Supplementary Document for the Manuscript entitled "Functional Linear Models with Latent Factors"

Zixuan Han, Tao Li, Jinhong You

Department of Statistics and Management

Shanghai University of Finance and Economics

Jiguo Cao

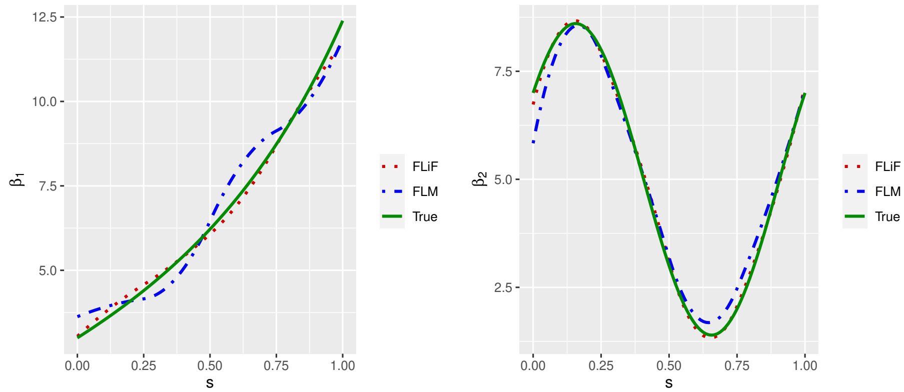
Department of Statistics and Actuarial Science, Simon Fraser University

In this supplementary document, we will provide the additional numerical results, empirical data analysis results and detailed proofs for theoretical results.

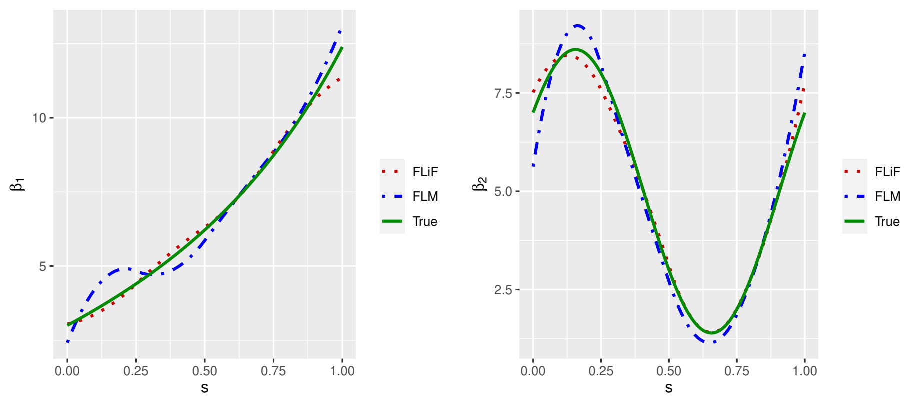
Supplementary Section A

We will first present the simulation results of the estimations under the cases where $c_1 = c_2 = 0, 0.5$ and relative pointwise bias of slope functions under three cases. We also present the RMSEs of the estimated loadings $\widehat{\boldsymbol{\lambda}}_l$ and factors $\widehat{\boldsymbol{F}}_l$, $l = 1, 2$, which are defined as $\mathtt{RMSE}(\hat{\boldsymbol{\lambda}}_l) = \left(\frac{1}{B}\sum_{b=1}^{B}(\frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\lambda}}_{li}^{b} - \boldsymbol{\lambda}_{li})^2)\right)^{1/2}$, $\mathtt{RMSE}(\hat{\boldsymbol{F}}_l) = \left(\frac{1}{B}\sum_{b=1}^{B}(\frac{1}{T}\sum_{t=1}^{T}(\hat{\boldsymbol{F}}_{lt}^{b} - \boldsymbol{F}_{lt})^2)\right)^{1/2}$, where $\boldsymbol{\lambda}_l =$

$(\boldsymbol{\lambda}_{l1}, \ldots, \boldsymbol{\lambda}_{lN})^\tau$ and $\boldsymbol{F}_l = (\boldsymbol{F}_{l1}, \ldots, \boldsymbol{F}_{lT})^\tau$.



(a) $c_1 = c_2 = 0$



(b) $c_1 = c_2 = 0.5$

Figure S1: The estimate for the slope functions $\boldsymbol{\beta}(s) = (\beta_1(s), \beta_2(s))^\tau$ using the functional linear model with latent factors (FLiF) and the conventional functional linear model (FLM) in the cases $c_1 = c_2 = 0$ and $c_1 = c_2 = 0.5$ obtained from one Monte Carlo run with the sample sizes $N = 100$ and the number of observations $T = 100$, where c_1 and c_2 are two constants indicating the correlation of scalar and functional covariates with the hiding factors, respectively.

RMSE of Estimated Loadings $\boldsymbol{\lambda}_l$ and Factors $\boldsymbol{F}_l$					
Sample Size		RMSE			
N	T	$\boldsymbol{\lambda}_1$	$\boldsymbol{\lambda}_2$	$\boldsymbol{F}_1$	$\boldsymbol{F}_2$
		$c_1 = c_2 = 0$			
50	50	0.275	0.258	0.195	0.216
50	100	0.238	0.224	0.154	0.163
100	100	0.172	0.193	0.128	0.119
		$c_1 = c_2 = 0.5$			
50	50	0.306	0.295	0.252	0.261
50	100	0.263	0.244	0.214	0.235
100	100	0.224	0.218	0.175	0.182
		$c_1 = c_2 = 1$			
50	50	0.357	0.368	0.302	0.282
50	100	0.315	0.309	0.253	0.247
100	100	0.269	0.251	0.212	0.229

Table S1: Root Mean Squared Errors (RMSE) of the estimates for factors $\boldsymbol{F}_l$ and the corresponding loadings $\boldsymbol{\lambda}_l$, $l = 1, 2$, when varying the sample sizes $N = 50, 100$ and the number of observations $T = 50, 100$, where c_1 and c_2 are two constants indicating the correlation of scalar and functional covariates with the hiding factors, respectively.

Supplementary Section B

When the common factors in our proposed model are known, the functional linear model with latent factors becomes the functional linear mixed model regardless of whether the factor loading is a one-dimensional or multi-dimensional vector. To compare these two models, in this section, we compare the proposed functional linear model with latent factors (FLiF) model with the conventional functional linear mixed model (FLMM).

The data are simulated based on the model $Y_{it} = \boldsymbol{\alpha}^{\tau} \boldsymbol{W_{it}} + \int_0^1 \boldsymbol{\beta}^{\tau}(s) \boldsymbol{X_{it}}(s) ds + \boldsymbol{\lambda_i^{\tau}} \boldsymbol{F_t} + \varepsilon_{it}, i = 1, \ldots, N$, and $t = 1, \ldots, T$, with the sample size $N = 50$ and the number of observations per subject $T = 50, 100$, where we set the factor structure as $\boldsymbol{\lambda_i} \sim N(0, \boldsymbol{I}_2)$ and $\boldsymbol{F_t} \sim N(0, 0.5\boldsymbol{I}_2)$. The scalar covariates $\boldsymbol{W_{it}}$ is set as $(W_{1it} + c_1\boldsymbol{\lambda_i^\tau F_t}, W_{2it})^\tau$, where c_1 is a constant indicating the correlation of scalar covariates with the hidden factors, and $W_{1it} \sim Exp(2)$, $W_{2it} \sim U(0, 1)$. The functional predictors $\boldsymbol{X_{it}}(\cdot) = (X_{1it}(\cdot), X_{2it}(\cdot))^\tau$ is set as $X_{1it}(s) = 1+c_2 \cdot \boldsymbol{\lambda}_i^\tau \boldsymbol{F}_t+\delta_{1it} \cdot s$, $X_{2it}(s) = c_2 \cdot \boldsymbol{\lambda}_i^\tau \boldsymbol{F}_t+\delta_{2it} \cdot \sin(2\pi s)$, where c_2 is another constant representing the correlation of functional covariates with the two hidden factors, $\delta_{1it} \sim U(-1, 1)$ and $\delta_{2it} \sim N(0, 2)$. We set the scalar coefficient $\boldsymbol{\alpha} = (1, 0.5)^\tau$, the functional coefficient $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \beta_2(\cdot))^\tau$ to be $\beta_1(s) = 2 + 3s + e^{2s}$, $\beta_2(s) = 5 + 3\sin(2\pi s) + 2\cos(2\pi s)$. The regression error ε_{it} are generated $i.i.d$ from the normal distribution $N(0, 1)$.

From the simulated data, we then estimate the proposed functional linear model with latent factors (FLiF) model, which assumes that the factors are unknown. We also estimate the conventional functional linear mixed model (FLMM): $Y_{it} = \boldsymbol{\alpha^\tau W_{it}} + \int_0^1 \boldsymbol{\beta}^\tau(s)\boldsymbol{X_{it}}(s)ds + \lambda_{1i}^\tau F_{1t} + \varepsilon_{it}$, which assumes the first factor F_{1t} is observed but the second factor F_{2t} is missing. This scenario mimics the problem of unobserved factors happening in many applications.

We use the relative root mean integrated squared errors (Re-RMISE) to measure the accuracy of the estimated functional coefficients under the two different models, which is defined as $\texttt{Re-RMISE}_l = \left(\frac{1}{B}\sum_{b=1}^{B}\int_S\left[\{\hat{\beta}_l^b(s) - \beta_l(s)\}/\beta_l(s)\right]^2 ds\right)^{1/2}$, $l = 1, 2$, where $\hat{\beta}_l^b(s)$ is the estimate of $\beta_l(s)$ in the b-th simulation replicate, $b = 1, \ldots, B$.

Average Re-RMISEs (%) for the estimated functional coefficient $\beta_1(s)$							
Sample Size		$c_1 = c_2 = 0$		$c_1 = c_2 = 0.5$		$c_1 = c_2 = 1$	
N	T	FLMM	FLiF	FLMM	FLiF	FLMM	FLiF
50	50	7.9	7.3	8.3	7.6	8.8	8.2
50	100	7.4	6.7	7.8	7.2	8.3	7.5
100	100	6.7	6.1	7.5	6.6	7.9	7.1

Table S2: The average of the relative root mean integrated squared errors (Re-RMISEs) (%) for the estimated functional coefficient $\hat{\beta}_1(s)$ over 100 simulation replicates for the proposed functional linear model with latent factors (FLiF) model and the functional linear mixed model (FLMM) when varying the sample sizes $N = 50, 100$ and the number of observations $T = 50, 100$, where c_1 and c_2 are two constants indicating the correlation of scalar and functional covariates with the hidden factors.

Table S2 shows the average Re-RMISEs of the estimated functional coefficient $\hat{\beta}_1(s)$ of two models under three different correlation settings. It shows that the estimates in the FLiF model have smaller Re-RMISE than those in the functional linear mixed model in almost all settings. Therefore, the performance of the proposed FLiF model is better than the FLMM model. This result is not surprising because the FLMM model assumes the first factor known but misses the second factor. Although the FLiF model requires identification and estimation of hidden factors, the FLiF model is more flexible because it can identify all hidden factors from data.

	fd	nvg	fin	edu	ener	med	eng	elec	consu	soft
food		0.87	0.92	0.03	0.79	0.98	0.81	0.12	**0.02**	0.96
navigation			0.93	**0.02**	0.86	0.82	0.73	0.45	0.12	0.91
finance				**6e-3**	0.82	0.63	0.74	0.31	0.07	0.83
education					0.18	**4e-5**	0.69	0.43	0.85	**8e-3**
energy						0.68	0.81	0.76	0.43	0.82
medicine							0.65	**0.02**	**3e-3**	0.93
engineering								0.80	0.65	0.74
electricity									0.84	0.28
consumption										0.06
software										

Table S3: P-value of multiple tests for loadings of ten sectors.

Supplementary Section C

Stock Price Analysis

An analysis of variance (ANOVA) test was conducted to determine whether the factor loadings of ten sectors differ significantly. The p-value of the ANOVA test is 1.197e-05, indicating significant differences in factor loadings among the ten sectors. To further identify which pairs of sectors differ significantly, we performed multiple tests on the factor loadings. Table S3 lists the p-values of the multiple test results for each pair of sectors. The table shows significant differences in factor loadings between the following pairs: 'education' and 'medicine' with a p-value of 4e-5, 'food' and 'consumption' with a p-value of 0.02, 'education' and 'navigation' with a p-value of 0.02, 'medicine' and 'electricity' with a p-value of 0.02, 'education' and 'finance' with a p-value of 6e-3, 'medicine' and 'consumption' with a p-value

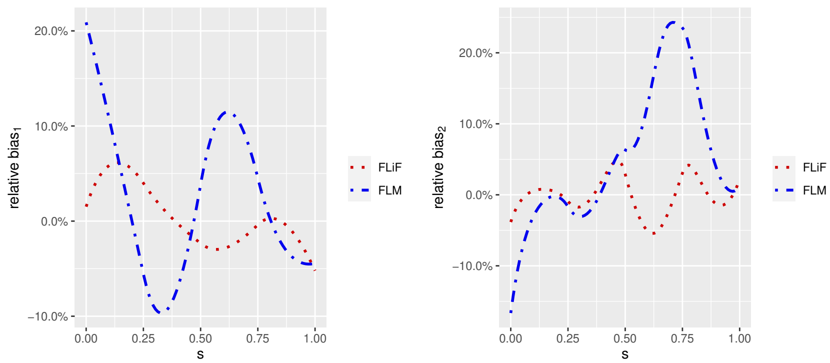
Re-RMSE for the estimated response $\hat{\boldsymbol{Y}}$		
Day	FLM	FLiF
70	0.076	0.068
71	0.081	0.072
72	0.086	0.075

Table S4: Relative Root Mean Square Error (Re-RMSE) for the estimated response variable $\hat{\boldsymbol{Y}}$ of day 70-72 using the proposed functional linear model with latent factors (FLiF) model and the conventional panel data model in air pollution analysis.

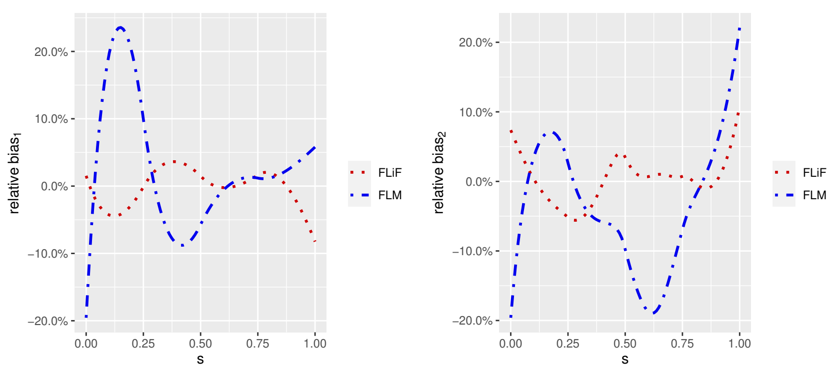
of 3e-3, and 'software' and 'education' with a p-value of 8e-3.

Air Pollution Analysis

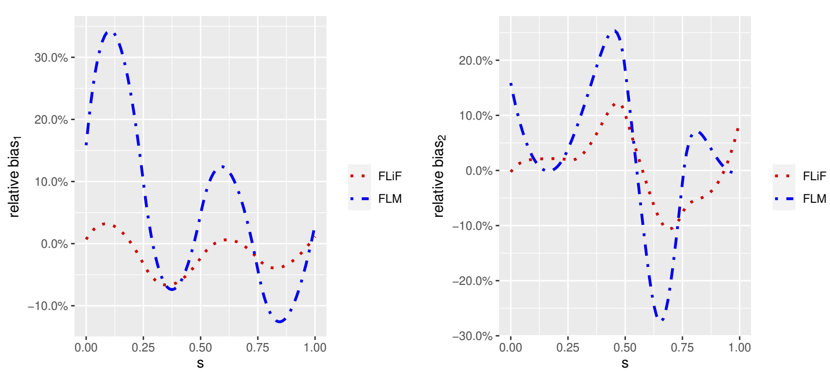
To assess whether latent factors improve prediction accuracy, we selected data from the first 69 days and estimated the proposed FLiF model. We then used this model to predict response variables for the subsequent three days. Simultaneously, we made the same predictions using the conventional functional linear model without latent factors. We measured the prediction performance of both models using the relative root mean square errors (Re-RMSEs). Table S4 presents the Re-RMSEs for the predicted responses $\hat{\boldsymbol{Y}}$ on days 70-72 using both the FLiF and FLM models. The results indicate that the Re-RMSEs for the predicted responses are reduced by 10.5%, 11.1%, and 12.8% on days 70, 71, and 72, respectively, when using the FLiF model in comparison with the FLM model. These findings suggest that incorporating hidden factors can enhance prediction accuracy.



(a) $c_1 = c_2 = 0$



(b) $c_1 = c_2 = 0.5$



(c) $c_1 = c_2 = 1$

Figure S2: The relative point-wise biases of the estimated slope functions $\hat{\beta}_1(s)$ (left panel) and $\hat{\beta}_2(s)$ (right panel) using the functional linear model with latent factors (FLiF) and the conventional functional linear model (FLM) in three cases $c_1 = c_2 = 0, 0.5, 1$ obtained from 200 Monte Carlo runs with the sample sizes $N = 100$ and the number of observations $T = 100$, where c_1 and c_2 are two constants indicating the correlation of scalar and functional covariates with the hiding factors, respectively.

Supplementary Section D

We will present the proof of theoretical results in this section. Some lemmas and auxiliary theories needed will also be given.

We use the following facts throughout the paper: $||\boldsymbol{F}|| = O_p(T^{1/2})$, $||\boldsymbol{Z}_i|| = O_p(T^{1/2})$ for all i, $\sum_{i=1}^N ||\boldsymbol{Z}_i||^2/NT = O_p(1)$, and $T^{-1/2}||\hat{\boldsymbol{F}}|| = \sqrt{r}$.

Lemma 1. *Under assumptions (A1)∼(A6), we have*

$$\sup_{\boldsymbol{F}} ||\frac{1}{NT}\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i|| = o_p(1).$$
$$\sup_{\boldsymbol{F}} ||\frac{1}{NT}\sum_{i=1}^N \boldsymbol{\lambda}_i^\tau \boldsymbol{F}_0^\tau \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i|| = o_p(1).$$
$$\sup_{\boldsymbol{F}} ||\frac{1}{NT}\sum_{i=1}^N \boldsymbol{\varepsilon}_i^\tau (\boldsymbol{P}_{\boldsymbol{F}} - \boldsymbol{P}_{\boldsymbol{F_0}}) \boldsymbol{\varepsilon}_i|| = o_p(1).$$

Theorem 1

Proof. Assume $\boldsymbol{\alpha}_0 = 0$ and $\boldsymbol{\beta}_0(\cdot) = 0$, $j = 1, \ldots, q$, namely, $\boldsymbol{\theta}_0 = 0$, then $\boldsymbol{Y}_i = \boldsymbol{F}_0 \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i$.

In the estimation procedure, the objective function is written as

$$Q(\boldsymbol{\theta}, \boldsymbol{F}) = \sum_{i=1}^N (\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\theta})^\tau \boldsymbol{M}_{\boldsymbol{F}} (\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\theta}) + \boldsymbol{\theta}^\tau \boldsymbol{G}_{\boldsymbol{\theta}} \boldsymbol{\theta}.$$

Define the centered objective function of the above equation as

$$\begin{aligned} Q_{NT}(\boldsymbol{\theta},\boldsymbol{F}) &= \frac{1}{NT}\left[\sum_{i=1}^{N}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\theta})^{\tau}\boldsymbol{M}_{\boldsymbol{F}}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\theta})\right] \\ &+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\theta}^{\tau}\boldsymbol{G}_{\boldsymbol{\theta}}\boldsymbol{\theta} - \frac{1}{NT}(\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{M}_{\boldsymbol{F}_0}\boldsymbol{\varepsilon}_i) \\ &= \tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F}) + Q_{\varepsilon} \end{aligned}$$

where

$$\begin{aligned} \tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F}) =& \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\theta}^{\tau}\boldsymbol{Z}_i^{\tau}\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{Z}_i\theta - \frac{2}{NT}\sum_{i=1}^{N}\boldsymbol{\theta}^{\tau}\boldsymbol{Z}_i^{\tau}\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F}_0\boldsymbol{\lambda}_i \\ &+ \operatorname{tr}\left[\frac{\boldsymbol{F}_0^{\tau}\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F}_0}{T}\cdot\frac{\boldsymbol{\Lambda}^{\tau}\boldsymbol{\Lambda}}{N}\right] + \frac{1}{NT}\boldsymbol{\theta}^{\tau}\boldsymbol{G}_{\boldsymbol{\theta}}\boldsymbol{\theta} \\ Q_{\varepsilon} =& \frac{2}{NT}\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}_0^{\tau}\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i - \frac{2}{NT}\sum_{i=1}^{N}\boldsymbol{\theta}^{\tau}\boldsymbol{Z}_i^{\tau}\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0}-\boldsymbol{P}_{\boldsymbol{F}})\boldsymbol{\varepsilon}_i \end{aligned}$$

Proof of (a) and (b)

By Lemma 1, $Q_{NT}(\boldsymbol{\theta},\boldsymbol{F}) = \tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F}) + o_p(1)$. For any invertible $\boldsymbol{H}$, $\tilde{Q}_{NT}(\boldsymbol{\theta}_0,\boldsymbol{F}_0\boldsymbol{H}) = 0$.

Next, show that for any $(\boldsymbol{\theta},\boldsymbol{F}) \neq (\boldsymbol{\theta}_0,\boldsymbol{F}_0\boldsymbol{H}) = (0,\boldsymbol{F}_0\boldsymbol{H})$, $\tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F}) > 0$.

Define $\boldsymbol{A} = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{Z}_i$, $\boldsymbol{B} = \dfrac{\boldsymbol{\Lambda}^{\tau}\boldsymbol{\Lambda}}{N}\otimes\boldsymbol{I}_T$, $\boldsymbol{C} = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{\tau}\otimes\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{Z}_i$, $\boldsymbol{\varphi} = vec(\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F}_0)$.

Then,

$$\begin{aligned}\tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F}) &= \boldsymbol{\theta}^\tau\boldsymbol{A}\boldsymbol{\theta} - 2\boldsymbol{\theta}^\tau\boldsymbol{C}^\tau\boldsymbol{\varphi} + \boldsymbol{\varphi}^\tau\boldsymbol{B}\boldsymbol{\varphi} + \frac{1}{NT}\boldsymbol{\theta}^\tau\boldsymbol{G}_{\boldsymbol{\theta}}\boldsymbol{\theta} \\ &= \boldsymbol{\theta}^\tau(\boldsymbol{A} + \frac{1}{NT}\boldsymbol{G}_{\boldsymbol{\theta}} - \boldsymbol{C}^\tau\boldsymbol{B}^{-1}\boldsymbol{C})\boldsymbol{\theta} + (\boldsymbol{\varphi}^\tau - \boldsymbol{\theta}^\tau\boldsymbol{C}^\tau\boldsymbol{B}^{-1})\boldsymbol{B}(\boldsymbol{\varphi} - \boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{\theta}) \\ &= \boldsymbol{\theta}^\tau(D(\boldsymbol{F}) + \frac{1}{NT}\boldsymbol{G}_{\boldsymbol{\theta}})\boldsymbol{\theta} + \boldsymbol{\eta}^\tau\boldsymbol{B}\boldsymbol{\eta}\end{aligned}$$

where $\boldsymbol{\eta} = \boldsymbol{\varphi} - \boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{\theta}$.

By Assumption 2, $D(\boldsymbol{F})$ is positive definite, $\boldsymbol{B}$ is positive definite and $\boldsymbol{G}_{\boldsymbol{\theta}}$ is semipositive definite by definition. If either $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0 = 0$ or $\boldsymbol{F} \neq \boldsymbol{F}_0\boldsymbol{H}$, $\tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F}) > 0$, indicating that $\tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F})$ achieves its unique minimum at $(\boldsymbol{\theta}_0, \boldsymbol{F}_0\boldsymbol{H}) = (0, \boldsymbol{F}_0\boldsymbol{H})$. Since $\hat{\beta}_j(s) = \hat{\gamma}_j^\tau\boldsymbol{b}(s)$, $j = 1,\ldots,q$, therefore, $\hat{\beta}_j(s)$ is uniquely defined with probability tending to one.

Furthermore,for any $||\boldsymbol{\alpha}|| \geq c > 0$, which implies that $||\boldsymbol{\theta}|| = ||(\boldsymbol{\alpha}^\tau,\boldsymbol{\gamma}^\tau)^\tau|| \geq c > 0$, $\tilde{Q}_{NT}(\boldsymbol{\theta},\boldsymbol{F}) \geq \lambda_{\boldsymbol{D}}c^2 > 0$, where $\lambda_{\boldsymbol{D}}$ is the minimum eigenvalue of $D(\boldsymbol{F})$. Hence, $\hat{\boldsymbol{\alpha}}$ is consistent for $\boldsymbol{\alpha}$.

Proof of (c)

The proof of part (c) is similar to that of Proposition 1 in Bai (2009) , and we do not present the detailed proof.

By definition, the centered objective function satisfies that $Q_{NT}(\boldsymbol{\theta}_0,\boldsymbol{F}_0) = 0$, and for estimation $(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{F}})$, it is obvious that $Q_{NT}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{F}}) \leq 0$.

However, since

$$Q_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{F}}) = \tilde{Q}_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{F}}) + o_p(1),$$

and $\tilde{Q}_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{F}}) \geq 0$, $\tilde{Q}_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{F}}) = o_p(1)$.

From the proof of part (a), $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$, combining with the equation of $\tilde{Q}_{NT}(\boldsymbol{\theta}, \boldsymbol{F})$, it is easy to obtain that

$$\mathrm{tr}\left[\frac{\boldsymbol{F}_0^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{F}_0}{T} \cdot \frac{\boldsymbol{\Lambda}^\tau \boldsymbol{\Lambda}}{N}\right] = o_p(1).$$

Since $\dfrac{\boldsymbol{F}_0^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{F}_0}{T} \geq 0$, $\dfrac{\boldsymbol{\Lambda}^\tau \boldsymbol{\Lambda}}{N} > 0$, then

$$\frac{\boldsymbol{F}_0^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{F}_0}{T} = \frac{\boldsymbol{F}_0^\tau \boldsymbol{F}_0}{T} - \frac{\boldsymbol{F}_0^\tau \boldsymbol{P}_{\hat{\boldsymbol{F}}} \boldsymbol{F}_0}{T} = \frac{\boldsymbol{F}_0^\tau \boldsymbol{F}_0}{T} - \frac{\boldsymbol{F}_0^\tau \hat{\boldsymbol{F}}}{T} \cdot \frac{\hat{\boldsymbol{F}}^\tau \boldsymbol{F}_0}{T} = o_p(1),$$

which also implies that $\dfrac{\hat{\boldsymbol{F}}^\tau \boldsymbol{P}_{\boldsymbol{F}_0} \hat{\boldsymbol{F}}}{T} \xrightarrow{P} \boldsymbol{I}_r$.

By Assumption 3, $\dfrac{\boldsymbol{F}_0^\tau \boldsymbol{F}_0}{T}$ is invertible, indicating that $\dfrac{\boldsymbol{F}_0^\tau \hat{\boldsymbol{F}}}{T}$ is invertible.

Then,

$$||\boldsymbol{P}_{\hat{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}_0}||^2 = \mathrm{tr}(\boldsymbol{P}_{\hat{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}_0})^2 = 2\mathrm{tr}(\boldsymbol{I}_r - \frac{\hat{\boldsymbol{F}}^\tau \boldsymbol{P}_{\boldsymbol{F}_0} \hat{\boldsymbol{F}}}{T}).$$

Therefore, $||\boldsymbol{P}_{\hat{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}_0}|| \xrightarrow{P} 0$. □

For ease of notation, define $\mathcal{D} = \{(\boldsymbol{W}_{it}, \boldsymbol{X}_{it}, \boldsymbol{\lambda}_i, \boldsymbol{f}_t), i = 1, \ldots, N, t = 1, \ldots, T\}$, $\delta_{NT} = \min[\sqrt{N}, \sqrt{T}]$, $\zeta_{LD} = qL^{-2d}$, $\varsigma = \min[\delta_{NT}^{-2}, L^{-2}]$. Meanwhile, we denote $a_n \asymp b_n$ as if a_n and b_n are both positive, and both a_n/b_n and b_n/a_n are bounded for all n.

Lemma 2. *Assume that assumptions (A1)∼(A6) hold, then*

(1) For any r, $1 \leq r \leq r_0$, $\boldsymbol{H}_k$ is an $r_0 \times r$ matrix, then $V(r, \hat{\boldsymbol{F}}_r) - V(r, \boldsymbol{F}_0\boldsymbol{H}_r) = O_p(\delta_{NT}^{-1})$.

(2) For each r with $1 \leq r < r_0$, and the matrix $\boldsymbol{H}_r$ is defined in (1), there assist a $c_r > 0$, such that plim $\inf_{N,T\to\infty}[V(r, \boldsymbol{F}_0\boldsymbol{H}_r) - V(r_0, \boldsymbol{F}_0)] = c_r$.

(3) For any fixed r with $r \geq r_0$, $V(r, \hat{\boldsymbol{F}}_r) - V(r_0, \hat{\boldsymbol{F}}_{r_0}) = O_p(\delta_{NT}^{-2})$.

Theorem 2

Proof. Since $BIC(r) = \ln V(r) + \rho r$ for any given r, then

$$BIC(r) - BIC(r_0) = \ln[V(r)/V(r_0)] + \rho(r - r_0).$$

We consider two cases as (1): $1 \leq r < r_0$, and (2): $r_0 < r \leq r_{max}$.

For case (1) $1 \leq r < r_0$, from Lemma (2), we can get that $V(r)/V(r_0) \geq 1+\epsilon_0$ for some $\epsilon_0 > 0$ with large probability, and thus $\ln[V(r)/V(r_0)] \geq \epsilon_0/2$. Combining with the fact that $\rho(r - r_0) \to 0$ where ρ is generally assumed to tend to zero at an appropriate rate for accurately determining the factor numbers, we have that $BIC(r) - BIC(r_0) \geq \epsilon_0/2 - \rho(r_0 - r) \geq \epsilon_0/3$ with large probability, then it is clear that

$$P(BIC(r) - BIC(r_0) < 0) \to 0, \quad N, T \to \infty.$$

For case (2) $r_0 < r \leq r_{max}$, from Lemma (2), it is obvious that $V(r)/V(r_0) = 1 + O_p(\delta_{NT}^{-2})$, thus $\ln[V(r)/V(r_0)] = O_p(\delta_{NT}^{-2})$. Then

$$\begin{aligned} P(BIC(r) - BIC(r_0) < 0) &= P(\ln[V(r)/V(r_0)] + \rho(r - r_0) < 0) \\ &\leq P(Op(\delta_{NT}^{-2}) + \rho < 0) \to 0, \quad N, T \to \infty. \end{aligned}$$

Therefore, $P(BIC(r) - BIC(r_0) > 0) \to 1$, namely, $BIC(r)$ achieves its minimum only at $r = r_0$ for any r in $1 \leq r \leq r_{max}$ as $N, T \to \infty$.

$\square$

From definition, we can rewrite $\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma} \end{pmatrix}$, $\boldsymbol{Z}_i = (\boldsymbol{W}_i \ \boldsymbol{B}_i)$, $\boldsymbol{Z}_i^\tau = (\boldsymbol{W}_i^\tau, \boldsymbol{B}_i^\tau)^\tau$.

Lemma 3. *Assume that assumptions (A1)∼(A6) hold, and denote* $\boldsymbol{H} =$

$(\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})(\frac{\boldsymbol{F}^\tau\hat{\boldsymbol{F}}}{T})\boldsymbol{V}_{NT}^{-1}$. *Then, we have*

$$(1)\quad T^{-1/2}||\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}|| = O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-1}) + Op(\zeta_{LD}^{1/2}) + O_p(\varsigma^{1/2}).$$

$$(2)\quad T^{-1}\boldsymbol{F}(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}) = O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-2}) + Op(\zeta_{LD}^{1/2}) + O_p(\varsigma).$$

$$(3)\quad T^{-1}\hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}) = O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-2}) + Op(\zeta_{LD}^{1/2}) + O_p(\varsigma).$$

$$(4)\quad T^{-1}\boldsymbol{Z}_k^\tau(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}) = O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-2}) + O_p(\zeta_{LD}^{1/2}) + O_p(\varsigma) \quad \textit{for all } k = 1,\ldots,N.$$

$$(5)\quad (NT)^{-1}\sum_{k=1}^{N}\boldsymbol{Z}_k^\tau(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}) = O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-2}) + O_p(\zeta_{LD}^{1/2}) + O_p(\varsigma).$$

Lemma 4. *Assume that assumptions (A1)∼(A6) hold, and denote* $\boldsymbol{H} = (\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})(\frac{\boldsymbol{F}^\tau\hat{\boldsymbol{F}}}{T})\boldsymbol{V}_{NT}^{-1}$. *Then, we have*

$$(1)\quad T^{-1}\boldsymbol{\varepsilon}_k^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}) = T^{-1/2}O_p(\|\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}\|) + O_p(\delta_{NT}^{-2}) + O_p(T^{-1/2}\zeta_{LD}^{1/2}) + O_p(\varsigma) \quad \textit{for all } k=1,\ldots,N.$$

$$(2)\quad (T\sqrt{N})^{-1}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}) = T^{-1/2}O_p(\|\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}\|) + N^{-1/2}O_p(\|\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}\|) + Op(N^{-1/2}) + O_p(\delta_{NT}^{-2}) + O_p(\zeta_{LD}^{1/2}) + O_p(\varsigma).$$

$$(3)\quad (NT)^{-1}\sum_{k=1}^{N}\boldsymbol{\lambda}_k^{\tau}\boldsymbol{\varepsilon}_k^{\tau}(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}) = (NT)^{-1/2}O_p(\|\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}\|) + O_p(N^{-1}) + N^{-1/2}O_p(\delta_{NT}^{-2}) + N^{-1/2}O_p(\zeta_{LD}^{1/2}) + N^{-1/2}O_p(\varsigma).$$

From Assumption (A6) and Corollary 6.21 in Chumaker (1981), there exists a constant M such that

$$\beta_j(s) = \sum_{l=1}^{L}\tilde{\gamma}_{jl}b_l(s) + c\epsilon_j(s),$$
$$\sup_{s\in\mathcal{S}}|ce_j(s)| \le ML^{-d}, j=1,\ldots,q. \quad \text{(S0.1)}$$

Let $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\gamma}}_1^{\tau},\ldots,\tilde{\boldsymbol{\gamma}}_q^{\tau})^{\tau}$ with $\tilde{\boldsymbol{\gamma}}_j = (\tilde{\gamma}_{j1},\ldots,\tilde{\gamma}_{jL})^{\tau}$, $\boldsymbol{e}_i = (e_{i1},\ldots,e_{iT})^{\tau}$ with $e_{it} = \sum_{j=1}^{q}\int cX_{itj}(s)\epsilon_j(s)ds$. Then, $\boldsymbol{Y}_i = \boldsymbol{W}_i\boldsymbol{\alpha} + \boldsymbol{B}_i\tilde{\boldsymbol{\gamma}} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i +$

$\boldsymbol{e}_i, i = 1, \ldots, N$. Denote $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\alpha}, \tilde{\boldsymbol{\gamma}})^\tau$, and the model can be written as

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i\tilde{\boldsymbol{\theta}} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i, \qquad i = 1, \ldots, N.$$

Theorem 3

Proof. Since $\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Z}_i + \boldsymbol{G}_{\boldsymbol{\theta}}\right)^{-1} \sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Y}_i$, then

$$\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma} \end{pmatrix} = \left(\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Z}_i + \boldsymbol{G}_{\boldsymbol{\theta}}\right)^{-1} \left(\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} (\boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i) - \boldsymbol{G}_{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}\right),$$

or equivalently,

$$\left(\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Z}_i + \boldsymbol{G}_{\boldsymbol{\theta}}\right)(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) = \left(\sum_{i=1}^N \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\hat{\boldsymbol{F}}} (\boldsymbol{W}_i \ \boldsymbol{B}_i) + \begin{pmatrix} \boldsymbol{0}_{\boldsymbol{p}\times\boldsymbol{p}} & \\ & \boldsymbol{G}_\beta \end{pmatrix}\right) \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma} \end{pmatrix} \tag{S0.2}$$

$$= \sum_{i=1}^N \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{F}\boldsymbol{\lambda}_i + \sum_{i=1}^N \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{\varepsilon}_i + \sum_{i=1}^N \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{e}_i - \begin{pmatrix} \boldsymbol{0}_{\boldsymbol{p}\times\boldsymbol{p}} & \\ & \boldsymbol{G}_\beta \end{pmatrix} \tilde{\boldsymbol{\theta}}.$$

For the third term, from assumptions (A3)∼(A5) and (S0.1), using the similar proofs to Lemma A.7 in Huang et al. (2004), and Lemmas 2 and 3 in Appendix, it is easy to obtain

$$\frac{1}{NT}||\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{e}_i||^2 = O_p(\zeta_{LD}/L).$$

For the first term of the right hand of(S0.2), since $\boldsymbol{M}_{\hat{\boldsymbol{F}}}\hat{\boldsymbol{F}} = 0$, then $\boldsymbol{M}_{\hat{\boldsymbol{F}}}\hat{\boldsymbol{F}} = \boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{F} - \hat{\boldsymbol{F}}\boldsymbol{H}^{-1})$. Meanwhile, for the last term, under the assumption (A7) of the smoothing parameter $\boldsymbol{\xi}$, we can obtain that it is $o_p(1)$.

From $\left[\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}})(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}})^\tau\right]\hat{\boldsymbol{F}} = \hat{\boldsymbol{F}}\boldsymbol{V}_{NT}$ and $\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}} =$

$\boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i$, we obtain the following expansion as

$$
\begin{aligned}
\hat{\boldsymbol{F}}\boldsymbol{V}_{NT} =& \Bigg[\frac{1}{NT}\sum_{i=1}^{N}\left((\boldsymbol{W}_i\ \boldsymbol{B}_i)\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i\right) \\
&\left((\boldsymbol{W}_i\ \boldsymbol{B}_i)\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i\right)^{\tau}\Bigg]\hat{\boldsymbol{F}} \\
=& \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{W}_i\ \boldsymbol{B}_i)\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}^{\tau}\begin{pmatrix}\boldsymbol{W}_i^{\tau} \\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{W}_i\ \boldsymbol{B}_i)\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{W}_i\ \boldsymbol{B}_i)\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}^{\tau}\begin{pmatrix}\boldsymbol{W}_i^{\tau} \\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}^{\tau}\begin{pmatrix}\boldsymbol{W}_i^{\tau} \\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{W}_i\ \boldsymbol{B}_i)\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma}\end{pmatrix}^{\tau}\begin{pmatrix}\boldsymbol{W}_i^{\tau} \\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} \\
=:& \boldsymbol{I}_1 + \ldots + \boldsymbol{I}_{16},
\end{aligned}
$$

where $\boldsymbol{I}_{16} = \dfrac{1}{NT}\sum_{i=1}^{N} \boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^\tau \boldsymbol{F}^\tau \hat{\boldsymbol{F}} = \boldsymbol{F}(\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})^{-1}\frac{\boldsymbol{F}^\tau\hat{\boldsymbol{F}}}{T}$.

Then, it is easy to get

$$\boldsymbol{F} - \hat{\boldsymbol{F}}\boldsymbol{H}^{-1} = -(\boldsymbol{I}_1 + \ldots + \boldsymbol{I}_{15})\boldsymbol{G},$$

where $\boldsymbol{H} = (\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})(\frac{\boldsymbol{F}^\tau\hat{\boldsymbol{F}}}{T})\boldsymbol{V}_{NT}^{-1}$, and $\boldsymbol{G} = (\frac{\boldsymbol{F}^\tau\hat{\boldsymbol{F}}}{T})^{-1}(\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})^{-1}$.

Furthermore,

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{F}\boldsymbol{\lambda}_i &= \frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{F} - \hat{\boldsymbol{F}}\boldsymbol{H}^{-1})\boldsymbol{\lambda}_i\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_1 + \ldots + \boldsymbol{I}_{15})\boldsymbol{G}\boldsymbol{\lambda}_i\\ &=: \boldsymbol{J}_1 + \ldots + \boldsymbol{J}_{15},\end{aligned}$$

where $\boldsymbol{J}_1 \sim \boldsymbol{J}_{15}$ are implicitly defined via $\boldsymbol{I}_1 \sim \boldsymbol{I}_{15}$ respectively.

For $\boldsymbol{J}_1$, we have

$$\begin{aligned}\boldsymbol{J}_1 &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_1)\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\Big[\frac{1}{NT}\sum_{k=1}^{N}(\boldsymbol{W}_k\ \boldsymbol{B}_k)\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}^{\tau}\begin{pmatrix}\boldsymbol{W}_k^{\tau}\\ \boldsymbol{B}_k^{\tau}\end{pmatrix}\hat{\boldsymbol{F}}\Big]\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= o_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}||).\end{aligned}$$

For $\boldsymbol{J}_2$, we have

$$\begin{aligned}\boldsymbol{J}_2 &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_2)\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\Big[\frac{1}{NT}\sum_{k=1}^{N}(\boldsymbol{W}_k\ \boldsymbol{B}_k)\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}}\Big](\frac{\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}}}{T})^{-1}(\frac{\boldsymbol{\Lambda}^{\tau}\boldsymbol{\Lambda}}{N})^{-1}\boldsymbol{\lambda}_i\\ &= -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{W}_k\ \boldsymbol{B}_k)\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}\Big(\boldsymbol{\lambda}_k^{\tau}(\frac{\boldsymbol{\Lambda}^{\tau}\boldsymbol{\Lambda}}{N})^{-1}\boldsymbol{\lambda}_i\Big)\\ &= -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{W}_k\ \boldsymbol{B}_k)a_{ik}\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix},\end{aligned}$$

where $a_{ik} = \boldsymbol{\lambda}_k^{\tau}(\frac{\boldsymbol{\Lambda}^{\tau}\boldsymbol{\Lambda}}{N})^{-1}\boldsymbol{\lambda}_i$.

For $\boldsymbol{J}_3$, we have

$$\begin{aligned}\boldsymbol{J}_3 &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_3)\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}(\boldsymbol{W}_k\ \boldsymbol{B}_k)\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}\boldsymbol{\varepsilon}_k^\tau\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{W}_k\ \boldsymbol{B}_k)\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}\frac{\boldsymbol{\varepsilon}_k^\tau\hat{\boldsymbol{F}}}{T}\boldsymbol{G}\boldsymbol{\lambda}_i\end{aligned}$$

By Lemma 3 and some calculation, we can have

$$\begin{aligned}T^{-1}\boldsymbol{\varepsilon}_k^\tau\hat{\boldsymbol{F}} =&T^{-1}\boldsymbol{\varepsilon}_k^\tau\boldsymbol{F}\boldsymbol{H} + T^{-1}\boldsymbol{\varepsilon}_k^\tau(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H})\\ =&O_p(T^{-1/2}) + T^{-1/2}O_p(\|\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}\|)\\ &+ O_p(\delta_{NT}^{-2}) + O_p(T^{-1/2}\zeta_{LD}^{1/2}) + O_p(T^{-1/2}\varsigma).\end{aligned}$$

Combining with the similar argument as the proof of Lemma 2, it is easy to get that $\boldsymbol{J}_3 = o_p(\|\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}\|)$.

Similarly, for $\boldsymbol{J}_5$, we have

$$\begin{aligned}
\boldsymbol{J}_5 &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_5)\boldsymbol{G}\boldsymbol{\lambda}_i\\
&= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}^{\tau}(\boldsymbol{W}_k\ \boldsymbol{B}_k)^{\tau}\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i\\
&= o_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}||).
\end{aligned}$$

For $\boldsymbol{J}_4$, we have

$$\begin{aligned}
\boldsymbol{J}_4 &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_4)\boldsymbol{G}\boldsymbol{\lambda}_i\\
&= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_k\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}^{\tau}(\boldsymbol{W}_k\ \boldsymbol{B}_k)^{\tau}\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i\\
&= -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{F}\boldsymbol{\lambda}_k\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}^{\tau}\frac{(\boldsymbol{W}_k\ \boldsymbol{B}_k)^{\tau}\hat{\boldsymbol{F}}}{T}\boldsymbol{G}\boldsymbol{\lambda}_i
\end{aligned}$$

Since $\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{F} = \boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{F}-\hat{\boldsymbol{F}}\boldsymbol{H}^{-1})$, using Lemma 3, we can get that $\boldsymbol{J}_4 = o_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}||)$.

For $\boldsymbol{J}_6$, we have

$$\begin{aligned}\boldsymbol{J}_6 &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_6)\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_k\boldsymbol{\varepsilon}_k^\tau\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{F}-\hat{\boldsymbol{F}}\boldsymbol{H}^{-1})\left(\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\lambda}_k\frac{\boldsymbol{\varepsilon}_k^\tau\hat{\boldsymbol{F}}}{T}\right)\boldsymbol{G}\boldsymbol{\lambda}_i.\end{aligned}$$

Here, by Lemma 3,

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\lambda}_k\frac{\boldsymbol{\varepsilon}_k^\tau\hat{\boldsymbol{F}}}{T} =&\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\lambda}_k\boldsymbol{\varepsilon}_k^\tau\boldsymbol{F}\boldsymbol{H}+\frac{1}{N}\sum_{k=1}^{NT}\boldsymbol{\lambda}_k\boldsymbol{\varepsilon}_k^\tau(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H})\\ =&O_p((NT)^{-1/2})+O_p(N^{-1})+N^{-1/2}O_p(\delta_{NT}^{-2})\\ &+N^{-1/2}O_p(\zeta_{LD}^{1/2})+N^{-1/2}O_p(\varsigma),\end{aligned}$$

and

$$-\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{F}-\hat{\boldsymbol{F}}\boldsymbol{H}^{-1}) = O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}||)+O_p(\delta_{NT}^{-2})+O_p(\zeta_{LD}^{1/2})+O_p(\varsigma).$$

Since $\boldsymbol{G}$ does not depend on i and $||\boldsymbol{G}|| = O_p(1)$, then

$$\begin{aligned}
\boldsymbol{J}_6 =& \Big[O_p(|| \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}} \end{pmatrix} ||) + O_p(\delta_{NT}^{-2}) + O_p(\zeta_{LD}^{1/2}) + O_p(\varsigma) \Big] \\
& \times \Big[O_p((NT)^{-1/2}) + O_p(N^{-1}) + N^{-1/2} O_p(\delta_{NT}^{-2}) \\
& \quad + N^{-1/2} O_p(\zeta_{LD}^{1/2}) + N^{-1/2} O_p(\varsigma) \Big] \\
=& o_p(|| \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}} \end{pmatrix} ||) + o_p((NT)^{-1/2}) + N^{-1} O_p(\delta_{NT}^{-2}) + N^{-1/2} O_p(\delta_{NT}^{-4}) \\
& + N^{-1} O_p(\zeta_{LD}^{1/2}) + N^{-1/2} O_p(\zeta_{LD}) + N^{-1} O_p(\varsigma) + N^{-1/2} O_p(\varsigma^2).
\end{aligned}$$

For $\boldsymbol{J}_7$, we have

$$\begin{aligned}
\boldsymbol{J}_7 &= -\frac{1}{NT} \sum_{i=1}^{N} \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\hat{\boldsymbol{F}}} (\boldsymbol{I}_7) \boldsymbol{G} \boldsymbol{\lambda}_i \\
&= -\frac{1}{NT} \sum_{i=1}^{N} \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\hat{\boldsymbol{F}}} (\frac{1}{NT} \sum_{k=1}^{N} \boldsymbol{\varepsilon}_k \boldsymbol{\lambda}_k^\tau \boldsymbol{F}^\tau \hat{\boldsymbol{F}}) (\frac{\boldsymbol{F}^\tau \hat{\boldsymbol{F}}}{T})^{-1} (\frac{\boldsymbol{\Lambda}^\tau \boldsymbol{\Lambda}}{N})^{-1} \boldsymbol{\lambda}_i \\
&= -\frac{1}{N^2 T} \sum_{i=1}^{N} \sum_{k=1}^{N} a_{ik} \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{\varepsilon}_k,
\end{aligned}$$

where $a_{ik} = \boldsymbol{\lambda}_k^\tau (\frac{\boldsymbol{\Lambda}^\tau \boldsymbol{\Lambda}}{N})^{-1} \boldsymbol{\lambda}_i$.

For $\boldsymbol{J}_8$, we have

$$\begin{aligned}\boldsymbol{J}_8 =& -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_8)\boldsymbol{G}\boldsymbol{\lambda}_i = -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\tau}\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i\\ =& -\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\sigma^2\boldsymbol{I}_T\hat{\boldsymbol{F}}\boldsymbol{G}\boldsymbol{\lambda}_i - \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\tau}-\sigma^2\boldsymbol{I}_T)\hat{\boldsymbol{F}}\boldsymbol{G}\boldsymbol{\lambda}_i\\ =&\boldsymbol{M}_{NT} + O_p((T\sqrt{N})^{-1}) + (NT)^{-1/2}\left[O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-1}) + O_p(\zeta_{LD}^{1/2}) + O_p(\varsigma^{1/2})\right]\\ &+ N^{-1/2}\left[O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-1}) + O_p(\zeta_{LD}^{1/2}) + O_p(\varsigma^{1/2})\right]^2,\end{aligned}$$

where $\boldsymbol{M}_{NT} = -\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\sigma^2\boldsymbol{I}_T\hat{\boldsymbol{F}}\boldsymbol{G}\boldsymbol{\lambda}_i$.

For $\boldsymbol{J}_9$ and $\boldsymbol{J}_{10}$ which depend on $\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}$, it is easy to prove that are bounded by $O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||)$.

For $\boldsymbol{J}_{11}$, using Lemma 3 and the equation $||\boldsymbol{e}_k^{\tau}\hat{\boldsymbol{F}}/T|| = ||\boldsymbol{e}_k||\sqrt{r}/\sqrt{T} =$

$O_p(\zeta_{LD}^{1/2})$, we have

$$\boldsymbol{J}_{11} = -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_{11})\boldsymbol{G}\boldsymbol{\lambda}_i = -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_k\boldsymbol{e}_k^\tau\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i$$

$$= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{F} - \hat{\boldsymbol{F}}\boldsymbol{H}^{-1})(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\lambda}_k\boldsymbol{e}_k^\tau\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i$$

$$= O_p(\zeta_{LD}^{1/2})\left[O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-2}) + O_p(\zeta_{LD}^{1/2}) + O_p(\varsigma)\right].$$

For $\boldsymbol{J}_{12}$, we have

$$\boldsymbol{J}_{12} = -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_{12})\boldsymbol{G}\boldsymbol{\lambda}_i$$

$$= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{e}_k\boldsymbol{\lambda}_k^\tau\boldsymbol{F}^\tau\hat{\boldsymbol{F}})(\frac{\boldsymbol{F}^\tau\hat{\boldsymbol{F}}}{T})^{-1}(\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})^{-1}\boldsymbol{\lambda}_i$$

$$= -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}a_{ik}\begin{pmatrix}\boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{e}_k$$

$$= O_p(L^{-1/2}\zeta_{LD}^{1/2}) + O_p(L^{-1/2}\varsigma),$$

where $a_{ik} = \boldsymbol{\lambda}_k^\tau(\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})^{-1}\boldsymbol{\lambda}_i$.

Similarly, we can get that $\boldsymbol{J}_{13} = -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\dfrac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k\boldsymbol{e}_k^\tau\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i$ is bounded by $(NT)^{-1/2}O_p(\zeta_{LD}^{1/2}) + (NT)^{-1/2}O_p(\varsigma)$.

For $\boldsymbol{J}_{14}$, we have

$$\begin{aligned}\boldsymbol{J}_{14} =& -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_{14})\boldsymbol{G}\boldsymbol{\lambda}_i = -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{e}_k\boldsymbol{\varepsilon}_k^\tau\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i\\ =& -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{e}_k(\frac{\boldsymbol{\varepsilon}_k^\tau\boldsymbol{F}\boldsymbol{H}}{T})\boldsymbol{G}\boldsymbol{\lambda}_i - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{e}_k(\frac{\boldsymbol{\varepsilon}_k^\tau(\hat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H})}{NT})\boldsymbol{G}\boldsymbol{\lambda}_i\\ =& T^{-1/2}O_p(\zeta_{LD}^{1/2}) + T^{-1/2}O_p(\varsigma)\\ &+ O_p(\zeta_{LD}^{1/2})\left[T^{-1/2}O_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\end{pmatrix}||) + O_p(\delta_{NT}^{-2}) + O_p(T^{-1/2}\zeta_{LD}^{1/2}) + O_p(T^{-1/2}\varsigma)\right].\end{aligned}$$

For $\boldsymbol{J}_{15}$, since $\boldsymbol{M}_{\hat{\boldsymbol{F}}}\hat{\boldsymbol{F}} = 0$, we have

$$\begin{aligned}\boldsymbol{J}_{15} &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{I}_{15})\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^\tau\\ \boldsymbol{B}_i^\tau\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{e}_k\boldsymbol{e}_k^\tau\hat{\boldsymbol{F}})\boldsymbol{G}\boldsymbol{\lambda}_i\\ &= o_p(\zeta_{LD}).\end{aligned}$$

Summarizing the results and we can obtain the following equation

$$\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{F}\boldsymbol{\lambda}_i=\boldsymbol{J}_2+\boldsymbol{J}_7+\boldsymbol{M}_{NT}+o_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}||)+O_p((NT)^{-1/2})$$

$$+O_p((T\sqrt{N})^{-1})+N^{-1/2}O_p(\delta_{NT}^{-2})+O_p(T^{-1/2}\zeta_{LD}^{1/2})+O_p(L^{-1/2}\zeta_{LD}^{1/2})$$

$$+N^{-1/2}O_p(\varsigma)+O_p(L^{-1/2}\varsigma).$$

Then,

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{W}_i\ \boldsymbol{B}_i)+\frac{1}{NT}\boldsymbol{G}_{\boldsymbol{\theta}}+o_p(1)\right)\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}-\boldsymbol{J}_2$$

$$=\left(\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{W}_i\ \boldsymbol{B}_i)+\frac{1}{NT}\boldsymbol{G}_{\boldsymbol{\theta}}+o_p(1)\right)\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}-\boldsymbol{J}_2$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i+\boldsymbol{J}_7+\boldsymbol{M}_{NT}+o_p(||\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}\\ \hat{\gamma}-\tilde{\gamma}\end{pmatrix}||)+O_p((NT)^{-1/2})$$

$$+O_p((T\sqrt{N})^{-1})+N^{-1/2}O_p(\delta_{NT}^{-2})+O_p(T^{-1/2}\zeta_{LD}^{1/2})+O_p(L^{-1/2}\zeta_{LD}^{1/2})$$

$$+N^{-1/2}O_p(\varsigma)+O_p(L^{-1/2}\varsigma).$$

Multiplying $(D(\hat{\boldsymbol{F}})+\frac{1}{NT}\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}$ on each side of the equation, where

$D(\boldsymbol{F}) = \frac{1}{NT}\sum_{i=1}^{N} \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\boldsymbol{F}}(\boldsymbol{W}_i\ \boldsymbol{B}_i) - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N} \begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix} \boldsymbol{M}_{\boldsymbol{F}}(\boldsymbol{W}_k\ \boldsymbol{B}_k) \cdot \boldsymbol{\lambda}_i^\tau \left(\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N}\right)^{-1}\boldsymbol{\lambda}_k$, leads to

$$
\begin{aligned}
\begin{pmatrix} \hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\gamma} - \tilde{\gamma} \end{pmatrix} &= (LD(\hat{\boldsymbol{F}}) + \frac{L}{NT}\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}\frac{L}{NT}\sum_{i=1}^{N}\left(\begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik}\begin{pmatrix} \boldsymbol{W}_k^\tau \\ \boldsymbol{B}_k^\tau \end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\right)\boldsymbol{\varepsilon}_i \\
&+ (LD(\hat{\boldsymbol{F}}) + \frac{L}{NT}\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}L\nu_{NT} \\
&+ (LD(\hat{\boldsymbol{F}}) + \frac{L}{NT}\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}\left(O_p(L(NT)^{-1/2}) + O_p(L(T\sqrt{N})^{-1}) + N^{-1/2}O_p(L\delta_{NT}^{-2})\right) \\
&+ (LD(\hat{\boldsymbol{F}}) + \frac{L}{NT}\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}\left(O_p(LT^{-1/2}\zeta_{LD}^{1/2}) + O_p(L^{1/2}\zeta_{LD}^{1/2}) + LN^{-1/2}O_p(\varsigma) + O_p(L^{1/2}\varsigma)\right).
\end{aligned}
$$

where

$$
\begin{aligned}
\nu_{NT} = &-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}(\begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}} - \boldsymbol{R}_i)\boldsymbol{F}(\frac{\boldsymbol{F}^\tau\boldsymbol{F}}{T})^{-1}(\frac{\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}}{N})^{-1}\boldsymbol{\lambda}_k(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}) \\
&-\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\begin{pmatrix} \boldsymbol{W}_i^\tau \\ \boldsymbol{B}_i^\tau \end{pmatrix}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\sigma^2\boldsymbol{I}_T\hat{\boldsymbol{F}}\boldsymbol{G}\boldsymbol{\lambda}_i,
\end{aligned}
$$

and $\boldsymbol{R}_i = \frac{1}{N}\sum_{k=1}^{N} a_{ik}(\boldsymbol{W}_i\ \boldsymbol{B}_i)$.

By Lemma 2 and 3, it can be shown that $D(\hat{\boldsymbol{F}}) = D(\boldsymbol{F}) + o_p(1)$. In addition, by Lemma 1 and Lemma A.6 in Bai (2009), it is easy to verify that $\nu_{NT} = op(1)$. By Lemma 2, we have

$$\left\|(D(\boldsymbol{F}) + \frac{1}{NT}\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}\frac{1}{NT}\sum_{i=1}^{N}\left(\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}\boldsymbol{M}_{\boldsymbol{F}} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}\begin{pmatrix}\boldsymbol{W}_k^{\tau}\\ \boldsymbol{B}_k^{\tau}\end{pmatrix}\boldsymbol{M}_{\boldsymbol{F}}\right)\boldsymbol{\varepsilon}_i\right\|^2 = O_p(L^2/NT)$$

uniformly for $\boldsymbol{F}$. Therefore, under the assumptions, we have

$$\begin{aligned}\left\|\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}\end{pmatrix}\right\| =& O_p(L(NT)^{-1/2}) + O_p(L/T) + O_p(L/N)\\ &+ O_p(LT^{-1/2}\zeta_{LD}^{1/2}) + O_p(L^{1/2}\zeta_{LD}^{1/2}) + O_p(L^{1/2}\varsigma).\end{aligned}$$

Since $\hat{\beta}_j(s) = \sum_{l=1}^{L}\hat{\gamma}_{jl}b_l(s)$ and $\tilde{\beta}_j(s) = \sum_{l=1}^{L}\tilde{\gamma}_{jl}b_l(s)$, from (S0.1), we can get that

$$||\hat{\beta}_j(\cdot) - \beta_j(\cdot)||_{L_2}^2 \leq 2||\hat{\beta}_j(\cdot) - \tilde{\beta}_j(\cdot)||_{L_2}^2 + ML^{-2d}$$

and

$$||\hat{\beta}_j(\cdot) - \tilde{\beta}_j(\cdot)||_{L_2}^2 = ||\hat{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\gamma}}_j||_A^2 \asymp L^{-1}||\hat{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\gamma}}_j||^2, \quad j = 1, \ldots, q,$$

where $||\boldsymbol{\gamma}_j||_A^2 = \boldsymbol{\gamma}_j^{\tau}\boldsymbol{A}\boldsymbol{\gamma}_j$, and $\boldsymbol{A} = (a_{uv})_{L\times L}$ is a matrix with entries $a_{uv} =$

$\int_{\mathcal{S}} b_u(s)b_v(s)ds$. Then, it is easy to obtain that

$$||\hat{\boldsymbol{\beta}}(\cdot) - \tilde{\boldsymbol{\beta}}(\cdot)||_{L_2}^2 = \sum_{j=1}^{q} ||\hat{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\gamma}}_j||_A^2 \asymp L^{-1}||\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}||^2.$$

Therefore, under the assumption that the error is identically and independent distributed, we can get the convergence rate of both $\hat{\boldsymbol{\alpha}}$ and $\beta_j(\cdot)$ as

$$||\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}||_{L_2}^2 = O_p(L^2(NT)^{-1} + L^{-2d+1} + L\varsigma^2),$$

and

$$\left|\left|\hat{\beta}_j(\cdot) - \beta_j(\cdot)\right|\right|_{L_2}^2 = O_p(L(NT)^{-1} + L^{-2d} + \varsigma^2), \quad j = 1, \ldots, q.$$

□

Theorem 4

Proof. Under some appropriate relative rate for T and N and some assumptions, we have

$$\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}} \end{pmatrix} = \left(\frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{V}_i^{\tau} \boldsymbol{V}_i + \frac{1}{NT} \boldsymbol{G}_{\boldsymbol{\theta}} \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{V}_i^{\tau} \boldsymbol{\varepsilon}_i + o_p(1),$$

where $\tilde{\boldsymbol{\theta}}$ is defined before.

As $N, T \to \infty$ simultaneously, the conditional variance matrix $\boldsymbol{\Phi}_{\boldsymbol{\theta}} = Var(\hat{\boldsymbol{\theta}}|\mathcal{D})$ of $\hat{\boldsymbol{\theta}}$ conditioning on $\mathcal{D}$ is the limit in probability of

$$\boldsymbol{\Phi}^*_{\boldsymbol{\theta}} = (\sum_{i=1}^{N} \boldsymbol{V}_i^\tau \boldsymbol{V}_i + \boldsymbol{G}_{\boldsymbol{\theta}})^{-1} (\sum_{i=1}^{N} \sigma^2 \boldsymbol{V}_i^\tau \boldsymbol{V}_i)(\sum_{i=1}^{N} \boldsymbol{V}_i^\tau \boldsymbol{V}_i + \boldsymbol{G}_{\boldsymbol{\theta}})^{-1}.$$

The results from the proof of Theorem 3 indicates the convergency of both $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$ and $\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}$. Since the proof of the asymptotic property of these two parts are similar, we just take the second part as example. Let $\bar{\boldsymbol{\gamma}} = E(\hat{\boldsymbol{\gamma}}|\mathcal{D})$, $\boldsymbol{\Phi}_{\boldsymbol{\gamma}} = Var(\hat{\boldsymbol{\gamma}}|\mathcal{D})$, from the Theorem 4.1 in Huang (2003), and invoking Lemma A.8 in Huang et al. (2004), we obtain that, for any vector $\boldsymbol{a}$ with dimension qL and whose components are not all zero,

$$(\boldsymbol{a}^\tau \boldsymbol{\Phi}_{\boldsymbol{\gamma}} \boldsymbol{a})^{-1/2} \boldsymbol{a}^\tau (\hat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}) \xrightarrow{L} N(0,1).$$

Then, any q-vector $\boldsymbol{w}$ whose components are not all zero, letting $\boldsymbol{w} = \boldsymbol{B}(s)^\tau \boldsymbol{a}$, we have

$$(\boldsymbol{w}^\tau Var(\hat{\boldsymbol{\beta}}(s)|\mathcal{D})\boldsymbol{w})^{-1/2} \boldsymbol{w}^\tau (\hat{\boldsymbol{\beta}}(s) - \bar{\boldsymbol{\beta}}(s)) \xrightarrow{L} N(0,1),$$

which yields the asymptotic result.

□

Theorem 5

Proof. From the definition in the article, $\hat{\boldsymbol{\beta}}(s) = \boldsymbol{B}(s)\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\gamma}} = \boldsymbol{c}_{\boldsymbol{\gamma}}\hat{\boldsymbol{\theta}}$ where $\boldsymbol{B}(s) = \boldsymbol{b}^{\tau}(s) \otimes \boldsymbol{I}_q$, then

$$\bar{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s) = \boldsymbol{B}(s)(\bar{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}) + \boldsymbol{B}(s)\tilde{\boldsymbol{\gamma}} - \boldsymbol{\beta}(s).$$

From (S0.1), we can get that $||\boldsymbol{B}(s)\tilde{\boldsymbol{\gamma}} - \boldsymbol{\beta}(s)|| = O_p(\zeta_{LD}^{1/2})$.

Meanwhile, from simple calculation, we can get that

$$\bar{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}} = \boldsymbol{c}_{\boldsymbol{\gamma}} \left(\sum_{i=1}^{N} \boldsymbol{Z}_i^{\tau} \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Z}_i + \boldsymbol{G}_{\boldsymbol{\theta}} \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{Z}_i^{\tau} \boldsymbol{M}_{\hat{\boldsymbol{F}}} (\boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{e}_i) - \boldsymbol{G}_{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}} \right),$$

First, from the similar proof of Lemma A.9 in Huang et al. (2004), we can obtain that

$$\left|\left| \left(\frac{L}{NT} \sum_{i=1}^{N} \boldsymbol{Z}_i^{\tau} \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Z}_i + \boldsymbol{G}_{\boldsymbol{\theta}}\right)^{-1} \right|\right|_{\infty} \leq C.$$

Next, under assumption (A3) and (A4), from $\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F} = 0$ and Lemma

1, it is clear that

$$
\begin{aligned}
&|\frac{L}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{F}\boldsymbol{\lambda}_i|_{\infty}\\
=&|\frac{L}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}(\boldsymbol{M}_{\hat{\boldsymbol{F}}}-\boldsymbol{M}_{\boldsymbol{F}})\boldsymbol{F}\boldsymbol{\lambda}_i|_{\infty}=|\frac{L}{NT}\sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{W}_i^{\tau}\\ \boldsymbol{B}_i^{\tau}\end{pmatrix}(\boldsymbol{P}_{\boldsymbol{F}}-\boldsymbol{P}_{\hat{\boldsymbol{F}}})\boldsymbol{F}\boldsymbol{\lambda}_i|_{\infty}\\
=&O_p(L^{1/2}\zeta_{LD}^{1/2}).
\end{aligned}
$$

Meanwhile, since $\boldsymbol{M}_{\hat{\boldsymbol{F}}}$ is an idempotent matrix, then from Lemma A.6 in Huang et al. (2004) and Lemma 1, we can get that

$$
\begin{aligned}
&|\frac{L}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{e}_i|_{\infty}=|\frac{L}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{pmatrix}\boldsymbol{W}_{it}^{\tau}\\ \boldsymbol{B}_{it}^{\tau}\end{pmatrix}(\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{e}_i)_t|_{\infty}\\
\le&\max_{1\le k\le p}\max_{1\le j\le q}|\frac{L}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{W}_{itk}+\boldsymbol{B}_{itj})(\boldsymbol{e}_i^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{e}_i)^{1/2}|_{\infty}\\
\le&\max_{k,j}\left|\frac{L}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{W}_{itk}+\boldsymbol{B}_{itj})||\boldsymbol{e}_i||\right|_{\infty}=O_p(\zeta_{LD}^{1/2}).
\end{aligned}
$$

The assumption of smoothing parameter $\boldsymbol{\xi}$ leads to the result that $||\frac{L}{NT}\boldsymbol{G}_{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}||\infty=o_p(1)$.

Moreover, similar to the proof of Corollary 1 in Huang et al. (2004), under assumption (A1) and (A6), from Lemma 1, for coefficient function

$\beta_j(\cdot)$, we can get that

$$
\begin{aligned}
&\boldsymbol{w}_j^\tau \boldsymbol{B}(s)(\sum_{i=1}^N \boldsymbol{V}_i^\tau \boldsymbol{V}_i + \boldsymbol{G}_{\boldsymbol{\theta}})^{-1}(\sum_{i=1}^N \sigma^2 \boldsymbol{V}_i^\tau \boldsymbol{V}_i)(\sum_{i=1}^N \boldsymbol{V}_i^\tau \boldsymbol{V}_i + \boldsymbol{G}_{\boldsymbol{\theta}})^{-1} \boldsymbol{B}^\tau(s) \boldsymbol{w}_j \\
\gtrsim& C \frac{L}{NT} \sum_{l=1}^L b_l(s) \gtrsim \frac{L}{NT}.
\end{aligned}
$$

We can similarly get the result of $\boldsymbol{\alpha}$, which proves the theorem.

□

Supplementary Section E

In this section, we will provide detailed proofs for lemmas needed for theoretical results.

Proof of Lemma 1

Proof. By assumptions (A1) and (A5), it is easy to obtain that $\frac{1}{NT}\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{\varepsilon}_i = o_p(1)$. Using $\boldsymbol{P}_{\boldsymbol{F}} = \boldsymbol{F}^\tau \boldsymbol{F}/T$, we can get

$$
\frac{1}{NT}\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i = \frac{1}{NT}\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{\varepsilon}_i - \frac{1}{NT}\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{P}_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i.
$$

Then,

$$\frac{1}{NT}||\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{P}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i|| = \frac{1}{N}||\sum_{i=1}^{N}\frac{\boldsymbol{Z}_i^{\tau}\boldsymbol{F}}{T}\cdot\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\varepsilon_{it}|| \\ \leq \frac{1}{N}\sum_{i=1}^{N}||\frac{\boldsymbol{Z}_i^{\tau}\boldsymbol{F}}{T}||\cdot||\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\varepsilon_{it}||.$$

Since $T^{-1/2}||\boldsymbol{F}|| = \sqrt{r}$, then $T^{-1}||\boldsymbol{Z}_i^{\tau}\boldsymbol{F}|| \leq T^{-1}||\boldsymbol{Z}_i||||\boldsymbol{F}|| = \sqrt{r}T^{-1/2}||\boldsymbol{Z}_i|| \leq \sqrt{r}(\frac{1}{T}\sum_{t=1}^{T}||Z_{it}||^2)^{1/2}$. Then, using the Cauchy–Schwarz inequality, the above equation is bounded by

$$\sqrt{r}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}||Z_{it}||^2\right)^{1/2}\cdot\left(\frac{1}{N}\sum_{i=1}^{N}||\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\varepsilon_{it}||^2\right)^{1/2}.$$

By $T^{1/2}||\boldsymbol{Z}_i = O_p(1)$, the first term of the expression is $O_p(1)$. Similar to the proof of Lemma A.1 in Bai (2009), it is sufficient to show that the second term is $o_p(1)$ uniformly in $\boldsymbol{F}$ as follows.

$$\frac{1}{N}\sum_{i=1}^{N}||\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\varepsilon_{it}||^2 = \mathrm{tr}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{F}_t\boldsymbol{F}_s^{\tau}\varepsilon_{it}\varepsilon_{is}\right) \\ = \mathrm{tr}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{F}_t\boldsymbol{F}_s^{\tau}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{is}\right).$$

Since $T^{-1}\sum_{t=1}^{T}||\boldsymbol{F}_t||^2 = r$, then by the Cauchy-Schwarz inequality and assumption $(A5)$, we can obtain that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}||\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t\varepsilon_{it}||^2\\
\le&\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}||\boldsymbol{F}_t||^2||\boldsymbol{F}_s||^2\right)^{1/2}\cdot N^{-1/2}\cdot\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{is}]^2\right)^{1/2}\\
=&rN^{-1/2}O_p(1).
\end{aligned}$$

Therefore, it shows that $\sup_{\boldsymbol{F}}||\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i^{\tau}\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i|| = o_p(1)$.

The proofs for the remaining statements are similar to that of the first one, and hence omitted.

□

Proof of Lemma 2

Proof. Driven by Bai and Ng (2002), to make better explanation, we denote

$$\begin{aligned}
V(r,\hat{\boldsymbol{F}}_r) &= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\dot{\theta}_r)^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}_r}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\dot{\boldsymbol{\theta}}_r),\\
V(r,\boldsymbol{F}_0\boldsymbol{H}_r) &= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\dot{\theta}_r)^{\tau}\boldsymbol{M}_{\boldsymbol{F}_0\boldsymbol{H}_r}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\dot{\boldsymbol{\theta}}_r),
\end{aligned}$$

where $\hat{\boldsymbol{F}}_r$ is the estimator of $\boldsymbol{F}$ given the estimated number of factors r, $\boldsymbol{M}_{\hat{\boldsymbol{F}}_r} = \boldsymbol{I}-\boldsymbol{P}_{\hat{\boldsymbol{F}}_r}$ and $\boldsymbol{M}_{\boldsymbol{F}_0\boldsymbol{H}_r} = \boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r}$ are the corresponding idempotent matrix spanned by the space of $\hat{\boldsymbol{F}}_r$ and $\boldsymbol{F}_0\boldsymbol{H}_r$ respectively.

(1) From the definition, we can get that

$$V(r,\hat{\boldsymbol{F}}_r)-V(r,\boldsymbol{F}_0\boldsymbol{H}_r)=\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\dot{\theta}_r)^\tau(\boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r}-\boldsymbol{P}_{\hat{\boldsymbol{F}}_r})(\boldsymbol{Y}_i-\boldsymbol{Z}_i\dot{\boldsymbol{\theta}}_r).$$

Denote $\boldsymbol{D}_r=\hat{\boldsymbol{F}}_r^\tau\hat{\boldsymbol{F}}_r/T$ and $\boldsymbol{D}_0=\boldsymbol{H}_r^\tau\boldsymbol{F}_0^\tau\boldsymbol{F}_0\boldsymbol{H}_r/T$, then

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r}-\boldsymbol{P}_{\hat{\boldsymbol{F}}_r}=&\frac{1}{T}\hat{\boldsymbol{F}}_r(\frac{\hat{\boldsymbol{F}}_r^\tau\hat{\boldsymbol{F}}_r}{T})\hat{\boldsymbol{F}}_r^\tau-\frac{1}{T}\boldsymbol{F}_0\boldsymbol{H}_r(\frac{\boldsymbol{H}_r^\tau\boldsymbol{F}_0^\tau\boldsymbol{F}_0\boldsymbol{H}_r}{T})\boldsymbol{H}_r^\tau\boldsymbol{F}_0^\tau\\ =&\frac{1}{T}[\hat{\boldsymbol{F}}_r\boldsymbol{D}_r^{-1}\hat{\boldsymbol{F}}_r^\tau-\boldsymbol{F}_0\boldsymbol{H}_r\boldsymbol{D}_0^{-1}\boldsymbol{H}_r^\tau\boldsymbol{F}_0^\tau]\\ =&\frac{1}{T}[(\hat{\boldsymbol{F}}_r-\boldsymbol{F}_0\boldsymbol{H}_r+\boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{D}_r^{-1}(\hat{\boldsymbol{F}}_r-\boldsymbol{F}_0\boldsymbol{H}_r+\boldsymbol{F}_0\boldsymbol{H}_r)^\tau-\boldsymbol{F}_0\boldsymbol{H}_r\boldsymbol{D}_0^{-1}\boldsymbol{H}_r^\tau\boldsymbol{F}_0^\tau]\\ =&\frac{1}{T}[(\hat{\boldsymbol{F}}_r-\boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{D}_r^{-1}(\hat{\boldsymbol{F}}_r-\boldsymbol{F}_0\boldsymbol{H}_r)^\tau+(\hat{\boldsymbol{F}}_r-\boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{D}_r^{-1}\boldsymbol{H}_r^\tau\boldsymbol{F}_0^\tau\\ &+\boldsymbol{F}_0\boldsymbol{H}_r\boldsymbol{D}_r^{-1}(\hat{\boldsymbol{F}}_r-\boldsymbol{F}_0\boldsymbol{H}_r)^\tau+\boldsymbol{F}_0\boldsymbol{H}_r(\boldsymbol{D}_r^{-1}-\boldsymbol{D}_0^{-1})\boldsymbol{H}_r^\tau\boldsymbol{F}_0^\tau]\end{aligned}$$

Denote $\boldsymbol{C}_i=\boldsymbol{Y}_i-\boldsymbol{Z}_i\dot{\theta}_r$, then

$$V(r,\hat{\boldsymbol{F}}_r)-V(r,\boldsymbol{F}_0\boldsymbol{H}_r)=\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{C}_i^\tau(\boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r}-\boldsymbol{P}_{\hat{\boldsymbol{F}}_r})\boldsymbol{C}_i=\boldsymbol{J}_1+\boldsymbol{J}_2+\boldsymbol{J}_3+\boldsymbol{J}_4.$$

For $\boldsymbol{J}_1$, by Theorem 1 and Lemma 1, we can get that

$$\begin{aligned}\boldsymbol{J}_1 &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0})^{\tau}\boldsymbol{D}_r^{-1}(\hat{\boldsymbol{F}}_{s,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{s,0})C_{it}C_{is}\\ &\le \left(T^{-2}[\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0})^{\tau}\boldsymbol{D}_r^{-1}(\hat{\boldsymbol{F}}_{s,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{s,0})]^2\right)^{1/2}\cdot\left(T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}[N^{-1}\sum_{i=1}^{N}C_{it}C_{is}]^2\right)^{1/2}\\ &\le \left(T^{-1}\sum_{t=1}^{T}||\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0}||^2\right)\cdot||\boldsymbol{D}_r^{-1}||\cdot Op(1)=O_p(\delta_{NT}^{-2}).\end{aligned}$$

For $\boldsymbol{J_2}$,

$$\begin{aligned}\boldsymbol{J}_2 &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0})^{\tau}\boldsymbol{D}_r^{-1}\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{s,0}C_{it}C_{is}\\ &\le \left(T^{-2}[\sum_{t=1}^{T}\sum_{s=1}^{T}||\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0}||^2\cdot||\boldsymbol{D}_r^{-1}||^2\cdot||\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{s,0}||^2\right)^{1/2}\cdot\left(T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}[N^{-1}\sum_{i=1}^{N}C_{it}C_{is}]^2\right)^{1/2}\\ &\le \left(T^{-1}\sum_{t=1}^{T}||\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0}||^2\right)^{1/2}\cdot||\boldsymbol{D}_r^{-1}||\cdot\left(T^{-1}\sum_{t=1}^{T}||\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{s,0}||^2\right)^{1/2}\cdot Op(1)\\ &= \left(T^{-1}\sum_{t=1}^{T}||\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0}||^2\right)^{1/2}\cdot Op(1)=O_p(\delta_{NT}^{-1})\end{aligned}$$

The same result can be obtained for $\boldsymbol{J}_3$. For $\boldsymbol{J}_4$,

$$\begin{aligned}\boldsymbol{J}_4 =&\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{F}_{t,0}^{\tau}\boldsymbol{H}_r^{\tau}(\boldsymbol{D}_r^{-1}-\boldsymbol{D}_0^{-1})\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{s,0}C_{it}C_{is}\\ &\le N^{-1}||\boldsymbol{D}_r^{-1}-\boldsymbol{D}_0^{-1}||\cdot\sum_{i=1}^{N}\left(T^{-1}\sum_{t=1}^{T}||\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0}||\cdot|C_{it}|\right)^2\\ &= ||\boldsymbol{D}_r^{-1}-\boldsymbol{D}_0^{-1}||\cdot Op(1),\end{aligned}$$

where $N^{-1}\sum_{i=1}^{N}\bigg(T^{-1}\sum_{t=1}^{T}||\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0}||\cdot|C_{it}|\bigg)^2$ is bounded by $N^{-1}T^{-2}||\boldsymbol{H}_r||^2\sum_{t=1}^{T}||\boldsymbol{F}_{t,0}||^2\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}|C_{it}|^2$ from the Assumptions.

Moreover, for the first term in $\boldsymbol{J}_4$,

$$
\begin{aligned}
||\boldsymbol{D}_r^{-1}-\boldsymbol{D}_0^{-1}|| &= ||\frac{\hat{\boldsymbol{F}}_r^{\tau}\hat{\boldsymbol{F}}_r}{T}-\frac{\boldsymbol{H}_r^{\tau}\boldsymbol{F}_0^{\tau}\boldsymbol{F}_0\boldsymbol{H}_r}{T}|| \\
&= T^{-1}||\sum_{t=1}^{T}(\hat{\boldsymbol{F}}_{t,r}\hat{\boldsymbol{F}}_{t,r}^{\tau}-\boldsymbol{H}_r^{\tau}\boldsymbol{F}_{t,0}\boldsymbol{F}_{t,0}^{\tau}\boldsymbol{H}_r)|| \\
&\le T^{-1}||\sum_{t=1}^{T}(\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0})(\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0})^{\tau}|| \\
&\quad + T^{-1}||\sum_{t=1}^{T}(\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0})\hat{\boldsymbol{F}}_{t,0}^{\tau}\boldsymbol{H}_r|| \\
&\quad + T^{-1}||\sum_{t=1}^{T}\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0}(\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0})^{\tau}|| \\
&\le T^{-1}\sum_{t=1}^{T}||\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0}||^2 \\
&\quad + 2\bigg(T^{-1}\sum_{t=1}^{T}||\hat{\boldsymbol{F}}_{t,r}-\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0}||^2\bigg)^{1/2}\cdot\bigg(T^{-1}\sum_{t=1}^{T}||\boldsymbol{H}_r^{\tau}\hat{\boldsymbol{F}}_{t,0}||^2\bigg)^{1/2} \\
&= O_p(\delta_{NT}^{-2})+O_p(\delta_{NT}^{-1})=O_p(\delta_{NT}^{-1}).
\end{aligned}
$$

Since the rank of $\boldsymbol{H}_r$ is $r\le r_0$, and from the above conclusion, D_0 and D_r both converges to a positive definite matrix, leading to the result that $||D_r||^{-1}=O_p(1)$. Meanwhile, since $\boldsymbol{D}_r^{-1}-\boldsymbol{D}_0^{-1}=\boldsymbol{D}_r^{-1}(\boldsymbol{D}_0-\boldsymbol{D}_r)\boldsymbol{D}_0^{-1}$, then $\boldsymbol{J}_4=O_p(\delta_{NT}^{-1})$. Therefore, for $1\le r\le r_0$, we can get that $V(r,\hat{\boldsymbol{F}}_r)-$

$V(r, \boldsymbol{F}_0\boldsymbol{H}_r) = O_p(\delta_{NT}^{-1})$.

(2)

$$
\begin{aligned}
V(r, \boldsymbol{F}_0\boldsymbol{H}_r) - V(r_0, \boldsymbol{F}_0) &= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\dot{\theta}_r)^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r})(\boldsymbol{Y}_i - \boldsymbol{Z}_i\dot{\boldsymbol{\theta}}_r) \\
&= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{F}_0\boldsymbol{\lambda}_{i,0} + \boldsymbol{\varepsilon}_i)^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r})(\boldsymbol{F}_0\boldsymbol{\lambda}_{i,0} + \boldsymbol{\varepsilon}_i) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0}^{\tau}\boldsymbol{F}_0^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r})\boldsymbol{F}_0\boldsymbol{\lambda}_{i,0} \\
&\quad + \frac{2}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r})\boldsymbol{F}_0\boldsymbol{\lambda}_{i,0} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r})\boldsymbol{\varepsilon}_i \\
&= \boldsymbol{J}_1 + \boldsymbol{J}_2 + \boldsymbol{J}_3.
\end{aligned}
$$

Then, we can get that

$$
\begin{aligned}
\boldsymbol{J}_1 &= tr\left(\frac{1}{NT}\boldsymbol{F}_0^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r})\boldsymbol{F}_0\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}_{i,0}^{\tau}\right) \\
&= tr\left(T^{-1}\left[\boldsymbol{F}_0^{\tau}\boldsymbol{F}_0 - \boldsymbol{F}_0^{\tau}\boldsymbol{F}_0\boldsymbol{H}_r(\boldsymbol{H}_r^{\tau}\boldsymbol{F}_0^{\tau}\boldsymbol{F}_0\boldsymbol{H}_r)^{-1}\boldsymbol{H}_r^{\tau}\boldsymbol{F}_0^{\tau}\boldsymbol{F}_0\right]\cdot N^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}_{i,0}^{\tau}\right) \\
&= tr(\boldsymbol{D}_{\boldsymbol{F}}\cdot\boldsymbol{D}_{\boldsymbol{\lambda}}),
\end{aligned}
$$

since the limit of matrix $\boldsymbol{D}_{\boldsymbol{F}}$ is semi-positive and $\boldsymbol{D}_{\boldsymbol{\lambda}}$ is positive, then $\boldsymbol{J}_1 \geq 0$.

For $\boldsymbol{J}_2 = \frac{2}{NT}\sum_{i=1}^{N}\varepsilon_i^{\tau}(\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r})\boldsymbol{F}_0\boldsymbol{\lambda}_{i,0}$, from Lemma 1,

$$|\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}\boldsymbol{P}_{\boldsymbol{F}_0}\boldsymbol{F}_0\boldsymbol{\lambda}_{i,0}| = |\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\boldsymbol{F}_{t,0}^{\tau}\boldsymbol{\lambda}_{i,0}|$$
$$\leq \left(T^{-1}\sum_{t=1}^{T}||\boldsymbol{F}_{t,0}||^2\right)^{1/2}\cdot\left(T^{-1}\sum_{t=1}^{T}||N^{-1/2}\sum_{i=1}^{N}||^2\varepsilon_{it}\boldsymbol{\lambda}_{i,0}\right)^{1/2} = O_p(N^{-1/2}),$$

similarly, it is easy to get that $|\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i^{\tau}\boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r}\boldsymbol{F}_0\boldsymbol{\lambda}_{i,0}| = O_p(N^{-1/2})$. Hence, it is obvious that $\boldsymbol{J}_3 \geq 0$.

For $\boldsymbol{J}_3$, it is obvious that $\boldsymbol{J}_3 \geq 0$ because of $\boldsymbol{P}_{\boldsymbol{F}_0} - \boldsymbol{P}_{\boldsymbol{F}_0\boldsymbol{H}_r} \geq 0$. Therefore, for each $1 \leq r < r_0$, we can get the conclusion that there assist a $c_r > 0$, such that plim $\inf_{N,T\to\infty}[V(r, \boldsymbol{F}_0\boldsymbol{H}_r) - V(r_0, \boldsymbol{F}_0)] = c_r$.

(3)

$$|V(r,\hat{\boldsymbol{F}}_r) - V(r_0,\hat{\boldsymbol{F}}_{r_0})| \leq |V(r,\hat{\boldsymbol{F}}_r) - V(r_0,\boldsymbol{F}_0)| + |V(r_0,\boldsymbol{F}_0) - V(r_0,\hat{\boldsymbol{F}}_{r_0})|$$
$$2\max_{r_0\leq r\leq r_{max}}|V(r,\hat{\boldsymbol{F}}_r) - V(r_0,\boldsymbol{F}_0)|.$$

Define $\boldsymbol{H}_r$ as in Theorem 1 with rank $r_0 \geq r$, and $\boldsymbol{H}_r^+$ be the generalized inverse of $\boldsymbol{H}_r$ such that $\boldsymbol{H}_r\boldsymbol{H}_r^+ = \boldsymbol{I}_{r_0}$. Meanwhile, since $\boldsymbol{C}_i = \boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta} = \boldsymbol{F}_0\boldsymbol{\lambda}_{i,0} + \boldsymbol{\varepsilon}_i$, then it is easy to get that

$$\boldsymbol{C}_i = \hat{\boldsymbol{F}}_r\boldsymbol{H}_r\boldsymbol{H}_r^+\boldsymbol{\lambda}_{i,0} + \boldsymbol{\varepsilon}_i - (\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{H}_r^+\boldsymbol{\lambda}_{i,0} = \hat{\boldsymbol{F}}_r\boldsymbol{H}_r\boldsymbol{H}_r^+\boldsymbol{\lambda}_{i,0} + \boldsymbol{e}_i,$$

where $\boldsymbol{e}_i = \boldsymbol{\varepsilon}_i - (\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{H}_r^+\boldsymbol{\lambda}_{i,0}$.

Then, from the following equations

$$V(r, \hat{\boldsymbol{F}}_r) = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{e}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}_r}\boldsymbol{e}_i,$$
$$V(r_0, \boldsymbol{F}_0) = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{\varepsilon}_i^\tau \boldsymbol{M}_{\boldsymbol{F}_{r_0}}\boldsymbol{\varepsilon}_i,$$

we can get

$$\begin{aligned} V(r, \hat{\boldsymbol{F}}_r) &= (NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\varepsilon}_i - (\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{H}_r^+\boldsymbol{\lambda}_{i,0})^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}_r}(\boldsymbol{\varepsilon}_i - (\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{H}_r^+\boldsymbol{\lambda}_{i,0}) \\ &= (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}_r}\boldsymbol{\varepsilon}_i - 2(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0}^\tau(\boldsymbol{H}_r^+)^\tau(\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}_r}\boldsymbol{\varepsilon}_i \\ &\quad + (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0}^\tau(\boldsymbol{H}_r^+)^\tau(\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}_r}(\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{H}_r^+\boldsymbol{\lambda}_{i,0} \\ &= \boldsymbol{J}_1 + \boldsymbol{J}_2 + \boldsymbol{J}_3. \end{aligned}$$

For $\boldsymbol{J}_2$, by the fact that $|tr(\boldsymbol{A})| \leq r||\boldsymbol{A}||$ for any $r \times r$ matrix $\boldsymbol{A}$,

$$\begin{aligned}
\boldsymbol{J}_2 &= 2T^{-1}tr\Bigg(\boldsymbol{H}_r^{+}(\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}_r}(N^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_{i,0})\Bigg) \\
&\leq 2r \cdot ||\boldsymbol{H}_r^{+}|| \cdot ||T^{-1/2}(\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)|| \cdot ||T^{-1/2}N^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_{i,0}|| \\
&\leq 2r \cdot ||\boldsymbol{H}_r^{+}|| \cdot \Bigg(T^{-1}\sum_{t=1}^{T}||\hat{\boldsymbol{F}}_{t,r} - \boldsymbol{H}_r^\tau\boldsymbol{F}_{t,0}||^2\Bigg)^{1/2} \cdot N^{1/2}\Bigg(T^{-1}\sum_{t=1}^{T}||N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{it}\boldsymbol{\lambda}_{i,0}||^2\Bigg)^{1/2} \\
&= O_p(\delta_{NT}^{-1}) \cdot N^{1/2} = O_p(\delta_{NT}^{-2}).
\end{aligned}$$

For $\boldsymbol{J}_3$, by Theorem 1,

$$\begin{aligned}
\boldsymbol{J}_3 &\leq (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0}^\tau(\boldsymbol{H}_r^{+})^\tau(\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)^\tau(\hat{\boldsymbol{F}}_r - \boldsymbol{F}_0\boldsymbol{H}_r)\boldsymbol{H}_r^{+}\boldsymbol{\lambda}_{i,0} \\
&\leq T^{-1}\sum_{t=1}^{T}||\hat{\boldsymbol{F}}_{t,r} - \boldsymbol{H}_r^\tau\boldsymbol{F}_{t,0}||^2 \cdot \Bigg(N^{-1}\sum_{i=1}^{N}||\boldsymbol{\lambda}_{i,0}||^2||\boldsymbol{H}_r^{+}||^2\Bigg) \qquad = O_p(\delta_{NT}^{-2}) \cdot O_p(1).
\end{aligned}$$

Then, $V(r, \hat{\boldsymbol{F}}_r) = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{M}_{\hat{\boldsymbol{F}}_r}\boldsymbol{\varepsilon}_i + O_p(\delta_{NT}^{-2})$. From the fact that for $r \geq r_0$, $V(r, \hat{\boldsymbol{F}}_r) - V(r_0, \boldsymbol{F}_{r_0}) \leq 0$, we can get that

$$V(r, \hat{\boldsymbol{F}}_r) - V(r_0, \boldsymbol{F}_{r_0}) = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{M}_{\hat{\boldsymbol{F}}_r}\boldsymbol{\varepsilon}_i - (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^\tau\boldsymbol{M}_{\boldsymbol{F}_{r_0}}\boldsymbol{\varepsilon}_i + O_p(\delta_{NT}^{-2}).$$

Meanwhile,

$$
\begin{aligned}
(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}\boldsymbol{M}_{\boldsymbol{F}_{r_0}}\boldsymbol{\varepsilon}_i &\leq ||(\boldsymbol{F}_{r_0}^{\tau}\boldsymbol{F}_{r_0}/T)^{-1}|| \cdot N^{-1}T^{-2}sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}\boldsymbol{F}_{r_0}\boldsymbol{F}_{r_0}^{\tau}\boldsymbol{\varepsilon}_i \\
&= O_p(1)\cdot(NT)^{-1}\sum_{i=1}^{N}||T^{-1/2}\sum_{t=1}^{T}\boldsymbol{F}_{t,r_0}\boldsymbol{\varepsilon}_{it}||^2 \\
&= O_p(T^{-1}) \leq O_p(\delta_{NT}^{-2}).
\end{aligned}
$$

Then, $(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}_r}\boldsymbol{\varepsilon}_i + O_p(\delta_{NT}^{-2}) \leq 0$, leading to the fact that $(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}_r}\boldsymbol{\varepsilon}_i = O_p(\delta_{NT}^{-2})$. Hence, $V(r,\hat{\boldsymbol{F}}_r)-V(r_0,\boldsymbol{F}_{r_0}) = O_p(\delta_{NT}^{-2})$.

Therefore, for any $r \geq r_0$, we can ge the result that $V(r,\hat{\boldsymbol{F}}_r)-V(r_0,\hat{\boldsymbol{F}}_{r_0}) = O_p(\delta_{NT}^{-2})$.

□

Proof of Lemma 3

Proof. (i) From $\left[\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}})(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}})^{\tau}\right]\hat{\boldsymbol{F}} = \hat{\boldsymbol{F}}\boldsymbol{V}_{NT}$ and $\boldsymbol{Y}_i -$

$\boldsymbol{Z}_i\hat{\boldsymbol{\theta}} = \boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i$, we obtain the following expansion as

$$
\begin{aligned}
\hat{\boldsymbol{F}}\boldsymbol{V}_{NT} =& \left[\frac{1}{NT}\sum_{i=1}^{N}\Big(\boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i\Big)\Big(\boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{e}_i\Big)^{\tau}\right]\hat{\boldsymbol{F}} \\
=& \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^{\tau}\boldsymbol{Z}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^{\tau}\boldsymbol{Z}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^{\tau}\boldsymbol{Z}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{Z}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^{\tau}\boldsymbol{Z}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{\varepsilon}_i^{\tau}\hat{\boldsymbol{F}} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{e}_i^{\tau}\hat{\boldsymbol{F}} \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} \\
=:& \boldsymbol{I}_1 + \ldots + \boldsymbol{I}_{16},
\end{aligned}
$$

where $\boldsymbol{I}_{16} = \dfrac{1}{NT}\sum_{i=1}^{N}\boldsymbol{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}} = \boldsymbol{F}(\frac{\boldsymbol{\Lambda}^{\tau}\boldsymbol{\Lambda}}{N})^{-1}\frac{\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}}}{T}$.

Then, it is easy to get that

$$
\hat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{H} = (\boldsymbol{I}_1 + \ldots + \boldsymbol{I}_{16})\boldsymbol{V}_{NT}^{-1}.
$$

By the fact that $T^{-1/2}||\hat{\boldsymbol{F}}|| = \sqrt{r}$ and $||\boldsymbol{Z}_i|| = O_p(T^{1/2} + T^{1/2}L^{-1})$, we can get that

$$T^{-1/2}||I_1|| \leq N^{-1}\sum_{i=1}^{N}(||\boldsymbol{Z}_i||^2/T)||\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}||^2\sqrt{r} = O_p(||\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}||^2)+O_p(\varsigma) = O_p(\varsigma),$$

$$T^{-1/2}||I_2|| \leq N^{-1}\sum_{i=1}^{N}(||\boldsymbol{Z}_i||/\sqrt{T})||\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}||\cdot||\boldsymbol{\lambda}_i||\cdot||\boldsymbol{F}^{\tau}\hat{\boldsymbol{F}}/T|| = O_p(||\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}||)+O_p(\varsigma^{1/2}).$$

Similarly, we can get that $T^{-1/2}||I_j|| = O_p(||\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}||) + O_p(\varsigma^{1/2})$ for $j = 3, 4, 5$, and $T^{-1/2}||I_j|| = O_p(\delta_{NT}^{-1}) + O_p(\varsigma^{1/2})$ for $j = 6, 7, 8$.

For I_9,

$$\begin{aligned} T^{-1/2}||I_9|| &\leq N^{-1}T^{-1/2}\sum_{i=1}^{N}(||\boldsymbol{Z}_i||/\sqrt{T})||\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}|| \cdot ||\hat{\boldsymbol{F}}/\sqrt{T}|| \cdot (\sum_{t=1}^{T}\varepsilon_{it}^2) \\ &= O_p(||\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}||) \cdot M\zeta_{LD}^{1/2} + O_p(\varsigma^{1/2}). \end{aligned}$$

Similarly, we can prove that $T^{-1/2}||I_{10}|| = O_p(||\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}||) \cdot M\zeta_{LD}^{1/2} + O_p(\varsigma^{1/2})$.

For I_{11},

$$\begin{aligned} T^{-1/2}||I_{11}|| &\leq N^{-1}T^{-1/2}\sum_{i=1}^{N}(||\boldsymbol{F}||/\sqrt{T}.||\boldsymbol{\lambda}_i|| \cdot (r\sum_{t=1}^{T}\varepsilon_{it}^2) \\ &= O_p(\zeta_{LD}^{1/2}). \end{aligned}$$

Similarly, we can prove that $T^{-1/2}||I_{12}|| = O_p(\zeta_{LD}^{1/2})$.

For I_{13},

$$T^{-1/2}||I_{13}|| \leq N^{-1}T^{-1}\sum_{i=1}^{N}||\boldsymbol{\varepsilon}_i|| \cdot (r\sum_{t=1}^{T}\varepsilon_{it}^2)$$
$$= O_p(\zeta_{LD}^{1/2}\delta_{NT}^{-1}).$$

Similarly, we can prove that $T^{-1/2}||I_{14}|| = O_p(\zeta_{LD}^{1/2}\delta_{NT}^{-1})$

For I_{15},

$$T^{-1/2}||I_{15}|| \leq N^{-1}T^{-1}\sum_{i=1}^{N}\sqrt{r} \cdot (\sum_{t=1}^{T}\varepsilon_{it}^2)$$
$$= O_p(\zeta_{LD}).$$

Therefore, following the proof of Proposition A.1 in Bai (2009), we can get the conclusion that

$$T^{-1/2}||\hat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{H}|| = O_p(||\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}||) + O_p(\delta_{NT}^{-1}) + Op(\zeta_{LD}^{1/2}) + O_p(\varsigma^{1/2}).$$

The similar proof of the other parts in this Lemma can be found in Feng et al. (2018).

□

Proof of Lemma 4

Proof. (i) From the discussion of Lemma 3, we have

$$T^{-1}\boldsymbol{\varepsilon}_k^{\tau}(\hat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{H}) = T^{-1}\boldsymbol{\varepsilon}_k^{\tau}(I_1 + \cdots + I_{15})\boldsymbol{V}_{NT}^{-1}.$$

For the first eight terms, using the similar arguments in the proof of Lemma A.4 in Bai (2009) and the results obtained in the Lemma 3, we can get that

$$T^{-1}\boldsymbol{\varepsilon}_k^{\tau}(I_1 + \cdots + I_8)\boldsymbol{V}_{NT}^{-1} = T^{-1/2}O_p(||\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}||) + O_p(\delta_{NT}^{-1}) + O_p(\varsigma^{1/2}).$$

Meanwhile, for the other terms, applying the similar proof of (i) in Lemma 3, it is easy to get that the order of dominant term as

$$\begin{aligned} ||T^{-1}\boldsymbol{\varepsilon}_k^{\tau}I_{11}\boldsymbol{V}_{NT}^{-1}|| &\le T^{-1} \cdot ||\boldsymbol{\varepsilon}_k^{\tau}\boldsymbol{F}||/\sqrt{T} \cdot N^{-1}\sum_{i=1}^{N}||\boldsymbol{\lambda}_i|| \cdot ||\boldsymbol{V}_{NT}^{-1}|| \cdot (r\sum_{t=1}^{T}\varepsilon_{it}^2)^{1/2} \\ &= O_p(T^{-1/2}\zeta_{LD}^{1/2}). \end{aligned}$$

Similarly, $||T^{-1}\boldsymbol{\varepsilon}_k^{\tau}I_{12}\boldsymbol{V}_{NT}^{-1}|| = O_p(T^{-1/2}\zeta_{LD}^{1/2})$.

Therefore, together with the above discussion, we can get that $T^{-1}\boldsymbol{\varepsilon}_k^{\tau}(\hat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{H}) = T^{-1/2}O_p(||\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}||) + O_p(\delta_{NT}^{-2}) + O_p(T^{-1/2}\zeta_{LD}^{1/2}) + O_p(\varsigma)$.

The similar proof of the other parts in this Lemma can be found in Feng et al. (2018).

□

Bibliography

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77:1229–1279.

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

Chumaker, L. (1981). *Spline functions: Basis theory*. New York: Wiley.

Feng, S., Li, G., Peng, H., and Tong, T. (2018). Varying coefficient panel data model with interactive fixed effects. *arXiv preprint*, page arXiv:1803.02714.

Huang, J. Z. (2003). Local asymptotic for polynomial spline regression. *The Annals of Statistics*, 31:1600–1635.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14:763–788.