	1.3	0.5	0.9	0.6	99.1	0.9	0	99.4	0.6	0
0.25	0.8	1.3	0	0.7	0.1	99.2	0.8	0	99.9	0.1	0
$N = 100, T = 200, \rho = 0.7, \alpha = 0, \beta = 0$											
0.05	12.7	25.9	23.4	15.9	17.5	87.3	11.8	0.8	82.5	16.1	0.13
0.10	5.3	8.4	8.8	6.4	7.5	94.7	5.1	0.2	92.5	7.2	0.3
0.15	4.5	5.9	4.2	5.1	5.2	95.5	4.5	0	94.8	5.0	0.2
0.20	3.4	4.2	4.0	3.3	3.4	96.6	3.4	0	96.6	3.4	0
0.25	3.6	3.5	0.3	2.8	2.1	96.4	3.6	0	97.9	2.1	0
$N = 100, T = 100, \rho = 0.7, \alpha = 0.3, \beta = 0.3$											
0.10	2.0	2.5	3.1	2.5	2.4	98.0	2.0	0	97.6	2.4	0
0.15	0.8	2.0	1.0	1.0	1.1	99.2	0.8	0	98.9	1.1	0
0.20	1.0	1.4	1.6	1.0	0.7	99.0	1.0	0	99.3	0.7	0
0.25	0.8	1.3	0.1	0.6	0.1	99.2	0.8	0	99.9	0.1	0
$N = 100, T = 200, \rho = 0.7, \alpha = 0.3, \beta = 0.3$											
0.05	12.5	26.8	23.8	16.3	17.8	87.5	11.7	0.7	82.2	16.5	1.2
0.10	5.4	8.0	8.2	6.2	7.3	94.6	5.2	0.2	92.7	7.0	0.3
0.15	4.6	5.6	4.2	5.3	5.3	95.4	4.6	0	94.7	5.2	0.1
0.20	3.7	4.0	1.9	3.6	3.2	96.3	3.7	0	96.8	3.2	0
0.25	3.6	3.5	0.3	2.9	2.1	96.4	3.6	0	97.9	2.0	0.1

Table 3: Power of tests and probabilities of selecting changes for $L = 2$

ϵ	$l\|0$			$Dmax$		$l+1\|l$			$\hat{L}_1$			$\hat{L}_2$	
	1	2	3	U	W	2|1	3|2	0	1	2	0	1	2
$N = 100, T = 100, \rho = 0, \alpha = 0, \beta = 0$													
0.10	100	100	100	98.4	100	23.4	0	0	76.6	23.4	0	76.6	23.4
0.15	100	100	100	23.1	100	12.4	0	0	87.6	12.4	0	87.6	12.4
0.20	100	100	100	4.9	99.9	9.6	0	0	90.4	9.6	0.1	90.3	9.6
0.25	100	100	100	3.6	3.7	11.1	0	0	88.9	11.1	96.3	3.3	0.4
$N = 100, T = 200, \rho = 0, \alpha = 0, \beta = 0$													
0.05	100	100	100	100	100	100	0.5	0	0	99.5	0	0	99.5
0.10	100	100	100	100	100	100	0	0	0	100	0	0	100
0.15	100	100	100	100	100	100	0	0	0	100	0	0	100
0.20	100	100	100	100	100	100	0	0	0	100	0	0	100
0.25	100	100	100	100	100	100	0	0	0	100	0	0	100
$N = 100, T = 100, \rho = 0.7, \alpha = 0, \beta = 0$													
0.10	100	100	100	98.9	100	41.9	0.1	0	58.1	41.8	0	58.1	41.8
0.15	100	100	100	28.7	100	23.3	0	0	76.7	23.3	0	76.7	23.3
0.20	100	100	100	5.9	100	15.8	0	0	84.2	15.8	0	84.2	15.8
0.25	100	100	100	4.3	4.3	15.5	0	0	84.5	15.5	95.7	3.6	0.7
$N = 100, T = 200, \rho = 0.7, \alpha = 0, \beta = 0$													
0.05	100	100	100	100	100	100	3.9	0	0	96.1	0	0	96.1
0.10	100	100	100	100	100	100	0.4	0	0	99.6	0	0	99.6
0.15	100	100	100	100	100	100	0.1	0	0	99.9	0	0	99.9
0.20	100	100	100	100	100	100	0	0	0	100	0	0	100
0.25	100	100	100	100	100	100	0	0	0	100	0	0	100
$N = 100, T = 100, \rho = 0.7, \alpha = 0.3, \beta = 0.3$													
0.10	97.3	98.5	99.9	78.5	97.7	37.0	0.3	2.7	60.6	36.5	2.3	60.9	36.6
0.15	97.5	98.9	100	16.9	96.9	19.6	0	2.5	78.0	19.5	3.1	77.4	19.5
0.20	97.5	99.9	100	1.3	95.1	15.3	0	2.5	82.2	15.3	4.9	80.1	15.0
0.25	97.5	99.9	99.2	0.1	1.4	15.7	0	2.5	81.9	15.6	98.6	1.2	0.2
$N = 100, T = 200, \rho = 0.7, \alpha = 0.3, \beta = 0.3$													
0.05	100	100	100	100	100	100	4.2	0	0	95.8	0	0	95.8
0.10	100	100	100	100	100	100	0.4	0	0	99.6	0	0	99.6
0.15	100	100	100	100	100	100	0.1	0	0	99.9	0	0	99.9
0.20	100	100	100	100	100	100	0	0	0	100	0	0	100
0.25	100	100	100	100	100	100	0	0	0	100	0	0	100

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

[5] Bai, J., Han, X., Shi, Y., 2016. Estimation and inference of structural changes in high dimensional factor models. Working Paper, City University of Hong Kong.

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

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

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

[9] Bai, J., Perron, P., 2003. Critical values for multiple structural change tests. Econometrics Journal 6, 72-78.

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

[11] 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.

[12] Bekaert, G., Ehrmann, M., Fratzscher, M., Mehl, A., 2014. The global crisis and equity market contagion. Journal of Finance 69, 2597-2649.

[13] Belvisi, M., Pianetti, R., Urga, G., 2015. Modelling financial markets comovements during crises: a dynamic multi-factor approach. Working Paper, Cass Business School.

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

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

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

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

[18] Cheng, X., Liao, Z., Schorfheide, F., 2016. Shrinkage estimation of high-dimensional factor models with structural instabilities. Review of Economic Studies 83, 1511-1543.

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

[20] Diebold, F.X., Chen, C., 1996. Testing structural stability with endogenous breakpoint a Size comparison of analytic and bootstrap procedures. Journal of Econometrics 70, 221-241.

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

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

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

[24] Ma, S., Su, L., 2018. Estimation of large dimensional factor models with an unknown number of breaks. Journal of Econometrics 207, 1-29.

[25] Massacci, D., 2017. Least squares estimation of large dimensional threshold factor models. Journal of Econometrics 197, 101-129.

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

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

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

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

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

[31] 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, pp. 173–205.

[32] Stock, J.H., Watson, M.W., 2012. Disentangling the channels of the 2007-09 recession. Brookings Papers on Economic Activity, pp. 81–156.

ESTIMATING AND TESTING HIGH DIMENSIONAL FACTOR MODELS WITH MULTIPLE STRUCTURAL CHANGES
APPENDIX (not intended for publication)

A HAJEK-RENYI INEQUALITY

Hajek-Renyi inequality is applicable to various settings, including i.i.d., martingale difference, martingale, vector-valued martingale, mixingale and linear process, also see Bai (1996) for more details. For a sequence of independent random variables $\{x_t, t = 1, ...\}$ with $\mathbb{E}x_t = 0$ and $\mathbb{E}x_t^2 = \sigma_t^2$, Hajek and Renyi (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. It is easy to see that if $\sigma_t^2 \leq \sigma^2$ for all t and $c_k = \frac{1}{k}$, $P(\sup_{m\leq k\leq T} \left|\frac{1}{k}\sum_{t=1}^{k} x_t\right| > M) \leq \frac{2\sigma^2}{M^2}\frac{1}{m}$, thus $\sup_{1\leq k\leq T} \left|\frac{1}{k}\sum_{t=1}^{k} x_t\right| = O_p(1)$ and $\sup_{T\tau\leq k\leq T} \left|\frac{1}{k}\sum_{t=1}^{k} x_t\right| = O_p(\frac{1}{\sqrt{T}})$. If $c_k = \frac{1}{\sqrt{k}}$, $P(\sup_{m\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| > M) \leq \frac{\sigma^2}{M^2}(1+\sum_{k=m+1}^{T}\frac{1}{k})$, thus $\sup_{1\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| = O_p(\sqrt{\log T})$ since $\sum_{k=1}^{T}\frac{1}{k} - \log T$ converges to the Euler constant and $\sup_{T\tau\leq k\leq T} \left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t\right| = O_p(1)$ since $\sum_{k=m+1}^{T}\frac{1}{k} = \sum_{k=1}^{T}\frac{1}{k} - \sum_{k=1}^{T\tau}\frac{1}{k} \to \log T - \log T\tau = \log\frac{1}{\tau}$.

B PROOF OF PROPOSITION 1

Proof. For any $\epsilon > 0$ and $\eta_1 > 0, ..., \eta_L > 0$, define $D = \{(k_1, ..., k_L) : (\tau_{\iota 0} - \eta_\iota)T \leq k_\iota \leq (\tau_{\iota 0}+\eta_\iota)T$ for $\iota = 1, ..., L\}$, we need to show $P((\tilde{k}_1, ..., \tilde{k}_L) \in D^c) < \epsilon$ as $(N, T) \to \infty$. Since $D^c = \cup_{\iota=1}^{L}\{(k_1, ..., k_L)$:for $\kappa = 1, ..., L$, either $k_\kappa < (\tau_{\iota 0}-\eta_\iota)T$ or $k_\kappa > (\tau_{\iota 0}+\eta_\iota)T\} = \cup_{\iota=1}^{L} D_{(\iota)}^c$, it suffices to show $P((\tilde{k}_1, ..., \tilde{k}_L) \in D_{(\iota)}^c) < \epsilon$ as $(N, T) \to \infty$ for all ι. Since $(\tilde{k}_1, ..., \tilde{k}_L) = \arg\min \tilde{S}(k_1, ..., k_L)$, $\tilde{S}(\tilde{k}_1, ..., \tilde{k}_L) \leq \tilde{S}(k_{1,0}, ..., k_{L,0}) \leq \sum_{t=1}^{T}(y_t + z_t)'(y_t + z_t)$. If $(\tilde{k}_1, ..., \tilde{k}_L) \in D_{(\iota)}^c$, then $\min_{(k_1,...,k_L)\in D_{(\iota)}^c} \tilde{S}(k_1, ..., k_L) = \tilde{S}(\tilde{k}_1, ..., \tilde{k}_L)$. Thus $(\tilde{k}_1, ..., \tilde{k}_L) \in D_{(\iota)}^c$ implies $\min_{(k_1,...,k_L)\in D_{(\iota)}^c} \tilde{S}(k_1, ..., k_L) \leq \sum_{t=1}^{T}(y_t + z_t)'(y_t + z_t)$ and it

suffices to show $P(\min_{(k_1,...,k_L)\in D^c_{(\iota)}} \tilde{S}(k_1,...,k_L) - \sum_{t=1}^T (y_t+z_t)'(y_t+z_t) \leq 0) < \epsilon$ as $(N,T)\to\infty$.

For any given partition $(k_1,...,k_L)$, let $\tilde{\Sigma}_\iota = \frac{1}{k_\iota - k_{\iota-1}} \sum_{t=k_{\iota-1}+1}^{k_\iota} \tilde{g}_t\tilde{g}_t'$ and $a_t = vech(\Sigma_\kappa - \tilde{\Sigma}_\iota)$ for $t \in [k_{\iota-1}+1, k_\iota] \cap [k_{\kappa-1,0}+1, ..., k_{\kappa,0}]$, $\iota,\kappa = 1,...,L+1$. It follows $vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\iota) = a_t + y_t + z_t$ and

$$\begin{aligned}\tilde{S}(k_1,...,k_L) &= \sum_{\iota=1}^{L+1}\sum_{t=k_{\iota-1}+1}^{k_\iota}[vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\iota)]'[vech(\tilde{g}_t\tilde{g}_t' - \tilde{\Sigma}_\iota)] \\ &= \sum_{t=1}^T (y_t+z_t)'(y_t+z_t) + \sum_{t=1}^T a_t'a_t + 2\sum_{t=1}^T a_t'(y_t+z_t) \quad \text{(A-2)}\end{aligned}$$

Thus it suffices to show $P(\min_{(k_1,...,k_L)\in D^c_{(\iota)}} [\sum_{t=1}^T a_t'a_t + 2\sum_{t=1}^T a_t'(y_t+z_t)] \leq 0) < \epsilon$ as $(N,T)\to\infty$. Since $\min_{(k_1,...,k_L)\in D^c_{(\iota)}} [\sum_{t=1}^T a_t'a_t + 2\sum_{t=1}^T a_t'(y_t+z_t)] \leq 0$ implies $\min_{(k_1,...,k_L)\in D^c_{(\iota)}} \sum_{t=1}^T a_t'a_t \leq 2 \sup_{(k_1,...,k_L)\in D^c_{(\iota)}} \left|\sum_{t=1}^T a_t'(y_t+z_t)\right|$, it suffices to show that the left hand side dominates the right hand side asymptotically.

Consider the left hand side first. For any $(k_1,...,k_L) \in D^c_{(\iota)}$, there exists κ^* such that $k_{\kappa^*-1} < (\tau_{\iota 0} - \eta_\iota)T$ and $k_{\kappa^*} > (\tau_{\iota 0} + \eta_\iota)T$, thus for $t \in [(\tau_{\iota 0} - \eta_\iota)T, \tau_{\iota 0}T]$, $a_t = vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_\iota)$ and for $t \in [\tau_{\iota 0}T + 1, (\tau_{\iota 0} + \eta_\iota)T]$, $a_t = vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_{\iota+1})$. So for any $(k_1,...,k_L) \in D^c_{(\iota)}$,

$$\begin{aligned}&\sum_{t=1}^T a_t'a_t \\ \geq\ &\sum_{t=(\tau_{\iota 0}-\eta_\iota)T}^{\tau_{\iota 0}T} a_t'a_t + \sum_{t=\tau_{\iota 0}T+1}^{(\tau_{\iota 0}+\eta_\iota)T} a_t'a_t \\ \geq\ &\eta_\iota T[vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_\iota)'vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_\iota) + vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_{\iota+1})'vech(\tilde{\Sigma}_{\kappa^*} - \Sigma_{\iota+1})] \\ \geq\ &\eta_\iota T\frac{vech(\Sigma_\iota - \Sigma_{\iota+1})'vech(\Sigma_\iota - \Sigma_{\iota+1})}{2}, \quad \text{(A-3)}\end{aligned}$$

where the last inequality is due to $(x-a)^2 + (x-b)^2 = 2(x - \frac{a+b}{2})^2 + \frac{(a-b)^2}{2}$ for any x. Thus $\min_{(k_1,...,k_L)\in D^c_{(\iota)}} \sum_{t=1}^T a_t'a_t \geq \eta_\iota T\frac{vech(\Sigma_\iota - \Sigma_{\iota+1})'vech(\Sigma_\iota - \Sigma_{\iota+1})}{2}$. Next, the right hand side

is no larger than

$$\left|\sum_{\kappa=1}^{L+1}\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'(y_t+z_t)\right| \tag{A-4}$$

$$+\sup_{(k_1,...,k_L)\in D^c_{(\iota)}}\left|\sum_{\iota=1}^{L+1}\sum_{t=k_{\iota-1}+1}^{k_\iota} vech(\tilde{\Sigma}_\iota)'(y_t+z_t)\right|. \tag{A-5}$$

For the first term,

$$\begin{aligned}
&\left|\sum_{\kappa=1}^{L+1}\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'y_t\right| \\
\leq\ & \sum_{\kappa=1}^{L+1}\left|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'y_t\right| \leq \sum_{\kappa=1}^{L+1}\|\Sigma_\kappa\|\left\|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} y_t\right\| \\
\leq\ & \sum_{\kappa=1}^{L+1}\|\Sigma_\kappa\|\,\|J_0\|^2\,\|R_\kappa\|^2\left\|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}}(f_tf_t'-\Sigma_F)\right\| = o_p(T),
\end{aligned} \tag{A-6}$$

where the last equality follows from Assumption 1; and

$$\left|\sum_{\kappa=1}^{L+1}\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} vech(\Sigma_\kappa)'z_t\right| \leq \sum_{\kappa=1}^{L+1}\|\Sigma_\kappa\|\left\|\sum_{t=k_{\kappa-1,0}+1}^{k_{\kappa,0}} z_t\right\| = o_p(T), \tag{A-7}$$

where the last equality follows from Lemma 5. For the second term, define $b_t = vech(\Sigma_\kappa)$ for $t \in [k_{\kappa-1,0}+1, ..., k_{\kappa,0}]$, $\kappa = 1, ..., L+1$, then $vech(\tilde{g}_t\tilde{g}_t') = b_t + y_t + z_t$ for all t and $vech(\tilde{\Sigma}_\iota) = \frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota} vech(\tilde{g}_t\tilde{g}_t') = \frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota}(b_t+y_t+z_t)$. It follows that the second term is no larger than

$$\begin{aligned}
&\sup_{(k_1,...,k_L)\in D^c_{(\iota)}}\left|\sum_{\iota=1}^{L+1}\frac{1}{k_\iota-k_{\iota-1}}(\sum_{t=k_{\iota-1}+1}^{k_\iota} b_t)'(\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t))\right| \\
&+\sup_{(k_1,...,k_L)\in D^c_{(\iota)}}\left|\sum_{\iota=1}^{L+1}\frac{1}{k_\iota-k_{\iota-1}}(\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t))'(\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t))\right| \\
\leq\ & (L+1)(\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t)}{\sqrt{k_\iota-k_{\iota-1}}}\right\|^2 + \\
&\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} b_t}{\sqrt{k_\iota-k_{\iota-1}}}\right\|\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(y_t+z_t)}{\sqrt{k_\iota-k_{\iota-1}}}\right\|) \\
=\ & (L+1)(B^2+AB).
\end{aligned} \tag{A-8}$$

For term A, we have $A \leq \sup_{1\leq k_{\iota-1}<k_\iota\leq T}\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|b_t\|}{\sqrt{k_\iota-k_{\iota-1}}} \leq \sup_{1\leq k_{\iota-1}<k_\iota\leq T}\sqrt{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|b_t\|^2} \leq$

$\sqrt{\sum_{t=1}^{T}\|b_t\|^2} = O(\sqrt{T})$. For term B, we have $B^2 \leq 2 \sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} y_t}{\sqrt{k_\iota-k_{\iota-1}}}\right\|^2 + 2\sup_{1\leq k_{\iota-1}<k_\iota\leq T}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} z_t}{\sqrt{k_\iota-k_{\iota-1}}}\right\|^2 = 2B_1^2 + 2B_2^2$. $B_1 = o_p(\sqrt{T})$, since

$$\begin{aligned} B_1 &\leq \sum_{\kappa=1}^{L+1} \sup_{k_{\kappa-1,0}<k<l\leq k_{\kappa,0}} \left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l} y_t\right\| \\ &\leq \sum_{\kappa=1}^{L+1} \|J_0\|^2 \|R_\kappa\|^2 \sup_{k_{\kappa-1,0}<k<l\leq k_{\kappa,0}} \left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l} \epsilon_t\right\|, \end{aligned} \tag{A-9}$$

and by Assumption 7,

$$\begin{aligned} &\mathbb{E}\Big(\sup_{k_{\kappa-1,0}<k<l\leq k_{\kappa,0}} \left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l} \epsilon_t\right\|^{4+\delta}\Big) \\ \leq\ & \sum_{k=k_{\kappa-1,0}}^{k_{\kappa,0}-1}\sum_{l=k+1}^{k_{\kappa,0}} \mathbb{E}\Big(\left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l} \epsilon_t\right\|^{4+\delta}\Big) \leq T^2 M. \end{aligned} \tag{A-10}$$

Using Lemma 5, $B_2 = o_p(\sqrt{T})$. Taking together, the right hand side is $o_p(T)$ and thus dominated by the left hand side. ■

C PROOF OF THEOREM 1

Proof. From Proposition 1, we know that for any $\epsilon > 0$ and $\eta_1 > 0, ..., \eta_L > 0$, $P((\tilde{k}_1, ..., \tilde{k}_L) \in D^c) < \epsilon$ as $(N,T) \to \infty$. Thus to show $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$ for any given $1 \leq \iota \leq L$, we need to show for any $\epsilon > 0$ and $\eta_1 > 0, ..., \eta_L > 0$, there exist $C > 0$ such that $P((\tilde{k}_1, ..., \tilde{k}_L) \in D, \left|\tilde{k}_\iota - k_{\iota 0}\right| > C) < \epsilon$ as $(N,T) \to \infty$. By symmetry, it suffices to show $P((\tilde{k}_1, ..., \tilde{k}_L) \in D, \tilde{k}_\iota < k_{\iota 0} - C) < \epsilon$ as $(N,T) \to \infty$. Define $D(C)_{(\iota)} = D \cap \{k_\iota < k_{\iota 0} - C\}$. Since $(\tilde{k}_1, ..., \tilde{k}_L) = \arg\min \tilde{S}(k_1, ..., k_L)$, $\tilde{S}(\tilde{k}_1, ..., \tilde{k}_L) \leq \tilde{S}(\tilde{k}_1, ..., k_{\iota 0}, ..., \tilde{k}_L)$. Thus if $(\tilde{k}_1, ..., \tilde{k}_L) \in D(C)_{(\iota)}$,

$$\begin{aligned} &\min_{(k_1,...,k_L)\in D(C)_{(\iota)}} [\tilde{S}(k_1, ..., k_L) - \tilde{S}(k_1, ..., k_{\iota 0}, ..., k_L)] \\ \leq\ & \tilde{S}(\tilde{k}_1, ..., \tilde{k}_L) - \tilde{S}(\tilde{k}_1, ..., k_{\iota 0}, ..., \tilde{k}_L) \leq 0. \end{aligned}$$

Therefore it suffices to show $P(\min_{(k_1,...,k_L)\in D(C)_{(\iota)}}[\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)]\leq 0)<\epsilon$ as $(N,T)\to\infty$.

We then show that the event $\min_{(k_1,...,k_L)\in D(C)_{(\iota)}}[\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)]\leq 0$ is just the event $\min_{(k_1,...,k_L)\in D(C)_{(\iota)}}\frac{\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)}{|k_\iota-k_{\iota 0}|}\leq 0$. Conditioning on the former, for any $(k_1^*,...,k_L^*)\in D(C)_{(\iota)}$, $\arg\min_{(k_1,...,k_L)\in D(C)_{(\iota)}}[\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)]=(k_1^*,...,k_L^*)$ implies $\tilde{S}(k_1^*,...,k_L^*)-\tilde{S}(k_1^*,...,k_{\iota 0},...,k_L^*)\leq 0$, and this further implies $\frac{\tilde{S}(k_1^*,...,k_L^*)-\tilde{S}(k_1^*,...,k_{\iota 0},...,k_L^*)}{|k_\iota^*-k_{\iota 0}|}\leq 0$. It follows that $\min_{(k_1,...,k_L)\in D(C)_{(\iota)}}\frac{\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)}{|k_\iota-k_{\iota 0}|}$, which is not larger than $\frac{\tilde{S}(k_1^*,...,k_L^*)-\tilde{S}(k_1^*,...,k_{\iota 0},...,k_L^*)}{|k_\iota^*-k_{\iota 0}|}$, has to be nonpositive. Note that the above argument holds for any $(k_1^*,...,k_L^*)\in D(C)_{(\iota)}$, thus the former implies the latter. Similarly, the latter also implies the former. Therefore, it suffices to show $P(\min_{(k_1,...,k_L)\in D(C)_{(\iota)}}\frac{\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)}{|k_\iota-k_{\iota 0}|}\leq 0)<\epsilon$ as $(N,T)\to\infty$.

Next, decompose $\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_{\iota 0},...,k_L)$ as

$$[\tilde{S}(k_1,...,k_L)-\tilde{S}(k_1,...,k_\iota,k_{\iota 0},...,k_L)] \tag{A-11}$$

$$-[\tilde{S}(k_1,...,k_{\iota 0},...,k_L)-\tilde{S}(k_1,...,k_\iota,k_{\iota 0},...,k_L)]. \tag{A-12}$$

Term (A-11) equals

$$\begin{aligned}
&\sum_{t=k_\iota+1}^{k_{\iota+1}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]\\
&-\sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^{\Delta})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^{\Delta})]\\
&-\sum_{t=k_{\iota 0}+1}^{k_{\iota+1}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1}^{*})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1}^{*})]\\
=\;& K_1-K_2-K_3,
\end{aligned} \tag{A-13}$$

and term (A-12) equals

$$\begin{aligned}
&\sum_{t=k_{\iota-1}+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^*)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^*)]\\
&-\sum_{t=k_{\iota-1}+1}^{k_\iota}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]\\
&-\sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^\Delta)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota^\Delta)]\\
=\;& L_1-L_2-L_3, \qquad \text{(A-14)}
\end{aligned}$$

where $\tilde{\Sigma}_\iota^\Delta=\frac{\sum_{t=k_\iota+1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}-k_\iota}$, $\tilde{\Sigma}_\iota^*=\frac{\sum_{t=k_{\iota-1}+1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}-k_{\iota-1}}$ and $\tilde{\Sigma}_{\iota+1}^*=\frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1}}\tilde{g}_t\tilde{g}_t'}{k_{\iota+1}-k_{\iota 0}}$. Note that $L_3=K_2$, thus $(K_1-K_2-K_3)-(L_1-L_2-L_3)=(K_1-K_3)-(L_1-L_2)$. Replacing $\tilde{\Sigma}_{\iota+1}^*$ by $\tilde{\Sigma}_{\iota+1}$, K_3 is magnified, thus $K_1-K_3\geq\sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]$; and replacing $\tilde{\Sigma}_\iota^*$ by $\tilde{\Sigma}_\iota$, L_1 is magnified, thus $L_1-L_2\leq\sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]$. Taken together,

$$\begin{aligned}
&(K_1-K_3)-(L_1-L_2)\\
\geq\;& \sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{\iota+1})]\\
&-\sum_{t=k_\iota+1}^{k_{\iota 0}}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_\iota)]\\
=\;& \sum_{t=k_\iota+1}^{k_{\iota 0}}vech(\Sigma_\iota-\tilde{\Sigma}_{\iota+1})'vech(\Sigma_\iota-\tilde{\Sigma}_{\iota+1})\\
&-\sum_{t=k_\iota+1}^{k_{\iota 0}}vech(\Sigma_\iota-\tilde{\Sigma}_\iota)'vech(\Sigma_\iota-\tilde{\Sigma}_\iota)\\
&+2\sum_{t=k_\iota+1}^{k_{\iota 0}}vech(\Sigma_\iota-\tilde{\Sigma}_{\iota+1})'(y_t+z_t)\\
&-2\sum_{t=k_\iota+1}^{k_{\iota 0}}vech(\Sigma_\iota-\tilde{\Sigma}_\iota)'(y_t+z_t)\\
=\;& K_{\Delta 1}-L_{\Delta 1}+K_{\Delta 2}-L_{\Delta 2}, \qquad \text{(A-15)}
\end{aligned}$$

thus it suffices to show $P(\min_{(k_1,\ldots,k_L)\in D(C)_{(\iota)}}\frac{K_{\Delta 1}-L_{\Delta 1}+K_{\Delta 2}-L_{\Delta 2}}{|k_\iota-k_{\iota 0}|}\leq 0)<\epsilon$ as $(N,T)\to\infty$.

We consider the case $k_{\iota-1}<k_{\iota-1,0}$ and $k_{\iota+1}>k_{\iota+1,0}$. In case $k_{\iota-1}\geq k_{\iota-1,0}$ or $k_{\iota+1}\leq k_{\iota+1,0}$, the proof is easier and therefore omitted. Plug in $\tilde{\Sigma}_{\iota+1}=\frac{1}{k_{\iota+1}-k_\iota}\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t+z_t)+vech(\frac{1}{k_{\iota+1}-k_\iota}[(k_{\iota 0}-k_\iota)\Sigma_\iota+(k_{\iota+1,0}-k_{\iota 0})\Sigma_{\iota+1}+(k_{\iota+1}-k_{\iota+1,0})\Sigma_{\iota+2}])$ and $\tilde{\Sigma}_\iota=\frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}}^{k_\iota}(y_t+z_t)+vech(\frac{1}{k_\iota-k_{\iota-1}}[(k_{\iota-1,0}-k_{\iota-1})\Sigma_{\iota-1}+(k_\iota-k_{\iota-1,0})\Sigma_\iota])$, and denote $\phi_{k_{\iota-1},k_\iota}=vech(\frac{1}{k_\iota-k_{\iota-1}}(k_{\iota-1,0}-k_{\iota-1})(\Sigma_{\iota-1}-\Sigma_\iota))$ and $\phi_{k_\iota,k_{\iota+1}}=vech(\frac{1}{k_{\iota+1}-k_\iota}[(k_{\iota+1,0}-$

$k_{\iota 0})(\Sigma_{\iota+1}-\Sigma_\iota)+(k_{\iota+1}-k_{\iota+1,0})(\Sigma_{\iota+2}-\Sigma_\iota)]$), we have

$$\begin{aligned}
\frac{1}{k_{\iota 0}-k_\iota}K_{\Delta 1} &= [\phi_{k_\iota,k_{\iota+1}}+\frac{\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t+z_t)}{k_{\iota+1}-k_\iota}]'[\phi_{k_\iota,k_{\iota+1}}+\frac{\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t+z_t)}{k_{\iota+1}-k_\iota}], && \text{(A-16)}\\
\frac{1}{k_{\iota 0}-k_\iota}L_{\Delta 1} &= [\phi_{k_{\iota-1},k_\iota}+\frac{\sum_{t=k_{\iota-1}}^{k_\iota}(y_t+z_t)}{k_\iota-k_{\iota-1}}]'[\phi_{k_{\iota-1},k_\iota}+\frac{\sum_{t=k_{\iota-1}}^{k_\iota}(y_t+z_t)}{k_\iota-k_{\iota-1}}], && \text{(A-17)}\\
\frac{1}{k_{\iota 0}-k_\iota}K_{\Delta 2} &= -2[\phi_{k_\iota,k_{\iota+1}}+\frac{\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t+z_t)}{k_{\iota+1}-k_\iota}]'\frac{\sum_{t=k_\iota+1}^{k_{\iota 0}}(y_t+z_t)}{k_{\iota 0}-k_\iota}, && \text{(A-18)}\\
\frac{1}{k_{\iota 0}-k_\iota}L_{\Delta 2} &= 2[\phi_{k_{\iota-1},k_\iota}+\frac{\sum_{t=k_{\iota-1}}^{k_\iota}(y_t+z_t)}{k_\iota-k_{\iota-1}}]'\frac{\sum_{t=k_\iota+1}^{k_{\iota 0}}(y_t+z_t)}{k_{\iota 0}-k_\iota}. && \text{(A-19)}
\end{aligned}$$

For $(k_1,...,k_L)\in D(C)_{(\iota)}$ and η_ι and $\eta_{\iota+1}$ small enough,

$$\begin{aligned}
&\left\|\phi_{k_\iota,k_{\iota+1}}\right\|\\
\geq\ & \frac{k_{\iota+1,0}-k_{\iota 0}}{k_{\iota+1}-k_\iota}\left\|vech(\Sigma_{\iota+1}-\Sigma_\iota)\right\|-\frac{k_{\iota+1}-k_{\iota+1,0}}{k_{\iota+1}-k_\iota}\left\|vech(\Sigma_{\iota+2}-\Sigma_\iota)\right\|\\
\geq\ & \frac{1}{1+\frac{\eta_{\iota+1}+\eta_\iota}{\tau_{\iota+1,0}-\tau_{\iota 0}}}\left\|vech(\Sigma_{\iota+1}-\Sigma_\iota)\right\|-\frac{\eta_{\iota+1}}{\eta_{\iota+1}+\tau_{\iota+1,0}-\tau_{\iota 0}}\left\|vech(\Sigma_{\iota+2}-\Sigma_\iota)\right\|\\
\geq\ & \frac{1}{2}\left\|vech(\Sigma_{\iota+1}-\Sigma_\iota)\right\|, && \text{(A-20)}
\end{aligned}$$

and for $\eta_{\iota-1}$ and η_ι small enough,

$$\left\|\phi_{k_{\iota-1},k_\iota}\right\|=\frac{k_{\iota-1,0}-k_{\iota-1}}{k_\iota-k_{\iota-1}}\left\|vech(\Sigma_{\iota-1}-\Sigma_\iota)\right\|\leq\frac{\eta_{\iota-1}}{\tau_{\iota 0}-\tau_{\iota-1,0}-\eta_\iota}\left\|vech(\Sigma_{\iota-1}-\Sigma_\iota)\right\| \qquad \text{(A-21)}$$

is arbitrarily small.

$$\begin{aligned}
&\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}}\left\|\frac{1}{k_{\iota+1}-k_\iota}\sum_{t=k_\iota+1}^{k_{\iota+1}}(y_t+z_t)\right\|\\
\leq\ & \frac{1}{\tau_{\iota+1,0}-\tau_{\iota 0}}(\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}}\left\|\frac{1}{T}\sum_{t=k_\iota+1}^{k_{\iota 0}}(y_t+z_t)\right\|\\
&+\left\|\frac{1}{T}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}(y_t+z_t)\right\|+\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}}\left\|\frac{1}{T}\sum_{t=k_{\iota+1,0}+1}^{k_{\iota+1}}(y_t+z_t)\right\|)\\
=\ & o_p(1), && \text{(A-22)}
\end{aligned}$$

where we have used

$$\begin{aligned}
\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T}\sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} y_t \right\| &= o_p(1), \left\| \frac{1}{T}\sum\nolimits_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} y_t \right\| = o_p(1), \\
\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T}\sum\nolimits_{t=k_{\iota+1,0}+1}^{k_{\iota+1}} y_t \right\| &= o_p(1), \qquad \text{(A-23)} \\
\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T}\sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} z_t \right\| &= o_p(1), \left\| \frac{1}{T}\sum\nolimits_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} z_t \right\| = o_p(1), \\
\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T}\sum\nolimits_{t=k_{\iota+1,0}+1}^{k_{\iota+1}} z_t \right\| &= o_p(1). \qquad \text{(A-24)}
\end{aligned}$$

The first three terms follow from Hajek-Renyi inequality, which is proved in Lemma 1 to be applicable to y_t within each regime while the last three terms follow from Lemma 5. Similarly,

$$\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_\iota - k_{\iota-1}} \sum\nolimits_{t=k_{\iota-1}}^{k_\iota} (y_t + z_t) \right\| = o_p(1), \qquad \text{(A-25)}$$

using Lemma 5 and

$$\begin{aligned}
\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T}\sum\nolimits_{t=k_{\iota-1,0}+1}^{k_{\iota-1,0}} y_t \right\| &= o_p(1), \qquad \text{(A-26)} \\
\sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{T}\sum\nolimits_{t=k_{\iota-1,0}+1}^{k_\iota} y_t \right\| &= o_p(1). \qquad \text{(A-27)}
\end{aligned}$$

Finally,

$$\begin{aligned}
& \sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_{\iota 0} - k_\iota} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} (y_t + z_t) \right\| \\
\leq & \sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_{\iota 0} - k_\iota} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} y_t \right\| + \sup_{(k_1,...,k_L)\in D(C)_{(\iota)}} \left\| \frac{1}{k_{\iota 0} - k_\iota} \sum\nolimits_{t=k_\iota+1}^{k_{\iota 0}} z_t \right\| \\
= & O_p(\frac{1}{\sqrt{C}}) + o_p(1), \qquad \text{(A-28)}
\end{aligned}$$

the first term follows from Hajek-Renyi inequality while the second terms follows from Lemma 5. Taken together and choosing sufficiently large C, the result follows. ■

D PROOF OF PROPOSITION 2

Proof. To simplify calculation, consider the case with two breaks. For any $\epsilon > 0$ and $\eta > 0$, define $W_\eta = \{k_1 : (\tau_{1,0} - \eta)T \leq k_1 \leq (\tau_{1,0} + \eta)T\}$, we need to show $P((\hat{k}_1 \in W_\eta^c) < \epsilon$ as $(N,T) \to \infty$. Since $\hat{k}_1 = \arg\min \tilde{S}(k_1)$, $\tilde{S}(\hat{k}_1) \leq \tilde{S}(k_{1,0})$. If $\hat{k}_1 \in W_\eta^c$, then $\min_{k_1 \in W_\eta^c} \tilde{S}(k_1) = \tilde{S}(\hat{k}_1)$. Thus $\hat{k}_1 \in W_\eta^c$ implies $\min_{k_1 \in W_\eta^c} \tilde{S}(k_1) \leq \tilde{S}(k_{1,0})$ and it suffices to show $P(\min_{k_1 \in W_\eta^c} \tilde{S}(k_1) - \tilde{S}(k_{1,0}) \leq 0) < \epsilon$ as $(N,T) \to \infty$.

For $k_1 < k_{1,0}$, after some calculation, we have:

$$\tilde{S}(k_1) - \tilde{S}(k_{1,0}) = \Pi_1(k_1) - \Pi(k_{1,0}) + \Psi_1(k_1) - \Psi(k_{1,0}), \tag{A-29}$$

where

$$\begin{aligned} & \Pi_1(k_1) - \Pi(k_{1,0}) \\ = \; & \frac{k_{1,0} - k_1}{(T - k_1)(T - k_{1,0})} \left\| vech[(T - k_{1,0})(\Sigma_1 - \Sigma_2) + (T - k_{2,0})(\Sigma_2 - \Sigma_3)] \right\|^2, \end{aligned} \tag{A-30}$$

$$\begin{aligned} & \Psi_1(k_1) \\ = \; & 2\varphi_{k_1}^{1\prime} \sum_{\iota=k_1+1}^{k_{1,0}} (y_t + z_t) + 2\varphi_{k_1}^{2\prime} \sum_{\iota=k_{1,0}+1}^{k_{2,0}} (y_t + z_t) + 2\varphi_{k_1}^{3\prime} \sum_{\iota=k_{2,0}+1}^{T} (y_t + z_t) \\ & -2[(k_{1,0} - k_1)\varphi_{k_1}^1 + (k_{2,0} - k_{1,0})\varphi_{k_1}^2 + (T - k_{2,0})\varphi_{k_1}^3]' \frac{1}{T - k_1} \sum_{\iota=k_1+1}^{T} (y_t + z_t) \\ & - \left\| \frac{1}{\sqrt{k_1}} \sum_{\iota=1}^{k_1} (y_t + z_t) \right\|^2 - \left\| \frac{1}{\sqrt{T - k_1}} \sum_{\iota=k_1+1}^{T} (y_t + z_t) \right\|^2, \end{aligned} \tag{A-31}$$

$$\begin{aligned} \Psi(k_{1,0}) \; = \; & 2\varphi_{k_{1,0}}^{2\prime} \sum_{\iota=k_{1,0}+1}^{k_{2,0}} (y_t + z_t) + 2\varphi_{k_{1,0}}^{3\prime} \sum_{\iota=k_{2,0}+1}^{T} (y_t + z_t) \\ & -2[(k_{2,0} - k_{1,0})\varphi_{k_{1,0}}^2 + (T - k_{2,0})\varphi_{k_{1,0}}^3]' \frac{1}{T - k_{1,0}} \sum_{\iota=k_{1,0}+1}^{T} (y_t + z_t) \\ & - \left\| \frac{1}{\sqrt{k_{1,0}}} \sum_{\iota=1}^{k_{1,0}} (y_t + z_t) \right\|^2 - \left\| \frac{1}{\sqrt{T - k_{1,0}}} \sum_{\iota=k_{1,0}+1}^{T} (y_t + z_t) \right\|^2, \end{aligned} \tag{A-32}$$

$$\varphi_{k_1}^1 = \frac{1}{T-k_1} vech[(k_{2,0}-k_{1,0})(\Sigma_1-\Sigma_2)+(T-k_{2,0})(\Sigma_1-\Sigma_3)], \quad \text{(A-33)}$$

$$\varphi_{k_1}^2 = \frac{1}{T-k_1} vech[(k_{1,0}-k_1)(\Sigma_2-\Sigma_1)+(T-k_{2,0})(\Sigma_2-\Sigma_3)], \quad \text{(A-34)}$$

$$\varphi_{k_1}^3 = \frac{1}{T-k_1} vech[(k_{1,0}-k_1)(\Sigma_3-\Sigma_1)+(k_{2,0}-k_{1,0})(\Sigma_3-\Sigma_2)]. \quad \text{(A-35)}$$

Since $\frac{1-\tau_{2,0}}{1-\tau_{1,0}} \|vech(\Sigma_2-\Sigma_3)\|^2 \leq \frac{\tau_{1,0}}{\tau_{2,0}} \|vech(\Sigma_1-\Sigma_2)\|^2$, $(1-\tau_{2,0})^2 \|vech(\Sigma_2-\Sigma_3)\|^2$ is smaller than $(1-\tau_{1,0})^2 \|vech(\Sigma_1-\Sigma_2)\|^2$, and thus for $k_1 \in W_\eta^c$ and $k_1 < k_{1,0}$, $\Pi_1(k_1) - \Pi(k_{1,0}) \geq cT$ for some c. On the other hand, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \Psi_1(k_1) = o_p(T)$ and $\Psi(k_{1,0}) = o_p(T)$ due to the following:

1. $\left\|\varphi_{k_1}^1\right\|$, $\left\|\varphi_{k_1}^2\right\|$ and $\left\|\varphi_{k_1}^3\right\|$ are uniformly bounded for $k_1 \in W_\eta^c$ and $k_1 < k_{1,0}$.
2. Using Hajek-Renyi inequality, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\sum_{\iota=k_1+1}^{k_{1,0}} y_t\right\|$, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\sum_{\iota=k_1+1}^{T} y_t\right\|$, $\left\|\sum_{\iota=k_{1,0}+1}^{k_{2,0}} y_t\right\|$ and $\left\|\sum_{\iota=k_{2,0}+1}^{T} y_t\right\|$ are all $O_p(\sqrt{T})$, $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\frac{1}{\sqrt{k_1}} \sum_{\iota=1}^{k_1} y_t\right\|$ is $O_p(\sqrt{\log T})$ and $\sup_{k_1 \in W_\eta^c, k_1 < k_{1,0}} \left\|\frac{1}{\sqrt{T-k_1}} \sum_{\iota=k_1+1}^{T} y_t\right\|$ is $O_p(1)$.
3. Using Lemma 5, $\sup_{1 \leq k < l \leq T} \left\|\sum_{\iota=k+1}^{l} z_t\right\|$ and $\sup_{1 \leq k < l \leq T} \left\|\frac{1}{\sqrt{l-k}} \sum_{\iota=k+1}^{l} z_t\right\|^2$ are both $o_p(T)$.

For $k_{1,0}+1 < k_1 \leq k_{2,0}$, after some calculation, we have:

$$\tilde{S}(k_1) - \tilde{S}(k_{1,0}) = \Pi_2(k_1) - \Pi(k_{1,0}) + \Psi_2(k_1) - \Psi(k_{1,0}), \quad \text{(A-36)}$$

where

$$\begin{aligned} & \Pi_2(k_1) - \Pi(k_{1,0}) \\ = \; & (k_1 - k_{1,0})[\frac{k_{1,0}}{k_1} \|vech(\Sigma_2-\Sigma_1)\|^2 - \frac{(T-k_{2,0})^2}{(T-k_1)(T-k_{1,0})} \|vech(\Sigma_3-\Sigma_2)\|^2] \\ \geq \; & (k_1 - k_{1,0})[\frac{k_{1,0}}{k_{2,0}} \|vech(\Sigma_2-\Sigma_1)\|^2 - \frac{T-k_{2,0}}{T-k_{1,0}} \|vech(\Sigma_3-\Sigma_2)\|^2], \end{aligned} \quad \text{(A-37)}$$

$$\begin{aligned}\Psi_2(k_1) &= 2\varphi_{k_1}^{4\prime}\sum_{\iota=1}^{k_{1,0}}(y_t+z_t)+2\varphi_{k_1}^{5\prime}\sum_{\iota=k_{1,0}+1}^{k_1}(y_t+z_t)+2\varphi_{k_1}^{6\prime}\sum_{\iota=k_1+1}^{k_{2,0}}(y_t+z_t)\\ &+2\varphi_{k_1}^{7\prime}\sum_{\iota=k_{2,0}+1}^{T}(y_t+z_t)-2[k_{1,0}\varphi_{k_1}^4+(k_1-k_{1,0})\varphi_{k_1}^5]'\frac{1}{k_1}\sum_{\iota=1}^{k_1}(y_t+z_t)\\ &-2[(k_{2,0}-k_1)\varphi_{k_1}^6+(T-k_{2,0})\varphi_{k_1}^7]'\frac{1}{T-k_1}\sum_{\iota=k_1+1}^{T}(y_t+z_t)\\ &-\left\|\frac{1}{\sqrt{k_1}}\sum_{\iota=1}^{k_1}(y_t+z_t)\right\|^2-\left\|\frac{1}{\sqrt{T-k_1}}\sum_{\iota=k_1+1}^{T}(y_t+z_t)\right\|^2, \qquad \text{(A-38)}\end{aligned}$$

$$\varphi_{k_1}^4 = \frac{k_1-k_{1,0}}{k_1}vech(\Sigma_1-\Sigma_2),\ \varphi_{k_1}^5=\frac{k_{1,0}}{k_1}vech(\Sigma_2-\Sigma_1), \qquad \text{(A-39)}$$

$$\varphi_{k_1}^6 = \frac{T-k_{2,0}}{T-k_1}vech(\Sigma_2-\Sigma_3),\ \varphi_{k_1}^7=\frac{k_{2,0}-k_1}{T-k_1}vech(\Sigma_3-\Sigma_2). \qquad \text{(A-40)}$$

The term in the bracket is positive, thus for $k_1 \in W_\eta^c$ and $k_{1,0}+1 < k_1 \le k_{2,0}$, $\Pi_2(k_1)-\Pi_2(k_{1,0}) \ge Tc$ for some c. Using the same argument as in the previous case, $\sup_{k_1\in W_\eta^c, k_{1,0}+1<k_1\le k_{2,0}} \Psi_2(k_1) = o_p(T)$.

For $k_{2,0} < k_1 \le T$, after some calculation, we have:

$$\tilde{S}(k_1)-\tilde{S}(k_{1,0}) = \Pi_3(k_1)-\Pi_3(k_{1,0})+\Psi_3(k_1)-\Psi_3(k_{1,0}). \qquad \text{(A-41)}$$

By symmetry, $\Pi_3(k_1)-\Pi_3(k_{2,0})$ has a similar expression as $\Pi_1(k_1)-\Pi_1(k_{1,0})$ and is positive. Thus $\Pi_3(k_1)-\Pi_3(k_{1,0}) \ge \Pi_3(k_{2,0})-\Pi_3(k_{1,0}) = (k_{2,0}-k_{1,0})[\frac{k_{1,0}}{k_{2,0}}\|vech(\Sigma_2-\Sigma_1)\|^2 - \frac{T-k_{2,0}}{T-k_{1,0}}\|vech(\Sigma_3-\Sigma_2)\|^2]$.

$$\begin{aligned}\Psi_3(k_1) &= 2\varphi_{k_1}^{8\prime}\sum_{\iota=1}^{k_{1,0}}(y_t+z_t)+2\varphi_{k_1}^{9\prime}\sum_{\iota=k_{1,0}+1}^{k_{2,0}}(y_t+z_t)+2\varphi_{k_1}^{10\prime}\sum_{\iota=k_{2,0}+1}^{k_1}(y_t+z_t)\\ &-2[k_{1,0}\varphi_{k_1}^8+(k_{2,0}-k_{1,0})\varphi_{k_1}^9+(k_1-k_{2,0})\varphi_{k_1}^{10}]'\frac{1}{k_1}\sum_{\iota=1}^{k_1}(y_t+z_t)\\ &-\left\|\frac{1}{\sqrt{k_1}}\sum_{\iota=1}^{k_1}(y_t+z_t)\right\|^2-\left\|\frac{1}{\sqrt{T-k_1}}\sum_{\iota=k_1+1}^{T}(y_t+z_t)\right\|^2, \qquad \text{(A-42)}\end{aligned}$$

and similarly $\sup_{k_1\in W_\eta^c, k_1>k_{2,0}} \Psi_3(k_1) = o_p(T)$. ■

E PROOF OF THEOREM 2

Proof. Using similar argument as proving Theorem 1, it suffices to show for any $\epsilon > 0$ and $\eta > 0$, there exist $C > 0$ such that $P(\min_{k_1 \in W_\eta, |k_1 - k_{1,0}| > C} \frac{\tilde{S}(k_1) - \tilde{S}(k_{1,0})}{|k_1 - k_{1,0}|} \leq 0) < \epsilon$ as $(N, T) \to \infty$.

First consider the case $k_1 < k_{1,0}$. Note that

$$\begin{aligned} \tilde{S}(k_1) &= \sum_{t=1}^{k_1} [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_{k_1})]' [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_{k_1})] \\ &\quad + \sum_{t=k_1+1}^{T} [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}^*_{k_1})]' [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}^*_{k_1})], \end{aligned} \tag{A-43}$$

where $\tilde{\Sigma}_{k_1} = \frac{1}{k_1} \sum_{t=1}^{k_1} \tilde{g}_t \tilde{g}_t'$ and $\tilde{\Sigma}^*_{k_1} = \frac{1}{T - k_1} \sum_{t=k_1+1}^{T} \tilde{g}_t \tilde{g}_t'$. Replacing $\tilde{\Sigma}_{k_{1,0}}$ by $\tilde{\Sigma}_{k_1}$ and $\tilde{\Sigma}^*_{k_{1,0}}$ by $\tilde{\Sigma}^*_{k_1}$ in the expression of $\tilde{S}(k_{1,0})$, $\tilde{S}(k_{1,0})$ is magnified. Thus

$$\begin{aligned} \frac{\tilde{S}(k_1) - \tilde{S}(k_{1,0})}{|k_1 - k_{1,0}|} &\geq \frac{1}{|k_1 - k_{1,0}|} \{ \sum_{t=k_1+1}^{k_{1,0}} [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}^*_{k_1})]' [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}^*_{k_1})] \\ &\quad - \sum_{t=k_1+1}^{k_{1,0}} [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_{k_1})]' [vech(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_{k_1})] \}. \end{aligned} \tag{A-44}$$

The right hand side equals

$$\begin{aligned} & vech(\Sigma_1 - \tilde{\Sigma}^*_{k_1})' vech(\Sigma_1 - \tilde{\Sigma}^*_{k_1}) - vech(\Sigma_1 - \tilde{\Sigma}_{k_1})' vech(\Sigma_1 - \tilde{\Sigma}_{k_1}) \\ & + 2 vech(\Sigma_1 - \tilde{\Sigma}^*_{k_1})' \frac{\sum_{t=k_1+1}^{k_{1,0}} (y_t + z_t)}{k_{1,0} - k_1} - 2 vech(\Sigma_1 - \tilde{\Sigma}_{k_1})' \frac{\sum_{t=k_1+1}^{k_{1,0}} (y_t + z_t)}{k_{1,0} - k_1} \\ = \; & \Xi_1 - \Xi_2 + \Xi_3 - \Xi_4. \end{aligned} \tag{A-45}$$

Plug in $\tilde{\Sigma}_{k_1}$ and $\tilde{\Sigma}^*_{k_1}$, we have

$$\begin{aligned}
\Xi_1 &= \left\| vech[\frac{k_{2,0}-k_{1,0}}{T-k_1}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_1}(\Sigma_1-\Sigma_3)] \right\|^2 + \left\| \frac{\sum_{t=k_1+1}^{T}(y_t+z_t)}{T-k_1} \right\|^2 \\
&\quad -2vech[\frac{k_{2,0}-k_{1,0}}{T-k_1}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_1}(\Sigma_1-\Sigma_3)]' \frac{\sum_{t=k_1+1}^{T}(y_t+z_t)}{T-k_1} \\
&= \Xi_{11}+\Xi_{12}-\Xi_{13}, \qquad \text{(A-46)} \\
\Xi_2 &= \left\| \frac{1}{k_1}\sum_{t=1}^{k_1}(y_t+z_t) \right\|^2, \qquad \text{(A-47)} \\
\Xi_3 &= 2vech[\frac{k_{2,0}-k_{1,0}}{T-k_1}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_1}(\Sigma_1-\Sigma_3)]' \frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t) \\
&\quad -2[\frac{1}{T-k_1}\sum_{t=k_1+1}^{T}(y_t+z_t)]' \frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t), \qquad \text{(A-48)} \\
\Xi_4 &= -2[\frac{1}{k_1}\sum_{t=1}^{k_1}(y_t+z_t)]' \frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}(y_t+z_t). \qquad \text{(A-49)}
\end{aligned}$$

If $vech[\frac{k_{2,0}-k_{1,0}}{T-k_{1,0}}(\Sigma_1-\Sigma_2)+\frac{T-k_{2,0}}{T-k_{1,0}}(\Sigma_1-\Sigma_3)]=0$, then $\Sigma_1-\Sigma_2=\frac{T-k_{2,0}}{T-k_{1,0}}(\Sigma_2-\Sigma_3)$, then $\frac{\tau_{1,0}}{\tau_{2,0}}\left\|vech(\Sigma_1-\Sigma_2)\right\|^2 = \frac{\tau_{1,0}(1-\tau_{2,0})^2}{\tau_{2,0}(1-\tau_{1,0})^2}\left\|vech(\Sigma_2-\Sigma_3)\right\|^2 < (\frac{1-\tau_{2,0}}{1-\tau_{1,0}})\left\|vech(\Sigma_2-\Sigma_3)\right\|^2$, this contradicts with Assumption 9. Thus $\Xi_{11}>c$ for some c. Using Hajek-Renyi inequality for y_t in each regime and Lemma 5 for z_t, $\sup_{k_1\in W_\eta, k_1<k_{1,0}-C}\left\|\frac{1}{T-k_1}\sum_{t=k_1+1}^{T}(y_t+z_t)\right\|$, $\sup_{k_1\in W_\eta, k_1<k_{1,0}-C}\left\|\frac{1}{k_1}\sum_{t=1}^{k_1}(y_t+z_t)\right\|$ and $\sup_{k_1\in W_\eta, k_1<k_{1,0}-C}\left\|\frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}z_t\right\|$ are all $o_p(1)$ while $\sup_{k_1\in W_\eta, k_1<k_{1,0}-C}\left\|\frac{1}{k_{1,0}-k_1}\sum_{t=k_1+1}^{k_{1,0}}y_t\right\|$ is $O_p(\frac{1}{\sqrt{C}})$. Thus for sufficiently large C, all the other terms are dominated by Ξ_{11}.

Next consider the case $k_1>k_{1,0}$. Using the same argument as the case $k_1<k_{1,0}$,

$$\begin{aligned}
\frac{\tilde{S}(k_1)-\tilde{S}(k_{1,0})}{|k_1-k_{1,0}|} &\geq \frac{1}{|k_1-k_{1,0}|}\{\sum_{t=k_{1,0}+1}^{k_1}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{k_1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}_{k_1})] \\
&\quad -\sum_{t=k_{1,0}+1}^{k_1}[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}^*_{k_1})]'[vech(\tilde{g}_t\tilde{g}_t'-\tilde{\Sigma}^*_{k_1})]\}, \qquad \text{(A-50)}
\end{aligned}$$

and the right hand side equals

$$
\begin{aligned}
& vech(\Sigma_2 - \tilde{\Sigma}_{k_1})' vech(\Sigma_2 - \tilde{\Sigma}_{k_1}) - vech(\Sigma_2 - \tilde{\Sigma}^*_{k_1})' vech(\Sigma_2 - \tilde{\Sigma}^*_{k_1}) \\
& +2vech(\Sigma_2 - \tilde{\Sigma}_{k_1})' \frac{\sum_{t=k_{1,0}+1}^{k_1}(y_t + z_t)}{k_1 - k_{1,0}} - 2vech(\Sigma_2 - \tilde{\Sigma}^*_{k_1})' \frac{\sum_{t=k_{1,0}+1}^{k_1}(y_t + z_t)}{k_1 - k_{1,0}} \\
= \; & \dot{\Xi}_1 - \dot{\Xi}_2 + \dot{\Xi}_3 - \dot{\Xi}_4. \qquad \text{(A-51)}
\end{aligned}
$$

Plug in $\tilde{\Sigma}_{k_1}$ and $\tilde{\Sigma}^*_{k_1}$, we have

$$
\begin{aligned}
\dot{\Xi}_1 &= \left\| \frac{k_{1,0}}{k_1} vech(\Sigma_2 - \Sigma_1) \right\|^2 + \left\| \frac{1}{k_1} \sum\nolimits_{t=1}^{k_1} (y_t + z_t) \right\|^2 \\
&\quad -2vech[\frac{k_{1,0}}{k_1}(\Sigma_2 - \Sigma_1)]' \frac{1}{k_1} \sum\nolimits_{t=1}^{k_1} (y_t + z_t) \\
&= \dot{\Xi}_{11} + \dot{\Xi}_{12} - \dot{\Xi}_{13}, \qquad \text{(A-52)} \\
\dot{\Xi}_2 &= \left\| \frac{T - k_{2,0}}{T - k_1} vech(\Sigma_2 - \Sigma_3) \right\|^2 + \left\| \frac{1}{T - k_1} \sum\nolimits_{t=k_1+1}^{T} (y_t + z_t) \right\|^2 \\
&\quad -2vech[\frac{T - k_{2,0}}{T - k_1}(\Sigma_2 - \Sigma_3)]' \frac{1}{T - k_1} \sum\nolimits_{t=k_1+1}^{T} (y_t + z_t) \\
&= \dot{\Xi}_{21} + \dot{\Xi}_{22} - \dot{\Xi}_{23}, \qquad \text{(A-53)} \\
\dot{\Xi}_3 &= 2vech[\frac{k_{1,0}}{k_1}(\Sigma_2 - \Sigma_1)]' \frac{1}{k_1 - k_{1,0}} \sum\nolimits_{t=k_{1,0}+1}^{k_1} (y_t + z_t) \\
&\quad -2[\frac{1}{k_1} \sum\nolimits_{t=1}^{k_1} (y_t + z_t)]' \frac{1}{k_1 - k_{1,0}} \sum\nolimits_{t=k_{1,0}+1}^{k_1} (y_t + z_t), \qquad \text{(A-54)} \\
\dot{\Xi}_4 &= 2vech[\frac{T - k_{2,0}}{T - k_1}(\Sigma_2 - \Sigma_3)]' \frac{1}{k_1 - k_{1,0}} \sum\nolimits_{t=k_{1,0}+1}^{k_1} (y_t + z_t) \\
&\quad -2[\frac{1}{T - k_1} \sum\nolimits_{t=k_1+1}^{T} (y_t + z_t)]' \frac{1}{k_1 - k_{1,0}} \sum\nolimits_{t=k_{1,0}+1}^{k_1} (y_t + z_t). \qquad \text{(A-55)}
\end{aligned}
$$

For $k_1 \in W_\eta$, $\dot{\Xi}_{11} - \dot{\Xi}_{21} \geq \left\| \frac{\tau_{1,0}}{\tau_{1,0}+\eta} vech(\Sigma_2 - \Sigma_1) \right\|^2 - \left\| \frac{1-\tau_{2,0}}{1-\tau_{1,0}-\eta} vech(\Sigma_2 - \Sigma_3) \right\|^2$. Thus by Assumption 9, $\dot{\Xi}_{11} - \dot{\Xi}_{21} \geq c$ for some $c > 0$ if η is sufficiently small. Again, using Hajek-Renyi inequality for y_t in each regime and Lemma 5 for z_t, all the other terms are dominated by $\dot{\Xi}_{11} - \dot{\Xi}_{21}$ for sufficiently large C. ■

F PROOF OF THEOREM 6

Proof. Since $\tilde{r}$ is consistent for r, we can treat r as known. It is not difficult to see that

$$\begin{aligned} SSNE_0 &= \sum_{t=1}^{T} vech(\tilde{f}_t\tilde{f}_t' - I_r)'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - I_r) \\ &\quad -Tvech(\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t' - I_r)'\tilde{\Omega}^{-1}vech(\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t' - I_r), \quad \text{(A-56)} \end{aligned}$$

and for any partition $(k_1, ..., k_l)$,

$$\begin{aligned} SSNE(k_1, ..., k_l) &= \sum_{t=1}^{T} vech(\tilde{f}_t\tilde{f}_t' - I_r)'\tilde{\Omega}^{-1}vech(\tilde{f}_t\tilde{f}_t' - I_r) - \sum_{\iota=1}^{l+1}(k_\iota - k_{\iota-1}) \\ &\quad vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\tilde{f}_t\tilde{f}_t'}{k_\iota - k_{\iota-1}} - I_r)'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\tilde{f}_t\tilde{f}_t'}{k_\iota - k_{\iota-1}} - I_r) \quad \text{(A-57)} \end{aligned}$$

Let $F_{NT}^* = SSNE_0 - SSNE(k_1, ..., k_l)$, it follows that

$$\begin{aligned} F_{NT}^* &= \sum_{\iota=1}^{l+1} vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{k_\iota - k_{\iota-1}}})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{k_\iota - k_{\iota-1}}}) \\ &\quad -vech(\frac{\sum_{t=1}^{T}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=1}^{T}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}}) \\ &= \sum_{\iota=1}^{l+1} D(k_{\iota-1}+1, k_\iota) - D(1, T) \\ &= \sum_{\iota=2}^{l+1}\{D(k_{\iota-1}+1, k_\iota) - [D(1, k_\iota) - D(1, k_{\iota-1})]\} \\ &= \sum_{\iota=1}^{l} F_{NT}^*(\iota+1). \quad \text{(A-58)} \end{aligned}$$

After some algebra, we have

$$\begin{aligned} F_{NT}^*(\iota+1) &= \frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1} - k_\iota)}vech[\frac{k_{\iota+1}}{T}\frac{\sum_{t=1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}} - \frac{k_\iota}{T}\frac{\sum_{t=1}^{k_{\iota+1}}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}}]' \\ &\quad \tilde{\Omega}^{-1}vech[\frac{k_{\iota+1}}{T}\frac{\sum_{t=1}^{k_\iota}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}} - \frac{k_\iota}{T}\frac{\sum_{t=1}^{k_{\iota+1}}(\tilde{f}_t\tilde{f}_t' - I_r)}{\sqrt{T}}] \\ &= \frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1} - k_\iota)}B(\tau_\iota, \tau_{\iota+1}; \tilde{F})'\tilde{\Omega}^{-1}B(\tau_\iota, \tau_{\iota+1}; \tilde{F}). \quad \text{(A-59)} \end{aligned}$$

Next, using four facts listed below, we have

$$\begin{aligned}
&\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\sum\nolimits_{\iota=1}^{l}\frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1}-k_\iota)}B(\tau_\iota,\tau_{\iota+1};\tilde{F})'(\tilde{\Omega}^{-1}-\Omega^{-1})B(\tau_\iota,\tau_{\iota+1};\tilde{F})\\
\leq\ &\frac{1}{\epsilon^3}\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\|\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\sum\nolimits_{\iota=1}^{l}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})\right\|^2=o_p(1)O_p(1)=o_p(1), \qquad \text{(A-60)}
\end{aligned}$$

$$\begin{aligned}
&\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\sum\nolimits_{\iota=1}^{l}\frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1}-k_\iota)}B(\tau_\iota,\tau_{\iota+1};\tilde{F})'\Omega^{-1}[B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0)]\\
\leq\ &\frac{l\left\|\Omega^{-1}\right\|}{\epsilon^3}\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})\right\|\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0)\right\|\\
=\ &O_p(1)o_p(1)=o_p(1). \qquad \text{(A-61)}
\end{aligned}$$

It follows that $F^*_{NT}=\sum_{\iota=1}^{l}\frac{T^3}{k_\iota k_{\iota+1}(k_{\iota+1}-k_\iota)}B(\tau_\iota,\tau_{\iota+1};FH_0)'\Omega^{-1}B(\tau_\iota,\tau_{\iota+1};FH_0)+o_p(1)$, where $o_p(1)$ is uniform and by Assumption 12 the first term converges weakly to $\sum_{\iota=1}^{l}\frac{1}{\tau_\iota\tau_{\iota+1}(\tau_{\iota+1}-\tau_\iota)}\left\|\tau_\iota W_{\frac{r(r+1)}{2}}(\tau_{\iota+1})-\tau_{\iota+1}W_{\frac{r(r+1)}{2}}(\tau_\iota)\right\|^2$.

1. $\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\|=o_p(1)$ if $\frac{d_T}{\delta_{NT}}\to 0$.
2. $\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0)\right\|=o_p(1)$ if $\frac{\sqrt{T}}{N}\to 0$.
3. $\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\|B(\tau_\iota,\tau_{\iota+1};FH_0)\|=O_p(1)$.
4. $\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})\right\|=O_p(1)$.

Fact (1) follows from Lemma 8.

Proof of (2): Note that

$$\begin{aligned}
&B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH_0)\\
=\ &B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH)+B(\tau_\iota,\tau_{\iota+1};FH)-B(\tau_\iota,\tau_{\iota+1};FH_0)\\
=\ &vech[\frac{k_{\iota+1}}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_\iota}(\tilde{f}_t\tilde{f}_t'-H^{'}f_tf_t'H)-\frac{k_\iota}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_{\iota+1}}(\tilde{f}_t\tilde{f}_t'-H'f_tf_t'H)]\\
&+vech[\frac{k_{\iota+1}}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_\iota}(H^{'}(f_tf_t'-\Sigma_F)H-H_0'(f_tf_t'-\Sigma_F)H_0)\\
&-\frac{k_\iota}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k_{\iota+1}}(H^{'}(f_tf_t'-\Sigma_F)H-H_0'(f_tf_t'-\Sigma_F)H_0)]. \qquad \text{(A-62)}
\end{aligned}$$

It is not difficult to see

$$\begin{aligned}
&\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\left\|B(\tau_\iota,\tau_{\iota+1};\tilde{F})-B(\tau_\iota,\tau_{\iota+1};FH)\right\| \\
\leq\ & 2\sup_{T\epsilon\leq k\leq T(1-\epsilon)}\left\|\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k}(\tilde{f}_t\tilde{f}_t'-H'f_tf_t'H)\right\|=O_p(\frac{\sqrt{T}}{\delta_{NT}^2}) \qquad \text{(A-63)}
\end{aligned}$$

by Lemma 7, and

$$\begin{aligned}
&\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\|B(\tau_\iota,\tau_{\iota+1};FH)-B(\tau_\iota,\tau_{\iota+1};FH_0)\| \\
\leq\ & 2\sup_{T\epsilon\leq k\leq T(1-\epsilon)}\left\|\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{k}(H^{'}(f_tf_t'-\Sigma_F)H-H_0'(f_tf_t'-\Sigma_F)H_0)\right\|=o_p(1) \qquad \text{(A-64)}
\end{aligned}$$

by part (2) of Lemma 6 and Assumption 12.

Proof of (3): Note that $B(\tau_\iota,\tau_{\iota+1};FH_0)=vech[\frac{k_{\iota+1}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_\iota}(H_0'f_tf_t'H_0-I_r)-\frac{k_\iota}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota+1}}(H_0'f_tf_t'H_0-I_r)]$, it is not difficult to see $\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}\|B(\tau_\iota,\tau_{\iota+1};FH_0)\|\leq 2\sup_{T\epsilon\leq k\leq T(1-\epsilon)}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}(H_0'f_tf_t'H_0-I_r)\right\|$, which is $O_p(1)$ by Assumption 12.

Proof of (4): It follows directly from (2) and (3). ■

G PROOF OF THEOREM 7

Proof. Under the alternative, the estimated number of factors converges to the number of pseudo factors and the estimated factors are pseudo factors, g_t. First note that $\sup_{(\tau_1,...,\tau_l)\in\Lambda_\epsilon}[SSNE_0-SSNE(k_1,...,k_l)]\geq SSNE_0-SSNE(k_{1,0},...,k_{l0})$, thus it suffices to show the latter goes to infinity in probability.

$$\begin{aligned}
SSNE(k_{1,0},...,k_{l0}) &= \sum\nolimits_{t=1}^{T}vech(\tilde{g}_t\tilde{g}_t')'\tilde{\Omega}^{-1}vech(\tilde{g}_t\tilde{g}_t')-\sum\nolimits_{\iota=1}^{l+1}(k_{\iota0}- \\
&\quad k_{\iota-1,0})vech(\frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota0}-k_{\iota-1,0}})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota0}-k_{\iota-1,0}}) \qquad \text{(A-65)} \\
SSNE_0 &= \sum\nolimits_{t=1}^{T}vech(\tilde{g}_t\tilde{g}_t')'\tilde{\Omega}^{-1}vech(\tilde{g}_t\tilde{g}_t') \\
&\quad -Tvech(\frac{\sum_{t=1}^{T}\tilde{g}_t\tilde{g}_t'}{T})'\tilde{\Omega}^{-1}vech(\frac{\sum_{t=1}^{T}\tilde{g}_t\tilde{g}_t'}{T}). \qquad \text{(A-66)}
\end{aligned}$$

Thus similar to (A-59), $SSNE_0 - SSNE(k_{1,0}, ..., k_{l0})$ can be written as

$$\sum_{\iota=1}^{l} \frac{T^3}{k_{\iota 0}k_{\iota+1,0}(k_{\iota+1,0}-k_{\iota 0})} vech(\frac{k_{\iota+1,0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t' - \frac{k_{\iota 0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota+1,0}}$$
$$\tilde{g}_t\tilde{g}_t')'\tilde{\Omega}^{-1}vech(\frac{k_{\iota+1,0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t' - \frac{k_{\iota 0}}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t')' \geq \frac{1}{\rho_{\max}(\tilde{\Omega})}$$
$$\sum_{\iota=1}^{l} \frac{T^3}{k_{\iota 0}k_{\iota+1,0}(k_{\iota+1,0}-k_{\iota 0})} \left\| vech(\frac{k_{\iota+1,0}}{T}\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}} - \frac{k_{\iota 0}}{T}\frac{\sum_{t=1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}}) \right\|^2 \text{(A-67)}$$

where $\rho_{\max}(\tilde{\Omega})$ is the maximal eigenvalue of $\tilde{\Omega}$. Note that $\frac{k_{\iota+1,0}}{T}\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}} - \frac{k_{\iota 0}}{T}\frac{\sum_{t=1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{\sqrt{T}} = \frac{(k_{\iota+1,0}-k_{\iota 0})k_{\iota 0}}{T^{\frac{3}{2}}}(\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}} - \frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota+1,0}-k_{\iota 0}})$, thus $SSNE_0 - SSNE(k_{1,0}, ..., k_{l0})$ is not smaller than

$$\sum_{\iota=1}^{l} \frac{(k_{\iota+1,0}-k_{\iota 0})k_{\iota 0}}{k_{\iota+1,0}\rho_{\max}(\tilde{\Omega})} \left\| vech(\frac{\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota 0}} - \frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t'}{k_{\iota+1,0}-k_{\iota 0}}) \right\|^2. \quad \text{(A-68)}$$

Recall that $vech(\tilde{g}_t\tilde{g}_t') = b_t + y_t + z_t$, by Assumption 1, $\frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} y_t = vech(J_0'R_\iota\frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}(f_tf_t' - \Sigma_F)R_\iota'J_0) = o_p(1)$ for each ι, and by Lemma 5, $\frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} z_t = o_p(1)$ for each ι. Thus $\frac{1}{k_{\iota 0}}\sum_{t=1}^{k_{\iota 0}}\tilde{g}_t\tilde{g}_t' - \frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}}\tilde{g}_t\tilde{g}_t' = \frac{1}{k_{\iota 0}}\sum_{t=1}^{k_{\iota 0}} b_t - \frac{1}{k_{\iota+1,0}-k_{\iota 0}}\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} b_t + o_p(1)$. Recall that $b_t = vech(J_0'R_\iota\Sigma_F R_\iota J_0)$ for $k_{\iota-1,0} < t \leq k_{\iota,0}$ and b_t is different in different regime, thus $\frac{\sum_{t=1}^{k_{\iota 0}} b_t}{k_{\iota 0}} - \frac{\sum_{t=k_{\iota 0}+1}^{k_{\iota+1,0}} b_t}{k_{\iota+1,0}-k_{\iota 0}} \neq 0$ for some ι. It follows that there exists some $c > 0$ such that $SSNE_0 - SSNE(k_{1,0}, ..., k_{l0}) \geq \frac{Tc}{\rho_{\max}(\tilde{\Omega})}$ with probability approaching one. Next, it is not difficult to see that under the alternative $\rho_{\max}(\tilde{\Omega}) = O_p(d_T)$ for the three kernel functions considered in this paper[18], since HAC method is used to estimate $\tilde{\Omega}$ while under the alternative $\tilde{g}_t\tilde{g}_t'$ is not properly centered. Noting that $\frac{d_T}{T} \to 0$, the result is proved. ■

H PROOF OF THEOREM 8

Proof. It is not difficult to see that $F_{NT}(l+1\,|l) = \sup_{1\leq\iota\leq l+1} \sup_{k\in\Lambda_{\iota,\eta}} [SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota)]$, where $SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota)$ is the sum of squared normalized error of the ι-th regime. Thus testing l versus $l+1$ changes is essentially testing jointly 0

[18] For more detailed procedure, also see the proof of Theorem 2 in Han and Inoue (2015).

versus 1 change in each regime. In what follows, we reestablish Theorem 6 with $l = 1$ but $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$. Similar to (A-58), we have

$$\begin{aligned}
& SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota) \\
= \; & vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{k - \tilde{k}_{\iota-1}}})'\tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{k - \tilde{k}_{\iota-1}}}) \\
& + vech(\frac{\sum_{t=k+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - k}})'\tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=k+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - k}}) \\
& - vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}})'\tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}}), \quad \text{(A-69)}
\end{aligned}$$

and similar to (A-59),

$$\begin{aligned}
& SSNE_\iota(\tilde{k}_{\iota-1}, \tilde{k}_\iota) - SSNE_\iota(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota) \\
= \; & \frac{1}{\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\frac{\tilde{k}_\iota - k}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}} - \frac{k - \tilde{k}_{\iota-1}}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}})' \\
& \tilde{\Omega}_\iota^{-1} vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}} - \frac{k - \tilde{k}_{\iota-1}}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - I_{r_\iota})}{\sqrt{\tilde{k}_\iota - \tilde{k}_{\iota-1}}}) \\
= \; & \frac{1}{\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\frac{\tilde{k}_\iota - k}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}} C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota)'\tilde{\Omega}_\iota^{-1} C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota). \quad \text{(A-70)}
\end{aligned}$$

Since $\tilde{k}_\iota - k_{\iota 0} = O_p(1)$, asymptotically it suffices to consider the case that $\left|\tilde{k}_\iota - k_{\iota 0}\right| \leq C$ for some integer C and all ι. And in such case $\Lambda_{\iota,\eta} \subset (k_{\iota-1,0}, k_{\iota 0}]$ for large T. Next, based on these two properties and using four facts listed below,

$\sup_{k\in\Lambda_{\iota,\eta}} \left\| (\frac{1}{\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\frac{\tilde{k}_\iota - k}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}} - \frac{1}{\frac{k-k_{\iota-1,0}}{k_{\iota 0} - k_{\iota-1,0}}\frac{k_{\iota 0} - k}{k_{\iota 0} - k_{\iota-1,0}}}) C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota)'\tilde{\Omega}_\iota^{-1} C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota) \right\|$,
$\sup_{k\in\Lambda_{\iota,\eta}} \left\| C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota)'(\tilde{\Omega}_\iota^{-1} - \Omega_\iota^{-1}) C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota) \right\|$,
$\sup_{k\in\Lambda_{\iota,\eta}} \left\| C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota)'\Omega_\iota^{-1}(C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota) - C(k_{\iota-1,0}, k, k_{\iota 0}; F_{\iota 0}H_{\iota 0})) \right\|$,
$\sup_{k\in\Lambda_{\iota,\eta}} \left\| C(k_{\iota-1,0}, k, k_{\iota 0}; F_{\iota 0}H_{\iota 0})'\Omega_\iota^{-1}(C(\tilde{k}_{\iota-1}, k, \tilde{k}_\iota; \tilde{F}_\iota) - C(k_{\iota-1,0}, k, k_{\iota 0}; F_{\iota 0}H_{\iota 0})) \right\|$ are

all $o_p(1)$. Thus

$$\begin{aligned}
&\sup_{k\in\Lambda_{\iota,\eta}}\left[SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota)-SSNE_\iota(\tilde{k}_{\iota-1},k,\tilde{k}_\iota)\right]\\
=&\sup_{k\in\Lambda_{\iota,\eta}}\frac{1}{\frac{k-k_{\iota-1,0}}{k_{\iota0}-k_{\iota-1,0}}\frac{k_{\iota0}-k}{k_{\iota0}-k_{\iota-1,0}}}C(k_{\iota-1,0},k,k_{\iota0};F_{\iota0}H_{\iota0})'\Omega_\iota^{-1}C(k_{\iota-1,0},k,k_{\iota0};F_{\iota0}H_{\iota0})+o_p(1)\\
=&\sup_{k\in\Lambda_{\iota,\eta}}F_{NT,\iota}(k)+o_p(1). \qquad \text{(A-71)}
\end{aligned}$$

By Assumption 12, with $k=[T\tau]$, $F_{NT,\iota}(k)\Rightarrow\frac{1}{\tau(1-\tau)}\left\|W_{\frac{r_\iota(r_\iota+1)}{2}}(\tau)-\tau W_{\frac{r_\iota(r_\iota+1)}{2}}(1)\right\|^2$ for $\tau\in(0,1)$. Furthermore, since Wiener process has independent increments, the limit process of $F_{NT,\iota}(k)$ is independent with each other for different ι. Finally, define $\Lambda^0_{\iota,\eta}=\{k:k_{\iota-1,0}+(k_{\iota0}-k_{\iota-1,0})\eta\le k\le k_{\iota0}-(k_{\iota0}-k_{\iota-1,0})\eta\}$. For any $\eta_1<\eta<\eta_2$, $\Lambda^0_{\iota,\eta_2}\subset\Lambda_{\iota,\eta}\subset\Lambda^0_{\iota,\eta_1}$ for large T, thus $\sup_{k\in\Lambda^0_{\iota,\eta_2}}F_{NT,\iota}(k)\le\sup_{k\in\Lambda_{\iota,\eta}}F_{NT,\iota}(k)\le\sup_{k\in\Lambda^0_{\iota,\eta_1}}F_{NT,\iota}(k)$. Since η_1 and η_2 can be arbitrarily close to η, $\sup_{k\in\Lambda_{\iota,\eta}}F_{NT,\iota}(k)$ has the same distribution as $\sup_{k\in\Lambda^0_{\iota,\eta}}F_{NT,\iota}(k)$. Taking together, we have the desired results.

1. $\left\|\tilde{\Omega}_\iota^{-1}-\Omega_\iota^{-1}\right\|=o_p(1)$ if $\frac{d_T}{T^{\frac{1}{4}}}\to0$ and $\frac{d_T}{\sqrt{N}}\to0$.
2. $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_\iota;\tilde{F}_\iota)-C(k_{\iota-1,0},k,k_{\iota0};F_{\iota0}H_{\iota0})\right\|=o_p(1)$ if $\frac{\sqrt{T}}{N}\to0$.
3. $\sup_{k\in\Lambda_{\iota,\eta}}\|C(k_{\iota-1,0},k,k_{\iota0};F_{\iota0}H_{\iota0})\|=O_p(1)$.
4. $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_\iota;\tilde{F}_\iota)\right\|=O_p(1)$.

Fact (1) follows from Lemma 11.

Proof of (2): Note that

$$
\begin{aligned}
& C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota 0}) \\
= \;& [C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})] \\
& +[C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota 0})] \\
= \;& vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}-\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}) \\
& +vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota 0}'(f_tf_t'-\Sigma_F)H_{\iota 0})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}} \\
& -\frac{k-\tilde{k}_{\iota-1}}{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota 0}'(f_tf_t'-\Sigma_F)H_{\iota 0})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}). \qquad \text{(A-72)}
\end{aligned}
$$

Thus it's not difficult to see $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};\tilde{F}_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})\right\|$ is not larger than $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}\right\|+\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(\tilde{f}_{\iota t}\tilde{f}_{\iota t}'-H_{\iota}'f_tf_t'H_{\iota})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}\right\|$, which is $O_p(\frac{\sqrt{T}}{\delta_{NT}^2})$ by Lemma 10. And $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota})-C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota 0})\right\|$ is not larger than $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota 0}'(f_tf_t'-\Sigma_F)H_{\iota 0})}{\sqrt{k-\tilde{k}_{\iota-1}}}\right\|+\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_{\iota}}(H_{\iota}'(f_tf_t'-\Sigma_F)H_{\iota}-H_{\iota 0}'(f_tf_t'-\Sigma_F)H_{\iota 0})}{\sqrt{\tilde{k}_{\iota}-\tilde{k}_{\iota-1}}}\right\|$, which is $o_p(1)$ by part (2) of Lemma 9 and Assumption 12. Finally, with $\left|\tilde{k}_{\iota}-k_{\iota 0}\right|\leq C$ for all ι, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|C(\tilde{k}_{\iota-1},k,\tilde{k}_{\iota};F_{\iota}H_{\iota 0})-C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})\right\|=o_p(1)$ is obvious.

Proof of (3): Note that $C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})=vech(\frac{\sum_{t=k_{\iota-1,0}+1}^{k}(H_{\iota 0}'f_tf_t'H_{\iota 0}-I_{r_\iota})}{\sqrt{k_{\iota 0}-k_{\iota-1,0}}}-\frac{k-k_{\iota-1,0}}{k_{\iota 0}-k_{\iota-1,0}}\frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota 0}}(H_{\iota 0}'f_tf_t'H_{\iota 0}-I_{r_\iota})}{\sqrt{k_{\iota 0}-k_{\iota-1,0}}})$, for some $\eta_1<\eta$, $\sup_{k\in\Lambda_{\iota,\eta}}\|C(k_{\iota-1,0},k,k_{\iota 0};F_{\iota 0}H_{\iota 0})\|\leq \sup_{k\in\Lambda_{\iota,\eta_1}^0}\left\|\frac{\sum_{t=k_{\iota-1,0}+1}^{k}(H_{\iota 0}'f_tf_t'H_{\iota 0}-I_{r_\iota})}{\sqrt{k_{\iota 0}-k_{\iota-1,0}}}\right\|+\left\|\frac{\sum_{t=k_{\iota-1,0}+1}^{k_{\iota 0}}(H_{\iota 0}'f_tf_t'H_{\iota 0}-I_{r_\iota})}{\sqrt{k_{\iota 0}-k_{\iota-1,0}}}\right\|$, which is $O_p(1)$ by Assumption 12.

Proof of (4): It follows directly from (2) and (3). ■

I PROOF OF THEOREM 9

Proof. The calculation of $SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota)-SSNE_\iota(\tilde{k}_{\iota-1},k,\tilde{k}_\iota)$ under the null is still valid under the alternative. Thus following (A-70) we have

$$\begin{aligned}
& F_{NT}(l+1\,|l) \\
\geq\ & \sup_{k\in\Lambda_{\iota,\eta}}[SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota)-SSNE_\iota(\tilde{k}_{\iota-1},k,\tilde{k}_\iota)] \\
\geq\ & SSNE_\iota(\tilde{k}_{\iota-1},\tilde{k}_\iota)-SSNE_\iota(\tilde{k}_{\iota-1},k_{\iota0},\tilde{k}_\iota) \\
\geq\ & \frac{1}{\frac{k_{\iota0}-\tilde{k}_{\iota-1}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\frac{\tilde{k}_\iota-k_{\iota0}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}}\frac{1}{\rho_{\max}(\tilde{\Omega}_\iota)}\left\|vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota0}}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\sqrt{\tilde{k}_\iota-\tilde{k}_{\iota-1}}}-\frac{k_{\iota0}-\tilde{k}_{\iota-1}}{\tilde{k}_\iota-\tilde{k}_{\iota-1}}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\sqrt{\tilde{k}_\iota-\tilde{k}_{\iota-1}}})\right\|^2 \\
=\ & \frac{(k_{\iota0}-\tilde{k}_{\iota-1})(\tilde{k}_\iota-k_{\iota0})}{(\tilde{k}_\iota-\tilde{k}_{\iota-1})\rho_{\max}(\tilde{\Omega}_\iota)}\left\|vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota0}}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{k_{\iota0}-\tilde{k}_{\iota-1}}-\frac{\sum_{t=k_{\iota0}+1}^{\tilde{k}_\iota}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\tilde{k}_\iota-k_{\iota0}})\right\|^2. \qquad \text{(A-73)}
\end{aligned}$$

Define $z_{\iota t}=vech(\tilde{g}_{\iota t}\tilde{g}'_{\iota t}-J'_{\iota0}g_{\iota t}g'_{\iota t}J_{\iota0})$. By Lemma 13 and Assumption 1,

$$\begin{aligned}
& vech(\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota0}}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{k_{\iota0}-\tilde{k}_{\iota-1}}) \\
=\ & \frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota0}}z_{\iota t}}{k_{\iota0}-\tilde{k}_{\iota-1}}+vech[J'_{\iota0}A_{\iota1}\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota0}}(f_tf'_t-\Sigma_F)}{k_{\iota0}-\tilde{k}_{\iota-1}}A'_{\iota1}J_{\iota0}]+vech(J'_{\iota0}A_{\iota1}\Sigma_FA'_{\iota1}J_{\iota0}) \\
=\ & vech(J'_{\iota0}A_{\iota1}\Sigma_FA'_{\iota1}J_{\iota0})+o_p(1), \qquad \text{(A-74)}
\end{aligned}$$

and similarly $vech(\frac{\sum_{t=k_{\iota0}+1}^{\tilde{k}_\iota}\tilde{g}_{\iota t}\tilde{g}'_{\iota t}}{\tilde{k}_\iota-k_{\iota0}})=vech(J'_{\iota0}A_{\iota2}\Sigma_FA'_{\iota2}J_{\iota0})+o_p(1)$. Since $A_{\iota1}\Sigma_FA'_{\iota1}\neq A_{\iota2}\Sigma_FA'_{\iota2}$ and $\rho_{\max}(\tilde{\Omega}_\iota)=O_p(d_T)$, there exists some $c>0$ such that $F_{NT}(l+1\,|l)\geq\frac{Tc}{d_T}$ with probability approaching one. ■

J PROOF OF LEMMAS

Lemma 1 *Under Assumption 7(1), Hajek-Renyi inequality applies to the process* $\{y_t, t=k_{\kappa-1,0}+1,...,k_{\kappa,0}\}$ *and* $\{y_t, t=k_{\kappa,0},...,k_{\kappa-1,0}+1\}$, $\kappa=1,...,L+1$.

Proof. Note that $y_t = vech(J_0' R_\kappa (f_t f_t' - \Sigma_F) R_\kappa' J_0)$ for $k_{\kappa-1,0} < k \leq k_{\kappa,0}$, thus $P(\sup_{k_{\kappa-1,0}+m \leq k \leq k_{\kappa,0}} c_k \left\| \sum_{t=k_{\kappa-1,0}+1}^{k} y_t \right\| > M)$ is controlled by

$$P(\|J_0' R_\kappa\|^2 \sup_{k_{\kappa-1,0}+m \leq k \leq k_{\kappa,0}} c_k \left\| \sum_{t=k_{\kappa-1,0}+1}^{k} \epsilon_t \right\| > M),$$

which is not larger than $\frac{C}{M^2}(mc_{k_{\kappa-1,0}+m}^2 + \sum_{k=k_{\kappa-1,0}+m+1}^{k_{\kappa,0}} c_k^2)$ by Hajek-Renyi inequality for process $\{\epsilon_t, t = k_{\kappa-1,0}+1, ..., k_{\kappa,0}\}$. Other processes can be proved similarly. ■

Lemma 2 *In case factor loadings have structural changes, under Assumptions 1-6,* $\|J - J_0\| = o_p(1)$ *and* $\|V_{NT} - V\| = o_p(1)$.

Proof. The proof follows similar procedure as Proposition 1 in Bai (2003), with J, J_0 and g_t corresponding to H, H_0 and f_t respectively. To avoid repetition, we will only sketch the main steps. In Bai (2003), proof of Proposition 1 relies on $d_{NT} = o_p(1)$ and $V_{NT}^* \xrightarrow{p} V$ (Bai's notation). The former relies on Lemma A.1 and A.3(i)[19] while the latter relies on Lemma A.3(ii). Lemma A.1 relies on Theorem 1 of Bai and Ng (2002) and Lemma A.3(i). Lemma A.3(ii) relies on Lemma A.3(i) and Lemma 1(ii) of Bai and Ng (2002). Thus it suffices to prove Lemma 1(ii) and Theorem 1 of Bai and Ng (2002) and Lemma A.3(i) of Bai (2003). In current context, the first can be proved using Assumption 2 and Assumption 4(2), the second can be proved using Assumptions 1-4, and the third can be proved using Assumption 5 and Weyl inequality. Finally, Assumption 6 ensures uniqueness of J_0. ■

Lemma 3 *Under Assumptions 1 and 7,*

1. $\sup_{0 \leq k < l \leq T} \frac{1}{\sqrt{T(l-k)}} \sum_{t=k+1}^{l} \|g_t\|^2 = O_p(1)$,
2. $\sup_{k_{\iota-1,0} < l \leq k_{\iota 0}} \frac{1}{l - k_{\iota-1,0}} \sum_{t=k_{\iota-1,0}+1}^{l} \|g_t\|^2 = O_p(1)$,
3. $\sup_{k_{\iota-1,0} \leq k < k_{\iota 0}} \frac{1}{k_{\iota 0} - k} \sum_{t=k+1}^{k_{\iota 0}} \|g_t\|^2 = O_p(1)$.

[19] In Bai (2003), Bai states that it relies on Lemma A.2, but in fact Lemma A.1 and A.3(i) is enough. This is because $d_{NT} = (\frac{\Lambda^{0\prime}\Lambda^0}{N})^{\frac{1}{2}} \frac{F^{0\prime}}{T} (\tilde{F} - F^0 H) V_{NT}$.

Proof. We first prove part (2). Recall that $g_t = R_\iota f_t$ for $k_{\iota-1,0} < t \leq k_{\iota 0}$, thus $\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}} \frac{\sum_{t=k_{\iota-1,0}+1}^{l}\|g_t\|^2}{l-k_{\iota-1,0}} \leq \|R_\iota\|^2 \mathbb{E}\|f_t\|^2 + \|R_\iota\|^2 \sup_{k_{\iota-1,0}<l\leq k_{\iota 0}} \frac{\sum_{t=k_{\iota-1,0}+1}^{l}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)}{l-k_{\iota-1,0}}$, where $\mathbb{E}\|f_t\|^2 = tr\Sigma_F$. It suffices to show the second term is $O_p(1)$. Let $D_l = \frac{\sum_{t=k_{\iota-1,0}+1}^{l}(f_t f_t' - \Sigma_F)}{l-k_{\iota-1,0}}$, it follows that $\left|\frac{\sum_{t=k_{\iota-1,0}+1}^{l}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)}{l-k_{\iota-1,0}}\right| = |trD_l| \leq \sqrt{r_\iota}(trD_l^2)^{\frac{1}{2}} = \sqrt{r_\iota}\|D_l\|$, thus $\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}}\left|\frac{\sum_{t=k_{\iota-1,0}+1}^{l}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)}{l-k_{\iota-1,0}}\right| \leq \sqrt{r_\iota}\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}}\left\|\frac{\sum_{t=k_{\iota-1,0}+1}^{l}\epsilon_t}{l-k_{\iota-1,0}}\right\|$, which is $O_p(1)$ by Hajek-Renyi inequality. Proof of part (3) is similar and omitted.

Now we prove part (1). The whole sample $t = 1, ..., T$ is divided into several nonoverlapping segments by the true change points. First consider the case that k and l lie in two different segments. Without loss of generality, suppose k lies in the ι-th segment and l lies in the κ-th segment, then $\sup_{k_{\iota-1,0}<k\leq k_{\iota 0};k_{\kappa-1,0}<l\leq k_{\kappa 0}} \frac{\sum_{t=k+1}^{l}\|g_t\|^2}{\sqrt{T(l-k)}}$ is no larger than $\sup_{k_{\iota-1,0}<k<k_{\iota 0}} \frac{\sum_{t=k+1}^{k_{\iota 0}}\|g_t\|^2}{k_{\iota 0}-k} + \frac{\sum_{t=k_{\iota 0}+1}^{k_{\kappa-1,0}}\|g_t\|^2}{k_{\kappa-1,0}-k_{\iota 0}} + \sup_{k_{\kappa-1,0}<l\leq k_{\kappa 0}} \frac{\sum_{t=k_{\kappa-1,0}+1}^{l}\|g_t\|^2}{l-k_{\iota-1,0}}$(If $\kappa - 1 = \iota$, the second term is zero). By parts (2) and (3), the first term and the third term are $O_p(1)$. The second term is no larger than $\sum_{\upsilon=\iota+1}^{\kappa-1}\|R_\upsilon\|^2 (\frac{1}{k_{\upsilon,0}-k_{\upsilon-1,0}}\sum_{t=k_{\upsilon-1,0}+1}^{k_{\upsilon,0}}\|f_t\|^2)$, which is $O_p(1)$. Next consider the case that k and l lie in the same segment. Without loss of generality, suppose they lie in the ι-th segment, then $\sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}} \frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\|g_t\|^2$ is no larger than $\|R_\iota\|^2 \mathbb{E}\|f_t\|^2 + \|R_\iota\|^2 \sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}}\left|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\|f_t\|^2 - \mathbb{E}\|f_t\|^2)\right|$. Similar to part (2), the second term is no larger than $\|R_\iota\|^2\sqrt{r_\iota}\sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}}\left\|\frac{\sum_{t=k+1}^{l}\epsilon_t}{\sqrt{T(l-k)}}\right\|$, which is $o_p(1)$ since by Assumption 7,

$$\begin{aligned} & \mathbb{E}\Big(\sup_{k_{\kappa-1,0}<k<l\leq k_{\iota 0}}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\epsilon_t\right\|^{4+\delta}\Big) \\ = & \frac{1}{T^{2+\frac{\delta}{2}}}\sum_{k=k_{\kappa-1,0}}^{k_{\kappa,0}-1}\sum_{l=k+1}^{k_{\kappa,0}}\mathbb{E}\Big(\left\|\frac{1}{\sqrt{l-k}}\sum_{t=k+1}^{l}\epsilon_t\right\|^{4+\delta}\Big) \leq \frac{M}{T^{\frac{\delta}{2}}}. \end{aligned} \quad \text{(A-75)}$$

Up to now, we have proved the desired result for each possible case. Since the number of cases is finite, the supremum among all $0 \leq k < l \leq T$ will also be $O_p(1)$. ■

Lemma 4 *Under Assumptions 1-8,*

1. $\sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| = O_p(\frac{1}{\delta_{NT}^2})$,

2. $\sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\tilde{g}_t - J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}})$,

3. $\sup_{k_{\iota-1,0}<l\le k_{\iota 0}}\left\|\frac{1}{l-k_{\iota-1,0}}\sum_{t=k_{\iota-1,0}+1}^{l}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| = O_p(\frac{1}{\delta_{NT}^2})$ *for each* ι,

4. $\sup_{k_{\iota-1,0}<l\le k_{\iota 0}}\left\|\frac{1}{l-k_{\iota-1,0}}\sum_{t=k_{\iota-1,0}+1}^{l}(\tilde{g}_t - J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}})$ *for each* ι,

5. $\sup_{k_{\iota-1,0}\le k<k_{\iota 0}}\left\|\frac{1}{k_{\iota 0}-k}\sum_{t=k+1}^{k_{\iota 0}}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| = O_p(\frac{1}{\delta_{NT}^2})$ *for each* ι,

6. $\sup_{k_{\iota-1,0}\le k<k_{\iota 0}}\left\|\frac{1}{k_{\iota 0}-k}\sum_{t=k+1}^{k_{\iota 0}}(\tilde{g}_t - J'g_t)g_t'J\right\| = O_p(\frac{1}{\delta_{NT}})$ *for each* ι.

Proof. Following Bai (2003), we have

$$\tilde{g}_t - J'g_t = V_{NT}^{-1}(\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\xi_{st}), \tag{A-76}$$

where $\zeta_{st} = \frac{e_s'e_t}{N} - \gamma_N(s,t)$, $\eta_{st} = \frac{g_s'\Gamma'e_t}{N}$and $\xi_{st} = \frac{g_t'\Gamma'e_s}{N}$. V_{NT} is the diagonal matrix of the first $\bar{r}$ largest eigenvalues of $\frac{1}{NT}XX'$ in decreasing order, $\tilde{G}$ is $\sqrt{T}$ times the corresponding eigenvector matrix, V is the diagonal matrix of eigenvalues of $\Sigma_\Gamma^{\frac{1}{2}}\Sigma_G\Sigma_\Gamma^{\frac{1}{2}}$ and Φ is the corresponding eigenvector matrix, $J = \frac{\Gamma'\Gamma}{N}\frac{G'\tilde{G}}{T}V_{NT}^{-1}$. First consider part (1).

$$\begin{aligned}
&\sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'\right\| \\
\le\ & 4\sup_{0\le k<l\le T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\gamma_N(s,t)\right\|^2 + \left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\zeta_{st}\right\|^2 \\
&+ \left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\eta_{st}\right\|^2 + \left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{g}_s\xi_{st}\right\|^2)\left\|V_{NT}^{-1}\right\|^2 \\
=\ & 4\left\|V_{NT}^{-1}\right\|^2(I + II + III + IV).
\end{aligned} \tag{A-77}$$

By part (1) of Lemma 2, $\left\|V_{NT}^{-1}\right\| \to \|V^{-1}\|$, thus it suffices to consider I, II, III and IV. By Assumption 4,

$$\begin{aligned} I &\leq \frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2 \sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\frac{1}{T}\sum_{s=1}^{T}\gamma_N(s,t)^2 \\ &\leq \bar{r}\frac{1}{T}\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(\sum_{s=1}^{T}M\left|\gamma_N(s,t)\right|) = O(\frac{1}{T}). \quad \text{(A-78)} \end{aligned}$$

By part (1) of Assumption 8**,**

$$\begin{aligned} II &\leq \frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2 \frac{1}{N}(\frac{1}{T}\sum_{s=1}^{T}\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}} \\ &\quad \sum_{t=k+1}^{l}\left|\frac{\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]}{\sqrt{N}}\right|^2) \\ &= \bar{r}\frac{1}{N}O_p(1). \quad \text{(A-79)} \end{aligned}$$

By part (2) of Assumption 8**,**

$$\begin{aligned} III &\leq \frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2 \sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\frac{1}{T}\sum_{s=1}^{T}\left|\frac{1}{N}\sum_{i=1}^{N}g_s'\gamma_i e_{it}\right|^2 \\ &\leq \bar{r}(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)\frac{1}{N}\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{it}\right\|^2 \\ &= \bar{r}O_p(1)\frac{1}{N}O_p(1). \quad \text{(A-80)} \end{aligned}$$

By part (1) of Lemma 3 and part (ii) of Lemma 1 in Bai and Ng (2002),

$$\begin{aligned} IV &\leq \frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2 \sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l} \\ &\quad \|g_t\|^2\frac{1}{N}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2 \\ &= \bar{r}O_p(1)\frac{1}{N}O_p(1). \quad \text{(A-81)} \end{aligned}$$

Next consider part (2).

$$
\begin{aligned}
& \sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}(\tilde{g}_t - J'g_t)g_t'J\right\| \\
\le\ & \left\|V_{NT}^{-1}\right\|\|J\|\sup_{0\le k<l\le T}\left\|\frac{1}{T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\sum\nolimits_{s=1}^{T}\tilde{g}_s g_t'\gamma_N(s,t)\right\| \\
& + \left\|V_{NT}^{-1}\right\|\|J\|\sup_{0\le k<l\le T}\left\|\frac{1}{T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\sum\nolimits_{s=1}^{T}\tilde{g}_s g_t'\zeta_{st}\right\| \\
& + \left\|V_{NT}^{-1}\right\|\|J\|\sup_{0\le k<l\le T}\left\|\frac{1}{T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\sum\nolimits_{s=1}^{T}\tilde{g}_s g_t'\eta_{st}\right\| \\
& + \left\|V_{NT}^{-1}\right\|\|J\|\sup_{0\le k<l\le T}\left\|\frac{1}{T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\sum\nolimits_{s=1}^{T}\tilde{g}_s g_t'\xi_{st}\right\| \\
=\ & \left\|V_{NT}^{-1}\right\|\|J\|\,(V+VI+VII+VIII). \qquad \text{(A-82)}
\end{aligned}
$$

By Lemma 2, $\left\|V_{NT}^{-1}\right\| \to \|V^{-1}\|$ and $\|J\| \to \|J_0\|$, thus it suffices to consider V, VI, VII and $VIII$. By part (1) of Lemma 3 and Assumption 4,

$$
\begin{aligned}
V \ \le\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s\|^2)^{\frac{1}{2}}\sup_{0\le k<l\le T}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}g_t'\gamma_N(s,t)\right\|^2)^{\frac{1}{2}} \\
\le\ & \bar{r}(\sup_{0\le k<l\le T}\frac{\sum_{t=k+1}^{l}\|g_t\|^2}{\sqrt{T(l-k)}})^{\frac{1}{2}}(\frac{1}{T}\sup_{0\le k<l\le T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\sum\nolimits_{s=1}^{T}|\gamma_N(s,t)|^2)^{\frac{1}{2}} \\
=\ & O_p(1)O_p(\frac{1}{\sqrt{T}}). \qquad \text{(A-83)}
\end{aligned}
$$

By part (1) of Lemma 3 and part (1) of Assumption 8,

$$
\begin{aligned}
VI \ \le\ & (\frac{1}{T}\sum\nolimits_{s=1}^{T}\|\tilde{g}_s\|^2)^{\frac{1}{2}}\sup_{0\le k<l\le T}(\frac{1}{T}\sum\nolimits_{s=1}^{T}\left\|\frac{\sum_{t=k+1}^{l}g_t'}{\sqrt{T(l-k)}}\frac{\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]}{N}\right\|^2)^{\frac{1}{2}} \\
\le\ & \bar{r}\frac{1}{\sqrt{N}}(\sup_{0\le k<l\le T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\|g_t\|^2)^{\frac{1}{2}}(\frac{1}{T}\sum\nolimits_{s=1}^{T} \\
& \sup_{0\le k<l\le T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\left\|\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]\right\|^2)^{\frac{1}{2}} \\
=\ & \frac{1}{\sqrt{N}}O_p(1)O_p(1). \qquad \text{(A-84)}
\end{aligned}
$$

By part (1) of Lemma 3 and part (2) of Assumption 8,

$$\begin{aligned} VII &\leq (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2)^{\frac{1}{2}} \sup_{0\leq k<l\leq T}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{\sum_{t=k+1}^{l}(\frac{1}{N}\sum_{i=1}^{N}g_s'\gamma_i e_{it})g_t'}{\sqrt{T(l-k)}}\right\|^2)^{\frac{1}{2}} \\ &\leq \bar{r}(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}}(\sup_{0\leq k<l\leq T}\left\|\frac{\sum_{t=k+1}^{l}\sum_{i=1}^{N}\gamma_i e_{it}g_t'}{\sqrt{NT(l-k)}}\right\|^2)^{\frac{1}{2}} \\ &\leq \bar{r}(\frac{1}{T}\sum_{s=1}^{T}\|g_s\|^2)^{\frac{1}{2}}\frac{1}{\sqrt{N}}(\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l} \\ &\quad \|g_t\|^2)^{\frac{1}{2}}(\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{it}\right\|^2)^{\frac{1}{2}} \\ &= O_p(1)\frac{1}{\sqrt{N}}O_p(1)O_p(1) \end{aligned} \tag{A-85}$$

By part (1) of Lemma 3 and part (ii) of Lemma 1 in Bai and Ng (2002),

$$\begin{aligned} VIII &\leq (\frac{1}{T}\sum_{s=1}^{T}\|\tilde{g}_s\|^2)^{\frac{1}{2}} \sup_{0\leq k<l\leq T}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{\sum_{t=k+1}^{l}g_t'(\frac{1}{N}\sum_{i=1}^{N}g_t'\gamma_i e_{is})}{\sqrt{T(l-k)}}\right\|^2)^{\frac{1}{2}} \\ &\leq \bar{r}(\sup_{0\leq k<l\leq T}\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}\|g_t\|^2)\frac{1}{\sqrt{N}}(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right\|^2)^{\frac{1}{2}} \\ &= O_p(1)\frac{1}{\sqrt{N}}O_p(1). \end{aligned} \tag{A-86}$$

For the other parts, proof of parts (3) and (5) are similar to proof of part (1), proof of parts (4) and (6) are similar to proof of part (2). ■

Lemma 5 *Under Assumptions 1-8,*

1. $\sup_{0\leq k<l\leq T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}z_t\right\| = o_p(1)$,
2. $\sup_{k_{\iota-1,0}<l\leq k_{\iota 0}}\left\|\frac{1}{l-k_{\iota-1,0}}\sum_{t=k_{\iota-1,0}+1}^{l}z_t\right\| = o_p(1)$ *for each* ι,
3. $\sup_{k_{\iota-1,0}\leq k<k_{\iota 0}}\left\|\frac{1}{k_{\iota 0}-k}\sum_{t=k+1}^{k_{\iota 0}}z_t\right\| = o_p(1)$ *for each* ι.

Proof. Recall that $z_t = vech[(\tilde{g}_t - J'g_t)(\tilde{g}_t - J'g_t)'] + vech[(\tilde{g}_t - J'g_t)g_t'J] + vech[J'g_t(\tilde{g}_t - J'g_t)'] + vech[(J'-J_0')g_tg_t'(J-J_0)] + vech[(J'-J_0')g_tg_t'J_0] + vech[J_0'g_tg_t'(J-J_0)]$. From

Lemma 2 and part (1) of Lemma 3, we have

$$\begin{aligned} & \sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}(J'-J_0')g_tg_t'(J-J_0)\right\| \\ \le\ & \|J-J_0\|^2\sup_{0\le k<l\le T}\frac{1}{\sqrt{T(l-k)}}\sum\nolimits_{t=k+1}^{l}\|g_t\|^2=o_p(1)O_p(1)=o_p(1), \text{(A-87)} \end{aligned}$$

and similarly $\sup_{0\le k<l\le T}\left\|\frac{1}{\sqrt{T(l-k)}}\sum_{t=k+1}^{l}(J'-J_0')g_tg_t'J_0\right\|=o_p(1)$. These together with parts (1) and (2) of Lemma 4 proves part (1). Part (2) can be proved similarly using Lemma 2, part (2) of Lemma 3 and parts (3) and (4) of Lemma 4. Part (3) can be proved similarly using Lemma 2, part (3) of Lemma 3 and parts (5) and (6) of Lemma 4. ■

Lemma 6 *In case factor loadings are stable, under Assumptions 1-6,* $\|H-H_0\|=o_p(1)$ *and* $\|U_{NT}-U\|=o_p(1)$.

Proof. The proof is similar to Lemma 2. ■

Lemma 7 *In case factor loadings are stable, under Assumptions 1-6 and 10-11,* $\sup_{T\epsilon\le k\le T(1-\epsilon)}\left\|\frac{1}{T}\sum_{t=1}^{k}(\tilde{f}_t\tilde{f}_t'-H'f_tf_t'H)\right\|=O_p(\frac{1}{\delta_{NT}^2})$.

Proof. It suffices to show $\sup_{T\epsilon\le k\le T(1-\epsilon)}\left\|\frac{1}{T}\sum_{t=1}^{k}(\tilde{f}_t-H'f_t)(\tilde{f}_t-H'f_t)'\right\|=O_p(\frac{1}{\delta_{NT}^2})$ and $\sup_{T\epsilon\le k\le T(1-\epsilon)}\left\|\frac{1}{T}\sum_{t=1}^{k}(\tilde{f}_t-H'f_t)f_t'H\right\|=O_p(\frac{1}{\delta_{NT}^2})$. The former is not larger than $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_t-H'f_t\right\|^2$, which is $O_p(\frac{1}{\delta_{NT}^2})$ by Lemma A.1 in Bai (2003). The latter is a refinement of part (2) of Lemma 4. For its proof, see Lemma 3 of Han and Inoue (2015), the required conditions (Assumptions 1-8(a) in Han and Inoue (2015)) can be verified. ■

Lemma 8 *In case factor loadings are stable, under Assumptions 1-6, 10 and 13,* $\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\|=o_p(1)$ *if* $\frac{d_T}{\delta_{NT}}\to 0$ *as* $(N,T)\to\infty$.

Proof. First note that $\left\|\tilde{\Omega}^{-1}-\Omega^{-1}\right\|\le\left\|\tilde{\Omega}^{-1}\right\|\left\|\tilde{\Omega}-\Omega\right\|\|\Omega^{-1}\|$, $\|\Omega^{-1}\|$ is constant, $\left\|\tilde{\Omega}^{-1}\right\|\le\sqrt{\frac{r(r+1)}{2}}\frac{1}{\rho_{\min}(\tilde{\Omega})}$ and $\left|\rho_{\min}(\tilde{\Omega})-\rho_{\min}(\Omega)\right|\le\left\|\tilde{\Omega}-\Omega\right\|$. Thus it remains to

show $\left\|\tilde{\Omega}-\Omega\right\| = o_p(1)$. By Assumption 13, $\left\|\tilde{\Omega}(FH_0)-\Omega\right\| = o_p(1)$. By second half of Theorem 2 in Han and Inoue (2015), $\left\|\tilde{\Omega}-\tilde{\Omega}(FH_0)\right\| = o_p(1)$ if $\frac{d_T}{\delta_{NT}} \to 0$. The required conditions in Han and Inoue (2015) can be verified. ■

Lemma 9 *In case factor loadings have structural changes, under Assumptions 1-5 and 14, with* $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$, *we have* $\|H_\iota - H_{\iota 0}\| = o_p(1)$ *and* $\|U_{\iota NT} - U_\iota\| = o_p(1)$.

Proof. First, Assumption 14 ensures uniqueness of $H_{\iota 0}$. The proof of $\|H_\iota - H_{\iota 0}\| = o_p(1)$ follows the same procedure as Proposition 1 in Bai (2003) which, as explained in Lemma 2, relies on Lemma 1(ii), Theorem 1 of Bai and Ng (2002) and Lemma A.3(i) of Bai (2003). Thus it suffices to reestablish these three with $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ and $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$. The first can be proved without adjustment. The second is proved in Theorem 5. The third ($\|U_{\iota NT} - U_\iota\| = o_p(1)$) is proved in Theorem 4. ■

Lemma 10 *Under Assumptions 1-5, 10-11 and 14, with* $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$, $\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - H'_\iota f_t f'_t H_\iota)}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\right\|$ *and* $\left\|\frac{\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}(\tilde{f}_{\iota t}\tilde{f}'_{\iota t} - H'_\iota f_t f'_t H_\iota)}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\right\|$ *are both* $O_p(\frac{1}{\delta_{NT}^2})$.

Proof. We will only show the first half, proof of the second half is the same. It suffices to prove $\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t} - H'_\iota f_t)(\tilde{f}_{\iota t} - H'_\iota f_t)'\right\| = O_p(\frac{1}{\delta_{NT}^2})$ and $\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\sum_{t=\tilde{k}_{\iota-1}+1}^{k}(\tilde{f}_{\iota t} - H'_\iota f_t) f'_t H_\iota\right\| = O_p(\frac{1}{\delta_{NT}^2})$ with $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ and $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$. The former is not larger than $\frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}}\sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota}\left\|\tilde{f}_{\iota t} - H'_\iota f_t\right\|^2$, which is $O_p(\frac{1}{\delta_{NT}^2})$ by Theorem 5 and $\|U_{\iota NT} - U_\iota\| = o_p(1)$ in Lemma 9. To prove the latter, it suffices to show $\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{\sum_{t=k_{\iota-1}+1}^{k}(\tilde{f}_{\iota t} - H'_\iota f_t) f'_t H_\iota}{k_\iota - k_{\iota-1}}\right\| = O_p(\frac{1}{\delta_{NT}^2})$ for each $k_{\iota-1} \in [k_{\iota-1,0} - C, k_{\iota-1,0} + C]$ and $k_\iota \in [k_{\iota,0} - C, k_{\iota,0} + C]$, where C is some positive integer (see Baltagi et al. (2017) for more details). For the case $k_{\iota-1} \in [k_{\iota-1,0}, k_{\iota-1,0} + C]$ and $k_\iota \in [k_{\iota,0} - C, k_{\iota,0}]$, Lemma 3 of Han and Inoue (2015) is applicable with T replaced by $k_\iota - k_{\iota-1}$. We next prove for the case $k_{\iota-1} \in [k_{\iota-1,0} - C, k_{\iota-1,0} - 1]$ and $k_\iota \in [k_{\iota,0} + 1, k_{\iota,0} + C]$. Proof of the other two cases are the same.

Note that in this case $x_{it} = f_t'\lambda_{0,\iota-1,i} + e_{it}$ for $t \in [k_{\iota-1}+1, k_{\iota-1,0}]$, $x_{it} = f_t'\lambda_{0,\iota,i} + e_{it}$ for $t \in [k_{\iota-1,0}+1, k_{\iota,0}]$ and $x_{it} = f_t'\lambda_{0,\iota+1,i} + e_{it}$ for $t \in [k_{\iota,0}+1, k_\iota]$. Define $w_{it} = f_t'(\lambda_{0,\iota-1,i} - \lambda_{0,\iota,i})$ for $t \in [k_{\iota-1}+1, k_{\iota-1,0}]$, $w_{it} = 0$ for $t \in [k_{\iota-1,0}+1, k_{\iota,0}]$ and $w_{it} = f_t'(\lambda_{0,\iota+1,i} - \lambda_{0,\iota,i})$ for $t \in [k_{\iota,0}+1, k_\iota]$, it follows that $x_{it} = f_t'\lambda_{0,\iota,i} + e_{it} + w_{it}$ for $t \in [k_{\iota-1}+1, k_\iota]$. Define $X_\iota = (x_{k_{\iota-1}+1}, ..., x_{k_\iota})'$, $w_t = (w_{1t}, ..., w_{Nt})'$, $W_\iota = (w_{k_{\iota-1}+1}, ..., w_{k_\iota})'$, $E_\iota = (e_{k_{\iota-1}+1}, ..., e_{k_\iota})'$ and recall $F_\iota = (f_{k_{\iota-1}+1}, ..., f_{k_\iota})'$, it follows that $X_\iota = F_\iota\Lambda_{0\iota}' + E_\iota + W_\iota$. Using the same decomposition as equation A.1 in Bai (2003), we have

$$\begin{aligned}
\tilde{f}_{\iota t} - H_\iota' f_t &= U_{\iota NT}^{-1}\frac{1}{N(k_\iota - k_{\iota-1})}[\tilde{F}_\iota'F_\iota\Lambda_{0\iota}'e_t + \tilde{F}_\iota'E_\iota\Lambda_{0\iota}f_t + \tilde{F}_\iota'E_\iota e_t \\
&\quad + \tilde{F}_\iota'F_\iota\Lambda_{0\iota}'w_t + \tilde{F}_\iota'W_\iota\Lambda_{0\iota}f_t + \tilde{F}_\iota'W_\iota w_t + \tilde{F}_\iota'E_\iota w_t + \tilde{F}_\iota'W_\iota e_t] \\
&= U_{\iota NT}^{-1}(Q_{1,t}^\iota + Q_{2,t}^\iota + Q_{3,t}^\iota + Q_{4,t}^\iota + Q_{5,t}^\iota + Q_{6,t}^\iota + Q_{7,t}^\iota + Q_{8,t}^\iota) \quad \text{(A-88)}
\end{aligned}$$

By Lemma 9, $\left\|U_{\iota NT}^{-1}\right\|$ and $\|H_\iota\|$ are both $O_p(1)$, thus it suffices to show for $m = 1, ..., 8$, $\sup_{k\in\Lambda_{\iota,\eta}} \left\|\frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k} Q_{m,t}^\iota f_t'\right\| = O_p(\frac{1}{\delta_{NT}^2})$.

For $m = 1, 2, 3$, the proof is the same as Lemma 3 of Han and Inoue (2015) except that in current case we use $\frac{1}{k_\iota - k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k_\iota}\left\|\tilde{f}_{\iota t} - H_\iota' f_t\right\|^2 = O_p(\frac{1}{\delta_{NT}^2})$ and $\|H_\iota\| = O_p(1)$ for $k_{\iota-1} \in [k_{\iota-1,0} - C, k_{\iota-1,0} - 1]$ and $k_\iota \in [k_{\iota,0}+1, k_{\iota,0}+C]$. These two are proved as intermediate result in Theorem 5 and Lemma 9, respectively. For $m = 4$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k} Q_{4,t}^\iota f_t'}{k_\iota - k_{\iota-1}}\right\|$ is not larger than $\left\|\frac{\tilde{F}_\iota'F_\iota\Lambda_{0\iota}'}{N(k_\iota - k_{\iota-1})}\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|w_t f_t'\|}{k_\iota - k_{\iota-1}})$ and

$$\begin{aligned}
\left\|\frac{\tilde{F}_\iota'F_\iota\Lambda_{0\iota}'}{N(k_\iota - k_{\iota-1})}\right\| &\le (\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota}\left\|\tilde{f}_{\iota s}\right\|^2}{k_\iota - k_{\iota-1}})^{\frac{1}{2}}(\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota}\|f_s\|^2}{k_\iota - k_{\iota-1}})^{\frac{1}{2}}\frac{1}{\sqrt{N}}(\frac{\sum_{i=1}^{N}\|\lambda_{0,\iota,i}\|^2}{N})^{\frac{1}{2}} \\
&= O_p(\frac{1}{\sqrt{N}}), \quad \text{(A-89)} \\
\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|w_t f_t'\|}{k_\iota - k_{\iota-1}} &\le \frac{\sum_{t=k_{\iota-1}+1}^{k_{\iota-1,0}}\|f_t f_t'\|}{k_\iota - k_{\iota-1}}(\sum_{i=1}^{N}\|\lambda_{0,\iota-1,i} - \lambda_{0,\iota,i}\|^2)^{\frac{1}{2}} \\
&\quad + \frac{\sum_{t=k_{\iota 0}+1}^{k_\iota}\|f_t f_t'\|}{k_\iota - k_{\iota-1}}(\sum_{i=1}^{N}\|\lambda_{0,\iota+1,i} - \lambda_{0,\iota,i}\|^2)^{\frac{1}{2}} \\
&= O_p(\frac{\sqrt{N}}{T}). \quad \text{(A-90)}
\end{aligned}$$

For $m=5$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k}Q_{5,t}^{\iota}f_t'}{k_\iota-k_{\iota-1}}\right\|$ is not larger than $\left\|\frac{\tilde{F}_\iota'W_\iota\Lambda_{0\iota}}{N(k_\iota-k_{\iota-1})}\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|f_tf_t'\|}{k_\iota-k_{\iota-1}})$ and

$$\frac{1}{k_\iota-k_{\iota-1}}\sum\nolimits_{t=k_{\iota-1}+1}^{k_\iota}\|f_tf_t'\|=O_p(1), \tag{A-91}$$

$$\begin{aligned}
&\left\|\frac{\tilde{F}_\iota'W_\iota\Lambda_{0\iota}}{N(k_\iota-k_{\iota-1})}\right\| \\
\leq\ &\left\|\frac{1}{k_\iota-k_{\iota-1}}\tilde{F}_\iota'W_\iota\right\|\frac{1}{\sqrt{N}}(\frac{\sum_{i=1}^{N}\|\lambda_{0,\iota,i}\|^2}{N})^{\frac{1}{2}} \\
\leq\ &[\frac{\sum_{s=k_{\iota-1}+1}^{k_{\iota-1,0}}\left\|\tilde{f}_{\iota s}f_s'\right\|}{k_\iota-k_{\iota-1}}(\sum\nolimits_{i=1}^{N}\|\lambda_{0,\iota-1,i}-\lambda_{0,\iota,i}\|^2)^{\frac{1}{2}} \\
&+\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota}\left\|\tilde{f}_{\iota s}f_s'\right\|}{k_\iota-k_{\iota-1}}(\sum\nolimits_{i=1}^{N}\|\lambda_{0,\iota+1,i}-\lambda_{0,\iota,i}\|^2)^{\frac{1}{2}}]\frac{1}{\sqrt{N}}(\frac{\sum_{i=1}^{N}\|\lambda_{0,\iota,i}\|^2}{N})^{\frac{1}{2}} \\
=\ &O_p(\frac{1}{T}).
\end{aligned} \tag{A-92}$$

The last equality is due to $\left\|\tilde{f}_{\iota s}-H_\iota'f_s\right\|=o_p(1)$ for $k_{\iota-1}+1\leq s\leq k_\iota$, which can be proved once Lemma A.2 in Bai (2003) is reestablished with $k_{\iota-1}\in[k_{\iota-1,0}-C,k_{\iota-1,0}-1]$ and $k_\iota\in[k_{\iota,0}+1,k_{\iota,0}+C]$. This is not difficult since in Bai (2003) Lemma A.2 is based on Lemma A.1 and Proposition 1, and as explained in the cases $m=1,2,3$, we have reestablished these two with $k_{\iota-1}\in[k_{\iota-1,0}-C,k_{\iota-1,0}-1]$ and $k_\iota\in[k_{\iota,0}+1,k_{\iota,0}+C]$. For $m=6$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k}Q_{6,t}^{\iota}f_t'\right\|$ is not larger than $\frac{1}{N}\left\|\frac{\tilde{F}_\iota'W_\iota}{k_\iota-k_{\iota-1}}\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|w_tf_t'\|}{k_\iota-k_{\iota-1}})$. The second and the third terms are both $O_p(\frac{\sqrt{N}}{T})$, as proved in $m=5$ and $m=4$ respectively. For $m=7$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{\sum_{t=k_{\iota-1}+1}^{k}Q_{7,t}^{\iota}f_t'}{k_\iota-k_{\iota-1}}\right\|$ is not larger than $\left\|\frac{1}{N(k_\iota-k_{\iota-1})}\tilde{F}_\iota'E_\iota\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|w_tf_t'\|}{k_\iota-k_{\iota-1}})$. The second term is $O_p(\frac{\sqrt{N}}{T})$, as proved in $m=4$. The first term is not larger than $\frac{1}{\sqrt{N}}(\frac{\sum_{s=k_{\iota-1}+1}^{k_\iota}\left\|\tilde{f}_{\iota s}\right\|^2}{k_\iota-k_{\iota-1}})^{\frac{1}{2}}(\frac{\sum_{i=1}^{N}\sum_{s=k_{\iota-1}+1}^{k_\iota}e_{is}^2}{N(k_\iota-k_{\iota-1})})^{\frac{1}{2}}$, which is $O_p(\frac{1}{\sqrt{N}})$. For $m=8$, $\sup_{k\in\Lambda_{\iota,\eta}}\left\|\frac{1}{k_\iota-k_{\iota-1}}\sum_{t=k_{\iota-1}+1}^{k}Q_{8,t}^{\iota}f_t'\right\|$ is not larger than $\frac{1}{N}\left\|\frac{1}{k_\iota-k_{\iota-1}}\tilde{F}_\iota'W_\iota\right\|(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|e_tf_t'\|}{k_\iota-k_{\iota-1}})$. The second term is $O_p(\frac{\sqrt{N}}{T})$, as proved in $m=5$. The third term is not larger than $(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\|f_t\|^2}{k_\iota-k_{\iota-1}})^{\frac{1}{2}}(\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota}\sum_{i=1}^{N}e_{it}^2}{k_\iota-k_{\iota-1}})^{\frac{1}{2}}$, which is

$O_p(\sqrt{N})$. Thus $\sup_{k\in\Lambda_{\iota,\eta}} \left\| \frac{1}{k_\iota - k_{\iota-1}} \sum_{t=k_{\iota-1}+1}^{k} Q_{m,t}^{\iota} f_t' \right\| = O_p(\frac{1}{T})$ for $m = 4, ..., 8$. ■

Lemma 11 *Under Assumptions 1-5 and 10-14, if* $\left|\tilde{k}_\iota - k_{\iota 0}\right|$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right|$ *are* $O_p(1)$, $\left\|\tilde{\Omega}_\iota^{-1} - \Omega_\iota^{-1}\right\| = o_p(1)$ *if* $\frac{d_T}{T^{\frac{1}{4}}} \to 0$ *and* $\frac{d_T}{\sqrt{N}} \to 0$ *as* $(N, T) \to \infty$.

Proof. Similar to Lemma 8, it suffices to show $\left\|\tilde{\Omega}_\iota - \tilde{\Omega}(F_\iota H_{\iota 0})\right\| = o_p(1)$, given $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$, $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$, $\frac{d_T}{T^{\frac{1}{4}}} \to 0$ and $\frac{d_T}{\sqrt{N}} \to 0$. This can be proved following the same procedure as Theorem 2 in Han and Inoue (2015). Here we present the adjustment. First, the notation should be replaced correspondingly, for example, in Han and Inoue (2015) the sample is $t = 1, ..., T$ while here the sample is $t = k_{\iota-1} + 1, ..., k_\iota$. Next, in Han and Inoue (2015) proof of Theorem 2 relies on their Lemma 7 and Lemma 8, which further relies on their Lemma 5 and Lemma 6 respectively. Once their Lemma 5 and Lemma 6 are reestablished given $\left|\tilde{k}_\iota - k_{\iota 0}\right| = O_p(1)$ and $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right| = O_p(1)$, the proof of Lemma 7, Lemma 8 and Theorem 2 need no adjustment.

We first reestablish parts (i) and (iii) of their Lemma 5. With $\frac{d_T}{T^{\frac{1}{4}}} \to 0$ and $\frac{d_T}{\sqrt{N}} \to 0$, they are enough. From equation (A-88), we have

$$\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| \tilde{f}_{\iota t} - H_\iota' f_t \right\|^4}{k_\iota - k_{\iota-1}} \leq 8^3 \left\| U_{\iota NT}^{-1} \right\|^4 \left(\sum_{m=1}^{8} \frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| Q_{m,t}^{\iota} \right\|^4}{k_\iota - k_{\iota-1}} \right). \tag{A-93}$$

Lemma 5 in Han and Inoue (2015) shows that $\sum_{m=1}^{3} \frac{1}{k_\iota - k_{\iota-1}} \sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| Q_{m,t}^{\iota} \right\|^4 = O_p(\frac{1}{T}) + O_p(\frac{1}{N^2})$, the proof need no adjustment. For $m = 4, ..., 8$, it can be shown that $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| Q_{4,t}^{\iota} \right\|^4}{k_\iota - k_{\iota-1}} = O_p(\frac{1}{T})$, $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| Q_{5,t}^{\iota} \right\|^4}{k_\iota - k_{\iota-1}} = O_p(\frac{1}{T})$, $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| Q_{6,t}^{\iota} \right\|^4}{k_\iota - k_{\iota-1}} = O_p(\frac{1}{T^3})$, $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| Q_{7,t}^{\iota} \right\|^4}{k_\iota - k_{\iota-1}} = O_p(\frac{1}{T})$ and $\frac{\sum_{t=k_{\iota-1}+1}^{k_\iota} \left\| Q_{8,t}^{\iota} \right\|^4}{k_\iota - k_{\iota-1}} = O_p(\frac{1}{T^2})$. The proof of Lemma 6 need no adjustment, but note that it utilized $\frac{1}{T} F'(\hat{F} - FH) V_{NT} = O_p(\frac{1}{\delta_{NT}^2})$. Its counterpart in current case is $\left\| \frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}} \sum_{t=\tilde{k}_{\iota-1}+1}^{\tilde{k}_\iota} (\tilde{f}_{\iota t} - H_\iota' f_t) f_t' \right\| = O_p(\frac{1}{\delta_{NT}^2})$, which is implicitly proved in Lemma 10. ■

Lemma 12 *Under Assumptions 1-5 and 14, if* $\left|\tilde{k}_\iota - k_{\iota+1,0}\right|$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right|$ *are* $O_p(1)$, *we have* $\|J_\iota - J_{\iota 0}\| = o_p(1)$ *and* $\|V_{\iota NT} - V_\iota\| = o_p(1)$.

Proof. The proof is similar to Lemma 9. ■

Lemma 13 *Under Assumptions 1-5, 10-11 and 14, if* $\left|\tilde{k}_\iota - k_{\iota+1,0}\right|$ *and* $\left|\tilde{k}_{\iota-1} - k_{\iota-1,0}\right|$ *are* $O_p(1)$, $\frac{1}{\tilde{k}_\iota - \tilde{k}_{\iota-1}} \sum_{t=\tilde{k}_{\iota-1}+1}^{k_{\iota 0}} z_{\iota t} = o_p(1)$ *and* $\frac{1}{\tilde{k}_\iota - k_{\iota 0}} \sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota} z_{\iota t} = o_p(1)$.

Proof. We will show the second half, the first half can be proved similarly. It suffices to show $\left\| \frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota} vech(\tilde{g}_{\iota t}\tilde{g}'_{\iota t} - J'_\iota g_{\iota t} g'_{\iota t} J_\iota)}{\tilde{k}_\iota - k_{\iota 0}} \right\|$ and $\left\| \frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota} vech(J'_\iota g_{\iota t} g'_{\iota t} J_\iota - J'_{\iota 0} g_{\iota t} g'_{\iota t} J_{\iota 0})}{\tilde{k}_\iota - k_{\iota 0}} \right\|$ are both $o_p(1)$. The first term can be proved similarly as Lemma 10. The second term is not larger than $\left\| \frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota} g_{\iota t} g'_{\iota t}}{\tilde{k}_\iota - k_{\iota 0}} \right\| \|J_\iota - J_{\iota 0}\|^2 + 2 \left\| \frac{\sum_{t=k_{\iota 0}+1}^{\tilde{k}_\iota} g_{\iota t} g'_{\iota t}}{\tilde{k}_\iota - k_{\iota 0}} \right\| \|J_\iota - J_{\iota 0}\| \, \|J_{\iota 0}\|$, which is $o_p(1)$ by Lemma 12. ■