Appendix for "Robust Factor Models with Explanatory Proxies"

Jianqing Fan* Yuan Ke† Yuan Liao‡

March 21, 2016

Abstract

This document contains an empirical study of testing proxy factors for S&P 500 returns, additional numerical results, and all the technical proofs of the main paper.

Contents

*Department of Operations Research and Financial Engineering, Bendheim Center for Finance, Princeton University

†Department of Operations Research and Financial Engineering, Princeton University

‡University of Maryland

A Testing proxy factors for S&P 500 returns

In this section, we use the *robust proxy-regressed* method and the test statistic proposed in Section 4 to test the explanatory power of the observable proxies for the true factors using S&P 500 returns. For each given group of observable proxies, we set the number of common factors K equals the number of observable proxies. The sieve basis is chosen as the additive Fourier basis with $J = 5$. We set the tuning parameter $\alpha_T = C\sqrt{\frac{T}{\log(NJ)}}$ with constant C selected by the 5-fold cross validation.

We calculate the daily excess returns for the stocks in S&P 500 index that have complete daily closing prices from January 2005 to December 2013. The data, collected from CRSP, contains 393 stocks with a time span of 2265 trading days. As we know, two stylized features of S&P 500 daily returns are asymmetry and heavy tails. The proxy factors ($\mathbf{w}_t$) are chosen to be the Fama-French 3/5 factors and the sector SPDR ETF's, which are intended to track the 9 largest S&P sectors (Fan et al. , 2015). The detailed descriptions of sector SPDR

ETF's are listed in Table A.1. In this study, we consider three groups of proxy factors with increasing information: (1) Fama-French 3 factors (FF3); (2) Fama-French 5 factors (FF5); and (3) Fama-French 5 factors plus 9 sector SPDR ETF's (FF5+ETF9).

We apply moving window tests with the window size (T) equals one month, three months or six months. The testing window moves one trading day forward per test. Within each testing window, we calculate the standardized test statistic S for three groups of proxy factors. The plots of S under various scenarios are reported in Figure A.1.

According to Figure A.1, under all scenarios, the null hypothesis ($H_0 : \text{cov}(\boldsymbol{\gamma}_t) = 0$) is rejected as S is always larger than the critical value 1.96. This suggests a strong evidence that the proxy factors can not fully explain the estimated common factors. Under all window sizes, a larger group of proxy factors tends to yield smaller statistics, demonstrating stronger explanatory power for estimated common factors. Also, we find the test statistics increase while the window size increases.

Table A.1: Sector SPDR ETF's (data available from Yahoo finance)

Code	Representative sector
XLE	Energy
XLB	Materials
XLI	Industrials
XLY	Consumer discretionary
XLP	Consumer staples
XLV	Health care
XLF	Financial
XLK	Information technology
XLU	Utilities

Moreover, we also used the monthly excess returns for the stocks in S&P 500 index that have complete record from January 1980 to December 2012, which contains 202 stocks with a time span of 396 months. Here we only consider the first two groups of proxy factors as sector SPDR ETF's are introduced since 1998. The window size equals sixty months and moves one month forward per test. Within each testing window, we also estimate the volatility of $\boldsymbol{\gamma}_t$, the part of factors that can not be explained by $\mathbf{w}_t$ as:

$$\widehat{Vol}(\boldsymbol{\gamma}_t) = \frac{1}{21T} \sum_{t=1}^{T} \widehat{\boldsymbol{\gamma}}_t' \widehat{\boldsymbol{\gamma}}_t,$$

where there are 21 trading days per month.

The results are reported in Figure A.2. For both Fama-French 3 factors and 5 factors, the null hypothesis is rejected most of the time except in early 1980s and 1990s. When the null hypothesis is accepted, Fama-French 5 factors tend to yield larger p-values. The estimated volatility of unexplained part are close to zero over these two periods. For the rest of the time, the standardized test statistics are much larger than the critical value 1.96 and hence the p-values are close to zero. Also the estimated volatilities are not close to zero. This indicates the proxy factors can not fully explain the estimated common factors during these testing periods.

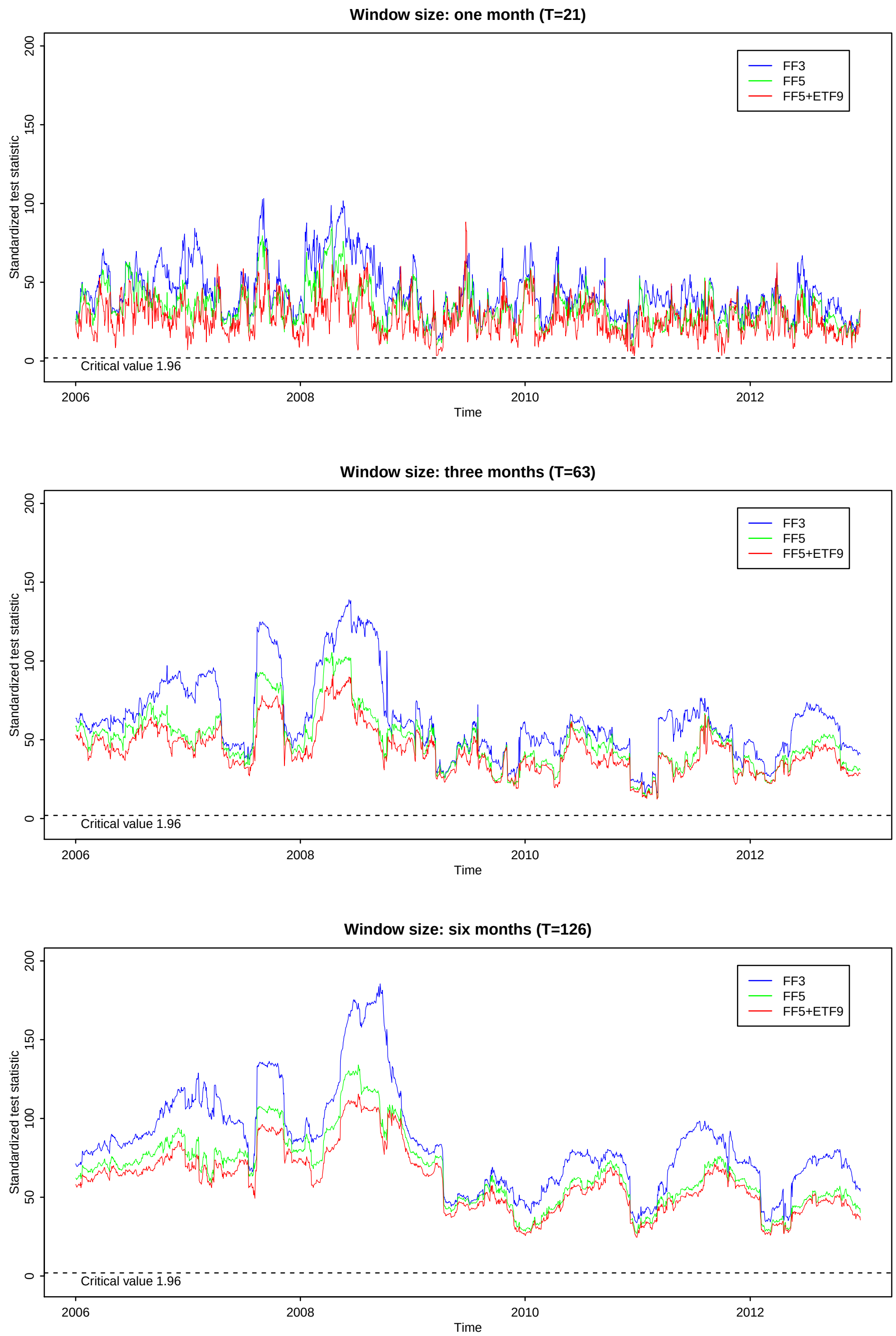

Figure A.1: S&P 500 daily returns: plots for standardized test statistic S for various window sizes. The dotted line is critical value 1.96.

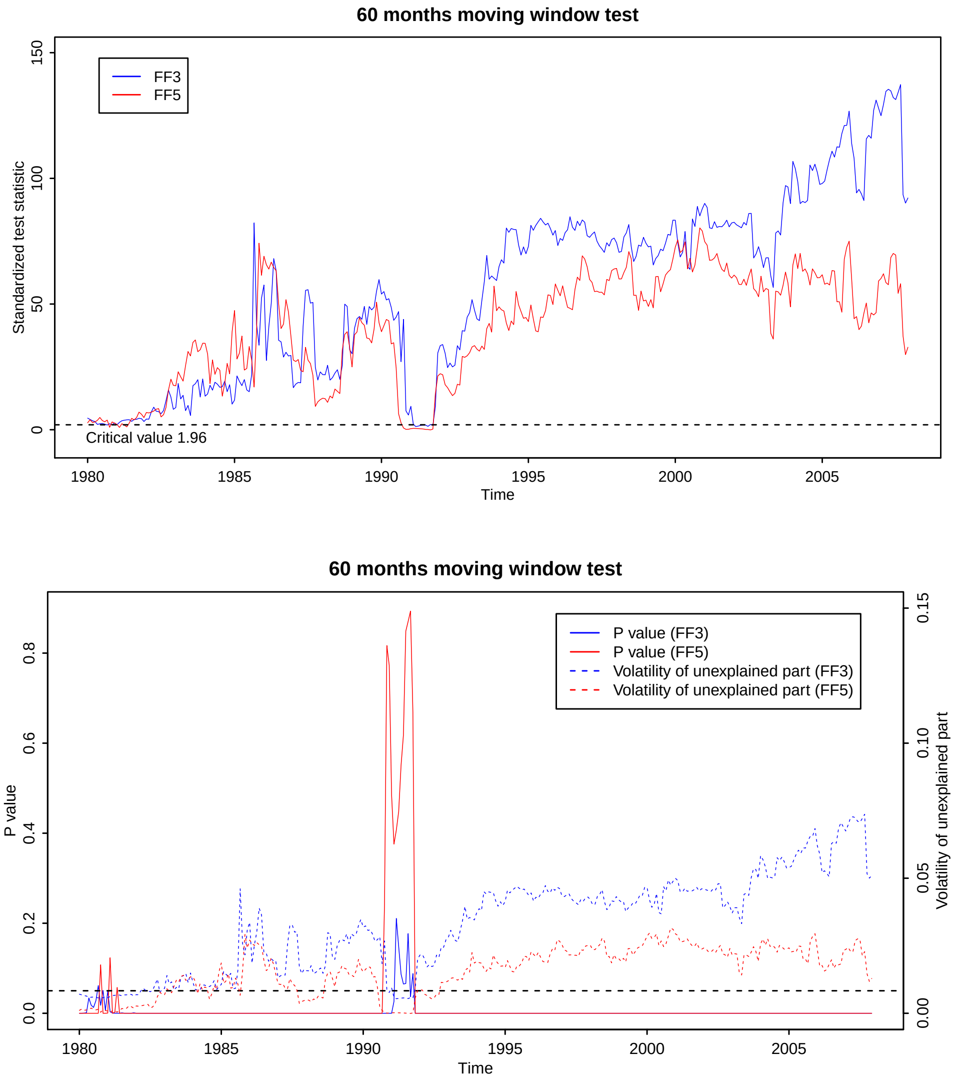

Figure A.2: S&P 500 monthly returns: plots for standardized test statistic S, P-value and the volatility of the part of factors that can not be explained by the proxy factors.

B Additional numerical results

In this section, we present some addition numerical results for Sections 6 and 7. Tables B.1–B.4 provide simulation results under additional sample sizes ($N = 100$, $T = 50$ and $N = 100$, $T = 100$), corresponding to the Tables 1–4 in Section 6 respectively. Figures B.1 and B.2 plots the forecast results of bond risk premia with maturity of 2 to 5 years. The factors are estimated by either PCA or our method. The forecast model is either linear model or multi-index model. See Section 7.3 for more details. Table B.5 and B.6 are the full versions of Table 7 and 8 and provide full forecast results for factor-augmented linear and multi-index models discussed in Section 7.4.

Table B.1: More results for mean relative estimation error of $\mathbf{\Lambda F'}$ (%): the smaller the better (with PCA as the benchmark)

Model (I)							
		$N = 100, T = 50$			$N = 100, T = 100$		
$\mathbf{u}_t$	σ	RPR	Sieve-LS	INT	RPR	Sieve-LS	INT
$N(0,\ 8)$	0.01	0.80	0.80	0.97	0.75	0.74	0.85
	0.3	0.88	0.88	0.99	0.89	0.88	0.95
	1.0	1.64	1.63	2.28	1.93	1.92	2.31
MixN	0.01	0.82	0.81	0.91	0.78	0.77	0.87
	0.3	0.95	0.94	0.98	0.99	0.98	1.02
	1.0	1.75	1.74	1.94	2.11	2.10	2.39
$2t_3$	0.01	0.58	0.87	0.97	0.73	0.83	0.94
	0.3	0.59	0.88	0.98	0.74	0.84	0.97
	1.0	0.60	0.89	0.98	0.75	0.86	0.99
LogN	0.01	0.63	0.84	0.97	0.86	0.78	0.93
	0.3	0.64	0.85	0.97	0.87	0.78	0.95
	1.0	0.65	0.86	0.99	0.89	0.78	1.00
Model (II)							
		$N = 100, T = 50$			$N = 100, T = 100$		
$\mathbf{u}_t$	σ	RPR	Sieve-LS	INT	RPR	Sieve-LS	INT
$N(0,\ 8)$	0.01	0.83	0.83	0.98	0.84	0.84	0.92
	0.3	0.92	0.91	1.03	0.93	0.93	0.95
	1.0	1.50	1.49	1.75	1.62	1.61	1.84
MixN	0.01	0.89	0.89	1.03	0.90	0.90	0.97
	0.3	0.98	0.97	1.06	0.99	0.99	0.99
	1.0	1.61	1.60	1.82	1.70	1.69	1.78
$2t_3$	0.01	0.53	0.88	0.97	0.66	0.83	0.94
	0.3	0.54	0.89	0.98	0.67	0.83	0.98
	1.0	0.56	0.90	1.00	0.68	0.84	0.99
LogN	0.01	0.63	0.86	0.96	0.87	0.80	0.93
	0.3	0.63	0.87	0.97	0.88	0.81	0.96
	1.0	0.64	0.88	0.99	0.88	0.83	0.99
Model (III)							
		$N = 100, T = 50$			$N = 100, T = 100$		
$\mathbf{u}_t$	σ	RPR	Sieve-LS	INT	RPR	Sieve-LS	INT
$N(0,\ 8)$	1.0	2.06	2.05	2.61	3.01	2.99	3.05
MixN	1.0	2.37	2.35	2.89	3.50	3.46	3.58
$2t_3$	1.0	1.19	1.18	1.09	1.19	1.18	1.14
LogN	1.0	1.17	1.17	1.18	1.18	1.17	1.19

Table B.2: More results for median canonical correlations of loadings: the larger the better

Model (I)									
		$N = 100, T = 50$				$N = 100, T = 100$			
$\mathbf{u}_t$	σ	RPR	Sieve-LS	PCA	INT	RPR	Sieve-LS	PCA	INT
$N(0,\ 8)$	0.01	0.89	0.89	0.75	0.84	0.93	0.93	0.85	0.89
	0.3	0.80	0.80	0.81	0.78	0.89	0.89	0.89	0.87
	1.0	0.79	0.79	0.94	0.77	0.89	0.89	0.97	0.87
MixN	0.01	0.93	0.93	0.84	0.90	0.95	0.95	0.91	0.91
	0.3	0.87	0.87	0.88	0.85	0.93	0.93	0.93	0.90
	1.0	0.86	0.86	0.96	0.84	0.92	0.92	0.98	0.88
$2t_3$	0.01	0.55	0.36	0.24	0.33	0.69	0.45	0.34	0.41
	0.3	0.53	0.33	0.25	0.30	0.68	0.44	0.35	0.40
	1.0	0.51	0.32	0.27	0.28	0.65	0.42	0.38	0.39
LogN	0.01	0.65	0.32	0.23	0.29	0.74	0.46	0.35	0.40
	0.3	0.62	0.29	0.24	0.25	0.71	0.45	0.37	0.39
	1.0	0.60	0.27	0.27	0.26	0.66	0.43	0.39	0.39

Model (II)									
		$N = 100, T = 50$				$N = 100, T = 100$			
$\mathbf{u}_t$	σ	RPR	Sieve-LS	PCA	INT	RPR	Sieve-LS	PCA	INT
$N(0,\ 8)$	0.01	0.93	0.94	0.90	0.92	0.94	0.94	0.92	0.91
	0.3	0.91	0.93	0.92	0.89	0.93	0.93	0.93	0.90
	1.0	0.89	0.90	0.95	0.87	0.93	0.93	0.97	0.89
MixN	0.01	0.90	0.91	0.88	0.89	0.96	0.95	0.94	0.93
	0.3	0.88	0.90	0.90	0.86	0.95	0.95	0.96	0.93
	1.0	0.85	0.87	0.93	0.84	0.95	0.95	0.98	0.92
$2t_3$	0.01	0.54	0.35	0.25	0.31	0.64	0.39	0.28	0.34
	0.3	0.52	0.33	0.26	0.30	0.61	0.37	0.33	0.33
	1.0	0.50	0.32	0.29	0.29	0.58	0.34	0.36	0.33
LogN	0.01	0.64	0.33	0.24	0.29	0.73	0.43	0.33	0.40
	0.3	0.62	0.30	0.25	0.29	0.70	0.41	0.34	0.37
	1.0	0.59	0.28	0.28	0.28	0.67	0.39	0.37	0.37

Table B.3: More results for median canonical correlations of factors: the larger the better

Model (I)									
		$N = 100, T = 50$				$N = 100, T = 100$			
$\mathbf{u}_t$	σ	RPR	Sieve-LS	PCA	INT	RPR	Sieve-LS	PCA	INT
$N(0,\ 8)$	0.01	0.96	0.96	0.83	0.88	0.97	0.97	0.85	0.94
	0.3	0.91	0.91	0.88	0.87	0.91	0.91	0.89	0.90
	1.0	0.91	0.91	0.97	0.85	0.91	0.91	0.96	0.88
MixN	0.01	0.98	0.98	0.89	0.94	0.98	0.98	0.91	0.94
	0.3	0.94	0.94	0.93	0.90	0.94	0.94	0.94	0.92
	1.0	0.93	0.94	0.98	0.88	0.93	0.93	0.98	0.90
$2t_3$	0.01	0.63	0.38	0.25	0.34	0.71	0.47	0.30	0.40
	0.3	0.62	0.35	0.26	0.32	0.70	0.45	0.33	0.38
	1.0	0.57	0.31	0.29	0.30	0.67	0.43	0.36	0.36
LogN	0.01	0.63	0.40	0.28	0.36	0.70	0.45	0.32	0.40
	0.3	0.60	0.36	0.31	0.35	0.66	0.44	0.34	0.39
	1.0	0.59	0.33	0.33	0.33	0.62	0.41	0.37	0.37

Model (II)									
		$N = 100, T = 50$				$N = 100, T = 100$			
$\mathbf{u}_t$	σ	RPR	Sieve-LS	PCA	INT	RPR	Sieve-LS	PCA	INT
$N(0,\ 8)$	0.01	0.97	0.97	0.91	0.93	0.97	0.97	0.97	0.92
	0.3	0.94	0.94	0.92	0.90	0.94	0.94	0.93	0.91
	1.0	0.94	0.94	0.97	0.88	0.94	0.94	0.96	0.90
MixN	0.01	0.98	0.98	0.94	0.92	0.98	0.98	0.95	0.94
	0.3	0.96	0.96	0.95	0.89	0.96	0.96	0.96	0.92
	1.0	0.96	0.96	0.98	0.88	0.96	0.96	0.98	0.90
$2t_3$	0.01	0.62	0.37	0.25	0.33	0.70	0.44	0.28	0.40
	0.3	0.60	0.35	0.25	0.29	0.68	0.41	0.30	0.38
	1.0	0.55	0.30	0.26	0.27	0.62	0.37	0.33	0.33
LogN	0.01	0.60	0.36	0.25	0.30	0.68	0.41	0.29	0.35
	0.3	0.56	0.35	0.26	0.28	0.65	0.38	0.31	0.34
	1.0	0.53	0.33	0.29	0.28	0.60	0.35	0.33	0.33

Table B.4: More results for mean relative mean squared error of forecast: the smaller the better (with PCA as the benchmark)

Model (I)							
		$N=100, T=50$			$N=100, T=100$		
$\mathbf{u}_t$	σ	RPR	Sieve-LS	INT	RPR	Sieve-LS	INT
$N(0,\ 8)$	0.01	0.90	0.90	0.99	0.95	0.95	0.98
	0.3	0.95	0.94	1.03	0.97	0.97	1.01
	1.0	1.03	1.02	1.07	1.03	1.02	1.04
MixN	0.01	0.76	0.76	0.95	0.74	0.73	0.93
	0.3	0.78	0.77	0.98	0.77	0.77	0.97
	1.0	0.95	0.95	1.02	1.02	1.01	1.04
$2t_3$	0.01	0.22	0.29	0.51	0.45	0.56	0.61
	0.3	0.23	0.30	0.57	0.24	0.41	0.68
	1.0	0.24	0.35	0.70	0.28	0.42	0.72
LogN	0.01	0.34	0.43	0.65	0.36	0.48	0.77
	0.3	0.34	0.44	0.71	0.37	0.48	0.80
	1.0	0.36	0.45	0.76	0.39	0.50	0.88
Model (II)							
		$N=100, T=50$			$N=100, T=100$		
$\mathbf{u}_t$	σ	RPR	Sieve-LS	INT	RPR	Sieve-LS	INT
$N(0,\ 8)$	0.01	0.89	0.89	0.97	0.94	0.93	0.97
	0.3	0.93	0.93	1.02	0.96	0.95	1.01
	1.0	1.10	1.09	1.11	1.04	1.04	1.07
MixN	0.01	0.75	0.75	0.91	0.80	0.79	0.92
	0.3	0.84	0.84	0.95	0.84	0.84	0.99
	1.0	1.12	1.07	1.09	1.21	1.19	1.19
$2t_3$	0.01	0.45	0.66	0.70	0.46	0.67	0.74
	0.3	0.47	0.68	0.76	0.47	0.68	0.79
	1.0	0.49	0.71	0.85	0.48	0.72	0.91
LogN	0.01	0.61	0.72	0.77	0.62	0.75	0.80
	0.3	0.62	0.74	0.82	0.63	0.78	0.84
	1.0	0.65	0.77	0.87	0.65	0.80	0.90
Model (III)							
		$N=100, T=50$			$N=100, T=100$		
$\mathbf{u}_t$	σ	RPR	Sieve-LS	INT	RPR	Sieve-LS	INT
$N(0,\ 8)$	1.0	1.50	1.49	1.33	1.43	1.42	1.30
MixN	1.0	1.48	1.45	1.32	1.72	1.68	1.37
$2t_3$	1.0	1.29	1.27	1.10	1.28	1.28	1.14
LogN	1.0	1.20	1.20	1.12	1.23	1.21	1.06

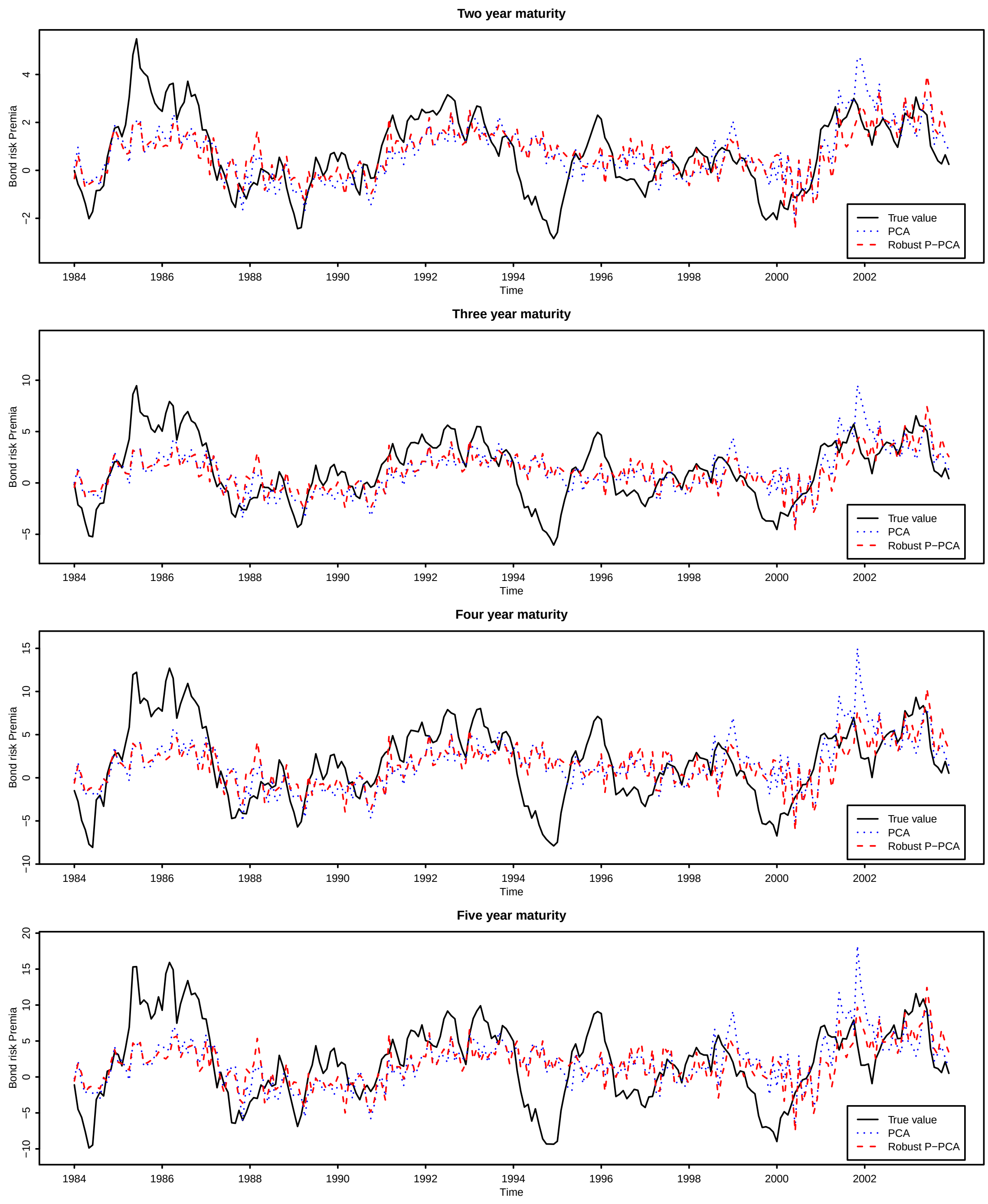

Figure B.1: Forecast with linear model

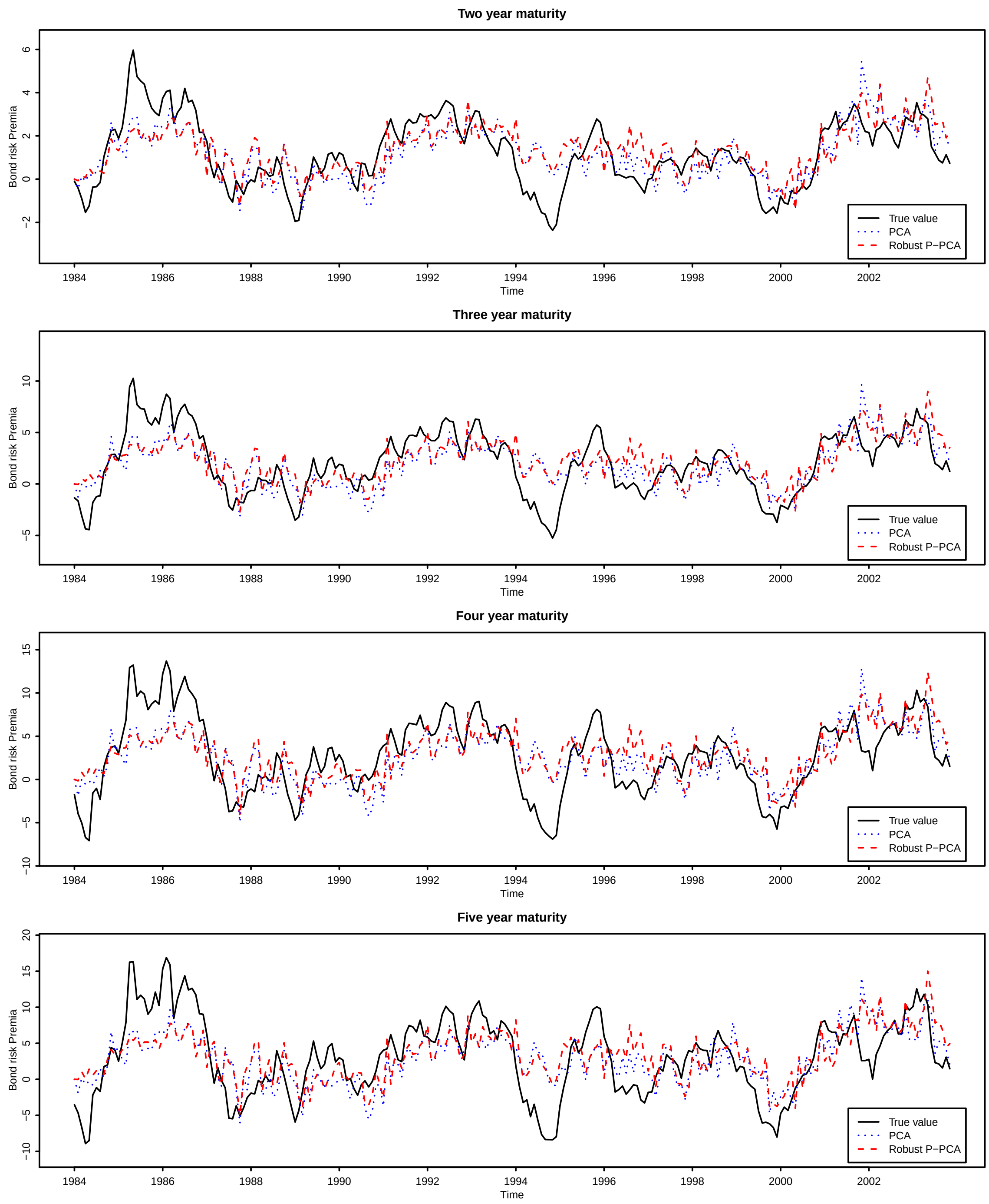

Figure B.2: Forecast with multi-index model

Table B.5: Full results for forecast out-of-sample R^2 (%) for factor-augmented linear model: the larger the better. The bold figures represent larger R^2 than forecast with factors alone under the same scenario.

$\mathbf{z}_t$	RPR				Sieve-LS				PCA			
	Maturity(Year)				Maturity(Year)				Maturity(Year)			
	2	3	4	5	2	3	4	5	2	3	4	5
$(\mathbf{f}_t', \ \mathbf{w}_t')'$	37.9	32.6	25.6	22.8	37.1	31.9	25.3	22.1	23.9	21.4	17.4	17.5
$(\mathbf{f}_t', w_{1,t})'$	38.0	32.8	25.5	22.9	37.1	29.2	24.3	21.9	**33.3**	27.9	22.9	19.4
$(\mathbf{f}_t', w_{2,t})'$	37.6	31.1	25.2	22.0	37.2	30.0	25.2	21.7	29.0	22.9	18.1	15.6
$(\mathbf{f}_t', w_{3,t})'$	36.9	29.0	24.0	21.8	36.1	28.4	23.9	21.5	28.9	22.6	17.9	15.5
$(\mathbf{f}_t', w_{4,t})'$	38.1	32.8	25.7	22.9	36.1	29.1	24.7	22.3	**33.6**	27.7	22.3	**20.0**
$(\mathbf{f}_t', w_{5,t})'$	37.6	31.3	25.4	22.5	37.2	30.6	25.4	22.3	29.4	23.1	18.5	16.0
$(\mathbf{f}_t', w_{6,t})'$	37.5	30.6	25.4	22.0	37.3	30.1	25.3	21.9	29.0	22.9	18.0	15.4
$(\mathbf{f}_t', w_{7,t})'$	37.8	32.5	25.4	22.3	37.4	32.2	25.3	22.0	28.7	22.8	18.5	16.1
$(\mathbf{f}_t', w_{8,t})'$	38.1	32.9	25.6	22.9	37.3	32.4	25.3	22.5	**34.8**	**32.0**	**27.1**	**24.3**
	Maturity(Year)											
	2	3	4	5								
$\mathbf{w}_t$	6.1	5.5	4.7	4.5								

Table B.6: Full results for forecast out-of-sample R^2 (%) for factor-augmented multi-index model: the larger the better. The bold figures represent larger R^2 than forecast with factors alone under the same scenario.

$\mathbf{z}_t$	RPR				Sieve-LS				PCA			
	Maturity(Year)				Maturity(Year)				Maturity(Year)			
	2	3	4	5	2	3	4	5	2	3	4	5
$(\mathbf{f}_t', \ \mathbf{w}_t')'$	41.7	39.0	35.6	34.1	41.1	35.7	32.2	30.0	30.8	26.3	24.6	22.0
$(\mathbf{f}_t', w_{1,t})'$	43.4	38.2	34.5	30.9	39.5	37.3	32.2	28.8	**39.4**	**36.9**	**31.7**	**28.5**
$(\mathbf{f}_t', w_{2,t})'$	39.5	33.6	28.4	25.1	36.6	30.5	25.3	23.4	31.4	28.6	23.8	21.0
$(\mathbf{f}_t', w_{3,t})'$	38.8	33.9	29.2	26.0	36.5	30.8	25.8	23.8	31.0	28.5	23.8	21.0
$(\mathbf{f}_t', w_{4,t})'$	41.5	39.8	35.4	33.2	38.3	35.6	32.0	29.1	**36.2**	**34.4**	**30.7**	**28.2**
$(\mathbf{f}_t', w_{5,t})'$	37.6	34.3	30.1	27.5	36.9	31.9	27.2	24.9	29.2	27.2	23.1	19.8
$(\mathbf{f}_t', w_{6,t})'$	41.2	38.4	33.5	30.6	38.5	32.8	27.8	25.5	30.8	28.2	23.7	20.5
$(\mathbf{f}_t', w_{7,t})'$	39.0	35.2	31.7	28.8	36.5	31.0	25.9	24.0	31.0	28.8	24.2	21.4
$(\mathbf{f}_t', w_{8,t})'$	41.1	38.9	34.6	30.2	39.0	36.3	31.6	26.8	**35.0**	**33.2**	**28.6**	**24.2**
	Maturity(Year)											
	2	3	4	5								
$\mathbf{w}_t$	13.6	10.8	10.0	6.8								

C Proofs of Theorem 2.1

The proof relies on the Weyl's theorem and Davis-Kahan theorem in the random matrix theory, as we now cite:

Lemma C.1 (Weyl's and Davis-Kahan)**.** *For two $N \times N$ symmetric matrices $\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}$ (possibly random), let $\{\lambda_i\}_{i=1}^N$ be the eigenvalues of $\mathbf{A}_1$ in descending order and $\{\boldsymbol{\xi}_i\}_{i=1}^N$ be their associated eigenvectors. Correspondingly, let $\{\widehat{\lambda}_i\}_{i=1}^N$ be the eigenvalues of $\widehat{\mathbf{\Sigma}}$ in descending order and $\{\widehat{\boldsymbol{\xi}}_i\}_{i=1}^N$ be their associated eigenvectors. We have:*

(i) (Weyl's)

$$|\widehat{\lambda}_i - \lambda_i| \le \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|.$$

(ii) (Davis-Kahan)

$$\|\widehat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i\| \le \frac{\sqrt{2}\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|}{\min(|\widehat{\lambda}_{i-1} - \lambda_i|, |\lambda_i - \widehat{\lambda}_{i+1}|)}.$$

C.1 Proof of Theorem 2.1

We shall denote by $\{\widehat{\lambda}_i, \widehat{\boldsymbol{\xi}}_i\}$'s as the eigenvalues-vectors of $\widehat{\boldsymbol{\Sigma}}$, where the eigenvalues are in descending order, and apply Lemma C.1.

Step 1 shows that $\lambda_i = (c_i + o(1))N$, $i = 1, ..., K$, for some constants $c_1 > c_2 > ... > c_K$ that are in $[\underline{c}, \bar{c}]$.

In fact, from (2.2), we have

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda} E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\boldsymbol{\Lambda}'.$$

Hence the first K eigenvalues are the same as those of

$$E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}^{1/2}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}^{1/2}$$

which are also the same as those of $N\boldsymbol{\Sigma}_{\Lambda,N}^{1/2} E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\boldsymbol{\Sigma}_{\Lambda,N}^{1/2}$. By Assumptions 2.1 and 2.2, these eigenvalues are distinct, and can be written as $N(c_i + o(1))$ for some $c_i > 0$, $i = 1, ..., K$. Let them be ordered so that $c_1 > ... > c_K$. This proves the claim.

Step 2 show that the first K eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ satisfy: with probability approaching one,

$$(c_i + c_{i+1})N/2 \leq \widehat{\lambda}_i \leq (c_i + c_{i-1})N/2, \quad i = 1, ..., K.$$

The remaining eigenvalues are either bounded or growing uniformly at rate $o_P(N)$.

In fact, by Weyl's theorem, uniformly in $i \leq N$, $\max_{i\leq N} |\widehat{\lambda}_i - \lambda_i| = o_P(N)$. Therefore, with probability approaching one, by step 1, for $i \leq K - 1$,

$$(c_i + c_{i+1})N/2 = \lambda_i - (c_i - c_{i+1})N/2 \leq \widehat{\lambda}_i \leq \lambda_i + (c_{i-1} - c_i)N/2 = (c_i + c_{i-1})N/2.$$

For $i = K$, we have $c_K N/2 \leq \widehat{\lambda}_i \leq (c_K + c_{K-1})N/2$.

In addition, note that $\lambda_{K+1} = ... = \lambda_N = 0$. Hence $\widehat{\lambda}_N \leq ... \leq \widehat{\lambda}_{K+1} = o_P(N)$.

Step 3 show that

$$\max_{i\leq K} \|\widehat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i\| = o_P(1).$$

We respectively lower bound $|\widehat{\lambda}_{i-1} - \lambda_i|$ and $|\lambda_i - \widehat{\lambda}_{i+1}|$ for $i \leq K$. As for the first term, for any $i \leq K$, by Step 2, $\widehat{\lambda}_{i-1} - \lambda_i \geq (c_{i-1} + c_i)N/2 - c_i N = (c_{i-1} - c_i)N/2$ with probability approaching one. As for the second term, when $i \leq K - 1$, we have $\lambda_i - \widehat{\lambda}_{i+1} \geq c_i N - (c_{i+1} + c_i)N/2 = (c_i - c_{i-1})N/2$. When $i = K$, we have $\lambda_i - \widehat{\lambda}_{i+1} \geq c_i N/2$ with probability approaching one.

Therefore, by Davis-Kahan Theorem, $\max_{i\le K}\|\widehat{\boldsymbol{\xi}}_i-\boldsymbol{\xi}_i\|\le o_P(N)/N=o_P(1)$.

Step 4 Complete the proof.

From (2.2), we see that

$$\boldsymbol{\Sigma}=\boldsymbol{\Lambda}E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\boldsymbol{\Lambda}'.$$

Let

$$\mathbf{L}=\boldsymbol{\Sigma}_{\Lambda,N}^{1/2}E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\boldsymbol{\Sigma}_{\Lambda,N}^{1/2}.$$

Let $\mathbf{M}$ be a $K\times K$ matrix, whose columns are the eigenvectors of $\mathbf{L}$. Then $\mathbf{D}:=\mathbf{M}'\mathbf{L}\mathbf{M}$ is a diagonal matrix, with diagonal elements being the eigenvalues of $\mathbf{L}$, which are distinct values bounded away from zero by Assumption 2.2. Let $\mathbf{H}=\boldsymbol{\Sigma}_{\Lambda,N}^{-1/2}\mathbf{M}$. Then

$$\frac{1}{N}\boldsymbol{\Sigma}\boldsymbol{\Lambda}\mathbf{H}=\boldsymbol{\Lambda}E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\boldsymbol{\Sigma}_{\Lambda,N}\mathbf{H}=\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\Lambda,N}^{-1/2}\mathbf{L}\boldsymbol{\Sigma}_{\Lambda,N}^{-1/2}\boldsymbol{\Sigma}_{\Lambda,N}\mathbf{H}=\boldsymbol{\Lambda}\mathbf{H}\mathbf{M}'\mathbf{L}\mathbf{M}=\boldsymbol{\Lambda}\mathbf{H}\mathbf{D}.$$

In addition, note that $(\boldsymbol{\Lambda}\mathbf{H})'(\boldsymbol{\Lambda}\mathbf{H})=N\mathbf{I}_K$, hence the columns of $\boldsymbol{\Lambda}\mathbf{H}/\sqrt{N}$ are the eigenvectors of $\boldsymbol{\Sigma}$, corresponding to the K nonzero eigenvalues. By step 3, we have

$$\frac{1}{\sqrt{N}}\|\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}\mathbf{H}\|=o_P(1).$$

In addition, when $\boldsymbol{\Sigma}_{\Lambda,N}=\mathbf{I}_K$ and $E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}$ is diagonal, $\mathbf{M}=\mathbf{I}_K$. Thus $\mathbf{H}=\mathbf{I}_K$.

D Proofs for Section 3

Before formally present our proof, let us first provide a guideline of our proof's strategy. First of all, define, for $i=1,...,N$,

$$\mathbf{b}_i:=\arg\min_{\mathbf{b}\in\mathbb{R}^J}E[x_{it}-\mathbf{b}'\Phi(\mathbf{w}_t)]^2,\quad \mathbf{b}_{i,\alpha}=\arg\min_{\mathbf{b}\in\mathbb{R}^J}E\alpha_T^2\rho\left(\frac{x_{it}-\Phi(\mathbf{w}_t)'\mathbf{b}}{\alpha_T}\right).$$

Recall that

$$\widehat{\mathbf{b}}_i=\arg\min_{\mathbf{b}\in\mathbb{R}^J}\frac{1}{T}\sum_{t=1}^T\alpha_T^2\rho\left(\frac{x_{it}-\Phi(\mathbf{w}_t)'\mathbf{b}}{\alpha_T}\right),$$

As $\alpha_T \to \infty$, $\mathbf{b}_{i,\alpha}$ is expected to converge to $\mathbf{b}_i$ uniformly in $i \leq N$. This is true given some moment conditions on

$$\mathbf{e}_t := \mathbf{x}_t - E(\mathbf{x}_t|\mathbf{w}_t).$$

We shall first prove this in Section D.1.

In addition, let $\widetilde{\mathbf{V}}$ be a $K \times K$ diagonal matrix, whose diagonal elements are the first K eigenvalues of $\widehat{\mathbf{\Sigma}}/N := \frac{1}{TN}\sum_{t=1}^{T}\widehat{E}(\mathbf{x}_t|\mathbf{w}_t)\widehat{E}(\mathbf{x}_t|\mathbf{w}_t)'$. By the definition of $\widehat{\mathbf{\Lambda}}$, $\frac{1}{N}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Lambda}} = \widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}$. To effectively use this equality, we need to obtain the Bahadur representations of $\widehat{\mathbf{b}}_i$ and $\widehat{E}(\mathbf{x}_t|\mathbf{w}_t)$. These are achieved in Section D.2.

Furthermore, Section D.3 proves the rates of convergence for the estimated loadings. Sections D.4 and D.5 respectively present the proofs for the rates of convergence for $\mathbf{g}(\mathbf{w}_t)$ and $\boldsymbol{\gamma}_t$. Finally, Section D.6 proves the asymptotic normality of $\widehat{\boldsymbol{\gamma}}_t$.

D.1 The approximation error of the robust estimator

Proposition D.1. *For any* $4 < k < \zeta_2 + 2$,

$$\max_{i \leq N} \|\mathbf{b}_{i,\alpha} - \mathbf{b}_i\| = O(\alpha_T^{-(k-1)}).$$

Proof. We first prove that for any $0 < k < \zeta_2 + 2$, $\max_{i\leq N}\sup_{\mathbf{w}} E(|e_{it}|^k|\mathbf{w}_t = \mathbf{w}) < \infty$. In fact, uniformly in $\mathbf{w}$ for $\mathbf{w}_t = \mathbf{w}$ and $i \leq N$, as long as $\zeta_2 + 2 > k$

$$\begin{aligned} E(|e_{it}|^k|\mathbf{w}_t) &= \int_0^\infty P(|e_{it}|^k > x|\mathbf{w}_t)dx \leq 1 + \int_1^\infty P(|e_{it}|^k > x|\mathbf{w}_t)dx \\ &\leq 1 + \int_1^\infty E(e_{it}^2 1\{|e_{it}| > x^{1/k}\}|\mathbf{w}_t)x^{-2/k}dx \leq 1 + \int_1^\infty Cx^{-(\zeta_2+2)/k}dx < \infty. \end{aligned}$$

Since $\zeta_2 > 2$ by assumption, there is $k > 4$ so that $\max_{i\leq N}\sup_{\mathbf{w}} E(|e_{it}|^k|\mathbf{w}_t = \mathbf{w}) < \infty$.

Now recall that $\mathbf{b}_i = \arg\min E(x_{it} - \mathbf{b}_i'\Phi(\mathbf{w}_t))^2$. Hence

$$E[(x_{it}-\mathbf{b}_{i,\alpha}'\Phi(\mathbf{w}_t))^2-(x_{it}-\mathbf{b}_i'\Phi(\mathbf{w}_t))^2] = (\mathbf{b}_{i,\alpha}'-\mathbf{b}_i)'E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'(\mathbf{b}_{i,\alpha}-\mathbf{b}_i) \geq \underline{c}\|\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|^2$$

On the other hand, let $g_\alpha(z) := z^2 - \alpha_T^2\rho(z/\alpha_T)$. Then for $C > 0$ as a generic constant,

$$\begin{aligned} & E[(x_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{w}_t))^2 - (x_{it} - \mathbf{b}_i'\Phi(\mathbf{w}_t))^2] = Eg_\alpha(x_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{w}_t)) - Eg_\alpha(x_{it} - \mathbf{b}_i'\Phi(\mathbf{w}_t)) \\ & +E[\alpha_T^2\rho(\alpha_T^{-1}(x_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{w}_t))) - \alpha_T^2\rho(\alpha_T^{-1}(x_{it} - \mathbf{b}_i'\Phi(\mathbf{w}_t)))] \\ \leq_{(1)} \; & Eg_\alpha(x_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{w}_t)) - Eg_\alpha(x_{it} - \mathbf{b}_i'\Phi(\mathbf{w}_t)) \leq_{(2)} E[|2\tilde{z} - \alpha_T\dot{\rho}(\alpha_T^{-1}\tilde{z})||\Phi(\mathbf{w}_t)'(\mathbf{b}_i - \mathbf{b}_{i,\alpha})|], \end{aligned}$$

$$\begin{aligned}
&\leq_{(3)} \quad 2\alpha_T^{-(k-1)}E|\tilde{z}|^k|\Phi(\mathbf{w}_t)'(\mathbf{b}_i-\mathbf{b}_{i,\alpha})| \leq_{(4)} 2\alpha_T^{-(k-1)}E|z_{it}+e_{it}+(\mathbf{b}_i-\widetilde{\mathbf{b}}_i)'\Phi(\mathbf{w}_t)|^k|\Phi(\mathbf{w}_t)'(\mathbf{b}_i-\mathbf{b}_{i,\alpha})| \\
&\leq \quad C\alpha_T^{-(k-1)}E(C+|(\mathbf{b}_i-\widetilde{\mathbf{b}}_i)'\Phi(\mathbf{w}_t)|^k)|\Phi(\mathbf{w}_t)'(\mathbf{b}_i-\mathbf{b}_{i,\alpha})|
\end{aligned}$$

where (1) is due to the definition of $\mathbf{b}_{i,\alpha}$; (2) is by the mean value representation: $g_\alpha(z_1)-g_\alpha(z_2)=(2\tilde{z}-\alpha_T\dot{\rho}(\tilde{z}/\alpha_T))(z_1-z_2)$, with $z_1=x_{it}-\mathbf{b}_{i,\alpha}'\Phi(\mathbf{w}_t)$, $z_2=x_{it}-\mathbf{b}_i'\Phi(\mathbf{w}_t)$, and $\tilde{z}=x_{it}-\widetilde{\mathbf{b}}_i'\Phi(\mathbf{w}_t)$ for some $\widetilde{\mathbf{b}}_i$ lying between $\mathbf{b}_i$ and $\mathbf{b}_{i,\alpha}$; (3) is due to

$$|2\tilde{z}-\alpha_T\dot{\rho}(\alpha_T^{-1}\tilde{z})| \leq 2|\tilde{z}|1\{|\tilde{z}|>\alpha_T\} \leq 2|\tilde{z}|\frac{|\tilde{z}|^{k-1}}{\alpha_T^{k-1}}1\{|\tilde{z}|>\alpha_T\} \leq 2|\tilde{z}|^k/\alpha_T^{k-1}.$$

(4) follows from $\tilde{z}=x_{it}-E(x_{it}|\mathbf{w}_t)+\mathbf{b}_i'\Phi(\mathbf{w}_t)+z_{it}-\widetilde{\mathbf{b}}_i'\Phi(\mathbf{w}_t)$, and that $e_{it}:=x_{it}-E(x_{it}|\mathbf{w}_t)$.

Next, for ease of presentation, we introduce $M_{it}:=C+|(\mathbf{b}_i-\widetilde{\mathbf{b}}_i)'\Phi(\mathbf{w}_t)|^k$ and $\boldsymbol{\Delta}_i:=\mathbf{b}_i-\mathbf{b}_{i,\alpha}$. Then the above inequality can be further written as:

$$\begin{aligned}
&= \quad C\alpha_T^{-(k-1)}EM_{it}|\Phi(\mathbf{w}_t)'\boldsymbol{\Delta}_i| = C\alpha_T^{-(k-1)}E[M_{it}^2\boldsymbol{\Delta}_i'\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\boldsymbol{\Delta}_i]^{1/2} \\
&\leq_{(5)} \quad C\alpha_T^{-(k-1)}[\boldsymbol{\Delta}_i'EM_{it}^2\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\boldsymbol{\Delta}_i]^{1/2} \leq C\alpha_T^{-(k-1)}\|EM_{it}^2\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|^{1/2}\|\boldsymbol{\Delta}_i\|.
\end{aligned}$$

We now bound $\max_{i\leq N}\|EM_{it}^2\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|=\max_{i\leq N}\sup_{\|\boldsymbol{\nu}\|=1}EM_{it}^2(\Phi(\mathbf{w}_t)'\boldsymbol{\nu})^2=O(1)$. By the Cauchy-Schwarz inequality, since $\Phi(\mathbf{w}_t)'\boldsymbol{\nu}$ is sub-Gaussian with the universal parameter,

$$\begin{aligned}
&\sup_{\|\boldsymbol{\nu}\|=1}[EM_{it}^2(\Phi(\mathbf{w}_t)'\boldsymbol{\nu})^2]^2 \leq EM_{it}^4\sup_{\|\boldsymbol{\nu}\|=1}E(\Phi(\mathbf{w}_t)'\boldsymbol{\nu})^4 \leq CEM_{it}^4 \leq C(C+E|(\mathbf{b}_i-\widetilde{\mathbf{b}}_i)'\Phi(\mathbf{w}_t)|^{4k}) \\
&\leq \quad C+CE\|\mathbf{b}_i-\widetilde{\mathbf{b}}_i\|^{4k}\left(\frac{(\mathbf{b}_i'-\widetilde{\mathbf{b}}_i)'}{\|\mathbf{b}_i-\widetilde{\mathbf{b}}_i\|}\Phi(\mathbf{w}_t)\right)^{4k} \leq C+C\|\boldsymbol{\Delta}_i\|^{4k}\sup_{\|\boldsymbol{\nu}\|=1}E(\boldsymbol{\nu}'\Phi(\mathbf{w}_t))^{4k} \leq C+C\|\boldsymbol{\Delta}_i\|^{4k}.
\end{aligned}$$

Therefore, we have proved that uniformly in i,

$$E[(x_{it}-\mathbf{b}_{i,\alpha}'\Phi(\mathbf{w}_t))^2-(x_{it}-\mathbf{b}_i'\Phi(\mathbf{w}_t))^2] \leq C\alpha_T^{-(k-1)}(C+C\|\boldsymbol{\Delta}_i\|^{4k})^{1/4}\|\boldsymbol{\Delta}_i\| \leq C\alpha_T^{-(k-1)}(1+\|\boldsymbol{\Delta}_i\|^k)\|\boldsymbol{\Delta}_i\|$$

We have also proved that the left hand side is lower bounded by $\underline{c}\|\boldsymbol{\Delta}_i\|^2$. Uniformly in i,

$$\|\boldsymbol{\Delta}_i\| \leq C\alpha_T^{-(k-1)}(1+\|\boldsymbol{\Delta}_i\|^k).$$

If $\max_i\|\boldsymbol{\Delta}_i\|=O(1)$, then $\|\boldsymbol{\Delta}_i\|\leq C\alpha_T^{-(k-1)}$. Otherwise, $\max_i\|\boldsymbol{\Delta}_i\|\leq C\alpha_T^{-(k-1)}\max_i\|\boldsymbol{\Delta}_i\|^k$, which then implies $1\leq C(\max_i\|\boldsymbol{\Delta}_i\|/\alpha_T)^{k-1}$. However, note that $\|\boldsymbol{\Delta}_i\|\leq\|\mathbf{b}_i\|+\|\mathbf{b}_{i,\alpha}\|\leq CJ^{1/2}$, and $J=o(\alpha_T^2)$, we have $\max_i\|\boldsymbol{\Delta}_i\|/\alpha_T=o(1)$, which is a contradiction. Therefore,

$\max_i \|\boldsymbol{\Delta}_i\| \leq C\alpha_T^{-(k-1)}$. Q.E.D.

The following lemma shows the sieve approximation error is uniformly controlled.

Lemma D.1. *Under Assumption 3.2, there is $\eta \geq 1$, as $J \to \infty$,*

$$\max_{i\leq N}\sup_{\mathbf{w}} |E(x_{it}|\mathbf{w}_t = \mathbf{w}) - \mathbf{b}_i'\Phi(\mathbf{w})| = O(J^{-\eta}).$$

Proof. Recall that for $k \leq K$, $\mathbf{v}_k = \arg\min_{\mathbf{v}} E(f_{kt} - \mathbf{v}'\Phi(\mathbf{w}_t))^2 = (E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')^{-1}E\Phi(\mathbf{w}_t)f_{kt}$, and that $\mathbf{b}_i = \arg\min_{\mathbf{b}\in\mathbb{R}^J} E[x_{it} - \mathbf{b}'\Phi(\mathbf{w}_t)]^2 = (E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')^{-1}E\Phi(\mathbf{w}_t)x_{it}$. Also note that $x_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it}$. We have $\mathbf{b}_i = \sum_{k=1}^K \mathbf{v}_k\lambda_{ik}$. Hence

$$\begin{aligned}
\max_{i\leq N}\sup_{\mathbf{w}} |E(x_{it}|\mathbf{w}_t = \mathbf{w}) - \mathbf{b}_i'\Phi(\mathbf{w})| &\leq \max_{i\leq N}\sup_{\mathbf{w}} |\sum_{k=1}^K \lambda_{ik}(E(f_{tk}|\mathbf{w}_t = \mathbf{w}) - \mathbf{v}_k'\Phi(\mathbf{w}))| \\
&\leq O(1)\max_k \sup_{\mathbf{w}} |E(f_{tk}|\mathbf{w}_t = \mathbf{w}) - \mathbf{v}_k'\Phi(\mathbf{w})| \\
&= O(J^{-\eta}).
\end{aligned}$$

D.2 Bahadur representation of the robust estimator

We now give the uniform convergence rate of $\widehat{\mathbf{b}}_i$ as well as its Bahadur representation.

Define

$$Q_i(\mathbf{b}) = \frac{1}{T}\sum_{t=1}^T \alpha_T^2\rho\left(\frac{x_{it} - \Phi(\mathbf{w}_t)'\mathbf{b}}{\alpha_T}\right).$$

Proposition D.2. *When $\alpha_T \leq C\sqrt{T/\log(NJ)}$ for any $C > 0$, and any $4 < k < \zeta_2 + 2$,*

$$\max_{i\leq N} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i\| = O_P(\sqrt{\frac{J\log N}{T}} + \alpha_T^{-(k-1)}).$$

Proof. Let $m_T = \sqrt{\frac{J\log N}{T}}$. We aim to show, for any $\epsilon > 0$, there is $\delta > 0$, when for all large N, T,

$$P(\min_{i\leq N}\inf_{\|\boldsymbol{\nu}\|=\delta} Q_i(\mathbf{b}_{i,\alpha} + m_T\boldsymbol{\nu}) - Q_i(\mathbf{b}_{i,\alpha}) > 0) > 1 - \epsilon.$$

This then implies $\max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}\| = O_P(m_T)$. The result then follows from Proposition D.1.

By the definition of $\mathbf{b}_{i,\alpha}$,

$$E[\Phi(\mathbf{w}_t)\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})] = 0, \qquad e_{it,\alpha} := x_{it} - \Phi(\mathbf{w}_t)'\mathbf{b}_{i,\alpha}.$$

In addition, we have $e_{it} = e_{it,\alpha} + \Delta_{it,\alpha}$, where $\Delta_{it,\alpha} := (\mathbf{b}_{i,\alpha} - \mathbf{b}_i)'\Phi(\mathbf{w}_t) - z_{it}$. Using the formula: $\rho(a+t) - \rho(a) = \dot{\rho}(a)t + \int_0^t (\dot{\rho}(a+x) - \dot{\rho}(a))dx$ for $a = \alpha_T^{-1} e_{it,\alpha}$ and $t = -m_T \alpha_T^{-1} \Phi(\mathbf{w}_t)'\boldsymbol{\nu}$,

$$
\begin{aligned}
& Q_i(\mathbf{b}_{i,\alpha} + m_T\boldsymbol{\nu}) - Q_i(\mathbf{b}_{i,\alpha}) = -\frac{1}{T}\sum_{t=1}^{T} m_T \alpha_T \dot{\rho}(\alpha_T^{-1} e_{it,\alpha})\Phi(\mathbf{w}_t)'\boldsymbol{\nu} \\
& + \frac{1}{T}\sum_{t=1}^{T} 1\{\Phi(\mathbf{w}_t)'\boldsymbol{\nu} < 0\}\alpha_T^2 \int_0^{-m_T\alpha_T^{-1}\Phi(\mathbf{w}_t)'\boldsymbol{\nu}} \dot{\rho}(\alpha_T^{-1} e_{it,\alpha} + x) - \dot{\rho}(\alpha_T^{-1} e_{it,\alpha})dx \\
& - \frac{1}{T}\sum_{t=1}^{T} 1\{\Phi(\mathbf{w}_t)'\boldsymbol{\nu} > 0\}\alpha_T^2 \int_{-m_T\alpha_T^{-1}\Phi(\mathbf{w}_t)'\boldsymbol{\nu}}^{0} \dot{\rho}(\alpha_T^{-1} e_{it,\alpha} + x) - \dot{\rho}(\alpha_T^{-1} e_{it,\alpha})dx.
\end{aligned}
$$

By the definition of $\dot{\rho}$, the integrant can be rewritten as:

$$
\begin{aligned}
& \dot{\rho}(\alpha_T^{-1} e_{it,\alpha} + x) - \dot{\rho}(\alpha_T^{-1} e_{it,\alpha}) = 2x1\{|\alpha_T^{-1} e_{it,\alpha} + x| < 1, |\alpha_T^{-1} e_{it,\alpha}| < 1\} \\
& + (\dot{\rho}(\alpha_T^{-1} e_{it,\alpha} + x) - \dot{\rho}(\alpha_T^{-1} e_{it,\alpha}))1\{|\alpha_T^{-1} e_{it,\alpha} + x| \geq 1, \text{ or } |\alpha_T^{-1} e_{it,\alpha}| \geq 1\} \\
= \; & 2x - (\dot{\rho}(\alpha_T^{-1} e_{it,\alpha} + x) - \dot{\rho}(\alpha_T^{-1} e_{it,\alpha}) - 2x)1\{|\alpha_T^{-1} e_{it,\alpha} + x| \geq 1, \text{ or } |\alpha_T^{-1} e_{it,\alpha}| \geq 1\}.
\end{aligned}
$$

In addition, note that

$$
|\dot{\rho}(x_1) - \dot{\rho}(x_1)| \leq 2|x_1 - x_2|, \quad \forall x_1, x_2.
$$

Thus we can further write:

$$
\begin{aligned}
& Q_i(\mathbf{b}_{i,\alpha} + m_T\boldsymbol{\nu}) - Q_i(\mathbf{b}_{i,\alpha}) = -\frac{1}{T}\sum_{t=1}^{T} m_T \alpha_T \dot{\rho}(\alpha_T^{-1} e_{it,\alpha})\Phi(\mathbf{w}_t)'\boldsymbol{\nu} + \frac{1}{T}\sum_{t=1}^{T}\alpha_T^2 \int_0^{-m_T\alpha_T^{-1}\Phi(\mathbf{w}_t)'\boldsymbol{\nu}} 2x dx \\
& - \frac{1}{T}\sum_{t=1}^{T} 1\{\Phi(\mathbf{w}_t)'\boldsymbol{\nu} < 0\}\alpha_T^2 \int_0^{-m_T\alpha_T^{-1}\Phi(\mathbf{w}_t)'\boldsymbol{\nu}} (\dot{\rho}(\alpha_T^{-1} e_{it,\alpha} + x) - \dot{\rho}(\alpha_T^{-1} e_{it,\alpha}) - 2x) \\
& \times 1\{|\alpha_T^{-1} e_{it,\alpha} + x| \geq 1, \text{ or } |\alpha_T^{-1} e_{it,\alpha}| \geq 1\}dx \\
& + \frac{1}{T}\sum_{t=1}^{T} 1\{\Phi(\mathbf{w}_t)'\boldsymbol{\nu} > 0\}\alpha_T^2 \int_{-m_T\alpha_T^{-1}\Phi(\mathbf{w}_t)'\boldsymbol{\nu}}^{0} (\dot{\rho}(\alpha_T^{-1} e_{it,\alpha} + x) - \dot{\rho}(\alpha_T^{-1} e_{it,\alpha}) - 2x) \\
& \times 1\{|\alpha_T^{-1} e_{it,\alpha} + x| \geq 1, \text{ or } |\alpha_T^{-1} e_{it,\alpha}| \geq 1\}dx \\
\geq \; & \inf_{\|\boldsymbol{\nu}\|=\delta} \frac{1}{T}\sum_{t=1}^{T}\alpha_T^2(-m_T\alpha_T^{-1}\Phi(\mathbf{w}_t)'\boldsymbol{\nu})^2 - \max_i \sup_{\|\boldsymbol{\nu}\|=\delta} \left|\frac{1}{T}\sum_{t=1}^{T} m_T\alpha_T\dot{\rho}(\alpha_T^{-1} e_{it,\alpha})\Phi(\mathbf{w}_t)'\boldsymbol{\nu}\right| \\
& - \max_i \sup_{\|\boldsymbol{\nu}\|=1} \frac{1}{T}\sum_{t=1}^{T}\alpha_T^2 \int_0^{m_T\alpha_T^{-1}|\Phi(\mathbf{w}_t)'\boldsymbol{\nu}|} 4x1\{|\alpha_T^{-1} e_{it,\alpha} + x| \geq 1, \text{ or } |\alpha_T^{-1} e_{it,\alpha}| \geq 1\}dx \\
:= \; & A_1 - A_2 - A_3.
\end{aligned}
$$

We now lower bound A_1 and upper bound A_2, A_3.

First of all, there is $c>0$ independent of δ, with probability approaching one,

$$A_1=\inf_{\|\boldsymbol{\nu}\|=\delta}\boldsymbol{\nu}'\frac{1}{T}\sum_{t=1}^T m_T^2\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\boldsymbol{\nu}\geq\lambda_{\min}(\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')m_T^2\delta^2\geq cm_T^2\delta^2.$$

As for A_2, note that $|\alpha_T\dot\rho(\alpha_T^{-1}e_{it,\alpha})|\leq|e_{it,\alpha}|\leq|e_{it}|+|\Delta_{it,\alpha}|$. Uniformly in $i\leq N, j\leq J$, by Holder's inequality, with an arbitrarily small $v>0$, and $p=(1+v)^{-1}$,

$$\begin{aligned}&E(\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{w}_t))^2\leq\alpha_T^{-2}E(\alpha_T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{w}_t))^2\leq2\alpha_T^{-2}E(e_{it}^2+\Delta_{it,\alpha}^2)\phi_j(\mathbf{w}_t)^2\\ \leq\;&2\alpha_T^{-2}EE\{e_{it}^2|\mathbf{w}_t\}\phi_j(\mathbf{w}_t)^2+2\alpha_T^{-2}E\Delta_{it,\alpha}^2\phi_j(\mathbf{w}_t)^2\leq C\alpha_T^{-2}((E\{e_{it}^2|\mathbf{w}_t\}^{1+v})^{1/p}+C)\leq C\alpha_T^{-2}.\end{aligned}$$

Note that $|\dot\rho|<2$ and $\{\phi_j(\mathbf{w}_t)\}$ is sub-Gaussian, thus by the Bernstein inequality, for $x=2\log(NJ)$,

$$P(|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{w}_t)|>\sqrt{\frac{2E(\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{w}_t))^2x}{T}}+\frac{Cx}{T})\leq2\exp(-x).$$

Note that when $\alpha_T\leq C\sqrt{T/\log(NJ)}$,

$$\sqrt{\frac{2E(\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{w}_t))^2x}{T}}+\frac{Cx}{T}\leq\sqrt{\frac{C\log(NJ)}{\alpha_T^2T}}+\frac{C\log(NJ)}{T}\leq2\sqrt{\frac{C\log(NJ)}{\alpha_T^2T}}.$$

Thus

$$P(\max_{ij}|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{w}_t)|>\sqrt{\frac{C\log(NJ)}{\alpha_T^2T}})\leq CNJ\exp(-2\log(NJ))=\frac{C}{NJ}.$$

Therefore, with probability approaching one,

$$\begin{aligned}A_2\;&\leq\;m_T\alpha_T\delta\max_i\|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\Phi(\mathbf{w}_t)'\|\leq m_T\alpha_T\sqrt{J}\delta\max_{i\leq N,j\leq J}|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{w}_t)|\\&\leq\;\delta m_T\alpha_T\sqrt{\frac{CJ\log(N)}{\alpha_T^2T}}=\delta m_T\sqrt{\frac{CJ\log(N)}{T}}.\end{aligned}$$

As for A_3, note that uniformly for $x\leq m_T\alpha_T^{-1}|\Phi(\mathbf{w}_t)'\boldsymbol{\nu}|$, and $e_{it}=e_{it,\alpha}+\Delta_{it,\alpha}$,

$$1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq1,\text{ or }|\alpha_T^{-1}e_{it,\alpha}|\geq1\}\leq1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq1\}+1\{|\alpha_T^{-1}e_{it,\alpha}|\geq1\}$$

$$\leq \quad 2\times 1\{|e_{it,\alpha}|>3\alpha_T/4\}+1\{m_T|\Phi(\mathbf{w}_t)'\boldsymbol{\nu}|>\alpha_T/4\}$$
$$\leq \quad 2\times 1\{|e_{it}|>\alpha_T/2\}+1\{m_T|\Phi(\mathbf{w}_t)'\boldsymbol{\nu}|>\alpha_T/4\}+1\{|\Delta_{it,\alpha}|>\alpha_T/4\}.$$

In addition, with probability at least $1-\epsilon/10$,

$$\begin{aligned}
\max_i \frac{1}{T}\sum_{t=1}^T 1\{|e_{it}|>\alpha_T/2\} &\leq_{(1)} \max_i P(|e_{it}|>\alpha_T/2)+\sqrt{\frac{\log N}{T}}\max_i P(|e_{it}|>\alpha_T/2)^{1/2},\\
\frac{1}{T}\sum_t 1\{m_T\|\Phi(\mathbf{w}_t)\|\delta>\alpha_T/4\} &\leq 10P(m_T\delta\|\Phi(\mathbf{w}_t)\|>\alpha_T/4)/\epsilon,\\
\max_i \frac{1}{T}\sum_{t=1}^T 1\{|\Delta_{it,\alpha}|>\alpha_T/4\} &\leq \max_i \frac{1}{T}\sum_{t=1}^T 1\{\|\Phi(\mathbf{w}_t)\|>C\alpha_T^k\}+1\{|z_{it}|>\alpha_T/4\}\\
&\leq 10P(\|\Phi(\mathbf{w}_t)\|>C\alpha_T^k)/\epsilon+CJ^{-2\eta}/\alpha_T^2+\sqrt{\frac{\log N}{T}}CJ^{-\eta}/\alpha_T,
\end{aligned}$$

where (1) follows from the triangular inequality,

$$\max_i \frac{1}{T}\sum_{t=1}^T 1\{|e_{it}|>\alpha_T/2\}\leq \max_i P(|e_{it}|>\alpha_T/2)+\max_i|\frac{1}{T}\sum_{t=1}^T 1\{|e_{it}|>\alpha_T/2\}-P(|e_{it}|>\alpha_T/2)|,$$

and we used Bernstein inequality+union bound to bound the second term since the indicator function is bounded. Hence for an arbitrarily small $v>0$, by Holder's inequality, for some generic constant $C>0$, independent of δ,

$$\begin{aligned}
A_3 &\leq \max_i \sup_{\|\boldsymbol{\nu}\|=\delta}\frac{1}{T}\sum_{t=1}^T 4(m_T|\Phi(\mathbf{w}_t)'\boldsymbol{\nu}|)^2[1\{|e_{it}|>\alpha_T/2\}+1\{m_T|\Phi(\mathbf{w}_t)'\boldsymbol{\nu}|>\alpha_T/4\}+1\{|\Delta_{it,\alpha}|>\alpha_T/4\}]\\
&\leq C\max_i(\frac{1}{T}\sum_{t=1}^T[1\{|e_{it}|>\alpha_T/2\}+1\{m_T\delta\|\Phi(\mathbf{w}_t)\|>\alpha_T/4\}+1\{|\Delta_{it,\alpha}|>\alpha_T/4\}])^{1-v}\\
&\quad\times(\frac{1}{T}\sum_{t=1}^T\|\Phi(\mathbf{w}_t)\|^{2/v})^v(m_T\delta)^2\\
&\leq (m_T\delta)^2C\Bigg(\max_i P(|e_{it}|>\alpha_T/2)+\sqrt{\frac{\log N}{T}}\max_i P(|e_{it}|>\alpha_T/2)^{1/2}+10P(m_T\delta\|\Phi(\mathbf{w}_t)\|>\alpha_T/4)/\epsilon\\
&\quad+10P(\|\Phi(\mathbf{w}_t)\|>C\alpha_T^k)/\epsilon+CJ^{-2\eta}/\alpha_T^2+\sqrt{\frac{\log N}{T}}CJ^{-\eta}/\alpha_T\Bigg)^{1-v}(C+E\|\Phi(\mathbf{w}_t)\|^{2/v})^v.
\end{aligned}$$

We now upper bound $E\|\Phi(\mathbf{w}_t)\|^{2/v}$ and $P(\|\Phi(\mathbf{w}_t)\|>x)$ for any x. Since $\{\phi_j(w_t)\}_{j\leq J}$ is sub-Gaussian, by Lemma 14.12 of Bühlmann and van de Geer (2011),

$$\begin{aligned}
E\|\Phi(\mathbf{w}_t)\|^{2/v} &\leq J^{1/v}E(\max_{j\leq J}\phi_j(\mathbf{w}_t)^{2/v}) \leq J^{1/v}E(\max_{j\leq J}|\phi_j(\mathbf{w}_t)^{2/v} - E\phi_j(\mathbf{w}_t)^{2/v}|) \\
&\quad + J^{1/v}\max_j E\phi_j(\mathbf{w}_t)^{2/v} \leq J^{1/v}C\log(J). \\
P(\|\Phi(\mathbf{w}_t)\| > x) &\leq P(\max_j |\phi_j(\mathbf{w}_t)|^2 J > x^2) \leq J\max_j P(|\phi_j(\mathbf{w}_t)| > x/J^{1/2}) \leq J\exp(-Cx^2/J).
\end{aligned}$$

Therefore,

$$\begin{aligned}
A_3 &\leq (m_T\delta)^2 C\Bigg(\max_i P(|e_{it}| > \alpha_T/2) + \sqrt{\frac{\log N}{T}}\max_i P(|e_{it}| > \alpha_T/2)^{1/2} + CJ\exp(-C\alpha_T^2/(Jm_T^2\delta^2))/\epsilon \\
&\quad + CJ\exp(-C\alpha_T^{2k}/J)/\epsilon + CJ^{-2\eta}/\alpha_T^2 + \sqrt{\frac{\log N}{T}}CJ^{-\eta}/\alpha_T\Bigg)^{1-v} J(\log J)^v := (m_T\delta)^2 C l_T.
\end{aligned}$$

Note that $l_T = o(1)$.

Consequently, for any $\epsilon > 0$, there are C, c, and c_ϵ independent of δ (may depend on ϵ), with probability at least $1-\epsilon$, uniformly in $i \leq N$ and $\|\boldsymbol{\nu}\| = \delta$, for $m_T = \sqrt{\frac{J\log N}{T}}$,

$$Q_i(\mathbf{b}_{i,\alpha} + m_T\boldsymbol{\nu}) - Q_i(\mathbf{b}_{i,\alpha}) \geq m_T^2\delta^2(c - c_\epsilon l_T) - \delta m_T C\sqrt{\frac{J\log N}{T}} \geq m_T\delta(m_T\delta c/2 - Cm_T) > 0$$

so long as $\delta c > 2C$. Thus $\max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}\| = O_P(m_T)$.

We now prove a simple lemma.

Lemma D.2. *There is $M > 0$ for all $x > M$,*

$$\begin{aligned}
\max_{i\leq N}\sup_{\mathbf{w}} P(|e_{it}| > x|\mathbf{w}_t = \mathbf{w}) &\leq Cx^{-\zeta_2-2} \\
\max_{i\leq N}\sup_{\mathbf{w}} E(|e_{it}|1\{|e_{it}| > x\}|\mathbf{w}_t = \mathbf{w}) &\leq Cx^{-\zeta_2-1}.
\end{aligned}$$

Proof. Uniformly in $\mathbf{w} = \mathbf{w}_t$ and $i \leq N$,

$$\begin{aligned}
P(|e_{it}| > x|\mathbf{w}_t) &= E(1\{|e_{it}| > x\}|\mathbf{w}_t) \leq E(e_{it}^2 1\{|e_{it}| > x\}|\mathbf{w}_t)x^{-2} \leq Cx^{-\zeta_2-2} \\
E(|e_{it}|1\{|e_{it}| > x\}|\mathbf{w}_t) &\leq E(e_{it}^2 1\{|e_{it}| > x\}|\mathbf{w}_t)x^{-1} \leq Cx^{-\zeta_2-1}.
\end{aligned}$$

Lemma D.3. *Uniformly for $i = 1, ..., N$,*

$$\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha} = (2E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')^{-1}\frac{1}{T}\sum_{t=1}^T \alpha_T\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})\Phi(\mathbf{w}_t) + \mathbf{R}_{i,b},$$

where $\max_{i\leq N}\|\mathbf{R}_{i,b}\| = O_P(\alpha_T^{-(\zeta_1-1)} + \sqrt{\frac{\log J}{T}})J\sqrt{\frac{J\log N}{T}}$.

Proof. Note that $\nabla Q_i(\mathbf{b}) = -\frac{1}{T}\sum_{t=1}^T \alpha_T\dot\rho(\alpha_T^{-1}(x_{it}-\Phi(\mathbf{w}_t)'\mathbf{b}))\Phi(\mathbf{w}_t)$. Define $\bar Q_i(\mathbf{b}) = EQ_i(\mathbf{b})$,

$$\boldsymbol{\mu}_i(\mathbf{b}) := \nabla Q_i(\mathbf{b}) - \nabla\bar Q_i(\mathbf{b}) = E\alpha_T\dot\rho(\alpha_T^{-1}(x_{it}-\Phi(\mathbf{w}_t)'\mathbf{b}))\Phi(\mathbf{w}_t) - \frac{1}{T}\sum_{t=1}^T \alpha_T\dot\rho(\alpha_T^{-1}(x_{it}-\Phi(\mathbf{w}_t)'\mathbf{b}))\Phi(\mathbf{w}_t).$$

The first order condition gives $\nabla Q_i(\widehat{\mathbf{b}}_i) = 0$. By the mean value expansion,

$$\begin{aligned} 0 &= \nabla Q_i(\widehat{\mathbf{b}}_i) - \nabla\bar Q_i(\widehat{\mathbf{b}}_i) + \nabla\bar Q_i(\widehat{\mathbf{b}}_i) - \nabla\bar Q_i(\mathbf{b}_{i,\alpha}) + \nabla\bar Q_i(\mathbf{b}_{i,\alpha}) - \nabla Q_i(\mathbf{b}_{i,\alpha}) + \nabla Q_i(\mathbf{b}_{i,\alpha}) \\ &= \boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i) + \nabla\bar Q_i(\widehat{\mathbf{b}}_i) - \nabla\bar Q_i(\mathbf{b}_{i,\alpha}) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha}) + \nabla Q_i(\mathbf{b}_{i,\alpha}) \\ &= \nabla^2\bar Q_i(\widetilde{\mathbf{b}}_i)(\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}) + \nabla Q_i(\mathbf{b}_{i,\alpha}) + \boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha}). \end{aligned}$$

for some $\widetilde{\mathbf{b}}_i$ in the segment joining $\widehat{\mathbf{b}}_i$ and $\mathbf{b}_{i,\alpha}$. We now proceed by: (i) upper bounding $\max_i\|\boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\|$, and (ii) finding the limit of $\nabla^2\bar Q_i(\widetilde{\mathbf{b}}_i)$ uniformly in i.

(i) Note that in the proof of Proposition D.2, we have proved that for any $\epsilon > 0$, there is $\delta > 0$, so that the following event holds with probability at least $1-\epsilon$:

$$\max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}\| \leq \delta m_T, \quad m_T = \sqrt{\frac{J\log N}{T}}.$$

We bound $E\max_i\sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}\|\boldsymbol{\mu}_i(\mathbf{b}) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\|$. Let $\mu_{ij}(\cdot)$ be the jth element of $\boldsymbol{\mu}_i$, $j\leq J$. Since $\{\mathbf{x}_t,\mathbf{w}_t\}_{t\leq T}$ are serially independent, there exists a Radamacher sequence $\{\varepsilon_t\}_{t\leq T}$ with $P(\varepsilon_t = 1) = P(\varepsilon_t = -1) = 1/2$, that is independent of $\{\mathbf{x}_t,\mathbf{w}_t\}$,

$$\begin{aligned} & E\max_{i\leq N, j\leq J}\sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}|\mu_{ij}(\mathbf{b}) - \mu_{ij}(\mathbf{b}_{i,\alpha})| \\ \leq_{(a)} &\ 2E\max_{i\leq N, j\leq J}\sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\alpha_T\left(\dot\rho(\alpha_T^{-1}(x_{it}-\Phi(\mathbf{w}_t)'\mathbf{b})) - \dot\rho(\alpha_T^{-1}(x_{it}-\Phi(\mathbf{w}_t)'\mathbf{b}_{i,\alpha}))\right)\phi_j(\mathbf{w}_t)| \\ \leq_{(b)} &\ 4E\max_{i\leq N, j\leq J}\sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\Phi(\mathbf{w}_t)'(\mathbf{b}_{i,\alpha}-\mathbf{b})\phi_j(\mathbf{w}_t)| \\ \leq &\ 4\delta m_T E\max_{j\leq J}\|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\phi_j(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\| \leq 4\delta m_T\sqrt{J}E\max_{l,j\leq J}|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\phi_j(\mathbf{w}_t)\phi_l(\mathbf{w}_t)| \\ \leq_{(c)} &\ 4\delta m_T\sqrt{J}\frac{L}{T}\log E\exp\left(L^{-1}\max_{l,j\leq J}|\sum_{t=1}^T\varepsilon_t\phi_j(\mathbf{w}_t)\phi_l(\mathbf{w}_t)|\right) \end{aligned}$$

$$
\begin{aligned}
&\leq_{(d)} \quad 4\delta m_T\sqrt{J}\frac{L}{T}\log\sum_{l,j\leq J} E\exp\left(L^{-1}|\sum_{t=1}^T \varepsilon_t\phi_j(\mathbf{w}_t)\phi_l(\mathbf{w}_t)|\right)\\
&\leq_{(e)} \quad 4\delta m_T\sqrt{J}\frac{L}{T}\log\sum_{l,j\leq J}\exp\left(\frac{T}{2(L^2-LK_0)}\right) = 4\delta m_T\sqrt{J}\frac{L}{T}\left(2\log J+\frac{T}{2(L^2-LK_0)}\right)\\
&= \quad 4\delta m_T\sqrt{J}\left(\frac{2L\log J}{T}+\sqrt{\frac{c_0\log J}{4T}}\right)\leq C\delta m_T\sqrt{\frac{J\log J}{T}}.
\end{aligned}
$$

Note that $|\dot{\rho}(\cdot)|\leq 2$ and $\{\phi_j(\cdot)\}$ is sub-Gaussian, hence (a) follows from the symmetrization theorem (see, e.g., Theorem 14.3 of Bühlmann and van de Geer (2011)); since $\dot{\rho}(\cdot)$ is Lipschitz continuous, (b) follows from the contraction theorem (e.g., Theorem 14.4 of Bühlmann and van de Geer (2011)). Let K_0 denote constant parameter of the sub-Gaussianity of $\{\phi_l(\mathbf{w}_t)\phi_j(\mathbf{w}_t)\}_{l,j\leq J}$; for some $c_0>0$, let

$$
L = K_0+\sqrt{\frac{T}{c_0\log J}}.
$$

Then (c) follows from the Jensen's inequality; (d) follows from the simple inequality that $\exp(\max)\leq\sum\exp$; (e) follows from an inequality of exponential moment of an average for sub-Gaussian random variables (Lemma 14.8 of Bühlmann and van de Geer (2011)).

Therefore,

$$
E\max_i \sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}\|\boldsymbol{\mu}_i(\mathbf{b})-\boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\|\leq CJm_T\sqrt{\frac{\log J}{T}}=\frac{CJ^{3/2}(\log N\log J)^{1/2}}{T}.
$$

Hence

$$
\max_i\|\boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i)-\boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\| = O_P(J^{3/2}(\log N\log J)^{1/2}/T).
$$

(ii) Note that

$$
\nabla\bar{Q}_i(\mathbf{b}) = -E\Phi(\mathbf{w}_t)\alpha_T\dot{\rho}(\alpha_T^{-1}(e_{it}+z_{it})+\alpha_T^{-1}\Phi(\mathbf{w}_t)'(\mathbf{b}_i-\mathbf{b})) = -E\Phi(\mathbf{w}_t)A_{it}(\mathbf{b})
$$

where $A_{it}(\mathbf{b})=E[\alpha_T\dot{\rho}(\alpha_T^{-1}(e_{it}+z_{it})+\alpha_T^{-1}\Phi(\mathbf{w}_t)'(\mathbf{b}_i-\mathbf{b}))|\mathbf{w}_t]$. Let $g_{e,i}$ denote the density of e_{it}, and let P_e denote the conditional probability measure conditioning on $\mathbf{w}_t$. Then careful calculations yield: $\nabla A_{it}(\mathbf{b})=-2\Phi(\mathbf{w}_t)'+\sum_{j=1}^8 B_{it,j}(\mathbf{b})\Phi(\mathbf{w}_t)'$, where

$$
B_{it,1}(\mathbf{b}) \quad = \quad -2\alpha_T g_{e,i}(\alpha_T-(\mathbf{b}_i-\mathbf{b})'\Phi(\mathbf{w}_t)-z_{it}),
$$

$$
\begin{aligned}
B_{it,2}(\mathbf{b}) &= -2\alpha_T g_{e,i}(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}), \\
B_{it,3}(\mathbf{b}) &= -2P_e((\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) + z_{it} + e_{it} > \alpha_T), \\
B_{it,4}(\mathbf{b}) &= 2((\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) + z_{it}) g_{e,i}(\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}), \\
B_{it,5}(\mathbf{b}) &= 2P_e(e_{it} < -\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}), \\
B_{it,6}(\mathbf{b}) &= -2((\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) + z_{it}) g_{e,i}(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}), \\
B_{it,7}(\mathbf{b}) &= 2[\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}] g_{e,i}(\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}), \\
B_{it,8}(\mathbf{b}) &= -2(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}) g_{e,i}(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{w}_t) - z_{it}).
\end{aligned}
$$

Since $\max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_i\| = o(m_T)$, $\max_{it} |z_{it}| = o_P(\alpha_T)$, $\Phi(\mathbf{w}_t)$ is sub-Gaussian and $J \log N \sqrt{\log T} = o(T)$, we have: with probability approaching one, for any $\epsilon > 0$,

$$
\max_{i,t} |(\mathbf{b}_i - \widetilde{\mathbf{b}}_i)'\Phi(\mathbf{w}_t)| + \max_{it} |z_{it}| < \epsilon\alpha_T.
$$

Hence with probability approaching one,

$$
\begin{aligned}
\max_i |\sum_{j\neq 3,5} B_{it,j}(\widetilde{\mathbf{b}}_i)| &\leq C\alpha_T \max_i \sup_{|x|<\epsilon\alpha_T} g_{e,i}(\pm\alpha_T + x) \leq C\alpha_T^{-(\zeta_1-1)}, \\
\max_i |B_{it,3}(\widetilde{\mathbf{b}}_i) + B_{it,5}(\widetilde{\mathbf{b}}_i)| &\leq C \max_i P(|e_{it}| > (1-\epsilon)\alpha_T) \leq C\alpha_T^{-(\zeta_2+2)}.
\end{aligned}
$$

Hence

$$
\|\nabla^2 \bar{Q}_i(\widetilde{\mathbf{b}}_i) - 2E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\| = \|\sum_{j=1}^{8} E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)' B_{it,j}(\widetilde{\mathbf{b}}_i)\| = O(J\alpha_T^{-(\zeta_1-1)} + J\alpha_T^{-(\zeta_2+2)}).
$$

Consequently, $\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha} = -(2E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')^{-1}\nabla Q_i(\mathbf{b}_{i,\alpha}) + \mathbf{R}_{i,b}$, where

$$
\begin{aligned}
\max_{i\leq N} \|\mathbf{R}_{i,b}\| &\leq \|(2E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')^{-1}\|(\|\nabla^2 \bar{Q}_i(\widetilde{\mathbf{b}}_i) - 2E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|\|\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}\| + \max_i \|\boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\| \\
&= O_P(\alpha_T^{-(\zeta_1-1)} + \alpha_T^{-(\zeta_2+2)} + \sqrt{\frac{\log J}{T}}) J m_T
\end{aligned}
$$

Proposition D.3. *Let $\widehat{E}(x_{it}|\mathbf{w}_t) = \widehat{\mathbf{b}}_i'\Phi(\mathbf{w}_t)$. Then for $\mathbf{A} = (2E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')^{-1}$,*

$$
\widehat{E}(x_{it}|\mathbf{w}_t) = E(x_{it}|\mathbf{w}_t) + \Phi(\mathbf{w}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^{T} \alpha_T \dot{\rho}(\alpha_T^{-1} e_{is})\Phi(\mathbf{w}_s) + R_{1,it} + R_{2,it} + R_{3,it},
$$

where

$$
\begin{aligned}
R_{1,it} &:= \Phi(\mathbf{w}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^{T}\alpha_T[\dot\rho(\alpha_T^{-1}e_{is,\alpha})-\dot\rho(\alpha_T^{-1}e_{is})]\Phi(\mathbf{w}_s)\\
R_{2,it} &:= \Phi(\mathbf{w}_t)'(\mathbf{R}_{i,b}+\mathbf{b}_{i,\alpha}-\mathbf{b}_i), \qquad R_{3,it}:=-z_{it}.
\end{aligned}
$$

Write $R_{it}:=R_{1,it}+R_{2,it}+R_{3,it}$, *then*

$$
\begin{aligned}
&\max_i \frac{1}{T}\sum_{t=1}^{T}R_{it}^2 = O_P(J^{1-2\eta}+\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T}+\frac{J^3\log N\log J}{T^2}),\\
&\max_i \frac{1}{T}\sum_{t=1}^{T}|\widehat{E}(x_{it}|\mathbf{w}_t)-E(x_{it}|\mathbf{w}_t)|^2 = O_P(\frac{J\log N}{T}+J^{-2\eta}).
\end{aligned}
$$

Proof. By Lemma D.3 and Proposition D.3,

$$
\begin{aligned}
&\widehat{E}(x_{it}|\mathbf{w}_t) = E(x_{it}|\mathbf{w}_t)+\Phi(\mathbf{w}_t)'(\widehat{\mathbf{b}}_i-\mathbf{b}_{i,\alpha})+\Phi(\mathbf{w}_t)'(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)-z_{it}\\
=\ & E(x_{it}|\mathbf{w}_t)+\Phi(\mathbf{w}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^{T}\alpha_T\dot\rho(\alpha_T^{-1}e_{is,\alpha})\Phi(\mathbf{w}_s)+\Phi(\mathbf{w}_t)'(\mathbf{R}_{i,b}+\mathbf{b}_{i,\alpha}-\mathbf{b}_i)-z_{it}\\
=\ & E(x_{it}|\mathbf{w}_t)+\Phi(\mathbf{w}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^{T}\alpha_T\dot\rho(\alpha_T^{-1}e_{is})\Phi(\mathbf{w}_s)+R_{it}.
\end{aligned}
$$

On the other hand, uniformly in i, for $a=\lambda_{\max}(\frac{1}{T}\sum_{t=1}^{T}\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')$,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}R_{it}^2 \le aC\|\mathbf{A}\|^2\|\frac{1}{T}\sum_{s=1}^{T}\alpha_T[\dot\rho(\alpha_T^{-1}e_{is,\alpha})-\dot\rho(\alpha_T^{-1}e_{is})]\Phi(\mathbf{w}_s)\|^2\\
&+aC\|\mathbf{R}_{i,b}+\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|^2+C\frac{1}{T}\sum_t z_{it}^2\\
\le\ & C(\frac{1}{T}\sum_s|e_{is,\alpha}-e_{is}|\|\Phi(\mathbf{w}_s)\|)^2+C\|\mathbf{R}_{i,b}\|^2+C\|\mathbf{b}_{i,\alpha}-\mathbf{b}\|^2+C\frac{1}{T}\sum_t z_{it}^2\\
\le\ & C\frac{1}{T}\sum_s(|z_{it}|^2+\|\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|^2\|\Phi(\mathbf{w}_t)\|^2)\frac{1}{T}\sum_t\|\Phi(\mathbf{w}_t)\|^2+C\|\mathbf{R}_{i,b}\|^2+C\|\mathbf{b}_{i,\alpha}-\mathbf{b}\|^2+O_P(J^{-2\eta})\\
=\ & O_P(J)(J^{-2\eta}+J\alpha_T^{-2(k-1)})+O_P(\alpha_T^{-2(\zeta_1-1)}+\alpha_T^{-2(\zeta_2+2)}+\frac{\log J}{T})J^2m_T^2.
\end{aligned}
$$

Also note that $\alpha_T^{-2\zeta_2-4} = O(\log N/T)$. Finally,

$$
\begin{aligned}
& \max_i \frac{1}{T}\sum_{t=1}^T |\widehat{E}(x_{it}|\mathbf{w}_t) - E(x_{it}|\mathbf{w}_t)|^2 \leq \max_i \frac{1}{T}\sum_{t=1}^T |\Phi(\mathbf{w}_t)'(\widehat{\mathbf{b}}_i - \mathbf{b}_i)|^2 + \max_i \frac{1}{T}\sum_{t=1}^T z_{it}^2 \\
\leq\ & a\|\widehat{\mathbf{b}}_i - \mathbf{b}_i\|^2 + \max_i \frac{1}{T}\sum_{t=1}^T z_{it}^2 = O_P(\frac{J\log N}{T} + \alpha_T^{-2(k-1)} + J^{-2\eta}).
\end{aligned}
$$

The term involving $\alpha_T^{-2(k-1)}$ is negligible since it is smaller than $(\log N/T)^3$.

D.3 Estimating the loadings

Write $\mathbf{M}_\alpha$ be an $N \times J$ matrix, whose ith row is given by

$$
\mathbf{M}'_{i,\alpha} := \frac{1}{T}\sum_{s=1}^T \alpha_T \dot{\rho}(\alpha_T^{-1} e_{is})\Phi(\mathbf{w}_s)'.
$$

Write $\mathbf{R}_t = (R_{1t}, ..., R_{Nt})'$. Then the Bahadur representation in Proposition D.3 can be written in the vector form: $\mathbf{A} = (2E\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)')^{-1}$,

$$
\widehat{E}(\mathbf{x}_t|\mathbf{w}_t) = E(\mathbf{x}_t|\mathbf{w}_t) + \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t) + \mathbf{R}_t = \mathbf{\Lambda}E(\mathbf{f}_t|\mathbf{w}_t) + \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t) + \mathbf{R}_t. \tag{D.1}
$$

Let $\widetilde{\mathbf{V}}$ be a $K \times K$ diagonal matrix, whose diagonal elements are the first K eigenvalues of $\widehat{\mathbf{\Sigma}}/N := \frac{1}{TN}\sum_{t=1}^T \widehat{E}(\mathbf{x}_t|\mathbf{w}_t)\widehat{E}(\mathbf{x}_t|\mathbf{w}_t)'$. By step 2 of the proof of Theorem 2.1, all the eigenvalues of $\widetilde{\mathbf{V}}$ are bounded away from both zero and infinity with probability approaching one. By the definition of $\widehat{\mathbf{\Lambda}}$, $\frac{1}{N}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Lambda}} = \widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}$. Plugging in (D.1), we have,

$$
\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H} = \sum_{i=1}^8 \mathbf{B}_i, \qquad \mathbf{H} = \frac{1}{TN}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1} \tag{D.2}
$$

where

$$
\begin{aligned}
\mathbf{B}_1 &= \frac{1}{TN}\sum_{t=1}^T \mathbf{\Lambda}E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'\mathbf{A}\mathbf{M}'_\alpha\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}, \qquad \mathbf{B}_2 = \frac{1}{TN}\sum_{t=1}^T \mathbf{\Lambda}E(\mathbf{f}_t|\mathbf{w}_t)\mathbf{R}'_t\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}, \\
\mathbf{B}_3 &= \frac{1}{TN}\sum_{t=1}^T \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1} \qquad \mathbf{B}_4 = \frac{1}{TN}\sum_{t=1}^T \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\mathbf{A}\mathbf{M}'_\alpha\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1},
\end{aligned}
$$

$$\mathbf{B}_5 = \frac{1}{TN}\sum_{t=1}^T \mathbf{M}_\alpha \mathbf{A}\Phi(\mathbf{w}_t)\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}, \qquad \mathbf{B}_6 = \frac{1}{TN}\sum_{t=1}^T \mathbf{R}_t E(\mathbf{f}_t|\mathbf{w}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1},$$
$$\mathbf{B}_7 = \frac{1}{TN}\sum_{t=1}^T \mathbf{R}_t\Phi(\mathbf{w}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}, \qquad \mathbf{B}_8 = \frac{1}{TN}\sum_{t=1}^T \mathbf{R}_t\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}.$$

Lemma D.4. *Recall that* $\mathbf{V}$ *is a* $K \times K$ *diagonal matrix, whose diagonal elements are the eigenvalues of* $\mathbf{\Sigma}_\Lambda^{1/2}E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\mathbf{\Sigma}_\Lambda^{1/2}$*. Then* $\widetilde{\mathbf{V}} \to^P \mathbf{V}$ *and* $\|\widehat{\mathbf{V}} - \mathbf{V}_N\| = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}})$.

Proof. Let $\mathbf{V}_N$ be a $K \times K$ diagonal matrix, whose diagonal elements are the first K eigenvalues of $\mathbf{\Sigma}/N$. From step 4 of the proof of Theorem 2.1, the diagonal entries are also the eigenvalues of $\{E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\{E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\}^{1/2}$. Since $\mathbf{\Sigma}_{\Lambda,N} \to \mathbf{\Sigma}_\Lambda$, by Weyl's theorem (Lemma C.1),

$$\|\mathbf{V} - \mathbf{V}_N\| \leq \|\{E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\}^{1/2}(\mathbf{\Sigma}_{\Lambda,N} - \mathbf{\Sigma}_\Lambda)\{E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}\}^{1/2}\| = o(1).$$

On the other hand, by Proposition D.3,

$$\begin{aligned}
\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\infty &\leq \max_{ij}\frac{1}{T}\sum_t |\widehat{E}(x_{it}|\mathbf{w}_t)\widehat{E}(x_{jt}|\mathbf{w}_t) - E(x_{it}|\mathbf{w}_t)E(x_{jt}|\mathbf{w}_t)| \\
&\quad + \max_{ij}|\frac{1}{T}\sum_t E(x_{it}|\mathbf{w}_t)E(x_{jt}|\mathbf{w}_t) - E\{E(x_{it}|\mathbf{w}_t)E(x_{jt}|\mathbf{w}_t)\}| = o_P(1) \\
&\leq \max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_i\| O_P(\|\frac{1}{T}\sum_t \Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|) + O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}}) \\
&= O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}}).
\end{aligned}$$

By Lemma C.1, $\|\widetilde{\mathbf{V}} - \mathbf{V}_N\| \leq \frac{1}{N}\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \leq \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\infty = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}}) = o_P(1)$. The result then follows from the triangular inequality.

D.3.1 Proof of Theorem 3.1: $\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}$

Result (3.4) follows from the following proposition.

Proposition D.4.

$$\frac{1}{N}\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|_F^2 = O_P(J/T + J^{1-2\eta} + \alpha_T^{-2(\zeta_1 - 1)}\frac{J^3\log N}{T}).$$

Proof. Under the assumptions, $J\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 = O_P(1)$, and $\|\mathbf{M}_\alpha\|^2/N = O_P(1)$. Hence from Lemma D.5 and Proposition D.3, and $\alpha_T = C\sqrt{T/\log(NJ)}$

$$\begin{aligned}\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|_F^2 &= O_P(\sum_{i=1}^8 \|\mathbf{B}_i\|_F^2) = O_P(\|\mathbf{M}_\alpha\|^2 J\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2) \\ &\quad + O_P(\|\mathbf{M}_\alpha\|^2 + \|\mathbf{M}_\alpha\|_F^4/N) + O_P(N\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 + N(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^2) \\ &= O_P(\|\mathbf{M}_\alpha\|^2) + O_P(N\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2).\end{aligned}$$

The result then follows from Lemma D.5 and Proposition D.3. Note that the term that reflects the effect of $\alpha_T^{-(k-1)}$ is removed, since it is negligible when $k > 4$, which is guaranteed by our assumption that $\zeta_2 > 2$. Furthermore, by assumptions $J^2\log^3 N = O(T)$ and $\zeta_1 > 2$,

$$\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T} = (\frac{\log(JN)}{T})^{\zeta_1} J^3 = O(\frac{J}{T}).$$

Q.E.D.

D.3.2 Proof of Theorem 3.1: $\max_{i\le N}\|\boldsymbol{\lambda}_i - \mathbf{H}'\boldsymbol{\lambda}_i\|$

Note in Lemma D.5 below that terms $\mathbf{B}_1, \mathbf{B}_4$ and $\mathbf{B}_7$ have two upper bounds, where the second bound uses a simple inequality $\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2 \le \|\mathbf{M}_\alpha\|^2\|\widehat{\mathbf{\Lambda}}\|^2$. Such a simple inequality is crude, but is sufficient to prove Proposition D.4. On the other hand, given Proposition D.4, a sharper rate for $\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2$ can be found. As a result, the first bounds for $\mathbf{B}_1, \mathbf{B}_4$ and $\mathbf{B}_7$ are used later to achieve sharp rates for $\widehat{\mathbf{g}}(\mathbf{w}_t) - \mathbf{g}(\mathbf{w}_t)$.

Lemma D.5. *(i)* $\|\mathbf{M}_\alpha\|^2 = O_P(NJ/T + NJ^{1-2\eta})$,

(ii) $\|\mathbf{B}_1\|_F^2 = O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2/N) = O_P(\|\mathbf{M}_\alpha\|^2)$, $\quad\|\mathbf{B}_3\|_F^2 = O_P(\|\mathbf{M}_\alpha\|^2)$.

(iii) $\|\mathbf{B}_2\|_F^2 = O_P(N\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2) = \|\mathbf{B}_6\|_F^2$,

(iv) $\|\mathbf{B}_4\|_F^2 = O_P(\|\mathbf{M}_\alpha\|^2\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2/N^2) = O_P(\|\mathbf{M}_\alpha\|_F^4/N)$, $\qquad\|\mathbf{B}_8\|_F^2 = O_P(N(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^2)$,

(v) $\|\mathbf{B}_5\|_F^2 = O_P(\|\mathbf{M}_\alpha\|^2 J\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)$.

$\|\mathbf{B}_7\|_F^2 = O_P(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 J\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2/N) = O_P(\|\mathbf{M}_\alpha\|^2 J\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)$.

Proof. (i) Recall that $e_{it} = e_{it,\alpha} + \Delta_{it,\alpha}$, where $\Delta_{it,\alpha} := (\mathbf{b}_{i,\alpha} - \mathbf{b}_i)'\Phi(\mathbf{w}_t) - z_{it}$.

$$E\|\mathbf{M}_\alpha\|_F^2 = E\sum_{i=1}^N \|\mathbf{M}_{i,\alpha}\|^2 = \sum_{i=1}^N\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T \alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\phi_j(\mathbf{w}_s))^2$$

$$
\begin{aligned}
&\leq 2\sum_{i=1}^{N}\sum_{j=1}^{J}E(\frac{1}{T}\sum_{s=1}^{T}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is,\alpha})\phi_j(\mathbf{w}_s))^2+2\sum_{i=1}^{N}\sum_{j=1}^{J}E(\frac{1}{T}\sum_{s=1}^{T}2|e_{is}-e_{is,\alpha}||\phi_j(\mathbf{w}_s)|)^2\\
&\leq 2\sum_{i=1}^{N}\sum_{j=1}^{J}\frac{1}{T}\mathrm{var}(\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is,\alpha})\phi_j(\mathbf{w}_s))\\
&\quad +C\sum_{i=1}^{N}\sum_{j=1}^{J}E(\frac{1}{T}\sum_{s=1}^{T}|(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'\Phi(\mathbf{w}_s)\phi_j(\mathbf{w}_s)|)^2+C\sum_{i=1}^{N}\sum_{j=1}^{J}E(\frac{1}{T}\sum_{s=1}^{T}|z_{is}\phi_j(\mathbf{w}_s)|)^2\\
&\leq O(NJ/T+NJ^2\alpha_T^{-2(k-1)}+NJ^{1-2\eta}),
\end{aligned}
$$

where the first inequality is due to the triangular inequality and $|\dot{\rho}(t_1)-\dot{\rho}(t_2)|\leq 2|t_1-t_2|$; the second inequality is due to $E\dot{\rho}(\alpha_T^{-1}e_{is,\alpha})\Phi(\mathbf{w}_s)=0$ and that $e_{is}-e_{is,\alpha}=(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'\Phi(\mathbf{w}_s)-z_{is}$.

(ii) Since $\frac{1}{T}\sum_t\Phi(\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'-(\|\frac{1}{T}\sum_t E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)')'\frac{1}{T}\sum_t E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'$ is semipositive definite, we have

$$
\|\frac{1}{T}\sum_t E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|^2\leq\|\frac{1}{T}\sum_t\Phi(\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|=O(\|E\Phi(\mathbf{w}_t)\mathbf{f}_t'\mathbf{f}_t\Phi(\mathbf{w}_t)\|).
$$

$\|\mathbf{B}_1\|_F^2\leq\frac{1}{N^2}\|\mathbf{\Lambda}\|^2\|\frac{1}{T}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'\mathbf{A}\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2\|\widetilde{\mathbf{V}}^{-1}\|^2=O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2/N)=O_P(\|\mathbf{M}_\alpha\|^2)$.

The bound for $\|\mathbf{B}_3\|_F^2$ is similar.

(iii) By Proposition D.3, $\frac{1}{T}\sum_t\|\mathbf{R}_t\|^2\leq N\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2$. Hence

$$
\|\mathbf{B}_2\|_F^2\leq\|\mathbf{\Lambda}\|_F^2\|\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|_F^2\frac{1}{N^2}\frac{1}{T}\sum_t\|E(\mathbf{f}_t|\mathbf{w}_t)\|^2\frac{1}{T}\sum_t\|\mathbf{R}_t\|^2=O_P(N\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2).
$$

The bound for $\|\mathbf{B}_6\|_F^2$ is similar.

(iv) $\|\mathbf{B}_4\|_F^2$ is upper bounded by

$$
\frac{1}{N^2}\|\mathbf{M}_\alpha\mathbf{A}\|^2\|\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\mathbf{A}\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2\|\widetilde{\mathbf{V}}^{-1}\|^2=O_P(\|\mathbf{M}_\alpha\|^2\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2/N^2)=O_P(\|\mathbf{M}_\alpha\|_F^4/N).
$$

Also, $\|\mathbf{B}_8\|_F^2\leq\frac{1}{N^2}(\frac{1}{T}\sum_{t=1}^T\|\mathbf{R}_t\|^2)^2\|\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2=O_P(N(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)^2)$.

(v) $\mathbf{B}_5$ and $\mathbf{B}_7$ are bounded similarly. We have,

$$
\|\mathbf{B}_7\|_F^2\leq\frac{1}{N^2}\|\frac{1}{T}\sum_{t=1}^T\mathbf{R}_t\Phi(\mathbf{w}_t)'\mathbf{A}\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2\|\widetilde{\mathbf{V}}^{-1}\|^2=O_P(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2J\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2/N)
$$

$$= O_P(\|\mathbf{M}_\alpha\|^2 J \max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)$$

.

Q.E.D.

Given Proposition D.4, due to

$$\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|_F^2 \le 2\|\mathbf{M}_\alpha\|^2\|\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}\|_F^2 + 2\|\mathbf{M}_\alpha'\mathbf{\Lambda}\|^2\|\mathbf{H}\|_F^2,$$

the rate of convergence for $\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|_F^2$ can be improved, reaching a sharper bound than $\|\mathbf{M}_\alpha\|^2\|\widehat{\mathbf{\Lambda}}\|_F^2$. This is given in Lemma D.6 below. As a result, rates for $\mathbf{B}_1, \mathbf{B}_4, \mathbf{B}_7$ can be improved as well.

Lemma D.6. *Given Proposition D.4, we have*

$$\begin{aligned}
\|\mathbf{M}_\alpha'\mathbf{\Lambda}\|_F^2 &= O_P(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T + N^2J\alpha_T^{-\zeta_2}/T),\\
\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|_F^2 &= O_P(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T + N^2J\alpha_T^{-\zeta_2}/T + NJ^{1-2\eta}(J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T})).
\end{aligned}$$

Proof. The proof is a straightforward calculation as follows:

$$\begin{aligned}
&E\|\mathbf{M}_\alpha'\mathbf{\Lambda}\|_F^2 = E\|\sum_{i=1}^N \boldsymbol{\lambda}_i\mathbf{M}_{i,\alpha}'\|_F^2 = E\|\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \boldsymbol{\lambda}_i\alpha_T\dot\rho(\alpha_T^{-1}e_{is})\Phi(\mathbf{w}_s)'\|_F^2\\
=&\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}\alpha_T\dot\rho(\alpha_T^{-1}e_{is})\phi_j(\mathbf{w}_s))^2 = \sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N 2\lambda_{ik}e_{is}1\{|e_{is}|<\alpha_T\}\phi_j(\mathbf{w}_s))^2\\
&+\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}\alpha_T\dot\rho(\alpha_T^{-1}e_{is})1\{|e_{is}|\ge\alpha_T\}\phi_j(\mathbf{w}_s))^2\\
\le& 8\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}e_{is}\phi_j(\mathbf{w}_s))^2 + 8\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s))^2\\
&+\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N 2|\lambda_{ik}|\alpha_T1\{|e_{is}|\ge\alpha_T\}\phi_j(\mathbf{w}_s))^2\\
\le& 8\sum_{k=1}^K\sum_{j=1}^J \operatorname{var}(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}e_{is}\phi_j(\mathbf{w}_s)) + 12\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N |\lambda_{ik}e_{is}|1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s))^2.
\end{aligned}$$

To bound the first term, Let E_w be the conditional expectation given $\mathbf{w}_s$. We need to bound $\sum_{i,l\le N}|E_we_{is}e_{ls}|$. Note that $e_{is} = x_{is} - E(x_{is}|\mathbf{w}_s) = \boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s + u_{is}$. Since $E(\mathbf{u}_s|\mathbf{f}_s,\mathbf{w}_t) = 0$, we have

$$E(\boldsymbol{\gamma}_s\mathbf{u}_s'|\mathbf{w}_s) = E(\mathbf{f}_s\mathbf{u}_s'|\mathbf{w}_s) - (E\mathbf{f}_s|\mathbf{w}_s)E(\mathbf{u}_s'|\mathbf{w}_s) = E(\mathbf{f}_s\mathbf{u}_s'|\mathbf{w}_s) = E(\mathbf{f}_sE(\mathbf{u}_s'|\mathbf{w}_s,\mathbf{f}_s)|\mathbf{w}_s) = 0.$$

Hence $E_w(e_{is}e_{ls}) = E_w(\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s + u_{is})(\boldsymbol{\lambda}_l'\boldsymbol{\gamma}_s + u_{ls}) = \boldsymbol{\lambda}_i'\operatorname{cov}(\boldsymbol{\gamma}_s)\boldsymbol{\lambda}_l + E_w(u_{is}u_{ls})$. Therefore,

$$\begin{aligned}
& 8\sum_{k=1}^K\sum_{j=1}^J \operatorname{var}(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}e_{is}\phi_j(\mathbf{w}_s)) = 8\sum_{k=1}^K\sum_{j=1}^J\frac{1}{T}\operatorname{var}(\sum_{i=1}^N\lambda_{ik}e_{is}\phi_j(\mathbf{w}_s)) \\
= \ & 8\sum_{k=1}^K\sum_{j=1}^J\frac{1}{T}\sum_{i=1}^N\sum_{l=1}^N\lambda_{ik}\lambda_{lk}E\{E_w(e_{is}e_{ls})\phi_j(\mathbf{w}_s)^2\} \leq C\sum_{j=1}^J\frac{1}{T}E\phi_j(\mathbf{w}_s)^2\sup_{\mathbf{w}}\sum_{i=1}^N\sum_{l=1}^N|E_w(e_{is}e_{ls})| \\
\leq \ & \frac{CJ}{T}\sum_{i=1}^N\sum_{l=1}^N|\boldsymbol{\lambda}_i'\operatorname{cov}(\boldsymbol{\gamma}_s)\boldsymbol{\lambda}_l| + \frac{CJ}{T}\sup_{\mathbf{w}}\sum_{i=1}^N\sum_{l=1}^N|E_w(u_{is}u_{ls})| \\
\leq \ & \frac{CJ}{T}N^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\| + \frac{CJN}{T}\sup_{\mathbf{w}}\max_{i\leq N}\sum_{l=1}^N|E_w(u_{is}u_{ls})| = O(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T).
\end{aligned}$$

Note that the second term is bounded by

$$\begin{aligned}
\leq \ & C\sum_{k=1}^K\sum_{j=1}^J\frac{1}{T}\sum_{i=1}^N\sum_{l=1}^N E|e_{is}|1\{|e_{is}|>\alpha_T\}|e_{ls}|1\{|e_{ls}|>\alpha_T\}\phi_j(\mathbf{w}_s)^2 \\
& +C\sum_{k=1}^K\sum_{j=1}^J\frac{1}{T^2}\sum_{s=1}^T\sum_{i=1}^N\sum_{t\neq s}^T\sum_{l=1}^N E|e_{is}|1\{|e_{is}|>\alpha_T\}|\phi_j(\mathbf{w}_s)|E|e_{lt}|1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{w}_t)| \\
\leq \ & C\sum_{k=1}^K\sum_{j=1}^J\frac{1}{T}\sum_{i=1}^N\sum_{l=1}^N\sup_{\mathbf{w}}E_w|e_{is}|1\{|e_{is}|>\alpha_T\}|e_{ls}|1\{|e_{ls}|>\alpha_T\} \\
& +C\sum_{k=1}^K\sum_{j=1}^J\sum_{i=1}^N\sum_{l=1}^N(\sup_{\mathbf{w}}E|e_{is}|1\{|e_{is}|>\alpha_T\})^2 \\
\leq \ & C\frac{KJ}{T}N^2\max_i\sup_{\mathbf{w}}E_we_{is}^21\{|e_{is}|>\alpha_T\} + CKJN^2(\max_i\sup_{\mathbf{w}}E|e_{is}|1\{|e_{is}|>\alpha_T\})^2 \\
= \ & O(N^2J\alpha_T^{-\zeta_2}/T + N^2J\alpha_T^{-2(\zeta_2+1)}).
\end{aligned}$$

Hence $E\|\mathbf{M}_\alpha'\boldsymbol{\Lambda}\|_F^2 = O(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T + N^2J\alpha_T^{-\zeta_2}/T + N^2J\alpha_T^{-2(\zeta_2+1)})$. The rate for $\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|_F^2$ comes from

$$\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|_F^2 \leq 2\|\mathbf{M}_\alpha\|^2\|\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H}\|_F^2 + 2\|\mathbf{M}_\alpha'\boldsymbol{\Lambda}\|^2\|\mathbf{H}\|_F^2.$$

Lemma D.7. $\max_{i\leq N}\|\mathbf{M}_{i,\alpha}\| = O_P(J^{-\eta}\sqrt{J} + \sqrt{J(\log N)/T})$

Proof. First, it follows from the proof of Proposition D.2 that

$$\max_i \|\frac{1}{T}\sum_{s=1}^T \alpha_T \dot{\rho}(\alpha_T^{-1} e_{is,\alpha})\Phi(\mathbf{w}_s)\| = O_P(\sqrt{\frac{J\log N}{T}}).$$

Secondly, since $|\dot{\rho}(t_1) - \dot{\rho}(t_2)| \leq 2|t_1 - t_2|$,

$$\begin{aligned}
&\max_i \|\frac{1}{T}\sum_{s=1}^T \alpha_T(\dot{\rho}(\alpha_T^{-1} e_{is}) - \dot{\rho}(\alpha_T^{-1} e_{is,\alpha}))\Phi(\mathbf{w}_s)\| \leq \max_i \|\frac{1}{T}\sum_{s=1}^T 2|e_{is} - e_{is,\alpha}|\Phi(\mathbf{w}_s)\| \\
\leq\ & \max_i \|\frac{1}{T}\sum_{s=1}^T 2|(\mathbf{b}_{i,\alpha} - \mathbf{b}_i)'\Phi(\mathbf{w}_t) - z_{it}|\Phi(\mathbf{w}_s)\| \leq 2\max_i \|\mathbf{b}_{i,\alpha} - \mathbf{b}_i\| O_P(J) + O_P(J^{-\eta}\sqrt{J}) \\
=\ & O_P(J^{-\eta}\sqrt{J} + J\alpha_T^{-(k-1)}).
\end{aligned}$$

The result then follows from the triangular inequality.

Q.E.D.

Proof of Theorem 3.1 result (3.5): $\max_{i\leq N} \|\boldsymbol{\lambda}_i - \mathbf{H}'\boldsymbol{\lambda}_i\|$

Proof. Let $\mathbf{B}_{i1}, ..., \mathbf{B}_{i8}$ respectively denote the ith row of $\mathbf{B}_1, ..., \mathbf{B}_8$. We have

$$\begin{aligned}
\max_i \|\mathbf{B}_{i1}\| &\leq O_P(\|\mathbf{M}_\alpha \widehat{\mathbf{\Lambda}}\|/N) \leq O_P(\max_i \|\mathbf{M}_{i,\alpha}\|) \\
\max_i \|\mathbf{B}_{i2}\| &\leq O_P(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}) \\
\max_i \|\mathbf{B}_{i3}\| &\leq O_P(\max_i \|\mathbf{M}_{i,\alpha}\|) = O_P(J^{-\eta}\sqrt{J} + J\alpha_T^{-(k-1)} + \sqrt{J(\log N)/T}) \\
\max_i \|\mathbf{B}_{i4}\| &\leq O_P(\max_i \|\mathbf{M}_{i,\alpha}\|) O_P(\|\mathbf{M}_\alpha' \widehat{\mathbf{\Lambda}}\|/N) \\
\max_i \|\mathbf{B}_{i5}\| &\leq O_P(\max_i \|\mathbf{M}_{i,\alpha}\|) O_P(\sqrt{J}\max_i \frac{1}{T}\sum_t R_{it}^2)^{1/2} \\
\max_i \|\mathbf{B}_{i6}\| &\leq O_P(\max_i \frac{1}{T}\sum_t R_{it}^2)^{1/2}) \\
\max_i \|\mathbf{B}_{i7}\| &\leq O_P(\max_i \frac{1}{T}\sum_t R_{it}^2)^{1/2}\sqrt{J}) O_P(\|\mathbf{M}_\alpha \widehat{\mathbf{\Lambda}}\|/N) \\
\max_i \|\mathbf{B}_{i8}\| &\leq O_P(\max_i \frac{1}{T}\sum_t R_{it}^2).
\end{aligned}$$

Hence

$$\max_{i\leq N} \|\boldsymbol{\lambda}_i - \mathbf{H}'\boldsymbol{\lambda}_i\| \leq O_P(\max_i \|\mathbf{B}_{i2}\| + \max_i \|\mathbf{B}_{i3}\|)$$

$$= O_P(J^{-\eta}\sqrt{J} + \sqrt{J(\log N)/T} + \alpha_T^{-(\zeta_1-1)}\sqrt{\frac{J^3\log N}{T}}) = O_P(J^{-\eta}\sqrt{J} + \sqrt{J(\log N)/T}),$$

where the last equality follows from

$$\alpha_T^{-(\zeta_1-1)}\sqrt{\frac{J^3\log N}{T}} = (\frac{\log(NJ)}{T})^{\zeta_1/2}J^{3/2} = O(\sqrt{\frac{J\log N}{T}})$$

under assumptions $(\log N)^3 J^2 = O(T)$ and $\zeta_1 > 2$.

D.4 Proof of Theorem 3.2: $\frac{1}{T}\sum_{t=1}^T \|\widehat{\mathbf{g}}(\mathbf{w}_t) - \mathbf{H}^{-1}\mathbf{g}(\mathbf{w}_t)\|^2$

Recall that $\widehat{\mathbf{g}}(\mathbf{w}_t) = \frac{1}{N}\widehat{\mathbf{\Lambda}}'\widehat{E}(\mathbf{x}_t|\mathbf{w}_t)$. By (D.1), $\widehat{\mathbf{g}}(\mathbf{w}_t) - \mathbf{H}^{-1}\mathbf{g}(\mathbf{w}_t) = \sum_{i=1}^4 \mathbf{C}_{ti}$, where

$$\mathbf{C}_{t1} = \frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'(\mathbf{\Lambda}\mathbf{H} - \widehat{\mathbf{\Lambda}})\mathbf{H}^{-1}E(\mathbf{f}_t|\mathbf{w}_t), \qquad \mathbf{C}_{t2} = -\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\mathbf{H}^{-1}E(\mathbf{f}_t|\mathbf{w}_t)$$
$$\mathbf{C}_{t3} = \frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t), \qquad \mathbf{C}_{t4} = \frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{R}_t.$$

Lemma D.8. *Assume* $J = O(N)$,

$$\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F = O_P(\sqrt{\frac{J}{TN}} + \frac{\sqrt{J^3\log N\log J}}{T} + \sqrt{\frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|}{T}} + J^{1/2-\eta} + \alpha_T^{-(\zeta_1-1)}\sqrt{\frac{J^3\log N}{T}}),$$

and $\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F$ *has the same rate of convergence.*

Proof. $\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}) = \sum_{i=1}^8 \mathbf{\Lambda}'\mathbf{B}_i$. Keep in mind that $\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|$ and $\|\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\|$ have sharper bounds than $\|\mathbf{\Lambda}\|\|\mathbf{M}_\alpha\|$, $\|\widehat{\mathbf{\Lambda}}\|\|\mathbf{M}_\alpha\|$, given in Lemma D.6.

For $i \neq 3, 4, 5$, we simply use $\|\mathbf{\Lambda}'\mathbf{B}_i\| \leq \|\mathbf{\Lambda}\|\|\mathbf{B}_i\| = O(\sqrt{N})\|\mathbf{B}_i\|$ and Lemma D.5. But note that for $\mathbf{B}_1, \mathbf{B}_7$, the first upper bound in the lemma is used.

$$\|\mathbf{\Lambda}'\mathbf{B}_1\| = O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|) \qquad \|\mathbf{\Lambda}'\mathbf{B}_2\| = O_P((\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}N) = \|\mathbf{\Lambda}'\mathbf{B}_6\|$$
$$\|\mathbf{\Lambda}'\mathbf{B}_7\| = O_P((\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}J^{1/2}\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|), \qquad \|\mathbf{\Lambda}'\mathbf{B}_8\| = O_P(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 N).$$

As for $\mathbf{B}_3, \mathbf{B}_4, \mathbf{B}_5$, we have

$$\|\mathbf{\Lambda}'\mathbf{B}_3\| \leq O_P(1)\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\frac{1}{T}\sum_{t=1}^T \Phi(\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\| = O_P(\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|).$$

$$\begin{aligned}
\|\mathbf{\Lambda}'\mathbf{B}_4\| &\leq O_P(1)\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\frac{1}{TN}\sum_{t=1}^T\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\| = O_P(\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|/N)\\
\|\mathbf{\Lambda}'\mathbf{B}_5\| &\leq O_P(1)\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{w}_t)\mathbf{R}_t'\|/\sqrt{N} = O_P(\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|(J\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}).
\end{aligned}$$

Note that $\sqrt{J}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|_F \leq N$. Hence

$$\begin{aligned}
\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|_F &\leq \sqrt{K}\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\| = \frac{1}{N}O_P(\|\mathbf{\Lambda}'\mathbf{B}_1\|+\|\mathbf{\Lambda}'\mathbf{B}_2\|+\|\mathbf{\Lambda}'\mathbf{B}_3\|+\|\mathbf{\Lambda}'\mathbf{B}_6\|)\\
&= O_P(\frac{1}{N}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|+(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}).
\end{aligned}$$

In addition, we have

$$\begin{aligned}
\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|_F &\leq \|\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|_F+\frac{1}{N}\|\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}\|_F^2\\
&= O_P(\frac{1}{N}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|+\frac{1}{N}\|\mathbf{M}_\alpha\|^2+(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}),
\end{aligned}$$

which is straightforward to be verified to have the same rate as $\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|_F$.

Q.E.D.

Proof of Theorem 3.2: first result

The convergence of $\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{g}}(\mathbf{w}_t)-\mathbf{H}^{-1}\mathbf{g}(\mathbf{w}_t)\|^2$ in this theorem is proved in the following proposition.

Proposition D.5.

$$\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{g}}(\mathbf{w}_t)-\mathbf{H}^{-1}\mathbf{g}(\mathbf{w}_t)\|^2 = O_P(\frac{J}{TN}+\frac{J^3\log N\log J}{T^2}+\frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|}{T}+J^{1-2\eta}+\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T}).$$

Proof. By the proof of Proposition D.4 and Lemma D.8, $\frac{1}{N}\|\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}\|_F^2 = O_P(\frac{1}{N}\|\mathbf{M}_\alpha\|^2)+O_P(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)$

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T\|\mathbf{C}_{t1}\|^2 &= O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha\|^4+(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)^2)\\
\frac{1}{T}\sum_{t=1}^T\|\mathbf{C}_{t2}\|^2 &= O_P(\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|^2) = O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2+\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)
\end{aligned}$$

$$\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{C}_{t4}\|^2 \quad = \quad O_P(\max_i \frac{1}{T}\sum_{t=1}^{T} R_{it}^2).$$

Finally, let $\boldsymbol{\beta}_i$ denote the ith row of $\frac{1}{N}\widehat{\boldsymbol{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}$, $i \leq K$. Then

$$\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{C}_{t3}\|^2 = \frac{1}{T}\sum_{t=1}^{T}\|\frac{1}{N}\widehat{\boldsymbol{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t)\|^2 = \sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\beta}_i'\Phi(\mathbf{w}_t))^2$$
$$\leq \quad \sum_{i=1}^{K}\|\boldsymbol{\beta}_i\|^2\|\frac{1}{T}\sum_{t=1}^{T}\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\| = O_P(1)\|\frac{1}{N}\widehat{\boldsymbol{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\|_F^2 = O_P(\frac{1}{N^2}\|\widehat{\boldsymbol{\Lambda}}'\mathbf{M}_\alpha\|^2).$$

Hence

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{g}}(\mathbf{w}_t) - \mathbf{g}(\mathbf{w}_t)\|^2 = O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha\|^4 + \frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|^2 + \max_i \frac{1}{T}\sum_{t=1}^{T} R_{it}^2).$$

$$= O_P(\frac{J}{TN} + \frac{J^3\log N\log J}{T^2} + \frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|}{T} + J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T}).$$

Finally, due to $\zeta_1 \geq 2.5$ and $\log N^2 = O(T)$,

$$\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T} = O(\frac{J^3\log N\log J}{T^2}).$$

D.5 Proof of Theorem 3.2: $\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2$

Note that $\mathbf{x}_t - E(\mathbf{x}_t|\mathbf{w}_t) = \boldsymbol{\Lambda}\boldsymbol{\gamma}_t + \mathbf{u}_t$. and $\widehat{\boldsymbol{\gamma}}_t = \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'(\mathbf{x}_t - \widehat{E}(\mathbf{x}_t|\mathbf{w}_t))$. Hence from (D.1)

$$\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t = \frac{1}{N}\mathbf{H}'\boldsymbol{\Lambda}'\mathbf{u}_t + \sum_{i=1}^{d}\mathbf{D}_{ti}, \tag{D.3}$$

where for $\mathbf{C}_{t3}, \mathbf{C}_{t4}$ defined earlier,

$$\mathbf{D}_{t1} \quad = \quad \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'(\boldsymbol{\Lambda}\mathbf{H} - \widehat{\boldsymbol{\Lambda}})\mathbf{H}^{-1}\boldsymbol{\gamma}_t, \qquad \mathbf{D}_{t2} = \frac{1}{N}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H})'\mathbf{u}_t$$
$$\mathbf{D}_{t3} \quad = \quad \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t) = \mathbf{C}_{t3}, \qquad \mathbf{D}_{t4} = \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'\mathbf{R}_t = \mathbf{C}_{t4}.$$

Hence for a constant $C > 0$,

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2 \leq C\sum_{i=1}^{4}\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{ti}\|^2.$$

We look at terms on the right hand side one by one. First of all,

$$\begin{aligned} & E\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t' - \mathrm{cov}(\boldsymbol{\gamma}_t)\|_F^2 = \sum_{i=1}^{K}\sum_{j=1}^{K}\mathrm{var}(\frac{1}{T}\sum_{t=1}^{T}\gamma_{it}\gamma_{jt}) = \sum_{i=1}^{K}\sum_{j=1}^{K}\frac{1}{T}\mathrm{var}(\gamma_{it}\gamma_{jt}) \\ = & O(T^{-1})\max_{i,j\leq K}\mathrm{var}(\gamma_{it}\gamma_{jt}). \end{aligned}$$

As for $\mathbf{D}_{t1}$, let $\mathbf{G} = \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'(\boldsymbol{\Lambda}\mathbf{H} - \widehat{\boldsymbol{\Lambda}})\mathbf{H}^{-1}$ and let $\mathbf{G}_i'$ denote its ith row, $i \leq K$. Then

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t1}\|^2 &= \sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{G}_i'\boldsymbol{\gamma}_t)^2 = \sum_{i=1}^{K}\mathbf{G}_i'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t'\mathbf{G}_i \leq \|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t'\|\|\mathbf{G}\|_F^2 \\ &= \|\mathbf{G}\|_F^2(\|\mathrm{cov}(\boldsymbol{\gamma}_t)\| + O(\frac{1}{T})\max_{i,j\leq K}\mathrm{var}(\gamma_{it}\gamma_{jt})) \\ &= O_P(\|\mathrm{cov}(\boldsymbol{\gamma}_t)\| + \frac{1}{T}\max_{i,j\leq K}\mathrm{var}(\gamma_{it}\gamma_{jt}))O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|^2 + \frac{1}{N^2}\|\mathbf{M}_\alpha\|^4 + \max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2) \end{aligned}$$

Terms $\mathbf{D}_{t3}$ and $\mathbf{D}_{t4}$ were bounded earlier. Term $\mathbf{D}_{t2}$ is given in Lemma D.11 below. Also note that the rate for $\frac{1}{N^2}\|\mathbf{M}_\alpha\|^4$ is dominated by the rate of $\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|^2 + \max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2$. Thus by Proposition D.3, Lemmas D.6, D.9, D.10,

$$\begin{aligned} \sum_{i=1}^{4}\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{ti}\|^2 &= O_P(\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2 + \frac{1}{N^2T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}|\mathbf{u}_s'\mathbf{R}_t|^2 + \frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|^2 + \max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2) \\ &= O_P(\frac{J}{TN} + \frac{J^4\log N\log J}{T^2} + \frac{J\|\mathrm{cov}(\boldsymbol{\gamma}_s)\|}{T} + J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^4\log N}{T}). \end{aligned}$$

Finally, $\frac{1}{T}\sum_{t=1}^{T}\|\frac{1}{N}\mathbf{H}'\boldsymbol{\Lambda}'\mathbf{u}_t\|^2 = O_P(\frac{1}{TN^2}\sum_{t=1}^{T}E\|\boldsymbol{\Lambda}'\mathbf{u}_t\|^2) = O_P(\frac{1}{N})$. Hence

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2 = O_P(\frac{1}{N} + \frac{J^4\log N\log J}{T^2} + \frac{J\|\mathrm{cov}(\boldsymbol{\gamma}_s)\|}{T} + J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^4\log N}{T}).$$

Lemma D.9.

$$\sum_{t=1}^T \|\mathbf{u}_t'\mathbf{M}_\alpha\|^2 = O_P(JN\|\operatorname{cov}(\boldsymbol{\gamma}_s)\| + JN^2/T + J + JN^2\alpha_T^{-\zeta_2}).$$

Proof. Note that $E\sum_{t=1}^T \|\mathbf{u}_t'\mathbf{M}_\alpha\|^2 = \frac{1}{T^2}\sum_{t=1}^T\sum_{j=1}^J E(\sum_{i=1}^N\sum_{s=1}^T u_{it}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\phi_j(\mathbf{w}_s))^2$. We now bound the right hand side. In fact, since $e_{is} = \boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s + u_{is}$,

$$\begin{aligned}
& E(\sum_{i=1}^N\sum_{s=1}^T u_{it}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\phi_j(\mathbf{w}_s))^2 \\
\leq\ & 8E(\sum_{i=1}^N\sum_{s=1}^T u_{it}e_{is}1\{|e_{is}|<\alpha_T\}\phi_j(\mathbf{w}_s))^2 + 2E(\sum_{i=1}^N\sum_{s=1}^T u_{it}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})1\{|e_{is}|\geq\alpha_T\}\phi_j(\mathbf{w}_s))^2 \\
\leq\ & CE(\sum_{i=1}^N\sum_{s=1}^T u_{it}e_{is}\phi_j(\mathbf{w}_s))^2 + CE(\sum_{i=1}^N\sum_{s=1}^T |u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)|)^2 \\
\leq\ & CE(\sum_{i=1}^N\sum_{s=1}^T u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s\phi_j(\mathbf{w}_s))^2 + CE(\sum_{i=1}^N\sum_{s=1}^T (u_{it}u_{is}-E(u_{it}u_{is}))\phi_j(\mathbf{w}_s))^2 + \\
& +CE(\sum_{i=1}^N\sum_{s=1}^T (Eu_{it}u_{is})\phi_j(\mathbf{w}_s))^2 + CE(\sum_{i=1}^N\sum_{s=1}^T |u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)|)^2.
\end{aligned}$$

The first term is bounded as: uniformly in t,

$$\begin{aligned}
& E(\sum_{i=1}^N\sum_{s=1}^T u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s\phi_j(\mathbf{w}_s))^2 = \sum_{i=1}^N\sum_{s=1}^T\sum_{l=1}^N\sum_{k=1}^T \boldsymbol{\lambda}_i'E\boldsymbol{\gamma}_s\phi_j(\mathbf{w}_s)u_{lt}u_{it}\phi_j(\mathbf{w}_k)\boldsymbol{\gamma}_k'\boldsymbol{\lambda}_l \\
=\ & \sum_{i=1}^N\sum_{l=1}^N \boldsymbol{\lambda}_i'E\boldsymbol{\gamma}_t u_{lt}u_{it}\phi_j(\mathbf{w}_t)^2\boldsymbol{\gamma}_t'\boldsymbol{\lambda}_l + \sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N \boldsymbol{\lambda}_i'\operatorname{cov}(\boldsymbol{\gamma}_s)E\phi_j(\mathbf{w}_s)^2\boldsymbol{\lambda}_l Eu_{lt}u_{it} \\
\leq\ & \sum_{i=1}^N\sum_{l=1}^N E[|(Eu_{lt}u_{it}|\mathbf{w}_t,\mathbf{f}_t)|\phi_j(\mathbf{w}_t)^2\|\boldsymbol{\gamma}_t\|^2]\max_i\|\boldsymbol{\lambda}_i\|^2 + T\sum_{i=1}^N\sum_{l=1}^N \|\operatorname{cov}(\boldsymbol{\gamma}_s)\|E\phi_j(\mathbf{w}_s)^2|Eu_{lt}u_{it}|\max_i\|\boldsymbol{\lambda}_i\|^2 \\
\leq\ & NC\sup_{\mathbf{w},\mathbf{f}}\max_i\sum_{l=1}^N|(Eu_{lt}u_{it}|\mathbf{w}_t,\mathbf{f}_t)|\sup_{\mathbf{w}}E(\|\boldsymbol{\gamma}_t\|^2|\mathbf{w}_t=\mathbf{w})E\phi_j(\mathbf{w}_t)^2 + \|\operatorname{cov}(\boldsymbol{\gamma}_s)\|TNC\max_i\sum_{l=1}^N|Eu_{lt}u_{it}| \\
\leq\ & NC\sup_{\mathbf{w},\mathbf{f}}\max_i\sum_{l=1}^N|(Eu_{lt}u_{it}|\mathbf{w}_t,\mathbf{f}_t)|\|\operatorname{cov}(\boldsymbol{\gamma}_t)\| + \|\operatorname{cov}(\boldsymbol{\gamma}_s)\|TNC\max_i\sum_{l=1}^N|Eu_{lt}u_{it}| \\
=\ & O(TN\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|).
\end{aligned}$$

The second term: note that for some $v > 1$, $E\{Eu_{it}^4|\mathbf{w}_t\}^v < \infty$, uniformly in t,

$$
\begin{aligned}
& E(\sum_{i=1}^N \sum_{s=1}^T (u_{it}u_{is} - E(u_{it}u_{is}))\phi_j(\mathbf{w}_s))^2 \\
= & \sum_{i=1}^N \sum_{s=1}^T \sum_{l=1}^N \sum_{k=1}^T E(u_{it}u_{is} - E(u_{it}u_{is}))(u_{lt}u_{lk} - E(u_{lt}u_{lk}))\phi_j(\mathbf{w}_k)\phi_j(\mathbf{w}_s) \\
= & \sum_{i=1}^N \sum_{l=1}^N E(u_{it}^2 - Eu_{it}^2)(u_{lt}^2 - Eu_{lt}^2)\phi_j(\mathbf{w}_t)^2 + \sum_{i=1}^N \sum_{s\neq t} \sum_{l=1}^N Eu_{it}(u_{lt}^2 - Eu_{lt}^2)\phi_j(\mathbf{w}_t)Eu_{is}\phi_j(\mathbf{w}_s) \\
& + \sum_{i=1}^N \sum_{l=1}^N \sum_{k\neq t} E(u_{it}^2 - Eu_{it}^2)u_{lt}\phi_j(\mathbf{w}_t)E\phi_j(\mathbf{w}_k)u_{lk} + \sum_{i=1}^N \sum_{s\neq t} \sum_{l=1}^N Eu_{it}u_{lt}Eu_{ls}u_{is}\phi_j(\mathbf{w}_s)^2 \\
& + \sum_{i=1}^N \sum_{s\neq t} \sum_{l=1}^N \sum_{k\neq t,s} Eu_{it}u_{lt}Eu_{is}\phi_j(\mathbf{w}_s)E\phi_j(\mathbf{w}_k)u_{lk} \\
= & \sum_{i=1}^N \sum_{l=1}^N E(u_{it}^2 - Eu_{it}^2)(u_{lt}^2 - Eu_{lt}^2)\phi_j(\mathbf{w}_t)^2 + \sum_{i=1}^N \sum_{s\neq t} \sum_{l=1}^N Eu_{it}u_{lt}Eu_{ls}u_{is}\phi_j(\mathbf{w}_s)^2 \\
\leq & \sum_{i=1}^N \sum_{l=1}^N E(u_{it}^2 - Eu_{it}^2)(u_{lt}^2 - Eu_{lt}^2)\phi_j(\mathbf{w}_t)^2 + CT(\max_i \sum_{l=1}^N |Eu_{it}u_{lt}|)(\sup_{\mathbf{w}} \max_l \sum_{i=1}^N |Eu_{ls}u_{is}|\mathbf{w})|E\phi_j(\mathbf{w}_s)^2) \\
= & O(N^2 + T).
\end{aligned}
$$

The third term is bounded as: uniformly in t,

$$
E(\sum_{i=1}^N \sum_{s=1}^T (Eu_{it}u_{is})\phi_j(\mathbf{w}_s))^2 = E(\sum_{i=1}^N (Eu_{it}^2)\phi_j(\mathbf{w}_t))^2 = O(N^2).
$$

Finally, the fourth term is :

$$
\begin{aligned}
& E(\sum_{i=1}^N \sum_{s=1}^T |u_{it}e_{is}1\{|e_{is}| > \alpha_T\}\phi_j(\mathbf{w}_s)|)^2 \\
= & E\sum_{i=1}^N \sum_{s=1}^T \sum_{l=1}^N \sum_{k=1}^T |u_{it}e_{is}1\{|e_{is}| > \alpha_T\}\phi_j(\mathbf{w}_s)||u_{lt}e_{lk}1\{|e_{lk}| > \alpha_T\}\phi_j(\mathbf{w}_k)| \\
= & \sum_{i=1}^N \sum_{s\neq t} \sum_{l=1}^N E|u_{it}u_{lt}e_{lt}1\{|e_{lt}| > \alpha_T\}\phi_j(\mathbf{w}_t)|E|e_{is}1\{|e_{is}| > \alpha_T\}\phi_j(\mathbf{w}_s)| \\
& + \sum_{i=1}^N \sum_{l=1}^N E|u_{it}e_{it}1\{|e_{it}| > \alpha_T\}u_{lt}e_{lt}1\{|e_{lt}| > \alpha_T\}|\phi_j(\mathbf{w}_t)^2
\end{aligned}
$$

$$+E\sum_{i=1}^{N}\sum_{l=1}^{N}\sum_{k\neq t}|u_{it}u_{lt}e_{it}1\{|e_{it}|>\alpha_T\}\phi_j(\mathbf{w}_t)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{w}_k)|$$
$$+\sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}E|u_{it}u_{lt}|E|e_{ls}1\{|e_{ls}|>\alpha_T\}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)^2|$$
$$+\sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}\sum_{k\neq s,t}E|u_{it}u_{lt}|E|e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{w}_k)|$$
$$:=\sum_{i=1}^{5}a_i.$$

We look at $a_i, i=1,...,5$ respectively. By Holder's inequality, and the assumption that $E\{E(u_{it}^4|\mathbf{w})\}^v<\infty$, and by repeatedly using Cauchy-Schwarz inequality, uniformly in t,

$$\begin{aligned}
a_1 &= \sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}E|u_{it}u_{lt}e_{lt}1\{|e_{lt}|>\alpha_T\}\phi_j(\mathbf{w}_t)|E|e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)| \\
&\leq \sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}(Ee_{lt}^2 1\{|e_{lt}|>\alpha_T\})^{1/2}(Eu_{it}^2u_{lt}^2\phi_j(\mathbf{w}_t)^2)^{1/2}\sup_{\mathbf{w}}E(|e_{is}|1\{|e_{is}|>\alpha_T\}|\mathbf{w})E|\phi_j(\mathbf{w}_s)| \\
&\leq CTN^2\max_i\{E[Eu_{it}^4|\mathbf{w}_t]^v\}^{1/(2v)}\alpha_T^{-(\zeta_2+1)-\zeta_2/2}=O(TN^2\alpha_T^{-(\zeta_2+1)-\zeta_2/2}) \\
a_2 &= \sum_{i=1}^{N}\sum_{l=1}^{N}E|u_{it}e_{it}1\{|e_{it}|>\alpha_T\}u_{lt}e_{lt}1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{w}_t)^2 \\
&\leq \sum_{i=1}^{N}\sum_{l=1}^{N}E|u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t1\{|e_{it}|>\alpha_T\}u_{lt}\boldsymbol{\lambda}_l'\boldsymbol{\gamma}_t1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{w}_t)^2 \\
&\quad+\sum_{i=1}^{N}\sum_{l=1}^{N}E|u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t1\{|e_{it}|>\alpha_T\}u_{lt}^2 1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{w}_t)^2 \\
&\quad+\sum_{i=1}^{N}\sum_{l=1}^{N}E|u_{it}^2 1\{|e_{it}|>\alpha_T\}u_{lt}\boldsymbol{\lambda}_l'\boldsymbol{\gamma}_t1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{w}_t)^2 \\
&\quad+\sum_{i=1}^{N}\sum_{l=1}^{N}E|u_{it}^2 1\{|e_{it}|>\alpha_T\}u_{lt}^2 1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{w}_t)^2 \\
&\leq C\sum_{i=1}^{N}\sum_{l=1}^{N}\max_i(Eu_{it}^4)^{1/2}(E\{E\|\boldsymbol{\gamma}_t\|^4|\mathbf{w}_t\}^v)^{1/(2v)}+C\sum_{i=1}^{N}\sum_{l=1}^{N}[E(u_{it}u_{lt}^2)^{4/3}]^{3/4}(E\|\boldsymbol{\gamma}_t\|^4\phi_j(\mathbf{w}_t)^8)^{1/4} \\
&\quad+C\sum_{i=1}^{N}\sum_{l=1}^{N}\{E[E(u_{it}^2u_{lt}^2|\mathbf{w}_t)]^v\}^{1/v}=O(N^2)
\end{aligned}$$

$$
\begin{aligned}
a_3 &= E\sum_{i=1}^N\sum_{l=1}^N\sum_{k\neq t}|u_{it}u_{lt}e_{it}1\{|e_{it}|>\alpha_T\}\phi_j(\mathbf{w}_t)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{w}_k)| \\
&\leq \sum_{i=1}^N\sum_{l=1}^N\sum_{k\neq t}(E|u_{it}u_{lt}\phi_j(\mathbf{w}_t)|^2)^{1/2}(Ee_{it}^2 1\{|e_{it}|>\alpha_T\})^{1/2}E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{w}_k)| \\
&\leq TC\sum_{i=1}^N\sum_{l=1}^N\{E[E(u_{it}^2u_{lt}^2|\mathbf{w}_t)]^v\}^{1/2v}\alpha_T^{-\zeta_2/2-(\zeta_2+1)} = O(N^2T\alpha_T^{-\zeta_2/2-(\zeta_2+1)}) \\
a_4 &= \sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N E|u_{it}u_{lt}|E|e_{ls}1\{|e_{ls}|>\alpha_T\}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)^2| = O(TN^2\alpha_T^{-\zeta_2}) \\
a_5 &= \sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N\sum_{k\neq s,t}E|u_{it}u_{lt}|E|e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{w}_k)| \\
&= O(N^2T^2\alpha_T^{-2(\zeta_2+1)}).
\end{aligned}
$$

Therefore, uniformly in $t\leq T$,

$$
E(\sum_{i=1}^N\sum_{s=1}^T|u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{w}_s)|)^2 = O(TN^2\alpha_T^{-(\zeta_2+1)-\zeta_2/2}+N^2+TN^2\alpha_T^{-\zeta_2}+N^2T^2\alpha_T^{-2(\zeta_2+1)}).
$$

Consequently, (note that $JN^2\alpha_T^{-\zeta_2}+JN^2T\alpha_T^{-2(\zeta_2+1)}\geq JN^2\sqrt{T}\alpha_T^{-(\zeta_2+1)-\zeta_2/2}$)

$$
E\sum_{t=1}^T\|\mathbf{u}_t'\mathbf{M}_\alpha\|^2 = O(JN\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|+JN^2/T+J+JN^2\alpha_T^{-\zeta_2}+JN^2T\alpha_T^{-2(\zeta_2+1)}).
$$

Lemma D.10.

$$
\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2 \quad = \quad O_P(N^2J^4\log N\log J+J^{2-2\eta}T^2N+T\alpha_T^{-2(\zeta_1-1)}N^2J^4\log N).
$$

Proof. Recall that $R_{it}=R_{1,it}+R_{2,it}+R_{3,it}$, where

$$
\begin{aligned}
R_{1,it} &:= \frac{1}{T}\sum_{k=1}^T\alpha_T[\dot\rho(\alpha_T^{-1}e_{ik,\alpha})-\dot\rho(\alpha_T^{-1}e_{ik})]\Phi(\mathbf{w}_k)'\mathbf{A}\Phi(\mathbf{w}_t) \\
R_{2,it} &:= \Phi(\mathbf{w}_t)'(\mathbf{R}_{i,b}+\mathbf{b}_{i,\alpha}-\mathbf{b}_i), \qquad R_{3,it}:=-z_{it}.
\end{aligned}
$$

In addition, recall $e_{it}=e_{it,\alpha}+\Delta_{it,\alpha}$, where $\Delta_{it,\alpha}=(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'\Phi(\mathbf{w}_t)-z_{it}$. For notational simplicity, we also write $H_{kt} := \Phi(\mathbf{w}_k)'\mathbf{A}\Phi(\mathbf{w}_t)$.

$$\sum_{s=1}^{T}\sum_{t=1}^{T}|\mathbf{u}_s'\mathbf{R}_t|^2 \leq C\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N}u_{is}R_{1,it})^2 + C\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N}u_{is}R_{2,it})^2 + C\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N}u_{is}R_{3,it})^2.$$

We look at these terms respectively.

bounding the first term

$$\begin{aligned}
&\sum_{s=1}^{T}\sum_{t=1}^{T}E(\sum_{i=1}^{N}u_{is}R_{1,it})^2 = \sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}u_{is}\frac{1}{T}\sum_{k=1}^{T}\alpha_T[\dot\rho(\alpha_T^{-1}e_{ik,\alpha}) - \dot\rho(\alpha_T^{-1}e_{ik})]\Phi(\mathbf{w}_k)'\mathbf{A}\Phi(\mathbf{w}_t)\}^2\\
\leq\ & C\sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}u_{is}\frac{1}{T}\sum_{k=1}^{T}\Delta_{ik,\alpha}H_{kt}\}^2\\
&+C\sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}|u_{is}|\frac{1}{T}\sum_{k=1}^{T}|\Delta_{ik,\alpha}|1\{|e_{ik}|>\alpha_T \text{ or } |e_{it,\alpha}|>\alpha_T\}|H_{kt}|\}^2\\
:=\ & Ca_1 + Ca_2.
\end{aligned}$$

For notational simplicity, let $I_{i,kt} := 1\{|e_{ik}| > \alpha_T \text{ or } |e_{it,\alpha}| > \alpha_T\}$.

$$\begin{aligned}
a_1 &= \sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}u_{is}\frac{1}{T}\sum_{k=1}^{T}\Delta_{ik,\alpha}H_{kt}\}^2\\
&= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{T}\sum_{k=1}^{T}E(Eu_{is}u_{js}|\{\mathbf{w}_l\}_{l\leq T})\Delta_{ik,\alpha}H_{kt}\Delta_{jm,\alpha}H_{mt}\\
&\leq \sup_{\mathbf{w}}\sum_{i=1}^{N}|E(u_{is}u_{js}|\mathbf{w}_s)|\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{m=1}^{T}\sum_{k=1}^{T}E\max_i|\Delta_{ik,\alpha}||H_{kt}\Delta_{jm,\alpha}H_{mt}|\\
&\leq CT^2N(\alpha_T^{-(k-1)}\sqrt{J}+J^{-\eta})^2J^2.\\
a_2 &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}\sum_{k=1}^{T}|u_{is}\Delta_{ik,\alpha}I_{i,kt}H_{kt}|\}^2 = \frac{1}{T^2}\sum_{s,t,k,l\leq T}\sum_{i=1}^{N}\sum_{j=1}^{N}E|u_{is}u_{js}\Delta_{ik,\alpha}H_{kt}\Delta_{jl,\alpha}I_{i,kt}I_{j,lt}H_{lt}|\\
&\leq \frac{1}{T^2}\sum_{t,k,l\leq T}\sum_{s=t \text{ or } k \text{ or } l}\sum_{i=1}^{N}\sum_{j=1}^{N}(E(u_{is}u_{js}\Delta_{ik,\alpha}H_{kt}\Delta_{jl,\alpha}H_{lt})^2)^{1/2}(EI_{i,kt}I_{j,lt})^{1/2}\\
&\quad+\frac{1}{T^2}\sum_{t,k,l\leq T}\sum_{s\neq t,k,l}\sum_{i=1}^{N}\sum_{j=1}^{N}E|u_{is}u_{js}|(E(\Delta_{ik,\alpha}H_{kt}H_{lt}\Delta_{jl,\alpha})^2)^{1/2}(EI_{i,kt}I_{j,lt})^{1/2}\\
&\leq \frac{CN(N+T)}{T^2}(\alpha_T^{-(k-1)}\sqrt{J}+J^{-\eta})^2J^2\sum_{t,k,l\leq T}(EI_{i,kt}I_{j,lt})^{1/2}
\end{aligned}$$

$$\leq \; CJ^2NT(N+T)(\alpha_T^{-(k-1)}\sqrt{J}+J^{-\eta})^2\alpha_T^{-(\zeta_2+2)/2}$$

where the last inequality is due to, uniformly in i, j,

$$\begin{aligned} P(|e_{it,\alpha}|>\alpha_T) &\leq P(|e_{it}|>3\alpha_T/4)+P(\|\Phi(\mathbf{w}_t)\|>C\alpha_T^k)\leq C\alpha_T^{-(\zeta_2+2)}, \\ \sum_{t,k,l\leq T}(EI_{i,kt}I_{j,lt})^{1/2} &\leq CT^3\alpha_T^{-(\zeta_2+2)/2}. \end{aligned}$$

Therefore, $\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}R_{1,it})^2 = O_P((\alpha_T^{-(k-1)}\sqrt{J}+J^{-\eta})^2J^2TN(T+N\alpha_T^{-(\zeta_2+2)/2}))$.

bounding the second term

By Lemma D.3, $\max_{i\leq N}\|\mathbf{R}_{i,b}\|^2 = O_P(\alpha_T^{-2(\zeta_1-1)}+\alpha_T^{-2(\zeta_2+2)}+\frac{\log J}{T})\frac{J^3\log N}{T}$. Hence

$$\begin{aligned} & \sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}R_{2,it})^2 = \sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}\Phi(\mathbf{w}_t)'(\mathbf{R}_{i,b}+\mathbf{b}_{i,\alpha}-\mathbf{b}_i))^2 \\ \leq \; & 2\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}\Phi(\mathbf{w}_t)'\mathbf{R}_{i,b})^2+2\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}\Phi(\mathbf{w}_t)'(\mathbf{b}_{i,\alpha}-\mathbf{b}_i))^2 := a_1+a_2, \text{ say} \\ a_1 \leq \; & 2\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N \|u_{is}\Phi(\mathbf{w}_t)\|)^2\max_i\|\mathbf{R}_{i,b}\|^2 = O_P(T^2N^2J)\max_i\|\mathbf{R}_{i,b}\|^2 \\ = \; & O_P((T\alpha_T^{-2(\zeta_1-1)}+T\alpha_T^{-2(\zeta_2+2)}+\log J)N^2J^4\log N). \\ E|a_2| = \; & 2\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'Eu_{is}u_{js}\sum_{t=1}^T\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'(\mathbf{b}_{j,\alpha}-\mathbf{b}_j) \\ \leq \; & 2\sup_{\mathbf{w}}\max_i\sum_{j=1}^N|(Eu_{is}u_{js}|\mathbf{w}_s)|\max_i\|\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|^2\sum_{s=1}^T\sum_{i=1}^N E\|\sum_{t=1}^T\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)\| \\ \leq \; & O(T^2\max_i\|\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|^2N) = O(T^2N\alpha_T^{-2(k-1)}). \end{aligned}$$

Therefore, $\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}R_{2,it})^2 = O_P((T\alpha_T^{-2(\zeta_1-1)}+T\alpha_T^{-2(\zeta_2+2)}+\log J)N^2J^4\log N + T^2N\alpha_T^{-2(k-1)})$.

bounding the third term

$$E\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}R_{3,it})^2 = \sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}z_{it})^2 = \sum_{s=1}^T\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N Eu_{is}u_{js}z_{it}z_{jt} = O(NT^2J^{-2\eta}).$$

Hence the result follows.

Lemma D.11.

$$\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t2}\|^2 = O_P(\frac{1}{N^3}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2+\frac{1}{N}\max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2+\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2+\frac{1}{N^2T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}|\mathbf{u}_s'\mathbf{R}_t|^2).$$

Proof. First of all, note that $\max_i\sum_j|Eu_{is}u_{js}|<\infty$, hence

$$E\frac{1}{T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{\Lambda}\|^2 = \sum_{j=1}^{K}E(\mathbf{u}_s'\boldsymbol{\lambda}_j)^2=O(N).$$

In addition, $\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t2}\|^2=\frac{1}{T}\sum_{t=1}^{T}\|\frac{1}{N}(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})'\mathbf{u}_t\|^2\le C\sum_{i=1}^{8}\frac{1}{N^2T}\sum_{t=1}^{T}\|\mathbf{u}_t'\mathbf{B}_i\|^2$.

$$\begin{aligned}
\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{B}_1\|^2 &= \frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{\Lambda}\frac{1}{TN}\sum_{t=1}^{T}E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2\\
&\le \frac{1}{N^4T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{\Lambda}\|^2\|\frac{1}{T}\sum_{t=1}^{T}E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2O_P(1)\\
&= O_P(1)\frac{1}{N^4T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{\Lambda}\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2=O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2/N^3),\\
\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{B}_2\|^2 &= \frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{\Lambda}\frac{1}{TN}\sum_{t=1}^{T}E(\mathbf{f}_t|\mathbf{w}_t)\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2\\
&\le \frac{1}{N^3T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{\Lambda}\|^2\|\frac{1}{T}\sum_{t=1}^{T}E(\mathbf{f}_t|\mathbf{w}_t)\mathbf{R}_t'\|^2O_P(1)=O_P(\frac{1}{N}\max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2),\\
\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{B}_3\|^2 &= \frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\mathbf{A}\frac{1}{TN}\sum_{t=1}^{T}\Phi(\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2\\
&\le O_P(1)\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\|\frac{1}{T}\sum_{t=1}^{T}\Phi(\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\|^2=\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2O_P(1),\\
\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{B}_4\|^2 &= \frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\mathbf{A}\frac{1}{TN}\sum_{t=1}^{T}\Phi(\mathbf{w}_t)\Phi(\mathbf{w}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2\\
&\le \frac{1}{N^4T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2O_P(1)\\
\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{B}_5\|^2 &= \frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\mathbf{A}\frac{1}{TN}\sum_{t=1}^{T}\Phi(\mathbf{w}_t)\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2
\end{aligned}$$

$$
\begin{aligned}
&\leq O_P(\frac{1}{N^3T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\|\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{w}_t)\mathbf{R}_t'\|^2) = O_P(\frac{J}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2) \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_6\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\frac{1}{TN}\sum_{t=1}^T\mathbf{u}_s'\mathbf{R}_tE(\mathbf{f}_t|\mathbf{w}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \\
&\leq \leq O_P(\frac{1}{N^2T^3}\sum_{s=1}^T\|\sum_{t=1}^T\mathbf{u}_s'\mathbf{R}_tE(\mathbf{f}_t|\mathbf{w}_t)'\|^2) \leq O_P(\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2) \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_7\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\frac{1}{TN}\sum_{t=1}^T\mathbf{u}_s'\mathbf{R}_t\Phi(\mathbf{w}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \leq O_P(\frac{J}{N^4T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2) \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_8\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\frac{1}{TN}\sum_{t=1}^T\mathbf{R}_t\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \leq O_P(1)\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2.
\end{aligned}
$$

Summarizing, we have

$$
\frac{1}{T}\sum_{t=1}^T\|\mathbf{D}_{t2}\|^2 = O_P(\frac{1}{N^3}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 + \frac{1}{N}\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2 + \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2 + \frac{1}{N^2T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2).
$$

D.6 Proof of Theorem 3.3: Limiting distribution for $\widehat{\boldsymbol{\gamma}}_t$

Recall (D.3), we shall examine $\mathbf{D}_{t1}, ..., \mathbf{D}_{t4}$ for each fixed t. As $\mathbf{D}_{t1}$ depends on $\frac{1}{N}\widehat{\mathbf{\Lambda}}'(\mathbf{\Lambda}\mathbf{H}-\widehat{\mathbf{\Lambda}})$, whose rate of convergence has a leading term $\sqrt{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|/T}$ as in Lemma D.8, therefore, we need to re-investigate the expansion of $\frac{1}{N}\widehat{\mathbf{\Lambda}}'(\mathbf{\Lambda}\mathbf{H}-\widehat{\mathbf{\Lambda}})$ in the proof of Lemma D.8. Similarly, we also need to bound $\frac{1}{N}\mathbf{u}_t'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})$ for a fixed t. For these purposes, we need to bound $\frac{1}{N}\|\mathbf{\Lambda}'\mathbf{B}_i\|$ and $\|\frac{1}{N}\mathbf{u}_t'\mathbf{B}_i\|$.

In fact, the rates for $\frac{1}{N}\|\mathbf{\Lambda}'\mathbf{B}_i\|, i\neq 1,3$ and $\frac{1}{N}\|\mathbf{u}_t'\mathbf{B}_i\|, i\neq 3$ are sufficient for the purpose here, which are found in the proof of Lemmas D.8 and D.11. However, $\frac{1}{N}\|\mathbf{\Lambda}'\mathbf{B}_i\|, i=1,3$ and $\frac{1}{N}\|\mathbf{u}_t'\mathbf{B}_3\|$ should be treated separately. Therefore we need to re-examine the expansion of $\frac{1}{N}\mathbf{M}_\alpha'\mathbf{\Lambda}$ and $\frac{1}{N}\mathbf{u}_t'\mathbf{M}_\alpha$. This is done in Lemma D.12 below.

In the lemmas below, for two matrices $\mathbf{A},\mathbf{B}$, we say $\mathbf{A}=\mathbf{B}+O_P(a_T)$ if $\|\mathbf{A}-\mathbf{B}\|=O_P(a_T)$.

Lemma D.12. *Write $\mathbf{\Sigma}_{\Lambda,N}=\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Lambda}$, $\mathbf{G}_T=\frac{1}{T}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{w}_t)\Phi(\mathbf{w}_t)'$.*

(i) Write $\xi_{NT}=\sqrt{\frac{J}{NT}}+\sqrt{\frac{J}{\alpha_T^{\zeta_2}T}}+\sqrt{\frac{J}{\alpha_T^{\zeta_2+1}}}$, and $\tilde{\xi}_{NT}=\xi_{NT}+\frac{J}{T}+\frac{1}{J^{2\eta-1}}+(\frac{\log N}{T})^{\zeta_1}J^3$.

$$
\frac{1}{N}\mathbf{\Lambda}'\mathbf{M}_\alpha = 2\mathbf{\Sigma}_{\Lambda,N}\frac{1}{T}\sum_{s=1}^T\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)' + O_P(\xi_{NT}),
$$

$$\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha = 2\mathbf{H}'\mathbf{\Sigma}_{\Lambda,N}\frac{1}{T}\sum_{s=1}^{T}\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)' + O_P(\tilde{\xi}_{NT}),$$
$$\|\frac{1}{N}\mathbf{u}_t'\mathbf{M}_\alpha\| = O_P(\tilde{\xi}_{NT}) \text{ for each fixed } t,$$
$$\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{R}_t\| = O_P(J^{1-\eta} + J^2\alpha_T^{-\zeta_1} + \frac{J^2\sqrt{\log J\log N}}{T}).$$

(ii) Let $\mathbf{M}_1 = \mathbf{\Sigma}_{\Lambda,N}\mathbf{G}_T\mathbf{A}$, $\mathbf{M}_2 = \mathbf{\Sigma}_{\Lambda,N}\mathbf{H}\widetilde{\mathbf{V}}^{-1}$, *and* $\mathbf{M}_3 = \mathbf{A}\mathbf{G}_T'\mathbf{\Sigma}_{\Lambda,N}\mathbf{H}\widetilde{\mathbf{V}}^{-1}$. *Then*

$$\frac{1}{N}\mathbf{\Lambda}'\mathbf{B}_1 = 2\mathbf{M}_1\frac{1}{T}\sum_{s=1}^{T}\Phi(\mathbf{w}_s)\boldsymbol{\gamma}_s'\mathbf{M}_2 + O_P(\tilde{\xi}_{NT}),$$
$$\frac{1}{N}\mathbf{\Lambda}'\mathbf{B}_3 = 2\mathbf{\Sigma}_{\Lambda,N}\frac{1}{T}\sum_{s=1}^{T}\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{M}_3 + O_P(\tilde{\xi}_{NT}).$$

(iii) For each fixed t,

$$\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}) = 2\mathbf{M}_1\frac{1}{T}\sum_{s=1}^{T}\Phi(\mathbf{w}_s)\boldsymbol{\gamma}_s'\mathbf{M}_2 + 2\mathbf{\Sigma}_{\Lambda,N}\frac{1}{T}\sum_{s=1}^{T}\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{M}_3$$
$$+O_P(\tilde{\xi}_{NT} + J^{1/2-\eta} + \alpha_T^{-\zeta_1}J^{3/2} + \frac{\sqrt{J^3\log N\log J}}{T}).$$
$$\|\frac{1}{N}\mathbf{u}_t'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| = O_P(\tilde{\xi}_{NT} + J^{1-\eta}\frac{1}{\sqrt{N}} + \frac{\sqrt{\log N\log J}}{T}J^2).$$

Proof. (i) Note that

$$\frac{1}{N}\mathbf{\Lambda}\mathbf{M}_\alpha = \frac{1}{TN}\sum_{s=1}^{T}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\Phi(\mathbf{w}_s)' = \frac{2}{TN}\sum_{s=1}^{T}\sum_{i=1}^{N}\boldsymbol{\lambda}_i e_{is}\Phi(\mathbf{w}_s)'$$
$$-\frac{2}{TN}\sum_{s=1}^{T}\sum_{i=1}^{N}\boldsymbol{\lambda}_i e_{is}\Phi(\mathbf{w}_s)'1(|e_{is}| > \alpha_T) + \frac{1}{TN}\sum_{s=1}^{T}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\Phi(\mathbf{w}_s)'1\{|e_{is}| > \alpha_T\}$$
$$= a_1 + a_2 + a_3, \text{ say.}$$

As in the proof of Lemma D.6, $E\|a_2\|^2 + E\|a_3\|^2 = O(J\alpha_T^{-\zeta_2}/T + J\alpha_T^{-2(\zeta_2+1)})$. As for a_1, note that $e_{is} = \boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s + u_{is}$. Hence

$$a_1 = \frac{2}{TN}\sum_{s=1}^{T}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)' + \frac{2}{TN}\sum_{s=1}^{T}\sum_{i=1}^{N}\boldsymbol{\lambda}_i u_{is}\Phi(\mathbf{w}_s)' = \frac{2}{NT}\mathbf{\Lambda}'\mathbf{\Lambda}\sum_{t=1}^{T}\boldsymbol{\gamma}_t\Phi(\mathbf{w}_t)' + O_P(\sqrt{J/(NT)}).$$

This gives the desired result for $\mathbf{\Lambda}'\mathbf{M}_\alpha$. On the other hand, the expansion for $\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha$ follows from Lemma D.5 for $\|\mathbf{M}_\alpha\|$ and $\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha = \mathbf{H}'\mathbf{\Lambda}'\mathbf{M}_\alpha + O_P(\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|\|\mathbf{M}_\alpha\|)$.

In the proof of Lemma D.9, we showed that $E\|\frac{1}{N}\mathbf{u}_t'\mathbf{M}_\alpha\|^2 = O(\frac{J}{T^2} + \frac{J}{NT} + \frac{J}{T}\alpha_T^{-\zeta_2})$. This implies $\frac{1}{N}\mathbf{u}_t'\mathbf{M}_\alpha = O_P(\frac{\sqrt{J}}{T} + \sqrt{\frac{J}{NT}} + \sqrt{\frac{J}{T}}\alpha_T^{-\zeta_2/2})$.

Finally, for each fixed t, $\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{R}_t\| = O_P(\frac{1}{N}\sum_{i=1}^N |R_{it}|) = O_P(J^{1-\eta} + J^2\alpha_T^{-\zeta_1} + \frac{J^2\sqrt{\log J \log N}}{T})$.

(ii) The first part follows from (i) and $\frac{1}{N}\mathbf{\Lambda}'\mathbf{B}_1 = \mathbf{\Sigma}_{\Lambda,N}\mathbf{G}_T\mathbf{A}\frac{1}{N}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}$.

The second part follows from (i), Lemmas D.6,D.8, and

$$\frac{1}{N}\mathbf{\Lambda}'\mathbf{B}_3 = \frac{1}{N}\mathbf{\Lambda}'\mathbf{M}_\alpha\mathbf{A}\mathbf{G}_T'\frac{1}{N}\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1} = \frac{1}{N}\mathbf{\Lambda}'\mathbf{M}_\alpha\mathbf{A}\mathbf{G}_T'\mathbf{\Sigma}_{\Lambda,N}\mathbf{H}\widetilde{\mathbf{V}}^{-1} + O_P(\|\frac{1}{N}\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|).$$

(iii) It follows from the proof of Lemma D.8 that

$$\sum_{i\neq 1,3}\|\frac{1}{N}\mathbf{\Lambda}'\mathbf{B}_i\| = O_P((\max_i \frac{1}{T}\sum_t R_{it}^2)^{1/2} + (\|\frac{1}{N}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|)^2) = O_P(J^{1/2-\eta} + \alpha_T^{-\zeta_1}J^{3/2} + \frac{\sqrt{J^3\log N\log J}}{T} + \tilde{\xi}_{NT}^2).$$

The result then follows from part (ii).

Finally, straightforward calculations very similar to the proof of Lemma D.11 give:

$$\|\frac{1}{N}\mathbf{u}_t'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| = O_P(\|\frac{1}{N}\mathbf{u}_t'\mathbf{\Lambda}\|(\|\frac{1}{N}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\| + (\max_i \frac{1}{T}\sum_t R_{it}^2)^{1/2}) + \|\frac{1}{N}\mathbf{u}_t'\mathbf{M}_\alpha\| + (\frac{1}{T}\sum_s(\frac{\mathbf{u}_t'\mathbf{R}_s}{N})^2)^{1/2}).$$

In the proof of Lemma D.10, we have already proved that

$$(\frac{1}{T}\sum_s(\frac{\mathbf{u}_t'\mathbf{R}_s}{N})^2)^{1/2} = O_P(J^{1-\eta}\frac{1}{\sqrt{N}} + (\alpha_T^{-\zeta_1} + \frac{\sqrt{\log N\log J}}{T})J^2).$$

On the other hand, the rates for other terms on the right hand side are found in part (i), and Proposition D.3.

Q.E.D.

We now study the limiting behaviors of $\mathbf{H}$. Define

$$\mathbf{\Sigma}_{F,T} = \frac{1}{T}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)', \qquad \mathbf{\Sigma}_F = E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}.$$

Let

$$\mathbf{C}_N := \frac{1}{N}\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}.$$

In addition, define and recall the following notation:

(i) $\mathbf{V}_N$ is a $K \times K$ diagonal matrix, consisting of the eigenvalues of $\mathbf{\Sigma}/N$, where $\mathbf{\Sigma} = E\{E(\mathbf{x}_t|\mathbf{w}_t)E(\mathbf{x}_t|\mathbf{w}_t)'\}$;

(ii) $\mathbf{V}^*$ is a $K \times K$ diagonal matrix, consisting of the eigenvalues of $\mathbf{C}_N'\mathbf{\Sigma}_{F,T}\mathbf{C}_N$;

(iii) $\mathbf{V}$ is a $K \times K$ diagonal matrix, consisting of the eigenvalues of $\mathbf{\Sigma}_\Lambda^{1/2}\mathbf{\Sigma}_F\mathbf{\Sigma}_\Lambda^{1/2}$;

(iv) $\widetilde{\mathbf{V}}$ is a $K \times K$ diagonal matrix, consisting of the eigenvalues of $\widehat{\mathbf{\Sigma}}/N$, where $\widehat{\mathbf{\Sigma}} = \frac{1}{T}\sum_t \widehat{E}(\mathbf{x}_t|\mathbf{w}_t)\widehat{E}(\mathbf{x}_t|\mathbf{w}_t)'$.

Denote

$$a_{NT} := \sqrt{\frac{J}{TN}} + \frac{\sqrt{J^3 \log N \log J}}{T} + \sqrt{\frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|}{T}} + J^{1/2-\eta} + \alpha_T^{-(\zeta_1-1)}\sqrt{\frac{J^3 \log N}{T}},$$

which is the rate of convergence for $\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F$ by Lemma D.8.

Lemma D.13. *(i)* $\mathbf{H} \to^P \mathbf{\Sigma}_F^{1/2}\mathbf{\Gamma}\mathbf{V}^{-1/2} := \mathbf{J}'$, $\quad \frac{1}{N}\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}} \to^P \mathbf{\Sigma}_F^{-1/2}\mathbf{\Gamma}\mathbf{V}^{1/2}$.

(ii) Let $\bar{\mathbf{\Gamma}}$ *be the eigenvector matrix of* $\mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_F^{1/2}$*. Then*

$$\|\mathbf{H} - \mathbf{\Sigma}_F^{1/2}\bar{\mathbf{\Gamma}}\mathbf{V}_N^{-1/2}\| = O_P(a_{NT} + J^{-\eta} + \sqrt{\frac{\log N}{T}}).$$

Note that result (i) provides the limit of $\mathbf{H}$ and is useful to derive the limiting distribution for $\widehat{\boldsymbol{\gamma}}$, while result (ii) provides a rate of convergence. But since $\mathbf{\Sigma}_F^{1/2}\bar{\mathbf{\Gamma}}\mathbf{V}_N^{-1/2}$ still depends on N, it is not the limit. Note that it is possible to obtain the rate of convergence for part (i) as well, but it then requires the knowledge of the rate for $\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Lambda} - \mathbf{\Sigma}_\Lambda$, and is not pursued here.

Proof. (i) By definition, $\mathbf{H} = \mathbf{\Sigma}_{F,T}\mathbf{C}_N\widetilde{\mathbf{V}}^{-1}$. Hence

$$\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}} - \mathbf{\Lambda}\mathbf{\Sigma}_{F,T}\mathbf{C}_N = (\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\widetilde{\mathbf{V}}. \tag{D.4}$$

Step 1 The limit of $\mathbf{C}_N'\mathbf{\Sigma}_{F,T}\mathbf{C}_N$.

Left multiply $\widehat{\mathbf{\Lambda}}'/N$ on (D.4), and note that $\frac{1}{N}\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Lambda}} = \mathbf{I}$, by Lemma D.8,

$$\widetilde{\mathbf{V}} - \mathbf{C}_N'\mathbf{\Sigma}_{F,T}\mathbf{C}_N = \frac{1}{N}\widehat{\mathbf{\Lambda}}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\widetilde{\mathbf{V}} = O_P(a_{NT}) = o_P(1).$$

By Lemma D.4, $\widetilde{\mathbf{V}} \to^P \mathbf{V}$. Thus $\mathbf{C}_N'\mathbf{\Sigma}_{F,T}\mathbf{C}_N \to^P \mathbf{V}$.

Step 2 The limit of $\mathbf{C}_N$.

Left multiply $\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{\Lambda}'/N$ on (D.4), $\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{C}_N\widetilde{\mathbf{V}}-\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_{F,T}\mathbf{C}_N = \mathbf{\Sigma}_{F,T}^{1/2}\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\widetilde{\mathbf{V}}$. Call the right hand side to be $\mathbf{D}$. By Lemma D.8, $\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|_F = O_P(a_{NT}) = o_P(1)$. Hence $\|\mathbf{D}\| = o_P(a_{NT})$. It also implies

$$\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{C}_N\widetilde{\mathbf{V}} = (\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_{F,T}^{1/2} + \mathbf{D}\mathbf{C}_N^{-1}\mathbf{\Sigma}_{F,T}^{-1/2})\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{C}_N.$$

Let $\widehat{\mathbf{\Gamma}} = \mathbf{\Sigma}_{F,T}^{1/2}\mathbf{C}_N\mathbf{V}^{*-1/2}$ so that each column of $\widehat{\mathbf{\Gamma}}$ has a unit length, and we have

$$\widehat{\mathbf{\Gamma}}\widetilde{\mathbf{V}} = (\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_{F,T}^{1/2} + \mathbf{D}\mathbf{C}_N^{-1}\mathbf{\Sigma}_{F,T}^{-1/2})\widehat{\mathbf{\Gamma}}.$$

Thus the columns of $\widehat{\mathbf{\Gamma}}$ are the eigenvectors of $\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_{F,T}^{1/2}+\mathbf{D}\mathbf{C}_N^{-1}\mathbf{\Sigma}_{F,T}^{-1/2}$. By $\|\mathbf{D}\| = o_P(1)$, Assumption 2.2, and Lemma C.1, the eigenvalues of $\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_{F,T}^{1/2}+\mathbf{D}\mathbf{C}_N^{-1}\mathbf{\Sigma}_{F,T}^{-1/2}$ are distinct, which are asymptotically those of $\mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_\Lambda\mathbf{\Sigma}_F^{1/2}$. This implies that $\widehat{\mathbf{\Gamma}}$ is uniquely determined except that each column can be replaced by the negative of itself. In addition, the kth column of $\widehat{\mathbf{\Gamma}}$ (and $\mathbf{\Sigma}_{F,T}^{1/2}\mathbf{C}_N$) depends on $\widehat{\mathbf{\Lambda}}$ only through the kth column of $\widehat{\mathbf{\Lambda}}$. Hence the sign of each column of $\widehat{\mathbf{\Gamma}}$ is also uniquely determined by the column signs of $\widehat{\mathbf{\Lambda}}$. Then applying Lemma C.1 on the eigenvectors, we conclude that there is a unique eigenvector matrix $\mathbf{\Gamma}$ of $\mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_\Lambda\mathbf{\Sigma}_F^{1/2}$, so that $\widehat{\mathbf{\Gamma}} \to^P \mathbf{\Gamma}$, which also implies $\mathbf{\Sigma}_F^{1/2}\mathbf{C}_N\mathbf{V}^{*-1/2} \to^P \mathbf{\Gamma}$.

By step 1, $\mathbf{V}^* \to^P \mathbf{V}$. Hence $\mathbf{C}_N \to^P \mathbf{\Sigma}_F^{-1/2}\mathbf{\Gamma}\mathbf{V}^{1/2}$. We have $\mathbf{H} = \mathbf{\Sigma}_{F,T}\mathbf{C}_N\widetilde{\mathbf{V}}^{-1} \to \mathbf{\Sigma}_F\mathbf{\Sigma}_F^{-1/2}\mathbf{\Gamma}\mathbf{V}^{1/2}\mathbf{V}^{-1} = \mathbf{\Sigma}_F^{1/2}\mathbf{\Gamma}\mathbf{V}^{-1/2} := \mathbf{J}'$.

(ii) Since $\|\mathbf{\Sigma}_{F,T}^{1/2}-\mathbf{\Sigma}_F^{1/2}\| = O_P(T^{-1/2})$, by Lemma C.1, and that $\mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_\Lambda\mathbf{\Sigma}_F^{1/2}$ have distinct eigenvalues, we have

$$\|\widehat{\mathbf{\Gamma}}-\bar{\mathbf{\Gamma}}\| \leq \|\mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_F^{1/2} - \mathbf{\Sigma}_{F,T}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\mathbf{\Sigma}_{F,T}^{1/2}\| + \|\mathbf{D}\mathbf{C}_N^{-1}\mathbf{\Sigma}_{F,T}^{-1/2}\| = O_P(\frac{1}{\sqrt{T}} + a_{NT}).$$

Also by Lemma D.4, $\|\widetilde{\mathbf{V}}-\mathbf{V}_N\| = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}})$. It then follows from step 1 that

$$\|\mathbf{C}_N'\mathbf{\Sigma}_{F,T}\mathbf{C}_N - \mathbf{V}_N\| = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}} + a_{NT}).$$

This implies $\|\mathbf{V}^*-\mathbf{V}_N\| = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}} + a_{NT})$. Lemma D.4 also states $\|\widehat{\mathbf{V}}-\mathbf{V}_N\| = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}})$. We also have $\mathbf{H} = \mathbf{\Sigma}_{FT}^{1/2}\widehat{\mathbf{\Gamma}}\mathbf{V}^{*1/2}\widetilde{\mathbf{V}}^{-1}$. Hence

$$\|\mathbf{H}-\mathbf{\Sigma}_F^{1/2}\bar{\mathbf{\Gamma}}\mathbf{V}_N^{-1/2}\| = \|\mathbf{\Sigma}_{FT}^{1/2}\widehat{\mathbf{\Gamma}}\mathbf{V}^{*1/2}\widetilde{\mathbf{V}}^{-1} - \mathbf{\Sigma}_F^{1/2}\bar{\mathbf{\Gamma}}\mathbf{V}_N^{-1/2}\| = O_P(a_{NT} + J^{-\eta} + \sqrt{\frac{\log N}{T}}).$$

Proof of Theorem 3.3

Proof. Recall (D.3),

$$\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t = \frac{1}{N}\mathbf{H}'\boldsymbol{\Lambda}'\mathbf{u}_t + \sum_{i=1}^{d}\mathbf{D}_{ti}.$$

Under our conditions, by Lemma D.12,

$$\begin{aligned}
\mathbf{D}_{t1} &= \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'(\boldsymbol{\Lambda}\mathbf{H} - \widehat{\boldsymbol{\Lambda}})\mathbf{H}^{-1}\boldsymbol{\gamma}_t = \frac{1}{N}\mathbf{H}'\boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\mathbf{H} - \widehat{\boldsymbol{\Lambda}})\mathbf{H}^{-1}\boldsymbol{\gamma}_t + O_P(\frac{1}{N}\|\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H}\|^2) \\
&= -\frac{2}{T}\sum_{s=1}^{T}\mathbf{H}'(\mathbf{M}_1\Phi(\mathbf{w}_s)\boldsymbol{\gamma}_s'\mathbf{M}_2 + \boldsymbol{\Sigma}_{\Lambda,N}\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{M}_3)\mathbf{H}^{-1}\boldsymbol{\gamma}_t + o_P(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}) \\
\mathbf{D}_{t2} &= \frac{1}{N}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H})'\mathbf{u}_t = o_P(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}), \\
\mathbf{D}_{t3} &= \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t) = 2\mathbf{H}'\boldsymbol{\Sigma}_{\Lambda,N}\frac{1}{T}\sum_{s=1}^{T}\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{A}\Phi(\mathbf{w}_t) + o_P(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}), \\
\mathbf{D}_{t4} &= \frac{1}{N}\widehat{\boldsymbol{\Lambda}}'\mathbf{R}_t = o_P(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}).
\end{aligned}$$

By the definitions of $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3$ in Lemma D.12 and Lemma D.13, and that $\mathbf{G}_T \to^P \mathbf{G} := E\mathbf{f}_t\Phi(\mathbf{w}_t)'$, we have

$$\mathbf{M}_1 \to^P \boldsymbol{\Sigma}_\Lambda\mathbf{G}\mathbf{A}, \quad \mathbf{M}_2 \to^P \boldsymbol{\Sigma}_\Lambda\boldsymbol{\Sigma}_F^{1/2}\boldsymbol{\Gamma}\mathbf{V}^{-3/2}, \quad \mathbf{M}_3 \to^P \mathbf{A}\mathbf{G}'\boldsymbol{\Sigma}_\Lambda\boldsymbol{\Sigma}_F^{1/2}\boldsymbol{\Gamma}\mathbf{V}^{-3/2}.$$

In addition, $\mathbf{M}_2\mathbf{H}^{-1} \to^P \boldsymbol{\Sigma}_F^{-1}$, and $\mathbf{M}_3\mathbf{H}^{-1} \to^P \mathbf{A}\mathbf{G}'\boldsymbol{\Sigma}_F^{-1}$.

Let $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$, and $\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}} \asymp \delta_{NT}^{-1}$. Let $a_{NT} = \frac{\delta_{NT}}{\sqrt{N}}$, $b_{NT} = \frac{\delta_{NT}}{\sqrt{T}}$. Note that $o(a_{NT}) = o(1)$ and $o(b_{NT}) = o(1)$. Since $\mathbf{H}' \to^P \mathbf{J} := \mathbf{V}^{-1/2}\boldsymbol{\Gamma}'\boldsymbol{\Sigma}_F^{1/2}$,

$$\begin{aligned}
&\delta_{NT}(\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t) = a_{NT}\mathbf{H}'\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_{it}\boldsymbol{\lambda}_i + b_{NT}\mathbf{H}'\boldsymbol{\Sigma}_{\Lambda,N}\frac{2}{\sqrt{T}}\sum_{s=1}^{T}\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{A}\Phi(\mathbf{w}_t) \\
&-b_{NT}\frac{2}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{H}'(\mathbf{M}_1\Phi(\mathbf{w}_s)\boldsymbol{\gamma}_s'\mathbf{M}_2 + \boldsymbol{\Sigma}_{\Lambda,N}\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{M}_3)\mathbf{H}^{-1}\boldsymbol{\gamma}_t + o_P(1), \\
= \;& b_{NT}\frac{2}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{J}\boldsymbol{\Sigma}_\Lambda(\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{A}\Phi(\mathbf{w}_t) - \mathbf{G}\mathbf{A}\Phi(\mathbf{w}_s)\boldsymbol{\gamma}_s'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{A}\mathbf{G}'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{\gamma}_t) \\
&+a_{NT}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{J}\boldsymbol{\lambda}_i u_{it} + o_P(1) = b_{NT}\mathbf{c}_T + a_{NT}\mathbf{d}_N + o_P(1),
\end{aligned}$$

where $\mathbf{d}_N = \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{J}\boldsymbol{\lambda}_i u_{it}$, and for $m_{ts} = [\Phi(\mathbf{w}_t) - \mathbf{G}'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{\gamma}_t]'\mathbf{A}\Phi(\mathbf{w}_s)$,

$$\begin{aligned}\mathbf{c}_T &= \frac{1}{\sqrt{T}}\sum_{s=1}^T 2\mathbf{J}\boldsymbol{\Sigma}_\Lambda\boldsymbol{\gamma}_s\Phi(\mathbf{w}_s)'\mathbf{A}[\Phi(\mathbf{w}_t) - \mathbf{G}'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{\gamma}_t] - 2\mathbf{J}\boldsymbol{\Sigma}_\Lambda\mathbf{G}\mathbf{A}\Phi(\mathbf{w}_s)\boldsymbol{\gamma}_s'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{\gamma}_t \\ &= \frac{2}{\sqrt{T}}\sum_{s=1}^T \mathbf{J}\boldsymbol{\Sigma}_\Lambda\{m_{ts}\mathbf{I}_K - \mathbf{G}\mathbf{A}\Phi(\mathbf{w}_s)\boldsymbol{\gamma}_t'\boldsymbol{\Sigma}_F^{-1}\}\boldsymbol{\gamma}_s.\end{aligned}$$

Then $\mathbf{d}_N \to^d \mathbf{d} =^d \mathcal{N}(0, \mathbf{JQJ}')$, $\mathbf{c}_T \to^d \mathbf{c} =^d \mathcal{N}(0, \mathbf{J}\boldsymbol{\Sigma}_\Lambda\mathbf{M}_t\boldsymbol{\Sigma}_\Lambda\mathbf{J}')$.

In addition, because $E(\mathbf{u}_t|\mathbf{w}_t, \mathbf{f}_t) = 0$, we have $E(\mathbf{u}_t h(\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_T)) = 0$ for any function h. This implies $\mathbf{d}_N$ and $\mathbf{c}_T$ are uncorrelated. Since they are both asymptotically normal, it implies that $\mathbf{c}_T$ and $\mathbf{d}_N$ are asymptotically independent. Thus, $(\mathbf{d}_N, \mathbf{c}_T)$ converges jointly to a multivariate normal distribution: $(\mathbf{d}_N, \mathbf{c}_T) \to^d (\mathbf{d}, \mathbf{c})$. Then, the almost sure representation theory (Pollard (1984), page 71) asserts that there exist random vectors $(\mathbf{d}_N^*, \mathbf{c}_T^*)$ and $(\mathbf{d}^*, \mathbf{c}^*)$ with the same distributions as $(\mathbf{d}_N, \mathbf{c}_T)$ and $(\mathbf{d}, \mathbf{c})$ such that $(\mathbf{d}_N^*, \mathbf{c}_T^*) \to (\mathbf{d}^*, \mathbf{c}^*)$ almost surely. Now, $b_{NT}\mathbf{c}^* + a_{NT}\mathbf{d}^* =^d b_{NT}\mathbf{c} + a_{NT}\mathbf{d} =^d \mathcal{N}(0, b_{NT}^2\mathbf{J}\boldsymbol{\Sigma}_\Lambda\mathbf{M}_t\boldsymbol{\Sigma}_\Lambda\mathbf{J}' + a_{NT}^2\mathbf{JQJ}')$. That is, $(b_{NT}^2\mathbf{J}\boldsymbol{\Sigma}_\Lambda\mathbf{M}_t\boldsymbol{\Sigma}_\Lambda\mathbf{J}' + a_{NT}^2\mathbf{JQJ}')^{-1/2}(b_{NT}\mathbf{c}^* + a_{NT}\mathbf{d}^*) =^d \mathcal{N}(0, \mathbf{I})$. Therefore,

$$(b_{NT}^2\mathbf{J}\boldsymbol{\Sigma}_\Lambda\mathbf{M}_t\boldsymbol{\Sigma}_\Lambda\mathbf{J}' + a_{NT}^2\mathbf{JQJ}')^{-1/2}(b_{NT}\mathbf{c}_T^* + a_{NT}\mathbf{d}_N^*) \to^d \mathcal{N}(0, \mathbf{I}).$$

The above is true with $(\mathbf{c}_T^*, \mathbf{d}_N^*)$ replaced with $(\mathbf{c}_T, \mathbf{d}_N)$ because they have the same distribution. Also, $(b_{NT}^2 + a_{NT}^2)^{-1/2} = O(1)$. Hence

$$(b_{NT}^2\mathbf{J}\boldsymbol{\Sigma}_\Lambda\mathbf{M}_t\boldsymbol{\Sigma}_\Lambda\mathbf{J}' + a_{NT}^2\mathbf{JQJ}')^{-1/2}\delta_{NT}(\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t) \to^d \mathcal{N}(0, \mathbf{I}),$$

which is equivalent to the desired result.

Proof of Corollary 3.1

It follows directly from Theorems 3.1, 3.2 and straightforward calcualtions that $\widehat{\boldsymbol{\Sigma}}_F = \mathbf{H}^{-1}\boldsymbol{\Sigma}_F\mathbf{H}'^{-1} + o_P(1)$, $\widehat{\boldsymbol{\Sigma}}_\Lambda = \mathbf{H}'\boldsymbol{\Sigma}_\Lambda\mathbf{H} + o_P(1)$, $\widehat{\mathbf{G}} = \mathbf{H}^{-1}\mathbf{G} + o_P(1)$, $\widehat{\mathbf{S}} = \mathbf{S} + o_P(1)$, $\widehat{\mathrm{cov}}(\boldsymbol{\gamma}_t) = \mathbf{H}^{-1}\,\mathrm{cov}(\boldsymbol{\gamma}_t)\mathbf{H}'^{-1} + o_P(1)$, $\widehat{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}_t + o_P(1)$, $\widehat{\boldsymbol{\beta}}_t = \mathbf{H}'\boldsymbol{\beta}_t + o_P(1)$. Hence $\widehat{\mathbf{M}}_t = \mathbf{H}^{-1}\mathbf{M}_t\mathbf{H}'^{-1} + o_P(1)$. In addition, by Lemma D.13, $\mathbf{H} \to^P \mathbf{J}'$. Hence

$$\widehat{\boldsymbol{\Sigma}}_\Lambda\widehat{\mathbf{M}}_t\widehat{\boldsymbol{\Sigma}}_\Lambda = \mathbf{H}'\boldsymbol{\Sigma}_\Lambda\mathbf{M}_t\boldsymbol{\Sigma}_\Lambda\mathbf{H} + o_P(1) = \mathbf{J}\boldsymbol{\Sigma}_\Lambda\mathbf{M}_t\boldsymbol{\Sigma}_\Lambda\mathbf{J}' + o_P(1).$$

Also, under cross-sectional correlations, by Lemma E.3 below, $\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\| = o_P(1)$. Hence

$$\widehat{\mathbf{Q}} = \mathbf{H}'\mathbf{Q}\mathbf{H} + o_P(1) = \mathbf{J}\mathbf{Q}\mathbf{J}' + o_P(1).$$

The result then follows from the Slutsky's theorem.

E Proofs for Section 4

The proof of the limiting distribution of S under the null is divided into two major steps.

step 1: Asymptotic expansion: under H_0,

$$S = \frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t + o_P(T^{-1/2}).$$

step 2: The effect of estimating $\mathbf{\Sigma}_u$ is first-order negligible:

$$\frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t = \frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t + o_P(T^{-1/2}).$$

The result then follows from the asymptotic normality of the first term on the right hand side. We shall prove this using Lindeberg's central limit theorem.

We achieve each step in the following subsections.

E.1 Step 1 asymptotic expansion of S

Proposition E.1. *Under H_0,*

$$S = \frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t + o_P(T^{-1/2})$$

Proof. Since $\|\widehat{\mathbf{W}}\| \leq \max_i \widehat{\sigma}_{ii} = O_P(1)$, it follows from (D.3) that it suffices to prove under H_0, $\frac{N}{T}\sum_{t=1}^{T}\mathbf{D}_{ti}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t = o_P(T^{-1/2})$, and $\frac{N}{T}\sum_{t=1}^{T}\|\mathbf{D}_{ti}\|^2 = o_P(T^{-1/2}), i = 2, 3, 4$.

By the proof of Propositions D.5, D.3, Lemmas D.6, D.11 and that $\mathbf{D}_{t3} = \mathbf{C}_{t3}$,$\mathbf{D}_{t4} = \mathbf{C}_{t4}$,

$$\frac{N}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t4}\|^2 = O_P(\max_i \frac{N}{T}\sum_{t=1}^{T}R_{it}^2) = O_P(NJ^{1-2\eta} + \frac{NJ^3\log N}{\alpha_T^{2(\zeta_1-1)}T} + \frac{NJ^3\log N\log J}{T^2}) = o_P(\frac{1}{\sqrt{T}})$$

$$\frac{N}{T}\sum_{t=1}^T\|\mathbf{D}_{t3}\|^2 = O_P(\frac{1}{N}\|\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\|^2)=O_P(\frac{J}{T}+\frac{NJ\alpha_T^{-\zeta_2}}{T}+J^{2-4\eta}+\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{TJ^{2\eta-1}})=o_P(\frac{1}{\sqrt{T}})$$

The last equality holds so long as $N\sqrt{T}=o(J^{2\eta-1})$, $NJ^4\log N\log J=o(T^{3/2})$, $\zeta_1>2$.

By Lemma D.10,

$$\frac{N}{T}\sum_{t=1}^T\|\mathbf{D}_{t2}\|^2=O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2+\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2+\frac{1}{NT}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2+\frac{1}{NT^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2)=o_P(\frac{1}{\sqrt{T}}).$$

The proof of $\frac{N}{T}\sum_{t=1}^T\mathbf{D}_{ti}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t=o_P(T^{-1/2})$ is given in Lemmas E.1 and E.2 below. It then leads to the desired result.

Lemma E.1. *Suppose* $(N+T)J^{1-2\eta}=o(1)$. *Then*

$$\mathrm{tr}(\frac{N}{T}\sum_{t=1}^T\mathbf{D}_{t2}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t)=o_P(T^{-1/2})$$

Proof. It suffics to prove $\|\frac{1}{T}\sum_{t=1}^T\mathbf{D}_{t2}\mathbf{u}_t'\mathbf{\Lambda}\|^2=\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|^2=o_P(\frac{1}{T})$. To this end, we need to decompose $\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}=\sum_{i=1}^8\mathbf{B}_i$ again as in (D.2). Every term can be bounded using established bounds except for the term involving $\mathbf{B}_3$. More specifically, for $i\neq 3$, we use $\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_i\|^2\leq\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\|_F^2\|\mathbf{B}_i\|^2$. On the other hand, $\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\|_F^2\leq 2\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Sigma}_u\|_F^2+2\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'(\mathbf{u}_t\mathbf{u}_t'-\mathbf{\Sigma}_u)\|_F^2$. The first term is $O_P(\frac{1}{N})$. As for the second tern,

$$\begin{aligned}
&E\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'(\mathbf{u}_t\mathbf{u}_t'-\mathbf{\Sigma}_u)\|_F^2=\sum_{k=1}^K\sum_{i=1}^N E(\frac{1}{TN}\sum_{t=1}^T\sum_{j=1}^N\lambda_{jk}(u_{jt}u_{it}-Eu_{jt}u_{it}))^2\\
=&\ \frac{1}{T^2N^2}\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T\mathrm{var}(\sum_{j=1}^N\lambda_{jk}(u_{jt}u_{it}-Eu_{jt}u_{it}))=\frac{1}{T^2N^2}\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T\sum_{j=1}^N\sum_{l=1}^N\lambda_{jk}\lambda_{lk}\,\mathrm{cov}(u_{jt}u_{it},u_{lt}u_{it})\\
=&\ O(\frac{1}{T})+\frac{1}{T^2N^2}\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T\sum_{j=1}^N\sum_{l\neq i,t}\lambda_{jk}\lambda_{lk}E(u_{jt}u_{it}-\sigma_{ij})u_{it}u_{lt}=O(\frac{1}{T}).
\end{aligned}$$

Hence $\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_i\|^2\leq O_P(\frac{1}{T}+\frac{1}{N})\|\mathbf{B}_i\|^2=o(\frac{1}{T})$, for $i\neq 3$, where the last equality holds by straightforward verifying $(\frac{T}{N}+1)\|\mathbf{B}_i\|^2=o(1)$ using Lemma D.5, assuming $(N+T)J^{1-2\eta}=o(1)$.

To allow $N/T \to \infty$, the term involving $\mathbf{B}_3$ requires a different and sharper bound:

$$\begin{aligned}
& \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_3\|^2 = \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{M}_\alpha\frac{1}{TN}\sum_{s=1}^{T}\mathbf{A}\Phi(\mathbf{w}_s)E(\mathbf{f}_s|\mathbf{w}_s)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \\
\leq\ & \|\frac{1}{TN}\sum_{t=1}^{T}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{M}_\alpha\|^2 O_P(1) = O_P(1)\|\frac{1}{TN}\sum_{t}\mathbf{\Lambda}'\mathbf{u}_t\sum_{i=1}^{N}u_{it}\frac{1}{T}\sum_{s=1}^{T}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\Phi(\mathbf{w}_s)\|^2 \\
\leq\ & O_P(1)\|\frac{1}{TN}\sum_{t}\mathbf{\Lambda}'\mathbf{u}_t\sum_{i=1}^{N}u_{it}\frac{1}{T}\sum_{s=1}^{T}u_{is}1\{|u_{is}|<\alpha_T\}\Phi(\mathbf{w}_s)\|^2 \\
& +O_P(1)\|\frac{1}{TN}\sum_{t}\mathbf{\Lambda}'\mathbf{u}_t\sum_{i=1}^{N}u_{it}\frac{1}{T}\sum_{s=1}^{T}\alpha_T\dot{\rho}(\alpha_T^{-1}u_{is})1\{|u_{is}|>\alpha_T\}\Phi(\mathbf{w}_s)\|^2 \\
\leq\ & O_P(1)\|\frac{1}{T^2N}\sum_{t}\sum_{s=1}^{T}\sum_{i=1}^{N}\mathbf{\Lambda}'\mathbf{u}_t u_{it}u_{is}\Phi(\mathbf{w}_s)\|^2 \\
& +O_P(1)\left(\frac{1}{T^2N}\sum_{t}\sum_{s=1}^{T}\sum_{i=1}^{N}\|\mathbf{\Lambda}'\mathbf{u}_t\||u_{it}||u_{is}|1\{|u_{is}|>\alpha_T\}\|\Phi(\mathbf{w}_s)\|\right)^2, \qquad \text{(E.1)}
\end{aligned}$$

where we used the fact that under H_0, $e_{is}=u_{is}$. We respectively bound the two terms on the right hand side.

First term in (E.1) Note that

$$\begin{aligned}
& E\|\frac{1}{T^2N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\mathbf{\Lambda}'\mathbf{u}_t u_{it}u_{is}\Phi(\mathbf{w}_s)\|^2 = \frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}E(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_{jk}u_{jt}u_{it}u_{is}\phi_l(\mathbf{w}_s))^2 \\
=\ & \frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{m=1}^{T}\sum_{n=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{v=1}^{N}\sum_{h=1}^{N}\lambda_{hk}\lambda_{jk}Eu_{jt}u_{it}u_{is}u_{hm}u_{vm}u_{vn}\phi_l(\mathbf{w}_n)\phi_l(\mathbf{w}_s) \\
=\ & \frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{m\neq s,t}\sum_{i=1}^{N}\sum_{v=1}^{N}\lambda_{vk}\lambda_{ik}Eu_{it}^2Eu_{vm}^2[E\phi_l(\mathbf{w}_s)^2(Eu_{is}u_{vs}|\mathbf{w}_s)] \\
& +\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{v=1}^{N}\lambda_{ik}\lambda_{ik}Eu_{it}^2Eu_{is}^2u_{vs}^2\phi_l(\mathbf{w}_s)^2 \\
& +\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{h\neq i}\lambda_{hk}\lambda_{ik}Eu_{it}^2Eu_{hs}u_{is}^3\phi_l(\mathbf{w}_s)^2 \\
& +\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{m\neq s,t}\sum_{i=1}^{N}\sum_{v=1}^{N}\lambda_{vk}\lambda_{ik}Eu_{it}^2Eu_{vm}^2E(\phi_l(\mathbf{w}_s)^2(Eu_{is}u_{vs}|\mathbf{w}_s)) \\
& +\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{v=1}^{N}\lambda_{ik}\lambda_{ik}Eu_{it}^2Eu_{is}^2u_{vs}^2\phi_l(\mathbf{w}_s)^2
\end{aligned}$$

$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{h\neq i}\lambda_{hk}\lambda_{ik}Eu_{it}^2Eu_{hs}u_{is}^3\phi_l(\mathbf{w}_s)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{h=1}^{N}\lambda_{hk}\lambda_{ik}Eu_{it}^3\phi_l(\mathbf{w}_t)Eu_{is}u_{hs}u_{is}\phi_l(\mathbf{w}_s)$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{h=1}^{N}\lambda_{hk}\lambda_{ik}Eu_{ht}u_{it}^3Eu_{is}^2\phi_l(\mathbf{w}_s)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{j\neq i}\lambda_{ik}\lambda_{jk}Eu_{jt}u_{it}^3Eu_{is}^2\phi_l(\mathbf{w}_s)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{j\neq i}\lambda_{jk}\lambda_{jk}Eu_{jt}^2u_{it}^2Eu_{is}^2\phi_l(\mathbf{w}_s)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{v\neq i}\sum_{h=1}^{N}\lambda_{hk}\lambda_{ik}Eu_{it}^2u_{ht}u_{vt}E(\phi_l(\mathbf{w}_s)^2(Eu_{vs}u_{is}|\mathbf{w}_s))$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s\neq t}\sum_{i=1}^{N}\sum_{j\neq i}\sum_{v\neq i}\lambda_{ik}\lambda_{jk}Eu_{jt}u_{it}^2u_{vt}(E\phi_l(\mathbf{w}_s)^2(Eu_{vs}u_{is}|\mathbf{w}_s))$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{v=1}^{N}\sum_{h=1}^{N}\lambda_{hk}\lambda_{ik}Eu_{it}^3u_{ht}u_{vt}^2\phi_l(\mathbf{w}_t)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}\lambda_{ik}\lambda_{jk}Eu_{jt}u_{it}^5\phi_l(\mathbf{w}_t)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}\sum_{v=1}^{N}\lambda_{jk}\lambda_{jk}Eu_{jt}u_{it}^2u_{jt}u_{vt}^2\phi_l(\mathbf{w}_t)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}\sum_{v\neq i}\lambda_{ik}\lambda_{jk}Eu_{jt}u_{it}^3u_{vt}^2\phi_l(\mathbf{w}_t)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}\sum_{h\neq i,j}\lambda_{hk}\lambda_{jk}Eu_{jt}u_{it}^2u_{ht}u_{jt}^2\phi_l(\mathbf{w}_t)^2$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{m\neq t}\sum_{i=1}^{N}\sum_{v=1}^{N}\sum_{h=1}^{N}\lambda_{hk}\lambda_{ik}Eu_{it}^3u_{vt}\phi_l(\mathbf{w}_t)^2Eu_{hm}u_{vm}$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{m\neq t}\sum_{i=1}^{N}\sum_{j\neq i}\sum_{h=1}^{N}\lambda_{hk}\lambda_{jk}Eu_{it}^2u_{jt}^2\phi_l(\mathbf{w}_t)^2Eu_{hm}u_{jm}$$
$$+\frac{1}{T^4N^2}\sum_{l=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{m\neq t}\sum_{i=1}^{N}\sum_{v=1}^{N}\lambda_{vk}\lambda_{jk}Eu_{it}^3\phi_l(\mathbf{w}_t)Eu_{vm}^3\phi_l(\mathbf{w}_m)=O_P(\frac{J}{TN}+\frac{J}{T^2}+\frac{JN}{T^3}).$$

Second term in (E.1) As for the second term, first note that under H_0, $u_{it} = e_{it}$. So Lemma D.2 implies $(E|u_{is}|1\{|u_{is}| > \alpha_T\}|\mathbf{w}_t = \mathbf{w}) \leq C\alpha_T^{-\zeta_2-1}$. On the other hand, by assumption, for some $C > 0$, $\sup_{\mathbf{w}} E(u_{it}^4 1\{|u_{it}| > \alpha_T\}|\mathbf{w}_t = \mathbf{w}) \leq \alpha_T^{-\zeta_5} C$, $E\|\mathbf{\Lambda}'\mathbf{u}_t\|^2 = O(N)$. Hence

$$
\begin{aligned}
&\frac{1}{T^2N}\sum_t\sum_{s=1}^T\sum_{i=1}^N E\|\mathbf{\Lambda}'\mathbf{u}_t\||u_{it}||u_{is}|1\{|u_{is}| > \alpha_T\}\|\Phi(\mathbf{w}_s)\| \\
= \;& \frac{1}{T^2N}\sum_t\sum_{i=1}^N E\|\mathbf{\Lambda}'\mathbf{u}_t\|u_{it}^2 1\{|u_{it}| > \alpha_T\}\|\Phi(\mathbf{w}_t)\| \\
&+\frac{1}{T^2N}\sum_t\sum_{s\neq t}\sum_{i=1}^N E\|\mathbf{\Lambda}'\mathbf{u}_t\||u_{it}|E|u_{is}|1\{|u_{is}| > \alpha_T\}\|\Phi(\mathbf{w}_s)\| \\
\leq \;& \frac{1}{T^2N}\sum_t\sum_{i=1}^N (E\|\mathbf{\Lambda}'\mathbf{u}_t\|^2)^{1/2}(E\|\Phi(\mathbf{w}_t)\|^2)^{1/2}\sup_{\mathbf{w}}(Eu_{it}^4 1\{|u_{it}| > \alpha_T\}|\mathbf{w}_t = \mathbf{w})^{1/2} \\
&+\frac{1}{T^2N}\sum_t\sum_{s\neq t}\sum_{i=1}^N (E\|\mathbf{\Lambda}'\mathbf{u}_t\|^2)^{1/2}(Eu_{it}^2)^{1/2}E\|\Phi(\mathbf{w}_s)\|\sup_{\mathbf{w}}(E|u_{is}|1\{|u_{is}| > \alpha_T\}|\mathbf{w}_t = \mathbf{w}) \\
= \;& O_P(\frac{\sqrt{JN}}{T}\alpha_T^{-\zeta_5/2} + \sqrt{NJ}\alpha_T^{-\zeta_2-1}).
\end{aligned}
$$

It then implies the second term in (E.1) is $O_P(\frac{JN}{T^2}\alpha_T^{-\zeta_5} + NJ\alpha_T^{-2\zeta_2-2})$.

Thus, when $\zeta_5 \geq 1$, $T = o(J^{2\eta-1})$

$$
\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_3\|^2 = O_P(\frac{J}{TN} + \frac{J}{T^2} + \frac{JN}{T^3} + \frac{JN}{T^2}\alpha_T^{-\zeta_5} + NJ\alpha_T^{-2\zeta_2-2}) = o_P(\frac{1}{T}).
$$

As a result, $\|\frac{1}{T}\sum_{t=1}^T \mathbf{D}_{t2}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = \|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|^2 = o_P(\frac{1}{T})$.

Lemma E.2. *For $i = 3, 4$,*

$$
\mathrm{tr}(\frac{N}{T}\sum_{t=1}^T \mathbf{D}_{ti}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t) = o_P(T^{-1/2})
$$

Proof. Again, it suffices to verify $\|\frac{1}{T}\sum_{t=1}^T \mathbf{D}_{ti}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = o_P(\frac{1}{T})$ for $i = 3, 4$. Note that $\|\frac{1}{T}\sum_{t=1}^T \Phi(\mathbf{w}_t)\mathbf{u}_t'\mathbf{\Lambda}\|^2 = O_P(\frac{NJ}{T})$. Then by definition,

$$
\|\frac{1}{T}\sum_{t=1}^T \mathbf{D}_{t3}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = \|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{w}_t)\mathbf{u}_t'\mathbf{\Lambda}\|^2 \leq O_P(\frac{1}{N^2})\|\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\|^2\|\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{w}_t)\mathbf{u}_t'\mathbf{\Lambda}\|^2 = o_P(\frac{1}{T}).
$$

On the other hand, recall the definition $R_{it} := R_{1,it} + R_{2,it} + R_{3,it}$, where

$$\begin{aligned} R_{1,it} &:= \Phi(\mathbf{w}_t)'\mathbf{\Lambda}\frac{1}{T}\sum_{s=1}^{T}\alpha_T[\dot\rho(\alpha_T^{-1}e_{is,\alpha}) - \dot\rho(\alpha_T^{-1}e_{is})]\Phi(\mathbf{w}_s) \\ R_{2,it} &:= \Phi(\mathbf{w}_t)'(\mathbf{R}_{i,b} + \mathbf{b}_{i,\alpha} - \mathbf{b}_i), \qquad R_{3,it} := -z_{it}. \end{aligned}$$

Thus it can be verified similarly that

$$\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{D}_{t4}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{R}_t\mathbf{u}_t'\mathbf{\Lambda}\|^2 = O_P(\frac{1}{NT^2})\sum_{i=1}^{N}\|\sum_{t=1}^{T}R_{it}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = o_P(\frac{1}{T}).$$

The verification is very similar as before, and is omitted here.

E.2 Step 2 Completion of the proof

We now aim to show $\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N = \mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N + o_P(T^{-1/2})$. Once this is done, it then follows from the facts that $\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N = O_P(1)$ and $(\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N)^{-1} = O_P(1)$,

$$(\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N)^{-1} = (\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N)^{-1} + o_P(T^{-1/2}).$$

As a result, by Proposition E.1,

$$\begin{aligned} S &= \frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}(\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N)^{-1}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t + o_P(T^{-1/2}) \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t + o_P(T^{-1/2}). \end{aligned}$$

Hence

$$\frac{TS - TK}{\sqrt{2TK}} = \frac{\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t - TK}{\sqrt{2TK}} + o_P(1) \to^d \mathcal{N}(0,1).$$

To finish the proof, we now show two claims:

(1)

$$\frac{\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t - TK}{\sqrt{2TK}} \to^d \mathcal{N}(0,1).$$

(2) $\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N = \mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N + o_P(T^{-1/2})$.

Proof of (1) We define $X_t = \mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t$ and $s_T^2 = \sum_{t=1}^{T}\text{var}(X_t)$. Then

$E(X_t) = \operatorname{tr} E((\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{\Lambda}) = K$. Also by Assumption 4.1, $s_T^2/T \to 2K$, hence we have $E\frac{1}{T}\sum_{t=1}^T(X_t - K)^2 < \infty$ for all large N, T. For any $\epsilon > 0$, by the dominated convergence theorem, for all large N, T,

$$\frac{1}{T}\sum_{t=1}^{T} E(X_t - K)^2 1\{|X_t - K| > \epsilon s_T\} \leq \frac{1}{T}\sum_{t=1}^{T} E(X_t - K)^2 1\{|X_t - K| > \epsilon\sqrt{KT}\} = o(1).$$

This then implies the Lindeberg condition, $\frac{1}{s_T^2}\sum_{t=1}^T E(X_t - K)^2 1\{|X_t - K| > \epsilon s_T\} = o(1)$. Hence by the Lindeberg central limit theorem,

$$\frac{\sum_t X_t - TK}{s_T} \to^d \mathcal{N}(0, 1).$$

The result then follows since $s_T^2/T \to 2K$.

Proof of (2) By the triangular inequality,

$$\begin{aligned}
&\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}} - \frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}\| \leq \|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\widehat{\mathbf{\Lambda}}\| + \|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'\mathbf{\Sigma}_u(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| \\
&+\|\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| + \|\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{\Lambda}\mathbf{H}\| + 2\|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}\|.
\end{aligned}$$

Using the established bounds for $\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|$ in Theorem 3.1, it is straightforward to verify $\|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'\mathbf{\Sigma}_u(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| = o_P(T^{-1/2})$. Other terms require sharper bounds yet to be established. These are given in Proposition E.2 below. It then follows that $\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N = \mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N + o_P(T^{-1/2})$. This completes the proof.

Proposition E.2. *(i)* $\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Sigma}_u(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}) = o_P(T^{-1/2})$;
(ii) $\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{\Lambda} = o_P(T^{-1/2})$;
(iii) $\|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{G}\| = o_P(T^{-1/2})$, *for either* $\mathbf{G} = \mathbf{\Lambda}$ *or* $\mathbf{G} = \widehat{\mathbf{\Lambda}}$.

Proof. Define $\widetilde{\mathbf{\Lambda}} = \mathbf{\Sigma}_u\mathbf{\Lambda}$. Note that we cannot simply bound these terms by $\frac{1}{N}\|\widetilde{\mathbf{\Lambda}}\|\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|$ or $\frac{1}{N}\|\mathbf{\Lambda}\|^2\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\|$, as these bounds are too crude to achieve the desired rate of convergence when $N/T \to \infty$. More careful analysis is called for.

(i) Proving $\frac{1}{N}\widetilde{\mathbf{\Lambda}}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}) = o_P(T^{-1/2})$ is exactly the same as that of Lemma D.8. Note that replacing $\mathbf{\Lambda}$ with $\widetilde{\mathbf{\Lambda}}$ does not introduce any complications as $\mathbf{\Sigma}_u$ is a diagonal matrix. Hence the proof is omitted here to avoid repetitions.

(ii) For any $k, l \leq K$, the (k,l) element of $\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{\Lambda}$ is given by

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}(\widehat{\sigma}_{ii}-\sigma_{ii}) = \frac{1}{N}\frac{1}{T}\sum_{t}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}(u_{it}^2 - Eu_{it}^2) + \frac{1}{N}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}\frac{1}{T}\sum_{t}(\widehat{u}_{it}^2 - u_{it}^2)$$

As for the first term,

$$\begin{aligned} E|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}(u_{it}^2-\sigma_{ii})| &\leq \left[E\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}(u_{it}^2-\sigma_{ii})\right)^2\right]^{1/2} \\ = & \left[\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\lambda_{ik}\lambda_{il}\lambda_{jk}\lambda_{jl}\operatorname{cov}(u_{it}^2, u_{js}^2)\right]^{1/2} \\ = & \left[\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}^2\lambda_{il}^2\operatorname{var}(u_{it}^2)\right]^{1/2} = o(\frac{1}{\sqrt{T}}). \end{aligned}$$

As for the second term, we have

$$\begin{aligned} & |\frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}(\widehat{u}_{it}^2-u_{it}^2)| \leq 2|\frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}(\widehat{u}_{it}-u_{it})u_{it}| + |\frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}(\widehat{u}_{it}-u_{it})^2| \\ \leq & O_P(1)|\frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{f}}_t-\mathbf{f}_t)'\frac{1}{N}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}u_{it}\boldsymbol{\lambda}_i| + O_P(1)|\frac{1}{TN}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}(\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}'\boldsymbol{\lambda}_i)'\sum_{t=1}^{T}u_{it}\mathbf{f}_t| \\ & +O_P(1)|\frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}u_{it}(\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}'\boldsymbol{\lambda}_i)'(\widehat{\mathbf{f}}_t-\mathbf{H}^{-1}\mathbf{f}_t)| + O(1)\max_i|\frac{1}{T}\sum_{t=1}^{T}(\widehat{u}_{it}-u_{it})^2| \\ \leq & O_P(1)(\frac{1}{T}\sum_{t}\|\widehat{\mathbf{f}}_t-\mathbf{H}^{-1}\mathbf{f}_t\|^2)^{1/2}(\frac{1}{T}\sum_{t}\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}u_{it}\boldsymbol{\lambda}_i\|^2)^{1/2} \\ & +O_P(1)(\frac{1}{N}\sum_{i}\|\lambda_{ik}\lambda_{il}(\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}'\boldsymbol{\lambda}_i)\|^2)^{1/2}(\frac{1}{N}\sum_{i}\|\frac{1}{T}\sum_{t}u_{it}\mathbf{f}_t\|^2)^{1/2} + O(1)\max_i\frac{1}{T}\sum_{t=1}^{T}(\widehat{u}_{it}-u_{it})^2 \\ & +O_P(1)(\frac{1}{T}\sum_{t}\|\widehat{\mathbf{f}}_t-\mathbf{H}^{-1}\mathbf{f}_t\|^2)^{1/2}(\frac{1}{TN}\sum_{it}u_{it}^2)^{1/2}(\frac{1}{N}\sum_{i}\|\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}'\boldsymbol{\lambda}_i\|^2)^{1/2}. \end{aligned}$$

Note that $\frac{1}{T}\sum_t\|\widehat{\mathbf{f}}_t-\mathbf{H}^{-1}\mathbf{f}_t\|^2 = O_P(\psi_{NT}^2)$, $\max_i\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2 = O_P(\psi_{NT}^2+\frac{J\log N}{T})$ by Lemma E.3. Also, $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}'\boldsymbol{\lambda}_i\|^2 = O_P\left(\frac{J}{T}+\frac{1}{J^{2\eta-1}}+(\frac{\log N}{T})^{\zeta_1}J^3\right)$ by Theorem 3.1. In addition,

$$
\begin{aligned}
&E\frac{1}{T}\sum_t\|\frac{1}{N}\sum_{i=1}^N\lambda_{ik}\lambda_{il}u_{it}\boldsymbol{\lambda}_i\|^2=\sum_{m=1}^K E(\frac{1}{N}\sum_i\lambda_{ik}\lambda_{il}\lambda_{im}u_{it})^2\\
=\ &\sum_{m=1}^K\frac{1}{N^2}\sum_i\sum_j\lambda_{ik}\lambda_{il}\lambda_{im}\lambda_{jk}\lambda_{jl}\lambda_{jm}Eu_{it}u_{jt}=\sum_{m=1}^K\frac{1}{N^2}\sum_i\lambda_{ik}^2\lambda_{il}^2\lambda_{im}^2Eu_{it}^2=O(\frac{1}{N}),\\
&E\frac{1}{N}\sum_i\|\frac{1}{T}\sum_t u_{it}\mathbf{f}_t\|^2=\frac{1}{N}\sum_i\sum_k E(\frac{1}{T}\sum_t u_{it}f_{kt})^2=\frac{1}{N}\sum_i\sum_k\frac{1}{T^2}\sum_t Eu_{it}^2f_{kt}^2=O(\frac{1}{T}).
\end{aligned}
$$

Hence it is straightforward to verify that $|\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T\lambda_{ik}\lambda_{il}(\widehat{u}_{it}^2-u_{it}^2)|=o_P(T^{-1/2})$ so long as $T=o(N^2)$, $T=o(J^{2\eta-1}N)$, $J^4\log N=o(NT)$.

(iii) Let G_{ik} denote the (i,k) element of $\mathbf{G}$, and let δ_{ik} denote the (i,k) element of $\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}\mathbf{H}$. Since $\max_i\|\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i\|=o_P(1)$, we have $\max_{ik}|G_{ik}|=O_P(1)$, regardless of $\mathbf{G}=\boldsymbol{\Lambda}$ or $\mathbf{G}=\widehat{\boldsymbol{\Lambda}}$. Then the (l,k) element of the $K\times K$ matrix $\frac{1}{N}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}\mathbf{H})'(\widehat{\boldsymbol{\Sigma}}_u-\boldsymbol{\Sigma}_u)\mathbf{G}$ is bounded by

$$
|\frac{1}{N}\sum_{i=1}^N\delta_{il}G_{ik}\frac{1}{T}\sum_t(\widehat{u}_{it}^2-\sigma_{ii})|\leq\max_{ilk}|\delta_{il}G_{ik}|\frac{1}{NT}\sum_{i=1}^N|\sum_{t=1}^T(\widehat{u}_{it}^2-u_{it}^2)+(u_{it}^2-\sigma_{ii})|.
$$

On one hand, by Lemma E.3,

$$
\max_{ilk}|\delta_{il}G_{ik}|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T|\widehat{u}_{it}^2-u_{it}^2|=O_P(\psi_{NT}+\sqrt{\frac{J\log N}{T}})\max_i\|\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}'\boldsymbol{\lambda}_i\|=o_P(\frac{1}{\sqrt{T}}).
$$

On the other hand, $E|\sum_{t=1}^T(u_{it}^2-\sigma_{ii})|\leq\mathrm{var}(\sum_{t=1}^T(u_{it}^2-\sigma_{ii}))^{1/2}=(T\mathrm{var}(u_{it}^2))^{1/2}=O(T^{1/2})$. Hence

$$
\max_{ilk}|\delta_{il}G_{ik}|\frac{1}{NT}\sum_{i=1}^N|\sum_{t=1}^T(u_{it}^2-\sigma_{ii})|=O_P(\frac{1}{\sqrt{T}})\max_i\|\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}'\boldsymbol{\lambda}_i\|=o_P(\frac{1}{\sqrt{T}}).
$$

Lemma E.3. *Define*

$$
\psi_{NT}=\frac{1}{J^{\eta-1/2}}+\frac{1}{\sqrt{N}}+\frac{J^2(\log N\log J)^{1/2}}{T}+(\frac{\log N}{T})^{\zeta_1/2}J^2.
$$

Under H_0, when $N=O(T^2)$,
(i) $\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{f}}_t-\mathbf{H}^{-1}\mathbf{f}_t\|^2=O_P\left(\psi_{NT}^2\right)$.
(ii) $\max_i\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2=O_P(\psi_{NT}^2+\frac{J\log N}{T})$.

(iii) $\frac{1}{NT}\sum_i\sum_t |\widehat{u}_{it}^2 - u_{it}^2| = O_P(\psi_{NT} + \sqrt{\frac{J\log N}{T}})$.

Proof. (i) By Theorem 3.2, under H_0,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T \|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\|^2 &\leq 2\frac{1}{T}\sum_{t=1}^T \|\widehat{\mathbf{g}}(\mathbf{w}_t) - \mathbf{H}^{-1}\mathbf{g}(\mathbf{w}_t)\|^2 + 2\frac{1}{T}\sum_{t=1}^T \|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2 \\ &= O_P\left(\frac{1}{N} + \frac{J^4\log N\log J}{T^2} + \frac{1}{J^{2\eta-1}} + (\frac{\log N}{T})^{\zeta_1}J^4\right). \qquad \text{(E.2)}\end{aligned}$$

(ii) Uniformly in i, by Theorem 3.1,

$$\frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2 \leq C\frac{1}{T}\sum_t \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\|^2\|\widehat{\mathbf{f}}_t\|^2 + C\frac{1}{T}\sum_t \|\boldsymbol{\lambda}_i\|^2\|\widehat{\mathbf{f}}_t - \mathbf{f}_t\|^2 = O_P(\psi_{NT}^2).$$

(iii) We have, using $|a^2 - b^2| \leq |a-b||a+b|$ and the Cauchy-Schwarz inequality,

$$\begin{aligned}&(\frac{1}{NT}\sum_i\sum_t|\widehat{u}_{it}^2 - u_{it}^2|)^2 \leq \frac{1}{NT}\sum_{it}(\widehat{u}_{it} - u_{it})^2\frac{1}{NT}\sum_{it}(\widehat{u}_{it} + u_{it})^2 \\ \leq\ & \max_i\frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2\frac{1}{NT}\sum_{it}[2(\widehat{u}_{it} - u_{it})^2 + 4u_{it}^2] \\ \leq\ & 2(\max_i\frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2)^2 + 4\max_i\frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2\frac{1}{NT}\sum_{it}u_{it}^2 = O_P(\max_i\frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2) \\ =\ & O_P(\psi_{NT}^2 + \frac{J\log N}{T}).\end{aligned}$$

F Omitted identification details and Proofs for Section 5

F.1 Identification details of Section 5

Note that there is a particular $\mathbf{M}$ so that $\mathbf{M}^{-1}\mathbf{z}_t$ is identified, and can be consistently estimated by $\widehat{\mathbf{M}}^{-1}\widehat{\mathbf{z}}_t$, defined in (5.4). To introduce this transformation matrix $\mathbf{M}$, recall that

$$\boldsymbol{\Sigma}_F = E\{E(\mathbf{f}_t|\mathbf{w}_t)E(\mathbf{f}_t|\mathbf{w}_t)'\}, \qquad \boldsymbol{\Sigma} = E\{E(\mathbf{x}_t|\mathbf{w}_t)E(\mathbf{x}_t|\mathbf{w}_t)'\}.$$

Let $\mathbf{V}_N$ denote a $K\times K$ diagonal matrix, whose diagonal elements are the largest eigenvalues of $N^{-1}\boldsymbol{\Sigma}$, and let $\bar{\boldsymbol{\Gamma}}$ be a $K\times K$ matrix whose columns are the eigenvectors corresponding to the largest K eigenvalues of $N^{-1}\mathbf{\Sigma}_F^{1/2}\mathbf{\Lambda}'\mathbf{\Lambda}\mathbf{\Sigma}_F^{1/2}$. Define $\bar{\mathbf{H}} = \mathbf{\Sigma}_F^{1/2}\bar{\mathbf{\Gamma}}\mathbf{V}_N^{-1/2}$, and

$$\mathbf{M} = \mathbf{G}(\mathbf{G}^{-1}E(\mathbf{z}_t\mathbf{z}_t')\mathbf{G}'^{-1})^{1/2}, \quad \text{where } \mathbf{G} = \begin{pmatrix} \bar{\mathbf{H}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\dim(\mathbf{z}_t)} \end{pmatrix}.$$

We then define

$$\widetilde{\mathbf{z}}_t = \mathbf{M}^{-1}\mathbf{z}_t, \quad \widetilde{\boldsymbol{\psi}}_i = \mathbf{M}'\boldsymbol{\psi}_i, \quad i = 1, ..., L.$$

Here $E(\widetilde{\mathbf{z}}_t\widetilde{\mathbf{z}}_t') = \mathbf{I}$. The model (5.1) is then observationally identical to

$$Y_t = h(\widetilde{\boldsymbol{\psi}}_1'\widetilde{\mathbf{z}}_t, ..., \widetilde{\boldsymbol{\psi}}_L'\widetilde{\mathbf{z}}_t) + \varepsilon_t, \tag{F.1}$$

but now the regressors are normalized and identifiable.

The following proposition is based on an argument of Li (1991), which shows that if $(\widetilde{\boldsymbol{\psi}}_1, ..., \widetilde{\boldsymbol{\psi}}_L)$ is a set of orthogonal basis, then span$\{\widetilde{\boldsymbol{\psi}}_1, ..., \widetilde{\boldsymbol{\psi}}_L\}$ is identified as the space spanned by the eigenvectors of $\mathbf{\Sigma}_{z|y}$.

Assumption F.1. *Suppose: (i)* $\widetilde{\boldsymbol{\psi}}_i \neq 0$ *for* $i = 1, ..., L$*, and* $\widetilde{\boldsymbol{\psi}}_i'\widetilde{\boldsymbol{\psi}}_j = 0$ *if* $i \neq j$*.*
(ii) For any $\mathbf{v} \in \mathbb{R}^{\dim(\mathbf{z}_t)}$*, there are* $(a_1, ..., a_L) \in \mathbb{R}^L$ *that depends on* $\mathbf{v}$ *such that*

$$\mathbf{v}'E(\widetilde{\mathbf{z}}_t|\widetilde{\boldsymbol{\psi}}_1'\widetilde{\mathbf{z}}_t, ..., \widetilde{\boldsymbol{\psi}}_L'\widetilde{\mathbf{z}}_t) = \sum_{i=1}^{L} a_i\widetilde{\boldsymbol{\psi}}_i'\widetilde{\mathbf{z}}_t;$$

(iii) The rank of $\mathbf{\Sigma}_{z|y}$ *equals L.*
(iv) ε_t *is independent of* $\widetilde{\mathbf{z}}_t$*.*

Proposition F.1 (Identification of span$\{\widetilde{\boldsymbol{\psi}}_1, ..., \widetilde{\boldsymbol{\psi}}_L\}$, Li (1991))**.** *Let* $\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_L$ *be the eigenvectors of* $\mathbf{\Sigma}_{z|y}$*, corresponding to the nonzero eigenvalues. Suppose Assumption F.1 holds. Then*

$$\text{span}\{\widetilde{\boldsymbol{\psi}}_1, ..., \widetilde{\boldsymbol{\psi}}_L\} = \text{span}\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_L\}.$$

In the literature on dimension reduction (Li, 1991), Assumption F.1 (ii),(iii) are essential for identifying the sufficient dimension reduction space. Condition (ii) is satisfied, for instance, when $\mathbf{z}_t$ follows a multivariate Gaussian distribution or more generally elliptical distributions. Condition (iii) is known as the "exhaustive condition". We refer to Li and Wang (2007) for many mild sufficient conditions for (iii).

Proof of Proposition F.1

Let $A^{\perp}$ denote the orthogonal complement of a space A so that $\forall \mathbf{v} \in A^{\perp}$, and $\mathbf{a} \in A$,

$\mathbf{v}'\mathbf{a} = 0$. Now for any $\mathbf{v} \in \text{span}\{\widetilde{\boldsymbol{\psi}}_1, ..., \widetilde{\boldsymbol{\psi}}_L\}^{\perp}$, we aim to show $\mathbf{v} \in \text{span}\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_L\}^{\perp}$. Let

$$m := \mathbf{v}'E(\widetilde{\mathbf{z}}_t | \widetilde{\boldsymbol{\psi}}_1'\widetilde{\mathbf{z}}_t, ..., \widetilde{\boldsymbol{\psi}}_L'\widetilde{\mathbf{z}}_t) = \sum_{i=1}^{L} a_i \widetilde{\boldsymbol{\psi}}_i'\widetilde{\mathbf{z}}_t,$$

for some $(a_1, ..., a_L)$ by Assumption F.1. Since $\mathbf{v}'\widetilde{\boldsymbol{\psi}}_i = 0$, $i = 1...L$, and $E(\widetilde{\mathbf{z}}_t\widetilde{\mathbf{z}}_t') = \mathbf{I}$,

$$Em^2 = E(m\mathbf{v}'\widetilde{\mathbf{z}}_t) = \sum_{i=1}^{L} a_i \widetilde{\boldsymbol{\psi}}_i' E(\widetilde{\mathbf{z}}_t\widetilde{\mathbf{z}}_t')\mathbf{v} = 0.$$

Hence $m = 0$ almost surely. This implies, almost surely,

$$\mathbf{v}'E(\widetilde{\mathbf{z}}_t|y_{t+1}) = \mathbf{v}'E[E(\widetilde{\mathbf{z}}_t|y_{t+1}, \widetilde{\boldsymbol{\psi}}_1'\widetilde{\mathbf{z}}_t, ..., \widetilde{\boldsymbol{\psi}}_L'\widetilde{\mathbf{z}}_t)|y_{t+1}] = E[m|y_{t+1}] = 0.$$

Hence

$$\mathbf{v}'E(\widetilde{\mathbf{z}}_t|y_{t+1} \in I_h) = \frac{E[\mathbf{v}'E(\widetilde{\mathbf{z}}_t|y_{t+1})1\{y_{t+1} \in I_h\}]}{P(y_{t+1} \in I_h)} = 0.$$

This further implies $\mathbf{v}'\boldsymbol{\Sigma}_{z|y}\mathbf{v} = 0$. Now let $\{v_i, \boldsymbol{\xi}_i\}$ be the eigenvalue-vectors of $\boldsymbol{\Sigma}_{z|y}$, where $v_1...v_L$ are the nonzero eigenvalues. Then from the eigen-decomposition,

$$0 = \mathbf{v}'\boldsymbol{\Sigma}_{z|y}\mathbf{v} = \sum_{i=1}^{L} v_i(\mathbf{v}'\boldsymbol{\xi}_i)^2.$$

It follows that $\mathbf{v}'\boldsymbol{\xi}_i = 0$, $i = 1, ..., L$. Hence $\mathbf{v} \in \text{span}\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_L\}^{\perp}$. Finally, as $\widetilde{\boldsymbol{\psi}}_i$'s are orthogonal bases of $\text{span}\{\widetilde{\boldsymbol{\psi}}_1, ..., \widetilde{\boldsymbol{\psi}}_L\}$, the desired result follows.

F.2 Proof of Theorem 5.1

We divide our proof of the convergence for the space of $\widehat{\boldsymbol{\psi}}_i$'s into the follow steps:

step 1: the convergence of $\widehat{\mathbf{M}}^{-1}\widehat{\mathbf{z}}_t - \widetilde{\mathbf{z}}_t$. Recall that $\mathbf{G} = \text{diag}\{\bar{\mathbf{H}}, \mathbf{I}\}$, $\widetilde{\mathbf{z}}_t = \mathbf{M}^{-1}\mathbf{z}_t$. Also $\|\widehat{\mathbf{M}}^{-1}\| = O_P(1)$ and $\mathbf{M} = \mathbf{G}(\mathbf{G}^{-1}E(\mathbf{z}_t\mathbf{z}_t')\mathbf{G}'^{-1})^{1/2}$. So

$$\begin{aligned} \frac{1}{T}\sum_t \|\widehat{\mathbf{M}}^{-1}\widehat{\mathbf{z}}_t - \widetilde{\mathbf{z}}_t\|^2 &\leq \frac{2}{T}\sum_t \|\mathbf{M}^{-1}\mathbf{z}_t - \widehat{\mathbf{M}}^{-1}\mathbf{G}^{-1}\mathbf{z}_t\|^2 + \frac{2}{T}\sum_t \|\widehat{\mathbf{M}}^{-1}\mathbf{G}^{-1}\mathbf{z}_t - \widehat{\mathbf{M}}^{-1}\widehat{\mathbf{z}}_t\|^2 \\ &\leq 2\|\mathbf{M}^{-1} - \widehat{\mathbf{M}}^{-1}\mathbf{G}^{-1}\|^2 + O_P(1)\frac{1}{T}\sum_t \|\mathbf{G}^{-1}\mathbf{z}_t - \widehat{\mathbf{z}}_t\|^2. \end{aligned}$$

Since $\mathbf{G}^{-1}\mathbf{z}_t - \widehat{\mathbf{z}}_t = \bar{\mathbf{H}}^{-1}\mathbf{f}_t - \widehat{\mathbf{f}}_t$, and by Theorem 3.2,

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\|^2 = O_P\left(\frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|}{T} + \frac{1}{N} + \frac{J^4\log N\log J}{T^2} + \frac{1}{J^{2\eta-1}} + (\frac{\log N}{T})^{\zeta_1}J^4\right);$$

by Lemma D.13, $\|\mathbf{H}-\bar{\mathbf{H}}\| = O_P(a_{NT} + J^{-2\eta} + \frac{\log N}{T})$, hence $\frac{1}{T}\sum_t\|\mathbf{G}^{-1}\mathbf{z}_t - \widehat{\mathbf{z}}_t\|^2 = O_P\left(b_{NT}^2\right)$, where

$$b_{NT}^2 = \frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_s)\| + \log N}{T} + \frac{1}{N} + \frac{1}{J^{2\eta-1}}.$$

Also, since $\widehat{\mathbf{M}} = (\frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{z}}_t\widehat{\mathbf{z}}_t')^{1/2}$, and

$$\begin{aligned}
&\|\mathbf{G}^{-1}E(\mathbf{z}_t\mathbf{z}_t')\mathbf{G}'^{-1} - \frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{z}}_t\widehat{\mathbf{z}}_t'\|^2 \\
\leq\ & 2\|\mathbf{G}^{-1}E(\mathbf{z}_t\mathbf{z}_t')\mathbf{G}'^{-1} - \mathbf{G}^{-1}\frac{1}{T}\sum_t\mathbf{z}_t\mathbf{z}_t'\mathbf{G}'^{-1}\|^2 + 2\|\frac{1}{T}\sum_t(\mathbf{G}^{-1}\mathbf{z}_t\mathbf{z}_t'\mathbf{G}'^{-1} - \widehat{\mathbf{z}}_t\widehat{\mathbf{z}}_t')\|^2 = O_P(b_{NT}^2),
\end{aligned}$$

hence $\|\mathbf{M}^{-1} - \widehat{\mathbf{M}}^{-1}\mathbf{G}^{-1}\|^2 \leq \|(\mathbf{G}^{-1}E(\mathbf{z}_t\mathbf{z}_t')\mathbf{G}'^{-1})^{-1/2} - \widehat{\mathbf{M}}^{-1}\|^2 O_P(1) = O_P(b_{NT}^2)$. This implies

$$\frac{1}{T}\sum_t\|\widehat{\mathbf{M}}^{-1}\widehat{\mathbf{z}}_t - \widetilde{\mathbf{z}}_t\|^2 = O_P(b_{NT}).$$

step 2: the convergence of $\widehat{\boldsymbol{\Sigma}}_{z|y} - \boldsymbol{\Sigma}_{z|y}$. This step follows from the following results:

(i) $\|\frac{1}{T}\sum_t\widehat{\mathbf{M}}^{-1}\widehat{\mathbf{z}}_t 1\{y_{t+1}\in I_h\} - \frac{1}{T}\sum_t\widetilde{\mathbf{z}}_t 1\{y_{t+1}\in I_h\}\|^2 = O_P(b_{NT}^2)$, which follows from step 1 and the Cauchy-Schwarz inequality.

(ii) $\|\frac{1}{T}\sum_t\widetilde{\mathbf{z}}_t 1\{y_{t+1}\in I_h\} - E(\widetilde{\mathbf{z}}_t 1\{y_{t+1}\in I_h\})\|^2 = O_P(T^{-1})$.

(iii) $|\frac{1}{T}\sum_{t=1}^{T}1\{y_{t+1}\in I_h\} - P(y_{t+1}\in I_h)| = O_P(T^{-1/2})$, and

(iv) $P(y_{t+1}\in I_h) = H^{-1}$, which is bounded away from zero since $H = O(1)$.

Hence

$$\begin{aligned}
&\|E(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h) - \widehat{E}(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h)\| = \|\frac{\sum_{t=1}^{T}\widehat{\mathbf{M}}^{-1}\widehat{\mathbf{z}}_t 1\{y_{t+1}\in I_h\}}{\sum_{t=1}^{T}1\{y_{t+1}\in I_h\}} - \frac{E(\widetilde{\mathbf{z}}_t 1\{y_{t+1}\in I_h\})}{P(y_{t+1}\in I_h)}\| \\
=\ & O_P(b_{NT}).
\end{aligned}$$

It then follows that

$$\|\widehat{\boldsymbol{\Sigma}}_{z|y} - \boldsymbol{\Sigma}_{z|y}\| \leq \|\frac{1}{H}\sum_h[E(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h) - \widehat{E}(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h)]\widehat{E}(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h)'\|$$

$$+\|\frac{1}{H}\sum_h [E(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h) - \widehat{E}(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h)]E(\widetilde{\mathbf{z}}_t|y_{t+1}\in I_h)'\| = O_P(b_{NT}).$$

step 3: convergence of the eigenvectors. Let $\{v_i\}$ and $\{\widehat{v}_i\}$ respectively denote the eigenvalues of $\mathbf{\Sigma}_{z|y}$ and $\widehat{\mathbf{\Sigma}}_{z|y}$. Then By Lemma C.1, $|v_i - \widehat{v}_i| = O_P(b_{NT})$. Since $v_i, i = 1, ..., L$ are distinct and bounded away from zero, $\min(|\widehat{v}_{i-1} - v_i|, |v_i - \widehat{v}_{i+1}|)$ is bounded away from zero for $i = 1, ..., L$. Hence still by Lemma C.1,

$$\|\widehat{\boldsymbol{\psi}}_i - \boldsymbol{\xi}_i\| = \|\widehat{\mathbf{\Sigma}}_{z|y} - \mathbf{\Sigma}_{z|y}\| = O_P(b_{NT}).$$

References

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for high-dimensional data, methods, theory and applications*. The first edition ed. Springer, New York.

FAMA, E. F. and FRENCH, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116** 1–22.

FAN, J., FURGER, A. and XIU, D. (2015). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. *Journal of Business* & *E*conomic Statistics, to appear.

FAN, J., LI, Q. and WANG, Y. (2014). Robust estimation of high-dimensional mean regression. *arXiv preprint arXiv:1410.2150* .

GIBBONS, M., ROSS, S. and SHANKEN, J. (1989). A test of the efficiency of a given portfolio. *Econometrica* **57** 1121–1152.

LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* **86** 316–327.

LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association* **102** 997–1008.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.