Detecting Structural Breaks in High-Dimensional Functional Time Series Factor Models

Caixia Xu[1], Huacheng Su[2,*], Xu Liu[1], Jinhong You[1]

[1]*Shanghai University of Finance and Economics*

and [2]*Ningbo University of Technology*

Supplementary Material

This is a supplement to the paper "Detecting Structural Breaks in High-Dimensional Functional Time Series Factor Models", in which it contains the numerical studies, additional results of the application, technical proofs of Theorems 1–8 and Lemmas 1–4 with their proofs.

S1 Simulation studies

In this section, we conduct simulation studies to evaluate the performance of the proposed method under finite sample.

S1.1 Data generation process

For each trajectory $X_{it}(s)$, 100 observation time points are uniformly generated within the interval $(0, 1)$ and 500 replicates are applied in each scenario unless otherwise stated. We generate data under the framework of high dimensional functional factor models with $r = 3$ and $r = 5$ common factors:

$$X_{it}(s) = \boldsymbol{\lambda}_{it}^T \boldsymbol{f}_t(s) + u_{it}(s), i = 1, \dots, N, t = 1, \dots, T,$$

where random error $\boldsymbol{u}_t(s) \sim N(0, I)$.

To construct $\boldsymbol{\lambda}_t$, we start by generating T samples of the N-dimensional vector $\boldsymbol{k}_t$ from a multivariate normal distribution $N(0, (\sigma_{ij})_{N\times N})$, where $\sigma_{ij} = a^{|i-j|}$ and $a \in (0, 1)$. We then define $\boldsymbol{K} = (\boldsymbol{k}_1, \dots, \boldsymbol{k}_T)^T$. Next, we perform eigen decomposition on the matrix $\boldsymbol{K}\boldsymbol{K}^T$ to obtain the eigenvalues and eigenvectors. The diagonal matrix $\boldsymbol{\lambda}_r$ consists of the first r largest eigenvalues, and $\boldsymbol{K}_r$ is a matrix of the corresponding eigenvectors in $\mathbb{R}^{T\times r}$. We calculate $\boldsymbol{\lambda}_T$ as the product $\boldsymbol{K}^T\boldsymbol{K}_r$, and then proceed to perform QR decomposition on $\boldsymbol{\lambda}_T$, resulting in $\boldsymbol{\lambda}_T = \boldsymbol{Q}_T\boldsymbol{R}_T$. Finally, to ensure the identification condition on $\boldsymbol{\lambda}$ is met, we set $\boldsymbol{\lambda} = \sqrt{N}\boldsymbol{Q}_T$. This construction process allows us to account for change points in the sequence by performing segment-wise operations on different intervals. In each time segment, the variance of the normal distribution from which $\boldsymbol{k}_t$ is drawn is adjusted, thereby modifying the parameter a and effectively changing the magnitude of $\boldsymbol{\lambda}$.

In order to account for the influence of inter-factor correlation on the model, we consider the following two scenarios when constructing $\boldsymbol{\xi}_t$ for

$\boldsymbol{f}_{tq}(s)$:

Scenario1: There is no correlation among the factors. We independently generate $\xi^*_{tk} \sim N(0, Krk^{-1})$, denote $\boldsymbol{\xi}^*_t = (\xi^*_{t,1}, \ldots, \xi^*_{t,Kr})^T$ and $\boldsymbol{\xi}^* = (\boldsymbol{\xi}^*_1, \ldots, \boldsymbol{\xi}^*_T)$. We then perform eigen decomposition of the matrix $\boldsymbol{\xi}^{*T}\boldsymbol{\xi}^*$ to obtain matrices Λ_{Kr} and M_{Kr}, which consist of the first Kr largest eigenvalues and their corresponding eigenvectors, respectively. As a result of this procedure, we obtain $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T)^T$.

Scenario2: There are correlations among the factors. Specifically, we generate $\boldsymbol{\xi}_{tk} = (\xi_{t1k}, \ldots, \xi_{trk})^T$ from a vector autoregressive model, given by $\boldsymbol{\xi}_{tk} = \boldsymbol{V}\boldsymbol{\xi}_{(t-1)k} + \boldsymbol{\epsilon}_t$, where $\boldsymbol{V}$ is a matrix with elements $0.46^{|q-q'|+1}$ for $1 \leq q, q' \leq r$. The innovation term $\boldsymbol{\epsilon}_{tk} = (\epsilon_{t1k}, \ldots, \epsilon_{trk})^T$ is composed of independent components drawn from a normal distribution $N(0, k^{-1.5})$.

Given $\boldsymbol{\xi}_t$, the factor process $\boldsymbol{f}_t(s) = (f_{t1}(s), \ldots, f_{tr}(s))^T$ is generated by $f_{tq}(s) = \sum_{k=1}^{K} \xi_{tqk}\phi_{qk}(s)$, where $\phi_{qk}(s) = \sqrt{2}\sin[\{2(q \neq 2) + (q = 2)\}k\pi s/10]$ if k is odd and $\phi_{qk}(s) = \sqrt{2}\cos[\{2(q \neq 2)k + (q = 2)(2k + 1)\}\pi s/10]$ if k is even. The setting is used to ensure the orthogonality of eigenfunctions.

We consider the following setups for the number of breaks.

DGP1 (Single structural break):

$$\boldsymbol{\lambda}_{it} = \begin{cases} \boldsymbol{\alpha}_{i1} & \text{for } t = 1, \ldots, t_1, \\ \boldsymbol{\alpha}_{i2} & \text{for } t = t_1 + 1, \ldots, T. \end{cases}$$

In this scenario, we set $t_1 = \lfloor T/2 \rfloor + 1$ for different sample cases (N, T).

DGP2 (Two structural breaks):

$$\boldsymbol{\lambda}_{it} = \begin{cases} \boldsymbol{\alpha}_{i1} & \text{for } t = 1, \ldots, t_1, \\ \boldsymbol{\alpha}_{i2} & \text{for } t = t_1 + 1, \ldots, t_2, \\ \boldsymbol{\alpha}_{i3} & \text{for } t = t_2 + 1, \ldots, T. \end{cases}$$

In this scenario, we set $t_1 = \lfloor T/3 \rfloor + 1, t_2 = \lfloor 2T/3 \rfloor + 1$ for different sample cases (N, T).

DGP3 (Three structural breaks):

$$\boldsymbol{\lambda}_{it} = \begin{cases} \boldsymbol{\alpha}_{i1} & \text{for } t = 1, \ldots, t_1, \\ \boldsymbol{\alpha}_{i2} & \text{for } t = t_1 + 1, \ldots, t_2, \\ \boldsymbol{\alpha}_{i3} & \text{for } t = t_2 + 1, \ldots, t_3, \\ \boldsymbol{\alpha}_{i4} & \text{for } t = t_3 + 1, \ldots, T. \end{cases}$$

In this scenario, we set $t_1 = \lfloor T/4 \rfloor + 1, t_2 = \lfloor 2T/4 \rfloor + 1, t_3 = \lfloor 3T/4 \rfloor + 1$ for different sample cases (N, T).

As the factor loadings are presumed to be nonrandom, we generate them once and maintain their fixed values throughout the 500 replications.

S1.2 Determination of the number of factors

To determine the number of factors, we employ the information criteria functions ER and GR in equation (2.3). The empirical probability of correct selection (PROB) and the average selected number of factors (AVE) determined by ER are presented in Table 1 and Table 3 respectively. The corresponding results of GR can be found in Table 2 and Table 4. It is noteworthy that we observe for both ER and GR estimators, PROB is consistently equal to one for all cases, indicating that the selected number of factors accurately reflects the true number of factors. Additionally, the AVE is either three or five, which aligns with the true number of factors. These results, detailed in Tables 1-4, provide evidence for the selection consistency of the information criteria discussed in Section 3.

Table 1: Empirical probability of correct selection (PROB) by ER

	scenario1						scenario2					
(N,T)	(100,200)		(100,400)		(200,400)		(100,200)		(100,400)		(200,400)	
r	3	5	3	5	3	5	3	5	3	5	3	5
DGP1	1	1	1	1	1	1	1	1	1	1	1	1
DGP2	1	1	1	1	1	1	1	1	1	1	1	1
DGP3	1	1	1	1	1	1	1	1	1	1	1	1

Table 2: Empirical probability of correct selection (PROB) by GR

	scenario1						scenario2					
(N,T)	(100,200)		(100,400)		(200,400)		(100,200)		(100,400)		(200,400)	
r	3	5	3	5	3	5	3	5	3	5	3	5
DGP1	1	1	1	1	1	1	1	1	1	1	1	1
DGP2	1	1	1	1	1	1	1	1	1	1	1	1
DGP3	1	1	1	1	1	1	1	1	1	1	1	1

Table 3: Average selected number of factors (AVE) by ER

	scenario1						scenario2					
(N,T)	(100,200)		(100,400)		(200,400)		(100,200)		(100,400)		(200,400)	
r	3	5	3	5	3	5	3	5	3	5	3	5
DGP1	3	5	3	5	3	5	3	5	3	5	3	5
DGP2	3	5	3	5	3	5	3	5	3	5	3	5
DGP3	3	5	3	5	3	5	3	5	3	5	3	5

Table 4: Average selected number of factors (AVE) by GR

	scenario1						scenario2					
(N,T)	(100,200)		(100,400)		(200,400)		(100,200)		(100,400)		(200,400)	
r	3	5	3	5	3	5	3	5	3	5	3	5
DGP1	3	5	3	5	3	5	3	5	3	5	3	5
DGP2	3	5	3	5	3	5	3	5	3	5	3	5
DGP3	3	5	3	5	3	5	3	5	3	5	3	5

S1.3 Estimation of the break points

Figure 1 and Figure 2 depict the distribution of d_t after 500 repeated experiments for scenario 1 and scenario 2, respectively, with the same sample size. In the absence of change points, the values of d_t should be close to zero. It can be observed that for both scenario 1 and scenario 2, regardless of whether the number of change points m is 1, 2, or 3, $\hat{d}_t$ effectively represents the positions of the change points. However, due to the interference of correlation, scenario 2 exhibits higher levels of noise compared to scenario 1.

Then we use WBS Algorithm to estimate the number and positions of breaks points. During our simulation, we set thresholds of the form $\Delta_T = \{C\sigma\sqrt{T}\}$, $C = [0.2, 6]$, and $\alpha = 1.05$ for sSIC information criterion, where σ is the median absolute deviation estimator from the r package *wbs*. Table 5 presents the accuracy of estimating the number of change points, and we can observe that the method proposed in this paper exhibits high accuracy. Additionally, for scenario 1, as the sample size increases, the accuracy improves accordingly. The increase in the number of factors also has a positive impact on the estimation accuracy. In scenario 2, we observe that increasing the sample size in the temporal dimension, T, can introduce interference in accurately identifying the number of change points. On the other hand, increasing the sample size in the factor dimension, N, has a positive impact on the accuracy of identification.

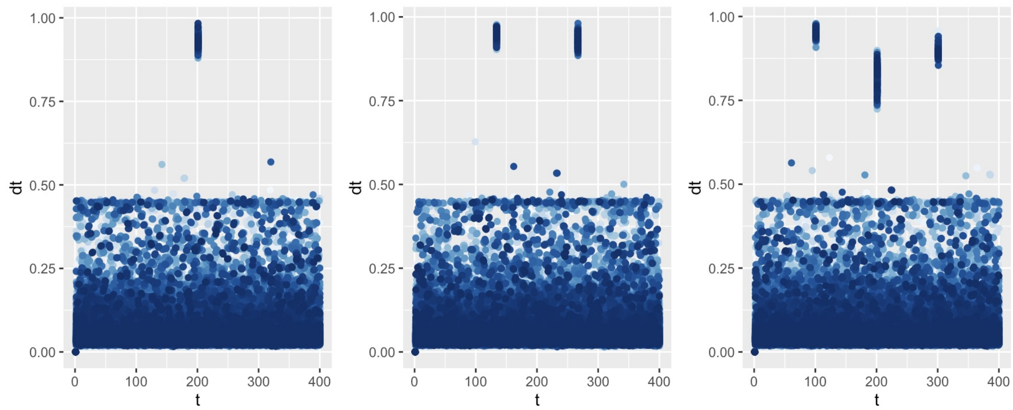

Figure 1: $\hat{d}_t$ when $N = 100, T = 400, r = 5$ for scenario1, from left to right, the samples represent DGP1, DGP2, and DGP3.

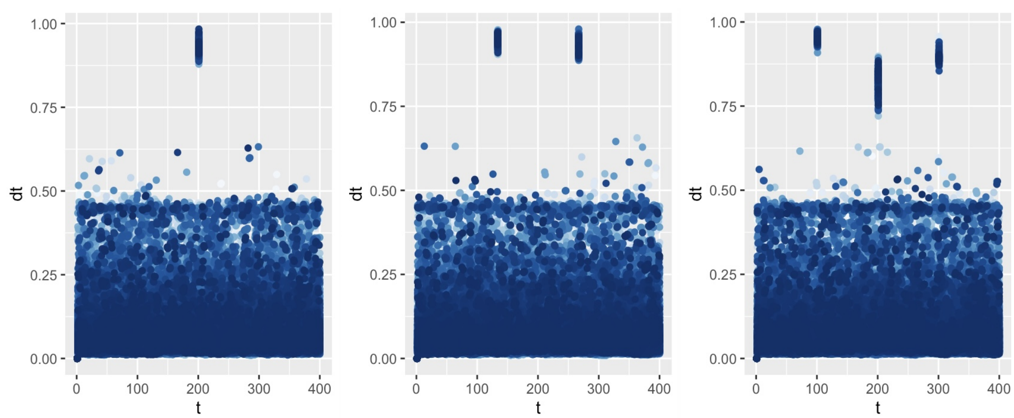

Figure 2: $\hat{d}_t$ when $N = 100, T = 400, r = 5$ for scenario2, from left to right, the samples represent DGP1, DGP2, and DGP3.

Table 6 shows the accuracy of break-point estimation, which is measured by normalised mutual information (NMI) measurements for cluster membership estimation defined by :

$$\text{NMI}(\hat{e}., e.) = 2\frac{MI(\hat{e}., e.)}{H(\hat{e}.) + H(e.)'}, \tag{S1.1}$$

where $\hat{e}. = \{I_1, \ldots, I_{\hat{m}+1}\}$ is the estimate of $e. = \{I_1, \ldots, I_{m_0+1}\}$, $H(e.)$ denotes the entropy of $e.$, $MI(\hat{e}., e.)$ is the mutual information between $\hat{e}.$ and $e.$ defined by

$$MI(\hat{e}., e.) = \sum_{k=1}^{\hat{m}+1}\sum_{j=1}^{m_0+1}(\frac{|\hat{I}_k \cap I_j|}{T})\log_2(\frac{T|\hat{I}_k \cap I_j|}{|\hat{I}_k||I_j|})$$

Table 5: Percentage of correct detection of the number of breaks

		scenario1			scenario2		
	(N,T)	(100,200)	(100,400)	(200,400)	(100,200)	(100,400)	(200,400)
DGP1	$r=3$	0.938	0.992	0.996	0.956	0.908	0.968
	$r=5$	0.980	0.996	0.998	0.904	0.964	0.972
DGP2	$r=3$	0.942	0.988	0.992	0.974	0.926	0.968
	$r=5$	0.988	0.990	0.994	0.900	0.976	0.972
DGP3	$r=3$	0.942	0.972	0.986	0.936	0.834	0.924
	$r=5$	0.976	0.994	0.996	0.884	0.964	0.972

Table 7 shows average Hausdorff distance of the estimated and true break points divided by $100 \times HD/T$. Let $D(A,B) \equiv \text{sup}_{b\in B}\text{inf}_{a\in A}|a-b|$ for any two sets A and B. The Hausdorff distance between A and B is defined as $\max\{\mathcal{D}(A,B), \mathcal{D}(B,A)\}$.

It is evident that the NMI values of the estimated time intervals before and after the change points in Table 6 are close to the ideal value of one and HD in Table 7 are close to the ideal value of zero. Furthermore, these

Table 6: NMI measurements for cluster membership estimation

		scenario1			scenario2		
	(N, T)	(100,200)	(100,400)	(200,400)	(100,200)	(100,400)	(200,400)
DGP1	$r = 3$	0.9868	0.9981	0.9984	0.9887	0.9837	0.9946
	$r = 5$	0.9898	0.9977	0.9993	0.9794	0.9929	0.9917
DGP2	$r = 3$	0.9908	0.9973	0.9971	0.9934	0.9936	0.9965
	$r = 5$	0.9952	0.9975	0.9977	0.9839	0.9948	0.9939
DGP3	$r = 3$	0.9915	0.9971	0.9972	0.9894	0.9885	0.9956
	$r = 5$	0.9945	0.9987	0.9989	0.9868	0.9962	0.9960

NMI and HD values demonstrate consistency with the accuracy of change point identification.

Table 7: The HD measurements for cluster membership estimation

		scenario1			scenario2		
	(N, T)	(100,200)	(100,400)	(200,400)	(100,200)	(100,400)	(200,400)
DGP1	$r = 3$	0.0430	0.0085	0.0165	0.0330	0.0140	0.0080
	$r = 5$	0.0560	0.0125	0.0090	0.0130	0.0015	0.0075
DGP2	$r = 3$	0.1020	0.0260	0.0220	0.0900	0.0185	0.0230
	$r = 5$	0.0870	0.0210	0.0110	0.0500	0.0070	0.0120
DGP3	$r = 3$	0.1430	0.0440	0.0335	0.1590	0.0265	0.0150
	$r = 5$	0.0990	0.0355	0.0210	0.0880	0.0255	0.0165

S1.4 Re-estimating the functional factor model

In Section 2.5 of the main paper, the determination of the value of K through principal components analysis, ultimately depends on the parameter K_{PCA}. Here, we provide detailed insights into the process of establishing

K_{PCA}, which is closely associated with a thorough sensitivity analysis. By varying the K_{PCA} values and evaluating the corresponding accuracy of estimating K, the numerical results are presented in Table 8. The data clearly indicate that 97.5% consistently delivers favorable results, with an average correct estimation rate exceeding 90% across all scenarios. This conclusion is further supported by Figure 3, illustrating that maintaining K_{PCA} within the range of 95% to 98% results in stable accuracy of estimation across all cases, with 97.5% emerging as one of the optimal choices.

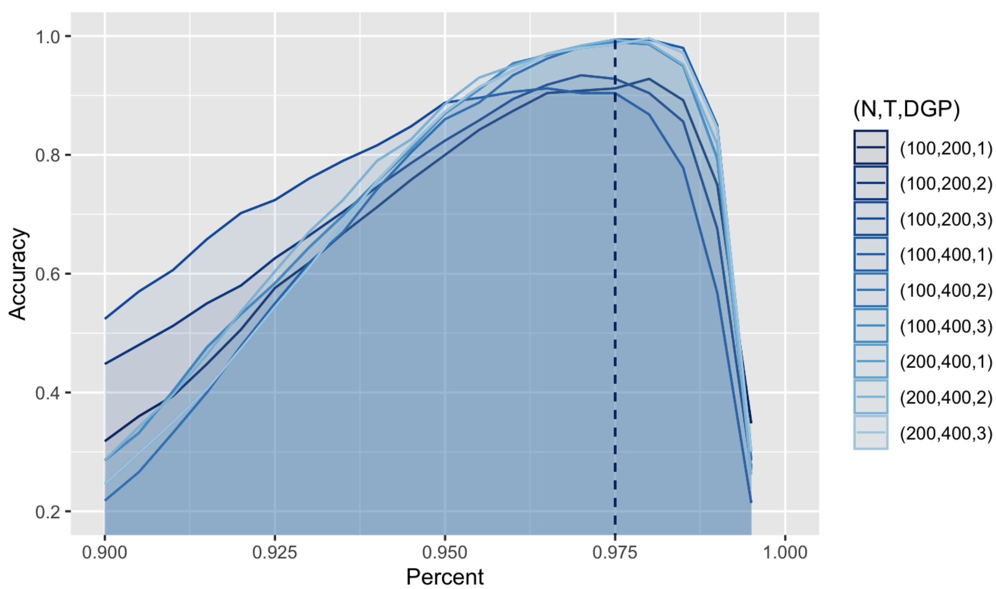

Figure 3: Accuracy in identifying the number of K Using various K_{PCA} values in Scenario 2, $r = 3$. (N, T, DGP) representing different sample cases.

In order to demonstrate the influence of the number of change points and the detection of change points on the normalized prediction error (PE),

Table 8: Accuracy in Determining the Number of K with Varied K_{PCA}

$K_{\text{PCA}}/(N,T)$	(100,200)			(100,400)			(200,400)		
$k=2, r=3$	DGP1	DGP2	DGP3	DGP1	DGP2	DGP3	DGP1	DGP2	DGP3
90	0.318	0.448	0.524	0.218	0.286	0.392	0.246	0.286	0.344
90.5	0.36	0.48	0.57	0.266	0.332	0.434	0.298	0.342	0.412
91	0.394	0.512	0.606	0.332	0.402	0.484	0.35	0.396	0.474
91.5	0.448	0.55	0.658	0.4	0.476	0.554	0.406	0.464	0.536
92	0.506	0.58	0.702	0.478	0.532	0.606	0.474	0.536	0.596
92.5	0.576	0.626	0.724	0.55	0.584	0.646	0.544	0.604	0.656
93	0.618	0.664	0.76	0.618	0.644	0.722	0.61	0.67	0.71
93.5	0.668	0.704	0.790	0.67	0.698	0.784	0.686	0.724	0.774
95	0.712	0.746	0.816	0.742	0.756	0.844	0.756	0.79	0.828
94.5	0.758	0.786	0.848	0.804	0.81	0.888	0.816	0.826	0.878
95	0.8	0.824	0.888	0.86	0.87	0.926	0.872	0.886	0.906
95.5	0.842	0.858	0.896	0.888	0.91	0.944	0.914	0.93	0.934
96	0.874	0.894	0.906	0.934	0.954	0.966	0.942	0.95	0.968
96.5	0.904	0.918	**0.912**	0.962	0.968	0.986	0.97	0.97	0.988
97	0.908	**0.934**	0.904	0.982	0.982	0.988	0.98	0.984	**0.998**
97.5	0.912	0.928	0.904	**0.994**	**0.99**	**0.992**	0.986	**0.994**	0.994
98	**0.928**	0.904	0.868	**0.994**	0.986	0.978	**0.996**	0.988	0.98
98.5	0.892	0.856	0.778	0.98	0.95	0.918	0.972	0.952	0.948
99	0.75	0.676	0.568	0.85	0.794	0.71	0.846	0.818	0.732
99.5	0.348	0.274	0.214	0.286	0.27	0.19	0.3	0.262	0.208

we compared our approach with Wen and Lin (2022), which corresponds to the scenario without any change points. The normalized prediction error (PE) is defined as follows:

$$PE = \sum_{i,t} n^{-1} \sum_{l=1}^{n} (\hat{X}_{it}(s_l) - X_{it}(s_l))^2 / \sum_{i,t} n^{-1} \sum_{l=1}^{n} X_{it}^2(s_l)$$

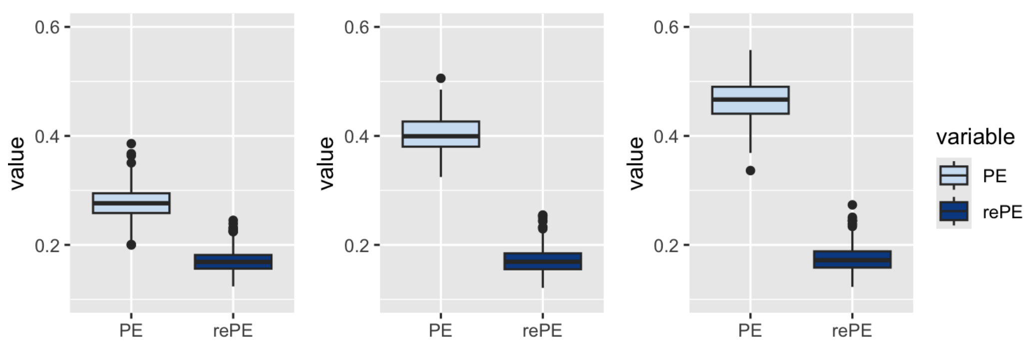

Figure 4: Normalized prediction error (PE) before and after the breaks when $N = 100, T = 200, r = 3$ for scenario2, from left to right, the samples represent DGP1, DGP2, and DGP3. PE represents the estimation error before identifying the change points, while rePE represents the estimation error after identifying the change points.

From Figure 4, We observe that before estimating the change points, the more change points actually present, the greater the interference introduced in the estimation, resulting in higher PE values. However, after estimating the change points, the influence of the change points on the estimation is mitigated, and the impact of the number of change points in the sequence on the estimation is eliminated.

S1.5 Comparisons between Change Points of Mean Functions and Loadings

In this section, we present comparisons between change points of mean functions and loadings. Our aim is to illustrate that when the factor structure exists and the change points of the mean function align with those of the loadings, detecting the change points of the loadings yields superior performance. This highlights how the loadings of factor models capture unique information compared to the original functional time series.

The dynamic nature of functional factors in factor models implies that the breakpoints in loadings may not align precisely with changes in mean functions. To begin, we contrast our approach with the method for identifying change points in mean functions based on Kovács et al. (2023) under scenario 1, where we assume a consistent distribution of factors, allowing us to consider the change-point locations in loading and mean function are identical. The subsequent tables present the outcomes of detecting change points in mean functions, offering a basis for comparison with the results of loading change-point detection outlined in subsection S1.3.

Upon comparing Table 9, 10, 11 with Table 5, 6, 7 respectively, the findings indicate that while the approach of identifying change points in mean functions shows some efficacy, it falls short compared to the method that incorporates the factor structure. This suggests that in the presence of a factor structure, relying solely on detecting change points in mean functions may not suffice.

Table 9: Percentage of correct detection of the number of breaks

	(N, J)	(100,200)	(100,400)	(200,400)
DGP1	$r = 3$	0.924	0.912	0.888
	$r = 5$	0.912	0.896	0.886
DGP2	$r = 3$	0.920	0.942	0.904
	$r = 5$	0.924	0.942	0.910
DGP3	$r = 3$	0.798	0.864	0.838
	$r = 5$	0.812	0.862	0.832

Table 10: NMI measurements for cluster membership estimation

	(N, J)	(100,200)	(100,400)	(200,400)
DGP1	$r = 3$	0.7504	0.7146	0.7341
	$r = 5$	0.7493	0.7193	0.7333
DGP2	$r = 3$	0.7332	0.7431	0.7296
	$r = 5$	0.7335	0.7433	0.7273
DGP3	$r = 3$	0.7333	0.7390	0.7460
	$r = 5$	0.7333	0.7369	0.7439

S2 Additional Results of Application

In this section we supplement some results for the application, which illustrate the presence of change points in the loading and the influence of climate factors.

Table 11: The HD measurements for cluster membership estimation

	(N, J)	(100,200)	(100,400)	(200,400)
DGP1	$r = 3$	0.0525	0.0590	0.0488
	$r = 5$	0.0517	0.0549	0.0548
DGP2	$r = 3$	0.06843	0.06560	0.0687
	$r = 5$	0.06844	0.06561	0.0692
DGP3	$r = 3$	0.0763	0.0768	0.0738
	$r = 5$	0.0764	0.0773	0.0745

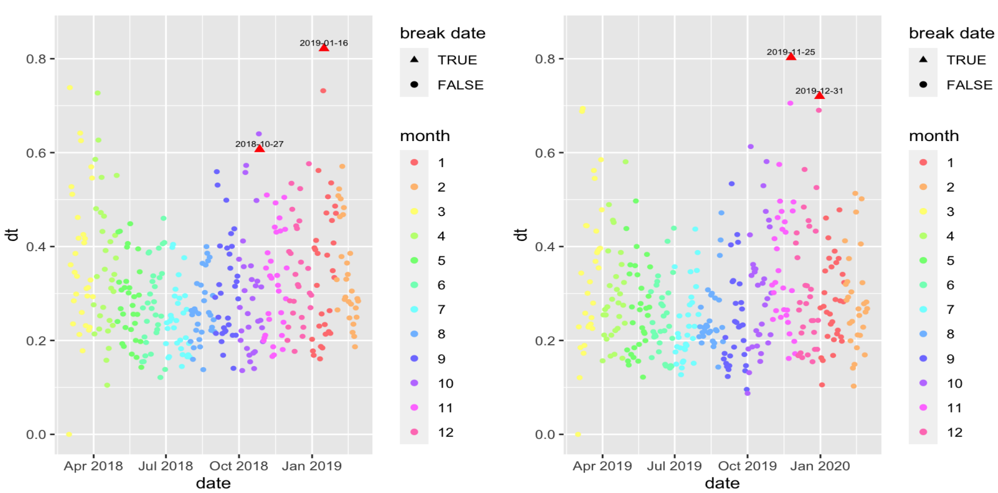

Figure 5: $\hat{d}_t$ for PM2.5 in 2018 and 2019 from left to right. The red triangles indicating the positions of change points, while the dots indicate the absence of detected change points.

Figure 5 presents the estimation of d_t and change points for PM2.5 in 2018 and 2019. The overall trend of d_t reveals the influence of climate and seasons on pollutant levels. In both 2018 and 2019, the d_t values for the months of June, July, and August were relatively small, indicating a relatively stable level of pollutant emissions during this period. However, in the months of March, April, September, and October, which fall between the seasons of summer and winter, the values of d_t were larger, indicating greater fluctuations in pollutant levels. According to Liang et al. (2016), Liang et al. (2015), during winter, the climatic conditions such as temperature, wind speed, and wind direction, are unfavorable for the dispersion of pollutants. Additionally, the use of heating systems during winters in northern regions leads to higher emissions of pollutants, resulting in significantly higher pollution levels compared to summer. Therefore, our estimation of the overall trend in $\hat{d}_t$ aligns with the actual variations in pollution levels. The detected change points illustrate the influence of human activities, as demonstrated in our manuscript.

S3 Proofs

In this section, we present the proofs of technical lemmas and theorems. The technical lemmas are used to prove the theorems of the paper. We first present some notations that will be used in the proofs of lemmas and theorems.

To fix notation, let $\|A\|_F = [tr(AA^T)]^{1/2}$ be the Frobenius norm of any matrix A, $\|A\|_2 = \sqrt{\psi_{\max}(A^TA)}$ as its spectral norm, $\|A\|_1$ be the 1-norm of any matrix A, $\|A\|_\infty$ be the sup-norm. Note that the Frobenius norm and spectral norm are equal when A is a vector and we always have $\|A\|_2 \leq \|A\|_F \leq \|A\|_2\sqrt{r(A)}$, $\|A\|_\infty = \|A^T\|_1$ and $\|A\|_2 \leq \sqrt{\|A\|_\infty \cdot \|A\|_1}$.

Proof of Theorem 1: Consider the inequalities

$$c/(1+c) < ln(1+c) < c \tag{A.1}$$

for $c > 0$. These inequalities lead to that

$$\frac{ln(1+\tilde{\psi}_{r'}/V(r'))}{ln(1+\tilde{\psi}_{r'+1}/V(r'+1))} < \frac{\tilde{\psi}_{r'}/V(r')}{(\tilde{\psi}_{r'+1}/V(r'+1))/(1+\tilde{\psi}_{r'+1}/V(r'+1))} = \frac{\tilde{\psi}_{r'}}{\tilde{\psi}_{r'+1}} = O_p(1)$$

for $r' = 1, \ldots, r-1, r+1, \ldots, r_{\max}$. Use equation A.1 again we have

$$\frac{ln(1+\tilde{\psi}_r/V(r))}{ln(1+\tilde{\psi}_{r+1}/V(r+1))} > \frac{(\tilde{\psi}_r/V(r))/(1+\tilde{\psi}_r/V(r))}{\tilde{\psi}_{r+1}/V(r+1)} = \frac{\tilde{\psi}_r V(r+1)}{\tilde{\psi}_{r+1}V(r-1)}$$

by Lemma 12 of Ahn and Horenstein (2013), under Assumptions 1-4, we have $V(r+1) = O_p(1)$, and then

$$\frac{V(r+1)}{V(r-1)} = \frac{V(r+1)}{\tilde{\psi}_r + \tilde{\psi}_{r+1} + V(r+1)} = O_p(1),$$

and

$$\frac{\tilde{\psi}_r V(r+1)}{\tilde{\psi}_{r+1}V(r-1)} = O_p(\eta_{N,n}^2)O_p(1) = O_p(\eta_{N,n}^2) = O_p(\min\{N,n\}).$$

These results indicate that the GR estimator is consistent.

Lemma C.1. *Suppose that Assumptions 1-4 hold. Then for all $t \in I_k$, as $(N, n) \to \infty$, it holds that*

$$\frac{1}{N}\|\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t\|_F^2 = O_p(\eta_{N,n}^{-2})$$

for $t = 1, \ldots, T$, where $\eta_{Nn} = min\{\sqrt{N}, \sqrt{n}\}$, $\boldsymbol{H}_t = (\sum_{l=1}^n \boldsymbol{f}_t(s_l)\boldsymbol{f}_t^T(s_l)/n)$ $(\boldsymbol{\alpha}_k^{0T}\hat{\boldsymbol{\lambda}}_t/N)\boldsymbol{V}_t^{-1}$ and $\boldsymbol{V}_t$ denote the $r \times r$ diagonal matrix of the first r largest eigenvalues of matrix $(Nn)^{-1}\sum_{l=1}^n \boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l)$ in descending order.

The convergence rate described in Lemma C.1 bears resemblance to those observed in the functional factor model Wen and Lin (2022).

Proof of Lemma C.1: By the definition of $\boldsymbol{V}_t$, we have $(Nn)^{-1}\boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l)\hat{\boldsymbol{\lambda}}_t =$

$\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t$. Then, we can decompose $\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t$ as follows:

$$
\begin{aligned}
&\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t \\
=&\frac{1}{Nn}\sum_{l=1}^{n} \boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l)\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t \\
=&\frac{1}{Nn}\sum_{l=1}^{n} (\boldsymbol{\alpha}_k^0 \boldsymbol{f}_t(s_l) + \boldsymbol{u}_t(s_l))(\boldsymbol{\alpha}_k^0 \boldsymbol{f}_t(s_l) + \boldsymbol{u}_t(s_l))^T \hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t \\
=&\frac{1}{Nn}\sum_{l=1}^{n} \boldsymbol{\alpha}_k^0 \boldsymbol{f}_t(s_l)\boldsymbol{f}_t^T(s_l)\boldsymbol{\alpha}_k^{0T}\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t + \frac{1}{Nn}\sum_{l=1}^{n} \boldsymbol{\alpha}_k^0 \boldsymbol{f}_t(s_l)\boldsymbol{u}_t^T(s_l)\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} \\
+&\frac{1}{Nn}\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l)\boldsymbol{f}_t^T(s_l)\boldsymbol{\alpha}_k^{0T}\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} + \frac{1}{Nn}\sum_{l=1}^{n} \boldsymbol{u}_t(s_l))\boldsymbol{u}_t^T(s_l))\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} \\
=&\frac{1}{Nn}\sum_{l=1}^{n} E[\boldsymbol{u}_t(s_l)\boldsymbol{u}_t^T(s_l)]\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} + \frac{1}{Nn}\sum_{l=1}^{n}[\boldsymbol{u}_t(s_l))\boldsymbol{u}_t^T(s_l)) - E(\boldsymbol{u}_t(s_l)\boldsymbol{u}_t^T(s_l))]\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} \\
+&\frac{1}{Nn}\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l)\boldsymbol{f}_t^T(s_l)\boldsymbol{\alpha}_k^{0T}\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1} + \frac{1}{Nn}\sum_{l=1}^{n} \boldsymbol{u}_t(s_l))\boldsymbol{u}_t^T(s_l))\hat{\boldsymbol{\lambda}}_t \boldsymbol{V}_t^{-1}
\end{aligned}
$$

or in vector form:

$$
\begin{aligned}
&\hat{\boldsymbol{\lambda}}_{it} - \boldsymbol{H}_t^T \boldsymbol{\alpha}_{i,k}^0 \\
=&V_t^{-1}[\frac{1}{Nn}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\sum_{l=1}^{n} E[u_{it}(s_l)u_{jt}(s_l)] + \frac{1}{Nn}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\sum_{l=1}^{n}(u_{it}(s_l)u_{jt}(s_l) - E[u_{it}(s_l)u_{jt}(s_l)]) \\
+&\frac{1}{Nn}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\boldsymbol{\alpha}_{j,k}^{0T}\sum_{l=1}^{n}\boldsymbol{f}_t^T(s_l)u_{it}(s_l) + \frac{1}{Nn}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\boldsymbol{\alpha}_{i,k}^{0T}\sum_{l=1}^{n}\boldsymbol{f}_t^T(s_l)u_{jt}(s_l)] \\
=&A_{1,ti} + A_{2,ti} + A_{3,ti} + A_{4,ti}. \qquad (*)
\end{aligned}
$$

Then by the Cauchy-Schwarz inequality, we have

$$
\frac{1}{N}\|\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t\|^2 = \frac{1}{N}\sum_{i=1}^{N}\|\hat{\boldsymbol{\lambda}}_{it} - \boldsymbol{H}_t^T\boldsymbol{\alpha}_{ik}^0\|^2 \leq \sum_{k=1}^{4}\|A_{k,ti}\|^2.
$$

Following Bai and Ng (2002), we can readily show that $N^{-1}\sum_{i=1}^{N}\|A_{1,ti}\|^2 = O_p(N^{-1})$ and $N^{-1}\sum_{i=1}^{N}\|A_{k,ti}\|^2 = O_p(n^{-1})$ for $k = 2,3,4$. Consequently, $N^{-1}\|\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0\boldsymbol{H}_t\|^2 = O_p(\eta_{N,n}^{-2})$.

Proof of Theorem 2: Noting that

$$
\begin{aligned}
&\frac{1}{N}\|\hat{\boldsymbol{\lambda}}_t - \hat{\boldsymbol{\lambda}}_{t-1}\|_F^2 \\
\leq&\frac{1}{N}\|\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0\boldsymbol{H}_t\|_F^2 + \frac{1}{N}\|\hat{\boldsymbol{\lambda}}_{t-1} - \boldsymbol{\alpha}_k^0\boldsymbol{H}_{t-1}\|_F^2 + \frac{1}{N}\|\hat{\boldsymbol{\alpha}}_k^0\boldsymbol{H}_t - \boldsymbol{\alpha}_k^0\boldsymbol{H}_{t-1}\|_F^2 \\
:=&A_{t1} + A_{t2} + A_{t3}
\end{aligned}
$$

by Lemma C.1, $A_{t1} = O_p(\eta_{N,n}^{-2})$, $A_{t2} = O_p(\eta_{N,n}^{-2})$. Since $N^{-1}\hat{\boldsymbol{\lambda}}_t^T\boldsymbol{\alpha}_k^0 = N^{-1}(\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0\boldsymbol{H}_t)^T\boldsymbol{\alpha}_k^0 + N^{-1}\boldsymbol{H}_t^T\hat{\boldsymbol{\alpha}}_k^{0T}\boldsymbol{\alpha}_k^0 = N^{-1}(\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0\boldsymbol{H}_t)^T\boldsymbol{\alpha}_k^0 + \boldsymbol{H}_t^T$ and $N^{-1}(\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0\boldsymbol{H}_t)^T\boldsymbol{\alpha}_k^0 = O_p(\eta_{N,n}^{-2})$ by Lemma B.2 in Bai (2003), we only need focus on the convergence properties of $\boldsymbol{H}_t$. Right multiply $\boldsymbol{H}_t$ to both sides of $N^{-1}\hat{\boldsymbol{\lambda}}_t^T\boldsymbol{\alpha}_k^0 = \boldsymbol{H}_t^T + O_p(\eta_{N,n}^{-2})$, we have

$$\frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T\boldsymbol{\alpha}_k^0\boldsymbol{H}_t = \boldsymbol{H}_t^T\boldsymbol{H}_t + O_p(\eta_{N,n}^{-2}).$$

Rewrite the left hand side of above as

$$
\begin{aligned}
\frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T\boldsymbol{\alpha}_k^0\boldsymbol{H}_t =&\frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T(\boldsymbol{\alpha}_k^0\boldsymbol{H}_t - \hat{\boldsymbol{\lambda}}_t + \hat{\boldsymbol{\lambda}}_t) \\
=&\frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T(\boldsymbol{\alpha}_k^0\boldsymbol{H}_t - \hat{\boldsymbol{\lambda}}_t) + \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T\hat{\boldsymbol{\lambda}}_t \\
=&\mathbb{I}_r + O_p(\eta_{N,n}^{-2}),
\end{aligned}
$$

according to $N^{-1}\hat{\boldsymbol{\lambda}}_t^T(\boldsymbol{\alpha}_k^0\boldsymbol{H}_t - \hat{\boldsymbol{\lambda}}_t) = O_p(\eta_{N,n}^{-2})$ and $N^{-1}\hat{\boldsymbol{\lambda}}_t^T\hat{\boldsymbol{\lambda}}_t = \mathbb{I}_r$, see Lemma

B.3 of Bai (2003). Equating the above two equations we obtain

$$\mathbb{I}_r = \boldsymbol{H}_t^T \boldsymbol{H}_t + O_p(\eta_{N,n}^{-2}).$$

By ignoring the term $O_p(\eta_{N,n}^{-2})$, the above shows that $\boldsymbol{H}_t$ is an orthogonal matrix so that its eigenvalues are either 1 or -1. We need to show that $\boldsymbol{H}_t$ is a diagonal matrix. From the definition of $\boldsymbol{H}_t$,

$$\begin{aligned}\boldsymbol{H}_t^T =&\boldsymbol{V}_t^{-1}(\frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T \boldsymbol{\alpha}_k^0)(\frac{1}{n}\sum_{l=1}^{n} \boldsymbol{f}_t(s_l)\boldsymbol{f}_t^T(s_l)) \\ =&\boldsymbol{V}_t^{-1}\boldsymbol{H}_t^T(\frac{1}{n}\sum_{l=1}^{n} \boldsymbol{f}_t(s_l)\boldsymbol{f}_t^T(s_l)) + O_p(\eta_{N,n}^{-2}),\end{aligned}$$

where we use the fact that $N^{-1}\hat{\boldsymbol{\lambda}}_t^T \boldsymbol{\alpha}_k^0 = \boldsymbol{H}_t^T + O_p(\eta_{N,n}^{-2})$ under $N^{-1}\hat{\boldsymbol{\lambda}}_t^T\hat{\boldsymbol{\lambda}}_t = \mathbb{I}_r$. Multiplying $\boldsymbol{V}_t$ on both sides and taking the transpose

$$\frac{1}{n}\sum_{l=1}^{n} \boldsymbol{f}_t(s_l)\boldsymbol{f}_t^T(s_l))\boldsymbol{H}_t = \boldsymbol{H}_t\boldsymbol{V}_t + O_p(\eta_{N,n}^{-2}).$$

This equation implies that $\boldsymbol{H}_t$ (up to a negligible term) is a matrix consisting of eigenvectors of $\frac{1}{n}\sum_{l=1}^{n} \boldsymbol{f}_t(s_l)\boldsymbol{f}_t^T(s_l))$. The latter matrix is diagonal and has distinct eigenvalues by assumption. Thus, each eigenvalue is associated with a unique unitary eigenvector (up to a sign change) and each eigenvector has a single non-zero element. This implies that $\boldsymbol{H}_t$ is a diagonal matrix up to an $O_p(\eta_{N,n}^{-2})$ order. It is already known that the eigenvalues of $\boldsymbol{H}_t$ are 1 or -1, $\boldsymbol{H}_t$ is a diagonal matrix with elements of 1 or -1 as its elements. Without loss of generality, we can assume all elements are 1.

This implies $\boldsymbol{H}_t = \mathbb{I}_r + O_p(\eta_{N,n}^{-2})$ and

$$\begin{aligned} A_{t3} =& \frac{1}{N}\|\hat{\boldsymbol{\alpha}}_k^0 \boldsymbol{H}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_{t-1}\|_F^2 \\ \leq& \frac{1}{N}\|\hat{\boldsymbol{\alpha}}_k^0\|_F^2 \|\boldsymbol{H}_t - \boldsymbol{H}_{t-1}\|_F^2 = O_p(\eta_{N,n}^{-2}). \end{aligned}$$

Combining results we obtain Theorem 1.

Lemma C.2. *Suppose that Assumptions 1-4 hold. Then for all* $t \in I_k$, *as* $(N, n) \to \infty$, *we have*

$$\hat{\boldsymbol{f}}_t(s) = \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T \boldsymbol{\alpha}_k^0 \boldsymbol{f}_t^0(s) + O_p(\frac{1}{\sqrt{N}} + \eta_{Nn}^{-1})$$

for $s = s_1, \ldots, s_n$.

Proof of Lemma C.2: we can decompose $\boldsymbol{f}_t(s)$ as follows:

$$\begin{aligned} \boldsymbol{f}_t(s) =& \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T \boldsymbol{X}_t(s) = \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T(\boldsymbol{\alpha}_k^0 \boldsymbol{f}_t^0(s) + \boldsymbol{u}_t(s)) \\ =& \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T \boldsymbol{\alpha}_k^0 \boldsymbol{f}_t^0(s) + \frac{1}{N}\boldsymbol{H}_t^T \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s) + \frac{1}{N}(\boldsymbol{\lambda}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t^T)\boldsymbol{u}_t(s) \\ :=& \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T \boldsymbol{\alpha}_k^0 \boldsymbol{f}_t^0(s) + D_{1,t} + D_{2,t}. \end{aligned}$$

First, $D_{1,t} = N^{-1}\boldsymbol{H}_t^T \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s) = O_p(N^{-1/2})$ for each t. Now, we use equation $(*)$ and make the following decomposition

$$\begin{aligned} D_{2,t} =& \frac{1}{N}(\boldsymbol{\lambda}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t^T)\boldsymbol{u}_t(s) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\lambda}}_{it} - \boldsymbol{H}_t^T \boldsymbol{\alpha}_{ik}^0)u_{it}(s) \\ =& \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{4} A_{kti}u_{it}(s) = \boldsymbol{V}_t^{-1}\sum_{k=1}^{4}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{V}_t A_{kti}u_{it}(s) \\ :=& \boldsymbol{V}_t^{-1}\sum_{k=1}^{4} D_{2tk}. \end{aligned}$$

We further decompose D_{2t1} as follows:

$$
\begin{aligned}
D_{2t1} =& \frac{1}{N}\boldsymbol{V}_t A_{1ti} u_{it}(s) = \frac{1}{N}\frac{1}{Nn}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\sum_{l=1}^{n} E[u_{it}(s_l)u_{jt}(s_l)]u_{it}(s) \\
=& \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\boldsymbol{\lambda}}_{jt} - \hat{\boldsymbol{H}}_t^T\boldsymbol{\alpha}_{j,k}^0)\sum_{l=1}^{n} E[u_{it}(s_l)u_{jt}(s_l)]u_{it}(s) \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\boldsymbol{H}}_t^T\boldsymbol{\alpha}_{j,k}^0\sum_{l=1}^{n} E[u_{it}(s_l)u_{jt}(s_l)]u_{it}(s) \\
:=& D_{2t1}^{(1)} + \hat{\boldsymbol{H}}_t^T D_{2t1}^{(2)}.
\end{aligned}
$$

By the repeated use of Cauchy-schwarz inequality, Lemma C.1, and Assumptions 4 (i) and (iii),

$$
\begin{aligned}
\|D_{2t1}^{(1)}\|_F \leq& \frac{1}{\sqrt{N}}[\frac{1}{N}\sum_{j=1}^{N}\|\hat{\boldsymbol{\lambda}}_{jt} - \hat{\boldsymbol{H}}_t^T\boldsymbol{\alpha}_{j,k}^0\|_F^2]^{1/2}[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{n}\sum_{l=1}^{n}\omega_{ij}^2(s_l)u_{it}^2(s)]^{1/2} \\
\leq& \frac{1}{\sqrt{N}}[\frac{1}{N}\sum_{j=1}^{N}\|\hat{\boldsymbol{\lambda}}_{jt} - \hat{\boldsymbol{H}}_t^T\boldsymbol{\alpha}_{j,k}^0\|_F^2]^{1/2}\max_i\sum_{j=1}^{N}\frac{1}{n}\sum_{l=1}^{n}\omega_{ij}^2(s_l)[\frac{1}{N}\sum_{i=1}^{N}u_{it}^2(s)]^{1/2} \\
=& N^{-1/2}O_p(\eta_{N,n}^{-1})O_p(1)O_p(1) = N^{-1/2}O_p(\eta_{N,n}^{-1}).
\end{aligned}
$$

Noting that

$$
\begin{aligned}
E\|D_{2t1}^{(2)}\| \leq& \frac{1}{N^2}E\|\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\alpha}_{j,k}^0\frac{1}{n}\sum_{l=1}^{n}E[u_{it}(s_l)u_{jt}(s_l)]u_{it}(s)\| \\
\leq& CN^{-1}\max_{i,t}E\|u_{it}(s)\|\cdot\frac{1}{n}\sum_{l=1}^{n}\max_j\sum_{i=1}^{N}\omega_{ij}(s_l) \\
=& O_p(\frac{1}{N}).
\end{aligned}
$$

By Markov inequality, $D_{2t1}^{(2)} = O_p(N^{-1})$. Then$D_{2t1} = N^{-1/2}O_p(\eta_{N,n}^{-1})$.

For D_{2t2}, we use $\bar{v}_{ij}(s_l) = u_{it}(s_l)u_{jt}(s_l) - E[u_{it}(s_l)u_{jt}(s_l)]$ and make the following decomposition:

$$\begin{aligned}
D_{2t2} &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{V}_t A_{2ti}u_{it}(s) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Nn}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\sum_{l=1}^{n}(u_{it}(s_l)u_{jt}(s_l) - E[u_{it}(s_l)u_{jt}(s_l)])u_{it}(s) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\frac{1}{n}\sum_{i=1}^{n}(u_{it}(s_l)u_{jt}(s_l) - E[u_{it}(s_l)u_{jt}(s_l)])u_{it}(s) \\
&= \frac{1}{N}\sum_{j=1}^{N}(\hat{\boldsymbol{\lambda}}_{jt} - \hat{\boldsymbol{H}}_t^T\boldsymbol{\alpha}_{j,k}^0)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{n}\sum_{l=1}^{n}\bar{v}_{ij}(s_l)u_{it}(s) \\
&+ \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{H}}_t^T\boldsymbol{\alpha}_{j,k}^0\frac{1}{N}\sum_{i=1}^{N}\frac{1}{n}\sum_{l=1}^{n}\bar{v}_{ij}(s_l)u_{it}(s) \\
&= D_{2t2}^{(1)} + D_{2t2}^{(2)}.
\end{aligned}$$

By Cauchy-schwarz inequality, $\|D_{2t2}^{(1)}\| \le (N^{-1}\sum_{i=1}^{N}\|\hat{\boldsymbol{\lambda}}_{jt} - \hat{\boldsymbol{H}}_t^T\boldsymbol{\alpha}_{j,k}^0\|^2)^{1/2}(\bar{D}_{2t2}^{(1)})^{1/2}$, where $\bar{D}_{2t2}^{(1)} = N^{-1}\sum_{j=1}^{N}(N^{-1}\sum_{i=1}^{N}n^{-1}\sum_{l=1}^{n}\bar{v}_{ij}(s_l)u_{it}(s))^2$. Noting that by Assumptions 4 (i) and (iv)

$$\begin{aligned}
E(\bar{D}_{2t2}^{(1)}) =& \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}E[\frac{1}{n}\sum_{l=1}^{n}\bar{v}_{ij}(s_l)\bar{v}_{ik}(s_l)u_{it}^2(s)] \\
&\le \frac{1}{n}\sum_{l=1}^{n}\max_{i,j}\|\bar{v}_{ij}(s_l)\|_4^2\max_{i,t}\|\bar{u}_{it}(s)\|_4^2 = O_p(\frac{1}{n}),
\end{aligned}$$

then we have $\|D_{2t2}^{(1)}\| = O_p(\eta_{N,n}^{-1})O_p(n^{-1})$ by Markov inequality and Lemma 1. By direct moment calculation and Chebyshev inequality, we can show that $D_{2t2}^{(2)} = O_p(N^{-1/2}\eta_{N,n}^{-1})$. It follows that $D_{2t2} = O_p(\eta_{N,n}^{-2})$.

Next

$$
\begin{aligned}
D_{2t3} &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{V}_t A_{3ti} u_{it}(s) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Nn}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\boldsymbol{\alpha}_{j,k}^{0T}\sum_{l=1}^{n}\boldsymbol{f}_t(s_l)u_{it}(s_l)u_{it}(s) \\
&= \frac{1}{N^2 n}\sum_{j=1}^{N}\hat{\boldsymbol{\lambda}}_{jt}\boldsymbol{\alpha}_{j,k}^{0T}\sum_{i=1}^{N}\sum_{l=1}^{n}\boldsymbol{f}_t(s_l)u_{it}(s_l)u_{it}(s) \\
&= \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T\boldsymbol{\alpha}_k^0\frac{1}{Nn}\sum_{l=1}^{n}\boldsymbol{f}_t(s_l)\boldsymbol{u}_t^T(s_l)\boldsymbol{u}_t(s) \\
&:= \frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T\boldsymbol{\alpha}_k^0 d_t.
\end{aligned}
$$

Note that

$$
\begin{aligned}
d_t =& \frac{1}{Nn}\sum_{l=1}^{n}\boldsymbol{f}_t(s_l)\boldsymbol{u}_t^T(s_l)\boldsymbol{u}_t(s) \\
=& \frac{1}{Nn}\sum_{l=1}^{n}E(\boldsymbol{f}_t(s_l)\boldsymbol{u}_t^T(s_l)\boldsymbol{u}_t(s)) + \frac{1}{Nn}\sum_{l=1}^{n}[\boldsymbol{f}_t(s_l)\boldsymbol{u}_t^T(s_l)\boldsymbol{u}_t(s) - E(\boldsymbol{f}_t(s_l)\boldsymbol{u}_t^T(s_l)\boldsymbol{u}_t(s))] \\
=& d_{t1} + d_{t2}.
\end{aligned}
$$

By Assumption 4 (ii), $|d_{t1}| \leq \frac{1}{n}\sum_{l'=1}^{n}|\gamma_{N,F}(l',l)| = O_p(n^{-1})$. By Assumption 4 (vi),

$$
E[|d_{t2}|^2] = \frac{1}{n^2}\sum_{l=1}^{n}\sum_{l'=1}^{n}E[\zeta_{F,ll'}\zeta_{F,ll''}] \leq \frac{1}{N}\max_{l,l'}\|\sqrt{N}\zeta_{F,ll'}\|_F^2 = O_p(\frac{1}{N}).
$$

It follows that $d_{t2} = O_p(N^{-1/2})$, $d_t = O_p(N^{-1/2}+n^{-1})$ and $D_{2t3} = O_p(N^{-1/2}+$

n^{-1}). Similarly, we can show that

$$
\begin{aligned}
D_{2t4} &= \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{V}_t A_{4ti} u_{it}(s) \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{Nn}\sum_{j=1}^{N} \hat{\boldsymbol{\lambda}}_{jt}\boldsymbol{\alpha}_{j,k}^{0T} \sum_{l=1}^{n} \boldsymbol{f}_t(s_l) u_{jt}(s_l) u_{it}(s) \\
&= \frac{1}{N^2 n}\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\boldsymbol{\lambda}}_{jt}\boldsymbol{\alpha}_{j,k}^{0T} \sum_{l=1}^{n} \boldsymbol{f}_t(s_l) u_{it}(s_l) u_{it}(s) \\
&= \frac{1}{N^2 n}\hat{\boldsymbol{\lambda}}_t^T \sum_{l=1}^{n} \boldsymbol{u}_t(s_l)\boldsymbol{f}_t^T(s_l)\boldsymbol{\alpha}_k^{0T}\boldsymbol{u}_t(s) \\
&= \frac{1}{N^2 n}(\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0 \boldsymbol{H}_t)^T \sum_{l=1}^{n} \boldsymbol{u}_t(s_l)\boldsymbol{f}_t^T(s_l)\boldsymbol{\alpha}_k^{0T}\boldsymbol{u}_t(s) \\
&+ \boldsymbol{H}_t^T \frac{1}{N^2 n}\boldsymbol{\alpha}_k^{0T} \sum_{l=1}^{n} \boldsymbol{u}_t(s_l)\boldsymbol{f}_t^T(s_l)\boldsymbol{\alpha}_k^{0T}\boldsymbol{u}_t(s) \\
&= O_p(\eta_{N,n}^{-1})O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{N}).
\end{aligned}
$$

It follows that $D_{2,t} = O_p(N^{-1/2} + \eta_{N,n}^{-2})$. Thus Lemma 2 follows.

Proof of Theorem 3: Noting that

$$
\begin{aligned}
&\|\hat{\boldsymbol{f}}_t(s) - \boldsymbol{H}_t^{-1}\boldsymbol{f}_t^0(s)\|_F^2 \\
=&\|\frac{1}{N}\hat{\boldsymbol{\lambda}}_t^T\boldsymbol{\alpha}_k^0\boldsymbol{f}_t^0(s) - \boldsymbol{H}_t^{-1}\boldsymbol{f}_t^0(s) + O_p(\frac{1}{\sqrt{N}} + \eta_{N,n}^{-2})\|_F^2 \\
\le& C\cdot\|\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\alpha}_k^0\boldsymbol{H}_t\|_F^2\|\boldsymbol{f}_t^0(s)\|_F^2 + O_p(\frac{1}{N} + \eta_{N,n}^{-4}) \\
\le& C\cdot tr(\boldsymbol{f}_t^0(s)\boldsymbol{f}_t^{0T}(s))O_p(\eta_{N,n}^{-2}) + O_p(\frac{1}{N} + \eta_{N,n}^{-4}) \\
=& C\cdot O_p(\eta_{N,n}^{-2}) + O_p(\frac{1}{N} + \eta_{N,n}^{-4}) \\
=& O_p(\eta_{N,n}^{-2})
\end{aligned}
$$

by Lemma C.2. Thus theorem 3 follows.

Proof of Theorem 4: We first state the following three Lemmas that are used in proving theorem 4.

Lemma C.3. *Suppose that Assumption 5 hold. There exist two constants C_1, C_2 such that if $C_1 \log^{1/2} T \leq \Delta_T \leq C_2 \delta_T^{1/2} \tilde{\mu}_{T,N}$, then for certain positive C_3, C_4, it holds that*

$$P(\mathcal{A}_T) \geq 1 - C_3/T - T\delta_T^{-1}(1 - \delta_T^2 T^{-2}/9)^M,$$

where $\mathcal{A}_T = \{\hat{m} = m; \max_{i=1,\dots,m} |\hat{t}'_i - t'_i| \leq C_4 \log T(\tilde{\mu}_T^{-2})\}$.

The aforementioned Lemma C.3 is derived from Theorem 2 in Fryzlewicz (2014), which demonstrates that when the minimum distance between change points is characterized and regulated, the WBS algorithm can ensure, with probability approaching one, the accuracy of both the number and the location of the change points. Building upon this Lemma, to prove Theorem 4, it suffices to demonstrate that the sSIC criterion is minimized when the number of change points is m.

Lemma C.4. *Use $\hat{\boldsymbol{X}}_t^K(s_l)$, $\hat{\boldsymbol{\alpha}}_t^K$ and $\hat{\boldsymbol{f}}_t^K(s_l)$ to denote the estimate of $\boldsymbol{X}_t(s_l)$, $\boldsymbol{\alpha}_t^0$ and $\boldsymbol{f}_t^0(s_l)$ when the number of change points m is estimated by K, where*

$$\hat{\boldsymbol{\alpha}}_t^K = N^{1/2} E_{eigen}(\frac{1}{nT_k} \sum_{t \in I_k} \sum_{l=1}^{n} \boldsymbol{X}_t(s_l) \boldsymbol{X}_t^T(s_l); r),$$

$$\hat{\boldsymbol{f}}_t^K(s_l) = \hat{\boldsymbol{\alpha}}_t^{KT}\boldsymbol{X}_t(s_l),$$

$$\hat{\sigma}_K^2 = \frac{1}{Tn}\sum_{t=1}^{T}\sum_{l=1}^{n}\|\hat{\boldsymbol{X}}_t^K(s_l) - \boldsymbol{X}_t(s_l)\|_F^2$$

for $t \in I_k$ and $|I_k| = \hat{t}'_{k+1} - \hat{t}'_k$. Then there are

(i) when $K \geq m$ and all the conditions in Lemma C.3 are satisfied, we can conclude that $(\hat{\sigma}_m^2 - \hat{\sigma}_K^2)/\hat{\sigma}_m^2 \leq M_1$ where $M_1 = O_p(\eta_{N\delta_T}^{-2})$.

(ii) When $K < m$, we have $(\hat{\sigma}_K^2 - \hat{\sigma}_m^2)/\hat{\sigma}_m^2 \geq M_2$ where $M_2 = O_p(\eta_{N\delta_T}^{-2})$.

Proof of Lemma C.4 (i): Define

$$A_T = \{|(c-b+1)^{-1/2}\sum_{t=b}^{c}\epsilon_t| \leq \kappa_1, 1 \leq b \leq c \leq T\}, \kappa_1 \leq (6\log T)^{-1/2},$$

$$B_T = \{\max_{a,b,c:1\leq a\leq b<c\leq T}|\tilde{D}_{a,c}^b\tilde{\mu}_{a,c}^b| \leq \kappa_2\}, \kappa_2 \geq 8\log T^{-1/2}$$

and

$$D_T^P = \{\forall i = 1,\ldots,m, \exists p = 1,\ldots,P, (\alpha_p,\beta_p) \in I_i \times I_{i+1}\},$$

where $I_i = [t'_{i-1} + \frac{1}{3}(t'_i - t'_{i-1}), t'_{i-1} + \frac{2}{3}(t'_i - t'_{i-1})], i = 1,\ldots,m+1$, by which we only have one change point in each I_i. From the proof of Lemma C.3 in Fryzlewicz (2014), we have $P(A_T \cap B_T \cap C_T) \geq 1 - C_3/T - T\delta_T^{-1}(1 - \delta_T^2 T^{-2}/9)^M$ and the following considerations are valid on the this set. First consider the case $K > m$, we have

$$\begin{aligned}
\frac{\hat{\sigma}_m^2 - \hat{\sigma}_k^2}{\hat{\sigma}_m^2} = & \frac{\sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{X}}_t^m(s_l) - \boldsymbol{X}_t(s_l)\|_F^2 - \sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{X}}_t^K(s_l) - \boldsymbol{X}_t(s_l)\|_F^2}{\sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{X}}_t^m(s_l) - \boldsymbol{X}_t(s_l)\|_F^2} \\
= & \frac{\sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT} \boldsymbol{X}_t(s_l) - \boldsymbol{X}_t(s_l)\|_F^2 - \sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT} \boldsymbol{X}_t(s_l) - \boldsymbol{X}_t(s_l)\|_F^2}{\sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{\alpha}}_t^K \hat{\boldsymbol{\alpha}}_t^{KT} \boldsymbol{X}_t(s_l) - \boldsymbol{X}_t(s_l)\|_F^2} \\
= & \frac{\sum_{t=1}^T \sum_{l=1}^n \boldsymbol{X}_t^T(s_l)[(\hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT})^2 - (\hat{\boldsymbol{\alpha}}_t^K \hat{\boldsymbol{\alpha}}_t^{KT})^2] \boldsymbol{X}_t(s_l)}{\sum_{t=1}^T \sum_{l=1}^n \boldsymbol{X}_t^T(s_l)(\hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT} - \mathbb{I}_N)^2 \boldsymbol{X}_t(s_l)} \\
= & \frac{(N-2)\sum_{t=1}^T \sum_{l=1}^n \boldsymbol{X}_t^T(s_l)[(\hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT} - \hat{\boldsymbol{\alpha}}_t^K \hat{\boldsymbol{\alpha}}_t^{KT})^2] \boldsymbol{X}_t(s_l)}{\sum_{t=1}^T \sum_{l=1}^n \boldsymbol{X}_t^T(s_l)((N-2)\hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT} + \mathbb{I}_N)^2 \boldsymbol{X}_t(s_l)}, \quad (**)
\end{aligned} \tag{A.2}$$

where the equation $(**)$ following by $\hat{\boldsymbol{\alpha}}_t^T \hat{\boldsymbol{\alpha}}_t / N = \mathbb{I}_r$. Without loss of generality, we can assume the length of each subset $T_k = \delta_T$. Furthermore, let

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_t \hat{\boldsymbol{\alpha}}_t^T = & (\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t + \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t)(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t + \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t)^T \\
= & (\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t)(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t)^T + 2\boldsymbol{\alpha}_t^0 \boldsymbol{H}_t(\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t)^T + \boldsymbol{\alpha}_t^0 \boldsymbol{\alpha}_t^{0T},
\end{aligned}$$

then according to Assumption 1 (i), it follows:

$$\begin{aligned}
(**) \leq & \frac{(N-2)\psi_{\max} T \|\hat{\boldsymbol{\alpha}}_t^m - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 + \|\hat{\boldsymbol{\alpha}}_t^K - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 + 2(N-2)\psi_{\max} \sum_{t=1}^T tr(\boldsymbol{\alpha}_t^0 \boldsymbol{H}_t(\hat{\boldsymbol{\alpha}}_t^m - \hat{\boldsymbol{\alpha}}_t^K)}{(N-2)\psi_{\min} T \|\hat{\boldsymbol{\alpha}}_t^m - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 + (N-2)\psi_{\min} T tr(\boldsymbol{\alpha}_t^0 \boldsymbol{\alpha}_t^{0T})} \\
\leq & \frac{\sqrt{N}\psi_{\max} O_p(\eta_{N\delta_T}^{-2}) + 2\psi_{\max} O_p(\eta_{N\delta_T}^{-1})}{\sqrt{N}\psi_{\min} O_p(\eta_{N\delta_T}^{-2}) + \sqrt{N}\psi_{\min}} \\
\sim & \frac{C}{\sqrt{N}} O_p(\eta_{N\delta_T}^{-1}),
\end{aligned}$$

where

$$tr(\boldsymbol{\alpha}_t^0 \boldsymbol{H}_t(\hat{\boldsymbol{\alpha}}_t^m - \hat{\boldsymbol{\alpha}}_t^K) \leq \sqrt{N} tr(\hat{\boldsymbol{\alpha}}_t^m - \hat{\boldsymbol{\alpha}}_t^K) = \sqrt{N}(tr(\hat{\boldsymbol{\alpha}}_t^m - \boldsymbol{\alpha}_t^0) - (\hat{\boldsymbol{\alpha}}_t^K - \boldsymbol{\alpha}_t^0)) \leq \sqrt{N} O_p(\eta_{N\delta_T}^{-1}).$$

We also use

$$\frac{1}{N}\|\hat{\boldsymbol{\alpha}}_t^m - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 = O_p(\eta_{N\delta_T}^{-2})$$

and

$$\frac{1}{N}\|\hat{\boldsymbol{\alpha}}_t^K - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 = O_p(\eta_{N\delta_T}^{-2})$$

following by Proposition 3.1 of Ma and Su (2018).

Proof of Lemma C.4 (ii): Similar to case (i), when $K < m$, we have

$$\begin{aligned}
&\frac{\hat{\sigma}_K^2 - \hat{\sigma}_m^2}{\hat{\sigma}_m^2} = \frac{\sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{X}}_t^K(s_l) - \boldsymbol{X}_t(s_l)\|_F^2 - \sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{X}}_t^m(s_l) - \boldsymbol{X}_t(s_l)\|_F^2}{\sum_{t=1}^T \sum_{l=1}^n \|\hat{\boldsymbol{X}}_t^m(s_l) - \boldsymbol{X}_t(s_l)\|_F^2} \\
=& \frac{(N-2)\sum_{t=1}^T \sum_{l=1}^n \boldsymbol{X}_t^T(s_l)[(\hat{\boldsymbol{\alpha}}_t^K \hat{\boldsymbol{\alpha}}_t^{KT} - \hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT})^2]\boldsymbol{X}_t(s_l)}{\sum_{t=1}^T \sum_{l=1}^n \boldsymbol{X}_t^T(s_l)((N-2)\hat{\boldsymbol{\alpha}}_t^m \hat{\boldsymbol{\alpha}}_t^{mT} + \mathbb{I}_N)^2 \boldsymbol{X}_t(s_l)} \\
\geq& \frac{\sum_{t=1}^T \psi_{\max}\|\hat{\boldsymbol{\alpha}}_t^K - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 - \sum_{t=1}^T \psi_{\min}\|\hat{\boldsymbol{\alpha}}_t^m - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 + 2\psi_{\min}\sum_{t=1}^T tr(\boldsymbol{\alpha}_t^0 \boldsymbol{H}_t(\hat{\boldsymbol{\alpha}}_t^K - \hat{\boldsymbol{\alpha}}_t^m)}{\sum_{t=1}^T \psi_{\max}\|\hat{\boldsymbol{\alpha}}_t^m - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 + 2\psi_{\max}\sum_{t=1}^T tr(\hat{\boldsymbol{\alpha}}_t^m \boldsymbol{H}_t^T \boldsymbol{\alpha}_t^0) + T\psi_{\max}} \\
\geq& \frac{\sqrt{N}\psi_{\min} O_p(\eta_{N\delta_T}^{-1} + \tau_1/\delta_T)^2 - \sqrt{N}\psi_{\max} O_p(\eta_{N\delta_T}^{-2}) + 2\psi_{\min} O_p(\eta_{N\delta_T}^{-1} + \tau_1/\delta_T) - 2\psi_{\max} O_p(\eta_{N\delta_T}^{-1})}{\psi_{\max} O_p(\eta_{N\delta_T}^{-2})/\sqrt{N} + 2\psi_{\max}\sqrt{N} - \psi_{\min}\sqrt{N} + \psi_{\max}/\sqrt{N}} \\
\sim& \frac{C}{\sqrt{N}} O_p(\eta_{N\delta_T}^{-1} + \tau_1/\delta_T)
\end{aligned}$$

on the set $A_T \cap B_T \cap C_T$. We also use

$$\frac{1}{N}\|\hat{\boldsymbol{\alpha}}_t^K - \boldsymbol{\alpha}_t^0 \boldsymbol{H}_t\|_F^2 = O_p(\eta_{N\delta_T}^{-1} + \tau_1/\delta_T)^2$$

when $K < m$ following by Lemma A.1 and A.2 of Ma and Su (2018).

Now we begin the proof of Theorem 4, which suffices to prove that the sSIC criterion achieves its minimum at m. When $K > m$, We have

$$
\begin{aligned}
sSIC(K) - sSIC(m) =& \frac{T}{2}\log\frac{\hat{\sigma}_K^2}{\hat{\sigma}_m^2} + (K-m)\log^{\alpha} T \\
=& \frac{T}{2}\log(1 - \frac{\hat{\sigma}_m^2 - \hat{\sigma}_K^2}{\hat{\sigma}_m^2}) + (K-m)\log^{\alpha} T \\
\geq& -\frac{T}{2}\frac{\hat{\sigma}_m^2 - \hat{\sigma}_K^2}{\hat{\sigma}_m^2} + (K-m)\log^{\alpha} T \\
\geq& -C\sqrt{\delta_T} + (K-m)\log^{\alpha} T
\end{aligned}
$$

which is guaranteed to be positive for T large enough by Lemma C.4. Conversely, if $K < m$, then we have

$$
\begin{aligned}
sSIC(K) - sSIC(m) =& \frac{T}{2}\log\frac{\hat{\sigma}_K^2}{\hat{\sigma}_m^2} + (K-m)\log^{\alpha} T \\
=& \frac{T}{2}\log(1 + \frac{\hat{\sigma}_K^2 - \hat{\sigma}_m^2}{\hat{\sigma}_m^2}) + (K-m)\log^{\alpha} T \\
\geq& \frac{T}{2}\frac{\hat{\sigma}_K^2 - \hat{\sigma}_m^2}{\hat{\sigma}_m^2 - m\log^{\alpha} T} \\
\geq& C{\delta_T}^{3/4} + (K-m)\log^{\alpha} T
\end{aligned}
$$

which is again guaranteed to be positive for T large enough by Lemma C.4 and Assumption 5. Hence for T large enough and on the set $A_T \cap B_T \cap C_T$, it follows that sSIC(k) is minimized at m, leading to the conclusion that $\hat{m} = m$, as desired.

Proof of Theorem 5: Denote $\check{\boldsymbol{V}}_k$ be the $r \times r$ diagonal matrix composed by the first r largest eigenvalues of $(NT_k)^{-1}\sum_{t\in I_k}\boldsymbol{X}_t(\boldsymbol{s})\boldsymbol{X}_t^T(\boldsymbol{s})$ and $\check{\boldsymbol{H}}_k = N^{-1}\check{\boldsymbol{V}}_k^{-1}\hat{\boldsymbol{\alpha}}_k^T\boldsymbol{\alpha}_k^0\boldsymbol{\Lambda}_{\boldsymbol{\xi}}$, where $\boldsymbol{X}_t(\boldsymbol{s}) = (\boldsymbol{X}_t(s_1), \ldots, \boldsymbol{X}_t(s_n))$ and $\boldsymbol{\Lambda}_{\boldsymbol{\xi}} = \mathrm{diag}(\sum_{k=1}^K \rho_{1k}, \ldots, \sum_{k=1}^K \rho_{rk})$, $\rho_{qk} = \mathrm{var}(\xi_{tqk})$. Since $(NT_k)^{-1}\sum_{t\in I_k} n^{-1}\boldsymbol{X}_t(\boldsymbol{s})$

$\boldsymbol{X}_t^T(\boldsymbol{s})\hat{\boldsymbol{\alpha}}_k = \hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k$, we have $\hat{\boldsymbol{\alpha}}_k = (NT_k)^{-1}\sum_{t\in I_k} n^{-1}\boldsymbol{X}_t(\boldsymbol{s})\boldsymbol{X}_t^T(\boldsymbol{s})\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1}$, thus

$$
\begin{aligned}
&\hat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T \\
=&\frac{1}{NT_k}\sum_{t\in I_k}\frac{1}{n}\boldsymbol{X}_t(\boldsymbol{s})\boldsymbol{X}_t^T(\boldsymbol{s})\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T \\
=&\frac{1}{NT_k}\sum_{t\in I_k}\frac{1}{n}\sum_{l=1}^{n}(\boldsymbol{\alpha}_k^0\boldsymbol{\Phi}^T(s_l)\boldsymbol{\xi}_t + \boldsymbol{u}_t(s_l))(\boldsymbol{\alpha}_k^0\boldsymbol{\Phi}^T(s_l)\boldsymbol{\xi}_t + \boldsymbol{u}_t(s_l))^T\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T \\
=&\frac{1}{NT_k}\sum_{t\in I_k}\frac{1}{n}\sum_{l=1}^{n}(\boldsymbol{\alpha}_k^0\boldsymbol{\Phi}^T(s_l)\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T\boldsymbol{\Phi}(s_l)\boldsymbol{\alpha}_k^{0T})\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T \\
+&\frac{1}{NT_k}\sum_{t\in I_k}\frac{1}{n}\sum_{l=1}^{n}\boldsymbol{u}_t(s_l)\boldsymbol{\xi}_t^T\boldsymbol{\Phi}(s_l)\boldsymbol{\alpha}_k^{0T}\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} + \frac{1}{NT_k}\sum_{t\in I_k}\frac{1}{n}\sum_{l=1}^{n}\boldsymbol{\alpha}_k^0\boldsymbol{\Phi}^T(s_l)\boldsymbol{\xi}_t\boldsymbol{u}_t^T(s_l)\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} \\
+&\frac{1}{NT_k}\sum_{t\in I_k}\frac{1}{n}\sum_{l=1}^{n}\boldsymbol{u}_t(s_l)\boldsymbol{u}_t^T(s_l)\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} \\
:=&A_{1k} + A_{2k} + A_{3k} + A_{4k}.
\end{aligned}
$$

First we prove $A_k := (NT_k)^{-1}\sum_{t\in I_k} n^{-1}\sum_{l=1}^{n}(\boldsymbol{\alpha}_k^0\boldsymbol{\Phi}^T(s_l)\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T\boldsymbol{\Phi}(s_l)\boldsymbol{\alpha}_k^{0T})\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T = O_p((NT_k)^{-1})$ by $\int_0^1 E[(NT_k)^{-1}\sum_{t\in I_k}(\boldsymbol{\alpha}_k^0\boldsymbol{\Phi}^T(s)\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T\boldsymbol{\Phi}(s)\boldsymbol{\alpha}_k^{0T})\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1}]ds = N^{-1}\boldsymbol{\alpha}_k^0\boldsymbol{\Lambda}_{\boldsymbol{\xi}}\boldsymbol{\alpha}_k^{0T}\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1} = \boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T$, $E(T_k^{-1}\sum_{t\in I_k} n^{-1}\sum_{l=1}^{n}[f_{tq}^2(s_l) - E(\int_0^1 f_{tq}^2(s_l)dt)]) = 0$ and $\mathrm{var}(T_k^{-1}\sum_{t\in I_k} n^{-1}\sum_{l=1}^{n}[f_{tq}^2(s_l) - E(\int_0^1 f_{tq}^2(s)dt)]) = O_p(1/T_k^2\sum_{t\in I_k}1/n) = O_p(1/nT_k)$.

For $\mathbb{I}_b := (A_{2k})_{ith}^T = (NT_k)^{-1}\check{\boldsymbol{V}}_k^{-1}\sum_{t\in I_k}\hat{\boldsymbol{\alpha}}_k^T\boldsymbol{\alpha}_k^0 n^{-1}\sum_{l=1}^{n}\boldsymbol{\Phi}^T(s_l)\boldsymbol{\xi}_t\boldsymbol{u}_{it}(s_l)$, we have $\|\mathbb{I}_b\|_F = \|\mathbb{I}_b\|_2 \le (NT_k)^{-1}\|\check{\boldsymbol{V}}_k^{-1}\|_2\cdot\|\hat{\boldsymbol{\alpha}}_k\|_F\cdot\|\boldsymbol{\alpha}_k^0\|_2\cdot\|\sum_{t\in I_k}\sum_{l=1}^{n}\boldsymbol{\Phi}^T(s_l)\boldsymbol{\xi}_t\boldsymbol{u}_{it}(s_l)\|_F$ and $\|\hat{\boldsymbol{\alpha}}_k\|_F = \sqrt{tr(\hat{\boldsymbol{\alpha}}_k\hat{\boldsymbol{\alpha}}_k^T)} = O_p(\sqrt{N})$. We first show $\|\check{\boldsymbol{V}}_k^{-1}\|_2 = O_p(1)$. Denote $\boldsymbol{\Sigma}_\xi$, $\boldsymbol{\Sigma}_{\boldsymbol{X}}(s)$ and $\boldsymbol{\Sigma}_{\boldsymbol{u}}(s)$ be the covariance matrix of $\boldsymbol{\xi}_t$, $\boldsymbol{X}_t(s)$ and

$\boldsymbol{u}_t(s)$ respectively for the fixed s. Further, denote $\tilde{\boldsymbol{\Sigma}}_X = \int \boldsymbol{\Sigma}_X(s)ds$ and $\tilde{\boldsymbol{\Sigma}}_u = \int \boldsymbol{\Sigma}_u(s)ds$, then

$$\tilde{\boldsymbol{\Sigma}}_X = \int \boldsymbol{\alpha}_k^0 \boldsymbol{\Phi}^T(s)\boldsymbol{\Sigma}_\xi \boldsymbol{\Phi}(s)\boldsymbol{\alpha}_k^{0T} ds + \tilde{\boldsymbol{\Sigma}}_u = \boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T} + \tilde{\boldsymbol{\Sigma}}_u.$$

Since $\sum_{k=1}^{\infty} \rho_{qk} < \infty$, we have $\|\boldsymbol{\Lambda}_{\boldsymbol{\xi}}\|_2 = \sqrt{\psi_{\max}(\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^T \boldsymbol{\Lambda}_{\boldsymbol{\xi}})} = \max_{1\le q\le r} \sum_{k=1}^{K} \rho_{qk} = O_p(1)$. Since Assumption 4 (iii), $\|\tilde{\boldsymbol{\Sigma}}_u\|_1 = \max_{j=1,\ldots,N} |\int \sum_{i'=1}^{N} E[u_{ij}(s)u_{i'j}(s)]ds| \le C \le \infty$, we obtain $N^{-1}\|\tilde{\boldsymbol{\Sigma}}_X - \boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T}\|_1 = N^{-1}\|\tilde{\boldsymbol{\Sigma}}_u\|_1 \to 0$. Let $\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^T$ be the SVD of $\tilde{\boldsymbol{\Sigma}}_X$, where $\boldsymbol{U} = (u_1, \ldots, u_N)$ and the first nonzero element of u_i is positive for $i = 1, \ldots, N$ and $\boldsymbol{\Sigma}^2 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$ with $\sigma_1^2 \ge \cdots \ge \sigma_N^2 \ge 0$. We further define $\boldsymbol{U}_r = (u_1, \ldots, u_r)$ and $\boldsymbol{\Sigma}_r^2 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_r^2)$. Letting QDK^T be the SVD of $\boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{1/2}$, $\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{1/2}$ can be well defined and $\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1/2}$ because $\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{1/2}$ is diagonal. Then $\boldsymbol{\alpha}_k^0 = QDK^T\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1/2}$ and $\boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T} = QDK^T\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1/2}\boldsymbol{\Lambda}_{\boldsymbol{\xi}}\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{1/2}KD^TQ^T = QDK^TKD^TQ^T = QDD^TQ^T = QD^2Q^T$, which implies QD^2Q^T is the eigenvalues decomposition of $\boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T}$. Thus

$$\begin{aligned} & N^{-1}\|\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^T - QD^2Q^T\|_2 = N^{-1}\|\tilde{\boldsymbol{\Sigma}}_X - \boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T}\|_2 \\ \le & N^{-1}\sqrt{\|\tilde{\boldsymbol{\Sigma}}_X - \boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T}\|_1 \|\tilde{\boldsymbol{\Sigma}}_X - \boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T}\|_\infty} \\ = & N^{-1}\|\tilde{\boldsymbol{\Sigma}}_X - \boldsymbol{\alpha}_k^0 \boldsymbol{\Lambda}_{\boldsymbol{\xi}} \boldsymbol{\alpha}_k^{0T}\|_1 \to 0. \end{aligned} \tag{A.3}$$

Since $\check{\boldsymbol{V}}_k^{-1} = \mathrm{diag}((NT_k)^{-1}\sum_{t\in I_k} \boldsymbol{X}_t(\boldsymbol{s})\boldsymbol{X}_t^T(\boldsymbol{s}); r)$ and $\tilde{\boldsymbol{\Sigma}}_X \approx 1/T_k \sum_{t\in I_k} \boldsymbol{X}_t(\boldsymbol{s})\boldsymbol{X}_t^T(\boldsymbol{s})$, we have $\check{\boldsymbol{V}}_k^{-1} = N^{-1}\boldsymbol{\Sigma}_r^2 + O_P(1)$ and by $\boldsymbol{\alpha}_k^0\boldsymbol{\alpha}_k^{0T} = N\mathbb{I}_r$, we have $\psi(\boldsymbol{\Lambda}_{\boldsymbol{\xi}}) =$

$\psi(\boldsymbol{\alpha}^0/\sqrt{N}\boldsymbol{\Lambda}_{\boldsymbol{\xi}}\boldsymbol{\alpha}^{0T}/\sqrt{N})$. Combined with equation A.3, we have

$$\begin{aligned}
&\|\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\Lambda}_{\boldsymbol{\xi}}\|_2 \\
=&\sqrt{\psi_{\max}(\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\Lambda}_{\boldsymbol{\xi}})(\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\Lambda}_{\boldsymbol{\xi}})^T} \\
=&\psi_{\max}(N^{-1}\boldsymbol{\Sigma}_r^2 - N^{-1}\boldsymbol{\alpha}_k^0\boldsymbol{\Lambda}_{\boldsymbol{\xi}}\boldsymbol{\alpha}_k^{0T}) = o_p(1)
\end{aligned}$$

by Lemma 2 of Jiang et al. (2019), we have $\|\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1}\|_2 = O_p(1) \cdot \|\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\Lambda}_{\boldsymbol{\xi}}\|_2 = o_p(1)$, it follows $\|\check{\boldsymbol{V}}_k^{-1}\| = \|\check{\boldsymbol{V}}_k^{-1} - \boldsymbol{\Lambda}_{\boldsymbol{\xi}} + \boldsymbol{\Lambda}_{\boldsymbol{\xi}}\|_2 = O_p(1)$. Moreover, note that

$$\sum_{q=1}^{r}\sum_{t\in I_k}\sum_{k=1}^{K}\xi_{tqk}^2(\frac{1}{n}\sum_{l=1}^{n}\Phi_{qk}^2(s_l))(\frac{1}{n}\sum_{l=1}^{n}u_{ij}^2(s_l)) = O_p(T_k).$$

Let $\eta = \sum_{q=1}^{r}\sum_{t\neq t'}\sum_{k\neq k'}\xi_{tqk}\xi_{t'qk}(1/n\sum_{l=1}^{n}\Phi_{qk}(s_l)u_{it}(s_l))(1/n\sum_{l=1}^{n}\Phi_{qk'}(s_l)u_{it'}(s_l))$, since $E[\xi_{tqk}\xi_{t'qk'}] = 0$, $E[\xi_{tqk}u_{it}(s_l)] = 0$ and $E[u_{it}(s)] = 0$, we have $E[\eta] = 0$, $\text{var}(\eta) = E(\eta) = O_p(T_k^2)$, which means $\eta = O_p(T_k)$, then

$$\begin{aligned}
&\|\sum_{t\in I_k}\sum_{l=1}^{n}\frac{1}{n}\Phi^T(s_l)\boldsymbol{\xi}_t u_{it}(s_l)\|_F^2 \\
=&\sum_{q=1}^{r}\sum_{t\in I_k}\sum_{k=1}^{K}\xi_{tqk}^2(\frac{1}{n}\sum_{l=1}^{n}\Phi_{qk}^2(s_l))(\frac{1}{n}\sum_{l=1}^{n}u_{ij}^2(s_l)) + \eta \\
=&O_p(T_k).
\end{aligned}$$

This gives $\|\mathbb{I}_b\|_F = \|\mathbb{I}_b\|_2 \leq (NT_k)^{-1}O_p(1)O_p(\sqrt{N})O_p(\sqrt{N})O_p(T_d) = O_p(1/\sqrt{T_k})$.

Let $\mathbb{III}_b := (A_{3k})_{i_{th}}^T = 1/T_k \sum_{t\in I_k} 1/n \sum_{l=1}^n \check{\boldsymbol{V}}_k^{-1}\hat{\boldsymbol{\alpha}}_k \boldsymbol{u}_t(s_l)\boldsymbol{\xi}_t^T \boldsymbol{\Phi}(s_l)\alpha_{ik}^0$ and

$$\begin{aligned}&\frac{1}{T_k}\sum_{t\in I_k}\frac{1}{n}\sum_{l=1}^n \check{\boldsymbol{V}}_k^{-1}\hat{\boldsymbol{\alpha}}_k \boldsymbol{u}_t(s_l)\boldsymbol{\xi}_t^T \boldsymbol{\Phi}(s_l)\alpha_{ik}^0\\ =&\frac{1}{NT_k}\check{\boldsymbol{V}}_k^{-1}\sum_{i'=1}^N(\hat{\alpha}_{i'k}^0-\check{\boldsymbol{H}}_k\alpha_{i'k}^0)(\sum_{t\in T_k}\frac{1}{n}\sum_{l=1}^n u_{ij}(s_l)\boldsymbol{\xi}_t^T\boldsymbol{\Phi}(s_l))\alpha_{i'k}^0\\ +&\frac{1}{NT_k}\check{\boldsymbol{V}}_k^{-1}\sum_{i'=1}^N\check{\boldsymbol{H}}_k\alpha_{i'k}^0\sum_{t\in T_k}\frac{1}{n}\sum_{l=1}^n u_{ij}(s_l)\boldsymbol{\xi}_t^T\boldsymbol{\Phi}(s_l)\alpha_{i'k}^0.\end{aligned}$$

By Lemma B.3 of Bai (2003), we have

$$\begin{aligned}&\frac{1}{N}\hat{\boldsymbol{\alpha}}_k^T\boldsymbol{\alpha}_k^0=\frac{1}{N}(\hat{\boldsymbol{\alpha}}_k^T-\check{\boldsymbol{H}}_k\boldsymbol{\alpha}_k^{0T}+\check{\boldsymbol{H}}_k\boldsymbol{\alpha}_k^{0T})\boldsymbol{\alpha}_k^0\\ =&\frac{1}{N}(\hat{\boldsymbol{\alpha}}_k^T-\check{\boldsymbol{H}}_k\boldsymbol{\alpha}_k^{0T})\boldsymbol{\alpha}_k^0+\check{\boldsymbol{H}}_k=\check{\boldsymbol{H}}_k+O_p(\frac{1}{N}+\frac{1}{T_k})\end{aligned} \tag{A.4}$$

and

$$\frac{1}{N}\hat{\boldsymbol{\alpha}}_k^T\boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T=\frac{1}{N}\hat{\boldsymbol{\alpha}}_k^T(\boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T-\hat{\boldsymbol{\alpha}}_k+\hat{\boldsymbol{\alpha}}_k)=\frac{1}{N}\hat{\boldsymbol{\alpha}}_k^T(\boldsymbol{\alpha}_k^0\check{\boldsymbol{H}}_k^T-\hat{\boldsymbol{\alpha}}_k)+\mathbb{I}_r=\mathbb{I}_r+O_p(\frac{1}{T_k}+\frac{1}{N}).$$

Thus, $\check{\boldsymbol{H}}_k\check{\boldsymbol{H}}_k^T+O_p({T_k}^{-1}+N^{-1})=\mathbb{I}_r$, This shows that $\check{\boldsymbol{H}}_k$ is an orthogonal matrix so that its eigenvalues are either 1 or -1 up to the order of $O_p({T_k}^{-1}+N^{-1})$. From the definition of $\check{\boldsymbol{H}}_k$

$$\check{\boldsymbol{H}}_k=\frac{1}{N}\check{\boldsymbol{V}}_k^{-1}\hat{\boldsymbol{\alpha}}_k^T\boldsymbol{\alpha}_k^0\boldsymbol{\Lambda}_{\boldsymbol{\xi}}=\check{\boldsymbol{V}}_k^{-1}\check{\boldsymbol{H}}_k\boldsymbol{\Lambda}_{\boldsymbol{\xi}}+O_p(\frac{1}{N}+\frac{1}{T_k}),$$

where we use equation A.4. Multiplying $\check{\boldsymbol{V}}_k^{-1}$ on both sides we have $\check{\boldsymbol{V}}_k\check{\boldsymbol{H}}_k=\check{\boldsymbol{H}}_k\boldsymbol{\Lambda}_{\boldsymbol{\xi}}$. This equation implies that $\check{\boldsymbol{H}}_k$ (up to a negligible term) is a matrix consisting of eigenvectors of the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{\xi}}$. It follows that

$$\|\check{\boldsymbol{H}}_k-\mathbb{I}_r\|_F=O_p(\frac{1}{N}+\frac{1}{T_k}). \tag{A.5}$$

Recalling $\hat{\boldsymbol{\alpha}}_k = (NT_k)^{-1}\sum_{t\in T_k} n^{-1}\sum_{l=1}^{n} \boldsymbol{X}_t(\boldsymbol{s})\boldsymbol{X}_t^T(\boldsymbol{s})\hat{\boldsymbol{\alpha}}_k\check{\boldsymbol{V}}_k^{-1}$, we can get that the first term of $\mathbb{II}_b$ is of the order $O_p(N^{-1}+T_k{}^{-1})$ and the second term is of the order $O_p(N^{-1/2}T_k^{-1/2})$ by noting that $\|\check{\boldsymbol{H}}_k-\mathbb{I}_r\|_F = O_p(N^{-1}+T_k{}^{-1})$.

At last, we consider $\mathbb{III}_b := (A_{4k})_{ith}^T = (NT_k)^{-1}\check{\boldsymbol{V}}_k^{-1}\sum_{i'=1}^{N}\hat{\alpha}_{i't}\sum_{t\in I_k} n^{-1}\sum_{l=1}^{n}(u_{it}(s_l)u_{i't}(s_l) - E[u_{it}(s_l)u_{i't}(s_l)]) + (NT_k)^{-1}\check{\boldsymbol{V}}_k^{-1}\sum_{i'=1}^{N}\hat{\alpha}_{i't}\sum_{t\in I_k} n^{-1}\sum_{l=1}^{n} E[u_{it}(s_l)u_{i't}(s_l)]$. Similar to the proof of part (b) in Lemma A.2 in Bai (2003), the first term is $O_p(T_k^{-1} + (NT_k)^{-1/2})$. Then we have

$$
\begin{aligned}
&\|\frac{1}{NT_k}\check{\boldsymbol{V}}_k^{-1}\sum_{i'=1}^{N}\hat{\alpha}_{i't}\sum_{t\in I_k}\frac{1}{n}\sum_{l=1}^{n}E[u_{it}(s_l)u_{i't}(s_l)]\|_2 \\
&\leq \frac{1}{NT_k}\|\check{\boldsymbol{V}}_k^{-1}\|_2\|\hat{\boldsymbol{\alpha}}_k\|_2(\sum_{t\in T_k}\sum_{i=1}^{N}E[u_{it}(s_l)u_{i't}(s_l)]) \\
&= O_p(\frac{1}{\sqrt{N}})
\end{aligned}
$$

by Assumption 4. Then $\|\mathbb{III}_b\|_2 = O_p(T_k^{-1}+(NT_k)^{-1/2}+N^{-1/2})$. Therefore, $\|\hat{\boldsymbol{\alpha}}_t - \check{\boldsymbol{H}}_k\boldsymbol{\alpha}_t^0\|_2 = O_p(N^{-1/2} + T_k^{-1/2} + N_0^{-1/2})$ where the approximation error $N_0^{-1/2} = (Nn)^{-1}$. Following by which we have $\|\hat{\boldsymbol{\alpha}}_{ik} - \check{\boldsymbol{H}}_k\boldsymbol{\alpha}_{it}^0\|_2 = O_p(N^{-1/2}+T_k^{-1/2})$ and $\|\hat{\boldsymbol{\alpha}}_{ik}-\boldsymbol{\alpha}_{ik}^0\|_2 = \|\hat{\boldsymbol{\alpha}}_{ik}-\check{\boldsymbol{H}}_k\boldsymbol{\alpha}_{ik}^0\|_2+\|\check{\boldsymbol{H}}_k-\mathbb{I}_r\|_F\|\boldsymbol{\alpha}_{ik}^0\|_2 = O_p(N^{-1/2} + T_k^{-1/2})$ by A.5. This completes the proof of Theorem 5.

Proof of Theorem 6: By the proof procedure in Theorem 1, for each t, the order of $\hat{\boldsymbol{\alpha}}_{ik} - \boldsymbol{\alpha}_{ik}^0$ is dominated by $\mathbb{I}_b$, thus we have

$$\begin{aligned}
&\sqrt{T_k}(\hat{\boldsymbol{\alpha}}_{ik} - \boldsymbol{\alpha}_{ik}^0) = \sqrt{T_k}(\hat{\boldsymbol{\alpha}}_{ik} - \check{\boldsymbol{H}}_k \boldsymbol{\alpha}_{ik}^0) + O_p(1) \\
=&\check{\boldsymbol{V}}_k^{-1} \frac{1}{N} \hat{\boldsymbol{\alpha}}_k^T \boldsymbol{\alpha}_k^0 \frac{1}{\sqrt{T_k}} \sum_{t \in I_k} \sum_{l=1}^{n} \frac{1}{n} \boldsymbol{\Phi}^T(s_l) \boldsymbol{\xi}_t u_{it}(s_l) + O_p(1) \\
=&\check{\boldsymbol{V}}_k^{-1} \frac{1}{N} \hat{\boldsymbol{\alpha}}_k^T \boldsymbol{\alpha}_k^0 \frac{1}{\sqrt{T_k}} \sum_{t \in I_k} \int \boldsymbol{\Phi}^T(s) \boldsymbol{\xi}_t u_{it}(s) ds + O_p(1)
\end{aligned}$$

based on numerical integration approximation. We have shown that $\check{\boldsymbol{V}}_k = \boldsymbol{\Lambda}_{\boldsymbol{\xi}} + o_p(1)$ and $\check{\boldsymbol{V}}_k^{-1} = \boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1} + o_p(1)$ in the proof of Theorem 5. Further, $\hat{\boldsymbol{\alpha}}_k^T \boldsymbol{\alpha}_k^0 / N = N^{-1}(\hat{\boldsymbol{\alpha}}_k^T - \boldsymbol{\alpha}_k^{0T}) \hat{\boldsymbol{\alpha}}_k^T + N^{-1} \boldsymbol{\alpha}_k^{0T} \boldsymbol{\alpha}_k^0 = \mathbb{I}_r + O_p(1/N + 1/T_k)$. Finally, we have

$$\sqrt{T_k}(\hat{\boldsymbol{\alpha}}_{ik} - \boldsymbol{\alpha}_{ik}^0) = \boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1} \frac{1}{\sqrt{T_k}} \sum_{t \in T_k} \int \boldsymbol{\Phi}^T(s) \boldsymbol{\xi}_t u_{it}(s) ds + O_p(1).$$

Since $E[\int \boldsymbol{\Phi}^T(s) \boldsymbol{\xi}_t u_{it}(s) ds] = 0$, denote $E[\int u_{it}(s) \boldsymbol{\Phi}^T(s) ds \boldsymbol{\xi}_t \boldsymbol{\xi}_t^T \int \boldsymbol{\Phi}^T(s) u_{it}(s) ds]$ as Ψ_i, we have the desired limiting distribution follows from the central limit theorem.

Proof of Theorem 7: Now we can proceed to demonstrate the convergence rate of $\hat{\boldsymbol{\xi}}_t$. Denote $\boldsymbol{M}^*(t) = \boldsymbol{I}_q \otimes \boldsymbol{M}(t)$, where $\otimes$ is Kronecker product, $\boldsymbol{M} = \sum_{l=1}^n \boldsymbol{M}^*(s_l) \boldsymbol{M}^{*T}(s_l)$ and $\tilde{\boldsymbol{V}}_k = \hat{\boldsymbol{\xi}} \hat{\boldsymbol{\xi}}^T / T_k$, we have

$$\hat{\boldsymbol{\xi}}_t = \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} (\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^{n} \boldsymbol{X}_t^T(s_l) \hat{\boldsymbol{\alpha}}_k \boldsymbol{M}^{*T}(s_l)) \boldsymbol{M}^{-1}) \boldsymbol{M}^{-1} (\sum_{l=1}^{n} \boldsymbol{M}^*(s_l) \hat{\boldsymbol{\alpha}}_k^T \boldsymbol{X}_t(s_l)).$$

We note that

$$\hat{\boldsymbol{\xi}}_t - \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} (\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^n \boldsymbol{X}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l)) \boldsymbol{M}^{-1}) \boldsymbol{M}^{-1} (\sum_{l=1}^n \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{X}_t(s_l))$$

$$= \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} (\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^n \boldsymbol{X}_t^T(s_l) (\hat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k^0) \boldsymbol{M}^{*T}(s_l)) \boldsymbol{M}^{-1}) \boldsymbol{M}^{-1} (\sum_{l=1}^n \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{X}_t(s_l)$$

$$+ \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} (\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^n \boldsymbol{X}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l)) \boldsymbol{M}^{-1}) \boldsymbol{M}^{-1} (\sum_{l=1}^n \boldsymbol{M}^*(s_l) (\hat{\boldsymbol{\alpha}}_k^T - \boldsymbol{\alpha}_k^{0T}) \boldsymbol{X}_t(s_l))$$

$$+ \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} (\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^n \boldsymbol{X}_t^T(s_l) (\hat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k^0) \boldsymbol{M}^{*T}(s_l)) \boldsymbol{M}^{-1}) \boldsymbol{M}^{-1} (\sum_{l=1}^n \boldsymbol{M}^*(s_l) (\hat{\boldsymbol{\alpha}}_k^T - \boldsymbol{\alpha}_k^{0T}) \boldsymbol{X}_t(s_l))$$

$$:= C_{t1} + C_{t2} + C_{t3}. \tag{A.6}$$

First, we consider C_{t1}. Note

$$\|C_{t1}\|_2 \leq \frac{1}{T_k N^2 \varphi_{T_k}} \|\tilde{\boldsymbol{V}}_k^{-1}\|_2 \cdot \|\hat{\boldsymbol{\xi}}\|_F \cdot \|(\sum_{l=1}^n \boldsymbol{X}_t^T(s_l) (\hat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k^0) \boldsymbol{M}^{*T}(s_l) \boldsymbol{M}^{-1})$$
$$\times \boldsymbol{M}^{-1} (\sum_{l=1}^n \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{X}_t(s_l))\|_F.$$

Then we have $\|C_{t1}\|_2 = O_p(T^{(e+v)/2}(T_k^{-1/2} + N^{-1/2} + N_0^{-1/2}))$ by noting $\|\hat{\boldsymbol{\xi}}\|_F = O_p(K^{1/2}\|\hat{\boldsymbol{\xi}}\|_2) = O_p(T^{(e+v)/2})$ and $\|\tilde{\boldsymbol{V}}_k^{-1}\|_2 = O_p(1)$.

In the following part, we will provide a detailed proof of $\|\hat{\boldsymbol{\xi}}\|_F = O_p(T^{(e+v)/2})$ and $\|\tilde{\boldsymbol{V}}_k^{-1}\|_2 = O_p(1)$. Denote $\boldsymbol{L}^* = (\boldsymbol{L}_{tt'}^*) \in \mathbb{R}^{T_k \times T_k}$ with

$$\boldsymbol{L}_{tt'}^* = \frac{K}{T_k N^2 \varphi_{T_k}} (\sum_{l=1}^n \boldsymbol{X}_t^T(s_l) \boldsymbol{\alpha}_k \boldsymbol{M}^{*T}(s_l)) \boldsymbol{M}^{-1} \boldsymbol{M}^{-1} (\sum_{l=1}^n \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^T \boldsymbol{X}_t(s_l)).$$

Then $\hat{\boldsymbol{\xi}} = \tilde{\boldsymbol{V}}_k^{-1} \hat{\boldsymbol{\xi}} \boldsymbol{L}^*$. Multiplying both sides of the equation by $T_k^{-1/2} \tilde{\boldsymbol{V}}_k^{1/2}$, we have $T_k^{-1/2} \tilde{\boldsymbol{V}}_k^{1/2} \hat{\boldsymbol{\xi}} = T_k^{-1/2} \tilde{\boldsymbol{V}}_k^{-1/2} \hat{\boldsymbol{\xi}} \boldsymbol{L}^*$. Since $(T_k^{-1/2} \tilde{\boldsymbol{V}}_k^{-1/2} \hat{\boldsymbol{\xi}})(T_k^{-1/2} \tilde{\boldsymbol{V}}_k^{-1/2} \hat{\boldsymbol{\xi}})^T =$

$\mathbb{I}_{Kr\times Kr}$, $\tilde{\boldsymbol{V}}_k$ is diagonal matrix with decreasing entries consisting of the first Kr largest eigenvalues of $\hat{\boldsymbol{L}}^*$ and $T_k^{-1/2}\tilde{\boldsymbol{V}}_k^{1/2}\hat{\boldsymbol{\xi}}$ is the corresponding eigenvector matrix. Note

$$\boldsymbol{X}_t(s) = \boldsymbol{\alpha}_k^0\boldsymbol{\Phi}^T(s)\boldsymbol{\xi}_t + \boldsymbol{u}_t(s) = \boldsymbol{\alpha}_k^0\boldsymbol{M}^{*T}(s))\boldsymbol{\Theta}^T\boldsymbol{\xi}_t + \boldsymbol{\alpha}_k^0\boldsymbol{e}^T(s)\boldsymbol{\xi}_t + \boldsymbol{u}_t(s), \quad \text{(A.7)}$$

where $\boldsymbol{e}(s) = \boldsymbol{\Phi}(s) - \boldsymbol{\Phi}_{T_k}(s) = \mathrm{diag}(\boldsymbol{e}_1(s), \ldots, \boldsymbol{e}_r(s))$ is a $Kr \times r$ block diagonal matrix with block q being $\boldsymbol{e}_q(s) = (e_{q1}(s), \ldots, e_{qK}(s))^T$ and $e_{qk}(s) = \phi_{qk}(s) - \phi_{qk,T_k}(s)$. By the asymptotic property of spline approximation, we have $\sup_{s\in[0,1]}|e_{qk}(s)| = \sup_{s\in[0,1]}|\phi_{qk,T_k}(s) - \phi_{qk}(s)| = O_p(T_k^{-\kappa v})$ for $q = 1\ldots, r, k = 1, \ldots, K$. Thus, $\|\hat{\boldsymbol{L}}^* - {T_k}^{-1}\boldsymbol{\xi}^T\boldsymbol{\xi}\|_2 \leq \|\hat{\boldsymbol{L}}^* - \boldsymbol{L}^*\|_2 + \|\boldsymbol{L}^* - {T_k}^{-1}\boldsymbol{\xi}^T\boldsymbol{\xi}\|_2 = o_p(1)$ similar to proof of Theorem 2 in Wen and Lin (2022). Since the first Kr largest eigenvalues of $\boldsymbol{\xi}^T\boldsymbol{\xi}/T_k$ equal to those of $\boldsymbol{\xi}\boldsymbol{\xi}^T/T_k$ and $\boldsymbol{\xi}\boldsymbol{\xi}^T/T_k$ is diagonal, $\boldsymbol{\xi}\boldsymbol{\xi}^T/T_k$ is the diagonal matrix with decreasing diagonal entries which consists of the first Kr largest eigenvalues of $\boldsymbol{\xi}^T\boldsymbol{\xi}/T_k$. Since $\tilde{\boldsymbol{V}}_k$ is the diagonal matrix with decreasing diagonal entries which consists of the first Kr largest eigenvalues of $\hat{\boldsymbol{L}}^*$, we then have $\|\tilde{\boldsymbol{V}}_k - \Sigma_{\xi,k}\|_2 \leq \max_{p=1,\ldots,Kr}|\psi_p(\hat{\boldsymbol{L}}^*) - \psi_p(\boldsymbol{\xi}^T\boldsymbol{\xi}/T_k)| + \|\boldsymbol{\xi}\boldsymbol{\xi}^T/T_k - \boldsymbol{\Sigma}_{\xi,k}\|_2 \leq \|\hat{\boldsymbol{L}}^* - \boldsymbol{\xi}^T\boldsymbol{\xi}/T_k\|_2 + \|\boldsymbol{\xi}\boldsymbol{\xi}^T/T_k - \boldsymbol{\Sigma}_\xi\|_2 = o_p(1)$ by Weyl's Theorem, where $\boldsymbol{\Sigma}_{\xi,k} = \mathrm{diag}(\boldsymbol{\Sigma}_{\xi,k1}, \ldots, \boldsymbol{\Sigma}_{\xi,kr})$, $\boldsymbol{\Sigma}_{\xi,kq} = \mathrm{diag}(\boldsymbol{\Sigma}_{\xi,kq1}, \ldots, \boldsymbol{\Sigma}_{\xi,kqK})$ and $\|\boldsymbol{\xi}\boldsymbol{\xi}^T/T_k - \boldsymbol{\Sigma}_{\xi,k}\|_2 \to 0$ by Assumption 2. In addition, Assumption 2 show $\|\boldsymbol{\Sigma}_{\xi,k}\|_2 = O_p(1)$, which indicates $\|\hat{\boldsymbol{\xi}}/T_k\| = O_p(1)$ because $\tilde{V}_k = \hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}^T/T_k$.

From Lemma 2 of Jiang et al. (2019) we have $\|\tilde{V}_k^{-1} - \boldsymbol{\Sigma}_{\xi,k}^{-1}\|_2 = O_p(1)$, which implies $\|\tilde{V}_k^{-1}\|_2 = O_p(1)$.

By following the clues presented in the proof of the term C_{t1}, we can deduce that C_{t2} exhibits the same order as C_{t1}. Since $\|\hat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k^0\| = o_p(\|\boldsymbol{\alpha}_k^0\|_F)$, we have $C_{3t} = o_p(C_{1t})$. Therefore,

$$\|\hat{\boldsymbol{\xi}}_t - 1/(T_k N^2)\varphi_{T_k} \tilde{\boldsymbol{V}}_k^{-1}(\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'}(\sum_{l=1}^{n} \boldsymbol{X}_t^T(s_l)\boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l))\boldsymbol{M}^{-1})\boldsymbol{M}^{-1}(\sum_{l=1}^{n} \boldsymbol{M}^*(s_l)\boldsymbol{\alpha}_k^{0T}\boldsymbol{X}_t(s_l)\|_2$$
$$= O_p(T_k^{(e+v)/2}(T_k^{-1/2} + N^{-1/2} + N_0^{-1/2})).$$

We denote $\boldsymbol{H}_k = \tilde{\boldsymbol{V}}_k^{-1}\hat{\boldsymbol{\xi}}\boldsymbol{\xi}^{0T}/T$ and get

$$\begin{aligned}\hat{\boldsymbol{\xi}}_t - \boldsymbol{H}_k\boldsymbol{\xi}_t = &\frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1}(\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'}(\sum_{l=1}^{n} \boldsymbol{X}_t^T(s_l)\hat{\boldsymbol{\alpha}}_k \boldsymbol{M}^{*T}(s_l))\boldsymbol{M}^{-1})\boldsymbol{M}^{-1}(\sum_{l=1}^{n} \boldsymbol{M}^*(s_l)\hat{\boldsymbol{\alpha}}_k^{T}\boldsymbol{X}_t(s_l))\\ &- \frac{1}{T_k}\tilde{\boldsymbol{V}}_k^{-1}\sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'}\boldsymbol{\xi}^T\boldsymbol{\xi} + O_p(T_k^{\frac{e+v}{2}}(T_k^{-1/2} + N^{-1/2} + N_0^{-1/2})).\end{aligned} \tag{A.8}$$

Substituting equation (A.7) into (A.8), we have

$$
\begin{aligned}
\hat{\boldsymbol{\xi}}_t - H_k \boldsymbol{\xi}_t = & \frac{1}{T_k \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} \boldsymbol{\xi}_{t'} (\sum_{l=1}^{n} \boldsymbol{e}(s_l) \boldsymbol{M}^*(s_l)) \boldsymbol{M}^{-1} \boldsymbol{\Theta}^T \boldsymbol{\xi} \\
& + \frac{1}{T_k N \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l) \boldsymbol{M}^{-1}) \boldsymbol{\Theta}^T \boldsymbol{\xi} \\
& + \frac{1}{T_k \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} \boldsymbol{\xi}_{t'} \boldsymbol{\Theta} (\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{e}^T(s_l)) \boldsymbol{\xi} \\
& + \frac{1}{T_k \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} \boldsymbol{\xi}_{t'} (\sum_{l=1}^{n} \boldsymbol{e}(s_l) \boldsymbol{M}^*(s_l) \boldsymbol{M}^{-1}) (\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{e}^T(s_l)) \boldsymbol{\xi} \\
& + \frac{1}{T_k N \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l) \boldsymbol{M}^{-1}) (\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{e}^T(s_l)) \boldsymbol{\xi}
\end{aligned}
$$

$$
\begin{aligned}
& + \frac{1}{T_k \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} \boldsymbol{\xi}_{t'} \boldsymbol{\Theta} (\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s_l)) \\
& + \frac{1}{T_k \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} \boldsymbol{\xi}_{t'} (\sum_{l=1}^{n} \boldsymbol{e}(s_l) \boldsymbol{M}^*(s_l) \boldsymbol{M}^{-1}) (\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s_l)) \\
& + \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} (\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l) \boldsymbol{M}^{-1}) (\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s_l)) \\
& + O_p(T_k^{\frac{e+v}{2}} (T_k^{-1/2} + N^{-1/2} + N_0^{-1/2})) \\
& := I + II + III + IV + V + VI + VII + VIII \\
& + O_p(T_k^{\frac{e+v}{2}} (T_k^{-1/2} + N^{-1/2} + N_0^{-1/2})). \qquad \text{(A.9)}
\end{aligned}
$$

For I, since for each t, we have

$$\begin{aligned}
&\|\sum_{l=1}^{n}\boldsymbol{e}(s_l)\boldsymbol{M}^*(s_l)\boldsymbol{M}^{-1}\boldsymbol{\Theta}^T\|_F\\
\le&(\sum_{q=1}^{r}\sum_{k=1}^{K}\sum_{k'=1}^{K}(\sum_{l=1}^{n}e_{qk}(s_l)\boldsymbol{\Theta}_{qk'}^T\boldsymbol{M}(s_l)\boldsymbol{M}^{-1})^2)^{-1/2}\\
=&O_p(T_k^{-\kappa v+v})
\end{aligned}$$

Then by Cauchy's inequality, we have $\|I\|_2 \le 1/(T_k\varphi_{T_k})\|\tilde{\boldsymbol{V}}_k^{-1}\|_2 \cdot \|\hat{\boldsymbol{\xi}}\|_F \cdot \|\boldsymbol{\xi}\|_2 \cdot \|\boldsymbol{\xi}_t\|_2 \cdot O_p(T_k^{-\kappa v+v})$ with $\|\boldsymbol{\xi}\|_2 = O_p(\sqrt{T}) = O_p(\sqrt{T_k})$ by condition (II) when $T_k \to \infty$. Therefore, $\|I\|_2 = O_p(T_k^{-\kappa v+e/2})$. Similarly, we can get $\|III\|_2 = O_p(T_k^{-\kappa v+e/2})$.

For II, since for each t, we have $\|\sum_{l=1}^{n}\boldsymbol{u}_t^T(s_l)\boldsymbol{\alpha}_k^0\boldsymbol{M}^{*T}(s_l)\boldsymbol{M}^{-1})\|_F = O_p(T_k^{v-1/2}N^{-1/2})$ by Assumption 4, then we have $\|II\|_2 \le K/(T_kN\varphi_{T_k})\|\tilde{\boldsymbol{V}}_k^{-1}\|_2 \cdot \|\hat{\boldsymbol{\xi}}\|_F \cdot \|\boldsymbol{\Theta}\|_F \cdot \|\boldsymbol{\xi}\|_2 \cdot O_p(T_k^{v+e/2-1/2}N^{-1/2})$ and $\|\boldsymbol{\Theta}\|_F$ by equation (2.8). Thus $\|II\|_2 = O_p(T_k^{e+v/2}N^{-1/2})$. Similarly, we have $\|VI\|_2 = O_p(T_k^{e+v/2}N^{-1/2})$.

For IV, since for each t, we have $\|\sum_{l=1}^{n}\boldsymbol{e}(s_l)\boldsymbol{M}^*(s_l)\boldsymbol{M}^{-1})\|_2 = O_p(T_k^{-\kappa v+v/2})$. Then we have

$$\|IV\|_2 \le \frac{1}{T_k\varphi_{T_k}}\|\tilde{\boldsymbol{V}}_k^{-1}\|_2 \cdot \|\hat{\boldsymbol{\xi}}\|_F \cdot \|\boldsymbol{\xi}\|_2 \cdot \|\boldsymbol{\xi}_t\|_2 \cdot O_p(T_k^{-2\kappa v+v}) = O_p(T_k^{-2\kappa v+\frac{e}{2}}) \tag{A.10}$$

and $\|\boldsymbol{V}\|_2 \le 1/(T_kN\varphi_{T_k})\|\tilde{\boldsymbol{V}}_k^{-1}\|_2 \cdot \|\hat{\boldsymbol{\xi}}\|_F \cdot \|\boldsymbol{\xi}_t\|_2 \cdot O_p(T_k^{v-1/2}N^{-1/2}) \cdot O_p(T_k^{-\kappa v+v/2}) = O_p(T_k^{-\kappa v+e/2+1/2}N^{-1/2})$. Similarly, $\|VII\|_2 = O_p(T_k^{-\kappa v+e/2+1/2}N^{-1/2})$.

Similar to C_{3t}, it follows

$$VIII = \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} E[(\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l) \boldsymbol{M}^{-1})(\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s_l))]$$
$$+ \frac{1}{T_k N^2 \varphi_{T_k}} \tilde{\boldsymbol{V}}_k^{-1} \sum_{t' \in I_k} \hat{\boldsymbol{\xi}}_{t'} \{(\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l) \boldsymbol{M}^{-1})(\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s_l))$$
$$- E[(\sum_{l=1}^{n} \boldsymbol{u}_t^T(s_l) \boldsymbol{\alpha}_k^0 \boldsymbol{M}^{*T}(s_l) \boldsymbol{M}^{-1})(\sum_{l=1}^{n} \boldsymbol{M}^{-1} \boldsymbol{M}^*(s_l) \boldsymbol{\alpha}_k^{0T} \boldsymbol{u}_t(s_l))]\}$$

by equation A.9, we can conclude that $((T_k)^{-1} \sum_{t \in T_k} \|\hat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t\|_2^2)^{1/2} = O_p(\delta_{T_k}) = O_p(T_k^{-\kappa v + e/2} + T_k^{(e+v)/2}(T_k^{-1/2} + N^{-1/2} + N_0^{-1/2}) + T_k^{-1+e/2+v} + T_k^{e/2+v} N^{-1})$. Then utilizing the analogous clues in Lemma A.2 of Bai (2003), we can get $\|VIII\|_2 = O_p(T_k^{-1+e/2+v} + \delta_{T_k} T_k^{-1/2+e/2+v} + \delta_{T_k} T_k^{e/2+v} N^{-1/2} + T_k^{v+e/2-1/2} N^{-1/2})$. Therefore, $\|\hat{\boldsymbol{\xi}}_t - \boldsymbol{H}_k \boldsymbol{\xi}_t\|_2 = O_p(T_k^{-1/2+e/2+v/2} + T_k^{e/2+v} N^{-1/2} + T_k^{e/2+v} N_0^{-1/2} + T_k^{-\kappa v + e})$.

Now we show $\boldsymbol{H}_k = \mathbb{I}_{Kr} + O_p(\delta_{T_k} T_k^{-1/2})$. In fact, since $\hat{\boldsymbol{\xi}} \boldsymbol{\xi}^T / T_k = {T_k}^{-1}(\hat{\boldsymbol{\xi}} - \boldsymbol{H}_k \boldsymbol{\xi} + \boldsymbol{H}_k \boldsymbol{\xi}) \boldsymbol{\xi}^T = \boldsymbol{H}_k \Sigma_{\boldsymbol{\xi},k} + O_p(\delta_{T_k} T_k^{-1/2})$ and $\hat{\boldsymbol{\xi}} \boldsymbol{\xi}^T H_k^T / T_k = {T_k}^{-1} \hat{\boldsymbol{\xi}}(\boldsymbol{\xi}^T \boldsymbol{H}_k^T - \hat{\boldsymbol{\xi}}^T + \hat{\boldsymbol{\xi}}^T) = \tilde{\boldsymbol{V}} + O_p(\delta_{T_k} T_k^{-1/2})$, we have $\boldsymbol{H}_k \boldsymbol{H}_k^T = {T_k}^{-1} \tilde{\boldsymbol{V}}_k^{-1} \hat{\boldsymbol{\xi}} \boldsymbol{\xi}^T \boldsymbol{H}_k^T + O_p(\delta_{T_k} T_k^{-1/2}) = \mathbb{I}_{Kr} + O_p(\delta_{T_k} T_k^{-1/2})$. This shows that $\boldsymbol{H}_k$ is an orthogonal matrix with the eigenvalues being either 1 or -1 up to the order of $O_p(\delta_{T_k} T_k^{-1/2})$. By the definition of $\boldsymbol{H}_k$, we have $\boldsymbol{H}_k = {T_k}^{-1} \tilde{\boldsymbol{V}}_k^{-1} \hat{\boldsymbol{\xi}} \boldsymbol{\xi}^T + O_p(\delta_{T_k} T_k^{-1/2}) = \tilde{\boldsymbol{V}}_k^{-1} \boldsymbol{H}_k \Sigma_{\boldsymbol{\xi},k}$, then $\tilde{\boldsymbol{V}}_k^{-1} \boldsymbol{H}_k = \boldsymbol{H}_k \Sigma_{\boldsymbol{\xi},k} + O_p(\delta_{T_k} T_k^{-1/2})$. It implies that $\boldsymbol{H}_k$ is a diagonal matrix up to the order of $O_p(\delta_{T_k} T_k^{-1/2})$. That is, $\|\boldsymbol{H}_k - \mathbb{I}_{Kr}\|_F = O_p(\delta_{T_k} T_k^{-1/2})$. Therefore, $\|\hat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t\|_2 = \|\hat{\boldsymbol{\xi}}_t - H_k \boldsymbol{\xi}_t\|_2 + \|\boldsymbol{H}_k -$

$\mathbb{I}_{Kr}\|_F\|\boldsymbol{\xi}_t\|_2 = O_p(T_k^{-1/2+e/2+v/2} + T_k^{-\kappa v+e/2} + T_k^{e/2+v}N^{-1/2} + T_k^{e/2+v}N_0^{-1/2})$.

Proof of Theorem 8: The proof follows a similar approach to that of Theorem 7. By utilizing the definition of eigenvectors, we can express

$$\hat{\boldsymbol{\Theta}} = \frac{1}{NT_k}\tilde{\boldsymbol{V}}^{-1}(\sum_{t\in I_k}\hat{\boldsymbol{\xi}}_t(\sum_{l=1}^{n}\hat{\boldsymbol{X}}_t^T(s_l)\hat{\boldsymbol{\alpha}}_k\boldsymbol{M}^{*T}(s_l))\boldsymbol{M}^{-1})$$

and

$$\begin{aligned}
&\|\hat{\boldsymbol{\Theta}} - \frac{1}{NT_k}\tilde{\boldsymbol{V}}^{-1}(\sum_{t\in I_k}\hat{\boldsymbol{\xi}}_t(\sum_{l=1}^{n}\hat{\boldsymbol{X}}_t^T(s_l)\boldsymbol{\alpha}_k^0\boldsymbol{M}^{*T}(s_l))\boldsymbol{M}^{-1})\|_F \\
=&\|\frac{1}{NT_k}\tilde{\boldsymbol{V}}^{-1}(\sum_{t\in I_k}\hat{\boldsymbol{\xi}}_t(\sum_{l=1}^{n}\hat{\boldsymbol{X}}_t^T(s_l)(\hat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k^0)\boldsymbol{M}^{*T}(s_l))\boldsymbol{M}^{-1})\|_F \qquad \text{(A.11)} \\
=&O_p((T_k^{-1/2} + N^{-1/2} + N_0^{-1/2})\varphi_{T_k}K^{1/2})
\end{aligned}$$

which is similar to the conclusion in equation (A.6). Then we show that

$$\begin{aligned}
&\|\frac{1}{NT_k}\tilde{\boldsymbol{V}}^{-1}(\sum_{t\in I_k}\hat{\boldsymbol{\xi}}_t(\sum_{l=1}^{n}\hat{\boldsymbol{X}}_t^T(s_l)\boldsymbol{\alpha}_k^0\boldsymbol{M}^{*T}(s_l))\boldsymbol{M}^{-1}) - H\boldsymbol{\Theta}\|_F \\
=&\|\frac{1}{T_k}\sum_{t\in I_k}\hat{\boldsymbol{\xi}}_t\boldsymbol{\xi}_t^T\sum_{l=1}^{n}\boldsymbol{e}(s_l)\boldsymbol{M}^{*T}(s_l))\boldsymbol{M}^{-1}\|_F \qquad \text{(A.12)} \\
+&\|\frac{1}{NT_k}\tilde{\boldsymbol{V}}^{-1}\sum_{t\in I_k}\hat{\boldsymbol{\xi}}_t\sum_{l=1}^{n}\boldsymbol{u}_t^T(s_l)\boldsymbol{\alpha}_k^0\boldsymbol{M}^{*T}(s_l)\boldsymbol{M}^{-1}\|_F.
\end{aligned}$$

The first term of the equation above is $O_p(K^{1/2}\varphi^{-\kappa+1/2})$ and the second term is $O_p(K^{1/2}\varphi_{T_k}N^{-1/2})$, which is similar to I and II in equation (A.9). We have shown that $\|H - I_{Kr}\|_F$ is sufficiently small, so we have $O_p((T_k^{-1/2} + N^{-1/2} + N_0^{-1/2})K^{1/2}\varphi_{T_k} + K^{1/2}\varphi_{T_k}^{-\kappa+1/2}$. And $\|\hat{\boldsymbol{\Theta}}_q - \boldsymbol{\Theta}_q\|_F$ is of the same order as $\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_F$ for each $q = 1, \ldots, r$ because r is finite. Finally we have $\|\hat{\boldsymbol{\Phi}}_t(s) - \boldsymbol{\Phi}_t(s)\|_F = O_p((T_k^{-1/2} + N^{-1/2} + N_0^{-1/2})K^{1/2}\varphi_{T_k} + K^{1/2}\varphi_{T_k}^{-\kappa})$ by noting $\varphi_{T_k} \int \boldsymbol{M}(s)\boldsymbol{M}^T(s)ds = \mathbb{I}_{\varphi_{T_k}}$. Hence, we have successfully completed the proof.

Bibliography

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

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

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*(6), 2243–2281.

Jiang, F., Y. Ma, and Y. Wei (2019). Sufficient direction factor model and its application to gene expression quantitative trait loci discovery. *Biometrika 106*(2), 417–432.

Kovács, S., P. Bühlmann, H. Li, and A. Munk (2023). Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika 110*(1), 249–256.

Liang, X., S. Li, S. Zhang, H. Huang, and S. X. Chen (2016). Pm2. 5 data reliability, consistency, and air quality assessment in five chinese cities. *Journal of Geophysical Research: Atmospheres 121*(17), 10–220.

Liang, X., T. Zou, B. Guo, S. Li, H. Zhang, S. Zhang, H. Huang, and S. X. Chen (2015). Assessing beijing's pm2. 5 pollution: severity, weather impact, apec and winter heating. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 471*(2182), 20150257.

Ma, S. and L. Su (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of econometrics 207*(1), 1–29.

Wen, S. and H. Lin (2022). Factor-guided functional pca for high-dimensional functional data. *arXiv preprint arXiv:2211.12012*.