Appendix for "Inferential Theory for Pricing Errors with Latent Factors and Firm Characteristics"

Jungjun Choi and Ming Yuan

July 23, 2026

Abstract

This appendix includes additional numerical results and proofs of theorems.

APPENDIX

A Table for characteristics

Table A.1: Firm Characteristics

Symbol	Description
INV	Investment, percentage year-on-year growth rate of total assets.
DPI2A	Changes in PPE and inventory scaled by lagged AT.
NOA	Net operating assets: operating assets minus operating liabilities, scaled by lagged AT.
LBM	Log book-to-market ratio of equity.
S2P	Sales-to-price ratio: net sales, scaled by market equity.
STREV	Short-term reversal: one-month return.
Q	Tobin's Q: market equity plus book debt, divided by AT.
IMOM	Intermediate momentum: cumulative return from month -12 to month -7.
PROF	Profitability: gross profitability scaled by book equity.
MOM	Momentum: cumulative return from month -12 to month -2.
OL	Operating leverage: sum of cost of goods sold and SG&A expenses scaled by AT.
D2A	Depreciation and amortization scaled by AT.
LME	Log market capitalization: log of market equity.
BIDASK	Bid-ask spread: average daily bid-ask spread in the month.
LTREV	Long-term reversal: cumulative return from month -36 to month -13.
LEV	Leverage: (LTD+DCL)/(LTD+CL+BE) with LTD: long-term debt, DCL: debt in current liabilities, and BE: book equity.
CTO	Capital turnover: ratio of net sales to lagged total assets.
CA	Cash and short-term investment, scaled by total assets.
SGA2S	SG&A-to-sales: selling, general and administrative expenses scaled by net sales.
AT	Total assets.
ATO	Net sales scaled by lagged, unscaled net operating assets.
FC2Y	Fixed costs-to-sales: SG&A plus advertising expenses and R&D expenses, divided by net sales.
E2P	Earnings-to-price ratio: income before extraordinary items divided by lagged market equity.
FCF	Net income, depreciation, and amortization less change in working capital and capital expenditure, scaled by book equity.
PM	Profit margin: operating income after depreciation scaled by sales.
LTURN	Turnover: dollar trading volume over market equity.
A2ME	Assets-to-market cap: total assets over lagged market equity.
ROE	Income before extraordinary items scaled by lagged book equity.
BETA	Market beta: market beta estimated with past one year's daily data.
SUV_m	Standardized unexplained volume: Standardized residual from regressing trading volume on absolute values of positive and negative returns.
OA	Operating accruals: changes in noncash working capital minus depreciation, scaled by lagged AT.
ROA	Return on assets: income before extraordinary items divided by lagged AT.
PCM	Price-to-cost margin: net sales minus costs of goods sold, scaled by net sales.
RNA	Return on net operating assets: operating income after depreciation scaled by lagged, unscaled net operating assets.
W52H	Stock price relative to its 52-week high price.
IVOL	Idiosyncratic volatility with respect to the Fama–French three-factor model.

B Time-invariant δ_o

Here, we provide the asymptotic theory for the case in which δ_o is time-invariant. In this setting, we do not require a sub-Gaussian assumption on the noise and allow for weak dependence across i.

Assumption B.1 (Noise)**.** Conditioning on $(x_{it}, f_{t+1})_{i\leq N, t\leq T}$, the following conditions hold: (i) $\mathbb{E}[\epsilon_{it+1}] = 0$ and $\mathbb{E}[\epsilon_{it+1}^4] \leq C_1$ for some constant $C_1 > 0$; (ii) there exists a

constant $C_2 > 0$ such that

$$\max_{i,t} \sum_{s=2}^{T+1} |\mathrm{Cov}(\epsilon_{it}, \epsilon_{is})| \leq C_2, \quad \max_{i,t} \sum_{j=1}^{N} |\mathrm{Cov}(\epsilon_{it}, \epsilon_{jt})| \leq C_2,$$

and

$$\max_{i,t} \sum_{s=2}^{T+1} \sum_{j=1}^{N} |\mathrm{Cov}(\epsilon_{it}, \epsilon_{js})| \leq C_2.$$

Assumption B.2 (Central Limit Theorem)**.** Conditioning on $(x_{it}, f_{t+1})_{i \leq N, t \leq T}$, we have

$$\frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{s=1}^{T} h_{js} \epsilon_{j,s+1} \to_d \mathcal{N}(0, \sigma_{o,it}^2), \qquad \text{where } h_{js} = \left\| B_{t,i}^o \right\|^{-1} B_{t,i}^{o\top} B_{s,j}^o,$$

for some positive constant $\sigma_{o,it}^2$.

For example, if ϵ_{it} are independent across i and t, with $\mathbb{E}[\epsilon_{it}] = 0$ and $\mathrm{Var}(\epsilon_{it}) = \sigma_t^2$, then this condition is satisfied by the Lindeberg central limit theorem, with limiting variance

$$\sigma_{o,it}^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{s=1}^{T} \sigma_{s+1}^2.$$

We then obtain the following asymptotic normality result.

Theorem B.1. *Suppose that Assumptions 1, 2, B.1, and B.2 hold. Then,*

$$\mathcal{V}_{o,it}^{-1/2} \left(\hat{\alpha}_{o,it} - \alpha_{o,it} \right) \to_d \mathcal{N}(0, 1),$$

where

$$\mathcal{V}_{o,it} = \sigma_{o,it}^2 \left\| B_{t,i}^o \right\|^2 / (NT).$$

Because

$$\frac{1}{N} \left\| B_t^o \right\|_F^2 = \frac{1}{N} \sum_{i=1}^{N} \left\| B_{t,i}^o \right\|^2 = N - L,$$

if $\left\| B_{t,i}^o \right\|^2$ is of similar magnitude across i, then the convergence rate of the estimator is roughly $1/\sqrt{T}$. Since the dimension of the parameter δ_o is large, namely $N - L$, we cannot achieve a faster convergence rate comparable to that for estimating $\alpha_{I,it}$.

C Comparison with KPS model

C.1 Model difference

The KPS model in Kelly et al. (2019) can be written in the following vector form:

$$r_{t+1} = \alpha_t + B_t f_{t+1} + E_{t+1}, \tag{C.1}$$

where

$$\alpha_t = X_t \eta, \qquad B_t = X_t \Gamma.$$

Because X_t contains a constant term, the KPS model may also be viewed as allowing a certain type of outside alpha, depending on whether one interprets the constant term as a characteristic. Specifically, when the first column of X_t is $\mathbf{1}_N$, that is, when $X_t = [\mathbf{1}_N \;\; X_t^\star]$,

$$\alpha_t = X_t \eta = \mathbf{1}_N \eta_1 + X_t^\star \eta^\star,$$

where $\eta = [\eta_1 \;\; \eta^{\star\top}]^\top$ and $\eta^\star$ is an $(L-1) \times 1$ vector. In this case, we may define the inside and outside alpha components as

$$\alpha_{I,t} = X_t^\star \eta^\star, \qquad \text{and} \qquad \alpha_{O,t} = \mathbf{1}_N \eta_1.$$

However, under this representation, the outside alpha is always equal to η_1 for all assets and all periods, which is quite restrictive. In particular, it cannot capture either the time variation or the cross-sectional heterogeneity of outside alphas. Moreover, this model does not satisfy the APT orthogonality condition $\alpha_t^\top B_t = 0$, since

$$\alpha_t^\top B_t = \eta^\top X_t^\top X_t \Gamma \neq 0.$$

By contrast, our model is given by

$$\alpha_t = \alpha_{I,t} + \alpha_{O,t}, \qquad B_t = X_t \Gamma,$$

with

$$\alpha_{I,t} = (I_N - P_{B_t}) X_t \eta, \qquad \text{and} \qquad \alpha_{O,t} = B_t^o \delta_{o,t},$$

where $P_{B_t} = B_t (B_t^\top B_t)^{-1} B_t^\top$, B_t^o is an $N \times (N-L)$ basis matrix for the subspace orthogonal to X_t, and $\delta_{o,t}$ is the corresponding coefficient vector.

Thus, the basic form in (C.1) and the specification of B_t are the same as in the KPS

model, whereas the specification of α_t is substantially different. First, $\alpha_{O,it}$ is allowed to vary across both assets and time, so our model can accommodate both the dynamics and the heterogeneity of outside alphas.

Moreover, since

$$B_t^\top (I_N - P_{B_t}) = 0 \qquad \text{and} \qquad X_t^\top B_t^o = 0,$$

our model satisfies the APT orthogonality condition $\alpha_t^\top B_t = 0$. As discussed in Section 5.4, this condition has an important economic interpretation.

In summary, our model allows for a much more general specification of outside alpha, including both time variation and cross-sectional heterogeneity. In addition, it satisfies the APT orthogonality condition $\alpha_t^\top B_t = 0$, which is not satisfied in the KPS model. At the same time, the two models share the same basic structure in (C.1) and the same specification for B_t.

C.2 Empirical study

We conduct an empirical analysis using the method in Kelly et al. (2019) on our data. We begin with the alpha tests. Table C.1 reports the p-values for the outside alpha test under various choices of the rank K. As shown in the table, except when the rank is 4, we cannot reject the null hypothesis at the 5% significance level. Moreover, at the 1% significance level, we cannot reject the null hypothesis in any case. In contrast, our method successfully detects the presence of outside pricing error in all cases.

Table C.1: Outside alpha test

	Rank				
Method	1	2	3	4	5*
Our method	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
KPS	0.740	0.929	0.077	0.025	0.231
	Rank				
Method	6	7	8	9	10
Our method	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
KPS	0.107	0.616	0.844	0.505	0.620

Table C.2 reports the p-values for the inside alpha test under various choices of the rank K. At the 1% significance level, the null hypothesis is rejected only when $K \leq 2$. When the rank exceeds 2, the method in Kelly et al. (2019) fails to detect the presence of inside pricing error. At the 5% significance level, the null hypothesis is rejected only when

$K = 1, 2, 6, 8$. For example, when the rank is set equal to the estimated one ($K = 5$), the KPS method does not detect inside pricing error. In contrast, our method successfully detects the presence of inside pricing error in all cases.

Table C.2: Inside alpha test

	Rank				
Method	1	2	3	4	5*
Our method	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
KPS	0.002	0.006	0.196	0.065	0.051
	Rank				
Method	6	7	8	9	10
Our method	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
KPS	0.017	0.089	0.042	0.225	0.053

In addition, the alpha and beta estimates obtained from the KPS method do not satisfy the APT orthogonality condition. When the rank is set equal to the estimated one ($K = 5$), the time average of $\left\|\hat{\alpha}_t^\top \hat{B}_t\right\|$ is 549.3988, whereas that of our method is 0. As a result, the pricing error may in fact load on systematic factors, which undermines its interpretation as pure mispricing.

Next, Figure C.1 illustrates the difference in the outside alpha estimators. As shown in the figure, the KPS method cannot capture either the time variation or the cross-sectional heterogeneity of outside alphas, whereas our method captures both. In addition, Figure C.2 compares the inside alpha estimators. We can see that the dynamics of the KPS estimator differ substantially from those of our estimator. Note again that the alpha estimates from the KPS method do not satisfy the APT orthogonality condition, which weakens their interpretation as pure mispricing.

In addition, Figure C.3 visualizes the estimated $\widehat{\Gamma}$ from the KPS method when $K = 5$, whereas Figure 9 in the main text is based on our method. Because of the rotational indeterminacy of factor models, even if two estimators share the same target Γ, they may differ by an invertible rotation matrix. For this reason, a direct comparison between Figure C.3 and Figure 9 of the main text is not fully rigorous. Still, a naive comparison suggests some similarities. For example, the loading vector for Factor 2 in the KPS estimator has large coefficients on fixed costs-to-sales (FC2Y) and SG&A-to-sales (SGA2S), which resembles the loading pattern for Factor 2 in our method, up to a sign change. Such a sign difference is not meaningful because of rotational indeterminacy. Likewise, the loading

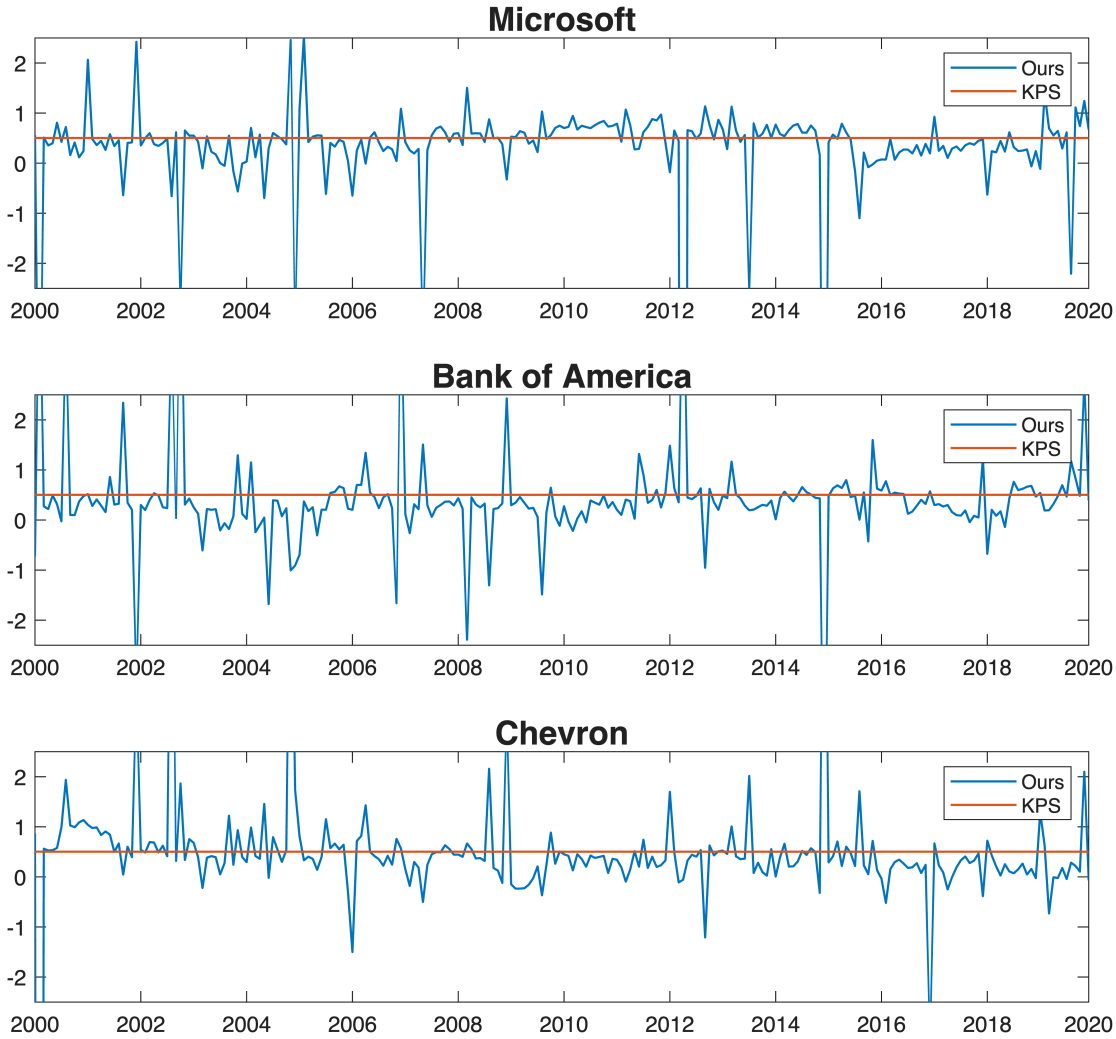


Figure C.1: Dynamics of α_O over time

vector for Factor 3 in the KPS estimator has large coefficients on total assets (AT) and log market capitalization (LME), which is similar to the loading vector for Factor 3 in our method. However, the loading patterns for the other factors quite differ. In fact, because of rotational indeterminacy, there is no reason to expect them to be close.

Additionally, Table C.3 reports the variables that are significant under the 5%-level Gamma test for the two methods, while Table C.4 reports the corresponding p-values for the KPS method when $K = 5$. As shown in these tables, the Gamma test based on the KPS method identifies far fewer significant variables than our method. This pattern suggests that, in this empirical application, the KPS bootstrap-type Gamma test rejects fewer null hypotheses than our proposed Gamma test, especially when pricing errors are present.

Moreover, we compare the total R^2 and predictive R^2 across the two methods. When $K = 5$, the total R^2 of the KPS method is 25.95%, whereas that of our method is 14.36%.

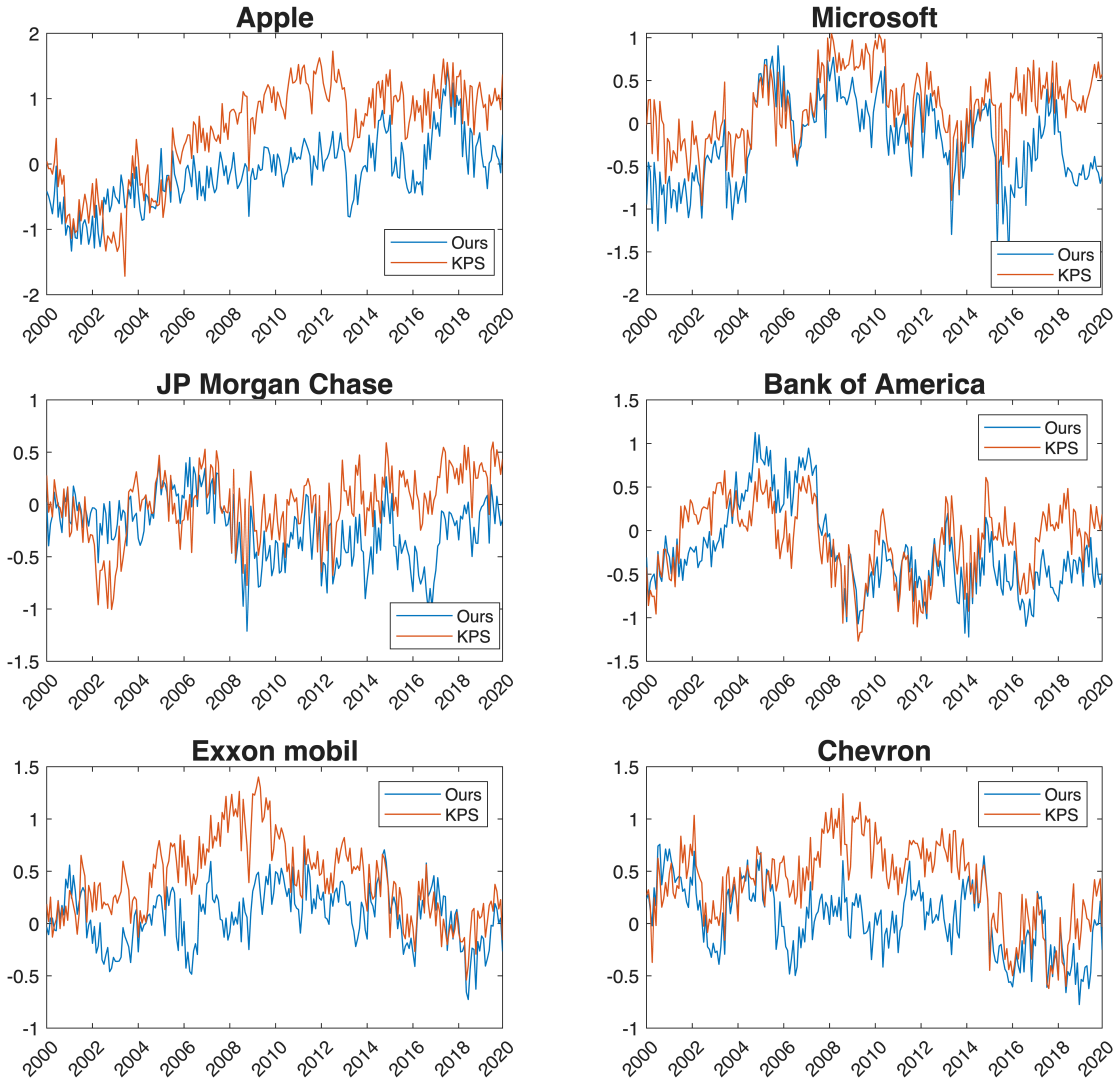


Figure C.2: Dynamics of α_I over time

Thus, in terms of total R^2, the KPS method performs better in this case. However, when $K = 10$, the total R^2 of the KPS method is 27.65%, while that of our method is 27.21%, so the two methods perform similarly. By contrast, for predictive R^2, when $K = 5$, the KPS method achieves 1.12%, whereas our method achieves 1.37%. Hence, our method performs slightly better in terms of predictive R^2. The results are similar when $K = 10$.

Lastly, Table C.5 reports the performance of the alpha portfolios. Here, the portfolio weights are given by α_{t-1}, where α_{t-1} is estimated using observations available up to time $t-1$. The out-of-sample evaluation period is the second half of our sample, namely from January 2010 to December 2019. For our method, we use the time-invariant δ_o model so that α_{t-1} remains close to α_t. As the table shows, the standard deviation under our method is much smaller than that under the KPS method. As discussed in Section 5.4,

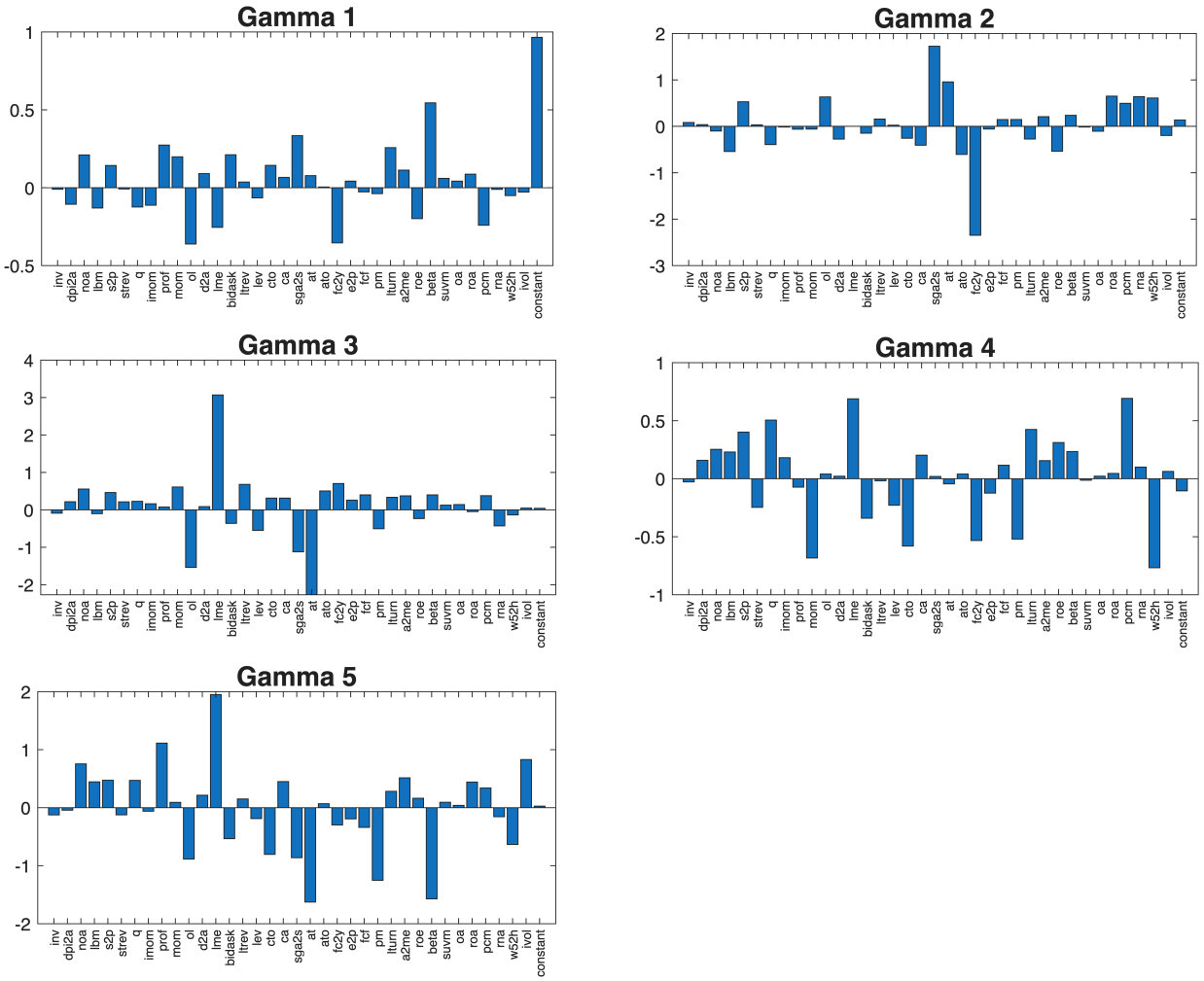


Figure C.3: Estimated $\widehat{\Gamma}$ from the KPS method

Table C.3: Significant variables in the Gamma test (5% level)

Method	Significant variables
Our method	lbm, strev, q, mom, ol, lme, bidask, ltrev, cto, sga2s, at, ato, fc2y, e2p, pm, lturn, a2me, roe, beta, pcm, rna, constant
KPS method	beta, ivol, constant

Table C.4: p-values of the Gamma test in the KPS method

Variable	p-value	Variable	p-value	Variable	p-value	Variable	p-value
INV	0.698	OL	0.285	ATO	0.326	OA	0.421
DPI2A	0.437	D2A	0.353	FC2Y	0.188	ROA	0.433
NOA	0.078	LME	0.104	E2P	0.569	PCM	0.425
LBM	0.629	BIDASK	0.259	FCF	0.302	RNA	0.218
S2P	0.276	LTREV	0.129	PM	0.255	W52H	0.169
STREV	0.479	LEV	0.197	LTURN	0.240	IVOL	0.050
Q	0.740	CTO	0.516	A2ME	0.502	Constant	0.000
IMOM	0.654	CA	0.062	ROE	0.495		
PROF	0.132	SGA2S	0.202	BETA	0.005		
MOM	0.489	AT	0.146	SUV_m	0.392		

Table C.5: Out-of-sample annualized Sharpe ratios for the alpha portfolios

Method	Mean	Standard deviation	Sharpe ratio
Our method	157.69	394.6	1.384
KPS method	403.03	1460.6	0.956

this is because our alpha portfolio has no exposure to the common factors due to the APT orthogonality condition, whereas the alpha portfolio constructed from the KPS method does not satisfy this property. As a result, although the average monthly (excess) return under the KPS method is higher than that under our method, the Sharpe ratio of our method is about 44.8% higher than that of the KPS method.

In summary, both methods have their respective strengths and weaknesses. Our method yields smaller p-values in the alpha tests and a better Sharpe ratio for the alpha portfolios, whereas the KPS model achieves a higher R^2.

D Simulated experiment

To demonstrate the finite sample performance of our methodology and the validity of our inferential theory, we conducted a simulation experiment.

D.1 Finite sample performance of inferential theory

We calibrate the simulated data to our model estimated from US monthly stock returns in the empirical study and set $N = 973$, $T = 240$, $L = 37$, and $K = 5$. For the parameters Γ and η, we use the estimated values from the model. For the characteristics, we first estimate the $(L-1) \times (L-1)$ covariance matrix Σ_x from the characteristics data excluding

the constant term, generate x_{it} from $\mathcal{N}(0, \Sigma_x)$, and include the constant. To generate the sparse ξ_t, we first randomly choose 71 periods and for each t in the chosen periods, we draw 3 values from $uniform[\xi_{center} - 0.5, \xi_{center} + 0.5]$ where ξ_{center} is the average of maximum and minimum of absolute value of nonzero estimated $\xi_{t,q}$s from the stock returns data. Then, we assign these values to 3 randomly chosen elements in ξ_t and set other elements to be zero. For the remaining (240 - 71) periods, we set ξ_t to be zero. Lastly, we set a sign of element randomly. Here, 71 is the number of periods where $\hat{\xi}_t \neq \mathbf{0}$ in the empirical study and 3 is the average of the number of nonzero $\hat{\xi}_{t,q}$ over the periods where $\hat{\xi}_t \neq \mathbf{0}$. In addition, for ζ, we use the estimated ζ from the empirical study. In addition, for each $1 \leq k \leq 5$, we generate $\breve{f}_{t+1,k}$ from a normal distribution whose mean and variance are the estimated values from the data. Lastly, we generate $\epsilon_{i,t+1}$ from $\mathcal{N}(0, \sigma^2)$ where σ is estimated from the data. The number of simulations is set to 1,000.

First, to study the advantage of debiased estimators for Γ and α_I, we compare the histograms and kernel density estimates of the t-statistics of the plain Γ estimator and the debiased Γ estimator. For the inference of Γ, we report the results of $\gamma_{1,1}$. For the inference of $\alpha_{I,it}$, we report the results of a randomly chosen i at the last period T.

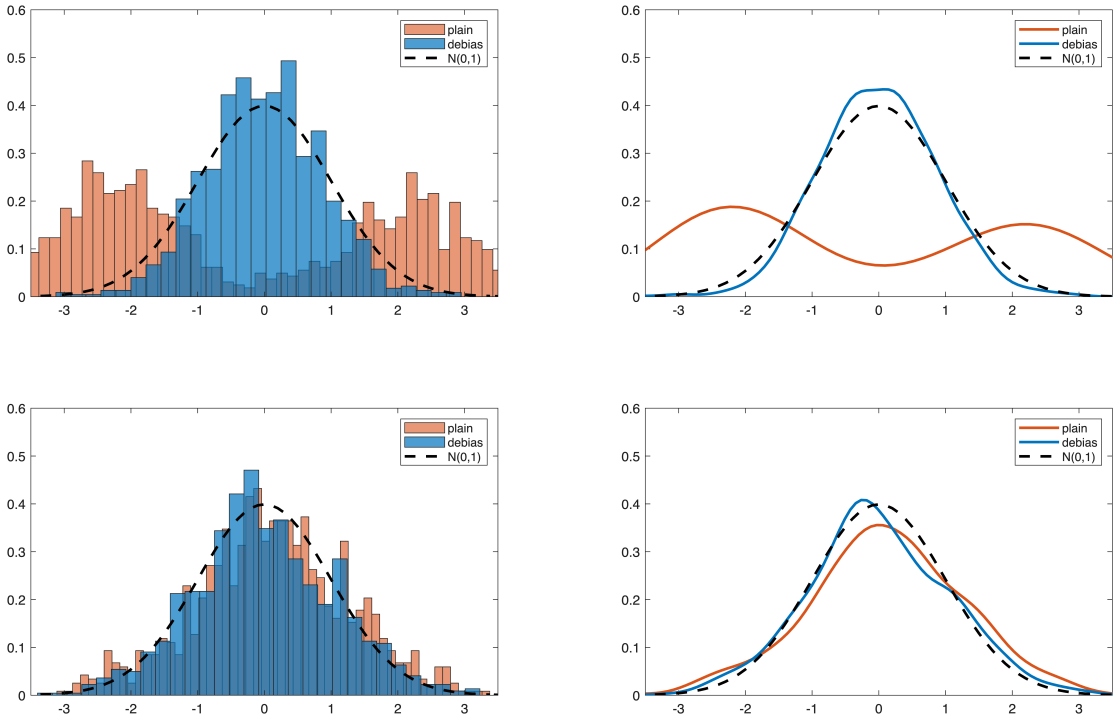


Figure D.1: Histograms for standardized estimates

Figure D.1 shows the histograms and kernel density estimates of the t-statistics. The upper panel shows the results for the Γ estimates and the the lower panel shows those for

the α_I estimates. In the case of Γ estimates, the plain estimator has a fairly large bias. Here, the sign of bias depends on that of rotation in each iteration. In addition, we can also check that the plain estimator of α_I is more biased compared to the debiased estimator. Besides, Table D.1 shows the coverage probabilities of the (asymptotic) confidence intervals. From the table, we can confirm that the coverage probabilities of the debiased estimators are closer to the target probabilities.

Table D.1: Coverage probability of the confidence interval.

Target prob.	90%	95%	99%
$\tilde{\alpha}_{I,it}$	83.8%	89.9%	96.5%
$\hat{\alpha}_{I,it}$	87.8%	93.4%	98.5%
$\tilde{\gamma}_{1,1}$	28.2%	40.6%	67.5%
$\hat{\gamma}_{1,1}$	94.7%	97.6%	99.3%

Next, we present the coverage probabilities of the confidence interval and the histograms for the standardized estimates (t-statistics) for α_O. For the inference of $\alpha_{O,it}$, we report the results of a randomly chosen i at the last period T. In the simulations, we always make T to be in the 71 chosen periods so that $\xi_T \neq \mathbf{0}$. Here, for the truncation level, we set $\rho = 1.5 \times \sigma \frac{\sqrt{\log NT}}{\sqrt{N}}$. For the inference of $\delta_{o,t,q}$, we report the results of a randomly chosen q at the last period T.

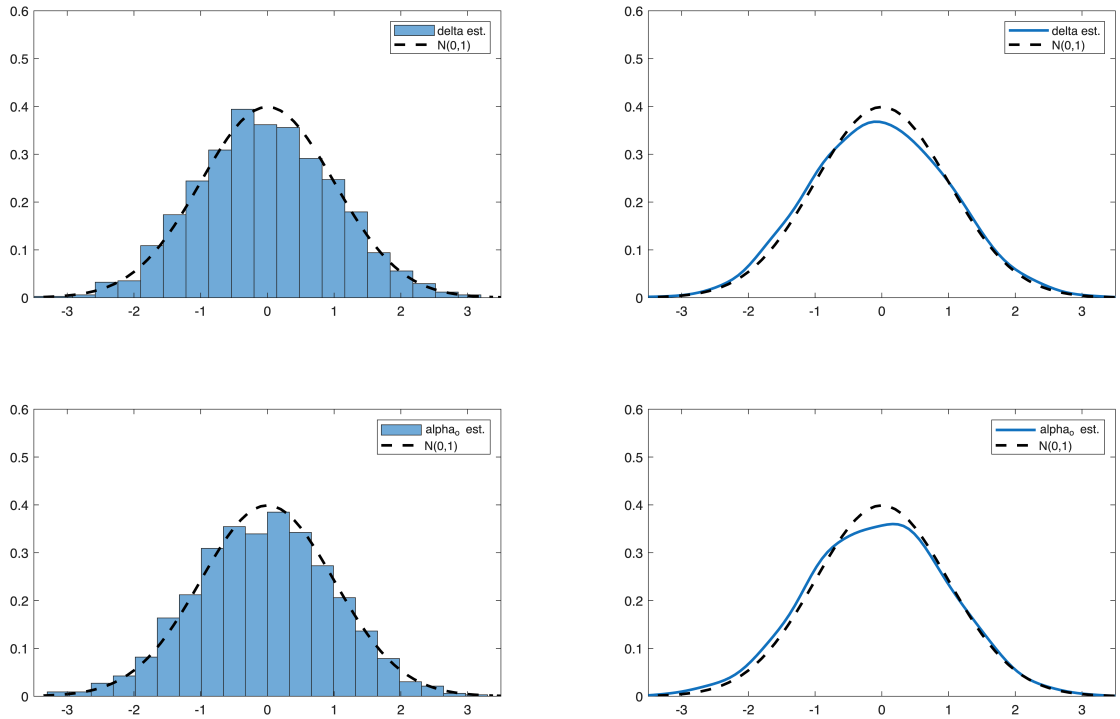


Figure D.2: Histograms for standardized estimates

Figure D.2 shows the histograms of the standardized estimates (t-statistics) along with the standard normal distribution and Table D.2 shows the coverage probabilities of the (asymptotic) confidence intervals. From the figure and table, we can confirm the asymptotic normality of our estimates.

Table D.2: Coverage probability of the confidence interval

Parameter	$\delta_{o,t,q}$			$\alpha_{O,it}$		
Target prob.	90%	95%	99%	90%	95%	99%
Coverage prob.	89.3%	95.3%	99.1%	89.8%	95.0%	98.7%

D.2 Power comparison with other methods

In this section, to show the relative advantage of our inference method, we compare the power of our alpha test with that of other methods in the case where the true model is close to the null hypothesis. Specifically, we consider the following model:

$$R_{t+1} = \alpha_{O,t} + B_t f_{t+1} + E_{t+1}, \qquad t = 1, \ldots, T,$$

where $B_t = X_t\Gamma$, $\alpha_{O,t} = B_t^o\delta_o$ and test the null hypothesis that there is no outside pricing error, $\alpha_{O,t}$. Here, the inside pricing error $\alpha_{I,t}$ is set to zero and δ_o is time invariant, so that the model can belong to both the model of Zhang (2024) and that of this paper. In addition, as in Zhang (2024), we define the basis B_t^o such that

$$B_t^o = X_t^o(X_t^{o\top}X_t^o/N)^{-1/2}, \quad X_t^o = [I_N - P_{X,t}]\begin{pmatrix} I_{N-L} \\ \mathbf{0}_{L\times(N-L)} \end{pmatrix}.$$

For the characteristics, we set $L = 10$ and generate x_{it} from the standard normal distribution and include the constant. For the factors, we set $K = 2$ and for each t, generate factors from $\mathcal{N}(0, \mathrm{diag}(2,1)^2)$. In addition, we generate each element of Γ from $\mathcal{N}(0, 1/L)$ and fix it for all iterations. Noises are generated from a standard normal distribution. On top of that, to generate the case where the true model is close to the null hypothesis, but, different from it, we set $\delta_1 = 0.01 \sim 0.06$ and $\delta_q = 0$ for all $2 \le q \le N - L$.

Figure D.3 shows the rejection probabilities of several tests when $N = T = 200$. Here, 'Our (formula)' uses the test statistics $\max_{i\le N, t\le T}\left|\hat{V}_{o,it}^{-1/2}\hat{\alpha}_{O,it}\right|$ and for the critical values,

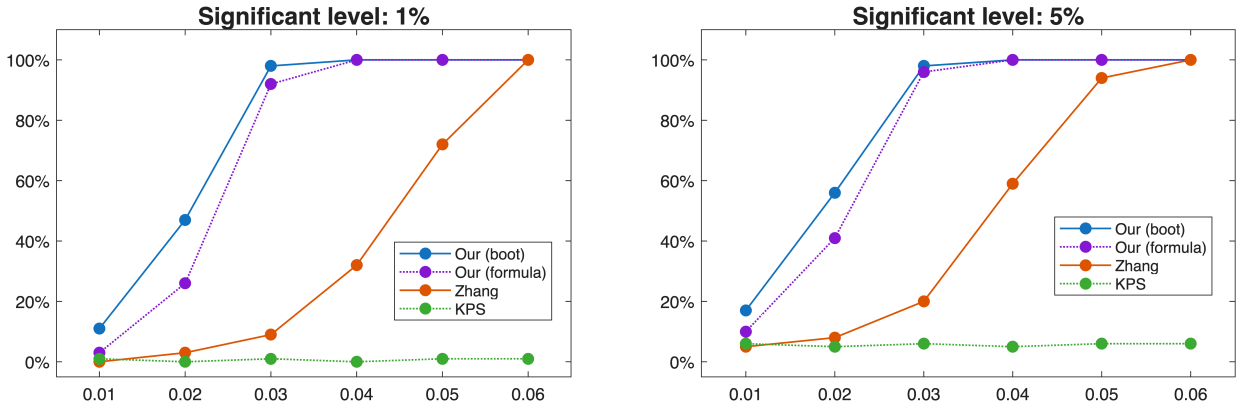


Figure D.3: Rejection probabilities with diverse δ_1

utilizes the relation

$$\mathbb{P}\left(\max_{i\leq N,t\leq T}\left|\hat{V}_{o,it}^{-1/2}\hat{\alpha}_{O,it}\right| > \Phi^{-1}(1-a/(2NT))\right) \leq a + o(1).$$

under the null. On the other hand, 'Ours (boot)' utilizes the bootstrap method in Belloni et al. (2018) and Chernozhuokov et al. (2022) to derive the distribution of the test statistics and the corresponding critical values, while it uses the same test statistics, $\max_{i\leq N,t\leq T}\left|\hat{V}_{o,it}^{-1/2}\hat{\alpha}_{O,it}\right|$. In addition, 'Zhang' denotes the method in Zhang (2024) using the test statistics $||\hat{\delta}_o||$ with the bootstrap method where $\hat{\delta}_o$ is the estimator from the iterative numerical algorithm. Lastly, 'KPS' denotes the alpha test method introduced in Kelly et al. (2019). Here, the number of simulations is 100 and the number of bootstraps is set to 500.

We can check that our methods have more power than other methods. In the case of the method in Kelly et al. (2019), it cannot detect the existence of alpha because their model only considers the inside pricing error. In addition, when it comes to the inference method in Zhang (2024), it seems to be weak in the tough case where the true model is close to the null. Among our methods, the bootstrap method is slightly better than the method based on the formula using the inverse cumulative distribution function.

Moreover, Tables D.3 and D.4 show the rejection probabilities with other sample sizes when the level is 99% and 95%, respectively. We can find similar patterns to Figure D.3. Our methods show the highest rejection probabilities and have better power than others. On the other hand, the method in Kelly et al. (2019) show the lowest rejection probabilities.

Table D.3: Rejection probabilities with level 99%

Sample Size (N,T)	Inference method	δ_1 0.01	0.02	0.03	0.04	0.05
(500,200)	Ours (boot)	18%	96%	100%	100%	100%
	Ous (formula)	4%	82%	100%	100%	100%
	Zhang	0%	0%	12%	56%	97%
	KPS	0%	0%	0%	0%	0%
(300,300)	Ours (boot)	7%	96%	100%	100%	100%
	Ous (formula)	0%	86%	100%	100%	100%
	Zhang	0%	0%	26%	90%	100%
	KPS	0%	0%	0%	0%	1%
(200,500)	Ours (boot)	32%	98%	100%	100%	100%
	Ous (formula)	15%	94%	100%	100%	100%
	Zhang	0%	10%	73%	100%	100%
	KPS	2%	1%	0%	0%	1%

D.3 Comparison with Kelly et al. (2019) (KPS)

In this section, we conduct a simple simulation study to illustrate the shortcomings of the KPS model. Specifically, we define the basis B_t^o by

$$B_t^o = X_t^o (X_t^{o\top} X_t^o / N)^{-1/2}, \qquad X_t^o = [I_N - P_{X,t}] \begin{pmatrix} I_{N-L} \\ \mathbf{0}_{L\times(N-L)} \end{pmatrix}.$$

For the characteristics, we set $L = 10$, generate x_{it} from the standard normal distribution, and include a constant term. For the factors $\breve{f}_{t+1,k}$, we set $K = 2$ and, for each t, generate them from $\mathcal{N}(0, \mathrm{diag}(2,1)^2)$. In addition, each element of Γ is generated from $\mathcal{N}(0, 1/L)$, η is generated from $Unif[-0.5, 0.5]$, and δ_o is generated from $\mathcal{N}(0, 1/(N-L))$; these quantities are then fixed across all simulation iterations. The noise terms are generated from the standard normal distribution. The number of simulation iterations is set to 100, and we take $N = T = 400$.

Table D.5 reports the outside alpha estimates for the first 10 stocks in the first and last periods. We can see that the KPS outside alpha estimate is 0.49 for all stocks and all periods. Therefore, it cannot capture either the time variation or the cross-sectional heterogeneity of outside alphas. In contrast, our method successfully captures both features.

In addition, Table D.6 reports the root mean squared error of the outside alpha estimators and the magnitude of $\|\hat{\alpha}_t^\top \hat{B}_t\|$. Because the KPS estimator cannot capture either the time variation or the cross-sectional heterogeneity of outside alphas, its RMSE is much

Table D.4: Rejection probabilities with level 95%

Sample Size (N,T)	Inference method	δ_1 0.01	0.02	0.03	0.04	0.05
(500,200)	Ours (boot)	22%	100%	100%	100%	100%
	Ours (formula)	8%	93%	100%	100%	100%
	Zhang	0%	8%	37%	86%	100%
	KPS	6%	8%	8%	10%	10%
(300,300)	Ours (boot)	15%	100%	100%	100%	100%
	Ours (formula)	3%	96%	100%	100%	100%
	Zhang	0%	12%	59%	100%	100%
	KPS	3%	4%	3%	4%	2%
(200,500)	Ours (boot)	41%	99%	100%	100%	100%
	Ours (formula)	19%	94%	100%	100%	100%
	Zhang	0%	27%	97%	100%	100%
	KPS	6%	3%	5%	6%	4%

Table D.5: Outside alpha ($\alpha_{O,it}$) of the first 10 stocks

Time	Method	Stocks (i) 1	2	3	4	5	6	7	8	9	10
$t=1$	Ours	-1.37	2.95	0.76	0.04	0.69	-0.11	0.02	1.63	1.62	1.73
	KPS	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	Target	-1.32	3.00	0.83	0.00	0.71	0.04	-0.01	1.56	1.66	1.58
$t=T$	Ours	-1.32	3.07	0.67	-0.02	0.67	-0.21	-0.07	1.89	1.71	1.61
	KPS	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	Target	-1.27	3.10	0.74	-0.06	0.70	-0.07	-0.10	1.81	1.76	1.45

larger than that of our estimator. Moreover, the KPS estimator does not satisfy the APT orthogonality condition, which has an important economic interpretation, as discussed in Section 5.4.

Table D.6: RMSEs for outside alpha and the APT orthogonality condition

	RMSE for outside alpha	Orthogonality condition, $\|\hat{\alpha}_t^\top \hat{B}_t\|$
Ours	0.05	0.00
KPS	1.12	28.39

E Interpretation of outside alpha

Our intention is not to interpret the outside alpha as a directly observable time-series average abnormal return in the same way as a conventional portfolio alpha. Rather, it represents the component of the conditional expected return that lies outside the span generated by the characteristics. In particular, in our model,

$$\mathbb{E}[\alpha_{O,t} \mid X_t] = \mathbb{E}[B_t^o \delta_{o,t} \mid X_t] = B_t^o \delta_{o,t} = \alpha_{O,t},$$

because B_t^o is a function of X_t, as noted in (1) of the main text. Hence, $\alpha_{O,t}$ is a component of the conditional expected return, rather than a residual term like E_{t+1} in our model. In this sense, $\alpha_{O,t}$ may be viewed as part of the pricing error.

Our use of the term "alpha" is motivated by the asset-pricing interpretation that this component captures the part of expected return that is not explained by the characteristic-related beta structure. As discussed in Zhang (2024), there may be several sources of such outside alpha. First, the true asset-pricing model may be rational, but not of the simple linear beta-pricing form. For example, some consumption-based or investment-based asset-pricing models imply more complicated, possibly nonlinear relationships between characteristics and expected returns. Second, investor biases may lead prices to deviate from rational asset-pricing predictions; if the model does not adequately capture these biases, the remaining discrepancy appears as outside pricing error. Third, the model may omit important return predictors; if relevant firm characteristics are excluded, their effects may also be absorbed into the outside pricing error. See Section 2.2 of Zhang (2024) for further discussion. These considerations help motivate an economic interpretation of the outside alpha.

Moreover, the time-invariant component ζ in $\delta_{o,t}$ introduces coherence over time in the outside alpha, which makes it difficult to regard $\alpha_{O,t}$ as a purely idiosyncratic residual. From this perspective, the outside alpha can be interpreted as an economically meaningful pricing-error object.

F Variance estimators

Here, we present the variance estimators we used in the empirical study and simulated experiment. We estimate the variances in our inferential theory under the assumption that

the noises are independent across i and t, and there is heteroskedasticity across t. We may also be able to consider more robust estimators as in Newey and West (1987); Hansen (2007); Bai et al. (2020).

Let $\hat{\sigma}^2_{t+1} = \frac{1}{N}\sum_{i=1}^N \hat{\epsilon}^2_{i,t+1}$, where $\hat{m}_{i,t+1} = \hat{\alpha}_{O,it} + \hat{\alpha}_{I,it} + x_{it}^\top \hat{\Gamma}\hat{f}_{t+1}$, and $\hat{\epsilon}_{i,t+1} = r_{i,t+1} - \hat{m}_{i,t+1}$. Then, we define

$$(\boldsymbol{H}^\top \widehat{\Sigma_f^{-1}\Sigma_{xf,l}\Sigma_f^{-1}}\boldsymbol{H}) = \left(\frac{1}{T}\sum_t \hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\right)^{-1} \frac{1}{T}\sum_{t=1}^T \hat{\sigma}^2_{t+1}[Q_t^{-1}]_{ll}\hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\left(\frac{1}{T}\sum_t \hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\right)^{-1},$$

$$\hat{\sigma}^2_{I,it} = \frac{1}{TL}\sum_{s=1}^T \hat{\sigma}^2_{s+1}(\hat{a}_s\hat{A} - \hat{b}_s\hat{B})Q_s(\hat{A}^\top\hat{a}_s - \hat{B}^\top\hat{b}_s),$$

$$\hat{a}_s = 1 - (\hat{\eta}^\top Q_t\hat{\Gamma}(\hat{\Gamma}^\top Q_t\hat{\Gamma})^{-1} + (\hat{\bar{f}})^\top)\left(\frac{1}{T}\sum_{u=1}^T \hat{f}^d_{u+1}\hat{f}^{d\top}_{u+1}\right)^{-1}\hat{f}^d_{s+1}, \quad \hat{\bar{f}} = (\hat{\Gamma}^\top\hat{\Gamma})^{-1}\hat{\Gamma}^\top\bar{\bar{R}},$$

$$\hat{b}_s = x_{it}^\top\hat{\Gamma}(\hat{\Gamma}^\top Q_t\hat{\Gamma})^{-1}\left(\frac{1}{T}\sum_{u=1}^T \hat{f}^d_{u+1}\hat{f}^{d\top}_{u+1}\right)^{-1}\hat{f}^d_{s+1},$$

$$\hat{A} = x_{it}^\top Q_t^{-1} - x_{it}^\top\hat{\Gamma}(\hat{\Gamma}^\top Q_t\hat{\Gamma})^{-1}\hat{\Gamma}^\top, \quad \hat{B} = \hat{\eta}^\top - \hat{\eta}^\top Q_t\hat{\Gamma}(\hat{\Gamma}^\top Q_t\hat{\Gamma})^{-1}\hat{\Gamma}^\top,$$

$$\hat{\bar{\sigma}}^2 = \frac{1}{T}\sum_{s=1}^T \hat{\sigma}^2_{s+1}, \quad \hat{\sigma}^2_{\delta,qt} = \hat{\sigma}^2_{t+1}.$$

Then, the following theorem shows the consistency of the variance estimators.

Theorem F.1. *Assume that the noises are independent across i and t, and there is heteroskedasticity across t. Then, we have*
(i) $\left\|(\boldsymbol{H}^\top \widehat{\Sigma_f^{-1}\Sigma_{xf,l}\Sigma_f^{-1}}\boldsymbol{H}) - \boldsymbol{H}^\top\Sigma_f^{-1}\Sigma_{xf,l}\Sigma_f^{-1}\boldsymbol{H}\right\| = o_p(1)$; *(ii)* $\left|\hat{\sigma}^2_{I,it} - \sigma^2_{I,it}\right| = o_p(1)$; *(iii)* $\left|\hat{\bar{\sigma}}^2 - \bar{\sigma}^2\right| = o_p(1)$; *(iv)* $\left|\hat{\sigma}^2_{\delta,qt} - \sigma^2_{\delta,qt}\right| = o_p(1)$ *and* $\left|\hat{\sigma}^2_{t+1} - \sigma^2_{t+1}\right| = o_p(1)$.

Here, we can use the variance estimators using 'tilde ($\tilde{}$)' estimators instead of the current estimators using 'hat ($\hat{}$)' estimators and will get the same result.

G Dependence on units of characteristics in Zhang (2024)

Zhang (2024) uses the following transformed characteristics instead of X_t:

$$X_t^\star = X_t + \mathbf{1}_N\psi^\top,$$

where ψ is some $L\times 1$ vector and $\mathbf{1}_N$ is the $N\times 1$ vector of ones. In addition, the systematic

risk has the following form:

$$B_t f_{t+1} = X_t^\star \Gamma f_{t+1} = (X_t + \mathbf{1}_N \psi^\top)\Gamma f_{t+1}$$

and the inside pricing error has the following form:

$$\alpha_{I,t} = B_t^I \delta_I, \qquad \text{where } B_t^I = S^\star (S^{\star\top} S^\star)^{-1/2},$$

$$S^\star = \left[X_t^\star \left(X_t^{\star\top} X_t^\star \right)^{-1} X_t^{\star\top} - X_t^\star \Gamma \left(\Gamma^\top X_t^{\star\top} X_t^\star \Gamma \right)^{-1} \Gamma^\top X_t^{\star\top} \right] \begin{pmatrix} I_{L-K} \\ O_{(N-L+K)\times(L-K)} \end{pmatrix}.$$

Let W_d be the $L \times L$ diagonal matrix representing the change in units of characteristics. Consider the case where we use characteristics with different units $\dot{X}_t = X_t W_d$ instead of X_t, and put 'dot' ($\dot{}$) on the parameters of this case. Then, to preserve the systematic risk and the inside pricing error, we need to have

$$\dot{\Gamma} = W_d^{-1}\Gamma, \quad \dot{\psi} = W_d \psi \quad \text{and} \quad \dot{\theta} = \theta, \tag{G.1}$$

where $\theta = \Gamma^\top \psi$. On the other hand, Zhang (2024) derives ψ such that

$$\psi = \Gamma(\Gamma^\top \Gamma)^{-1}\theta.$$

So, if we use $\dot{X}_t$, then we would have $\dot{\psi} = \dot{\Gamma}(\dot{\Gamma}^\top \dot{\Gamma})^{-1}\dot{\theta}$. However, it doesn't satisfy (G.1) because

$$\dot{\psi} = \dot{\Gamma}(\dot{\Gamma}^\top \dot{\Gamma})^{-1}\dot{\theta} = W_d^{-1}\Gamma \left(\Gamma^\top W_d^{-2} \Gamma \right)^{-1} \theta \quad \neq \quad W_d \Gamma (\Gamma^\top \Gamma)^{-1}\theta = W_d \psi,$$

unless $W_d = a I_L$ for some $a \neq 0$. Hence, the sizes of the systematic risk and the inside pricing error are changed depending on the units we use for the characteristics, which is undesirable.

On the other hand, in the case of our paper, because

$$P_{X,t} = P_{\dot{X},t} \quad \text{and} \quad B_t = X_t \Gamma = \dot{X}_t \dot{\Gamma} = \dot{B}_t$$

where $\dot{X}_t = X_t W_d$ and $\dot{\Gamma} = W_d^{-1}\Gamma$, we have

$$(I_N - P_{\dot{B},t}) X_t \eta = (I_N - P_{\dot{B},t}) \dot{X}_t \dot{\eta}$$

where $\dot{\eta} = W_d^{-1}\eta$. Hence, our model is robust to the change of units of the characteristics.

H Ω matrix

This section presents the Ω matrix we used in our empirical analysis. Here, we consider a different Ω matrix for $\alpha_{O,t}$ from that in Zhang (2024) because Ω in Zhang (2024) with our data makes the sizes of some $B^o_{t,jq}$ quite large. To remedy this issue, we introduce a new Ω such that:

$$B^o_t = X^o_t(X^{o\top}_t X^o_t/N)^{-1/2}, \quad X^o_t = [I_N - P_{X,t}]\,\Omega, \quad \Omega_{N\times(N-L)} = \begin{pmatrix} \Psi_{(N-L)\times(N-L)} \\ \Theta_{L\times(N-L)} \end{pmatrix}$$

where

$$\Psi = \begin{bmatrix} 1 & 1.01 & 1 & 1.01 & 1 & 1.01 & 0 & 0 & 0 \\ 0 & 1 & 1.01 & 1 & 1.01 & 1 & 1.01 & 0 & 0 \\ 0 & 0 & 1 & 1.01 & 1 & 1.01 & 1 & 1.01 & 0 \\ 0 & 0 & 0 & 1 & 1.01 & 1 & 1.01 & 1 & 1.01 \\ 1 & 0 & 0 & 0 & 1 & 1.01 & 1 & 1.01 & 1 \\ 1.01 & 1 & 0 & 0 & 0 & 1 & 1.01 & 1 & 1.01 \\ 1 & 1.01 & 1 & 0 & 0 & 0 & 1 & 1.01 & 1 \\ 1.01 & 1 & 1.01 & 1 & 0 & 0 & 0 & 1 & 1.01 \\ 1 & 1.01 & 1 & 1.01 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \otimes I_{(N-L)/9},$$

$$\Theta = \begin{bmatrix} \mathbf{1}^\top_{\lfloor (N-L)/L \rfloor} \otimes I_L & \mathbf{0}_{L\times(N-L-\lfloor (N-L)/L \rfloor \cdot L)} \end{bmatrix},$$

which is designed to make $\Omega = [\Psi^\top \quad \Theta^\top]^\top$ to have the same number of nonzero elements in each column and it makes $B^o_{t,jq}$s not too large in our data. Here, we put 1.01 in some elements instead of 1 just to make Ω have a full column rank. Importantly, the representable set of $\alpha_{O,t}$ is not changed even if we use this basis, because $\alpha_{O,t} = B^o_t \delta_{o,t}$ can be represented with any other basis $\dot{B}^o_t$ such that $\alpha_{O,t} = \dot{B}^o_t \dot{\delta}_{o,t}$ with $\dot{\delta}_{o,t} = (\dot{B}^{o\top}_t \dot{B}^o_t)^{-1} \dot{B}^{o\top}_t B^o_t \delta_{o,t}$.

I Proofs

I.1 Proof of main results

I.1.1 Proof of Theorem 1 (a)

First of all, by Lemma I.4, we have

$$\left\| \tilde{f}^d_{t+1} - H_F f^d_{t+1} \right\|, \quad \frac{1}{T}\sum_{t=1}^{T} \left\| \tilde{f}^d_{t+1} - H_F f^d_{t+1} \right\| = O_p\left(\frac{L^{10}}{N^{10}} + \frac{L}{N\sqrt{T}} + \frac{1}{\sqrt{N}} \right),$$

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right\|^2=O_p\left(\left(\frac{L^{10}}{N^{10}}+\frac{L}{N\sqrt{T}}+\frac{1}{\sqrt{N}}\right)^2\right).$$

In addition, by Lemma I.10, $\|H_F\|, \left\|H_F^{-1}\right\| = O_p(1)$. Note that

$$\begin{aligned}\tilde{\gamma}_l - H_\Gamma^\top \gamma_l &= \left(e_l^\top \otimes \left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1}\right)\frac{1}{T}\sum_{t=1}^{T}\ddot{E}_{t+1}^d \otimes \left(\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right) \\ &+ \left(e_l \otimes \left[\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1} - \left(\frac{1}{T}\sum_t H_F f_{t+1}^d f_{t+1}^{d\top} H_F^\top\right)^{-1}\right]\right)\frac{1}{T}\sum_{t=1}^{T}\ddot{E}_{t+1}^d \otimes \left(H_F f_{t+1}^d\right) \\ &+ H_F^{-\top}\left(e_l \otimes \left(\frac{1}{T}\sum_t f_{t+1}^d f_{t+1}^{d\top}\right)^{-1}\right)\frac{1}{T}\sum_{t=1}^{T}\ddot{E}_{t+1}^d \otimes \ f_{t+1}^d,\end{aligned} \tag{I.1}$$

where $H_\Gamma^\top = \left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1}\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d f_{t+1}^{d\top}\right)$. For the first term, because

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T} e_l^\top \ddot{E}_{t+1} \otimes \left(\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right)\right\| &\le \left(\frac{1}{T}\sum_{t=1}^{T}(e_l^\top \ddot{E}_{t+1})^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right\|^2\right)^{1/2} \\ &= O_p\left(\frac{1}{\sqrt{N}}\left(\frac{L^{10}}{N^{10}}+\frac{L}{N\sqrt{T}}+\frac{1}{\sqrt{N}}\right)\right),\end{aligned}$$

by Lemma I.9, we know the order of the first term is $O_p\left(\frac{1}{\sqrt{N}}\left(\frac{L^{10}}{N^{10}}+\frac{L}{N\sqrt{T}}+\frac{1}{\sqrt{N}}\right)\right)$. For the second term, note that

$$\begin{aligned}&\left\|\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1} - \left(\frac{1}{T}\sum_t H_F f_{t+1}^d f_{t+1}^{d\top} H_F^\top\right)^{-1}\right\| \\ &\lesssim \left\|\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1}\right\|\left\|\left(\frac{1}{T}\sum_t H_F f_{t+1}^d f_{t+1}^{d\top} H_F^\top\right)^{-1}\right\| \\ &\quad\times\left\|\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top} - \frac{1}{T}\sum_t H_F f_{t+1}^d f_{t+1}^{d\top} H_F^\top\right\| \\ &= O_p\left(\frac{L^{10}}{N^{10}}+\frac{L}{N\sqrt{T}}+\frac{1}{\sqrt{N}}\right)\end{aligned}$$

as mentioned in the proof of Lemma I.3. Moreover,

$$\frac{1}{T}\sum_{t=1}^{T} e_l^\top \ddot{E}_{t+1}^d \otimes \left(H_F f_{t+1}^d\right) = H_F \frac{1}{T}\sum_{t=1}^{T} e_l^\top \ddot{E}_{t+1} \otimes f_{t+1}^d$$

and $\left\| \frac{1}{T}\sum_{t=1}^{T} e_l^\top \ddot{E}_{t+1} \otimes f_{t+1}^d \right\| = O_p\left(\frac{1}{\sqrt{NT}}\right)$ because

$$\frac{1}{T}\sum_{t=1}^{T} e_l^\top \ddot{E}_{t+1} \otimes f_{t+1}^d = \frac{1}{NT}\sum_{it}(e_l^\top Q_t^{-1} x_{it}) f_{t+1}^d \epsilon_{i,t+1} = \frac{1}{NT} A^\top \mathrm{vec}(E)$$

where A is the $NT \times K$ matrix of $(e_l^\top Q_t^{-1} x_{it}) f_{t+1}^d$ and

$$\mathbb{E}\left\| \frac{1}{NT} A^\top \mathrm{vec}(E) \right\|_F^2 \lesssim \frac{1}{N^2T^2} \left\| \mathbb{E}[\mathrm{vec}(E)\mathrm{vec}(E)^\top] \right\| \|A\|_F^2 = O_p\left(\frac{1}{NT}\right)$$

since $\|A\|_F^2 \leq \max_t |e_l^\top Q_t^{-1} e_l| \sum_t ||f_{t+1}^d||^2 = O_p(NT)$. Hence, the order of the second term is $O_p\left(\frac{1}{\sqrt{NT}}\left(\frac{L^{10}}{N^{10}} + \frac{L}{N\sqrt{T}} + \frac{1}{\sqrt{N}}\right)\right)$. Lastly, we show that the third term converges to a normal distribution. Note that

$$\sqrt{NT}\frac{1}{T}\sum_{t=1}^{T} e_l^\top \ddot{E}_{t+1}^d \otimes f_{t+1}^d = \frac{1}{\sqrt{NT}}\sum_{it}(e_l^\top Q_t^{-1} x_{it}) f_{t+1}^d \epsilon_{i,t+1} \to_d \mathcal{N}(0, \Sigma_{xf,l}).$$

In addition, because $H_F^{-\top} \to_p \boldsymbol{H}^\top$ by Lemma I.10 and $\frac{1}{T}\sum_{t=1}^{T} f_t^d f_t^{d\top} \to_p \Sigma_f$ by Assumption 2, we have

$$\sqrt{NT} H_F^{-\top} \left(e_l \otimes \left(\frac{1}{T}\sum_t f_{t+1}^d f_{t+1}^{d\top} \right)^{-1} \right) \frac{1}{T}\sum_{t=1}^{T} \ddot{E}_{t+1}^d \otimes \ f_{t+1}^d \to_d \mathcal{N}\left(0, \boldsymbol{H}^\top \Sigma_f^{-1} \Sigma_{xf,l} \Sigma_f^{-1} \boldsymbol{H}\right).$$

Since the first and second terms are $o_p(1/\sqrt{NT})$ under our assumption, we have

$$\sqrt{NT}\left(\hat{\gamma}_l - H_\Gamma^\top \gamma_l\right) \to_d \mathcal{N}\left(0, \boldsymbol{H}^\top \Sigma_f^{-1} \Sigma_{xf,l} \Sigma_f^{-1} \boldsymbol{H}\right). \ \square$$

I.1.2 Proof of Theorem 1 (b)

The first term of (I.1) can be represented as:

$$\frac{1}{T}\sum_{t=1}^{T} e_l^\top \ddot{E}_{t+1}^d \otimes \left(\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right) = \tilde{\Gamma}^\top \frac{1}{T}\sum_{t=1}^{T}(e_l^\top \ddot{E}_{t+1}) \times \ddot{E}_{t+1} + \tilde{\Gamma}^\top \bar{\ddot{E}} \times \frac{1}{T}\sum_{t=1}^{T}(e_l^\top \ddot{E}_{t+1}).$$

By Lemma I.9, we have $||\bar{\ddot{E}}|| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$, $||\frac{1}{T}\sum_{t=1}^{T}(e_l^\top \ddot{E}_{t+1})|| = O_p\left(\frac{1}{\sqrt{NT}}\right)$. Hence, the second term of the above equation is $O_p\left(\frac{\sqrt{L}}{NT}\right) = o_p\left(\frac{1}{\sqrt{NT}}\right)$. For the first term, we use the debasing method. Note that

$$\tilde{\Gamma}^\top \frac{1}{T}\sum_{t=1}^{T}(e_l^\top \ddot{E}_{t+1}) \times \ddot{E}_{t+1} - \tilde{\Gamma}^\top \frac{1}{T}\sum_{t=1}^{T} \hat{\sigma}_{t+1}^2 (X_t^\top X_t)^{-1} e_l$$

$$= \tilde{\Gamma}^\top \frac{1}{N^2T} \sum_{i,j,t} (e_l^\top Q_t^{-1} x_{it}) Q_t^{-1} x_{jt} \epsilon_{i,t+1} \epsilon_{j,t+1} - \tilde{\Gamma}^\top \frac{1}{N^2T} \sum_{i,t} (e_l^\top Q_t^{-1} x_{it}) Q_t^{-1} x_{it} \hat{\sigma}_{t+1}^2$$
$$= \tilde{\Gamma}^\top \frac{1}{N^2T} \sum_{i,j,t} a_{ijt} (u_{ij,t+1} - \mathbb{E}[u_{ij,t+1}]) + \tilde{\Gamma}^\top \frac{1}{N^2T} \sum_{i,t} a_{iit} (\mathbb{E}[\epsilon_{i,t+1}^2] - \hat{\sigma}_{t+1}^2),$$

where $u_{ij,t+1} = \epsilon_{i,t+1}\epsilon_{j,t+1}$ and $a_{ijt} = (e_l^\top Q_t^{-1} x_{it}) Q_t^{-1} x_{jt}$. Then, by Lemma I.11, the first term of the last equation is $O_p\left(\frac{\sqrt{L}}{N\sqrt{T}}\right) = o_p\left(\frac{1}{\sqrt{NT}}\right)$. In addition, the second term of the last equation is also $o_p\left(\frac{1}{\sqrt{NT}}\right)$ by Lemma I.11. Hence, we have

$$\sqrt{NT}\left(\hat{\gamma}_l - H_\Gamma^\top \gamma_l\right) = \sqrt{NT} H_F^{-\top} \left(e_l \otimes \left(\frac{1}{T} \sum_t f_{t+1}^d f_{t+1}^{d\top} \right)^{-1} \right) \frac{1}{T} \sum_{t=1}^{T} \ddot{E}_{t+1}^d \otimes \ f_{t+1}^d + o_p(1)$$
$$\to_d \mathcal{N}\left(0, \boldsymbol{H}^\top \Sigma_f^{-1} \Sigma_{xf,l} \Sigma_f^{-1} \boldsymbol{H}\right). \ \square$$

I.1.3 Proof of Theorem 2

(a) CLT for $\tilde{\alpha}_{I,it}$: Note that

$$\begin{aligned}\tilde{\alpha}_{I,it} - \alpha_{I,it} &= e_i^\top \left(P_{B,t} - P_{\tilde{B},t}\right) X_t \bar{\bar{R}} + \left(X_{it}^\top - e_i^\top P_{B,t} X_t\right) \bar{\bar{E}} \\ &= e_i^\top \left(P_{B,t} - P_{\tilde{B},t}\right) X_t (\eta + \Gamma \bar{\bar{f}}) + \left(x_{it}^\top - e_i^\top P_{B,t} X_t\right) \bar{\bar{E}} + e_i^\top \left(P_{B,t} - P_{\tilde{B},t}\right) X_t \bar{\bar{E}}. \end{aligned} \tag{I.2}$$

By Lemma I.13, the first term can be represented like

$$\begin{aligned} e_i^\top \left(P_{B,t} - P_{\tilde{B},t}\right) X_t (\eta + \Gamma \bar{\bar{f}}) = &-B_{it}^\top \left(B_t^\top B_t\right)^{-1} H_\Gamma^{-\top} \left(\tilde{B}_t - B_t H_\Gamma\right)^\top (I_N - P_{B,t}) X_t (\eta + \Gamma \bar{\bar{f}}) \\ &- e_i^\top (I_N - P_{B,t}) \left(\tilde{B}_t - B_t H_\Gamma\right) H_\Gamma^{-1} \left(B_t^\top B_t\right)^{-1} B_t^\top X_t (\eta + \Gamma \bar{\bar{f}}) \\ &+ O_p\left(\frac{L}{NT} + \frac{L}{N^2} + \left(\frac{L}{N}\right)^{20+1}\right). \end{aligned}$$

Here, we use Lemma I.12 to show the higher order terms are $O_p\left(\frac{L}{NT} + \frac{L}{N^2} + \left(\frac{L}{N}\right)^{20+1}\right)$. Then, the first part of the dominating term can be represented as

$$\begin{aligned} &B_{it}^\top \left(B_t^\top B_t\right)^{-1} H_\Gamma^{-\top} \left(\tilde{B}_t - B_t H_\Gamma\right)^\top (I_N - P_{B,t}) X_t (\eta + \Gamma \bar{\bar{f}}) \\ &= B_{it}^\top \left(B_t^\top B_t\right)^{-1} H_\Gamma^{-\top} \left(\tilde{\Gamma}_t - \Gamma_t H_\Gamma\right)^\top X_t^\top (I_N - P_{B,t}) X_t (\eta + \Gamma \bar{\bar{f}})) \\ &= \left((\eta + \Gamma \bar{\bar{f}}))^\top X_t^\top (I_N - P_{B,t}) X_t \otimes B_{it}^\top \left(B_t^\top B_t\right)^{-1} H_\Gamma^{-\top}\right) \text{vec}\left[\left(\tilde{\Gamma} - \Gamma H_\Gamma\right)^\top\right] \\ &= \left(\eta^\top X_t^\top (I_N - P_{B,t}) X_t \otimes B_{it}^\top \left(B_t^\top B_t\right)^{-1} H_\Gamma^{-\top}\right) \text{vec}\left[\left(\tilde{\Gamma} - \Gamma H_\Gamma\right)^\top\right]. \end{aligned}$$

Here, we use the relation that $\Gamma^\top X_t^\top M_{B,t} = 0$ where $M_{B,t} = I_N - P_{B,t}$. In addition, we have

$$\begin{aligned}
\text{vec}\left[\left(\tilde{\Gamma} - \Gamma H_\Gamma\right)^\top\right] &= \left(I_L \otimes \left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1}\right) \frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1}^d \otimes \left(\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right) \\
&\quad + \left(I_L \otimes \left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1}\right) \frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1}^d \otimes \left(H_F f_{t+1}^d\right) \\
&= \left(I_L \otimes H_F^{-\top}\left(\frac{1}{T}\sum_t f_{t+1}^d f_{t+1}^{d\top}\right)^{-1}\right) \frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1}^d \otimes f_{t+1}^d \\
&\quad + O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\left(\frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10} + \frac{1}{\sqrt{N}}\right)\right).
\end{aligned}$$

Here, we derive the order of the residual terms using the proof of Lemma I.3 with $a_{NT} = \frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10} + \frac{1}{\sqrt{N}}$. Then, since $\left\|\left(\eta^\top X_t^\top M_{B,t} X_t \otimes B_{it}^\top \left(B_t^\top B_t\right)^{-1} H_\Gamma^{\top -1}\right)\right\| = O_p(1)$, we have

$$\begin{aligned}
&B_{it}^\top \left(B_t^\top B_t\right)^{-1} H_\Gamma^{-\top} \left(\tilde{B}_t - B_t H_\Gamma\right)^\top (I_N - P_{B,t}) X_t(\eta + \Gamma \bar{f}) \\
&= \left(\eta^\top (X_t^\top M_{B,t} X_t/N) \otimes B_{it}^\top \left(B_t^\top B_t/N\right)^{-1} \left(\frac{1}{T}\sum_{s=1}^T f_{s+1}^d f_{s+1}^{d\top}\right)^{-1}\right) \frac{1}{NT}\sum_{j=1}^N \sum_{s=1}^T \left(Q_t^{-1} x_{js} \otimes f_{s+1}^d\right) \epsilon_{j,s+1} \\
&\quad + O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\left(\frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10} + \frac{1}{\sqrt{N}}\right)\right) \\
&= \frac{1}{NT}\sum_{j=1}^N \sum_{s=1}^T \left(\eta^\top - \eta^\top Q_t \Gamma (Q_t^B)^{-1} \Gamma^\top\right) x_{js} \left(B_{it}^\top (Q_t^B)^{-1} (Q^f)^{-1} f_{s+1}^d\right) \epsilon_{j,s+1} \\
&\quad + O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\left(\frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10} + \frac{1}{\sqrt{N}}\right)\right),
\end{aligned}$$

where $Q_t^B = B_t^\top B_t/N$ and $Q^f = F^{d\top} F^d/T$ since $\eta^\top (X_t^\top M_{B,t} X_t/N) = \eta^\top Q_t - \eta^\top Q_t \Gamma (Q_t^B)^{-1} \Gamma^\top Q_t$. Similarly, the second part of the dominating term can be represented as

$$\begin{aligned}
&e_i^\top (I_N - P_{B,t}) \left(\tilde{B}_t - B_t H_\Gamma\right) H_\Gamma^{-1} \left(B_t^\top B_t\right)^{-1} B_t^\top X_t(\eta + \Gamma \bar{\bar{f}}) \\
&= \left(e_i^\top M_{B,t} X_t \otimes \left(\eta^\top Q_t \Gamma (Q_t^B)^{-1} + \bar{\bar{f}}^\top\right) H_\Gamma^{-\top}\right) \text{vec}\left[\left(\hat{\Gamma} - \Gamma H_\Gamma\right)^\top\right] \\
&= \left(e_i^\top M_{B,t} X_t \otimes \left(\eta^\top Q_t \Gamma (Q_t^B)^{-1} + \bar{\bar{f}}^\top\right) (Q^f)^{-1}\right) \frac{1}{NT}\sum_{j=1}^N \sum_{s=1}^T \left(Q_t^{-1} x_{js} \otimes f_{s+1}^d\right) \epsilon_{j,s+1}
\end{aligned}$$

$$
\begin{aligned}
&+ O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\left(\frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10} + \frac{1}{\sqrt{N}}\right)\right) \\
&= \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\left(x_{it}^{\top}Q_t^{-1} - B_{it}^{\top}(Q_t^B)^{-1}\Gamma^{\top}\right)x_{js}\left(\eta^{\top}Q_t\Gamma(Q_t^B)^{-1} + \bar{\bar{f}}^{\top}\right)\left(Q^f\right)^{-1}f_{s+1}^d\epsilon_{j,s+1} \\
&+ O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\left(\frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10} + \frac{1}{\sqrt{N}}\right)\right),
\end{aligned}
$$

because $e_i^{\top}M_{B,t}X_t = x_{it}^{\top} - B_{it}^{\top}(Q_t^B)^{-1}\Gamma^{\top}Q_t$. In addition, the second term in (I.2) can be represented as

$$
\left(x_{it}^{\top} - e_i^{\top}P_{B,t}X_t\right)\bar{\bar{E}} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\left(x_{it}^{\top}Q_t^{-1} - B_{it}^{\top}(Q_t^B)^{-1}\Gamma^{\top}\right)x_{js}\epsilon_{j,s+1}.
$$

Moreover, the third term in (I.2) can be bounded like

$$
\left\|e_i^{\top}\left(P_{B,t} - P_{\tilde{B},t}\right)X_t\bar{\bar{E}}\right\| \leq \left\|e_i^{\top}\left(P_{B,t} - P_{\tilde{B},t}\right)X_t\right\|\left\|\bar{\bar{E}}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\left(\frac{\sqrt{L}}{\sqrt{NT}} + \frac{\sqrt{L}}{N} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}}\right)\right),
$$

by Lemma I.9 and the bound that $\left\|e_i^{\top}\left(P_{B,t} - P_{\tilde{B},t}\right)X_t\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \frac{\sqrt{L}}{N} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}}\right)$ since

$$
\begin{aligned}
e_i^{\top}(P_{B,t} - P_{\tilde{B},t})X_t &= x_{it}^{\top}\Gamma\left(\Gamma^{\top}Q_t\Gamma\right)^{-1}\Gamma^{\top}Q_t - x_{it}^{\top}\tilde{\Gamma}\left(\tilde{\Gamma}^{\top}Q_t\tilde{\Gamma}\right)^{-1}\tilde{\Gamma}^{\top}Q_t \\
&\lesssim \left\|x_{it}^{\top}\Gamma H_\Gamma - x_{it}^{\top}\tilde{\Gamma}\right\|\left\|\left(H_\Gamma^{\top}\Gamma^{\top}Q_t\Gamma H_\Gamma\right)^{-1}\right\|\left\|H_\Gamma^{\top}\Gamma^{\top}Q_t\right\| \\
&+ \left\|B_{it}^{\top}H_\Gamma\right\|\left\|\left(H_\Gamma^{\top}\Gamma^{\top}Q_t\Gamma H_\Gamma\right)^{-1} - \left(\tilde{\Gamma}^{\top}Q_t\tilde{\Gamma}\right)^{-1}\right\|\left\|H_\Gamma^{\top}\Gamma^{\top}Q_t\right\| \\
&+ \left\|B_{it}^{\top}H_\Gamma\right\|\left\|\left(\tilde{\Gamma}^{\top}Q_t\tilde{\Gamma}\right)^{-1}\right\|\left\|\tilde{\Gamma}^{\top} - H_\Gamma^{\top}\Gamma^{\top}\right\|\|Q_t\| \\
&= O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \frac{\sqrt{L}}{N} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}}\right)
\end{aligned}
$$

by Lemma I.12. In summary, we have

$$
\frac{\sqrt{NT}}{\sqrt{L}}\sigma_{I,it}^{-1}\left(\tilde{\alpha}_{I,it} - \alpha_{I,it}\right) = \sigma_{I,it}^{-1}\frac{1}{\sqrt{NTL}}\sum_{j=1}^{N}\sum_{s=1}^{T}g_{it,js}\epsilon_{j,s+1} + o_p(1).
$$

where

$$
\begin{aligned}
g_{it,js} = &\left[1 - \left(\eta^{\top}Q_t\Gamma(Q_t^B)^{-1} + \bar{\bar{f}}^{\top}\right)\left(Q^f\right)^{-1}f_{s+1}^d\right]\left(x_{it}^{\top}Q_t^{-1} - B_{it}^{\top}(Q_t^B)^{-1}\Gamma^{\top}\right)x_{js} \\
&- \left(B_{it}^{\top}(Q_t^B)^{-1}(Q^f)^{-1}f_{s+1}^d\right)\left(\eta^{\top} - \eta^{\top}Q_t\Gamma(Q_t^B)^{-1}\Gamma^{\top}\right)x_{js}.
\end{aligned}
$$

By Assumption 5, the first term converges to a standard normal distribution. It completes the proof.

(b) CLT for $\hat{\alpha}_{I,it}$: The proof is basically the same as that of (a). The only difference is that we use the bounds from Lemma I.12 like $\left\|\hat{B}_t - B_t H_\Gamma\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{T}}\right)$, $\left\|\hat{B}_{it} - H_\Gamma^\top B_{it}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$, $\left\|\left(\hat{B}_t^\top \hat{B}_t/N\right)^{-1} - \left(H_\Gamma^\top B_t^\top B_t H_\Gamma/N\right)^{-1}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$ to show the higher order terms are $O_p\left(\frac{L}{N^{3/2}\sqrt{T}}\right)$. In addition, the order of the residuals in the dominating parts becomes $o_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$. Hence, we have

$$\frac{\sqrt{NT}}{\sqrt{L}}\sigma_{I,it}^{-1}\left(\hat{\alpha}_{I,it} - \alpha_{I,it}\right) = \sigma_{I,it}^{-1}\frac{1}{\sqrt{NTL}}\sum_{j=1}^{N}\sum_{s=1}^{T} g_{it,js}\epsilon_{j,s+1} + o_p(1).$$

By Assumption 5, the first term converges to a standard normal distribution, and the second term converges to 0. It completes the proof. □

I.1.4 Proof of Theorem 3

Because

$$\sqrt{N}\left(\tilde{\delta}_{o,tq} - \delta_{o,tq}\right) = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} B_{t,jq}^o \epsilon_{j,s+1}$$

where $B_{t,jq}^o = e_j^\top B_t^o e_q$, we have by Assumption 5 that

$$\sqrt{N}\sigma_{\delta,qt}^{-1}\left(\tilde{\delta}_{o,tq} - \delta_{o,tq}\right) \to_d \mathcal{N}(0,1). \quad \square$$

I.1.5 Proof of Theorem 4

Let $\dot{\xi}_t = \tilde{\delta}_{o,t} - \tilde{\zeta}$. First, we denote by $\mathcal{W}$ the event that for all t,

$$\max_{1\le q\le N-L}\left|\frac{1}{N}\sum_{j=1}^{N} B_{t,jq}^o\epsilon_{j,t+1} - \frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B_{s,jq}^o\epsilon_{j,s+1}\right| \le (C_u + 0.05)\sigma_{t+1}\frac{\sqrt{\log NT}}{\sqrt{N}}$$

where $C_u > 0$ is the universal constant in Lemma I.15. Then, by Lemmas I.15 and I.16, we know $\Pr(\mathcal{W}) \to 1$. Set $\rho_t = C_\rho\sigma_{t+1}\frac{\sqrt{\log NT}}{\sqrt{N}}$ where $C_\rho = C_u + 0.1$. Then, because $\xi_{t,q} = 0$ if $q \notin D_t$, on the event $\mathcal{W}$, we have for all t,

$$\begin{aligned}\max_{q\notin D_t}\left|\dot{\xi}_{t,q}\right| &\le \max_{q\notin D_t}\left|\frac{1}{N}\sum_{j=1}^{N} B_{t,jq}^o\epsilon_{j,t+1} - \frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B_{s,jq}^o\epsilon_{j,s+1}\right| + \max_{q\notin D_t}\left|\frac{1}{T}\sum_{s=1}^{T}\xi_{s,q}\right| \\ &\le C_\rho\sigma_{t+1}\frac{\sqrt{\log NT}}{\sqrt{N}} = \rho_t.\end{aligned}$$

Hence, on the event $\mathcal{W}$, we have $\tilde{\xi}_{t,q} = 0$ for all t and $q \notin D_t$. Here, we use the relation that

$$\dot{\xi}_t = \xi_t + \frac{1}{N} B_t^{o\top} E_{t+1} - \frac{1}{NT} \sum_{s=1}^{T} B_s^{o\top} E_{s+1} - \frac{1}{T} \sum_{s=1}^{T} \xi_s.$$

In addition, on the event $\mathcal{W}$, we have for all t,

$$\begin{aligned}
\min_{q \in D_t} \left| \dot{\xi}_{t,q} \right| &\geq \min_{q \in D_t} |\xi_{t,q}| - \max_{q \in D_t} \left| \frac{1}{N} \sum_{j=1}^{N} B_{t,jq}^o \epsilon_{j,t+1} - \frac{1}{NT} \sum_{s=1}^{T} \sum_{j=1}^{N} B_{s,jq}^o \epsilon_{j,s+1} \right| - \max_{q \in D_t} \left| \frac{1}{T} \sum_{s=1}^{T} \xi_{s,q} \right| \\
&\geq \min_{q \in D_t} |\xi_{t,q}| - C_\rho \sigma_{t+1} \frac{\sqrt{\log NT}}{\sqrt{N}} \\
&\gg C_\rho \sigma_{t+1} \frac{\sqrt{\log NT}}{\sqrt{N}},
\end{aligned}$$

by Assumption 4. So, on the event $\mathcal{W}$, we have $\tilde{\xi}_{t,q} = \dot{\xi}_{t,q}$ for all t and $q \in D_t$. In addition, on the event $\mathcal{W}$, D_t is the same as $\mathcal{D}_t$ for all t, where $\mathcal{D}_t = \{1 \leq q \leq N - L : \tilde{\xi}_{t,q} \neq 0\}$. Note that

$$V_{o,it}^{-1/2} \left(\hat{\alpha}_{O,it} - \alpha_{O,it} \right) = V_{o,it}^{-1/2} \sum_{q=1}^{N-L} B_{t,iq}^o \left(\hat{\zeta}_q - \zeta_q \right) + V_{o,it}^{-1/2} \sum_{q=1}^{N-L} B_{t,iq}^o \left(\tilde{\xi}_{t,q} - \xi_{t,q} \right) + V_{o,it}^{-1/2} \sum_{q \in \mathcal{D}_t} B_{t,iq}^o \bar{\tilde{\xi}}_q,$$

where

$$\hat{\zeta} - \zeta = \frac{1}{NT} B_s^{o\top} E_{s+1} + \frac{1}{T} \sum_{s=1}^{T} (\xi_s - \tilde{\xi}_s).$$

Then, the first part can be decomposed into

$$V_{o,it}^{-1/2} \sum_{q=1}^{N-L} B_{t,iq}^o \left(\hat{\zeta}_q - \zeta_q \right) = V_{o,it}^{-1/2} \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} B_{t,i}^{o\top} B_{s,j}^o \epsilon_{j,s+1} - V_{o,it}^{-1/2} \sum_{q=1}^{N-L} B_{t,iq}^o \left(\bar{\tilde{\xi}}_q - \bar{\xi}_q \right).$$

In addition, the second part can be decomposed into

$$\begin{aligned}
&V_{o,it}^{-1/2} \sum_{q=1}^{N-L} B_{t,iq}^o \left(\tilde{\xi}_{t,q} - \xi_{t,q} \right) \\
&= V_{o,it}^{-1/2} \sum_{q \in D_t} B_{t,iq}^o \left(\tilde{\xi}_{t,q} - \xi_{t,q} \right) + V_{o,it}^{-1/2} \sum_{q \notin D_t} B_{t,iq}^o \left(\tilde{\xi}_{t,q} - \xi_{t,q} \right) \\
&= V_{o,it}^{-1/2} \sum_{q \in D_t} B_{t,iq}^o \left(\dot{\xi}_{t,q} - \xi_{t,q} \right) + V_{o,it}^{-1/2} \sum_{q \in D_t} B_{t,iq}^o \left(\tilde{\xi}_{t,q} - \dot{\xi}_{t,q} \right) + V_{o,it}^{-1/2} \sum_{q \notin D_t} B_{t,iq}^o \left(\tilde{\xi}_{t,q} - \xi_{t,q} \right).
\end{aligned}$$

Note that, on the event $\mathcal{W}$, the second term is zero since $\tilde{\xi}_{t,q} = \dot{\xi}_{t,q}$ for all $q \in D_t$. In addition, $\Pr(\mathcal{W}) \to 1$. Hence, w.p.c. to 1, the second term is zero, and for any $\varepsilon > 0$, we

have

$$\Pr\left(\left|V_{o,it}^{-1/2}\sum_{q\in D_t}B_{t,iq}^{o}\left(\tilde{\xi}_{t,q}-\dot{\xi}_{t,q}\right)\right|\geq\varepsilon\right)\leq\Pr\left(\left|V_{o,it}^{-1/2}\sum_{q\in D_t}B_{t,iq}^{o}\left(\tilde{\xi}_{t,q}-\dot{\xi}_{t,q}\right)\right|\neq 0\right)\to 0.$$

Hence, the second term is $o_p(1)$. Similarly, on the event $\mathcal{W}$, the third term is zero, since $\tilde{\xi}_{t,q}=\xi_{t,q}=0$ for all $q\notin D_t$. Hence, we know that the third term is $o_p(1)$. Hence, we have

$$\begin{aligned}
&V_{o,it}^{-1/2}\sum_{q=1}^{N-L}B_{t,iq}^{o}\left(\tilde{\xi}_{t,q}-\xi_{t,q}\right)=V_{o,it}^{-1/2}\sum_{q\in D_t}B_{t,iq}^{o}\left(\dot{\xi}_{t,q}-\xi_{t,q}\right)+o_p(1)\\
&=V_{o,it}^{-1/2}\frac{1}{N}\sum_{j=1}^{N}\sum_{q\in D_t}B_{t,iq}^{o}B_{t,jq}^{o}\epsilon_{j,t+1}-V_{o,it}^{-1/2}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{q\in\mathcal{D}_t}B_{t,iq}^{o}B_{s,jq}^{o}\epsilon_{j,s+1}-V_{o,it}^{-1/2}\sum_{q\in D_t}B_{t,iq}^{o}\bar{\xi}_q+o_p(1)\\
&=V_{o,it}^{-1/2}\frac{1}{N}\sum_{j=1}^{N}\sum_{q\in D_t}B_{t,iq}^{o}B_{t,jq}^{o}\epsilon_{j,t+1}-V_{o,it}^{-1/2}\sum_{q\in D_t}B_{t,iq}^{o}\bar{\xi}_q+o_p(1).
\end{aligned}$$

Here, the last equation comes from the fact that $\mathbb{E}\left[\left(\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{q\in D_t}B_{t,iq}^{o}B_{s,jq}^{o}\epsilon_{j,s+1}\right)^2\right]\lesssim\frac{|D_t|}{NT}\left(\frac{1}{|D_t|}\sum_{q\in D_t}B_{t,iq}^{o2}\right)$. Therefore, we have

$$\begin{aligned}
V_{o,it}^{-1/2}\left(\hat{\alpha}_{O,it}-\alpha_{O,it}\right)=V_{o,it}^{-1/2}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}B_{t,i}^{o\top}B_{s,j}^{o}\epsilon_{j,s+1}+V_{o,it}^{-1/2}\frac{1}{N}\sum_{j=1}^{N}\sum_{q\in D_t}B_{t,iq}^{o}B_{t,jq}^{o}\epsilon_{j,t+1}\\
+V_{o,it}^{-1/2}\sum_{q\notin D_t}B_{t,iq}^{o}\left(\tilde{\bar{\xi}}_q-\bar{\xi}_q\right)+o_p(1).
\end{aligned}$$

Here, we use the fact that $V_{o,it}^{-1/2}\sum_{q\in D_t}B_{t,iq}^{o}\tilde{\bar{\xi}}_q-V_{o,it}^{-1/2}\sum_{q\in\mathcal{D}_t}B_{t,iq}^{o}\tilde{\bar{\xi}}_q=o_p(1)$ because, on the event $\mathcal{W}$, $D_t=\mathcal{D}_t$ and $V_{o,it}^{-1/2}\sum_{q\in D_t}B_{t,iq}^{o}\tilde{\bar{\xi}}_q=V_{o,it}^{-1/2}\sum_{q\in\mathcal{D}_t}B_{t,iq}^{o}\tilde{\bar{\xi}}_q$. Lastly, we show that $V_{o,it}^{-1/2}\sum_{q\notin D_t}B_{t,iq}^{o}\left(\tilde{\bar{\xi}}_q-\bar{\xi}_q\right)\to_p 0$. By using the same argument as above, we have

$$\begin{aligned}
&V_{o,it}^{-1/2}\sum_{q\notin D_t}B_{t,iq}^{o}\left(\tilde{\bar{\xi}}_q-\bar{\xi}_q\right)\\
&=V_{o,it}^{-1/2}\frac{1}{T}\sum_{s=1}^{T}\sum_{q\notin D_t}B_{t,iq}^{o}\left(\tilde{\xi}_{s,q}-\xi_{s,q}\right)\\
&=V_{o,it}^{-1/2}\frac{1}{T}\sum_{s=1}^{T}\sum_{q\in D_s/D_t}B_{t,iq}^{o}\left(\dot{\xi}_{s,q}-\xi_{s,q}\right)+o_p(1)\\
&=V_{o,it}^{-1/2}\frac{1}{T}\sum_{s=1}^{T}\sum_{q\in D_s/D_t}B_{t,iq}^{o}\left(\frac{1}{N}\sum_{j=1}^{N}B_{s,jq}^{o}\epsilon_{j,s+1}-\frac{1}{NT}\sum_{s'=1}^{T}\sum_{j=1}^{N}B_{s',jq}^{o}\epsilon_{j,s'+1}-\frac{1}{T}\sum_{s'=1}^{T}\xi_{s',q}\right)+o_p(1).
\end{aligned}$$

Some calculation shows that

$$\mathbb{E}\left[\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{q\in D_s/D_t} B^o_{t,iq}\frac{1}{N}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right)^2\right] \lesssim \frac{1}{NT}\frac{1}{T}\sum_{s=1}^{T}\sum_{q\in D_s/D_t} B^{o2}_{t,iq}$$

Hence, the first term is $o_p(1)$. In addition, we can also show that

$$\mathbb{E}\left[\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{q\in D_s/D_t} B^o_{t,iq}\frac{1}{NT}\sum_{s'=1}^{T}\sum_{j=1}^{N} B^o_{s',jq}\epsilon_{j,s'+1}\right)^2\right] \lesssim \frac{1}{NT}\frac{1}{T}\sum_{s=1}^{T}\sum_{q\in D_s/D_t} B^{o2}_{t,iq}.$$

So, the second term is $o_p(1)$. Moreover, the third term also converges to 0 under our assumption. Hence, we have

$$V^{-1/2}_{o,it}\left(\hat{\alpha}_{O,it}-\alpha_{O,it}\right) = V^{-1/2}_{o,it}\left(\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} B^{o\top}_{t,i}B^o_{s,j}\epsilon_{j,s+1} + \frac{1}{N}\sum_{j=1}^{N}\sum_{q\in D_t} B^o_{t,iq}B^o_{t,jq}\epsilon_{j,t+1}\right) + o_p(1)$$
$$\to_d \mathcal{N}(0,1) \quad \square$$

I.1.6 Proof of Theorem B.1

Because

$$\hat{\alpha}_{o,it}-\alpha_{o,it} = B^{o\top}_{t,i}(\hat{\delta}_o-\delta_o) = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} B^{o\top}_{t,i}B^o_{s,j}\epsilon_{j,s+1},$$

we have

$$\sqrt{NT}\left\|B^{o\top}_{t,i}\right\|^{-1}\left(\hat{\alpha}_{o,it}-\alpha_{o,it}\right) = \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T} h_{js}\epsilon_{j,s+1},$$

where $h_{js} = \left\|B^o_{t,i}\right\|^{-1} B^{o\top}_{t,i}B^o_{s,j}$. Hence, by Assumption 5,

$$\sqrt{NT}\sigma^{-1}_{o,it}\left\|B^{o\top}_{t,i}\right\|^{-1}\left(\hat{\alpha}_{o,it}-\alpha_{o,it}\right) \to_d \mathcal{N}(0,1). \quad \square$$

I.1.7 Proof of Theorem F.1

(i) Let $\varpi_t = [Q_t^{-1}]_{ll}$. Then, we know $\max_t|\varpi_t|$ is bounded. Note that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}^2_{t+1}\varpi_t\hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1} - \frac{1}{T}\sum_{t=1}^{T}\sigma^2_{t+1}\varpi_t H_F f^d_{t+1}f^{d\top}_{t+1}H_F^\top\right\|$$
$$\lesssim \left\|\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}^2_{t+1}-\sigma^2_{t+1})\varpi_t f^d_{t+1}f^{d\top}_{t+1}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T}\sigma^2_{t+1}\varpi_t\left(H_F f^d_{t+1}f^{d\top}_{t+1}H_F^\top - \hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\right)\right\|$$
$$+ \left\|\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}^2_{t+1}-\sigma^2_{t+1})\varpi_t\left(H_F f^d_{t+1}f^{d\top}_{t+1}H_F^\top - \hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\right)\right\|.$$

By Lemma I.8, the first term can be bounded like

$$\left\|\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_{t+1}^2-\sigma_{t+1}^2)\varpi_t f_{t+1}^d f_{t+1}^{d\top}\right\| \le \max_t|\varpi_t|\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_{t+1}^2-\sigma_{t+1}^2)^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|f_{t+1}^d f_{t+1}^{d\top}\right\|^2\right)^{1/2}$$
$$= o_p(1).$$

In addition, the second term can be bound like

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\sigma_{t+1}^2\varpi_t\left(H_F f_{t+1}^d f_{t+1}^{d\top}H_F^\top-\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}\right)\right\|$$
$$\lesssim \max_t\left|\sigma_{t+1}^2\varpi_t\right|\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_{t+1}^d-H_F f_{t+1}^d\right\|\left\|f_{t+1}^d\right\| = o_p(1).$$

Moreover, the third term can be bounded like

$$\left\|\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_{t+1}^2-\sigma_{t+1}^2)\varpi_t\left(H_F f_{t+1}^d f_{t+1}^{d\top}H_F^\top-\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}\right)\right\|$$
$$\le \max_t|\varpi_t|\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_{t+1}^2-\sigma_{t+1}^2)^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|H_F f_{t+1}^d f_{t+1}^{d\top}H_F^\top-\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}\right\|^2\right)^{1/2} = o_p(1)$$

by Lemma I.8. Here, $\frac{1}{T}\sum_{t=1}^{T}\left\|H_F f_{t+1}^d f_{t+1}^{d\top}H_F^\top-\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}\right\|^2 = o_p(1)$ because

$$\frac{1}{T}\sum_{t=1}^{T}\left\|H_F f_{t+1}^d-\hat{f}_{t+1}^d\right\|^2\left\|f_{t+1}^d\right\|^2 = o_p(1), \quad \frac{1}{T}\sum_{t=1}^{T}\left\|H_F f_{t+1}^d-\hat{f}_{t+1}^d\right\|^4 = o_p(1).$$

Hence, we have

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_{t+1}^2\varpi_t\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}-\frac{1}{T}\sum_{t=1}^{T}\sigma_{t+1}^2\varpi_t H_F f_{t+1}^d f_{t+1}^{d\top}H_F^\top\right\| = o_p(1).$$

In addition, we have

$$\left\|\left(\frac{1}{T}\sum_t\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}\right)^{-1}-\left(H_F\frac{1}{T}\sum_t f_{t+1}^d f_{t+1}^{d\top}H_F^\top\right)^{-1}\right\|$$
$$\le\left\|\left(\frac{1}{T}\sum_t\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}\right)^{-1}\right\|\left\|\left(H_F\frac{1}{T}\sum_t f_{t+1}^d f_{t+1}^{d\top}H_F^\top\right)^{-1}\right\|$$
$$\times\left\|\frac{1}{T}\sum_t\hat{f}_{t+1}^d\hat{f}_{t+1}^{d\top}-H_F\frac{1}{T}\sum_t f_{t+1}^d f_{t+1}^{d\top}H_F^\top\right\|$$
$$= o_p(1)$$

as noted in the proof of Lemma I.3. Hence, we have

$$\left(\frac{1}{T}\sum_t \hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\right)^{-1}\frac{1}{T}\sum_{t=1}^T \hat{\sigma}^2_{t+1}\varpi_t \hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\left(\frac{1}{T}\sum_t \hat{f}^d_{t+1}\hat{f}^{d\top}_{t+1}\right)^{-1}$$
$$\to_p \left(H_F\frac{1}{T}\sum_t f^d_{t+1}f^{d\top}_{t+1}H_F^\top\right)^{-1}\frac{1}{T}\sum_{t=1}^T \sigma^2_{t+1}\varpi_t H_F f^d_{t+1}f^{d\top}_{t+1}H_F^\top\left(H_F\frac{1}{T}\sum_t f^d_{t+1}f^{d\top}_{t+1}H_F^\top\right)^{-1}$$
$$= H_F^{-\top}\left(\frac{1}{T}\sum_t f^d_{t+1}f^{d\top}_{t+1}\right)^{-1}\frac{1}{T}\sum_{t=1}^T \sigma^2_{t+1}\varpi_t f^d_{t+1}f^{d\top}_{t+1}\left(\frac{1}{T}\sum_t f^d_{t+1}f^{d\top}_{t+1}\right)^{-1}H_F^{-1}.$$

Since $H_F^{-1}\to_p \mathcal{H}$ by Lemma I.10, we have the desired result.

(ii) Let $g_{js} = (a_sA - b_sB)x_{js}$ where $a_s = 1-(\eta^\top Q_t\Gamma(Q_t^B)^{-1} + \bar{\breve{f}}^\top)(Q^f)^{-1}f^d_{s+1}$, $b_s = B_{it}^\top(Q_t^B)^{-1}(Q_f)^{-1}f^d_{s+1}$, $A = x_{it}^\top Q_t^{-1} - B_{it}^\top(Q_t^B)^{-1}\Gamma^\top$, and $B = (\eta^\top - \eta^\top Q_t\Gamma(Q_t^B)^{-1}\Gamma^\top$. Then, we can have

$$\frac{1}{NTL}\sum_{j=1}^N\sum_{s=1}^T \sigma^2_{s+1}g^2_{js} = \frac{1}{NTL}\sum_{j=1}^N\sum_{s=1}^T \sigma^2_{s+1}(a_sA-b_sB)x_{js}x_{js}^\top(A^\top a_s - B^\top b_s)$$
$$= \frac{1}{TL}\sum_{s=1}^T \sigma^2_{s+1}(a_sA-b_sB)Q_s(A^\top a_s - B^\top b_s).$$

First, note that

$$\left\|\frac{1}{TL}\sum_{s=1}^T \hat{\sigma}^2_{s+1}\hat{a}_s^2\hat{A}Q_s\hat{A}^\top - \frac{1}{TL}\sum_{s=1}^T \sigma^2_{s+1}a_s^2AQ_sA^\top\right\| \tag{I.3}$$
$$\lesssim \left(\frac{1}{T}\sum_{s=1}^T(\hat{\sigma}^2_{s+1}-\sigma^2_{s+1})^2\right)^{1/2}\left(\frac{1}{TL^2}\sum_{s=1}^T\left\|a_s^2AQ_sA^\top\right\|^2\right)^{1/2}$$
$$+\max_s \sigma^2_{s+1}\frac{1}{TL}\sum_{s=1}^T\left\|\hat{a}_s^2\hat{A}Q_s\hat{A}^\top - a_s^2AQ_sA^\top\right\|$$
$$+\left(\frac{1}{T}\sum_{s=1}^T(\hat{\sigma}^2_{s+1}-\sigma^2_{s+1})^2\right)^{1/2}\left(\frac{1}{TL^2}\sum_{s=1}^T\left\|\hat{a}_s^2\hat{A}Q_s\hat{A}^\top - a_s^2AQ_sA^\top\right\|^2\right)^{1/2},$$

where $\hat{a}_s = 1-(\hat{\eta}^\top Q_t\hat{\Gamma}(\hat{\Gamma}^\top Q_t\hat{\Gamma})^{-1} + (\hat{\bar{f}})^\top)(\frac{1}{T}\sum_{u=1}^T \hat{f}^d_{u+1}\hat{f}^{d\top}_{u+1})^{-1}\hat{f}^d_{s+1}$ and $\hat{A} = x_{it}^\top Q_t^{-1} - x_{it}^\top\hat{\Gamma}(\hat{\Gamma}^\top Q_t\hat{\Gamma})^{-1}\hat{\Gamma}^\top$. Note that

$$\frac{1}{TL^2}\sum_{s=1}^T\left\|a_s^2AQ_sA^\top\right\|^2 \le \max_s\|Q_s\|^2\left(\frac{\|A\|}{\sqrt{L}}\right)^4\frac{1}{T}\sum_{s=1}^T a_s^2 = O_p(1)$$

since $\frac{1}{T}\sum_{s=1}^{T}||f_{s+1}^{d}||^2 = O_p(1)$. Hence, the first term converges to 0 by Lemma I.8. In addition, note that

$$\frac{1}{TL^2}\sum_{s=1}^{T}\left\|\hat{a}_s^2\hat{A}Q_s\hat{A}^\top - a_s^2AQ_sA^\top\right\|^2$$
$$\lesssim \max_s\|Q_s\|^2\left(\frac{\|A\|}{\sqrt{L}}\right)^4\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{a}_s^2-a_s^2\right\|^2 + \max_s\|Q_s\|^2\left(\frac{\left\|\hat{A}-A\right\|}{\sqrt{L}}\right)^2\left(\frac{\|A\|}{\sqrt{L}}\right)^2\frac{1}{T}\sum_{s=1}^{T}a_s^4.$$

By using similar bounds as in Lemmas I.4, I.5, and I.6, we have

$$\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{a}_s^2-a_s^2\right\|^2 \lesssim \frac{1}{T}\sum_{s=1}^{T}\|\hat{a}_s-a_s\|^2\|a_s\|^2 + \frac{1}{T}\sum_{s=1}^{T}\|\hat{a}_s-a_s\|^4 = o_p(1),$$

because

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\check{f}}_{t+1}-H_\Gamma^{-1}\check{f}_{t+1}\right\|^2\left\|f_{t+1}^{d}\right\|^2 = o_p(1), \quad \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\check{f}}_{t+1}-H_\Gamma^{-1}\check{f}_{t+1}\right\|^4 = o_p(1).$$

Moreover, we can easily check that $||\hat{A}-A||/\sqrt{L} = o_p(1)$ by Lemmas I.4 and I.12. Hence, we have

$$\frac{1}{TL^2}\sum_{s=1}^{T}\left\|\hat{a}_s^2\hat{A}Q_s\hat{A}^\top - a_s^2AQ_sA^\top\right\|^2 = o_p(1)$$

and, similarly, we can show $\frac{1}{TL}\sum_{s=1}^{T}\left\|\hat{a}_s^2\hat{A}Q_s\hat{A}^\top - a_s^2AQ_sA^\top\right\| = o_p(1)$. Then, with the aid of Lemma I.8, we can show that the second third terms of (I.3) are $o_p(1)$. Using a similar way, we can also show that

$$\left\|\frac{1}{TL}\sum_{s=1}^{T}\hat{\sigma}_{s+1}^2\hat{a}_s\hat{b}_s\hat{A}Q_s\hat{B}^\top - \frac{1}{TL}\sum_{s=1}^{T}\sigma_{s+1}^2a_sb_sAQ_sB^\top\right\| = o_p(1),$$
$$\left\|\frac{1}{TL}\sum_{s=1}^{T}\hat{\sigma}_{s+1}^2\hat{b}_s^2\hat{B}Q_s\hat{B}^\top - \frac{1}{TL}\sum_{s=1}^{T}\sigma_{s+1}^2b_s^2BQ_sB^\top\right\| = o_p(1).$$

(iii) By Lemma I.8, we have

$$\frac{1}{T}\sum_{s=1}^{T}\hat{\sigma}_{s+1}^2 - \frac{1}{T}\sum_{s=1}^{T}\sigma_{s+1}^2 = o_p(1).$$

(iv) By Lemma I.8, we have $\hat{\sigma}_{t+1}^2 - \sigma_{t+1}^2 = o_p(1)$. □

I.2 Auxiliary lemmas

Let $U_K D_K V_K^\top$ be the top-K singular value decomposition of $\ddot{R}^d$.

Lemma I.1 (naive bound of $\tilde{\Gamma}$)**.** *We have*

$$\left\|\tilde{\Gamma} - \Gamma H_\Gamma\right\|_F = O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\right).$$

Proof of Lemma I.1. By the equation (4), we have

$$\tilde{\Gamma} - \Gamma H_\Gamma = \ddot{E}^d F^{d\top} H_F^\top (\tilde{F}^d \tilde{F}^{d\top})^{-1} + \ddot{E}^d \ddot{E}^{d\top} \tilde{\Gamma} (\tilde{F}^d \tilde{F}^{d\top})^{-1},$$

since $\tilde{\Gamma}^\top \tilde{\Gamma} = U_K^\top U_K = I_K$. Note that $\|H_F\| = \left\|\tilde{\Gamma}^\top \Gamma\right\| \leq \|\Gamma\|$ is bounded and w.h.p.,

$$\sqrt{\psi_K\left(\tilde{F}^d \tilde{F}^{d\top}\right)} = \psi_K(D_K) \geq \psi_K(\Gamma F^d) - \left\|\ddot{E}^d\right\| \geq c\sqrt{T}$$

for some constant $c > 0$ since we have $\left\|\ddot{R}^d - \Gamma \mathcal{F}^d\right\| = \left\|\ddot{E}^d\right\| = O_p\left(\frac{\sqrt{LT}}{\sqrt{N}}\right)$ by Lemma I.9. Hence, we have

$$\left\|\tilde{\Gamma} - \Gamma H_\Gamma\right\| \leq \left\|\ddot{E}^d F^{d\top} H_F^\top (\tilde{F}^d \tilde{F}^{d\top})^{-1}\right\| + \left\|\ddot{E}^d \ddot{E}^{d\top} \tilde{\Gamma} (\tilde{F}^d \tilde{F}^{d\top})^{-1}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}} + \frac{L}{N}\right).$$

Because K is finite, we have the desired result. □

Lemma I.2. *Assume that* $\left\|\tilde{\Gamma} - \Gamma H_\Gamma\right\|_F = O_p\left(b_{NT}\right)$ *for some sequence* $b_{NT} \to 0$*. Then, we have*

$$\left\|\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}} b_{NT} + \frac{1}{\sqrt{N}}\right), \quad \frac{1}{T}\sum_{t=1}^T \left\|\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}} b_{NT} + \frac{1}{\sqrt{N}}\right),$$
$$\frac{1}{T}\sum_{t=1}^T \left\|\tilde{f}_{t+1}^d - H_F f_{t+1}^d\right\|^2 = O_p\left(\left(\frac{\sqrt{L}}{\sqrt{N}} b_{NT} + \frac{1}{\sqrt{N}}\right)^2\right).$$

Proof of Lemma I.2. We have

$$\tilde{f}_{t+1}^d - H_F f_{t+1}^d = \left(\tilde{\Gamma} - \Gamma H_\Gamma\right)^\top \ddot{E}_{t+1}^d + H_\Gamma^\top \Gamma^\top \ddot{E}_{t+1}^d. \tag{I.4}$$

By Lemma I.9, the order of the first term in (I.4) is $O_p\left(\frac{\sqrt{L}}{\sqrt{N}} b_{NT}\right)$ and that of the second term is $O_p\left(\frac{1}{\sqrt{N}}\right)$. Here, we use the fact that

$$\|H_\Gamma\| = \left\|F^d \tilde{F}^{d\top} (\tilde{F}^d \tilde{F}^{d\top})^{-1}\right\| \leq \left\|F^d\right\| \|V_K\| \left\|D_K^{-1}\right\| = O_p(1).$$

In addition, note that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^{d}-H_F^{-1}f_{t+1}^{d}\right\| \leq \left\|\tilde{\Gamma}-\Gamma H_\Gamma\right\|\frac{1}{T}\sum_{t=1}^{T}\left\|\ddot{E}_{t+1}^{d}\right\| + \|H_\Gamma\|\frac{1}{T}\sum_{t=1}^{T}\left\|\Gamma^\top\ddot{E}_{t+1}^{d}\right\|.$$

Hence, using the same method as above with Lemma I.9, we can have $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}}b_{NT}+\frac{1}{\sqrt{N}}\right)$. In addition, by Lemma I.9, we have

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left\|\ddot{E}_{t+1}^{d}\right\|^2\middle| X\right] = O_p\left(\frac{L}{N}\right), \quad \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left\|\Gamma^\top\ddot{E}_{t+1}^{d}\right\|^2\middle| X\right] = O_p\left(\frac{1}{N}\right).$$

Using this result, we have $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right\|^2 = O_p\left(\left(\frac{\sqrt{L}}{\sqrt{N}}b_{NT}+\frac{1}{\sqrt{N}}\right)^2\right)$. □

Lemma I.3. *Assume that there is a sequence $a_{NT}\to 0$ such that*

$$\left\|\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right\| = O_p\left(a_{NT}\right), \quad \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right\| = O_p\left(a_{NT}\right),$$
$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right\|^2 = O_p\left(a_{NT}^2\right)$$

Then, we have

$$\left\|\tilde{\Gamma}-\Gamma H_\Gamma\right\|_F = O_p\left(\frac{\sqrt{L}}{\sqrt{N}}a_{NT}+\frac{\sqrt{L}}{\sqrt{NT}}\right).$$

Proof of Lemma I.3. We have

$$\begin{aligned}\operatorname{vec}\left(\tilde{\Gamma}^\top\right)-\operatorname{vec}\left(H_\Gamma^\top\Gamma^\top\right) = &\left(I_L\otimes\left(\frac{1}{T}\sum_t\tilde{f}_{t+1}^{d}\tilde{f}_{t+1}^{d\top}\right)^{-1}\right)\frac{1}{T}\sum_{t=1}^{T}\ddot{E}_{t+1}^{d}\otimes\left(\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right)\\ &+\left(I_L\otimes\left(\frac{1}{T}\sum_t\tilde{f}_{t+1}^{d}\tilde{f}_{t+1}^{d\top}\right)^{-1}\right)\frac{1}{T}\sum_{t=1}^{T}\ddot{E}_{t+1}^{d}\otimes\left(H_F f_{t+1}^{d}\right). \quad \text{(I.5)}\end{aligned}$$

First, Note that

$$\begin{aligned}\psi_{\min}\left(\frac{1}{T}\sum_t\tilde{f}_{t+1}^{d}\tilde{f}_{t+1}^{d\top}\right) &\geq \psi_{\min}\left(H_F\frac{1}{T}\sum_t f_{t+1}^{d}f_{t+1}^{d\top}H_F^\top\right)\\ &\quad-\left\|\frac{1}{T}\sum_t\tilde{f}_{t+1}^{d}\tilde{f}_{t+1}^{d\top}-H_F\frac{1}{T}\sum_t f_{t+1}^{d}f_{t+1}^{d\top}H_F^\top\right\|\\ &\geq c_1,\end{aligned}$$

for some constant $c_1 > 0$, w.p.c. to 1, because

$$\begin{aligned}
&\left\| \frac{1}{T}\sum_t \tilde{f}^d_{t+1}\tilde{f}^{d\top}_{t+1} - H_F \frac{1}{T}\sum_t f^d_{t+1} f^{d\top}_{t+1} H_F^\top \right\| \\
&\lesssim \left\| H_F \frac{1}{T}\sum_t f^d_{t+1}(\tilde{f}^d_{t+1} - H_F f^d_{t+1})^\top \right\| \\
&\leq \|H_F\| \left(\frac{1}{T}\sum_t \left\| f^d_{t+1} \right\|^2 \right)^{1/2} \left(\frac{1}{T}\sum_t \left\| \tilde{f}^d_{t+1} - H_F f^d_{t+1} \right\|^2 \right)^{1/2} = O_p(a_{NT}) = o_p(1),
\end{aligned}$$

and $\psi_{\min}\left(H_F \frac{1}{T}\sum_t f^d_{t+1} f^{d\top}_{t+1} H_F^\top\right) \geq \psi^2_{\min}(H_F)\,\psi_{\min}\left(\frac{1}{T}\sum_t f^d_{t+1} f^{d\top}_{t+1}\right) \geq c_2$ for some constant $c_2 > 0$ w.p.c. to 1. Here, $\psi_{\min}(H_F) = \psi_{\min}\left(\tilde{\Gamma}^\top \Gamma\right) > c$ comes from a typical assertion such as Proposition 1 of Bai (2003). Hence, $\left\| \left(\frac{1}{T}\sum_t \tilde{f}^d_{t+1}\tilde{f}^{d\top}_{t+1} \right)^{-1} \right\| = O_p(1)$. In addition, we have $\frac{1}{T}\sum_{t=1}^T \ddot{E}^d_{t+1} \otimes \left(\tilde{f}^d_{t+1} - H_F f^d_{t+1}\right) = \frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1} \otimes \left(\tilde{f}^d_{t+1} - H_F f^d_{t+1}\right)$ and

$$\begin{aligned}
\left\| \frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1} \otimes \left(\tilde{f}^d_{t+1} - H_F f^d_{t+1}\right) \right\| &\leq \left(\frac{1}{T}\sum_{t=1}^T \left\| \ddot{E}_{t+1} \right\|^2 \right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^T \left\| \tilde{f}^d_{t+1} - H_F f^d_{t+1} \right\|^2 \right)^{1/2} \\
&= O_p\left(\frac{\sqrt{L}}{\sqrt{N}} a_{NT} \right),
\end{aligned}$$

by Lemma I.9. Moreover, $\frac{1}{T}\sum_{t=1}^T \ddot{E}^d_{t+1} \otimes \left(H_F f^d_{t+1}\right) = (I_L \otimes H_F)\frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1} \otimes f^d_{t+1}$ and $\left\| \frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1} \otimes f^d_{t+1} \right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$ because

$$\frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1} \otimes f^d_{t+1} = \frac{1}{NT}\sum_{it} (Q_t^{-1} x_{it} \otimes f^d_{t+1}) \epsilon_{i,t+1} = \frac{1}{NT} A^\top \mathrm{vec}(E)$$

where A is the $NT \times LK$ matrix of $Q_t^{-1} x_{it} \otimes f^d_{t+1}$ and

$$\mathbb{E}\left\| \frac{1}{NT} A^\top \mathrm{vec}(E) \right\|_F^2 \lesssim \frac{1}{N^2 T^2} \left\| \mathbb{E}[\mathrm{vec}(E)\mathrm{vec}(E)^\top] \right\| \|A\|_F^2 = O_p\left(\frac{L}{NT}\right)$$

since $\|A\|_F^2 \leq \max_t ||Q_t^{-1}||^2 \sum_t ||f^d_{t+1}||^2 \sum_i ||x_{it}||^2 = O_p(NTL)$. Then, applying these bounds to (I.5), we have

$$\left\| \tilde{\Gamma} - \Gamma H_\Gamma \right\|_F = \left\| \mathrm{vec}\left(\tilde{\Gamma}^\top\right) - \mathrm{vec}\left(H_\Gamma^\top \Gamma^\top\right) \right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}} a_{NT} + \frac{\sqrt{L}}{\sqrt{NT}} \right). \quad \square$$

Lemma I.4. *We have*

$$\left\| \tilde{f}^d_{t+1} - H_F f^d_{t+1} \right\|, \quad \frac{1}{T}\sum_{t=1}^T \left\| \tilde{f}^d_{t+1} - H_F f^d_{t+1} \right\| = O_p\left(\left(\frac{L}{N}\right)^{10} + \frac{L}{N\sqrt{T}} + \frac{1}{\sqrt{N}} \right),$$

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_{t+1}^{d}-H_F f_{t+1}^{d}\right\|^2=O_p\left(\left(\left(\frac{L}{N}\right)^{10}+\frac{L}{N\sqrt{T}}+\frac{1}{\sqrt{N}}\right)^2\right),$$
$$\left\|\tilde{\Gamma}-\Gamma H_\Gamma\right\|_F=O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}+\frac{\sqrt{L}}{N}\right).$$

Proof of Lemma I.4. Starting from the naive bound in Lemma I.1, by applying Lemmas I.2 and I.3 recursively, we can derive sharper and sharper bounds. But the number of recursions should be bounded. We apply Lemmas I.2 and I.3 10 times here. □

Lemma I.5. *Let* $\breve{\tilde{f}}_{t+1}=(\tilde{\Gamma}^\top\tilde{\Gamma})^{-1}\tilde{\Gamma}^\top\ddot{R}_{t+1}$. *We have*

$$\left\|\breve{\tilde{f}}_{t+1}-H_\Gamma^{-1}\breve{f}_{t+1}\right\|=O_p\left(\frac{1}{\sqrt{N}}+\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}\right),$$
$$\frac{1}{T}\sum_{t=1}^{T}\left\|\breve{\tilde{f}}_{t+1}-H_\Gamma^{-1}\breve{f}_{t+1}\right\|=O_p\left(\frac{1}{\sqrt{N}}+\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}\right),$$
$$\frac{1}{T}\sum_{t=1}^{T}\left\|\breve{\tilde{f}}_{t+1}-H_\Gamma^{-1}\breve{f}_{t+1}\right\|^2=O_p\left(\left[\frac{1}{\sqrt{N}}+\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}\right]^2\right).$$

Proof of Lemma I.5. Note that

$$\breve{\tilde{f}}_{t+1}-H_\Gamma^{-1}\breve{f}_{t+1}=(\tilde{\Gamma}-\Gamma H_\Gamma)^\top\eta-\tilde{\Gamma}^\top(\tilde{\Gamma}-\Gamma H_\Gamma)H_\Gamma^{-1}\breve{f}_{t+1}+\tilde{\Gamma}^\top\ddot{E}_{t+1}.$$

We have $\left\|\tilde{\Gamma}-\Gamma H_\Gamma\right\|_F=O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}+\frac{\sqrt{L}}{N}\right)$ by Lemma I.4. In addition, using the same token as in the proof of Lemma I.2, we have

$$\tilde{\Gamma}^\top\ddot{E}_{t+1}=O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\left(\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}+\frac{\sqrt{L}}{N}\right)+\frac{1}{\sqrt{N}}\right).$$

Hence, we have

$$\left\|\breve{\tilde{f}}_{t+1}-H_\Gamma^{-1}\breve{f}_{t+1}\right\|=O_p\left(\frac{1}{\sqrt{N}}+\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}\right).$$

Similarly, we can show the other bounds. □

Lemma I.6. *We have (i)* $\|\tilde{\eta}-\eta\|=O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}+\frac{\sqrt{L}}{N}\right)$; *(ii)* $\left\|x_{it}^\top\tilde{\eta}-x_{it}^\top\eta\right\|=O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}+\left(\frac{L}{N}\right)^{10+\frac{1}{2}}+\frac{\sqrt{L}}{N}+\frac{L^{3/2}}{N^{3/2}\sqrt{T}}\right)$.

Proof of Lemma I.6. (i) Note that

$$\tilde{\eta}-\eta=(P_\Gamma-P_{\tilde{\Gamma}})\left(\eta+\Gamma\bar{\breve{f}}+\bar{\ddot{E}}\right)+(I_L-P_\Gamma)\,\bar{\ddot{E}}$$

where $P_\Gamma = \Gamma\left(\Gamma^\top\Gamma\right)^{-1}\Gamma^\top$ and $P_{\tilde{\Gamma}} = \tilde{\Gamma}\left(\tilde{\Gamma}^\top\tilde{\Gamma}\right)^{-1}\tilde{\Gamma}^\top$. From Lemma I.4, we have $\left\|\tilde{\Gamma} - H_\Gamma\Gamma\right\|_F = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N}\right)$. So, a simple calculation using Lemma I.13 shows that $\|P_\Gamma - P_{\tilde{\Gamma}}\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N}\right)$. In addition, we have $\left\|\ddot{\bar{E}}\right\| = \left\|\frac{1}{T}\sum_{t=1}^T \ddot{E}_{t+1}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$ by Lemma I.9. Then, since $\|P_\Gamma\| \leq 1$, we have

$$\|\tilde{\eta} - \eta\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N}\right).$$

(ii) The proof is similar to (i). Here, we use Lemma I.12 (i). □

Lemma I.7. *Let* $\hat{\sigma}^2_{t+1} = \frac{1}{N}\sum_{i=1}^N \hat{\epsilon}^2_{i,t+1}$, *where* $m_{i,t+1} = \alpha_{O,it} + x_{it}^\top\eta + x_{it}^\top\Gamma\breve{f}_{t+1}$, $\tilde{m}_{i,t+1} = \hat{\alpha}_{O,it} + x_{it}^\top\tilde{\eta} + x_{it}^\top\tilde{\Gamma}\tilde{\breve{f}}_{t+1}$, *and* $\hat{\epsilon}_{i,t+1} = r_{i,t+1} - \tilde{m}_{i,t+1}$. *Then, we have (i)* $\left|\hat{\sigma}^2_{t+1} - \sigma^2_{t+1}\right| = o_p(1)$, *(ii)* $\frac{1}{T}\sum_{t=1}^T \left|\hat{\sigma}^2_{t+1} - \sigma^2_{t+1}\right| = o_p(1)$, *(iii)* $\frac{1}{T}\sum_{t=1}^T \left|\hat{\sigma}^2_{t+1} - \sigma^2_{t+1}\right|^2 = o_p(1)$.

Proof of Lemma I.7. (i) Note that

$$\begin{aligned}
\hat{\sigma}^2_{t+1} - \mathbb{E}[\epsilon^2_{j,t+1}] &= \frac{1}{N}\sum_{j=1}^N \left(\epsilon^2_{j,t+1} - \mathbb{E}[\epsilon^2_{j,t+1}]\right) + \frac{1}{N}\sum_{j=1}^N \left(\hat{\epsilon}^2_{j,t+1} - \epsilon^2_{j,t+1}\right) \\
&= \frac{1}{N}\sum_{j=1}^N \left(\epsilon^2_{j,t+1} - \mathbb{E}[\epsilon^2_{j,t+1}]\right) + 2\frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}\left(\tilde{m}_{j,t+1} - m_{j,t+1}\right) \\
&\quad + \frac{1}{N}\sum_{j=1}^N \left(\tilde{m}_{j,t+1} - m_{j,t+1}\right)^2.
\end{aligned}$$

The first term is $o_p(1)$ by the concentration inequality. For the second term, we can derive

$$\frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}\left(\hat{\alpha}_{O,jt} - \alpha_{O,jt}\right) = o_p(1)$$

by the same token as in the proof of Lemma I.11. In addition, we have

$$\frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}\left(x_{jt}^\top\eta - x_{jt}^\top\tilde{\eta}\right) = \frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}x_{jt}^\top\left(\eta - \tilde{\eta}\right) = o_p(1)$$

by Lemma I.6 since $\left\|\frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}x_{jt}^\top\right\| = O_p(\frac{\sqrt{L}}{\sqrt{N}})$. Moreover, we have

$$\frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}x_{it}^\top(\tilde{\Gamma} - \Gamma H_\Gamma)H_\Gamma^{-1}\breve{f}_{t+1} = o_p(1), \quad \frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}B_{jt}^\top H_\Gamma(\tilde{\breve{f}}_{t+1} - H_\Gamma^{-1}\breve{f}_{t+1}) = o_p(1)$$

by Lemmas I.4 and I.5. Hence, we have $\frac{1}{N}\sum_{j=1}^N \epsilon_{j,t+1}\left(\tilde{m}_{j,t+1} - m_{j,t+1}\right) = o_p(1)$. For the third term, we have $\frac{1}{N}\sum_{j=1}^N \left(\hat{\alpha}_{O,jt} - \alpha_{O,jt}\right)^2 = o_p(1)$ by using the bound of $\mathbb{E}\left[\left(\hat{\alpha}_{O,jt} - \alpha_{O,jt}\right)^2 \middle| X\right]$

as in the proof of Lemma I.11. In addition, we have

$$\frac{1}{N}\sum_{j=1}^{N}\left(x_{jt}^{\top}\eta - x_{jt}^{\top}\tilde{\eta}\right)^2 = (\eta-\tilde{\eta})^{\top} Q_t (\eta-\tilde{\eta}) = o_p(1).$$

Moreover, we have

$$\frac{1}{N}\sum_{j=1}^{N}(x_{it}^{\top}(\tilde{\Gamma}-\Gamma H_{\Gamma})H_{\Gamma}^{-1}\breve{f}_{t+1})^2 = \breve{f}_{t+1}^{\top}H_{\Gamma}^{-\top}(\tilde{\Gamma}-\Gamma H_{\Gamma})^{\top}Q_t(\tilde{\Gamma}-\Gamma H_{\Gamma})H_{\Gamma}^{-1}\breve{f}_{t+1} = o_p(1).$$

Lastly, we have

$$\frac{1}{N}\sum_{j=1}^{N}(B_{jt}^{\top}H_{\Gamma}(\tilde{\breve{f}}_{t+1}-H_{\Gamma}^{-1}\breve{f}_{t+1}))^2 = (\tilde{\breve{f}}_{t+1}-H_{\Gamma}^{-1}\breve{f}_{t+1})^{\top}H_{\Gamma}^{\top}Q_B H_{\Gamma}(\tilde{\breve{f}}_{t+1}-H_{\Gamma}^{-1}\breve{f}_{t+1}) = o_p(1).$$

Hence, we have $\frac{1}{N}\sum_{j=1}^{N}(\tilde{m}_{j,t+1}-m_{j,t+1})^2 = o_p(1)$.

(ii) The proof is similar to that of (iii).

(iii) First, note that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\epsilon_{j,t+1}^2 - \mathbb{E}[\epsilon_{j,t+1}^2]\right)\right\|^2 = o_p(1)$$

because $\mathbb{E}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\epsilon_{j,t+1}^2-\mathbb{E}[\epsilon_{j,t+1}^2]\right)\right\|^2 = O\left(\frac{1}{N}\right)$. In addition, we can show that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}\left(\hat{\alpha}_{O,jt}-\alpha_{O,jt}\right)\right\|^2 = o_p(1)$$

by using the similar method as in the proof of Lemma I.11. Moreover, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}\left(x_{jt}^{\top}\eta - x_{jt}^{\top}\tilde{\eta}\right)\right\|^2 \le \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}x_{jt}\right\|^2\|\eta-\tilde{\eta}\|^2 = o_p(1)$$

since $\mathbb{E}[||\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}x_{jt}||^2|X] \lesssim \frac{L}{N}$. In addition, we can bound

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}x_{it}^{\top}(\tilde{\Gamma}-\Gamma H_{\Gamma})H_{\Gamma}^{-1}\breve{f}_{t+1}\right\|^2 \le \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}x_{jt}\right\|^2\left\|\breve{f}_{t+1}\right\|^2\left\|\tilde{\Gamma}-\Gamma H_{\Gamma}\right\|^2 = o_p(1)$$

since $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[||\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}x_{jt}||^2||\breve{f}_{t+1}||^2|X,F] \lesssim \frac{L}{N}$. Besides, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}B_{jt}^{\top}H_{\Gamma}(\hat{f}_{t+1}-H_{\Gamma}^{-1}f_{t+1})\right\|^2$$

$$\leq \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{j=1}^{N} \epsilon_{j,t+1} B_{jt} \right\|^4 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1} \right\|^4 \right)^{1/2} = o_p(1)$$

since $\mathbb{E}[||\frac{1}{N}\sum_{j=1}^{N} \epsilon_{j,t+1} B_{jt}||^4 | X] \lesssim \frac{1}{N}$. Here, we can show that $\frac{1}{T}\sum_{t=1}^{T} \left\| \tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1} \right\|^4 = o_p(1)$ in a similar way to the proof of Lemmas I.5 and I.2 with a weak dependence of noises across i. Hence, we have

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{j=1}^{N} \epsilon_{j,t+1} \left(\tilde{m}_{j,t+1} - m_{j,t+1} \right) \right\|^2 = o_p(1).$$

Moreover, we can show that $\frac{1}{T}\sum_{t=1}^{T} \left\| \frac{1}{N}\sum_{j=1}^{N} \left(\hat{\alpha}_{O,jt} - \alpha_{O,jt} \right)^2 \right\|^2 = o_p(1)$ by using a concentration inequality for the sub-Gaussian random variable like Lemma I.15. In addition, we have $\frac{1}{T}\sum_{t=1}^{T} \left\| \frac{1}{N}\sum_{j=1}^{N} (x_{jt}^\top (\tilde{\eta} - \eta))^2 \right\|^2 \lesssim \|\tilde{\eta} - \eta\|^4 = o_p(1)$. Similarly, we have $\frac{1}{T}\sum_{t=1}^{T} \left\| \frac{1}{N}\sum_{j=1}^{N} (x_{jt}^\top (\tilde{\Gamma} - \Gamma H_\Gamma) \breve{f}_{t+1})^2 \right\|^2 \lesssim \frac{1}{T}\sum_{t=1}^{T} \left\| \breve{f}_{t+1} \right\|^4 \left\| \tilde{\Gamma} - \Gamma H_\Gamma \right\|^4 = o_p(1)$. Lastly, we have

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{j=1}^{N} (x_{jt}^\top \Gamma H_\Gamma (\tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1}))^2 \right\|^2 \lesssim \frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1}) \right\|^4 = o_p(1).$$

Hence, we have

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{j=1}^{N} \left(\tilde{m}_{j,t+1} - m_{j,t+1} \right)^2 \right\|^2 = o_p(1). \quad \square$$

Lemma I.8. *Let* $\hat{\sigma}_{t+1}^2 = \frac{1}{N}\sum_{i=1}^{N} \hat{\epsilon}_{i,t+1}^2$, *where* $m_{i,t+1} = \alpha_{O,it} + x_{it}^\top \eta + x_{it}^\top \Gamma \breve{f}_{t+1}$, $\hat{m}_{i,t+1} = \hat{\alpha}_{O,it} + x_{it}^\top \hat{\eta} + x_{it}^\top \hat{\Gamma} \hat{\tilde{f}}_{t+1}$, *and* $\hat{\epsilon}_{i,t+1} = r_{i,t+1} - \hat{m}_{i,t+1}$. *Then, we have (i)* $\left| \hat{\sigma}_{t+1}^2 - \sigma_{t+1}^2 \right| = o_p(1)$, *(ii)* $\frac{1}{T}\sum_{t=1}^{T} \left| \hat{\sigma}_{t+1}^2 - \sigma_{t+1}^2 \right| = o_p(1)$, *(iii)* $\frac{1}{T}\sum_{t=1}^{T} \left| \hat{\sigma}_{t+1}^2 - \sigma_{t+1}^2 \right|^2 = o_p(1)$.

Proof of Lemma I.8. The proof is the same as that of Lemma I.7 and we omit it here. $\square$

I.3 Technical lemmas

Lemma I.9. *(i)* $\left\| (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}} \right)$; *(ii)* $\frac{1}{T}\sum_{t=1}^{T} \left\| (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\|^2 = O_p\left(\frac{L}{N} \right)$; *(iii)* $\frac{1}{T}\sum_{t=1}^{T} \left\| (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}} \right)$; *(iv)* $\left\| \frac{1}{T}\sum_t (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} \right)$; *(v)* $\left\| \Gamma^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\| = O_p\left(\frac{1}{\sqrt{N}} \right)$, *(vi)* $\frac{1}{T}\sum_{t=1}^{T} \left\| \Gamma^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\| = O_p\left(\frac{1}{\sqrt{N}} \right)$; *(vii)* $\frac{1}{T}\sum_{t=1}^{T} \left\| \Gamma^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\|^2 = O_p\left(\frac{1}{N} \right)$;
(viii) $\max_t \mathbb{E}[(e_l^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1})^2 | X] = O_p(\frac{1}{N})$; *(ix)* $||\frac{1}{T}\sum_{t=1}^{T} (e_l^\top \ddot{E}_{t+1})|| = O_p\left(\frac{1}{\sqrt{NT}} \right)$.

Proof of Lemma I.9. (i) Note that

$$\begin{aligned}\mathbb{E}\left[\left\|(X_t^\top X_t)^{-1}X_t^\top E_t\right\|^2\middle| X\right] &= \sum_{l=1}^{L} e_l^\top (X_t^\top X_t)^{-1}X_t^\top \mathbb{E}\left[E_{t+1}E_{t+1}^\top\middle| X\right] X_t (X_t^\top X_t)^{-1} e_l \\ &\le \sum_{l=1}^{L} \left\|e_l^\top (X_t^\top X_t)^{-1}X_t^\top\right\|^2 \left\|\mathbb{E}\left[E_{t+1}E_{t+1}^\top\middle| X\right]\right\| \\ &= \left\|(X_t^\top X_t)^{-1}X_t^\top\right\|_F^2 \left\|\mathbb{E}\left[E_{t+1}E_{t+1}^\top\middle| X\right]\right\| = O_p\left(\frac{L}{N}\right).\end{aligned}$$

by Assumptions 1 and 3. Hence, we have $\left\|(X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\right)$.

(ii) Because

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\left\|(X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right\|^2\middle| X\right] &= \sum_{t=1}^{T}\mathbb{E}\left[\left\|(X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right\|^2\middle| X\right] \\ &\le \max_t \left\|(X_t^\top X_t)^{-1}\right\|^2 \max_t \left\|\mathbb{E}\left[E_{t+1}E_{t+1}^\top\middle| X\right]\right\| \sum_{t=1}^{T}\|X_t\|_F^2 \\ &= O_p\left(\frac{LT}{N}\right),\end{aligned}$$

we have $\frac{1}{T}\sum_{t=1}^{T}\left\|(X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right\|^2 = O_p\left(\frac{L}{N}\right)$.

(iii) Note that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left\|(X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right\|\middle| X\right] &\le \frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[\left\|(X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right\|^2\middle| X\right]\right)^{1/2} \\ &\le \max_t \left\|(X_t^\top X_t)^{-1}\right\| \max_t \left\|\mathbb{E}\left[E_{t+1}E_{t+1}^\top\middle| X\right]\right\|^{1/2} \frac{1}{T}\sum_{t=1}^{T}\|X_t\|_F \\ &= O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\right).\end{aligned}$$

(iv) Note that

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T}(X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right\|^2 &= \sum_{l=1}^{L}\left(\frac{1}{T}\sum_{t=1}^{T}e_l^\top (X_t^\top X_t)^{-1}X_t^\top E_{t+1}\right)^2 \\ &= \frac{1}{N^2T^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}e_l^\top (X_t^\top X_t/N)^{-1}x_{it}\epsilon_{i,t+1}\right)^2 \\ &= \frac{1}{N^2T^2}\sum_{l=1}^{L}\left(A_l^\top \mathrm{vec}(E)\right)^2,\end{aligned}$$

where A_l is the $NT \times 1$ vector whose (i,t)-th element is $e_l^\top (X_t^\top X_t/N)^{-1} x_{it}$. Hence, we have

$$\mathbb{E}\left\| \frac{1}{T}\sum_{t=1}^{T} (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\|^2 \leq \frac{1}{N^2T^2} \sum_{l=1}^{L} \|A_l\|^2 \left\| \mathbb{E}\left[\mathrm{vec}(E)\mathrm{vec}(E)^\top | X \right] \right\| = O_p\left(\frac{L}{NT}\right)$$

because $\max_t e_l^\top (X_t^\top X_t/N)^{-1} e_l \leq \max_t \left\|(X_t^\top X_t/N)^{-1}\right\| \|e_l\|^2 \leq C$ for some constant $C > 0$ and

$$\sum_{l=1}^{L} \|A_l\|^2 = \sum_{l=1}^{L}\sum_{i=1}^{N}\sum_{t=1}^{T} e_l^\top (X_t^\top X_t/N)^{-1} x_{it} x_{it}^\top (X_t^\top X_t/N)^{-1} e_l = N \sum_{l=1}^{L}\sum_{t=1}^{T} e_l^\top (X_t^\top X_t/N)^{-1} e_l.$$

(v) Note that

$$\begin{aligned}
\mathbb{E}\left[\left\| \Gamma^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\|^2 \middle| X \right] &= \sum_{k=1}^{K} e_k^\top \Gamma^\top (X_t^\top X_t)^{-1} X_t^\top \mathbb{E}\left[E_{t+1}E_{t+1}^\top \middle| X \right] X_t (X_t^\top X_t)^{-1} \Gamma e_k \\
&\leq \left\| \mathbb{E}\left[E_{t+1}E_{t+1}^\top \middle| X \right] \right\| \left\| \Gamma^\top (X_t^\top X_t)^{-1} X_t^\top \right\|_F^2 \\
&\lesssim \left\| \mathbb{E}\left[E_{t+1}E_{t+1}^\top \middle| X \right] \right\| \left\| \Gamma^\top (X_t^\top X_t)^{-1} X_t^\top \right\|^2 \\
&\lesssim \|\Gamma\|^2 \left\| (X_t^\top X_t)^{-1} X_t^\top \right\|^2 = O_p\left(\frac{1}{N}\right).
\end{aligned}$$

(vi), (vii) trivially follow from (v).

(viii) We have

$$\begin{aligned}
\max_t \mathbb{E}[(e_l^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1})^2 | X] &= \max_t e_l^\top (X_t^\top X_t)^{-1} X_t^\top \mathbb{E}[E_{t+1}E_{t+1}^\top] X_t (X_t^\top X_t)^{-1} e_l \\
&\leq \max_t \left\| \mathbb{E}[E_{t+1}E_{t+1}^\top] \right\| \max_t \left\| X_t (X_t^\top X_t)^{-1} \right\|^2 \\
&= O_p\left(\frac{1}{N}\right).
\end{aligned}$$

(ix) Note that

$$\begin{aligned}
\left\| \frac{1}{T}\sum_{t=1}^{T} e_l^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\|^2 &= \frac{1}{N^2T^2} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} e_l^\top (X_t^\top X_t/N)^{-1} x_{it} \epsilon_{i,t+1} \right)^2 \\
&= \frac{1}{N^2T^2} \left(A_l^\top \mathrm{vec}(E) \right)^2,
\end{aligned}$$

where A_l is the $NT \times 1$ vector whose (i,t)-th element is $e_l^\top (X_t^\top X_t/N)^{-1} x_{it}$. Hence, we have

$$\mathbb{E}\left\| \frac{1}{T}\sum_{t=1}^{T} e_l^\top (X_t^\top X_t)^{-1} X_t^\top E_{t+1} \right\|^2 \leq \frac{1}{N^2T^2} \|A_l\|^2 \left\| \mathbb{E}\left[\mathrm{vec}(E)\mathrm{vec}(E)^\top | X \right] \right\| = O_p\left(\frac{1}{NT}\right),$$

because $\max_t e_l^\top (X_t^\top X_t/N)^{-1} e_l \leq \max_t \left\|(X_t^\top X_t/N)^{-1}\right\| \|e_l\|^2 \leq C$ for some constant $C > 0$

and $\|A_l\|^2 = \sum_{i=1}^N \sum_{t=1}^T e_l^\top (X_t^\top X_t/N)^{-1} x_{it} x_{it}^\top (X_t^\top X_t/N)^{-1} e_l = N \sum_{t=1}^T e_l^\top (X_t^\top X_t/N)^{-1} e_l$.
□

Lemma I.10. *(i) W.h.p., $\|H_F\|$ and $\left\|H_F^{-1}\right\|$ are bounded. (ii) W.h.p., $\|H_\Gamma\|$ and $\left\|H_\Gamma^{-1}\right\|$ are bounded. (iii) $\left\|H_F - H_\Gamma^{-1}\right\| = o_p(1)$ and $\left\|H_\Gamma - H_F^{-1}\right\| = o_p(1)$. (iv) $H_F \to_p I_{sgn} \boldsymbol{G}^{-1} \left(\Gamma^\top \Gamma\right)^{1/2}$ and $H_F^{-1} \to_p \left(\Gamma^\top \Gamma\right)^{-1/2} \boldsymbol{G} I_{sgn}$.*

Proof of Lemma I.10. (i) First, $\|H_F\| = \left\|\tilde{\Gamma}^\top \Gamma\right\| \leq \|\Gamma\|$ is bounded. In addition, by a typical assertion for the spectral method such as Proposition 1 of Bai (2003), we have $\psi_{\min}(H_F) = \psi_{\min}\left(\tilde{\Gamma}^\top \Gamma\right) > c$. Hence, $\left\|H_F^{-1}\right\|$ is bounded.
(ii) Simple calculation shows that

$$H_\Gamma^\top = \left[\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1} \left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d (H_F f_{t+1}^d - \tilde{f}_{t+1}^d)^\top\right) + I_K\right] H_F^{-\top}.$$

Let $U_K D_K V_K^\top$ be the top-K singular value decomposition of $\ddot{R}^d$. Note that w.h.p.,

$$\sqrt{\psi_K\left(\tilde{F}^d \tilde{F}^{d\top}\right)} = \psi_K(D_K) \geq \psi_K(\Gamma F^d) - \left\|\ddot{E}^d\right\| \geq c\sqrt{T}$$

for some constant $c > 0$ since we have $\left\|\ddot{R}^d - \Gamma \mathcal{F}^d\right\| = \left\|\ddot{E}^d\right\| = O_p\left(\frac{\sqrt{LT}}{\sqrt{N}}\right)$ by Lemma I.9. Hence,

$$\psi_{\min}\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right) = \frac{1}{T}\psi_{\min}\left(\tilde{F}^d \tilde{F}^{d\top}\right) = \frac{1}{T}\psi_{\min}(D_K)^2 > c,$$

for some constant $c > 0$ and $\left\|\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1}\right\|$ is bounded. In addition,

$$\left\|\frac{1}{T}\sum_t \tilde{f}_{t+1}^d (H_F f_{t+1}^d - \tilde{f}_{t+1}^d)^\top\right\| = o_p(1)$$

by Lemma I.4. Hence, w.h.p.,

$$\begin{aligned}
&\psi_{\min}\left[\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1} \left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d (H_F f_{t+1}^d - \tilde{f}_{t+1}^d)^\top\right) + I_K\right] \\
&\geq 1 - \left\|\left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top}\right)^{-1} \left(\frac{1}{T}\sum_t \tilde{f}_{t+1}^d (H_F f_{t+1}^d - \tilde{f}_{t+1}^d)^\top\right)\right\| \\
&\geq \frac{1}{2}.
\end{aligned}$$

Then, $\psi_{\min}(H_\Gamma) \geq \frac{1}{2}\psi_{\min}(H_F^{-1}) > c$ for some $c > 0$. Hence, $\left\|H_\Gamma^{-1}\right\|$ is bounded. In addition,

because $\left\| \left(\frac{1}{T} \sum_t \tilde{f}^d_{t+1} \tilde{f}^{d\top}_{t+1} \right)^{-1} \left(\frac{1}{T} \sum_t \tilde{f}^d_{t+1} (H_F f^d_{t+1} - \tilde{f}^d_{t+1})^\top \right) \right\| = o_p(1)$, $\left\| H_\Gamma - H_F^{-1} \right\| = o_p(1)$ and $\|H_\Gamma\|$ is also bounded.

(iii) We prove $\left\| H_\Gamma - H_F^{-1} \right\| = o_p(1)$ in (ii). In addition, we have

$$\left\| H_\Gamma^{-1} - H_F \right\| \leq \|H_F\| \left\| H_\Gamma - H_F^{-1} \right\| \left\| H_\Gamma^{-1} \right\| = o_p(1).$$

(iv) Let $\Omega = \left(\Gamma^\top \Gamma\right)^{1/2} \left(\frac{1}{T} F^{d\top} F^d \right) \left(\Gamma^\top \Gamma\right)^{1/2}$ and G be a $K \times K$ matrix whose columns are the eigenvectors of Ω such that $\Lambda = G^\top \Omega G$ is the descending order diagonal matrix of the eigenvalues of Ω. Define $H = \left(\Gamma^\top \Gamma\right)^{-1/2} G$. Then, we have

$$\begin{aligned}
(\Gamma F^{d\top} F^d \Gamma^\top) \Gamma H &= \Gamma \left(\Gamma^\top \Gamma\right)^{-1/2} \left(\Gamma^\top \Gamma\right)^{1/2} F^{d\top} F^d \left(\Gamma^\top \Gamma\right)^{1/2} \left(\Gamma^\top \Gamma\right)^{1/2} H \\
&= \Gamma \left(\Gamma^\top \Gamma\right)^{-1/2} \left[\left(\Gamma^\top \Gamma\right)^{1/2} F^{d\top} F^d \left(\Gamma^\top \Gamma\right)^{1/2} G \right] \\
&= \Gamma \left(\Gamma^\top \Gamma\right)^{-1/2} T \Omega G = \Gamma \left(\Gamma^\top \Gamma\right)^{-1/2} G T \Lambda \\
&= \Gamma H T \Lambda.
\end{aligned}$$

In addition, note that $(\Gamma H)^\top (\Gamma H) = H^\top \Gamma^\top \Gamma H = G^\top G = I_K$. Therefore, ΓH is the eigenvector of $\Gamma F^{d\top} F^d \Gamma^\top$ and the left singular vector of $\Gamma F^{d\top}$. Let U be the left singular vector of $\Gamma F^{d\top}$. Then, we have $H_F = U_K^\top U H^{-1}$ since $\Gamma = U H^{-1}$. Since $\left\| \Omega - \left(\Gamma^\top \Gamma\right)^{1/2} \Sigma_f \left(\Gamma^\top \Gamma\right)^{1/2} \right\| = o_p(1)$ and the eigenvalues of $\left(\Gamma^\top \Gamma\right)^{1/2} \Sigma_f \left(\Gamma^\top \Gamma\right)^{1/2}$ are distinct, by the eigenvector perturbation theory, there is a unique eigenvector of $\left(\Gamma^\top \Gamma\right)^{1/2} \Sigma_f \left(\Gamma^\top \Gamma\right)^{1/2}$, says, $\boldsymbol{G}$, such that $\|G - \boldsymbol{G}\| = o_p(1)$. Therefore, $\left\| H - \left(\Gamma^\top \Gamma\right)^{-1/2} \boldsymbol{G} \right\| \to_p 0$. Moreover, because $\|G^{-1} - \boldsymbol{G}^{-1}\| = \left\| G^\top - \boldsymbol{G}^\top \right\| = o_p(1)$, we also have $\left\| H^{-1} - \boldsymbol{G}^{-1} \left(\Gamma^\top \Gamma\right)^{1/2} \right\| = o_p(1)$. By the same method in Claim E.1 of Choi and Yuan (2024), we know $U_K^\top U \to_p I_{sgn}$ where I_{sgn} is the $K \times K$ diagonal matrix consisting of the diagonal elements of ± 1 and the sign of these are determined by the sign alignment between U_K and U. Hence, we have $H_F \to_p I_{sgn} \boldsymbol{G}^{-1} \left(\Gamma^\top \Gamma\right)^{1/2}$ and $H_F^{-1} \to_p \left(\Gamma^\top \Gamma\right)^{-1/2} \boldsymbol{G} I_{sgn}$. □

Lemma I.11. *Let $a_{ijt} = (e_l^\top Q_t^{-1} x_{it}) Q_t^{-1} x_{jt}$ where $Q_t = X_t^\top X_t / N$ and $u_{ijt} = \epsilon_{it} \epsilon_{jt}$. Then, we have (i) $\frac{1}{TN^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T a_{ijt} (u_{ij,t+1} - \mathbb{E}[u_{ij,t+1}]) = O_p\left(\frac{\sqrt{L}}{N\sqrt{T}} \right)$;*
(ii) $\frac{1}{TN^2} \sum_{i=1}^N \sum_{t=1}^T a_{iit} \left(\mathbb{E}[\epsilon^2_{i,t+1}] - \hat{\sigma}^2_{t+1} \right) = o_p\left(\frac{1}{\sqrt{NT}} \right)$.

Proof of Lemma I.11. (i) Let A be the $N^2 T \times L$ matrix whose (i,j,t)-th row is $a_{ijt}^\top$ and U be the $N^2 T \times 1$ vector whose (i,j,t)-th element is $u_{ijt} - \mathbb{E}[u_{ijt}]$. Then, we have

$$\frac{1}{TN^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T a_{ijt} (u_{ijt} - \mathbb{E}[u_{ijt}]) = \frac{1}{TN^2} A^\top U$$

and

$$\mathbb{E}\left[\left(\frac{1}{TN^2}\left\|A^\top U\right\|\right)^2\middle| X\right] = \frac{1}{T^2N^4}\sum_{r=1}^{L} A_r^\top \mathbb{E}\left[UU^\top \middle| X\right] A_r \leq \frac{1}{T^2N^4}\|A\|_F^2 \left\|\mathbb{E}\left[UU^\top \middle| X\right]\right\|$$
$$= O_p\left(\frac{1}{N^2T}\right),$$

where $A_r = Ae_r$, because $\left\|\mathbb{E}\left[UU^\top \middle| X\right]\right\|$ is bounded and

$$\|A\|_F^2 = \sum_{r,i,j,t}(e_l^\top Q_t^{-1}x_{it})^2(e_r^\top Q_t^{-1}x_{jt})^2 = N^2\sum_t e_l^\top Q_t^{-1}e_l \sum_r e_r^\top Q_t^{-1}e_r \lesssim LN^2\sum_t \left|e_l^\top Q_t^{-1}e_l\right|$$
$$= O_p\left(LN^2T\right),$$

since $\|e_l\| = \|e_r\| = 1$ where $1 \leq l, r \leq L$. Note that, when ϵ_{it} is independent across i, $\mathrm{Cov}(\epsilon_{i,t+1}\epsilon_{j,t+1}, \epsilon_{i',s+1}\epsilon_{j',s+1})$ is nonzero only when $i \neq j$, $i = i'$, $j = j'$ or $i \neq j$, $i = j'$, $j = i'$ except for the case $i = j = i' = j'$. Hence, the condition $\max_{i,t}\sum_{s=1}^T |\mathrm{Cov}(\epsilon_{i,t+1}, \epsilon_{i,s+1})| \leq C_1$ and $\max_{i,t}\sum_{s=1}^T \left|\mathrm{Cov}(\epsilon_{i,t+1}^2, \epsilon_{i,s+1}^2)\right| \leq C_1$ for some constant $C_1 > 0$ is enough to have $\left\|\mathbb{E}\left[UU^\top \middle| X\right]\right\| < C_2$ for some constant $C_2 > 0$ because $\max_{ijt}\sum_{i',j',s} |\mathrm{Cov}(\epsilon_{i,t+1}\epsilon_{j,t+1}, \epsilon_{i',s+1}\epsilon_{j',s+1})|$ is bounded.

(ii) Let $m_{i,t+1} = \alpha_{O,it} + x_{it}^\top\eta + x_{it}^\top\Gamma\breve{f}_{t+1}$, $\tilde{m}_{i,t+1} = \hat{\alpha}_{O,it} + x_{it}^\top\tilde{\eta} + x_{it}^\top\tilde{\Gamma}\breve{\tilde{f}}_{t+1}$, and $\hat{\epsilon}_{i,t+1} = r_{i,t+1} - \tilde{m}_{i,t+1}$. Note that

$$\hat{\sigma}_{t+1}^2 - \mathbb{E}[\epsilon_{j,t+1}^2] = \frac{1}{N}\sum_{j=1}^N\left(\hat{\epsilon}_{j,t+1}^2 - \epsilon_{j,t+1}^2\right) + \frac{1}{N}\sum_{j=1}^N\left(\epsilon_{j,t+1}^2 - \mathbb{E}[\epsilon_{j,t+1}^2]\right).$$

First of all, by the concentration inequality with the weak dependent error condition, the part related to the second term is bounded as

$$\frac{1}{TN^2}\sum_{i=1}^N\sum_{t=1}^T a_{iit}\frac{1}{N}\sum_{j=1}^N\left(\epsilon_{j,t+1}^2 - \mathbb{E}[\epsilon_{j,t+1}^2]\right) = \frac{1}{TN^2}\sum_{j=1}^N\sum_{t=1}^T \bar{a}_t\left(\epsilon_{j,t+1}^2 - \mathbb{E}[\epsilon_{j,t+1}^2]\right) = O_p\left(\frac{1}{\sqrt{T}N^{3/2}}\right),$$

where $\bar{a}_t = \frac{1}{N}\sum_{i=1}^N a_{iit} = Q_t^{-1}e_l$. Here, we use the bounds $\max_t ||\bar{a}_t|| = O_p(1)$ and $\frac{1}{T}\sum_{t=1}^T ||\bar{a}_t||^2 = O_p(1)$. For the part related to the first term, note that

$$\frac{1}{N}\sum_{j=1}^N\left(\hat{\epsilon}_{j,t+1}^2 - \epsilon_{j,t+1}^2\right) = 2\frac{1}{N}\sum_{j=1}^N\epsilon_{j,t+1}\left(\tilde{m}_{j,t+1} - m_{j,t+1}\right) + \frac{1}{N}\sum_{j=1}^N\left(\tilde{m}_{j,t+1} - m_{j,t+1}\right)^2,$$

and

$$\frac{1}{N}\sum_{j=1}^N\epsilon_{j,t+1}\left(\tilde{m}_{j,t+1} - m_{j,t+1}\right)$$

$$= \frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}(\hat{\alpha}_{O,jt}-\alpha_{O,jt}) + \frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}x_{jt}^{\top}(\tilde{\eta}-\eta) + \frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}x_{jt}^{\top}(\tilde{\Gamma}\tilde{\tilde{f}}_{t+1}-\Gamma f_{t+1}).$$

Note that the dominating terms of $\hat{\alpha}_{O,jt}-\alpha_{O,jt}$ are

$$\frac{1}{NT}\sum_{k=1}^{N}\sum_{s=1}^{T}B_{t,j}^{o\top}B_{s,k}^{o}\epsilon_{k,s+1}, \quad \frac{1}{N}\sum_{k=1}^{N}\sum_{q\in D_t}B_{t,jq}^{o\top}B_{t,kq}^{o}\epsilon_{k,t+1}, \quad \sum_{q\notin D_t}B_{t,jq}^{o}\left(\tilde{\tilde{\xi}}_q-\bar{\xi}_q\right).$$

The part related to the first term is bounded as

$$\begin{aligned}
&\left\|\frac{1}{TN^2}\sum_{i=1}^{N}\sum_{t=1}^{T}a_{iit}\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t+1}[\hat{\alpha}_{O,jt}-\alpha_{O,jt}]_{part_1}\right\| \\
&\leq \frac{1}{TN^2}\left\|\sum_{t=1}^{T}\sum_{j=1}^{N}\epsilon_{j,t+1}\bar{a}_t B_{t,j}^{o\top}\right\|\left\|\frac{1}{NT}\sum_{k=1}^{N}\sum_{s=1}^{T}B_{s,k}^{o}\epsilon_{k,s+1}\right\| = O_P\left(\frac{1}{NT}\right).
\end{aligned}$$

Here we use the relation that

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{s=1}^{T}\sum_{k=1}^{N}B_{s,k}^{o}\epsilon_{k,s+1}\right\|^2\middle| X\right] &= \sum_{q=1}^{N-L}\mathbb{E}\left[\left(\sum_{s=1}^{T}\sum_{k=1}^{N}B_{s,kq}^{o}\epsilon_{k,s+1}\right)^2\middle| X\right] \\
&= \sum_{q=1}^{N-L}\mathcal{B}_q^{\top}\mathbb{E}\left[\mathrm{vec}(E)\mathrm{vec}(E)^{\top}\middle| X\right]\mathcal{B}_q \\
&\leq \sum_{q=1}^{N-L}\|\mathcal{B}_q\|^2\left\|\mathbb{E}\left[\mathrm{vec}(E)\mathrm{vec}(E)^{\top}\middle| X\right]\right\| \\
&= \sum_{s=1}^{T}\|B_s^{o}\|_F^2\left\|\mathbb{E}\left[\mathrm{vec}(E)\mathrm{vec}(E)^{\top}\middle| X\right]\right\| \\
&\lesssim N(N-L)T
\end{aligned}$$

since $\|B_s^{o}\|_F^2 = N(N-L)$ by Lemma I.14 and $\left\|\mathbb{E}\left[\mathrm{vec}(E)\mathrm{vec}(E)^{\top}\middle| X\right]\right\|$ is bounded. In addition, we use the relation that

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{t=1}^{T}\sum_{j=1}^{N}\epsilon_{j,t+1}\bar{a}_t B_{t,j}^{o\top}\right\|_F^2\middle| X\right] &= \sum_{r=1}^{L}\sum_{k=1}^{N-L}\mathbb{E}\left[\left(\sum_{t=1}^{T}\sum_{j=1}^{N}\epsilon_{j,t+1}\bar{a}_{t,r}B_{t,jk}^{o}\right)^2\middle| X\right] \\
&= \sum_{r=1}^{L}\sum_{k=1}^{N-L}A_{rk}^{\top}\mathbb{E}\left[EE^{\top}\middle| X\right]A_{rk} \\
&\leq \left\|\mathbb{E}\left[EE^{\top}\middle| X\right]\right\|\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{r=1}^{L}\sum_{k=1}^{N-L}\bar{a}_{t,r}^2B_{t,jk}^{o2}
\end{aligned}$$

$$= \left\|\mathbb{E}\left[EE^\top \middle| X\right]\right\| \sum_{t=1}^{T} \|\bar{a}_t\|^2 \|B_t^o\|_F^2 = O_p\left(TN(N-L)\right)$$

where A_{rk} is the $NT \times 1$ vector whose (i,t)-th element is $\bar{a}_{t,r} B^o_{t,ik}$, because $\left\|\mathbb{E}\left[EE^\top \middle| X\right]\right\|$ is bounded, $\sum_{t=1}^{T} \|\bar{a}_t\|^2 = O_p(T)$, and $\|B_t^o\|_F^2 = N(N-L)$. In addition, for the second part, we have

$$\begin{aligned}
&\left\| \frac{1}{TN^2} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{iit} \frac{1}{N} \sum_{j=1}^{N} \epsilon_{j,t+1} [\hat{\alpha}_{O,jt} - \alpha_{O,jt}]_{part_2} \right\| \\
&\leq \frac{1}{N} \left(\frac{1}{T} \sum_{t=1}^{T} \|\bar{a}_t\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} \epsilon_{j,t+1} \epsilon_{k,t+1} \left(\sum_{q \in D_t} B^o_{t,jq} B^o_{t,kq} \right) \right)^2 \right)^{1/2} \\
&= o_p\left(\frac{1}{\sqrt{NT}} \right)
\end{aligned}$$

since by Lemma I.15, w.h.p., for all t,

$$\begin{aligned}
\frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} \epsilon_{j,t+1} \epsilon_{k,t+1} \left(\sum_{q \in D_t} B^o_{t,jq} B_{t,kq} \right) &= \sum_{q \in D_t} \left(\frac{1}{N} \sum_{j=1}^{N} \epsilon_{j,t+1} B^o_{t,jq} \right) \left(\frac{1}{N} \sum_{k=1}^{N} \epsilon_{k,t+1} B^o_{t,kq} \right) \\
&\lesssim \frac{\log N}{N} |D_t| \ll \frac{\sqrt{N}}{\sqrt{T}}.
\end{aligned}$$

For the third term, the dominating parts of $[\hat{\alpha}_{O,jt} - \alpha_{O,jt}]_{part_3}$ are

$$\frac{1}{T} \sum_{s=1}^{T} \sum_{q \in D_s / D_t} B^o_{t,jq} \frac{1}{N} \sum_{k=1}^{N} B^o_{s,kq} \epsilon_{k,s+1}, \quad \frac{1}{T} \sum_{s=1}^{T} \sum_{q \in D_s / D_t} B^o_{t,jq} \frac{1}{T} \sum_{s'=1}^{T} \xi_{s',q}.$$

For the first part, we have

$$\begin{aligned}
&\frac{1}{NT} \sum_{t=1}^{T} \bar{a}_t \frac{1}{N} \sum_{j=1}^{N} \epsilon_{j,t+1} \frac{1}{T} \sum_{s=1}^{T} \sum_{q \in D_s / D_t} B^o_{t,jq} \frac{1}{N} \sum_{k=1}^{N} B^o_{s,kq} \epsilon_{k,s+1} \\
&\leq \max_t \|\bar{a}_t\| \frac{1}{N^3 T^2} \sum_{t=1}^{T} \left\| \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{s=1}^{T} \epsilon_{j,t+1} \epsilon_{k,s+1} \sum_{q \in D_s / D_t} B^o_{t,jq} B^o_{s,kq} \right\| \\
&= o_p\left(\frac{1}{\sqrt{NT}} \right)
\end{aligned}$$

since by Lemma I.15, w.h.p., for all t,

$$\begin{aligned}
\sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{s=1}^{T} \epsilon_{j,t+1} \epsilon_{k,s+1} \sum_{q \in D_s / D_t} B^o_{t,jq} B^o_{s,kq} &= \sum_{s=1}^{T} \sum_{q \in D_s / D_t} \left(\sum_{k=1}^{N} \epsilon_{k,s+1} B^o_{s,kq} \right) \left(\sum_{j=1}^{N} \epsilon_{j,t+1} B^o_{t,jq} \right) \\
&\lesssim T |\bar{D}| N \log N
\end{aligned}$$

where $|\bar{D}| = \frac{1}{T}\sum_{s=1}^{T} \bar{D}_s$. In addition, we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{t=1}^{T} \bar{a}_t \frac{1}{N}\sum_{j=1}^{N} \epsilon_{j,t+1} \frac{1}{T}\sum_{s=1}^{T} \sum_{q \in D_s/D_t} B^o_{t,jq} \frac{1}{T}\sum_{s'=1}^{T} \xi_{s',q} \\
&\leq \frac{1}{N} \max_t \|\bar{a}_t\| \max_{t,q} \left\| \frac{1}{N}\sum_{j=1}^{N} \epsilon_{j,t+1} B^o_{t,jq} \right\| \frac{1}{T^2} \sum_{t=1}^{T}\sum_{s=1}^{T} \sum_{q \in D_s/D_t} \|\bar{\xi}_q\| \\
&= o_p\left(\frac{1}{\sqrt{NT}}\right)
\end{aligned}$$

by Lemma I.15 and the sparsity condition. Moreover, the part related to $\tilde{\eta} - \eta$ is bounded as

$$\begin{aligned}
\left\| \frac{1}{TN^2}\sum_{i=1}^{N}\sum_{t=1}^{T} a_{iit} \frac{1}{N}\sum_{j=1}^{N} \epsilon_{j,t+1} x_{jt}^\top (\tilde{\eta} - \eta) \right\| &\leq \frac{1}{TN^2} \left\| \sum_{t=1}^{T}\sum_{j=1}^{N} \epsilon_{j,t+1} \bar{a}_t x_{jt}^\top \right\| \|\tilde{\eta} - \eta\| \\
&= O_P\left(\frac{1}{\sqrt{NT}} \frac{\sqrt{L}}{N} \left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N} \right) \right)
\end{aligned}$$

by Lemma I.6 and the fact that $\left\| \sum_{t=1}^{T}\sum_{j=1}^{N} \epsilon_{j,t+1} \bar{a}_t x_{jt} \right\| = O_p\left(\sqrt{NTL}\right)$. For the part related to $\tilde{\Gamma}\tilde{\breve{f}}_{t+1} - \Gamma \breve{f}_{t+1}$, note that

$$\begin{aligned}
&\left\| \frac{1}{TN^2}\sum_{i=1}^{N}\sum_{t=1}^{T} a_{iit} \frac{1}{N}\sum_{j=1}^{N} \epsilon_{j,t+1} x_{jt}^\top (\tilde{\Gamma}\tilde{\breve{f}}_{t+1} - \Gamma \breve{f}_{t+1}) \right\| \\
&\lesssim \left\| \frac{1}{TN^2}\sum_{t=1}^{T}\sum_{j=1}^{N} \bar{a}_t \epsilon_{j,t+1} x_{jt}^\top (\tilde{\Gamma} - \Gamma H_\Gamma) H_\Gamma^{-1} \breve{f}_{t+1} \right\| + \left\| \frac{1}{TN^2}\sum_{t=1}^{T}\sum_{j=1}^{N} \bar{a}_t \epsilon_{j,t+1} x_{jt}^\top \Gamma H_\Gamma (\tilde{\breve{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1}) \right\|.
\end{aligned}$$

The first term can be bounded as

$$\begin{aligned}
&\left\| \frac{1}{TN^2}\sum_{t=1}^{T}\sum_{j=1}^{N} \bar{a}_t \epsilon_{j,t+1} x_{jt}^\top (\tilde{\Gamma} - \Gamma H_\Gamma) H_\Gamma^{-1} \breve{f}_{t+1} \right\| \\
&\leq \frac{1}{N} \left(\frac{1}{T}\sum_{t=1}^{T} \left\| \frac{1}{N}\sum_{j=1}^{N} \epsilon_{j,t+1} x_{jt}^\top \right\|^2 \right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^{T} \|\bar{a}_t\|^2 \left\| \breve{f}_{t+1} \right\|^2 \right)^{1/2} \left\| \tilde{\Gamma} - \Gamma H_\Gamma \right\| \\
&= O_p\left(\frac{\sqrt{L}}{N^{3/2}} \left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N} \right) \right)
\end{aligned}$$

using the bound $\left\| \tilde{\Gamma} - \Gamma H_\Gamma \right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N} \right)$ from Lemma I.4. In addition,

the second term is bounded as

$$\begin{aligned}
&\left\| \frac{1}{TN^2} \sum_{t=1}^{T} \sum_{j=1}^{N} \bar{a}_t \epsilon_{j,t+1} x_{jt}^\top \Gamma H_\Gamma (\tilde{\breve{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1}) \right\| \\
&\leq \frac{1}{N} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{N} \sum_{j=1}^{N} \epsilon_{j,t+1} \bar{a}_t B_{jt}^\top \right\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{\breve{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1} \right\|^2 \right)^{1/2} \\
&= O_p \left(\frac{1}{N^{3/2}} \left(\frac{1}{\sqrt{N}} + \frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N} \right)^{10+\frac{1}{2}} \right) \right),
\end{aligned}$$

by Lemma I.5. Next, we bound the parts related to the following term:

$$\begin{aligned}
&\frac{1}{N} \sum_{j=1}^{N} (\tilde{m}_{j,t+1} - m_{j,t+1})^2 \\
&\lesssim \frac{1}{N} \sum_{j=1}^{N} \left[(\hat{\alpha}_{O,jt} - \alpha_{O,jt})^2 + (x_{jt}^\top (\tilde{\eta} - \eta))^2 + (x_{jt}^\top (\tilde{\Gamma} - \Gamma H_\Gamma) H_\Gamma^{-1} \breve{f}_{t+1})^2 + (x_{jt}^\top \Gamma H_\Gamma (\tilde{\breve{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1})^2 \right].
\end{aligned}$$

The first term is bounded as

$$\frac{1}{TN^2} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{iit} \frac{1}{N} \sum_{j=1}^{N} (\hat{\alpha}_{O,jt} - \alpha_{O,jt})^2 = \frac{1}{TN^2} \sum_{t=1}^{T} \sum_{j=1}^{N} \bar{a}_t (\hat{\alpha}_{O,jt} - \alpha_{O,jt})^2 = o_p \left(\frac{1}{\sqrt{NT}} \right),$$

by using the bound from the independent sub-Gaussian assumption like Lemma I.15 as above. The second term is bounded like

$$\begin{aligned}
&\frac{1}{TN^2} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{iit} \frac{1}{N} \sum_{j=1}^{N} (x_{jt}^\top (\tilde{\eta} - \eta))^2 = \frac{1}{TN} \sum_{t=1}^{T} \bar{a}_t (\tilde{\eta} - \eta)^\top Q_t (\tilde{\eta} - \eta) \\
&= O_p \left(\frac{1}{N} \left(\frac{L}{NT} + \left(\frac{L}{N} \right)^{20+1} + \frac{L}{N^2} \right) \right).
\end{aligned}$$

by Lemma I.6. The third term is bounded like

$$\begin{aligned}
&\frac{1}{TN^2} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{iit} \frac{1}{N} \sum_{j=1}^{N} (x_{jt}^\top (\tilde{\Gamma} - \Gamma H_\Gamma) H_\Gamma^{-1} \breve{f}_{t+1})^2 \\
&= \frac{1}{TN} \sum_{t=1}^{T} \bar{a}_t \breve{f}_{t+1}^\top H_\Gamma^{-\top} (\tilde{\Gamma} - \Gamma H_\Gamma)^\top Q_t (\tilde{\Gamma} - \Gamma H_\Gamma) H_\Gamma^{-1} \breve{f}_{t+1} \\
&= O_p \left(\frac{1}{N} \left(\frac{L}{NT} + \left(\frac{L}{N} \right)^{20+1} + \frac{L}{N^2} \right) \right).
\end{aligned}$$

using the bound $\left\| \tilde{\Gamma} - \Gamma H_\Gamma \right\| = O_p \left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N} \right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N} \right)$ from Lemma I.4. Lastly, the

fourth term is bounded like

$$\begin{aligned}
&\left\| \frac{1}{TN^2} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{iit} \frac{1}{N} \sum_{j=1}^{N} (x_{jt}^\top \Gamma H_\Gamma (\tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1})^2 \right\| \\
&= \left\| \frac{1}{TN} \sum_{t=1}^{T} \bar{a}_t (\tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1})^\top H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma (\tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1}) \right\| \\
&\lesssim \frac{1}{TN} \max_t \|\bar{a}_t\| \sum_{t=1}^{T} \left\| \tilde{\tilde{f}}_{t+1} - H_\Gamma^{-1} \breve{f}_{t+1} \right\|^2 = O_p \left(\frac{1}{N} \left(\frac{L}{NT} + \left(\frac{L}{N} \right)^{20+1} + \frac{L}{N^2} \right) \right),
\end{aligned}$$

since a simple calculation with Lemma I.5. Here, we use the fact that $\max_t ||\bar{a}_t|| = O_p(1)$ To sum up, we have

$$\frac{1}{TN^2} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{iit} \left(\mathbb{E}[\epsilon_{i,t+1}^2] - \hat{\sigma}_{t+1}^2 \right) = o_p \left(\frac{1}{\sqrt{NT}} \right). \quad \square$$

Lemma I.12. *(i)* $\left\| x_{it}^\top \Gamma H_\Gamma - x_{it}^\top \tilde{\Gamma} \right\| = O_p \left(\frac{L^{3/2}}{N^{3/2}\sqrt{T}} + \frac{\sqrt{L}}{N} + \left(\frac{L}{N} \right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{\sqrt{NT}} \right)$; *(ii)* $\left\| \left(H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma \right)^{-1} - \left(\tilde{\Gamma}^\top Q_t \tilde{\Gamma} \right)^{-1} \right\| = O_p \left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N} \right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N} \right)$; *(iii)* $\left\| X_t^\top \Gamma H_\Gamma - X_t^\top \tilde{\Gamma} \right\| = O_p \left(\sqrt{N} \left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N} \right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N} \right) \right)$; *(iv)* $\left\| x_{it}^\top \Gamma H_\Gamma - x_{it}^\top \hat{\Gamma} \right\| = O_p \left(\frac{\sqrt{L}}{\sqrt{NT}} \right)$; *(v)* $\left\| \Gamma H_\Gamma - \hat{\Gamma} \right\| = O_p \left(\frac{\sqrt{L}}{\sqrt{NT}} \right)$; *(vi)* $\left\| \left(H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma \right)^{-1} - \left(\hat{\Gamma}^\top Q_t \hat{\Gamma} \right)^{-1} \right\| = O_p \left(\frac{\sqrt{L}}{\sqrt{NT}} \right)$.

Proof of Lemma I.12. (i) Note that

$$\begin{aligned}
\tilde{\Gamma}^\top x_{it} - H_\Gamma^\top \Gamma^\top x_{it} &= \left(x_{it}^\top \otimes I_K \right) \operatorname{vec} \left(\tilde{\Gamma}^\top - H_\Gamma^\top \Gamma^\top \right) \\
&= \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top} \right)^{-1} H_F \frac{1}{NT} \sum_{js} x_{it}^\top Q_s^{-1} x_{js} f_{s+1}^d \epsilon_{j,s+1} \\
&+ \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{f}_{t+1}^d \tilde{f}_{t+1}^{d\top} \right)^{-1} \frac{1}{NT} \sum_{s} x_{it}^\top Q_s^{-1} X_s^\top E_{s+1} \left(\tilde{f}_{s+1}^d - H_F f_{s+1}^d \right).
\end{aligned}$$

By the weak dependence assumption, we have $\frac{1}{NT} \sum_{js} x_{it}^\top Q_s^{-1} x_{js} f_{s+1}^d \epsilon_{j,s+1} = O_p \left(\frac{\sqrt{L}}{\sqrt{NT}} \right)$. In addition, we have

$$\begin{aligned}
&\frac{1}{NT} \sum_{s} x_{it}^\top Q_s^{-1} X_s^\top E_{s+1} \left(\tilde{f}_{s+1}^d - H_F f_{s+1}^d \right) \\
&\leq \left(\frac{1}{T} \sum_{s} \left\| \frac{1}{N} \sum_{j} x_{it}^\top Q_s^{-1} x_{js} \epsilon_{j,s+1} \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{s} \left\| \tilde{f}_{s+1}^d - H_F f_{s+1}^d \right\|^2 \right)^{\frac{1}{2}}
\end{aligned}$$

$$= O_p\left(\frac{\sqrt{L}}{\sqrt{N}}\left(\frac{1}{\sqrt{N}} + \frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10}\right)\right),$$

since $\frac{1}{T}\sum_s \left\|\tilde{f}^d_{s+1} - H_F f^d_{s+1}\right\|^2 = O_p\left(\left(\frac{1}{\sqrt{N}} + \frac{L}{N\sqrt{T}} + \left(\frac{L}{N}\right)^{10}\right)^2\right)$ by Lemma I.4 and $\frac{1}{T}\sum_s \mathbb{E}[||\frac{1}{N}\sum_j x_{it}^\top Q_s^{-1} x_{js}\epsilon_{j,s+1}||^2|X] = O_p\left(\frac{L}{N}\right)$.

(ii) First, note that

$$\left\|H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma - \tilde{\Gamma}^\top Q_t \tilde{\Gamma}\right\| \lesssim \left\|\tilde{\Gamma} - \Gamma H_\Gamma\right\| \|Q_t\| \|\Gamma H_\Gamma\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N}\right) = o_p(1)$$

by using the bound $\left\|\tilde{\Gamma} - \Gamma H_\Gamma\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N}\right)$ from Lemma I.4. Then, since $\psi_{\min}\left(H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma\right) > c$ for some constant $c > 0$ we have $\left\|\left(\tilde{\Gamma}^\top Q_t \tilde{\Gamma}\right)^{-1}\right\| = O_p(1)$. Then, we have

$$\begin{aligned}\left\|\left(H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma\right)^{-1} - \left(\tilde{\Gamma}^\top Q_t \tilde{\Gamma}\right)^{-1}\right\| &\leq \left\|\left(H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma\right)^{-1}\right\| \left\|\left(\tilde{\Gamma}^\top Q_t \tilde{\Gamma}\right)^{-1}\right\| \left\|H_\Gamma^\top \Gamma^\top Q_t \Gamma H_\Gamma - \tilde{\Gamma}^\top Q_t \tilde{\Gamma}\right\| \\ &= O_p\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N}\right).\end{aligned}$$

(iii) It follows from

$$\left\|X_t\tilde{\Gamma} - X_t\Gamma H_\Gamma\right\| \leq \|X_t\| \left\|\tilde{\Gamma} - \Gamma H_\Gamma\right\| = O_p\left(\sqrt{N}\left(\frac{\sqrt{L}}{\sqrt{NT}} + \left(\frac{L}{N}\right)^{10+\frac{1}{2}} + \frac{\sqrt{L}}{N}\right)\right).$$

(iv) As noted in the proof of Theorem 1 (b), we need to bound the following term:

$$\begin{aligned}&\tilde{\Gamma}^\top \frac{1}{T}\sum_{s=1}^T (x_{it}^\top \ddot{E}_{t+1}) \times \ddot{E}_{t+1} - \tilde{\Gamma}^\top \frac{1}{T}\sum_{t=1}^T \hat{\sigma}^2_{t+1}(X_t^\top X_t)^{-1} x_{it} \\ &= \tilde{\Gamma}^\top \frac{1}{N^2 T}\sum_{j,j',t}(x_{it}^\top Q_t^{-1} x_{jt}) Q_t^{-1} x_{j't}\epsilon_{j,t+1}\epsilon_{j',t+1} - \tilde{\Gamma}^\top \frac{1}{N^2 T}\sum_{j,t}(x_{it}^\top Q_t^{-1} x_{jt}) Q_t^{-1} x_{jt}\hat{\sigma}^2_{t+1} \\ &= \tilde{\Gamma}^\top \frac{1}{N^2 T}\sum_{j,j',t} a_{jj't}(u_{jj',t+1} - \mathbb{E}[u_{jj',t+1}]) + \tilde{\Gamma}^\top \frac{1}{N^2 T}\sum_{j,t} a_{jjt}(\mathbb{E}[\epsilon^2_{j,t+1}] - \hat{\sigma}^2_{t+1}),\end{aligned}$$

where $a_{jj't} = (x_{it}^\top Q_t^{-1} x_{jt}) Q_t^{-1} x_{j't}$. Using a similar method as in the proof of Lemma I.11, we can show that it is $o_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$. In addition, we have

$$\frac{1}{T}\sum_{t=1}^T x_{it}^\top \ddot{E}_{t+1} \otimes f^d_{t+1} = \frac{1}{NT}\sum_{j=1}^N \sum_{t=1}^T (x_{it}^\top Q_t^{-1} x_{jt}) f^d_{t+1}\epsilon_{j,t+1} = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right).$$

Hence, we have $\left\|x_{it}^\top \Gamma H_\Gamma - x_{it}^\top \hat{\Gamma}\right\| = O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$.

(v) By the similar method as in the proof of Theorem 1 (b), we need to bound the following term:

$$\begin{aligned}&\frac{1}{T}\sum_{s=1}^{T}\ddot{E}_{t+1}\otimes\ddot{E}_{t+1}-\frac{1}{TN^2}\sum_{i=1}^{N}\sum_{t=1}^{T}a_{iit}\hat{\sigma}_{t+1}^2\\&=\frac{1}{N^2T}\sum_{i,j,t}a_{ijt}(u_{ij,t+1}-\mathbb{E}[u_{ij,t+1}])+\frac{1}{N^2T}\sum_{i,t}a_{iit}(\mathbb{E}[\epsilon_{i,t+1}^2]-\hat{\sigma}_{t+1}^2),\end{aligned}$$

where $a_{ijt}=(Q_t^{-1}\otimes Q_t^{-1})(x_{it}\otimes x_{jt})$. Here, we have

$$\max_t\|\bar{a}_t\|=\max_t\left\|\mathrm{vec}(Q_t^{-1})\right\|=\max_t\left\|Q_t^{-1}\right\|_F\leq\sqrt{L}\max_t\left\|Q_t^{-1}\right\|=O_p\left(\sqrt{L}\right)$$

since $\bar{a}_t=\frac{1}{N}\sum_i a_{iit}=\mathrm{vec}(Q_t^{-1}\frac{1}{N}\sum_i x_{it}x_{it}^\top Q_t^{-1})=\mathrm{vec}(Q_t^{-1})$. Then, using a similar method as in the proof of Lemma I.11, we can show that it is $o_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$. In addition, we have

$$\frac{1}{T}\sum_{t=1}^{T}\ddot{E}_{t+1}\otimes f_{t+1}^d=\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}(Q_t^{-1}x_{jt}\otimes f_{t+1}^d)\epsilon_{j,t+1}=O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right).$$

Hence, we have $\left\|\Gamma H_\Gamma-\hat{\Gamma}\right\|=O_p\left(\frac{\sqrt{L}}{\sqrt{NT}}\right)$.
(vi) With the aid of (v), we can proof it in the same way as that of (ii). □

Lemma I.13. *For an $a\times b$ matrix Φ and its estimator $\hat{\Phi}$, we have the following decomposition:*

$$\begin{aligned}P_{\hat{\Phi}}-P_\Phi&=\hat{\Phi}\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}\hat{\Phi}^\top-\Phi\left(\Phi^\top\Phi\right)^{-1}\Phi^\top\\&=\Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi}-\Phi\right)^\top(I-P_\Phi)+(I-P_\Phi)\left(\hat{\Phi}-\Phi\right)\left(\Phi^\top\Phi\right)^{-1}\Phi^\top+\textit{higher order terms},\end{aligned}$$

where

$$\begin{aligned}\textit{higher order terms}=&(\hat{\Phi}-\Phi)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}-\left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top+(\hat{\Phi}-\Phi)\left(\Phi^\top\Phi\right)^{-1}(\hat{\Phi}-\Phi)^\top\\&+\Phi\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}-\left(\Phi^\top\Phi\right)^{-1}\right](\hat{\Phi}-\Phi)^\top+(\hat{\Phi}-\Phi)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}-\left(\Phi^\top\Phi\right)^{-1}\right](\hat{\Phi}-\Phi)^\top\\&-\Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi}-\Phi\right)^\top\Phi\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}-\left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top\\&-\Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi}-\Phi\right)^\top\left(\hat{\Phi}-\Phi\right)\left(\Phi^\top\Phi\right)^{-1}\Phi^\top\\&-\Phi\left(\Phi^\top\Phi\right)^{-1}\Phi^\top\left(\hat{\Phi}-\Phi\right)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}-\left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top\\&-\Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi}-\Phi\right)^\top\left(\hat{\Phi}-\Phi\right)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}-\left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top.\end{aligned}$$

Proof of Lemma I.13. First, a simple calculation shows that

$$\begin{aligned}
P_{\hat{\Phi}} - P_{\Phi} = {} & (\hat{\Phi} - \Phi)\left(\Phi^\top\Phi\right)^{-1}\Phi^\top + \Phi\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top + \Phi\left(\Phi^\top\Phi\right)^{-1}(\hat{\Phi} - \Phi)^\top \\
& + (\hat{\Phi} - \Phi)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top + (\hat{\Phi} - \Phi)\left(\Phi^\top\Phi\right)^{-1}(\hat{\Phi} - \Phi)^\top \\
& + \Phi\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right](\hat{\Phi} - \Phi)^\top + (\hat{\Phi} - \Phi)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right](\hat{\Phi} - \Phi)^\top.
\end{aligned}$$

In addition, a simple calculation shows that

$$\begin{aligned}
\Phi\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top &= \Phi\left(\Phi^\top\Phi\right)^{-1}\left[\Phi^\top\Phi - \hat{\Phi}^\top\hat{\Phi}\right]\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1}\Phi^\top \\
&= -\Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi} - \Phi\right)^\top P_{\Phi} - P_{\Phi}\left(\hat{\Phi} - \Phi\right)\left(\Phi^\top\Phi\right)^{-1}\Phi^\top \\
&\quad - \Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi} - \Phi\right)^\top\Phi\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top \\
&\quad - \Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi} - \Phi\right)^\top\left(\hat{\Phi} - \Phi\right)\left(\Phi^\top\Phi\right)^{-1}\Phi^\top \\
&\quad - \Phi\left(\Phi^\top\Phi\right)^{-1}\Phi^\top\left(\hat{\Phi} - \Phi\right)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top \\
&\quad - \Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi} - \Phi\right)^\top\left(\hat{\Phi} - \Phi\right)\left[\left(\hat{\Phi}^\top\hat{\Phi}\right)^{-1} - \left(\Phi^\top\Phi\right)^{-1}\right]\Phi^\top.
\end{aligned}$$

Therefore, we have

$$P_{\hat{\Phi}} - P_{\Phi} = \Phi\left(\Phi^\top\Phi\right)^{-1}\left(\hat{\Phi} - \Phi\right)^\top(I - P_{\Phi}) + (I - P_{\Phi})\left(\hat{\Phi} - \Phi\right)\left(\Phi^\top\Phi\right)^{-1}\Phi^\top + \text{higher order terms.}\ \square$$

Lemma I.14. *(i) $B_t^{o\top}B_t^o = N \cdot I_{N-L}$ and $\|B_t^o\| = \sqrt{N}$; (ii) For all $1 \le q \le N - L$, we have $\left\|B_{t,q}^o\right\|^2 = N$ where $B_{t,q}^o = B_t^o e_q$ and e_q is an $(N-L)\times 1$ unit vector. That is, $\sum_{i=1}^N (B_{t,iq}^o)^2 = N$ where $B_{t,iq}^o = e_i^\top B_t^o e_r$; (iii) $\|B_t^o\|_F^2 = \sum_{i=1}^N \left\|B_{t,i}^o\right\|^2 = N(N-L)$ where $B_{t,i}^o = B_t^{o\top}e_i$ and e_i is an $N\times 1$ unit vector.*

Proof of Lemma I.14. (i) $B_t^{o\top}B_t^o = (X_t^{o\top}X_t^o/N)^{-1/2}X_t^{o\top}X_t^o(X_t^{o\top}X_t^o/N)^{-1/2} = N \cdot I_{N-L}$. (ii) $\left\|B_{t,q}^o\right\|^2 = e_q^\top B_t^{o\top}B_t^o e_q = N \cdot e_q^\top e_q = N$. (iii) $\|B_t^o\|_F^2 = \sum_{q=1}^{N-L}\left\|B_{t,q}^o\right\|^2 = N(N-L)$ by (i). $\square$

Lemma I.15. *We have w.p.c. to 1, for all t,*

$$\max_{1\le q\le N-L}\left|\frac{1}{N}\sum_{j=1}^N B_{t,jq}^o\epsilon_{j,t+1}\right| \le C_u\sigma_{t+1}\frac{\sqrt{\log NT}}{\sqrt{N}}$$

for some constant $C_u > 0$. In addition, if T/N^a is bounded for some $a \ge 1$, then we can have the same result with $\sqrt{\log N}$ in place of $\sqrt{\log NT}$.

Proof of Lemma I.15. For each $1 \leq q \leq N-L$, we have $\sum_{j=1}^{N} \left(B^o_{t,jq}\right)^2 = \left\|B^o_{t,q}\right\|^2 = N$ by Lemma I.14. Hence, by Hoeffding's inequality (e.g., Theorem 2.6.3 of Vershynin (2018)), for each q and t, we have with probability exceeding $1 - O((NT)^{-9})$ that

$$\frac{1}{N}\sum_{j=1}^{N} B^o_{t,jq}\epsilon_{j,t+1} \leq C_u \sigma_{t+1}\frac{\sqrt{\log NT}}{\sqrt{N}}$$

for some universal constants $C_u > 0$. Hence, we have w.p.c. to 1, for all t,

$$\max_{1\leq q\leq N-L}\left|\frac{1}{N}\sum_{j=1}^{N} B^o_{t,jq}\epsilon_{j,t+1}\right| \leq C_u \sigma_{t+1}\frac{\sqrt{\log NT}}{\sqrt{N}}. \quad \square$$

Lemma I.16. *We have w.p.c. to 1,*

$$\max_{1\leq q\leq N-L}\left|\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right| \ll \sigma_{t+1}\frac{\sqrt{\log NT}}{\sqrt{N}}$$

for all $1 \leq t \leq T$. In addition, if T/N^a is bounded for some $a \geq 1$, then we can have the same result with $\sqrt{\log N}$ in place of $\sqrt{\log NT}$.

Proof of Lemma I.16. We want to show that w.p.c. to 1,

$$\max_{1\leq q\leq N-L}\left|\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right| \leq \sigma_{\min}\frac{(\log NT)^{1/4}}{\sqrt{N}}$$

where $\sigma_{\min}$ is some constant such that $\sigma_{\min} \leq \sigma_t$ for all t. Then, Lemma I.16 is followed from it. By the Markov's inequality, we have

$$\Pr\left(\left|\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right| \geq \sigma_{\min}\frac{(\log NT)^{1/4}}{\sqrt{N}}\right) \leq \frac{\mathbb{E}\left[\left|\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right|^{\alpha}\right]}{\left(\sigma_{\min}\frac{(\log NT)^{1/4}}{\sqrt{N}}\right)^{\alpha}}$$

for some integer $\alpha \geq 1$. Then, because

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right| \geq \sigma_{\min}\frac{(\log NT)^{1/4}}{\sqrt{N}} \text{ at least one } q\right) \\
&\leq \Pr\left(\bigcup_{1\leq q\leq N-L}\left\{\left|\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right| \geq \sigma_{\min}\frac{(\log NT)^{1/4}}{\sqrt{N}}\right\}\right) \\
&\leq (N-L) \times \frac{\frac{1}{(NT)^{\alpha/2}}\mathbb{E}\left[\left|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{j=1}^{N} B^o_{s,jq}\epsilon_{j,s+1}\right|^{\alpha}\right]}{\left(\sigma_{\min}\frac{(\log NT)^{1/4}}{\sqrt{N}}\right)^{\alpha}} \longrightarrow 0
\end{aligned}$$

under our assumptions, we can say that w.p.c. to 1,

$$\max_{1\leq q\leq N-L}\left|\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N}B^{o}_{s,jq}\epsilon_{j,s+1}\right|\leq\sigma_{\min}\frac{(\log NT)^{1/4}}{\sqrt{N}}.\quad\square$$

References

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

Bai, J., Choi, S. H., and Liao, Y. (2020). Standard errors for panel data models with unknown clusters. *Journal of Econometrics*, page 105004.

Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C., and Kato, K. (2018). High-dimensional econometrics and regularized gmm. *arXiv preprint arXiv:1806.01888*.

Chernozhuokov, V., Chetverikov, D., Kato, K., and Koike, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *The Annals of Statistics*, 50(5):2562–2586.

Choi, J. and Yuan, M. (2024). High dimensional factor analysis with weak factors. *arXiv preprint arXiv:2402.05789*.

Hansen, C. B. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics*, 141(2):597–620.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, pages 703–708.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

Zhang, C. (2024). Testing pricing errors of models with latent factors and firm characteristics as covariances. *Management Science*, 70(3):1706–1728.