FACTOR-AUGMENTED MODEL FOR FUNCTIONAL DATA

Yuan Gao†, Han Lin Shang* and Yanrong Yang†

†Australian National University

** Macquarie University*

This Supplement is structured as follows. In Section S1, we extend the proposed model to a nonparametric smoothing approach. In Section S2, we propose a covariance matrix estimator for the discrete raw data. We show the results of extended simulations on the proposed model under different settings in Section S3, and the Canadian weather data are analysed in Section S4. Lastly, Section S5 contains the proofs for the theorems stated in the main paper.

S1 Extending the Model to Nonparametric Smoothing

In the FASM, we use the basis smoothing method where we assume the basis functions $\{\phi_k(u) : k = 1, 2, \ldots, K\}$ are known, and we show the asymptotic properties in the main paper. However, the basis functions are usually unknown in practice, and nonparametric smoothing techniques are frequently used. In this section, we extend the proposed model to a

smoothing spline.

In spline smoothing, a spline basis is used to model functions. We consider smoothing splines, where regularized regression is performed, and the knots are placed on all the observed discrete points. The most commonly considered basis is the cubic smoothing splines, where the order is 4. The order of a polynomial as the spline function is the number of constants required to define it, and is one more than its degree, its highest power. The number of basis functions equals the number of interior knots plus the order. Thus, we use $(p+2)$ spline basis functions. Denote the kth basis function as $\psi_k(u)$, $k = 1, \ldots, p+2$. Let the $p \times (p+2)$ matrix $\boldsymbol{\Psi}$ denote the basis matrix, where the (j,k)th element is $\psi_k(u_j)$. The objective function can be written as

$$\mathrm{SSR}(\boldsymbol{w}_i, \boldsymbol{A}, \boldsymbol{f}) = \frac{1}{np} \sum_{i=1}^{n} \left[(\boldsymbol{Y}_i - \boldsymbol{\Psi}\boldsymbol{w}_i - \boldsymbol{A}\boldsymbol{f}_i)^\top (\boldsymbol{Y}_i - \boldsymbol{\Psi}\boldsymbol{w}_i - \boldsymbol{A}\boldsymbol{f}_i) + \alpha \boldsymbol{w}_i^\top \boldsymbol{U} \boldsymbol{w}_i \right],$$

where $\boldsymbol{w}_i$ is the vector of smoothing coefficients and the matrix $\boldsymbol{U} = \int_{\mathcal{I}} D^2\boldsymbol{\Psi}(s) D^2\boldsymbol{\Psi}^\top(s) ds$ is similarly defined as the matrix $\boldsymbol{R}$ in (3.5). The

estimators are the solution of the equation system

$$\begin{cases} \widehat{\boldsymbol{w}}_i = \left(\boldsymbol{\Psi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}} \boldsymbol{\Psi} + \alpha \boldsymbol{U}^\top\right)^{-1} \boldsymbol{\Psi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}} \boldsymbol{Y}_i, \quad i = 1, \ldots, n \\ \left[\frac{1}{np} \sum_{i=1}^n (\boldsymbol{Y}_i - \boldsymbol{\Psi}\widehat{\boldsymbol{w}}_i)(\boldsymbol{Y}_i - \boldsymbol{\Psi}\widehat{\boldsymbol{w}}_i)^\top\right] \widehat{\boldsymbol{A}} = \widehat{\boldsymbol{A}} \boldsymbol{V}_{np}, \end{cases}$$

which can be solved by the iteration approach in Section 3.2. The function estimator is $\widehat{\mathcal{X}}_i(u) = \widehat{\boldsymbol{w}}_i^\top \boldsymbol{\Psi}(u)$, where $\boldsymbol{\Psi}(u)$ is a vector containing all the $(p+2)$ basis functions $\{\psi_k(u) : k = 1, 2, \ldots, p+2\}$.

It could be seen that the model is almost the same as the model proposed in Section 3. The difference lies in the dimensions of the matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. In the parametric model, we assume the basis functions are known, and the number of basis functions is fixed, so $\boldsymbol{\Phi}$ is a $p \times K$ matrix. While in smoothing spline modeling, the matrix $\boldsymbol{\Psi}$ is of dimension $p \times (p+2)$. The number of basis functions $(p+2)$ goes to infinity. However, this does not render our model estimation infeasible because we have included a penalty term.

The finite-sample performance of this factor-augmented nonparametric smoothing method is shown in S3.3.

S2 Statistical Inference on Covariance Matrix Estimation

Having presented the model estimation approach and the estimators' asymptotic properties, we now consider some common statistical inferences with the proposed FASM. Our model serves as a dimension-reduction technique and avoids the curse of dimensionality, making inferences from the model convenience.

Covariance estimation is fundamental in both FDA and high-dimensional data analysis. Data are of high dimension in these areas, which brings many challenges. The number of discrete points on each curve is often larger than the number of curves in the FDA. Similarly, the dimension p of high-dimensional data is typical of the same order or larger than the sample size n. In this case, the traditional sample covariance estimator no longer works. Dimension reduction by imposing some structures on the data is one of the main ways to solve this problem (see, e.g., Wong et al., 2003; Bickel and Levina, 2008; Fan et al., 2008). We propose an alternative covariance matrix estimator by reducing the data dimension with a smoothing model and a factor model in FASM.

We consider the covariance matrix of the observed high-dimensional

data $\boldsymbol{Y}_i$. Let

$$\boldsymbol{\Sigma}_Y \equiv \mathrm{cov}(\boldsymbol{Y}).$$

Based on the FASM where

$$\boldsymbol{Y}_i = \boldsymbol{\Phi}\boldsymbol{c}_i + \boldsymbol{A}\boldsymbol{f}_i + \boldsymbol{\epsilon}_i,$$

we obtain

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\boldsymbol{c}}\boldsymbol{\Phi}^\top + \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{f}}\boldsymbol{A}^\top + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}, \tag{S2.1}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{c}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{f}}$ are covariance matrices of the vectors $\boldsymbol{c}_i$ and $\boldsymbol{f}_i$ respectively and $\boldsymbol{\Sigma}_\epsilon$ denotes the error variance structure which is a diagonal matrix under Assumption 4. Based on the above equation, we have an estimator

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}} = \boldsymbol{\Phi}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{c}}\boldsymbol{\Phi}^\top + \widehat{\boldsymbol{A}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}\widehat{\boldsymbol{A}}^\top + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}, \tag{S2.2}$$

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{c}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}$ can be calculated by

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{c}} &= \frac{1}{n-1}\boldsymbol{C}\boldsymbol{C}^\top - \frac{1}{n(n-1)}\boldsymbol{C}\boldsymbol{1}\boldsymbol{1}^\top\boldsymbol{C}^\top \\
\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}} &= \frac{1}{n-1}\boldsymbol{F}\boldsymbol{F}^\top - \frac{1}{n(n-1)}\boldsymbol{F}\boldsymbol{1}\boldsymbol{1}^\top\boldsymbol{F}^\top,
\end{aligned}$$

where $\boldsymbol{1}$s are vectors containing ones, the dimensions of which depend on the matrices multiplied before and after the vectors. The diagonal error

covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is estimated by

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \operatorname{diag}\left(n^{-1}\widehat{\boldsymbol{E}}\widehat{\boldsymbol{E}}^{\top}\right),$$

where $\widehat{\boldsymbol{E}}$ is the residual matrix calculated as $\widehat{\boldsymbol{E}} = \boldsymbol{Y} - \boldsymbol{\Phi}\widehat{\boldsymbol{C}} - \widehat{\boldsymbol{A}}\widehat{\boldsymbol{F}}^{\top}$.

Remark 1. In functional data analysis where the functional signal is the focus, the estimation of the covariance function $\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\boldsymbol{c}}\boldsymbol{\Phi}^{\top}$ is of main interest. In this paper, we study the covariance structure of discrete functional data with contamination, described as the mixture of functional data and high-dimensional data. Previous literature has also used this type of covariance estimator based on factor models. For example, Fan et al. (2008) employed a multi-factor model where the factors are assumed observable. In contrast, Fan et al. (2011) considered an extension to approximate factor models where cross-sectional correlation is allowed in the error terms.

We compare the finite-sample performances using the mean squared error (MSE) of the proposed covariance estimator with the ordinary sample covariance estimator. When the factor structure is ignored, the sample covariance estimator is expected to have a larger variance than our estimator. The advantage of the proposed estimator is shown in Section S3.2.

S3 Simulation Results

This section provides some simulation results on various statistical inferences mentioned previously or in the main paper.

S3.1 Comments on choices for the number of factors

The numeric iteration procedure for finding $(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}}, \widehat{\boldsymbol{f}})$ is introduced in Section 3 of the main paper. We now discuss how the number of factors r is selected in each iteration. As mentioned earlier, there are some common criteria on the decision of the number of factors, including the information criteria in Bai and Ng (2002); and the identification of spiked eigenvalues in Onatski (2010) and Ahn and Horenstein (2013). In practice here, we adopt a modified ratio criterion based on Ahn and Horenstein (2013), which is more sensitive to capturing the case of no existing common factors.

In details, at the t-th iteration, let $\upsilon_k^{(t)}$ be the k-th largest eigenvalue of the matrix $(np)^{-1}\sum_{i=1}^n(\boldsymbol{Y}_i - \boldsymbol{\Phi}\widehat{\boldsymbol{c}}_i^{(t+1)})(\boldsymbol{Y}_i - \boldsymbol{\Phi}\widehat{\boldsymbol{c}}_i^{(t+1)})^\top$. Define

$$\mathrm{ER}^{(t)}(k) \equiv \frac{\upsilon_k^{(t)}}{\upsilon_{k+1}^{(t)}}, \quad k = 1, \ldots, kmax, \tag{S3.3}$$

where "ER" refers to "eigenvalue ratio" and $kmax$ is the maximum possible number of factors. Let $\mathrm{ER}_1^{(t)}$ and $\mathrm{ER}_2^{(t)}$ denote the maximum and the second

maximum of $\mathrm{ER}^{(t)}(k), k = 1, \ldots, kmax$. The estimator for r we propose is

$$\widehat{r}^{(t)} = \begin{cases} \mathrm{argmax}_{1 \le k \le kmax} \mathrm{ER}^{(t)}(k), & \text{if } (\mathrm{ER}_1^{(t)} - \mathrm{ER}_2^{(t)})/\mathrm{ER}_1^{(t)} > q; \\ 0, & \text{if } (\mathrm{ER}_1^{(t)} - \mathrm{ER}_2^{(t)})/\mathrm{ER}_1^{(t)} \le q. \end{cases} \tag{S3.4}$$

Remark 2. The ratio criterion (S3.3) is the same as Ahn and Horenstein (2013). However, to expand the gap among values of the ratio $ER^t(k)$ over $k = 1, 2, \ldots, kmax$, we propose a ratio criterion (S3.4) for $ER^{(t)}(k)$. In practice, we find that this criterion is more sensitive to detecting the case of no existing common factors. (S3.4) uses q as a threshold. In practice, q is set to be 0.5. Given this, we will only assume a factor model structure when the largest eigenvalue ratio exceeds the second largest eigenvalue ratio by 50%.

S3.2 Covariance matrix estimation

This section shows the finite sample performance of the covariance estimator defined in (S2.2). We also calculate the regular sample covariance estimator $\widehat{\boldsymbol{\Sigma}}_Y^*$ using

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}^* = \frac{1}{n-1} \left(\boldsymbol{Y} - \overline{\boldsymbol{Y}}\right) \left(\boldsymbol{Y} - \overline{\boldsymbol{Y}}\right)^\top,$$

where the $p \times n$ matrix $\overline{\boldsymbol{Y}}$ is the sample mean matrix whose jth row elements are $n^{-1} \sum_{i=1}^{n} Y_{ij}$.

Both estimators are compared with the population covariance matrix, calculated using (S2.1). We calculate the estimation errors under the Frobenius norm as

$$\text{MSE} = \frac{1}{p} \left\| \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}} - \boldsymbol{\Sigma}_{\boldsymbol{Y}} \right\|_F^2 .$$

We show the MSE results in Table 2. It can be seen that the FASM produces smaller MSE values under almost all cases.

S3.3 Nonparametric smoothing model

In this section, we apply the factor-augmented nonparametric smoothing model introduced in Section S1 to simulated data and compare the results with using nonparametric smoothing models without the factor component.

We generate simulated data Y_{ij}, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ from the following model:

$$\begin{aligned} Y_{ij} &= \mathcal{X}_i(u_j) + \eta_{ij} + \epsilon_{ji} \\ &= \sum_{k=1}^{9} c_{ik} \phi_k(u_j) + \sum_{k=1}^{4} \lambda_{jk} F_{ki} + \epsilon_{ji}, \end{aligned}$$

where $\{\phi_k(u) : k = 1, 2, \ldots, 9\}$ are Fourier basis functions and the smooth-

Table 1: The MSE of the two covariance estimators with different sample sizes and dimensions. The adjustment of σ^2 value is used to control the signal-to-noise ratio.

			MSE	
Sample Size	Dimension	Size of η	FASM	Sample variance
$n = 20$	$p = 51$	$\sigma = 0$	**0.076**	0.104
		$\sigma = 0.5$	**0.108**	0.165
		$\sigma = 0.75$	**0.230**	0.264
		$\sigma = 1$	0.476	**0.450**
$n = 20$	$p = 101$	$\sigma = 0$	**0.066**	0.096
		$\sigma = 0.5$	**0.107**	0.163
		$\sigma = 0.75$	**0.199**	0.271
		$\sigma = 1$	**0.359**	0.442
$n = 50$	$p = 51$	$\sigma = 0$	**0.030**	0.041
		$\sigma = 0.5$	**0.059**	0.063
		$\sigma = 0.75$	0.128	**0.102**
		$\sigma = 1$	0.232	**0.177**
$n = 100$	$p = 101$	$\sigma = 0$	**0.014**	0.019
		$\sigma = 0.5$	**0.022**	0.031
		$\sigma = 0.75$	**0.047**	0.049
		$\sigma = 1$	0.110	**0.085**

ing coefficients c_{ik} are generated from $\mathcal{N}(0, 1.5^2)$. The factors $\{F_{ki}, k = 1, 2, 3, 4\}$ follow $\mathcal{N}(0, 0.5^2)$ and the factor loadings $(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4})^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a 4 by 4 covariance matrix. The random error terms ϵ_{ji} follow $\mathcal{N}(0, 0.5^2)$. We set the multivariate mean term $\boldsymbol{\mu} = \mathbf{0}$ and variance $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_4$. We adjust the value of σ^2 to control the signal-to-noise ratio. When σ^2 is large, the signal-to-noise level is low, and when σ^2 is small, the signal-to-noise level is high.

Smoothing spline

We use order 4 B-spline basis with knots at every data point. With data of dimension p, we use $(p+2)$ basis functions. The tuning parameter λ is selected by the mean generalized cross-validation (3.11) at each iteration step. The covariance estimate is calculated using (S2.2). Table 2 presents the results. Apparently, the proposed factor-augmented nonparametric smoothing performs better than the pure smoothing method as the signal-to-noise ratio level is higher. Mean integrated squared error (MISE) is defined in 5.26.

Table 2: Using smoothing splines: The MISE of the estimated functions and the MSE of the two covariance estimators with different sample sizes and dimensions. The adjustment of σ^2 value is used to control the signal-to-noise ratio.

			MISE		MSE	
Sample Size	Dimension	σ	FASM	Smoothing	FASM	Sample covariance
$n = 20$	$p = 51$	$\sigma = 0.5$	0.154	**0.142**	**1.535**	1.882
		$\sigma = 0.75$	**0.212**	0.219	**1.485**	1.931
		$\sigma = 1$	**0.293**	0.317	**1.649**	2.336
$n = 20$	$p = 101$	$\sigma = 0.5$	**0.076**	**0.076**	**1.479**	1.814
		$\sigma = 0.75$	**0.112**	0.121	**1.593**	2.086
		$\sigma = 1$	**0.159**	0.175	**1.587**	2.268
$n = 50$	$p = 51$	$\sigma = 0.5$	**0.138**	0.142	**0.612**	0.713
		$\sigma = 0.75$	**0.192**	0.213	**0.630**	0.765
		$\sigma = 1$	**0.270**	0.313	**0.690**	0.869
$n = 100$	$p = 101$	$\sigma = 0.5$	**0.070**	0.075	**0.298**	0.349
		$\sigma = 0.75$	**0.104**	0.119	**0.309**	0.377
		$\sigma = 1$	**0.153**	0.177	**0.352**	0.429

S3.4 Misidentification of the basis function

We elaborate on the example presented in Section 2.2. We generate data from

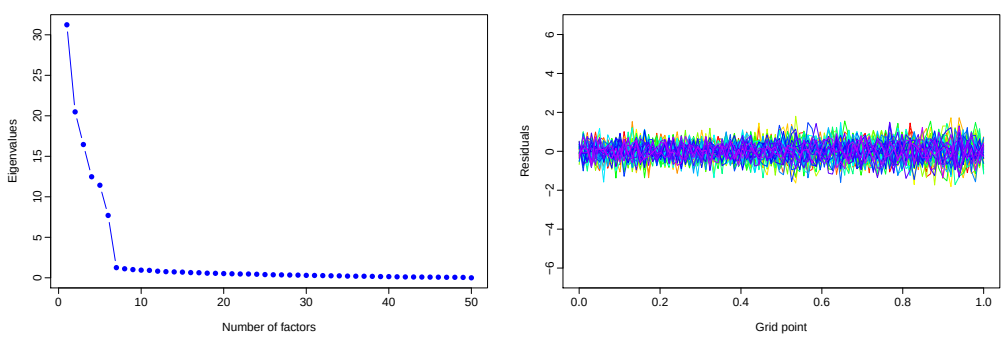
$$Y_{ij} = \sum_{k=1}^{7} c_{ik}\phi_k(u_j) + \epsilon_{ji}, \quad i = 1, \ldots, n, \; j = 1, \ldots, p, \tag{S3.5}$$

where $\{\phi_k(u), k = 1, \ldots, 7\}$ are a set of Fourier basis functions. The first Fourier basis function $\phi_1(u)$ is the constant function; the remainder are sine and cosine pairs with integer multiples of the base period. We generate the Fourier functions with doubled frequencies in the second half to simulate the change in the basis functions. In particular, when $u \in [0, 0.5]$, $\phi_k(u) = 2\sin(k\pi u)$, for $k = 2, 4, 6$, and $\phi_k(u) = 2\cos[(k-1)\pi u]$, for $k = 3, 5, 7$, and when $u \in (0.5, 1]$, $\phi_k(u) = 2\sin(2k\pi u)$, for $k = 2, 4, 6$, and $\phi_k(u) = 2\cos[2(k-1)\pi u]$, for $k = 3, 5, 7$. The coefficients $\{c_{ik} : k = 1, 2, \ldots, 7\}$ are generated from the normal distribution with mean 0 and variance 0.5^2. The error terms are also drawn from the normal distribution with mean 0 and variance 0.5^2. The generated data Y_{ij} are shown in Figure 2a in the main article. It can be seen that the data exhibit more variation in the second half of the interval.

Suppose we were unaware of the change in the frequencies of the basis

functions and used the bases in the first half to fit the data on the whole interval. The smoothing model residuals, shown in Figure 2b in the main article, are large in the second half. When the frequency of the basis functions is misidentified, a smoothing model with the wrong set of bases is inadequate. We conduct principal component analysis on the smoothing residuals; the eigenvalues in descending order are shown in Figure 1a. The residuals preserve a spiked structure, where six common factors explain most of the variation.

We also apply FASM to the same data with the wrong set of basis functions. According to the eigenvalue scree plot, we retain six factors in the model ($r = 6$). The resulting residuals are shown in Figure 1b. The large residuals in the second part of Figure 2b are removed. When the basis functions are misidentified, the FASM serves as a remedy.



(a) Spikes of the smoothing residuals (b) Residuals of FASM

Figure 1: Applying FASM to the data generated by (S3.5).

S3.5 Functional data with step jumps

We study the case where the functional data exhibit a dramatic change in the mean level within a small window. We generate data from the following model

$$Y_{ij} = \mu(u_j) + \sum_{k=1}^{7} c_{ik}\phi_k(u_j) + \epsilon_{ji}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, p,$$

where the basis functions $\phi_k(u)$ are order 4 B-spline bases. The coefficients c_{ik} come from $\mathcal{N}(0, 1.5^2)$ and the error terms from $\mathcal{N}(0, 0.5^2)$. The mean function $\mu(u)$ is generated by a linear combination of 25 B-spline basis functions. Figure 2 shows an example of the mean function-there is a sharp increase in the mean function at around $u = 0.5$.

The change in the mean level happens at $u = 0.5$, and δ denotes the amount of change. Figure 3 is generated using $\delta = 2$. Figure 3 compares the residuals from the smoothing model and the FASM. With the smoothing model, the residuals around the jump are large. In contrast, our model explains the large residuals around the structural break very well. In the aspect of model selection, we consider the trade-off between model fit and model flexibility. We first define a notion of flexibility for a fitted model with the degrees of freedom. We use the same concept as in most textbooks:

degrees of freedom measures the number of parameters estimated from the data required to define the model. The degrees of freedom for the smoothing model is calculated by (3.12) of the last step of convergence. The degree of freedom for the FASM is

$$\mathrm{df} = \mathrm{trace}\left[\boldsymbol{\Phi}(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(t)}}\boldsymbol{\Phi} + \alpha \boldsymbol{R})^{-1}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(t)}}\right] + r,$$

where r is the number of factors retained in the fitted model (see. Green and Silverman, 1993). The larger the degrees of freedom, the more flexible the fitted models are. To quantify the model fitting, we use

$$\mathrm{RMSE} = \sqrt{\frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(Y_{ij} - \widehat{Y}_{ij}\right)^2},$$

where $\widehat{Y}_{ij} = \sum_{k=1}^{K}\widehat{c}_{ik}\phi_k(u_j) + \widehat{\eta}_{ij}$. In Table 3, we show the simulation results by changing the value of the mean shift δ. The RMSE of the FASM is always smaller than the compared model. The degrees of freedom when $\delta = 1$ are similar. When δ increases, the degree of freedom is smaller for the proposed model. Therefore, the FASM achieves better fit with less flexibility.

Table 3: The trade-off between model fitting and flexibility.

	RMSE		DF	
	Smoothing	FASM	Smoothing	FASM
$\delta = 1$	0.2045	**0.1631**	**10.68**	11.15
$\delta = 2$	0.2063	**0.1640**	17.59	**11.03**
$\delta = 3$	0.3308	**0.1647**	14.23	**10.94**

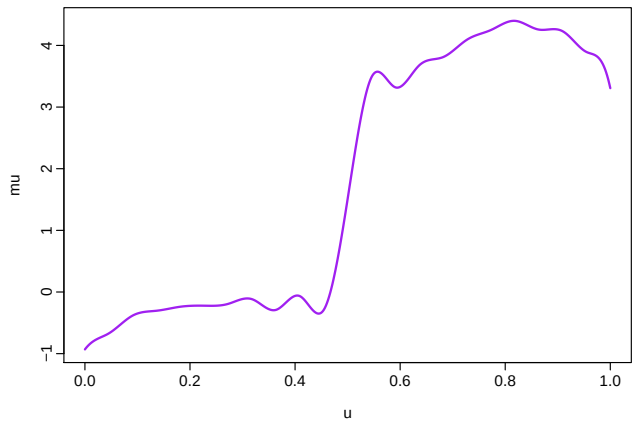
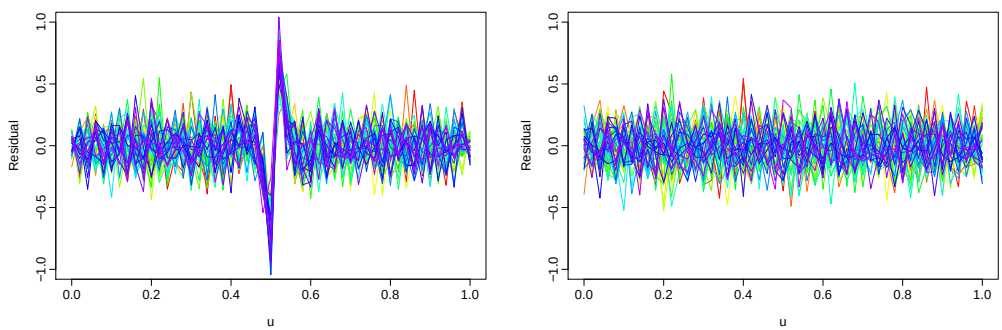


Figure 2: The mean function $\mu(u)$.



(a) Residuals from applying smoothing model (b) Residuals from FASM

Figure 3: Residual plots of the two models.

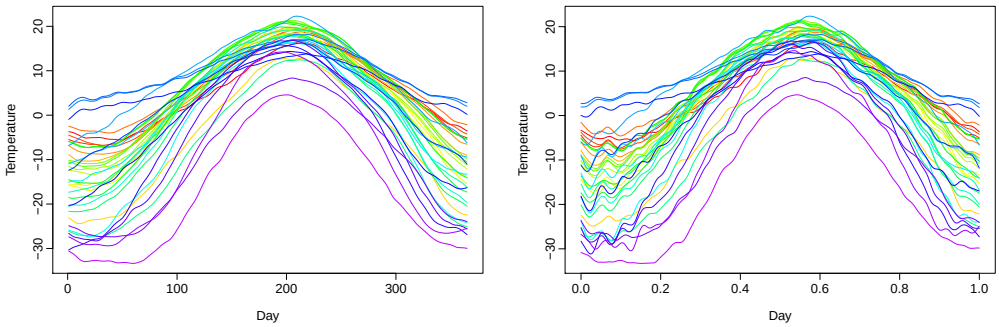
S4 Canadian Weather Data

In Section 2.1, we introduced Canadian weather data. Raw observations of daily temperature and precipitation data are presented in Figure 1. Since the true basis functions are unknown, we apply the FASM with nonparametric smoothing splines introduced in Section S1 to these two datasets.

We use order 4 B-spline basis functions with knots at every data point.

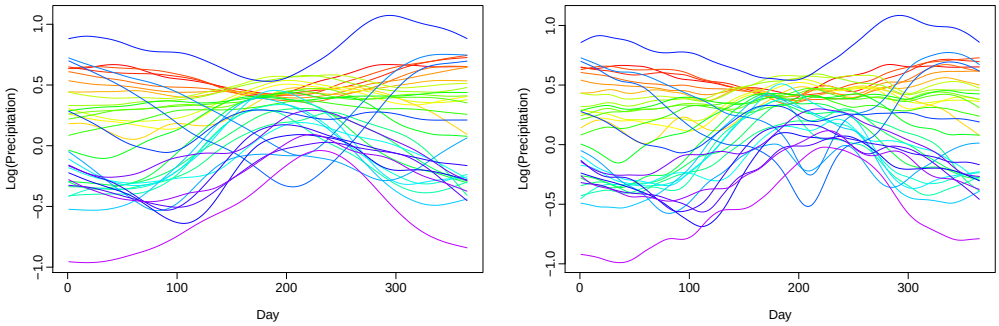
Thus, when the number of data points is 365, we use 367 basis functions. The number of factors r is chosen with the scree plot showing the fraction of variation explained. For temperature data, we presume the measurement error is small. The resulting smoothed curves are shown in Figure 4. Compared with the smoothing model introduced in Section S3.1, the FASM generates similar results. This meets our expectation that our model should work the same as a simple smoothing model when measurement error does not exist.

In Section 2.1, we suspect large measurement errors are contained in the raw log precipitation data. We apply the two models to the log precipitation data; the resulting smoothed curves are presented in Figure 5. The plot on the right shows smoother curves, especially at the drop in the blue curve (the 'Victoria' Station) at around day 200. Looking at the residual plots in Figure 6, our model mainly explains some extreme residuals left out from solely applying the smoothing model. As in Section S2, we also compare the RMSE and degrees of freedom of the two fitted models; they are 0.1933 and 14.41 for the smoothing model and 0.1659 and 12.71, respectively for the proposed model. Thus, in terms of model selection, our model performs better across both model fit and simplicity.



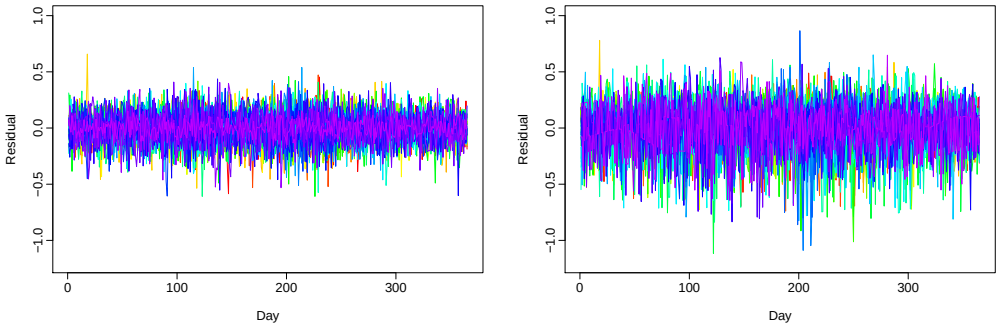
(a) Smoothed curves using FASM (b) Smoothed curves using Bsmooth

Figure 4: Comparison between the smoothed temerature curves.



(a) Smoothed curves using FASM (b) Smoothed curves using Bsmooth

Figure 5: Comparison between the smoothed log(precipitation) curves.



(a) Residuals from FASM (b) Residuals from the FASM

Figure 6: Comparison between the residuals.

S5 Proofs

This section contains the proofs for the theorems in the main article. In S5.1 Appendix A, we provide the proofs for the theorems in Section 4. In S5.2 Appendix B, we include the results of a proposition and its proof. In S5.3 Appendix C, the lemmas used for the proofs in S5.1 Appendix A and S5.2 Appendix B are stated, and their proofs are provided as well.

S5.1 Appendix A

Theorem 2 is the main result of the asymptotic theories, and its proof is lengthy. Thus, we include the outlines for the proof in the following before we show the details.

Outlines for proof of Theorem 2

In Theorem 2, we find the rate of convergence of the estimated coefficient matrix $\widehat{\boldsymbol{C}}$. The difference between $\widehat{\boldsymbol{C}}$ and $\boldsymbol{C}^0$ could be decomposed into three terms:

$$\frac{1}{p}\left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \alpha \boldsymbol{R}\right)\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right) = \frac{1}{p}\alpha \boldsymbol{R}\boldsymbol{C}^0 + \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0\boldsymbol{F}^\top + \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}. \tag{S5.6}$$

Based on Assumption 1, the term $\left|\left|\left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}} \boldsymbol{\Phi} + \alpha \boldsymbol{R}\right)/p\right|\right|$ is $O_p(1)$. The first term on the right-hand side of (4.23) comes from the penalty, and the order can be found easily from Assumption 6. The third term contains the random error matrix $\boldsymbol{E}$, and the order can be found using the result in Lemma 9. The second term is the most complicated one, and we show in the following proof that it could be further broken down into eight terms. We find the order of each of the eight terms using the lemmas in Appendix C. Most of the terms can be shown to be $o_p\left(\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right)$ and thus can be omitted. Combining the remaining terms, we arrive at the result

$$\begin{aligned}\left|\left|\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right)\boldsymbol{M}_{\boldsymbol{F}}\right|\right| \leq \left|\left|\boldsymbol{Q}^{-1}\left(\widehat{\boldsymbol{A}}\right)\frac{1}{p}\alpha \boldsymbol{R}\boldsymbol{C}^0\right|\right| + \left|\left|\boldsymbol{Q}^{-1}\left(\widehat{\boldsymbol{A}}\right)\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\boldsymbol{M}_{\boldsymbol{F}}\right|\right| \\ + O_p\left(\frac{1}{\min(n,p)}\right) + O_p\left(\frac{\sqrt{n}}{p\sqrt{p}}\right) + O_p\left(\frac{1}{\sqrt{np}}\right),\end{aligned} \tag{S5.7}$$

where matrix $\boldsymbol{Q}\left(\widehat{\boldsymbol{A}}\right)$ and $\boldsymbol{M}_{\boldsymbol{F}}$ are

$$\boldsymbol{Q}\left(\widehat{\boldsymbol{A}}\right) = \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} \qquad \boldsymbol{M}_{\boldsymbol{F}} = \boldsymbol{I}_n - \boldsymbol{F}\left(\boldsymbol{F}^\top \boldsymbol{F}\right)^{-1}\boldsymbol{F}^\top.$$

The first term on the right-hand side of (S5.7) is $O_p(1)$ using the assumption on the tuning parameter α. We also show the second term is $O_p(1)$ using results from the lemmas. When n and p are of the same order, we are able

to show the spectral norm of $\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right) \boldsymbol{M}_{\boldsymbol{F}}$ is $O_p(1)$.

Next begins the formal proofs.

Proof of Theorem 1

Proof. The concentrated objective function defined in Section 4.2 is

$$S_{np}(\boldsymbol{c}_i, \boldsymbol{A}) = \frac{1}{np} \sum_{i=1}^{n} \left[(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)^\top \boldsymbol{M}_{\boldsymbol{A}} (\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i) + \alpha \boldsymbol{c}_i^\top \boldsymbol{R}\boldsymbol{c}_i \right] - \frac{1}{np} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{\epsilon}_i.$$

Assume $\boldsymbol{c}_i^0 = \boldsymbol{0}$ for simplicity without loss of generality. From $\boldsymbol{Y_i} = \boldsymbol{\Phi}\boldsymbol{c}_i^0 + \boldsymbol{A}^0 \boldsymbol{f}_i + \boldsymbol{\epsilon}_i = \boldsymbol{A}^0 \boldsymbol{f}_i + \boldsymbol{\epsilon}_i$, we have

$$\begin{aligned}
S_{np}(\boldsymbol{c}_i, \boldsymbol{A}) =& \frac{1}{np} \sum_{i=1}^{n} \left[(\boldsymbol{A}^0 \boldsymbol{f}_i + \boldsymbol{\epsilon}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)^\top \boldsymbol{M}_{\boldsymbol{A}} (\boldsymbol{A}^0 \boldsymbol{f}_i + \boldsymbol{\epsilon}_i - \boldsymbol{\Phi}\boldsymbol{c}_i) + \alpha \boldsymbol{c}_i^\top \boldsymbol{R}\boldsymbol{c}_i \right] - \frac{1}{np} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{\epsilon}_i \\
=& \frac{1}{np} \sum_{i=1}^{n} \boldsymbol{f}_i^\top \boldsymbol{A}^{0\top} \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{A}^0 \boldsymbol{f}_i + \frac{1}{np} \sum_{i=1}^{n} \boldsymbol{c}_i^\top \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{\Phi}\boldsymbol{c}_i - \frac{2}{np} \sum_{i=1}^{n} \boldsymbol{f}_i^\top \boldsymbol{A}^{0\top} \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{\Phi}\boldsymbol{c}_i \\
&+ \frac{2}{np} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i^\top \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{A}^0 \boldsymbol{f}_i - \frac{2}{np} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i^\top \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{\Phi}\boldsymbol{c}_i + \frac{1}{np} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i^\top (\boldsymbol{M}_{\boldsymbol{A}} - \boldsymbol{M}_{\boldsymbol{A}^0}) \boldsymbol{\epsilon}_i + \frac{\alpha}{np} \sum_{i=1}^{n} \boldsymbol{c}_i^\top \boldsymbol{R}\boldsymbol{c}_i.
\end{aligned}$$

Denote the first three terms in the above equation as

$$\widetilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A}) = \frac{1}{np} \sum_{i=1}^{n} \boldsymbol{f}_i^\top \boldsymbol{A}^{0\top} \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{A}^0 \boldsymbol{f}_i + \frac{1}{np} \sum_{i=1}^{n} \boldsymbol{c}_i^\top \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{\Phi}\boldsymbol{c}_i - \frac{2}{np} \sum_{i=1}^{n} \boldsymbol{f}_i^\top \boldsymbol{A}^{0\top} \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{\Phi}\boldsymbol{c}_i.$$

Then by Lemma 4,

$$S_{np}(\boldsymbol{c}_i, \boldsymbol{A}) = \widetilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A}) + o_p(1).$$

It is easy to see that $\widetilde{S}_{np}(\boldsymbol{c}_i^0 = \boldsymbol{0}, \boldsymbol{A}^0\boldsymbol{H}) = 0$ for any $r \times r$ invertible $\boldsymbol{H}$, because $\boldsymbol{M}_{\boldsymbol{A}^0\boldsymbol{H}} = \boldsymbol{M}_{\boldsymbol{A}^0}$ and $\boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{A}^0 = \boldsymbol{0}$.

Here we define two matrix operations before further analysis on $\tilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A})$. For an $m \times n$ matrix $\boldsymbol{U}$ and a $p \times q$ matrix $\boldsymbol{V}$, the vectorization of $\boldsymbol{U}$ is defined as

$$\mathrm{vec}(\boldsymbol{U}) \equiv (u_{1,1}, \ldots, u_{m,1}, u_{1,2}, \ldots, u_{m,2}, u_{1,n}, \ldots, u_{m,n})^\top,$$

and the Kronecker product $\boldsymbol{U} \otimes \boldsymbol{V}$ is the $pm \times qn$ block matrix defined as

$$\boldsymbol{U} \otimes \boldsymbol{V} \equiv \begin{bmatrix} u_{1,1}\boldsymbol{V} & \ldots & u_{1,n}\boldsymbol{V} \\ \vdots & & \vdots \\ u_{m,1}\boldsymbol{V} & \ldots & u_{m,n}\boldsymbol{V} \end{bmatrix},$$

where u_{ij} represents the element on the ith row and jth column of matrix $\boldsymbol{U}$.

Next we can further write $\tilde{S}_{n,p}(\boldsymbol{c}_i, \boldsymbol{A})$ as

$$\widetilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A}) = \mathrm{vec}(\boldsymbol{M_A}\boldsymbol{A}^0)^\top \left(\frac{\boldsymbol{F}^\top \boldsymbol{F}}{np} \otimes \boldsymbol{I}_p\right) \mathrm{vec}(\boldsymbol{M_A}\boldsymbol{A}^0) + \frac{1}{n}\sum_{i=1}^n \boldsymbol{c}_i^\top \left(\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M_A}\boldsymbol{\Phi}\right)\boldsymbol{c}_i$$
$$- \frac{1}{n}\sum_{i=1}^n 2\boldsymbol{c}_i^\top \left(\frac{1}{p}\boldsymbol{f}_i \otimes \boldsymbol{M_A}\boldsymbol{\Phi}\right) \mathrm{vec}(\boldsymbol{M_A}\boldsymbol{A}^0).$$

If we denote

$$\boldsymbol{P} = \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M_A}\boldsymbol{\Phi},$$
$$\boldsymbol{W} = \frac{\boldsymbol{F}^\top \boldsymbol{F}}{np} \otimes \boldsymbol{I}_p,$$
$$\boldsymbol{V}_i = \frac{1}{p}\boldsymbol{f}_i \otimes \boldsymbol{M_A}\boldsymbol{\Phi},$$

and $\boldsymbol{\gamma} = \mathrm{vec}(\boldsymbol{M_A}\boldsymbol{A}^0)$, then we can write

$$\begin{aligned}
\widetilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A}) &= \frac{1}{n}\sum_{i=1}^n \left[\boldsymbol{c}_i^\top \boldsymbol{P}\boldsymbol{c}_i + \boldsymbol{\gamma}^\top \boldsymbol{W}\boldsymbol{\gamma} - 2\boldsymbol{c}_i^\top \boldsymbol{V}_i^\top \boldsymbol{\gamma}\right] \\
&= \frac{1}{n}\sum_{i=1}^n \left[\boldsymbol{c}_i^\top \left(\boldsymbol{P} - \boldsymbol{V}_i^\top \boldsymbol{W}^{-1}\boldsymbol{V}_i\right)\boldsymbol{c}_i + (\boldsymbol{\gamma}^\top - \boldsymbol{c}_i^\top \boldsymbol{V}_i^\top \boldsymbol{W}^{-1})\boldsymbol{W}(\boldsymbol{\gamma}^\top - \boldsymbol{W}^{-1}\boldsymbol{V}_i\boldsymbol{c}_i)\right] \\
&\equiv \frac{1}{n}\sum_{i=1}^n \left[\boldsymbol{c}_i^\top \boldsymbol{D}_i(\boldsymbol{A})\boldsymbol{c}_i + \boldsymbol{\theta}_i^\top \boldsymbol{W}\boldsymbol{\theta}_i\right].
\end{aligned}$$

In the last equation,

$$\boldsymbol{D}_i(\boldsymbol{A}) \equiv \boldsymbol{P} - \boldsymbol{V}_i^\top \boldsymbol{W}^{-1} \boldsymbol{V}_i = \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{\Phi} - \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{\Phi} \boldsymbol{f}_i^\top \left(\frac{\boldsymbol{F}^\top \boldsymbol{F}}{n}\right)^{-1} \boldsymbol{f}_i,$$

$$\boldsymbol{\theta}_i \equiv \boldsymbol{\gamma}^\top - \boldsymbol{W}^{-1} \boldsymbol{V}_i \boldsymbol{c}_i.$$

By Assumption 2 and 3, the matrices $\boldsymbol{D}_i$ and $\boldsymbol{W}$ are positive definite for each i. Thus we have $\widetilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A}) \geq 0$. In addition, if either $\boldsymbol{c}_i \neq \boldsymbol{c}_i^0$ or $\boldsymbol{A} \neq \boldsymbol{A}^0 \boldsymbol{H}$, then $\widetilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A}) > 0$. Thus $\widetilde{S}_{np}(\boldsymbol{c}_i, \boldsymbol{A})$ achieves its unique minimum at $(\boldsymbol{c}_i^0, \boldsymbol{A}^0)$.

This leads us to the result that

$$\left|\left|\widehat{\boldsymbol{c}}_i - \boldsymbol{c}_i^0\right|\right| = o_p(1), \quad \text{uniformly for all } i = 1, \ldots, n$$

Combining the i, we have

$$\frac{\left\|\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right\|}{\sqrt{n}} = o_p(1).$$

To prove part (ii), note that the centred objective function satisfies $S_{np}(\boldsymbol{c}_i^0 = \boldsymbol{0}, \boldsymbol{A}^0) = 0$ and, by definition in (4.24), we have $S_{np}(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}}) \leq 0$. Therefore,

$$0 \geq S_{np}(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}}) = \widetilde{S}_{np}(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}}) + o_p(1).$$

Combined with $\widetilde{S}_{np}(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}}) \geq 0$, it must be true that

$$\widetilde{S}_{np}(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}}) = o_p(1).$$

This implies that

$$\frac{1}{np}\sum_{i=1}^{n} \boldsymbol{F}_i^\top \boldsymbol{A}^{0\top} \boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{A}^0 \boldsymbol{F}_i = \operatorname{tr}\left[\frac{\boldsymbol{A}^{0\top}\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0}{p}\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right] = o_p(1).$$

Since $\boldsymbol{F}^\top\boldsymbol{F}/n = O_p(1)$, it must be true that

$$\frac{\boldsymbol{A}^{0\top}\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0}{p} = \frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p} - \frac{\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}}{p}\frac{\widehat{\boldsymbol{A}}^\top\boldsymbol{A}^0}{p} = o_p(1). \tag{S5.8}$$

By Assumption 4, $\boldsymbol{A}^{0\top}\boldsymbol{A}^0/p$ is invertible. Thus $\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p$ is also invertible. Next,

$$\|\boldsymbol{P}_{\widehat{\boldsymbol{A}}} - \boldsymbol{P}_{\boldsymbol{A}^0}\|^2 \leq \operatorname{tr}[(\boldsymbol{P}_{\widehat{\boldsymbol{A}}} - \boldsymbol{P}_{\boldsymbol{A}^0})^2] \leq 2\operatorname{tr}(\boldsymbol{I}_r - \widehat{\boldsymbol{A}}^\top\boldsymbol{P}_{\boldsymbol{A}^0}\widehat{\boldsymbol{A}}/p).$$

But (S5.8) implies $\widehat{\boldsymbol{A}}^\top\boldsymbol{P}_{\boldsymbol{A}^0}\widehat{\boldsymbol{A}}/p \to \boldsymbol{I}_r$, which means $\|\boldsymbol{P}_{\widehat{\boldsymbol{A}}} - \boldsymbol{P}_{\boldsymbol{A}}\| \to 0$. □

Proof of Theorem 2

Proof. Writing the first equation in (3.10) in matrix notation, we have

$$\widehat{\boldsymbol{C}} = \left(\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \alpha\boldsymbol{R}\right)^{-1}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{Y}. \tag{S5.9}$$

Substitute $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{C}^0 + \boldsymbol{A}^0\boldsymbol{f}^\top + \boldsymbol{E}$ into (S5.9) and subtract the matrix $\boldsymbol{C}^0$ on both sides, we get

$$\begin{aligned}\widehat{\boldsymbol{C}} - \boldsymbol{C}^0 = &\left[\left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \alpha\boldsymbol{R}\right)^{-1}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} - \boldsymbol{I}_K\right]\boldsymbol{C}^0 \\ &+ \left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \alpha\boldsymbol{R}\right)^{-1}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0\boldsymbol{F}^\top \\ &+ \left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \alpha\boldsymbol{R}\right)^{-1}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E},\end{aligned}$$

or

$$\frac{1}{p}\left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \alpha\boldsymbol{R}\right)\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right) = \frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0 + \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0\boldsymbol{F}^\top + \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E} \tag{S5.10}$$

We first look at the second term on the right-hand side of (S5.10). Recall that $\boldsymbol{M}_{\widehat{\boldsymbol{A}}} = \boldsymbol{I}_p - \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top/p$. We have $\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\widehat{\boldsymbol{A}} = \boldsymbol{0}$. Thus

$$\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0 = \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(\boldsymbol{A}^0 - \widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} + \widehat{\boldsymbol{A}}\boldsymbol{H}^{-1}\right) = \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(\boldsymbol{A}^0 - \widehat{\boldsymbol{A}}\boldsymbol{H}^{-1}\right),$$

where $\boldsymbol{H}$ is defined as

$$\boldsymbol{H} = (\boldsymbol{F}^\top\boldsymbol{F}/n)^{-1}(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)^{-1}\boldsymbol{V}_{np}, \tag{S5.11}$$

Using (S5.37), it follows that

$$\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0\boldsymbol{F}^\top = -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(I1+\cdots+I8\right)\left(\frac{\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}}{p}\right)^{-1}\left(\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right)^{-1}\boldsymbol{F}^\top \tag{S5.12}$$

$$\equiv J1+\cdots+J8.$$

In the following, we calculate the order for each from $J1$ to $J8$. Note that $I1$ to $I8$ are defined in (S5.35). Before we begin, for simplicity, denote

$$\boldsymbol{G} \equiv \left(\frac{\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}}{p}\right)^{-1}\left(\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right)^{-1}. \tag{S5.13}$$

We prove in Lemma 5 that $\boldsymbol{G} = O_p(1)$. We also use the fact that $\left\|\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\right\| = O_p(1)$. Now

$$J1 = -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(I1\right)\boldsymbol{G}\boldsymbol{F}^\top. \tag{S5.14}$$

Since $I1 = O_p\left(\sqrt{p}n^{-1}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|^2\right)$, using the result from Lemma 2 (i), the term $J1$ is bounded in norm by $O_p\left(\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|^2/\sqrt{n}\right)$. Thus it is also $o_p\left(\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)$.

$$\begin{aligned} J2 &= -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(I2\right)\left(\frac{\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}}{p}\right)^{-1}\left(\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right)^{-1}\boldsymbol{F}^\top \\ &= \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi}\left(\widehat{\boldsymbol{C}}-\boldsymbol{C}^0\right)\boldsymbol{F}\left(\boldsymbol{F}^\top\boldsymbol{F}\right)^{-1}\boldsymbol{F}^\top. \end{aligned} \tag{S5.15}$$

For the term $J2$, since it is not a small order term, we keep it as what it is.

Then

$$\begin{aligned} J3 =& -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(I3\right)\left(\frac{\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}}{p}\right)^{-1}\left(\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right)^{-1}\boldsymbol{F}^\top \\ =&\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi}\left(\widehat{\boldsymbol{C}}-\boldsymbol{C}^0\right)\boldsymbol{E}^\top\widehat{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{F}^\top \end{aligned} \tag{S5.16}$$

We consider

$$\begin{aligned} \left\|\boldsymbol{E}^\top\widehat{\boldsymbol{A}}\right\| \leq& \left\| \boldsymbol{E}^\top\left(\widehat{\boldsymbol{A}}-\boldsymbol{A}^0\boldsymbol{H}\right)\right\| + \left\|\boldsymbol{E}^\top\boldsymbol{A}^0\boldsymbol{H}\right\| \\ =& O_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{p}{\sqrt{n}}\right) + O_p\left(\sqrt{np}\right), \end{aligned} \tag{S5.17}$$

where Lemma 6 is used. Then we can calculate $\|J3\| = O_p\left(\frac{\sqrt{K}}{\sqrt{p}}\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\|\right) = o_p\left(\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)$ under Assumptin 1.

Next

$$\begin{aligned} \|J4\| =& \left\|-\frac{1}{p}\boldsymbol{\Phi}\boldsymbol{M}_{\widehat{\boldsymbol{A}}}I4\left(\frac{\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}}{p}\right)^{-1}\left(\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right)^{-1}\boldsymbol{F}^\top\right\| \\ =& O_p\left(\frac{\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0}{\sqrt{p}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right). \end{aligned} \tag{S5.18}$$

Using Proposition 1, we have $\left|\left|\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0/\sqrt{p}\right|\right| = \left\|\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(\boldsymbol{A}^0-\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1}\right)/\sqrt{p}\right\| = o_p\left(1\right)$, Thus, $\|J4\| = o_p\left(\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)$.

Next

$$
\begin{aligned}
J5 &= -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}(I5)\boldsymbol{G}\boldsymbol{F}^\top \\
&= -\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\left(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right)^\top \boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{F}^\top.
\end{aligned}
$$

Let's consider

$$
\begin{aligned}
\left\|\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\right\| &\leq \left\|\boldsymbol{\Phi}^\top \boldsymbol{E}\right\| + \left\|\boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top \boldsymbol{E}/p\right\| \\
&\leq \left\|\boldsymbol{\Phi}^\top \boldsymbol{E}\right\| + \frac{1}{p}\|\boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}}\|\|\widehat{\boldsymbol{A}}^\top \boldsymbol{E}\| \\
&= O_p(\sqrt{npK}) + O_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{p}{\sqrt{n}}\right) \\
&= O_p(\sqrt{npK}) + O_p\left(\frac{p}{\sqrt{n}}\right),
\end{aligned}
$$

where (S5.17) is used. We can then calculate $\|J5\| = O_p\left(\frac{\sqrt{K}}{\sqrt{p}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) = o_p\left(\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right)$ under Assumption 1.

Then we consider

$$
\begin{aligned}
J6 &= -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}(I6)\boldsymbol{G}\boldsymbol{F}^\top \\
&= -\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{A}^0\boldsymbol{F}^\top \boldsymbol{E}^\top \widehat{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{F}^\top \\
&= -\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(\boldsymbol{A}^0 - \widehat{\boldsymbol{A}}\boldsymbol{H}^{-1}\right)\boldsymbol{F}^\top \boldsymbol{E}^\top \widehat{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{F}^\top,
\end{aligned}
$$

where the last equation comes from $\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} = \boldsymbol{0}$. using Lemma 6. Thus,

$$
\begin{aligned}
\|J6\| &\leq \left\|\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\left(\boldsymbol{A}^0 - \widehat{\boldsymbol{A}}\boldsymbol{H}^{-1}\right)\right\| \left\|\boldsymbol{F}^\top \boldsymbol{E}^\top \widehat{\boldsymbol{A}}\right\| \|\boldsymbol{G}\| \left\|\boldsymbol{F}^\top\right\| \\
&= -\frac{1}{np^2} \times O_p(\sqrt{p}) \times \left[O_p\left(\frac{\sqrt{p}}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{p}}{\min\left(\sqrt{n}, \sqrt{p}\right)}\right)\right] \\
&\times \left[O_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{p}{\sqrt{n}}\right) + O_p\left(\sqrt{np}\right)\right] \times O_p\left(\sqrt{n}\right) \\
&= o_p\left(\frac{\sqrt{K}}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{n\sqrt{n}}\right) + O_p\left(\frac{1}{p\sqrt{p}}\right) \qquad \text{(S5.19)} \\
&= o_p\left(\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{n\sqrt{n}}\right) + O_p\left(\frac{1}{p\sqrt{p}}\right), \qquad \text{(S5.20)}
\end{aligned}
$$

where Proposition 1 is used in the first equation, and we use $\left\|\boldsymbol{F}^\top \boldsymbol{E}^\top \widehat{\boldsymbol{A}}\right\| / \left\|\boldsymbol{E}^\top \widehat{\boldsymbol{A}}\right\| = O_p(1)$.

Next

$$
\begin{aligned}
J7 &= -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}(I7)\boldsymbol{G}\boldsymbol{F}^\top \\
&= -\frac{1}{np}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\boldsymbol{F}\left(\frac{\boldsymbol{F}^\top \boldsymbol{F}}{n}\right)^{-1}\boldsymbol{F}^\top. \qquad \text{(S5.21)}
\end{aligned}
$$

This term is not a small order term, so we keep it as what it is. And lastly, the proof of order for the term $J8$ is too long, so we show in Lemma 10

that

$$J8 = o_p\left(\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\right) + O_p\left(\frac{\sqrt{nK}}{p\sqrt{p}}\right) + O_p\left(\frac{1}{n}\right), \quad \text{(S5.22)}$$

Collecting terms from $J1$ to $J8$, we can write (S5.10) as

$$\left(\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \frac{1}{p}\alpha\boldsymbol{R}\right)\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right) = \frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0 + J1 + \cdots + J8 + \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}.$$

Combining the results we have found for $J1$, $J3$, $J4$, $J5$, $J6$ and $J8$ in eqs. (S5.16), (S5.18), (S5.19) and (S5.22),

$$\left(\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + o_p(1)\right)\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right) - J2 = \frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0 + \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E} + J7 + \boldsymbol{\Delta}_n^{(1)}, \quad \text{(S5.23)}$$

, where $\boldsymbol{\Delta}_n^{(1)}$ satisfies

$$\left|\left|\boldsymbol{\Delta}_n^{(1)}\right|\right| = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{\sqrt{nK}}{p\sqrt{p}}\right) + O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\right). \quad \text{(S5.24)}$$

Substitute $J2$ and $J7$ from eqs. (S5.15) and (S5.21) into (S5.23), we have

$$\begin{aligned}&\left(\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + o_p(1)\right)\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right) - \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi}\left(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0\right)\boldsymbol{F}\left(\boldsymbol{F}^\top\boldsymbol{F}\right)^{-1}\boldsymbol{F}^\top \\ &= \frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0 + \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E} - \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\boldsymbol{F}\left(\boldsymbol{F}^\top\boldsymbol{F}\right)^{-1}\boldsymbol{F}^\top + \boldsymbol{\Delta}_n^{(1)}. \quad \text{(S5.25)}\end{aligned}$$

We combine the two terms on the left-hand side of (S5.25) and also combine the second and third term on the right-hand side of (S5.25), then we get

$$\begin{aligned}&\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi}(\widehat{\boldsymbol{C}}-\boldsymbol{C}^0)\left(\boldsymbol{I}_n-\boldsymbol{F}\left(\boldsymbol{F}^\top\boldsymbol{F}\right)^{-1}\boldsymbol{F}^\top\right)\\ =&\frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0+\frac{1}{p}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\left(\boldsymbol{I}_n-\boldsymbol{F}(\boldsymbol{F}^\top\boldsymbol{F})^{-1}\boldsymbol{F}^\top\right)+\boldsymbol{\Delta}_n^{(1)}.\end{aligned}$$

Let $\boldsymbol{Q}(\widehat{\boldsymbol{A}})\equiv\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi}/p$, and $\boldsymbol{M}_{\boldsymbol{F}}\equiv\boldsymbol{I}_n-\boldsymbol{F}(\boldsymbol{F}^\top\boldsymbol{F})^{-1}\boldsymbol{F}^\top$. Left multiplying $\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}$ to both sides of the equation above, we have

$$(\widehat{\boldsymbol{C}}-\boldsymbol{C}^0)\boldsymbol{M}_{\boldsymbol{F}}=\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0+\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\frac{1}{p}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\boldsymbol{M}_{\boldsymbol{F}}+\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\boldsymbol{\Delta}_n^{(1)} \tag{S5.26}$$

$$=\boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0+\boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{p}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\boldsymbol{M}_{\boldsymbol{F}}+\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\boldsymbol{\Delta}_n^{(1)}+\boldsymbol{\Delta}_n^{(2)}, \tag{S5.27}$$

where, by Lemma 7, we have

$$\left|\left|\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\boldsymbol{\Delta}_n^{(1)}+\boldsymbol{\Delta}_n^{(2)}\right|\right|=O_p\left(\frac{1}{n}\right)+O_p\left(\frac{\sqrt{nK}}{p\sqrt{p}}\right)+O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\right). \tag{S5.28}$$

Scale the equation (S5.26) with $\sqrt{p}/\sqrt{n}$ and substitute $(\sqrt{np})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}$

with $(\sqrt{np})^{-1}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{E}$ using Lemma 9. We can write

$$\begin{aligned}\frac{\sqrt{p}}{\sqrt{n}}(\widehat{\boldsymbol{C}}-\boldsymbol{C}^0)\boldsymbol{M}_{\boldsymbol{F}} =& \boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0 + \boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{E}\boldsymbol{M}_{\boldsymbol{F}} \\ &+ \frac{\sqrt{p}}{\sqrt{n}}\left(\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\boldsymbol{\Delta}_n^{(1)} + \boldsymbol{\Delta}_n^{(2)}\right) + \boldsymbol{\Delta}_n^{(3)}, \end{aligned} \tag{S5.29}$$

where

$$\left|\left|\frac{\sqrt{p}}{\sqrt{n}}\left(\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\boldsymbol{\Delta}_n^{(1)} + \boldsymbol{\Delta}_n^{(2)}\right) + \boldsymbol{\Delta}_n^{(3)}\right|\right| = O_p\left(\frac{p}{n^2}\right) + O_p\left(\frac{\sqrt{K}}{\min(n,p)}\right), \tag{S5.30}$$

Here we use the assumption $K = o(\min(n,p))$ and that the terms $\sqrt{p}\frac{\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\|^2}{n}$, $\frac{\sqrt{p}}{\sqrt{n}}\frac{\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\|}{\sqrt{n}}$, and $\sqrt{K}\frac{\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\|}{\sqrt{n}}$ from Lemma 9 are dominated by $\frac{\sqrt{p}}{\sqrt{n}}\left|\left|\widehat{\boldsymbol{C}}-\boldsymbol{C}^0\right|\right|$. Thus, by (S5.29) and (S5.30), we have when $p/n^2 \to 0$

$$\frac{\sqrt{p}}{\sqrt{n}}(\widehat{\boldsymbol{C}}-\boldsymbol{C}^0)\boldsymbol{M}_{\boldsymbol{F}} = \boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{E}\boldsymbol{M}_{\boldsymbol{F}} + \boldsymbol{\Delta}_n^{(4)}, \tag{S5.31}$$

where

$$||\boldsymbol{\Delta}_n^{(4)}|| := \left|\left|\frac{\sqrt{p}}{\sqrt{n}}\left(\boldsymbol{Q}(\widehat{\boldsymbol{A}})^{-1}\boldsymbol{\Delta}_n^{(1)} + \boldsymbol{\Delta}_n^{(2)}\right) + \boldsymbol{\Delta}_n^{(3)} + \boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{p}\alpha\boldsymbol{R}\boldsymbol{C}^0\right|\right| = o_p(1). \tag{S5.32}$$

□

Proof of Theorem 3

From (S5.31) and (S5.32), we have when K is fixed,

$$\frac{\sqrt{p}}{\sqrt{n}}(\widehat{\boldsymbol{C}} - \boldsymbol{C}^0)\boldsymbol{M_F} = \boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}}\boldsymbol{E}\boldsymbol{M_F} + o_p(1).$$

Using Lemma 1 we have, for any vector $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$,

$$\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}}\boldsymbol{E}\boldsymbol{M_F}\boldsymbol{b} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{L}), \tag{S5.33}$$

where $\boldsymbol{L}$ is defined in (4.17).

Multiplying the constant matrix $\boldsymbol{Q}(\boldsymbol{A}^0)^{-1}$ to (S5.33), we have the result

$$\boldsymbol{Q}(\boldsymbol{A}^0)^{-1}\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}}\boldsymbol{E}\boldsymbol{M_F}\boldsymbol{b} \xrightarrow{d} \mathcal{N}\left(0, \widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{L}\widetilde{\boldsymbol{Q}}^{-1}\right).$$

Proof of Theorem 4

Proof. For any vectors $\boldsymbol{\gamma} \in \mathbb{R}^K$ and $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$,

$$\boldsymbol{\gamma}^\top \left(\frac{1}{p}\Phi^\top \boldsymbol{M_{A^0}}\boldsymbol{\Phi}\right)^{-1} \frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}}\boldsymbol{E}\boldsymbol{M_F}\mathbf{b} = \frac{1}{\sqrt{np}}\sum_i^n\sum_j^p \boldsymbol{\gamma}^\top \widetilde{\boldsymbol{\omega}}_j \epsilon_{ji} \sum_k^n \psi_{ik} b_k \equiv \frac{1}{\sqrt{np}}\sum_i^n\sum_j^p \widetilde{x}_{ij}.$$

Since we assume ϵ_{ji} are i.i.d., the variance of the above quantity is given by

$$\operatorname{var}\left(\frac{1}{\sqrt{np}}\sum_i^n\sum_j^p \widetilde{x}_{ij}\right) = \frac{\sigma^2}{np}\sum_i^n\sum_j^p \left(\boldsymbol{\gamma}^\top \widetilde{\boldsymbol{\omega}}_j\right)^2 \left(\sum_k^n \psi_{ik} b_k\right)^2.$$

The Lindeberg condition is assumed to hold in Assumption 9. Thus we have a central limit theorem result

$$\frac{1}{\sqrt{np}} \sum_{i}^{n} \sum_{j}^{p} \widetilde{x}_{ij} \xrightarrow{d} \mathcal{N}(0, \widetilde{L}),$$

where $\widetilde{L}$ is defined in (4.20). □

S5.2 Appendix B

In this section, we provide the proposition used in Appendix A, along with its proof.

Proposition 1. *Under Assumptions 1 to 4, we have the following statements:*

(i) The matrix $\boldsymbol{V}_{np}$ defined in (3.9) *is invertible and $\boldsymbol{V}_{np} \xrightarrow{p} \boldsymbol{V}$, where the $r \times r$ matrix $\boldsymbol{V}$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma_f \Sigma_a}$;*

(ii) The matrix $\boldsymbol{H}$ defined in (S5.11) is an $r \times r$ invertible matrix and

$$\frac{1}{p}\|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\|^2 = O_p\left(\frac{1}{n}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|^2\right) + O_p\left(\frac{1}{\min(n,p)}\right).$$

Proof. Write the second equation in (3.10) in a matrix form, we have

$$\frac{1}{np}(\boldsymbol{Y}-\boldsymbol{\Phi}\widehat{\boldsymbol{C}})(\boldsymbol{Y}-\boldsymbol{\Phi}\widehat{\boldsymbol{C}})^\top\widehat{\boldsymbol{A}}=\widehat{\boldsymbol{A}}\boldsymbol{V}_{np}.$$

By (3.2), we also have

$$\boldsymbol{Y}-\boldsymbol{\Phi}\widehat{\boldsymbol{C}}=\boldsymbol{\Phi}(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})+\boldsymbol{A}^0\boldsymbol{F}^\top+\boldsymbol{E}. \tag{S5.34}$$

Plugging it in (S5.34) and by expanding terms, we obtain

$$\begin{aligned}
\widehat{\boldsymbol{A}}\boldsymbol{V}_{np}=&\frac{1}{np}\left[\boldsymbol{\Phi}(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})+\boldsymbol{A}^0\boldsymbol{F}^\top+\boldsymbol{E}\right]\left[\boldsymbol{\Phi}(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})+\boldsymbol{A}^0\boldsymbol{F}^\top+\boldsymbol{E}\right]^\top\widehat{\boldsymbol{A}}\\
=&\frac{1}{np}\boldsymbol{\Phi}(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})^\top\boldsymbol{\Phi}^\top\widehat{\boldsymbol{A}}+\frac{1}{np}\boldsymbol{\Phi}(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})\boldsymbol{F}\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}\\
&+\frac{1}{np}\boldsymbol{\Phi}(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})\boldsymbol{E}^\top\widehat{\boldsymbol{A}}+\frac{1}{np}\boldsymbol{A}^0\boldsymbol{F}^\top(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})^\top\boldsymbol{\Phi}^\top\widehat{\boldsymbol{A}},\\
&+\frac{1}{np}\boldsymbol{E}(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})^\top\boldsymbol{\Phi}^\top\widehat{\boldsymbol{A}}+\frac{1}{np}\boldsymbol{A}^0\boldsymbol{F}^\top\boldsymbol{E}^\top\widehat{\boldsymbol{A}}\\
&+\frac{1}{np}\boldsymbol{E}\boldsymbol{F}\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}+\frac{1}{np}\boldsymbol{E}\boldsymbol{E}^\top\widehat{\boldsymbol{A}}\\
&+\frac{1}{np}\boldsymbol{A}^0\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}\\
\equiv&I1+\cdots+I9.
\end{aligned} \tag{S5.35}$$

The above can be rewritten as

$$\widehat{\boldsymbol{A}}\boldsymbol{V}_{np}-\boldsymbol{A}^0(\boldsymbol{F}^\top\boldsymbol{F}/n)(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)=I1+\cdots+I8. \tag{S5.36}$$

Right multiplying $(\boldsymbol{F}^\top \boldsymbol{F}/n)^{-1}(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)^{-1}$ on each side, we obtain

$$\widehat{\boldsymbol{A}}\left[\boldsymbol{V}_{np}(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)^{-1}(\boldsymbol{F}^\top \boldsymbol{F}/n)^{-1}\right]-\boldsymbol{A}^0 = (I1+\cdots+I8)(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)^{-1}(\boldsymbol{F}^\top \boldsymbol{F}/n)^{-1}. \tag{S5.37}$$

Note that the matrix in the square bracket is $\boldsymbol{H}^{-1}$, but the invertibility of $\boldsymbol{V}_{np}$ hasn't been proved yet. We can write

$$\frac{1}{\sqrt{p}}\left\|\widehat{\boldsymbol{A}}\left[\boldsymbol{V}_{np}(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)^{-1}(\boldsymbol{F}^\top \boldsymbol{F}/n)^{-1}\right]-\boldsymbol{A}^0\right\| \leq \frac{1}{\sqrt{p}}(\|I1\|+\cdots+\|I8\|)\|\boldsymbol{G}\|, \tag{S5.38}$$

where $\boldsymbol{G}$ is defined in (S5.13) and $\|\boldsymbol{G}\|$ is proved to be $O_p(1)$ in Lemma 5. In the following, we find the order for each term on the right-hand side of (S5.38). We repeatedly use results from Lemma 2, where the orders of the norms of matrices $\boldsymbol{\Phi}, \boldsymbol{A}$ and $\boldsymbol{F}$ are given. The first term

$$\begin{aligned}\frac{1}{\sqrt{p}}\|I1\| &\leq \frac{1}{\sqrt{p}}\frac{1}{np}\|\boldsymbol{\Phi}\|\|(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})^\top\|\|\boldsymbol{\Phi}^\top\|\|\widehat{\boldsymbol{A}}\| \\ &= O_p\left(\frac{1}{n}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|^2\right) = o_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right).\end{aligned}$$

For the second term

$$\begin{aligned}\frac{1}{\sqrt{p}}\|I2\| &\leq \frac{1}{\sqrt{p}}\frac{1}{np}\|\boldsymbol{\Phi}\|\|(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})\|\|\boldsymbol{F}\|\|\boldsymbol{A}^{0\top}\|\|\widehat{\boldsymbol{A}}\| \\ &= O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right).\end{aligned}$$

The terms $I3$ to $I5$ are all $O_p\left(\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|/\sqrt{n}\right)$. The proofs are similar to the proof for $I2$ since they are only a switch in the order of the matrices. For the sixth term

$$\frac{1}{\sqrt{p}}\|I6\| \leq \frac{1}{\sqrt{p}}\frac{1}{np}\|\boldsymbol{A}^0\|\|\boldsymbol{F}^\top\boldsymbol{E}^\top\|\|\widehat{\boldsymbol{A}}\| = O_p\left(\frac{1}{\sqrt{n}}\right),$$

by Lemma 3 (i). Similarly, for the next term

$$\frac{1}{\sqrt{p}}\|I7\| \leq \frac{1}{\sqrt{p}}\frac{1}{np}\|\boldsymbol{E}\boldsymbol{F}\|\|\boldsymbol{A}^{0\top}\|\|\widehat{\boldsymbol{A}}\| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

For the last term

$$\frac{1}{\sqrt{p}}\|I8\| \leq \frac{1}{\sqrt{p}}\frac{1}{np}\|\boldsymbol{E}\boldsymbol{E}^\top\|\|\widehat{\boldsymbol{A}}\| = O_p\left(\frac{1}{\sqrt{n}}\right)+\left(\frac{1}{\sqrt{p}}\right),$$

where Lemma 3 (iv) is used.

Putting all the terms above together, we have

$$\frac{1}{\sqrt{p}}\left\|\widehat{\boldsymbol{A}}\left[\boldsymbol{V}_{np}(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)^{-1}(\boldsymbol{F}^\top\boldsymbol{F}/n)^{-1}\right]-\boldsymbol{A}^0\right\| = O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n},\sqrt{p})}\right). \tag{S5.39}$$

To show (i), left multiply (S5.36) by $p^{-1}\widehat{\boldsymbol{A}}^\top$. Using $\widehat{\boldsymbol{A}}^\top\widehat{\boldsymbol{A}}/p=\boldsymbol{I}_r$, we have

$$\boldsymbol{V}_{np}-(\widehat{\boldsymbol{A}}^\top\boldsymbol{A}^0/p)(\boldsymbol{F}^\top\boldsymbol{F}/n)(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)=\frac{1}{p}\widehat{\boldsymbol{A}}^\top(I1+\cdots+I8)=o_p(1),$$

where the last equality is using Lemma 2 (v) and that $p^{-1/2}(\|I1\|+\cdots+\|I8\|)=o_p(1)$ from (S5.39). Thus,

$$\boldsymbol{V}_{np}=(\widehat{\boldsymbol{A}}^\top\boldsymbol{A}^0/p)(\boldsymbol{F}^\top\boldsymbol{F}/n)(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)+o_p(1).$$

We have shown in (S5.8) that $\widehat{\boldsymbol{A}}^\top\widehat{\boldsymbol{A}}^0$ is invertible, thus $\boldsymbol{V}_{np}$ is invertible. To obtain the limit of $\boldsymbol{V}_{np}$, left multiply (S5.36) by $p^{-1}\boldsymbol{A}^{0\top}$ to yield

$$(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)\boldsymbol{V}_{np}-(\boldsymbol{A}^{0\top}\boldsymbol{A}^0/p)(\boldsymbol{F}^\top\boldsymbol{F}/n)(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)=o_p(1),$$

or

$$(\boldsymbol{A}^{0\top}\boldsymbol{A}^0/p)(\boldsymbol{F}^\top\boldsymbol{F}/n)(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p) + o_p(1) = (\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)\boldsymbol{V}_{np} \tag{S5.40}$$

because $p^{-1}\boldsymbol{A}^{0\top}(\|I1\| + \ldots \|I8\|) = o_p(1)$. Equation (S5.40) shows that the columns of $(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p)$ are the eigenvectors of the matrix $(\boldsymbol{A}^{0\top}\boldsymbol{A}^0/p)(\boldsymbol{F}^\top\boldsymbol{F}/n)$, and that $\boldsymbol{V}_{np}$ consists of the eigenvalues of the same matrix. Thus, $\boldsymbol{V}_{np} \xrightarrow{p} \boldsymbol{V}$, where the $r \times r$ matrix $\boldsymbol{V}$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma_f}\boldsymbol{\Sigma_a}$.

For (ii), since $\boldsymbol{V}_{np}$ is invertible, $\boldsymbol{H}$ is also invertible we can write (S5.39) as

$$\frac{1}{\sqrt{p}}\left\|\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} - \boldsymbol{A}^0\right\| = O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n}, \sqrt{p})}\right).$$

By right multiplying the matrix $\boldsymbol{H}$, we obtain (ii).

□

S5.3 Appendix C

In this section, we state all the lemmas used for previous theorems and propositions, along with the proofs of the lemmas.

Let $\boldsymbol{\omega}_j$ denote the jth column of the $K \times p$ matrix $\boldsymbol{\Phi}^\top\boldsymbol{M}_{\boldsymbol{A}^0}$. We then

have the following lemma.

Lemma 1. *In addition to Assumptions 1 - 8, we suppose K is fixed. Then for any vector $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$, we have*

$$\frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{E} \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{b} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{L}), \quad as \quad n, p \to \infty,$$

where $\boldsymbol{L}$ is defined in (4.17).

This lemma paves the way for Theorem 3 on asymptotic normality.

Proof. For any vector $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$,

$$\frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{E} \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{b} = \frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \boldsymbol{\omega}_j \epsilon_{ji} b_i \equiv \frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \boldsymbol{x}_{ij},$$

where $\boldsymbol{\omega}_j$ is the jth column in the matrix $\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0}$. Since we assume ϵ_{ji} are i.i.d., the variance of the above quantity is given by

$$\operatorname{var}\left(\frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{E} \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{b}\right) = \operatorname{var}\left(\frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \boldsymbol{x}_{ij}\right) = \frac{1}{np} \sum_i^p \sum_j^n b_j b_i \sigma^2 E\left(\boldsymbol{\omega}_i \boldsymbol{\omega}_j^\top\right).$$

The Lindeberg condition is assumed to hold in Assumption 8. Thus we have a central limit theorem result

$$\frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{E} \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{b} = \frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \boldsymbol{x}_{ij} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{L}),$$

where $\boldsymbol{L}$ is defined in (4.17). □

Lemma 2. *Under Assumptions 1-3, we have*

(i) $\frac{1}{\sqrt{p}}\|\boldsymbol{\Phi}\| = O_p(1)$

(ii) $\frac{1}{\sqrt{p}}\|\boldsymbol{A}\| = O_p(1)$

(iii) $\frac{1}{\sqrt{n}}\|\boldsymbol{F}\| = O_p(1)$

(iv) $\frac{1}{\sqrt{np}}\|\boldsymbol{E}\| = O_p(1)$

(v) $\frac{1}{\sqrt{p}}\|\widehat{\boldsymbol{A}}\| = O_p(1)$

Proof. We have (i) from Assumption 1, (ii) and (iii) from Assumption 4, and (iv) from Assumption 5. Lastly, (v) is directly from the restriction $\widehat{\boldsymbol{A}}^\top\widehat{\boldsymbol{A}}/p = \boldsymbol{I}_r$. □

Lemma 3. *Under Assumptions 1 to 5, we have*

(i) $\frac{1}{np}\|\boldsymbol{E}\boldsymbol{F}\|^2 = O_p(1)$ *and* $\frac{1}{np}\|\boldsymbol{E}^\top\boldsymbol{A}^0\|^2 = O_p(1)$

(ii) $\frac{1}{npK}\|\boldsymbol{E}^\top\boldsymbol{\Phi}\|^2 = O_p(1)$

(iii) $\frac{1}{np}\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{A}^0\|^2 = O_p(1)$ *and* $\frac{1}{npK}\|\boldsymbol{\Phi}^\top\boldsymbol{E}^\top\boldsymbol{F}\|^2 = O_p(1)$

(iv) $\|\boldsymbol{E}^\top\boldsymbol{E}\|^2 = O_p(n^2p) + O_p(p^2n)$;

$\|\boldsymbol{E}\boldsymbol{E}^\top\|^2 = O_p(n^2p) + O_p(p^2n)$;

$\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{E}\|^2 = O_p(n^2p) + O_p(p^2n)$;

$$\|\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top\|^2 = O_p(n^2pK) + O_p(p^2nK);$$

$$\|\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top \boldsymbol{A}^0\|^2 = O_p(n^2pK) + O_p(p^2nK);$$

$$\|\boldsymbol{F}^\top \boldsymbol{E}^\top \boldsymbol{E}\boldsymbol{F}\|^2 = O_p(n^2p) + O_p(p^2n).$$

Proof. For (i),

$$\begin{aligned}\mathbb{E}\left(\frac{1}{np}\|\boldsymbol{E}\boldsymbol{F}\|^2\right) &\leq \mathbb{E}\left(\frac{1}{np}\sum_{k=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\epsilon_{ki}\epsilon_{kj}\boldsymbol{f}_i^\top\boldsymbol{f}_j\right)\\ &= \frac{1}{np}\sum_{k=1}^{p}\sum_{i=j}^{n}\mathbb{E}(\epsilon_{ki}\epsilon_{kj})\mathbb{E}(\boldsymbol{f}_i^\top\boldsymbol{f}_j) = O(1),\end{aligned}$$

where the second equation uses the independence between ϵ_{ki} and $\boldsymbol{f}_j$ assumed in Assumption 5. The proof for $\|\boldsymbol{E}^\top\boldsymbol{A}^0\|$ is similar.

For (ii),

$$\begin{aligned}\mathbb{E}\left(\frac{1}{npK}\|\boldsymbol{E}^\top\boldsymbol{\Phi}\|^2\right) &\leq \mathbb{E}\left(\frac{1}{npK}\sum_{k=1}^{n}\sum_{i=1}^{p}\sum_{j=1}^{p}\epsilon_{ki}\epsilon_{kj}\boldsymbol{\phi}_i^\top\boldsymbol{\phi}_j\right)\\ &\leq \frac{1}{npK}\sum_{k=1}^{n}\sum_{i=j}^{p}\mathbb{E}(\epsilon_{ki}\epsilon_{kj})\|\boldsymbol{\phi}_i^\top\boldsymbol{\phi}_j\| = O(1),\end{aligned}$$

where we use the $K \times 1$ vector norm $\|\boldsymbol{\phi}_i\| = O(\sqrt{K})$.

The proof of (iii) is similar to (i) and (ii). For (iv),

$$\begin{aligned}
\mathbb{E}\left(\|\boldsymbol{E}^\top\boldsymbol{E}\|^2\right) &\leq \mathbb{E}\left(\sum_{ij}^{n}\sum_{kl}^{p}\epsilon_{kj}\epsilon_{lj}\epsilon_{ki}\epsilon_{li}\right) \\
&= \sum_{i\neq j}^{n}\sum_{k=l}^{p}\mathbb{E}(\epsilon_{kj}^2)\mathbb{E}(\epsilon_{ki}^2) + \sum_{i=j}^{n}\sum_{k\neq l}^{p}\mathbb{E}(\epsilon_{kj}^2)\mathbb{E}(\epsilon_{lj}^2) + \sum_{i=j}^{n}\sum_{k=l}^{p}\mathbb{E}(\epsilon_{kj}^4) \\
&= O(n^2p) + O(p^2n) + O(np) \\
&= O(n^2p) + O(p^2n),
\end{aligned}$$

where Assumption 4 is used. The proof of $\|\boldsymbol{E}\boldsymbol{E}^\top\|$ is the same. The orders of $\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{E}\|$ and $\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{E}\boldsymbol{F}\|^2$ are the same since

$$\mathbb{E}\left(\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{E}\|^2\right) \leq \mathbb{E}\left(\sum_{ij}^{n}\sum_{kl}^{p}\epsilon_{kj}\epsilon_{lj}\epsilon_{ki}\epsilon_{li}\|\boldsymbol{f}_i\|^2\right),$$

and

$$\mathbb{E}\left(\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{E}\boldsymbol{F}\|^2\right) \leq \mathbb{E}\left(\sum_{ij}^{n}\sum_{kl}^{p}\epsilon_{kj}\epsilon_{lj}\epsilon_{ki}\epsilon_{li}\|\boldsymbol{f}_i\|^4\right),$$

where the order of $\boldsymbol{f}_i$ is assumed to be $O_p(1)$ in Assumption 3.

□

Lemma 4. *Under Assumptions 1-6,*

(i) $\frac{1}{np}\left\|\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top\boldsymbol{M}_{\boldsymbol{A}}\boldsymbol{A}^0\boldsymbol{F}_i\right\| = o_p(1)$

(ii) $\frac{1}{np}\left\|\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top\boldsymbol{M}_{\boldsymbol{A}}\boldsymbol{\Phi}\boldsymbol{c}_i\right\| = o_p(1)$

(iii) $\frac{1}{np}\left\|\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top(\boldsymbol{M}_{\boldsymbol{A}}-\boldsymbol{M}_{\boldsymbol{A}^0})\boldsymbol{\epsilon}_i\right\| = o_p(1)$

(iv) $\frac{\alpha}{np}\left\|\sum_{i=1}^{n}\boldsymbol{c}_i^\top\boldsymbol{R}\boldsymbol{c}_i\right\| = o_p(1)$

Proof. We prove (ii). First, we have

$$\mathbb{E}\left(\left\|\sum_{i=1}^{n}\boldsymbol{\epsilon}_i\right\|^2\right) = \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{p}\epsilon_{ik}\epsilon_{jk}\right) = \sum_{i=j}^{n}\sum_{k=1}^{p}\mathbb{E}\left(\epsilon_{ik}^2\right) = O(np).$$

Since $\boldsymbol{M}_{\boldsymbol{A}} = \boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}^\top/p$, we have

$$\frac{1}{np}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top\boldsymbol{M}_{\boldsymbol{A}}\boldsymbol{\Phi}\boldsymbol{c}_i = \frac{1}{np}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top\boldsymbol{\Phi}\boldsymbol{c}_i - \frac{1}{np^2}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{\Phi}\boldsymbol{c}_i. \tag{S5.41}$$

The first term on the right of (S5.41) is $o_p(1)$ since

$$\begin{aligned}
\mathbb{E}\left(\left\|\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top\boldsymbol{\Phi}\boldsymbol{c}_i\right\|^2\right) &= \mathbb{E}\left(\left\|\sum_{j=1}^{p}\sum_{i=1}^{n}\epsilon_{ji}\boldsymbol{\phi}_j^\top\boldsymbol{c}_i\right\|^2\right)\\
&= \mathbb{E}\left(\sum_{t}^{p}\sum_{s}^{n}\sum_{j}^{p}\sum_{i}^{n}\epsilon_{ji}\epsilon_{ts}\boldsymbol{c}_i^\top\boldsymbol{\phi}_j\boldsymbol{\phi}_t^\top\boldsymbol{c}_s\right)\\
&= \sum_{t}^{p}\sum_{s}^{n}\sum_{j}^{p}\sum_{i}^{n}\mathbb{E}(\epsilon_{ji}\epsilon_{ts})\mathbb{E}(\boldsymbol{c}_i^\top\boldsymbol{\phi}_j\boldsymbol{\phi}_t^\top\boldsymbol{c}_s)\\
&= \sum_{j}^{p}\sum_{i}^{n}\sigma^2\mathbb{E}(\boldsymbol{c}_i^\top\boldsymbol{\phi}_j\boldsymbol{\phi}_j^\top\boldsymbol{c}_i)\\
&= O(npK),
\end{aligned}$$

where the third equation uses Assumption 5; the fourth equation uses the

assumption that ϵ_{ji} are independent in both directions.

The second term on the right-hand side of (S5.41) is also $o_p(1)$ since

$$\begin{aligned}
\mathbb{E}\left(\left\|\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top \boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{\Phi}\boldsymbol{c}_i\right\|^2\right) &= \mathbb{E}\left(\left\|\sum_{j=1}^{p}\sum_{i=1}^{n}\epsilon_{ji}\boldsymbol{a}_j^\top \boldsymbol{A}^\top\boldsymbol{\Phi}\boldsymbol{c}_i\right\|^2\right)\\
&= \mathbb{E}\left(\sum_{t}^{p}\sum_{s}^{n}\sum_{j}^{p}\sum_{i}^{n}\epsilon_{ji}\epsilon_{ts}\boldsymbol{c}_i^\top\boldsymbol{\Phi}^\top\boldsymbol{A}\boldsymbol{a}_j\boldsymbol{a}_t^\top\boldsymbol{A}^\top\boldsymbol{\Phi}\boldsymbol{c}_s\right)\\
&= \sum_{t}^{p}\sum_{s}^{n}\sum_{j}^{p}\sum_{i}^{n}\mathbb{E}(\epsilon_{ji}\epsilon_{ts})\mathbb{E}(\boldsymbol{c}_i^\top\boldsymbol{\Phi}^\top\boldsymbol{A}\boldsymbol{a}_j\boldsymbol{a}_t^\top\boldsymbol{A}^\top\boldsymbol{\Phi}\boldsymbol{c}_s)\\
&= \sum_{j}^{p}\sum_{i}^{n}\sigma^2\mathbb{E}(\boldsymbol{c}_i^\top\boldsymbol{\Phi}^\top\boldsymbol{A}\boldsymbol{a}_j\boldsymbol{a}_j^\top\boldsymbol{A}^\top\boldsymbol{\Phi}\boldsymbol{c}_i)\\
&= O(np^3),
\end{aligned}$$

where the third equality uses the independence in Assumption 5 and the last equality uses the results in Lemma 2, where $\boldsymbol{\Phi}$ and $\boldsymbol{A}$ are both $O_p(\sqrt{p})$.

With $K = o(p)$ in Assumption 1, we have $\frac{1}{np}\left\|\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top\boldsymbol{M}_{\boldsymbol{A}}\boldsymbol{\Phi}\boldsymbol{c}_i\right\| = o_p(1)$.

The proofs for (i) and (iii) are similar. And (iv) is a direct result from Assumption 6. □

Lemma 5. *Under Assumptions 1-5 , we have*

$$\boldsymbol{G} \equiv \left(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p\right)^{-1}\left(\boldsymbol{F}^\top\boldsymbol{F}/n\right)^{-1} = O_p(1).$$

Proof. The matrix $\boldsymbol{F}^\top \boldsymbol{F}/n$ is asymptotically positive definite by Assumption 4. We have shown in the proof of Theorem 1 in (S5.8) that the matrix $\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p$ is invertible, thus is also positive definite. Therefore, $\lambda_{\min}\left(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p\right) > 0$, and $\lambda_{\min}\left(\boldsymbol{F}^\top \boldsymbol{F}/n\right) > 0$, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix. So we have

$$\left(\boldsymbol{A}^{0\top}\widehat{\boldsymbol{A}}/p\right)^{-1} = O_p(1), \quad \left(\boldsymbol{F}^\top \boldsymbol{F}/n\right)^{-1} = O_p(1).$$

□

Lemma 6. *We have the following*

(i)

$$\left\|\boldsymbol{E}^\top(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})\right\| = O_p\left(\frac{p}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p(\sqrt{n}) + O_p\left(\frac{p}{\sqrt{n}}\right).$$

(ii)

$$\left\|\boldsymbol{F}^\top\boldsymbol{E}^\top(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})\right\| = O_p\left(\frac{p}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p(\sqrt{n}) + O_p\left(\frac{p}{\sqrt{n}}\right).$$

Proof. For (i), from Proposition 1, we can write

$$\begin{aligned}\|\boldsymbol{E}^\top(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})\| &= \|\boldsymbol{E}^\top(I1 + \ldots, I8)\boldsymbol{G}\| \\ &\leq \|\boldsymbol{E}^\top I1\boldsymbol{G}\| + \cdots + \|\boldsymbol{E}^\top I8\boldsymbol{G}\| \\ &= \|a1\| + \cdots + \|a8\|.\end{aligned}$$

To find the order for each term, the results from Lemma 2 are repeatedly used where the order of the matrices $\boldsymbol{\Phi}, \boldsymbol{A}, \widehat{\boldsymbol{A}}$ and $\boldsymbol{F}$ are given.

$$\begin{aligned}\|a1\| &= \left\|\boldsymbol{E}^\top \frac{1}{np}\boldsymbol{\Phi}(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})^\top\boldsymbol{\Phi}^\top\widehat{\boldsymbol{A}}\boldsymbol{G}\right\| \\ &\leq \frac{1}{np}\|\boldsymbol{E}^\top\boldsymbol{\Phi}\| \left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|^2 \|\boldsymbol{\Phi}\|\|\widehat{\boldsymbol{A}}\|\|\boldsymbol{G}\| \\ &= O_p\left(\frac{\sqrt{pK}}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|^2\right) = o_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right),\end{aligned}$$

where the order of $\|\boldsymbol{E}^\top\boldsymbol{\Phi}\|$ is from Lemma 3 (ii). The orders of $\|\boldsymbol{\Phi}\|$, $\|\widehat{\boldsymbol{A}}\|$ and $\|\boldsymbol{G}\|$can be found from Lemmas 2 and 5. Similarly,

$$\begin{aligned}\|a2\| &= \left\|\boldsymbol{E}^\top \frac{1}{n}\boldsymbol{\Phi}(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})\boldsymbol{F}\left(\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right)^{-1}\right\| \\ &\leq \frac{1}{n}\|\boldsymbol{E}^\top\boldsymbol{\Phi}\| \left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\| \|\boldsymbol{F}\| \left\|\left(\frac{\boldsymbol{F}^\top\boldsymbol{F}}{n}\right)^{-1}\right\| \\ &= O_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right).\end{aligned}$$

$$
\begin{aligned}
\|a3\| &= \left\| \boldsymbol{E}^\top \frac{1}{np} \boldsymbol{\Phi}(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}) \boldsymbol{E}^\top \widehat{\boldsymbol{A}} \boldsymbol{G} \right\| \\
&\leq \frac{1}{np} \|\boldsymbol{E}^\top \boldsymbol{\Phi}\| \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \|\boldsymbol{E}^\top\| \|\widehat{\boldsymbol{A}}\| \|\boldsymbol{G}\| \\
&= O_p \left(\sqrt{pK} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right).
\end{aligned}
$$

$$
\begin{aligned}
\|a4\| &= \left\| \boldsymbol{E}^\top \frac{1}{np} \boldsymbol{A}^0 \boldsymbol{F}^\top (\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})^\top \boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}} \boldsymbol{G} \right\| \\
&\leq \frac{1}{np} \|\boldsymbol{E}^\top \boldsymbol{A}^0\| \|\boldsymbol{F}^\top\| \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \|\boldsymbol{\Phi}\| \|\widehat{\boldsymbol{A}}\| \|\boldsymbol{G}\| \\
&= O_p \left(\sqrt{p} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right),
\end{aligned}
$$

where Lemma 3 (ii) is used.

$$
\begin{aligned}
\|a5\| &= \|\boldsymbol{E}^\top \frac{1}{np} \boldsymbol{E} (\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})^\top \boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}} \boldsymbol{G}\| \\
&\leq \frac{1}{np} \|\boldsymbol{E}^\top \boldsymbol{E}\| \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \|\boldsymbol{\Phi}\| \|\widehat{\boldsymbol{A}}\| \|\boldsymbol{G}\| \\
&= O_p \left(\sqrt{p} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) + O_p \left(\frac{p}{\sqrt{n}} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right),
\end{aligned}
$$

where Lemma 3 (iv) is used.

$$
\begin{aligned}
\|a6\| &= \|\boldsymbol{E}^\top \frac{1}{np} \boldsymbol{A}^0 \boldsymbol{F}^\top \boldsymbol{E}^\top \widehat{\boldsymbol{A}} \boldsymbol{G}\| \\
&\leq \frac{1}{np} \|\boldsymbol{E}^\top \boldsymbol{A}^0 \boldsymbol{F}^\top \boldsymbol{E}^\top (\widehat{\boldsymbol{A}} - \boldsymbol{A}^0 \boldsymbol{H}) \boldsymbol{G}\| + \frac{1}{np} \|\boldsymbol{E}^\top \boldsymbol{A}^0 \boldsymbol{F}^\top \boldsymbol{E}^\top \boldsymbol{A}^0 \boldsymbol{H} \boldsymbol{G}\| \\
&\leq \frac{1}{np} \|\boldsymbol{E}^\top \boldsymbol{A}^0\| \|\boldsymbol{F}^\top \boldsymbol{E}^\top\| \|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0 \boldsymbol{H}\| \|\boldsymbol{G}\| + \frac{1}{np} \|\boldsymbol{E}^\top \boldsymbol{A}^0\| \|\boldsymbol{F}^\top \boldsymbol{E}^\top \boldsymbol{A}^0\| \|\boldsymbol{H} \boldsymbol{G}\| \\
&= O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{p}}{\min(\sqrt{n}, \sqrt{p})}\right) + O_p(1) \\
&= O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{p}}{\min(\sqrt{n}, \sqrt{p})}\right),
\end{aligned}
$$

where the order of $\|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0 \boldsymbol{H}\|$ is proved in Proposition 1 and the order of other matrix norms can be found in Lemma 3 $(i), (ii)$ and (iii).

$$
\begin{aligned}
\|a7\| &= \|\boldsymbol{E}^\top \frac{1}{np} \boldsymbol{E} \boldsymbol{F} \boldsymbol{A}^{0\top} \widehat{\boldsymbol{A}} \boldsymbol{G}\| \\
&= \left\| \frac{1}{n} \boldsymbol{E}^\top \boldsymbol{E} \boldsymbol{F} \left(\frac{\boldsymbol{F}^\top \boldsymbol{F}}{n}\right)^{-1} \right\| \\
&\leq \frac{1}{n} \left\|\boldsymbol{E}^\top \boldsymbol{E} \boldsymbol{F}\right\| \left\| \left(\frac{\boldsymbol{F}^\top \boldsymbol{F}}{n}\right)^{-1} \right\| \\
&= O_p\left(\sqrt{p}\right) + O_p\left(\frac{p}{\sqrt{n}}\right),
\end{aligned}
$$

where Lemma 3 (iv) is used.

$$
\begin{aligned}
\|a8\| =& \frac{1}{np}\left\|\boldsymbol{E}^\top\boldsymbol{E}\boldsymbol{E}^\top\widehat{\boldsymbol{A}}\boldsymbol{G}\right\| \\
\leq& \frac{1}{np}\left\|\boldsymbol{E}^\top\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{A}^0\boldsymbol{H}\boldsymbol{G}\right\| + \frac{1}{np}\left\|\boldsymbol{E}^\top\boldsymbol{E}\boldsymbol{E}^\top(\widehat{\boldsymbol{A}}-\boldsymbol{A}^0\boldsymbol{H})\boldsymbol{G}\right\| \\
\leq& \frac{1}{np}\|\boldsymbol{E}^\top\boldsymbol{E}\|\|\boldsymbol{E}^\top\boldsymbol{A}^0\|\|\boldsymbol{H}\|\|\boldsymbol{G}\| + \frac{1}{np}\|\boldsymbol{E}^\top\boldsymbol{E}\|\|\boldsymbol{E}^\top\|\|\widehat{\boldsymbol{A}}-\boldsymbol{A}^0\boldsymbol{H}\|\|\boldsymbol{G}\| \\
=& \frac{1}{np}\left[O_p(n\sqrt{p})+O_p(p\sqrt{n})\right]O_p(\sqrt{np}) \\
&+\frac{1}{np}\left[O_p(n\sqrt{p})+O_p(p\sqrt{n})\right]O_p(\sqrt{np})\left[O_p\left(\frac{\sqrt{p}}{\sqrt{n}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)+O_p\left(\frac{\sqrt{p}}{\min(\sqrt{n},\sqrt{p})}\right)\right] \\
=& O_p(\sqrt{n})+O_p(\sqrt{p})+O_p\left(\frac{p}{\sqrt{n}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)+O_p\left(\sqrt{p}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)+O_p\left(\sqrt{n}\right)+O_p\left(\frac{p}{\sqrt{n}}\right) \\
=& O_p\left(\frac{p}{\sqrt{n}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)+O_p\left(\sqrt{p}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)+O_p\left(\sqrt{n}\right)+O_p\left(\frac{p}{\sqrt{n}}\right),
\end{aligned}
$$

where the order of $\widehat{\boldsymbol{A}}-\boldsymbol{A}^0\boldsymbol{H}$ is proved in Proposition 1 and the order of other matrix norms can be found in Lemma 3 $(i),(ii)$ and (iii).

Combining all the terms, we have

$$
\|\boldsymbol{E}^\top(\widehat{\boldsymbol{A}}-\boldsymbol{A}^0\boldsymbol{H})\| = O_p\left(\frac{p}{\sqrt{n}}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)+O_p\left(\sqrt{pK}\left\|\boldsymbol{C}^0-\widehat{\boldsymbol{C}}\right\|\right)+O_p(\sqrt{n})+O_p\left(\frac{p}{\sqrt{n}}\right).
$$

For (ii), multiplying the matrix $\boldsymbol{F}^\top$ in the front does not change the order, using the fact that $\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{\Phi}\|$ is of the same order as $\|\boldsymbol{E}^\top\boldsymbol{\Phi}\|$ and that $\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{E}\|$ and $\|\boldsymbol{F}^\top\boldsymbol{E}^\top\boldsymbol{E}\boldsymbol{F}\|$ are of the same order as $\|\boldsymbol{E}^\top\boldsymbol{E}\|$, as proved in Lemma 3. $\square$

Lemma 7. *Define the matrix*

$$\boldsymbol{Q}(\boldsymbol{A}) = \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}}\boldsymbol{\Phi}.$$

Under Assumptions 1-4, it holds

$$\left\|\boldsymbol{Q}\left(\widehat{\boldsymbol{A}}\right)^{-1} - \boldsymbol{Q}\left(\boldsymbol{A}^0\right)^{-1}\right\| = o_p(1).$$

Proof. We have

$$\begin{aligned}\left\|\boldsymbol{Q}\left(\widehat{\boldsymbol{A}}\right) - \boldsymbol{Q}\left(\boldsymbol{A}^0\right)\right\| &= \left\|\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} - \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{\Phi}\right\| = \left\|\frac{1}{p}\boldsymbol{\Phi}^\top\left(\boldsymbol{M}_{\widehat{\boldsymbol{A}}} - \boldsymbol{M}_{\boldsymbol{A}^0}\right)\boldsymbol{\Phi}\right\| \\ &= \left\|\frac{1}{p}\boldsymbol{\Phi}^\top\left(\boldsymbol{P}_{\boldsymbol{A}^0} - \boldsymbol{P}_{\widehat{\boldsymbol{A}}}\right)\boldsymbol{\Phi}\right\| = O_p\left(\left\|\boldsymbol{P}_{\boldsymbol{A}^0} - \boldsymbol{P}_{\widehat{\boldsymbol{A}}}\right\|\right) = o_p(1),\end{aligned}$$

using Theorem 1 (ii). In Assumption 3, we have assumed $\inf_{\boldsymbol{A}} \boldsymbol{D}(\boldsymbol{A}) > 0$, since the second term in $\boldsymbol{D}(\boldsymbol{A})$ is nonnegative, we have $\inf_{\boldsymbol{A}} \boldsymbol{Q}(\boldsymbol{A}) > 0$, so the matrix $\boldsymbol{Q}(\boldsymbol{A}^0)$ is invertible and its inverse is bounded under the spectral norm. Therefore,

$$\left\|\boldsymbol{Q}\left(\widehat{\boldsymbol{A}}\right)^{-1} - \boldsymbol{Q}\left(\boldsymbol{A}^0\right)^{-1}\right\| = \left\|\boldsymbol{Q}\left(\widehat{\boldsymbol{A}}\right)^{-1}\left[\boldsymbol{Q}\left(\boldsymbol{A}^0\right) - \boldsymbol{Q}\left(\widehat{\boldsymbol{A}}\right)\right]\boldsymbol{Q}\left(\boldsymbol{A}^0\right)^{-1}\right\| = o_p(1).$$

□

Lemma 8. *Recall $\boldsymbol{H}$ defined in (S5.11), then*

$$\boldsymbol{H}\boldsymbol{H}^\top = \left(\frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p}\right)^{-1} + O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n},\sqrt{p})}\right)$$

Proof. We have

$$\frac{1}{p}\boldsymbol{A}^{0\top}(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}) = O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n},\sqrt{p})}\right), \tag{S5.42}$$

and

$$\frac{1}{p}\widehat{\boldsymbol{A}}^\top(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}) = \boldsymbol{I}_r - \frac{1}{p}\widehat{\boldsymbol{A}}^\top\boldsymbol{A}^0\boldsymbol{H} = O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n},\sqrt{p})}\right). \tag{S5.43}$$

Left multiply (S5.42) by $\boldsymbol{H}^\top$ and sum with the transpose of (S5.43) to obtain

$$\boldsymbol{I}_r - \frac{1}{p}\boldsymbol{H}^\top\boldsymbol{A}^{0\top}\boldsymbol{A}^0\boldsymbol{H} = O_p\left(\frac{1}{n}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n},\sqrt{p})}\right).$$

Right multiplying by $\boldsymbol{H}^\top$ and left multiplying by $\boldsymbol{H}^{\top-1}$, we obtain

$$\boldsymbol{I}_r - \frac{1}{p}\boldsymbol{A}^{0\top}\boldsymbol{A}^0\boldsymbol{H}\boldsymbol{H}^\top = O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n},\sqrt{p})}\right).$$

Then left multiplying $\left(\boldsymbol{A}^{0\top}\boldsymbol{A}^0/p\right)^{-1}$, we have

$$\boldsymbol{H}\boldsymbol{H}^\top = \left(\frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p}\right)^{-1} + O_p\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(\sqrt{n},\sqrt{p})}\right).$$

□

Lemma 9. *Under Assumptions 1-5, when $p/n \to \rho > 0$,*

$$\begin{aligned}\left\|\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E} - \frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{E}\right\| =& O_p\left(\sqrt{p}\frac{\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|^2}{n}\right) + O_p\left(\frac{\sqrt{p}}{\sqrt{n}}\frac{\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|}{\sqrt{n}}\right)\\ &+ O_p\left(\sqrt{K}\frac{\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|}{\sqrt{n}}\right) + O_p\left(\frac{\sqrt{p}}{\min(n,p)}\right) + o_p(1).\end{aligned}$$

Proof. Using

$$\boldsymbol{M}_{\boldsymbol{A}^0} = \boldsymbol{I}_p - \boldsymbol{A}^0\left(\boldsymbol{A}^{0\top}\boldsymbol{A}^0\right)^{-1}\boldsymbol{A}^{0\top}, \qquad \boldsymbol{M}_{\widehat{\boldsymbol{A}}} = \boldsymbol{I}_p - \left(\widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top\right)/p,$$

we calculate

$$
\begin{aligned}
\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{E} - \frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E} =& \frac{1}{p\sqrt{np}}\boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top \boldsymbol{E} - \frac{1}{p\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{A}^0 \left(\frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p}\right)^{-1} \boldsymbol{A}^{0\top}\boldsymbol{E} \\
=& \frac{1}{p\sqrt{np}}\Bigg\{\boldsymbol{\Phi}^\top(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})\boldsymbol{H}^\top \boldsymbol{A}^{0\top}\boldsymbol{E} \\
&+ \boldsymbol{\Phi}^\top(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})^\top \boldsymbol{E} \\
&+ \boldsymbol{\Phi}^\top \boldsymbol{A}^0\boldsymbol{H}(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})^\top \boldsymbol{E} \\
&+ \boldsymbol{\Phi}^\top \boldsymbol{A}^0 \left[\boldsymbol{H}\boldsymbol{H}^\top - \left(\frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p}\right)^{-1}\right] \boldsymbol{A}^{0\top}\boldsymbol{E}\Bigg\} \\
\equiv& a + b + c + d,
\end{aligned}
$$

where we substitute $\widehat{\boldsymbol{A}}$ with $\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H} + \boldsymbol{A}^0\boldsymbol{H}$ in the second equality. So the first term on the right-hand side of the first equality is broken down into four terms, one of which is combined with the second term in the right-hand side of the first equality.

We calculate each term:

$$
\begin{aligned}
\|a\| &= \left\|\frac{1}{p\sqrt{np}}\boldsymbol{\Phi}^\top(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})\boldsymbol{H}^\top \boldsymbol{A}^\top \boldsymbol{E}\right\| \\
&= \frac{1}{p\sqrt{np}} \times O_p(\sqrt{p}) \times \left[O_p\left(\frac{\sqrt{p}}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{p}}{\min(\sqrt{n}, \sqrt{p})}\right)\right] \times O_p(\sqrt{np}) \\
&= O_p\left(\frac{\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|}{\sqrt{n}}\right) + \left(\frac{1}{\min(\sqrt{n}, \sqrt{p})}\right) = o_p(1),
\end{aligned}
$$

where the order of $\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}$ is $\sqrt{p}q$ as proved in Proposition 1 and the order of $\|\boldsymbol{\Phi}\|\ \|\boldsymbol{A}^\top\boldsymbol{E}\|$ can be found in Lemma 2 and 3 (ii) respectively. And

$$
\begin{aligned}
\|b\| &= \left\| \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top (\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})^\top \boldsymbol{E} \right\| \\
&\leq \frac{1}{p\sqrt{np}} \times O_p(\sqrt{p}) \times \left[O_p\left(\frac{p \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|^2}{n} \right) + O_p\left(\frac{p}{\min(n,p)} \right) \right] \times O_p(\sqrt{np}) \\
&= \sqrt{p} \times O_p\left(\frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|^2}{n} \right) + O_p\left(\frac{\sqrt{p}}{\min(n,p)} \right),
\end{aligned}
$$

where again Proposition 1 and Lemma 2 are used. And

$$
\begin{aligned}
c =& \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{A}^0\boldsymbol{H}(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H})^\top \boldsymbol{E} \\
=& \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{A}^0\boldsymbol{H}\boldsymbol{H}^\top(\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} - \boldsymbol{A}^0)^\top \boldsymbol{E} \\
=& \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{A}^0 \left[\boldsymbol{H}\boldsymbol{H}^\top - \left(\frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p} \right)^{-1} \right] (\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} - \boldsymbol{A}^0)^\top \boldsymbol{E} \\
&+ \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{A}^0 \left(\frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p} \right)^{-1} (\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} - \boldsymbol{A}^0)^\top \boldsymbol{E} \\
\equiv& c1 + c2,
\end{aligned}
$$

where the second equality is using $\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H} = (\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} - \boldsymbol{A}^0)\boldsymbol{H}$. In the third equality, we subtract $(\boldsymbol{A}^{0\top}\boldsymbol{A}^0/p)^{-1}$ from $\boldsymbol{H}\boldsymbol{H}^\top$ and then add it back.

For $c1$,

$$
\begin{aligned}
\|c1\| =& \left\| \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{A}^0 \left[\boldsymbol{H}\boldsymbol{H}^\top - \left(\frac{\boldsymbol{A}^{0\top}\boldsymbol{A}^0}{p} \right)^{-1} \right] (\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} - \boldsymbol{A}^0)^\top \boldsymbol{E} \right\| \\
\leq& \frac{1}{p\sqrt{np}} \times O_p(\sqrt{p}) \times O_p(\sqrt{p}) \times \left[O_p\left(\frac{1}{\sqrt{n}} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) + O_p\left(\frac{1}{\min(\sqrt{n}, \sqrt{p})} \right) \right] \\
& \times \left[O_p\left(\frac{p}{\sqrt{n}} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) + O_p\left(\sqrt{pK} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) + O_p(\sqrt{n}) + O_p\left(\frac{p}{\sqrt{n}} \right) \right] \\
=& O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|^2}{n} \right) + O_p\left(\sqrt{K} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|^2}{n} \right) O_p\left(\frac{1}{\min(\sqrt{n}, \sqrt{p})} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{n} \right) \\
& + O_p\left(\frac{\sqrt{p}}{n} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{\sqrt{n}} \right) + O_p\left(\frac{\sqrt{K}}{\min(\sqrt{n}, \sqrt{p})} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{n} \right) \\
& + O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \frac{1}{\min(n, p)} \right) + O_p\left(\frac{1}{\min(n, p)} \right) \\
=& O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|^2}{n} \right) + O_p\left(\sqrt{K} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|^2}{n} \right) + O_p\left(\frac{\sqrt{p}}{n} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{\sqrt{n}} \right) \\
& + O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \frac{1}{\min(n, p)} \right) + o_p(1),
\end{aligned}
$$

where the order of $\boldsymbol{\Phi}$, $\boldsymbol{A}^0$ and $\boldsymbol{E}$ are found in Lemma 2; the order of $\boldsymbol{H}\boldsymbol{H}^\top - \left(\boldsymbol{A}^{0\top}\boldsymbol{A}^0/p \right)^{-1}$ is found in Lemma 8; and the order of $\widehat{\boldsymbol{A}}\boldsymbol{H}^{-1} - \boldsymbol{A}^0$ is

found in Proposition 1. Now for $c2$, using the same lemmas and proposition,

$$\begin{aligned}
\|c2\| &= \left\| \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{A}^0 \left(\frac{\boldsymbol{A}^{0\top} \boldsymbol{A}^0}{p} \right)^{-1} (\widehat{\boldsymbol{A}} \boldsymbol{H}^{-1} - \boldsymbol{A}^0)^\top \boldsymbol{E} \right\| \\
&\leq \frac{1}{p\sqrt{np}} O_p(\sqrt{p}) \times O_p(\sqrt{p}) \times \left[O_p\left(\frac{p}{\sqrt{n}} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) + O_p\left(\sqrt{pK} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) \right. \\
&\left. + O_p(\sqrt{n}) + O_p\left(\frac{p}{\sqrt{n}} \right) \right] \\
&= O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{\sqrt{n}} \right) + O_p\left(\sqrt{K} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{\sqrt{n}} \right) + O_p\left(\frac{\sqrt{p}}{\min(n,p)} \right).
\end{aligned}$$

And lastly, we have

$$\begin{aligned}
\|d\| &= \left\| \frac{1}{p\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{A}^0 \left[\boldsymbol{H}\boldsymbol{H}^\top - \left(\frac{\boldsymbol{A}^{0\top} \boldsymbol{A}^0}{p} \right)^{-1} \right] \boldsymbol{A}^{0\top} \boldsymbol{E} \right\| \\
&\leq \frac{1}{p\sqrt{np}} O_p(\sqrt{p}) \times O_p(\sqrt{p}) \times \left[O_p\left(\frac{1}{\sqrt{n}} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) + O_p\left(\frac{1}{\min(\sqrt{n}, \sqrt{p})} \right) \right] \times O_p(\sqrt{np}) \\
&= O_p\left(\frac{1}{\sqrt{n}} \left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\| \right) + O_p\left(\frac{1}{\min(\sqrt{n}, \sqrt{p})} \right) = o_p(1),
\end{aligned}$$

where again Lemma 8 is used.

Thus combining the above terms and using $K = o(\min(n,p)$, we have

$$\begin{aligned}
\left\| \frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}} \boldsymbol{E} - \frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{E} \right\| =& O_p\left(\sqrt{p} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|^2}{n} \right) + O_p\left(\frac{\sqrt{p}}{\sqrt{n}} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{\sqrt{n}} \right) \\
&+ O_p\left(\sqrt{K} \frac{\left\| \boldsymbol{C}^0 - \widehat{\boldsymbol{C}} \right\|}{\sqrt{n}} \right) + O_p\left(\frac{\sqrt{p}}{\min(n,p)} \right) + o_p(1).
\end{aligned}$$

□

Lemma 10. *Recall J8 defined in* (S5.12)*, we have*

$$\|J8\| = o_p\left(\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{\min(n,p)}\right) + O_p\left(\frac{\sqrt{n}}{\sqrt{p}}\frac{1}{\min(n,p)}\right).$$

Proof.

$$\begin{aligned}
J8 &= -\frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}(I8)\boldsymbol{G}\boldsymbol{F}^\top \\
&= -\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{E}\boldsymbol{E}^\top \widehat{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{F}^\top \\
&= -\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top \widehat{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{F}^\top + \frac{1}{np^3}\boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top \boldsymbol{E}\boldsymbol{E}^\top \widehat{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{F}^\top \\
&\equiv I + II,
\end{aligned}$$

where we use $\boldsymbol{M}_{\widehat{\boldsymbol{A}}} = \boldsymbol{I}_p - \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top/p$. For I,

$$I = -\frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top\left(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\right)\boldsymbol{G}\boldsymbol{F}^\top - \frac{1}{np^2}\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top \boldsymbol{A}^0\boldsymbol{H}\boldsymbol{G}\boldsymbol{F}^\top,$$

then

$$
\begin{aligned}
\|I\| \leq & \frac{1}{np^2}\left\|\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top\right\| \|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\|\|\boldsymbol{G}\|\|\boldsymbol{F}\| + \frac{1}{np^2}\left\|\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top \boldsymbol{A}^0\right\| \|\boldsymbol{H}\|\|\boldsymbol{G}\|\|\boldsymbol{F}\| \\
= & \frac{1}{np^2} \times \left[O_p\left(p\sqrt{nK}\right) + O_p\left(n\sqrt{pK}\right)\right] \times \left[O_p\left(\frac{\sqrt{p}}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{p}}{\min(\sqrt{n},\sqrt{p})}\right)\right] \\
& \times O_p\left(\sqrt{n}\right) + \frac{1}{np^2} \times \left[O_p\left(p\sqrt{nK}\right) + O_p\left(n\sqrt{pK}\right)\right] \times O_p\left(\sqrt{n}\right) \\
= & O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{K}}{p}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\right) + O_p\left(\frac{\sqrt{nK}}{p\sqrt{p}}\right)
\end{aligned}
$$

where the order of $\left\|\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top\right\|$ and $\left\|\boldsymbol{\Phi}^\top \boldsymbol{E}\boldsymbol{E}^\top \boldsymbol{A}^0\right\|$ are found in Lemma 3; the order of $\|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\|$ is from Proposition 1; and the orders of $\|\boldsymbol{F}\|$ and $\|\boldsymbol{G}\|$ are found in Lemma 2 (iii) and Lemma 5 respectively.

For II,

$$
\begin{aligned}
II = & \frac{1}{np^3}\boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}}\left(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H} + \boldsymbol{A}^0\boldsymbol{H}\right)^\top \boldsymbol{E}\boldsymbol{E}^\top\left(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H} + \boldsymbol{A}^0\boldsymbol{H}\right)\boldsymbol{G}\boldsymbol{F}^\top \\
= & \frac{1}{np^3}\boldsymbol{\Phi}^\top \widehat{\boldsymbol{A}}\left[\left(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\right)^\top \boldsymbol{E}\boldsymbol{E}^\top\left(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\right) + \left(\boldsymbol{A}^0\boldsymbol{H}\right)^\top \boldsymbol{E}\boldsymbol{E}^\top\left(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\right)\right. \\
& \left. + \left(\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\right)^\top \boldsymbol{E}\boldsymbol{E}^\top \boldsymbol{A}^0\boldsymbol{H} + \left(\boldsymbol{A}^0\boldsymbol{H}\right)^\top \boldsymbol{E}\boldsymbol{E}^\top \boldsymbol{A}^0\boldsymbol{H}\right]\boldsymbol{G}\boldsymbol{F}^\top,
\end{aligned}
$$

then

$$
\begin{aligned}
\|II\| \leq & \frac{1}{np^3}\|\boldsymbol{\Phi}^\top\|\|\widehat{\boldsymbol{A}}\| \left[\|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\|^2\|\boldsymbol{E}\boldsymbol{E}^\top\| + \|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\|\|\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{A}^0\| + \|\boldsymbol{A}^{0\top}\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{A}^0\|\right]\|\boldsymbol{G}\|\|\boldsymbol{F}\| \\
= & \frac{1}{np^2}\left[O_p\left(p\sqrt{n}\right) + O_p\left(n\sqrt{p}\right)\right] \times \left[O_p\left(\frac{p}{n}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|^2\right) + O_p\left(\frac{\sqrt{p}}{\sqrt{n}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right)\right. \\
& \left. + O_p\left(\frac{p}{\min(n,p)}\right)\right] \times O_p\left(\sqrt{n}\right) \\
= & O_p\left(\frac{1}{\sqrt{np}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|^2\right) + O_p\left(\frac{1}{n}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|^2\right) + O_p\left(\frac{1}{\sqrt{np}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) \\
& + O_p\left(\frac{1}{p}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{\sqrt{n}}{p\sqrt{p}}\right)
\end{aligned}
$$

where the order of $\left\|\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{A}^0\right\|$ and $\left\|\boldsymbol{A}^{0\top}\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{A}^0\right\|$ are found in Lemma 3; the order of $\|\widehat{\boldsymbol{A}} - \boldsymbol{A}^0\boldsymbol{H}\|$ is from Proposition 1; and the orders of $\|\boldsymbol{\Phi}\|$, $\|\boldsymbol{F}\|$ and $\|\boldsymbol{G}\|$ are found in Lemma 2 (i), (iii) and Lemma 5 respectively.

Combining I and II, we have

$$
\begin{aligned}
\|J8\| &= O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{K}}{p}\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\right) + O_p\left(\frac{\sqrt{nK}}{p\sqrt{p}}\right) + O_p\left(\frac{1}{n}\right) \\
&= o_p\left(\left\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\right\|\right) + O_p\left(\frac{\sqrt{K}}{\sqrt{np}}\right) + O_p\left(\frac{\sqrt{nK}}{p\sqrt{p}}\right) + O_p\left(\frac{1}{n}\right),
\end{aligned}
$$

under Assumption 1.

□

Bibliography

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

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

Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*(1), 199–227.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*(1), 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics 39*(6), 3320.

Green, P. J. and B. W. Silverman (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Crc Press.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*(4), 1004–1016.

Wong, F., C. K. Carter, and R. Kohn (2003). Efficient estimation of covariance selection models. *Biometrika 90*(4), 809–830.