Supplementary Appendix to

The Incidental Parameters Problem in Testing for Remaining Cross-section Correlation

by

Artūras Juodis and Simon Reese

January 2021

Contents

S.1 Empirical application

All of the empirical results presented assume the country-industry pair as the unit of analysis (panel group member i), of which we have $N = 119$. This panel is unbalanced where for Germany, Portugal, and Sweden the length of the available time-series is substantially shorter than for other countries.

Some of our theoretical results can be easily extended to cover unbalanced datasets. However in order to simplify our analysis we disregard from using the observations for Germany, Portugal, Sweden in the construction of the test statistics. Additionally, we remove two sectors from the Great Britain where observations for $t = 2004, 2005$ are missing. This way we are left with $N = 82$ and $T = 25$.

S.2 Additional MC results

This section reports additional results for the Monte Carlo experiments in Section 5 of the main paper. The five tables below are equivalents to Tables 1-5 and their underlying setup differs only in that the errors in $y_{i,t}$ and $x_{i,t}$ are drawn from a standardized $\chi^2(2)$ distribution with zero expected value and unit variance.

Table S.1: Sample moments of original CD test statistic for remaining CSD under $\mathbb{H}_0$

Part A: Application of CD to 2WFE residuals

	c_σ:	0.1	0.5	1	1.5	0.1	0.5	1	1.5
N	T	Mean				Variance $\times 100$			
25	25	-3.35	-3.31	-3.20	-2.97	1.28	1.77	3.94	9.92
25	50	-4.91	-4.87	-4.72	-4.43	0.54	0.89	2.70	8.46
25	100	-7.08	-7.02	-6.82	-6.43	0.23	0.60	2.71	9.99
25	200	-10.11	-10.02	-9.75	-9.21	0.10	0.49	3.35	14.57
50	25	-3.30	-3.26	-3.14	-2.91	1.03	1.38	2.91	7.33
50	50	-4.86	-4.80	-4.65	-4.33	0.38	0.65	1.99	6.37
50	100	-7.00	-6.94	-6.73	-6.32	0.14	0.36	1.69	6.41
50	200	-10.00	-9.91	-9.63	-9.06	0.06	0.28	1.94	8.66
100	25	-3.28	-3.24	-3.11	-2.86	0.88	1.16	2.35	5.87
100	50	-4.83	-4.78	-4.62	-4.30	0.26	0.45	1.38	4.50
100	100	-6.97	-6.90	-6.68	-6.26	0.09	0.23	1.06	4.22
100	200	-9.95	-9.86	-9.57	-8.99	0.03	0.16	1.07	4.77
200	25	-3.27	-3.23	-3.10	-2.85	0.73	0.96	1.95	4.90
200	50	-4.82	-4.76	-4.60	-4.28	0.21	0.35	1.01	3.24
200	100	-6.95	-6.88	-6.66	-6.23	0.07	0.17	0.75	2.95
200	200	-9.92	-9.83	-9.54	-8.94	0.02	0.11	0.71	3.21

Part B: Application of CD to CCE residuals

	c_σ:	0.1	0.5	1	1.5	0.1	0.5	1	1.5
N	T	Mean				Variance $\times 100$			
25	25	-3.35	-3.31	-3.21	-3.02	1.37	1.86	3.90	9.30
25	50	-4.92	-4.87	-4.74	-4.47	0.50	0.86	2.66	8.14
25	100	-7.08	-7.02	-6.84	-6.48	0.22	0.57	2.63	9.45
25	200	-10.11	-10.03	-9.77	-9.28	0.10	0.50	3.26	13.22
50	25	-3.28	-3.24	-3.13	-2.90	1.20	1.60	3.30	8.04
50	50	-4.85	-4.81	-4.66	-4.37	0.38	0.60	1.73	5.41
50	100	-7.00	-6.94	-6.74	-6.35	0.13	0.34	1.54	5.83
50	200	-10.00	-9.91	-9.64	-9.09	0.06	0.26	1.76	7.69
100	25	-3.23	-3.19	-3.07	-2.83	1.23	1.59	2.97	6.77
100	50	-4.82	-4.77	-4.61	-4.31	0.31	0.48	1.32	4.09
100	100	-6.97	-6.90	-6.69	-6.27	0.09	0.22	1.01	4.01
100	200	-9.95	-9.86	-9.57	-9.01	0.03	0.17	1.11	4.89
200	25	-3.16	-3.12	-3.00	-2.75	1.64	1.90	3.04	6.35
200	50	-4.80	-4.75	-4.59	-4.27	0.27	0.40	1.05	3.22
200	100	-6.94	-6.88	-6.67	-6.24	0.07	0.16	0.71	2.86
200	200	-9.92	-9.83	-9.54	-8.96	0.02	0.10	0.69	3.17

Notes. *$\varepsilon_{i,t}$ and $e_{i,t}$ are standardized $\chi^2(2)$ with zero expected value and unit variance. $\sigma^2_{i,y} = c_\sigma \left(\varsigma^2_{i,y} - 2\right)/4 + 1$; $\varsigma^2_{i,y}$ is $\chi^2(2)$. In part B, $r = 2$ and loadings $\boldsymbol{\lambda}_i$ are drawn from $U(0.5, 1.5)$. $\boldsymbol{\Lambda}_i$ has first element from $U(0.5, 1.5)$ and second from $U(-0.5, 0.5)$. Factors $\boldsymbol{f}_t$ drawn from $N(0, 1)$. In part A, we restrict $\boldsymbol{f}_t = \left[f_t^{(1)}, 1\right]'$, $\boldsymbol{\lambda}_i = \left[1, \lambda_i^{(2)}\right]'$ and $\boldsymbol{\Lambda}_i = \left[1, \Lambda_i^{(2)}\right]'$.*

Table S.2: Sample moments of original CD test statistic applied to 2WFE residuals under $\mathbb{H}_1$

	$\boldsymbol{\lambda}_i$:	Mean				Variance			
	$\boldsymbol{\lambda}_i$:	symmetric		skewed		symmetric		skewed	
	σ_i^2:	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$
N	T								
25	25	-3.37	-3.09	-3.30	-2.68	0.02	0.10	0.02	0.41
25	50	-4.86	-4.50	-4.78	-3.89	0.01	0.13	0.02	0.70
25	100	-6.95	-6.47	-6.86	-5.63	0.02	0.25	0.02	1.27
25	200	-9.88	-9.25	-9.76	-8.06	0.02	0.39	0.02	2.23
50	25	-3.32	-2.87	-3.25	-2.19	0.01	0.14	0.02	0.58
50	50	-4.81	-4.23	-4.73	-3.26	0.01	0.18	0.01	0.94
50	100	-6.88	-6.09	-6.78	-4.78	0.01	0.31	0.01	1.55
50	200	-9.78	-8.73	-9.65	-6.85	0.01	0.52	0.02	2.80
100	25	-3.30	-2.49	-3.23	-1.34	0.01	0.27	0.01	0.98
100	50	-4.78	-3.76	-4.70	-2.13	0.01	0.32	0.01	1.42
100	100	-6.83	-5.47	-6.73	-3.19	0.01	0.49	0.01	2.41
100	200	-9.73	-7.84	-9.59	-4.65	0.01	0.83	0.01	4.14
200	25	-3.30	-1.80	-3.22	0.33	0.01	0.56	0.01	1.97
200	50	-4.77	-2.88	-4.68	0.07	0.01	0.60	0.01	2.68
200	100	-6.82	-4.28	-6.71	-0.12	0.01	0.94	0.01	4.25
200	200	-9.70	-6.22	-9.56	-0.43	0.01	1.48	0.01	7.28

Notes. *The model has a factor error structure with 3 factors.* $\varepsilon_{i,t}$ *and* $e_{i,t}$ *are standardized* $\chi^2(2)$ *with zero expected value and unit variance. "*σ_i^2*:* $\perp \boldsymbol{\lambda}_i$*" means that* $\sigma_{i,y}^2 = \left(\varsigma_{i,y}^2 - 2\right)/4 + 1$ *where* $\varsigma_{i,y}^2$ *is* $\chi^2(2)$*. For "*σ_i^2*:* $f(\boldsymbol{\lambda}_i)$*", we let* $\sigma_{i,y}^2 = d_\sigma T^{-1} \sum_{t=1}^{T} (\boldsymbol{\lambda}_i' \boldsymbol{f}_t)^2$.
The case "$\boldsymbol{\lambda}_i$*: symmetric" corresponds to drawing* $\boldsymbol{\lambda}_i$ *from* $U(0.5, 1.5)$*. "*$\boldsymbol{\lambda}_i$*: skewed" loadings are from a standardized* $\chi^2(2)$ *distribution with mean and variance equal to 1.*
$\boldsymbol{\Lambda}_i$ *has first element from* $U(0.5, 1.5)$ *and all others from* $U(-0.5, 0.5)$*. Factors* $\boldsymbol{f}_t$ *are drawn from* $N(0, 1)$*.*

Table S.3: Sample moments of original CD test statistic applied to CCE residuals under $\mathbb{H}_1$

		Mean				Variance			
	$\boldsymbol{\lambda}_i$:	symmetric		skewed		symmetric		skewed	
	σ_i^2:	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$	$\perp \boldsymbol{\lambda}_i$	$f(\boldsymbol{\lambda}_i)$
N	T								
25	25	-3.27	-3.21	-3.25	-3.01	0.03	0.04	0.03	0.19
25	50	-4.78	-4.72	-4.76	-4.42	0.02	0.03	0.02	0.25
25	100	-6.87	-6.80	-6.86	-6.40	0.02	0.04	0.02	0.41
25	200	-9.80	-9.70	-9.78	-9.16	0.03	0.06	0.02	0.68
50	25	-3.21	-3.14	-3.19	-2.93	0.02	0.04	0.02	0.15
50	50	-4.71	-4.65	-4.70	-4.37	0.02	0.02	0.01	0.17
50	100	-6.78	-6.70	-6.78	-6.32	0.01	0.03	0.01	0.26
50	200	-9.68	-9.58	-9.67	-9.04	0.02	0.04	0.01	0.42
100	25	-3.16	-3.10	-3.14	-2.90	0.02	0.03	0.02	0.10
100	50	-4.67	-4.61	-4.66	-4.34	0.01	0.02	0.01	0.11
100	100	-6.73	-6.66	-6.73	-6.28	0.01	0.02	0.01	0.16
100	200	-9.62	-9.52	-9.61	-8.99	0.01	0.03	0.01	0.27
200	25	-3.11	-3.05	-3.09	-2.84	0.02	0.03	0.02	0.09
200	50	-4.66	-4.58	-4.64	-4.31	0.01	0.02	0.01	0.09
200	100	-6.71	-6.64	-6.71	-6.26	0.01	0.02	0.01	0.11
200	200	-9.59	-9.48	-9.58	-8.96	0.01	0.02	0.01	0.21

Notes. *See Table S.2.*

Table S.4: Rejection rates for weighted CD test statistic when applied to 2WFE residuals

Part A: Size

	$\boldsymbol{\lambda}_i$:	symmetric						skewed					
	σ_i^2:	$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$			$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$		
N	T	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}
25	25	4.9	5.4	2.5	5.1	5.6	2.9	5.0	5.9	2.5	6.1	6.6	2.7
25	50	6.5	7.0	3.9	6.8	7.4	3.9	5.7	6.5	3.5	6.2	6.8	3.3
25	100	8.8	9.0	5.0	8.0	8.2	4.6	8.4	8.8	5.1	8.0	8.1	3.9
25	200	10.8	10.9	5.7	9.7	9.8	6.1	11.0	11.0	7.0	10.9	11.1	4.3
50	25	5.1	5.9	2.8	5.7	6.2	3.6	5.7	6.7	2.3	5.1	5.7	2.9
50	50	5.0	6.6	4.3	4.6	5.9	4.0	6.5	7.6	3.4	5.4	6.8	3.1
50	100	6.6	7.1	3.9	5.9	6.7	3.2	6.6	7.0	4.6	5.9	6.5	3.5
50	200	6.6	6.8	4.7	5.7	5.9	4.4	5.8	6.2	4.4	6.1	6.3	3.5
100	25	5.3	5.8	2.8	5.3	5.5	3.1	4.6	4.7	3.5	5.0	5.4	3.7
100	50	5.9	8.7	5.1	4.5	6.7	4.1	4.6	7.4	3.9	4.8	7.4	4.0
100	100	5.6	6.6	4.5	5.2	7.0	4.3	4.9	6.1	5.0	4.2	5.9	2.8
100	200	5.8	6.3	4.0	5.5	6.2	4.8	5.5	5.9	3.1	5.9	6.3	4.1
200	25	4.4	4.5	2.9	4.8	4.8	3.5	5.0	5.1	3.2	4.8	4.9	3.1
200	50	4.6	7.4	3.6	4.8	8.6	3.8	4.5	8.0	3.7	5.8	8.7	4.0
200	100	4.7	8.3	4.2	5.5	9.7	4.7	5.2	8.6	4.1	5.7	9.4	4.1
200	200	4.2	5.4	5.0	5.4	7.4	5.0	5.9	6.8	5.2	4.9	6.6	3.6

Part B: Power

	$\boldsymbol{\lambda}_i$:	symmetric						skewed					
	σ_i^2:	$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$			$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$		
N	T	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}
25	25	12.6	26.6	8.0	10.7	14.0	5.8	12.3	19.6	14.7	14.3	20.0	24.0
25	50	21.2	87.4	14.8	17.5	46.1	12.7	18.0	59.6	32.9	23.2	63.7	48.4
25	100	34.6	99.9	26.4	27.3	95.2	20.9	32.3	90.8	50.1	36.3	96.7	66.6
25	200	50.4	100	39.7	43.6	100	34.7	44.1	99.3	64.4	51.3	100	80.4
50	25	11.7	21.1	8.1	10.7	11.7	5.7	11.7	20.9	35.2	12.6	16.0	58.9
50	50	19.6	98.3	15.7	17.5	56.0	12.8	18.7	85.5	59.5	23.0	79.7	86.0
50	100	35.4	100	27.0	26.3	99.7	22.2	29.7	99.8	80.2	37.5	100	95.4
50	200	49.3	100	40.6	43.5	100	37.4	45.3	100	90.7	53.2	100	99.1
100	25	11.1	14.4	6.5	10.6	10.9	5.1	10.4	19.1	73.4	12.0	12.9	93.5
100	50	19.9	99.9	15.0	16.3	66.4	13.3	18.1	98.9	92.9	22.3	94.6	99.7
100	100	33.3	100	25.8	26.2	100	23.8	30.8	100	98.8	36.7	100	100
100	200	48.7	100	41.3	41.9	100	37.0	46.2	100	100	52.3	100	100
200	25	12.7	12.7	8.2	10.8	10.8	6.6	11.9	13.9	96.5	12.9	13.0	100
200	50	18.9	100	16.6	17.9	73.1	13.3	18.8	100	100	22.1	98.2	100
200	100	33.9	100	27.4	26.3	100	21.5	32.8	100	100	36.9	100	100
200	200	50.9	100	41.9	43.6	100	37.2	46.7	100	100	54.6	100	100

Notes. *In Part A the model has two factors which are restricted as noted in Table 1. Part B corresponds to a model with factor error structure and 3 factors. For details on all other model parameters, see Table S.2.*
CD_W is the weighted CD test statistic introduced in Theorem 3. CD_{W+} is its power-enhanced refinement. CD_{BC} is a CD test statistic with analytic bias correction.

Table S.5: Rejection rates for weighted CD test statistic when applied to CCE residuals

Part A: Size

	$\boldsymbol{\lambda}_i$:	symmetric						skewed					
	σ_i^2:		$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$			$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$	
N	T	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}
25	25	6.2	6.8	4.6	5.7	6.4	4.4	5.6	6.1	4.7	5.6	6.6	5.9
25	50	7.4	8.1	5.7	5.8	6.6	5.9	7.0	7.9	6.1	6.7	7.3	8.9
25	100	8.3	8.5	6.6	8.0	7.9	7.8	9.0	9.7	6.4	8.2	8.8	14.5
25	200	9.2	10.1	7.4	11.3	11.6	8.8	10.9	12.2	9.5	13.0	14.0	24.2
50	25	5.1	5.8	4.1	4.4	5.0	4.8	5.9	6.6	4.9	5.5	6.7	8.0
50	50	5.1	6.5	5.1	6.1	7.0	4.2	5.5	7.0	5.8	7.2	8.0	8.1
50	100	5.5	6.0	4.7	5.2	5.5	6.3	6.3	6.6	4.7	6.6	7.2	11.8
50	200	6.4	6.5	5.9	6.9	6.9	6.7	6.0	6.6	6.9	7.4	7.9	21.9
100	25	5.2	5.4	5.9	5.5	5.7	6.7	5.4	5.8	6.6	5.5	5.7	8.9
100	50	5.0	7.4	4.3	5.3	8.3	5.9	5.3	7.0	4.5	5.1	6.9	8.0
100	100	5.5	6.6	5.1	5.1	7.0	5.2	5.8	7.0	5.4	5.1	6.9	11.3
100	200	5.0	5.5	4.8	6.0	6.2	5.1	5.2	5.9	5.7	5.8	6.1	19.2
200	25	5.8	5.8	11.7	5.1	5.1	12.7	6.3	6.3	11.7	4.9	5.0	16.3
200	50	6.3	9.4	4.5	5.6	9.7	5.3	5.0	9.1	5.0	5.6	9.8	9.5
200	100	5.4	8.5	5.5	5.1	8.3	5.8	4.9	8.3	4.9	5.1	8.4	10.8
200	200	5.9	7.0	5.3	4.8	5.9	5.9	4.8	6.7	4.7	5.0	7.1	15.8

Part B: Power

	$\boldsymbol{\lambda}_i$:	symmetric						skewed					
	σ_i^2:		$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$			$\perp \boldsymbol{\lambda}_i$			$f(\boldsymbol{\lambda}_i)$	
N	T	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}	CD_W	CD_{W+}	CD_{BC}
25	25	10.4	16.3	5.7	9.2	10.7	5.6	8.5	13.4	6.2	9.7	12.8	7.1
25	50	15.0	47.6	9.9	10.2	19.8	8.9	12.5	37.4	11.8	15.2	31.4	14.5
25	100	22.6	83.9	14.0	17.6	50.6	13.2	20.5	64.4	18.7	23.3	67.0	24.8
25	200	33.6	97.6	22.5	26.0	85.7	22.5	28.2	86.6	28.8	34.4	91.1	42.4
50	25	9.7	12.6	5.7	7.8	8.5	4.9	8.7	15.0	5.9	8.6	10.3	7.4
50	50	11.5	67.0	8.6	11.5	23.5	8.2	10.9	58.8	9.4	13.5	42.7	14.9
50	100	20.6	98.4	14.7	14.7	68.5	13.9	18.1	91.6	21.6	21.2	90.4	28.1
50	200	28.0	100	23.5	24.3	98.6	22.9	26.6	99.0	32.1	33.2	99.6	43.0
100	25	9.3	9.9	4.8	7.5	7.8	5.5	9.0	13.0	6.2	9.0	9.3	8.5
100	50	12.6	80.4	6.4	9.3	25.4	7.6	12.2	82.9	11.9	13.1	54.9	15.9
100	100	18.2	100	12.5	15.2	85.8	12.8	17.6	99.7	21.8	22.2	98.9	29.5
100	200	29.4	100	21.9	23.6	99.9	24.3	28.6	100	36.2	29.7	100	46.7
200	25	9.4	9.7	7.9	8.1	8.1	6.6	7.5	8.1	7.6	10.7	10.8	8.6
200	50	12.1	90.8	9.4	10.9	33.4	7.2	11.9	95.2	11.2	12.8	64.5	17.2
200	100	20.0	100	13.0	14.8	96.3	13.0	16.6	100	21.8	21.1	99.8	29.8
200	200	27.2	100	22.6	23.5	100	21.7	26.6	100	42.4	31.0	100	49.0

Notes. *In Part A the model has two factors which are restricted as noted in Table S.1. Part B corresponds to a model with factor error structure and 3 factors. For details on all other model parameters, see Table S.2.*
CD_W is the weighted CD test statistic introduced in Theorem 3. CD_{W+} is its power-enhanced refinement. CD_{BC} is a CD test statistic with analytic bias correction.

S.3 Additional theoretical contributions

S.3.1 Power analysis: Additive

To simplify the discussion, we disregard from the effect of covariates on the variable of interest $y_{i,t}$ and assume that the true model is given by a pure static factor model, amounting to model (16) with $\boldsymbol{\beta} = \mathbf{0}$. Extending our results to the case of general $\boldsymbol{\beta}$ is possible, although formally cumbersome.[1] In a first instance, suppose that a researcher erroneously assumes cross-section dependence to stem from time fixed effects. The corresponding mis-specified model is formally given by

$$y_{i,t} = \mu_{y,t} + \nu_{i,t}, \tag{S.1}$$

where $\mu_{y,t} = \boldsymbol{f}_t' \operatorname{E}[\boldsymbol{\lambda}_i]$ and $\nu_{i,t} = \boldsymbol{f}_t' (\boldsymbol{\lambda}_i - \operatorname{E}[\boldsymbol{\lambda}_i]) + \tilde{\varepsilon}_{i,t}$. Deviations of the data from their cross-section averages can accordingly be written as

$$\begin{aligned} \widehat{\nu}_{i,t} &= y_{i,t} - N^{-1} \sum_{i=1}^{N} y_{i,t} \\ &= \boldsymbol{f}_t' \tilde{\boldsymbol{\lambda}}_i + \tilde{\varepsilon}_{i,t} \end{aligned} \tag{S.2}$$

where $\tilde{\boldsymbol{\lambda}}_i = \boldsymbol{\lambda}_i - N^{-1} \sum_{i=1}^{N} \boldsymbol{\lambda}_i$ and $\tilde{\varepsilon}_{i,t} = \varepsilon_{i,t} - N^{-1} \sum_{i=1}^{N} \varepsilon_{i,t}$. These deviations are sample equivalents of the composite error term $\nu_{i,t}$ in the mis-specified time fixed effects model (S.1). Let the variance of this composite error be defined as $\varsigma_{\nu,i}^2 = \operatorname{E}\left[\nu_{i,t}^2 \mid \sigma_i, \boldsymbol{\lambda}_i\right] = (\boldsymbol{\lambda}_i - \operatorname{E}[\boldsymbol{\lambda}_i])' \boldsymbol{\Sigma}_F (\boldsymbol{\lambda}_i - \operatorname{E}[\boldsymbol{\lambda}_i]) + \sigma_i^2$ and let $\overline{\varsigma_\nu^k} = N^{-1} \sum_{i=1}^{N} \varsigma_{\nu,i}^k,\ k \in \mathbb{Z}$ be shorthand notation for cross-section averages of $\varsigma_{\nu,i}^k$. Finally, define the stacked vector of the residual error term as $\boldsymbol{\nu}_i = [\nu_{i,1}, \nu_{i,2}, \ldots, \nu_{i,T}]'$.

For the sake of simplicity, we assume that $\varsigma_{\nu,i}^2$ is known and used to standardize regression residuals when constructing the CD test statistic. This greatly simplifies the proofs of the following proposition while leaving the main results of this section qualitatively unaffected (as we formally show for results under $\mathbb{H}_0$ in the main text). We confirm this conjecture with corresponding simulations in Section 5, allowing for both unknown error variances and unknown, general slope coefficients $\boldsymbol{\beta}$. Given our current setup, let the CD test statistic constructed from $\widehat{\nu}_{i,t}$ be expressed by

$$CD_{\mathbb{H}_1} = \sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \varsigma_{\nu,i}^{-1} \widehat{\nu}_{i,t} \widehat{\nu}_{j,t} \varsigma_{\nu,j}^{-1}. \tag{S.3}$$

The properties of this test statistic are characterized as follows:

[1] To appreciate this point note that, even if the unobserved heterogeneity driving the data is mis-specified, least squares estimates of $\boldsymbol{\beta}$ are consistent for the pseudo-true parameter vector $\boldsymbol{\delta} = \operatorname*{plim}_{N,T\to\infty} \left[\left(\sum_{i=1}^{N} \check{\boldsymbol{X}}_i' \check{\boldsymbol{X}}_i \right)^{-1} \left(\sum_{i=1}^{N} \check{\boldsymbol{X}}_i' \check{\boldsymbol{y}}_i \right) \right]$ with $\check{\boldsymbol{X}}_i$ and $\check{\boldsymbol{y}}_i$ being orthogonal to estimates of the assumed sources of unobserved heterogeneity.

Proposition S.1. *Suppose that the true model is given by* (S.1) *and that its components satisfy Assumptions 2–6 and $\varsigma_{\nu,i} \in [\delta; M]$. Let the CD test statistic $CD_{\mathbb{H}_1}$ be constructed from the cross-sectionally demeaned data* (S.2)*. Then,*

$$CD_{\mathbb{H}_1} = \sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \left(\varsigma_{\nu,i}^{-1} - \mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right] \right) \boldsymbol{\nu}_i' \boldsymbol{\nu}_j \left(\varsigma_{\nu,j}^{-1} - \mathrm{E}\left[\varsigma_{\nu,j}^{-1}\right] \right) + \sqrt{T}\Xi_{\mathbb{H}_1} + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right), \tag{S.4}$$

where

$$\Xi_{\mathbb{H}_1} = \sqrt{\frac{N}{2(N-1)}} \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{\nu}_i' \boldsymbol{\nu}_i \left[\left(\overline{\varsigma_\nu^{-1}} \right)^2 - 2\overline{\varsigma_\nu^{-1}} \varsigma_{\nu,i}^{-1} \right].$$

Furthermore, the leading stochastic term $\sqrt{\frac{1}{2TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \left(\varsigma_{\nu,i}^{-1} - \mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right] \right) \boldsymbol{\nu}_i' \boldsymbol{\nu}_j \left(\varsigma_{\nu,j}^{-1} - \mathrm{E}\left[\varsigma_{\nu,j}^{-1}\right] \right)$ has expected value $N\sqrt{T}\,\mathrm{cov}\left[\varsigma_{\nu,1}^{-1}, \boldsymbol{\lambda}_1'\right] \boldsymbol{\Sigma}_F\, \mathrm{cov}\left[\boldsymbol{\lambda}_1, \varsigma_{\nu,1}^{-1}\right]$. In the special case $\mathrm{cov}\left[\boldsymbol{\lambda}_1, \varsigma_{\nu,1}^{-1}\right] = \mathbf{0}$ the order of this leading term is determined by its variance which diverges at rate T.

Proposition S.1 establishes a decomposition of the CD test statistic that is almost identical to that obtained under the null hypothesis: Analogous to Theorem 1, $CD_{\mathbb{H}_1}$ consists of a leading stochastic component which reflects a CD test statistic with incorrect normalization as well as an equivalent to the bias term in Eq. (12) which we denote by $\sqrt{T}\Xi_{\mathbb{H}_1}$. The key difference is that both components are functions of the composite model error $\nu_{i,t} = \boldsymbol{f}_t' \left(\boldsymbol{\lambda}_i - \mathrm{E}\left[\boldsymbol{\lambda}_i\right]\right) + \varepsilon_{i,t}$ rather than the true errors $\varepsilon_{i,t}$.

Noticeable differences with regards to Theorem 1 are given as far as the leading stochastic component of $CD_{\mathbb{H}_1}$, the first term in Eq. (S.4), is concerned. Its mean is a positive-definite quadratic form and will therefore generally diverge to $+\infty$ at rate $N\sqrt{T}$. This result is interesting in the context of Sarafidis et al. (2009) who conjecture "*[...] that the CD test will have poor power properties when it is applied to a regression with time dummies or on cross-sectionally demeaned data*". First, the cited statement does not acknowledge the bias term $\Xi_{\mathbb{H}_1}$ which generally leads $CD_{\mathbb{H}_1}$ to diverge. However, it is reasonable not to consider this term as a source of power towards the alternative hypothesis since it does not involve sample estimates of cross-section covariances (see Remark S.2 below). Second, Proposition S.1 indicates that the CD test indeed has power since remaining sources of co-movements across cross-sections generally affect the mean of its leading stochastic component. The failure to appreciate this property is rooted in a convention to treat error variances as fixed parameters, thereby ruling out any correlation between them and factor loadings. Extending this understanding of error variances to the variance of the composite error term $\nu_{i,t}$ (i.e. $\varsigma_{\nu,i}^2$) leads to the special case $\mathrm{cov}\left[\boldsymbol{\lambda}_1, \varsigma_{\nu,1}^{-1}\right] = \mathbf{0}$ in which the power of CD is indeed considerably reduced.

Still, while the rate at which $CD_{\mathbb{H}_1}$ diverges is in general unaffected by the inclusion of time fixed effects, power in small samples may be compromised because of the presence of two diverging components with opposite signs in Eq. (S.4). While the mean of the leading stochastic component dominates in large samples, it may be cancelled out by $\Xi_{\mathbb{H}_1}$ if the number of observations available to the researcher is rather small.

Remark S.1. It is straightforward to extend the results of Proposition S.1 to a model specification without time fixed effects and to address a common concern about the power of CD that has repeatedly been made in the literature. As argued in Pesaran et al. (2008) and Sarafidis et al. (2009), the power of this test is reduced substantially if unaccounted sources of cross-section correlation average out, as for example in a factor model with zero mean factor loadings. This property is given by construction for two-way fixed effects or time fixed effects residuals by the mere fact that the within transformation involves cross-sectionally demeaning the data. While the bias term in Proposition S.1 and the scaling effect on the leading stochastic component in $CD_{\mathbb{H}_1}$ are exclusively a consequence of accounting for time fixed effects, the defining role of $\operatorname{cov}\left[\boldsymbol{\lambda}_1, \varsigma_{\nu,1}^{-1}\right]$ for the rate at which $CD_{\mathbb{H}_1}$ diverges is a general property in models with zero mean loadings. It follows that existing claims about the power losses of CD in the presence of factor with zero mean loadings merely address the special case with $\operatorname{cov}\left[\boldsymbol{\lambda}_1, \varsigma_{\nu,1}^{-1}\right] = \mathbf{0}$. In fact, Monte Carlo results that have been reported to support these claims, as e.g. in Pesaran (2004, 2015), Sarafidis et al. (2009) and Baltagi et al. (2012), all use specifications which entail a zero covariance between factor loadings and the inverse error variances.[2] The CD test statistic is very likely to have much better power properties in simulation experiments that enforce a non-zero correlation between factor loadings and error variances. A simple way of achieving this would be to impose a constant ratio between the magnitude of common variation and that of idiosyncratic variation which holds for all cross-sections. See Parker and Sul (2016, Eq. (27)) or Section 5 in this study for a formal definition.

Remark S.2. An additional similarity of Proposition S.1 with Theorem 1 is that the term $\Xi_{\mathbb{H}_1}$ does not involve sample estimates of cross-section covariances, implying that this term is not indicative of the degree of cross-section co-movement in the data. Moreover, it can be shown that $\Xi_{\mathbb{H}_1} \xrightarrow{p} \left(\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right]\right)^2 \mathrm{E}\left[\varsigma_{\nu,i}^{2}\right] - 2\,\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right] \mathrm{E}\left[\varsigma_{\nu,i}\right]$, so that the leading term in $\sqrt{T}\Xi_{\mathbb{H}_1}$ would approach $-\sqrt{T/2}$ as $\varsigma_{\nu,i} \to \varsigma_\nu$, for some constant $\varsigma_\nu > 0$.

S.3.2 Power analysis: Multifactor

A mis-specified latent common factor model can be characterized by equation (16) such that Assumptions 2–5 as well as the following Assumption S.1 hold.

Assumption S.1. $\operatorname{rk}\left([\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \boldsymbol{\mu}_{\boldsymbol{\Lambda}}]\right) = m + 1 < r$.

[2] This follows because all cited articles draw $\boldsymbol{\lambda}_i$ and σ_i independently of each other and assume a symmetric distribution for $\boldsymbol{\lambda}_i$. Using the integer representation of the expected value it is possible to show that $\operatorname{cov}\left[\boldsymbol{\lambda}_1, \varsigma_{\nu,1}^{-1}\right] = \mathbf{0}$ in this case.

Assumption S.1 enforces a failure of the rank condition set up by Pesaran (2006) to ensure that the space spanned by factor estimates can be consistent for the space spanned by the true factors. Under failure of the rank condition, a fraction of the common variation affecting the dependent variable $y_{i,t}$ remains asymptotically unaccounted for. As proved in Lemma S.18, we can assume without loss of generality that accounted and unaccounted sources of cross-section dependence are due to two uncorrelated sets of unobserved factors. Formally, we decompose

$$\boldsymbol{F}\boldsymbol{\lambda}_i = \boldsymbol{F}^{(1)}\boldsymbol{\lambda}_i^{(1)} + \boldsymbol{F}^{(2)}\boldsymbol{\lambda}_i^{(2)}, \tag{S.5}$$

where a rotation of the $m+1$ factors $\boldsymbol{F}^{(1)}$ is consistently estimated by cross-section averages $\widehat{\boldsymbol{F}} = [\overline{\boldsymbol{y}}, \overline{\boldsymbol{X}}]$ whereas the remaining $r-m-1$ factors $\boldsymbol{F}^{(2)}$ are asymptotically orthogonal to $\widehat{\boldsymbol{F}}$. An analogous decomposition can be applied to the matrix product $\boldsymbol{F}\boldsymbol{\Lambda}_i$, allowing us to split the loadings matrix into two blocks $\boldsymbol{\Lambda}_i^{(1)}$ and $\boldsymbol{\Lambda}_i^{(2)}$. Most importantly, this decomposition implies that $\mathrm{E}\left[\boldsymbol{\lambda}_i^{(2)}\right] = \mathbf{0}$ and $\mathrm{E}\left[\boldsymbol{\Lambda}_i^{(2)}\right] = \mathbf{O}$.

We can accordingly express the defactored data as

$$\begin{aligned}\widehat{\boldsymbol{\nu}}_i &= \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{y}_i \\ &= \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{F}^{(1)}\boldsymbol{\lambda}_i^{(1)} + \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{F}^{(2)}\boldsymbol{\lambda}_i^{(2)} + \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i.\end{aligned} \tag{S.6}$$

where Assumption S.1 and Lemma S.18 ensure that $\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{F}^{(2)}\boldsymbol{\lambda}_i^{(2)}$ does not vanish as the sample size increases. Again, we denote the variance of the composite error term $\nu_{i,t} = \boldsymbol{f}_t^{(2)\prime}\boldsymbol{\lambda}_i^{(2)} + \varepsilon_{i,t}$ by $\varsigma_{\nu,i}^2 = \mathrm{E}\left[\nu_{i,t}^2 \mid \boldsymbol{\lambda}_i^{(2)}, \sigma_i^2\right] = \boldsymbol{\lambda}_i^{(2)\prime}\boldsymbol{\Sigma}_{F,22}\boldsymbol{\lambda}_i^{(2)} + \sigma_i^2$, where the $r-m-1$-dimensional square matrix $\boldsymbol{\Sigma}_{F,22}$ is the lower-right block of $\boldsymbol{\Sigma}_F$. The order in probability of the CD test statistic can then be characterized as follows.

Proposition S.2. *Suppose that the true model is given by* (S.1) *and that its components satisfy Assumptions 2–5 and S.1, and* $\varsigma_{\nu,i} \in [\delta; M]$. *Let the CD test statistic* $CD_{\mathbb{H}_1}$, *defined in* (S.3), *be constructed from defactored data* (S.6). *Then,*

$$CD_{\mathbb{H}_1} = \sqrt{\frac{2}{TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j + \sqrt{T}\left(\Phi_{1,\mathbb{H}_1} - 2\Phi_{2,\mathbb{H}_1}\right) + \mathcal{O}_P\left(N^{-1/2}\sqrt{T}\right) + \mathcal{O}_P\left(T^{-1/2}\right) \tag{S.7}$$

where

$$\begin{aligned}\boldsymbol{\varpi}_i &= \left[\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i - \boldsymbol{D}_i\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right], \\ \Phi_{1,\mathbb{H}_1} &= \left(N^{-1}\sum_{\ell=1}^{N}\varsigma_{\nu,\ell}^{-1}\boldsymbol{\lambda}_\ell^{(1)\prime}\right)\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{D}_i'\boldsymbol{D}_i\right)\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right), \\ \Phi_{2,\mathbb{H}_1} &= N^{-1}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{D}_i\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right),\end{aligned}$$

where $\overline{\boldsymbol{C}}^{(1)} = \left[\overline{\boldsymbol{\lambda}}^{(1)},\ \overline{\boldsymbol{\Lambda}}^{(1)}\right]$ *and* $\boldsymbol{D}_i = \boldsymbol{F}\left[\boldsymbol{\lambda}_i^{(2)},\ \boldsymbol{\Lambda}_i^{(2)}\right] + \left[\boldsymbol{\varepsilon}_i,\ \boldsymbol{e}_i\right]$ *. Furthermore, it holds that*

$$\underset{N,T\to\infty}{\text{plim}}\ N^{-1}T^{-1/2}\sqrt{\frac{2}{TN\left(N-1\right)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j = \text{cov}\left[\varsigma_{\nu,1}^{-1},\ \boldsymbol{\lambda}_1^{(2)\prime}\right]\boldsymbol{\Sigma}_{F,22}\,\text{cov}\left[\boldsymbol{\lambda}_1^{(2)},\ \varsigma_{\nu,1}^{-1}\right].$$

In the special case $\text{cov}\left[\varsigma_{\nu,1}^{-1},\ \boldsymbol{\lambda}_1^{(2)\prime}\right] = \boldsymbol{0}$ *, the term* $\sqrt{\frac{2}{TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j$ *is asymptotically centered around zero and has a variance that diverges at rate* T*.*

The remarks made concerning Proposition S.1 apply in an identical fashion to Proposition S.2. That is, the CD test statistic allows for the same type of decomposition under the null hypothesis as it does under the alternative, differing only in that the expressions in Proposition S.2 contain the composite error term $\nu_{i,t}$ rather than the true model errors $\varepsilon_{i,t}$.

S.3.3 Weighted CD statistic and power enhancement

Asymptotic unbiasedness of CD_W comes at the cost of power. More specifically, our approach to bias correction centers the leading components of CD_W around zero, irrespective of whether cross-section correlation in the data is completely controlled for or not. As a consequence, only increases in the variance of CD_W under its alternative hypothesis lead to power against the null hypothesis of this test.

This point can be formally illustrated by a minor modification of Proposition S.1 which consists of replacing $\varsigma_{\nu,i}^{-1} - \text{E}\left[\varsigma_{\nu,i}^{-1}\right]$ with w_i. Given random weights, the leading stochastic component in Eq. (S.4),

$$\sqrt{\frac{2}{TN\left(N-1\right)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}w_i\boldsymbol{\nu}_i'\boldsymbol{\nu}_j w_j, \tag{S.8}$$

has an asymptotic mean of zero if $N^{-1}\sqrt{T} \to 0$ since independent Rademacher distributed weights ensure that $\text{E}\left[w_i\nu_{i,t}\nu_{j,t}w_j\right] = 0$ holds for all $i \neq j$. Thus, remaining sources of cross-section co-movement in the composite error term $\nu_{i,t}$ will not shift the location of expression (S.8). By contrast, it can be shown that the variance of this term continues to diverge at rate T. This entails that CD_W diverges at rate $\sqrt{T}$ under its alternative hypothesis, a property that holds under the conditions of Proposition S.1 as well.

Using Lemma S.5 we can establish that:

$$\widehat{\rho}_{ij} = 0 + \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P((NT)^{-1/2}) + \mathcal{O}_P(N^{-1}), \tag{S.9}$$

for all pairs i, j. The threshold $2\sqrt{\ln(N)/T}$ is justified by the behavior of the maximum out of $N(N-1)/2$ estimated correlation coefficients in a model with $\rho_{ij} = 0$ for all i, j. Under the assumption that a CLT holds for $\widehat{\rho}_{ij}$ and that $\sqrt{T}N^{-1} \to 0$ as $N, T, \to \infty$, this maximum should diverge at rate $2\sqrt{\ln(N)/T}$. Multiplication of each indicator function in the screening statistic

$\sum_{i=2}^{N}\sum_{j=1}^{i-1}|\widehat{\rho}_{ij}|\,\mathbf{1}\left(|\widehat{\rho}_{ij}|>2\sqrt{\ln(N)/T}\right)$ with the absolute value of its corresponding correlation coefficient then ensures that the power enhancement component in (32) converges to 0 if $\rho_{ij}=0$ for all pairs i,j, subject to additional tail regularity conditions.

S.4 Further discussions

S.4.1 An analytically bias-corrected CD statistic

Section 4.1 introduced parametric bias-correction as a feasible approach to re-establish asymptotically normal inference for the CD tests statistic under its null hypothesis. This section provides details on its implementation as a benchmark CD test statistic in the Monte Carlo experiments of Section 5. The analytically bias-corrected CD test statistic CD_{BC} is corrected with a plug-in estimate of the bias terms of Theorems 1 and 2. That is, whenever CD_{BC} is applied to 2WFE residuals, we construct it as

$$CD_{BC}=\widehat{\Omega}_{FE}^{-1/2}\left(CD-\sqrt{T}\widehat{\Xi}\right)$$

where

$$\widehat{\Xi}=\sqrt{\frac{1}{2N(N-1)}}\left[\sum_{i=1}^{N}\left(1-\sqrt{\widehat{\sigma}_i^2}N^{-1}\sum_{\ell=1}^{N}\left(\widehat{\sigma}_\ell^2\right)^{-1/2}\right)^2-N\right],$$
$$\widehat{\Omega}_{FE}=\frac{2\,(N-1)}{N}\left(\widehat{\Xi}+1\right)^2$$

with $\widehat{\sigma}_i^2=T^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_{i,t}^2$ for 2WFE residuals $\widehat{\varepsilon}_{i,t}$. The expressions above arise from rearranging the terms of Theorem 1 so that their estimates can be obtained with minimal computational burden. In this regard, it is helpful to note that the asymptotic variance of Ω can be written as a function of the bias term Ξ. Likewise, for application of CD_{BC} to CCE residuals, we set

$$CD_{BC}=\widehat{\Omega}_{CCE}^{-1/2}\left[CD-\sqrt{T}\left(\widehat{\Phi}_1-2\widehat{\Phi}_2\right)\right],$$

where

$$\widehat{\Phi}_1=\sqrt{\frac{1}{2N(N-1)}}\left[N^{-1}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2\right)^{-1/2}\right]^2\overline{\widehat{\boldsymbol{\lambda}}}'\widehat{\boldsymbol{B}}'\begin{bmatrix}\sum_{i=1}^{N}\widehat{\sigma}_i^2, & \mathbf{0}_m'\\ \mathbf{0}_m, & \sum_{i=1}^{N}\widehat{\boldsymbol{\Sigma}}_i\end{bmatrix}\widehat{\boldsymbol{B}}\overline{\widehat{\boldsymbol{\lambda}}},$$
$$\widehat{\Phi}_2=\sqrt{\frac{1}{2N(N-1)}}\left[N^{-1}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2\right)^{-1/2}\right]\overline{\widehat{\boldsymbol{\lambda}}}'\begin{bmatrix}\sum_{i=1}^{N}\widehat{\sigma}_i^2\\ \mathbf{0}_m\end{bmatrix},$$
$$\widehat{\Omega}_{CCE}=\left[1+\sqrt{\frac{2\,(N-1)}{N}}\left(\widehat{\Phi}_1-2\widehat{\Phi}_2\right)\right]^2$$

with

$$
\begin{aligned}
\widehat{\boldsymbol{B}} &= \begin{bmatrix} 1, & \mathbf{0}'_m \\ \widehat{\boldsymbol{\beta}}^{CCE}, & \boldsymbol{I}_m \end{bmatrix}, \\
\widehat{\boldsymbol{\beta}}^{CCE} &= \left(\frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{X}'_i \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \boldsymbol{X}_i \right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{X}'_i \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \boldsymbol{y}_i \right), \\
\overline{\widehat{\boldsymbol{\lambda}}} &= \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i, \\
\widehat{\boldsymbol{\lambda}}_i &= \left(\widehat{\boldsymbol{F}}' \widehat{\boldsymbol{F}} \right)^{-1} \widehat{\boldsymbol{F}}' \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}^{CCE} \right), \\
\widehat{\boldsymbol{\Sigma}}_i &= \frac{1}{T} \boldsymbol{X}'_i \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \boldsymbol{X}_i,
\end{aligned}
$$

as well as $\widehat{\sigma}_i^2 = T^{-1} \widehat{\varepsilon}'_i \widehat{\varepsilon}_i$ and $\widehat{\varepsilon}_i = \boldsymbol{M}_{\widehat{\boldsymbol{F}}}(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}^{CCE})$. The plug-in estimators $\widehat{\Phi}_1$ and $\widehat{\Phi}_2$ are calculated under the assumption of independence between factor loadings and error variances so that we can write $\mathrm{E}\left[\boldsymbol{\lambda}_i \sigma_i^{-1}\right] = \mathrm{E}\left[\boldsymbol{\lambda}_i\right] \mathrm{E}\left[\sigma_i^{-1}\right]$. While this assumption allows making the estimates of both bias terms sufficiently precise in the setup of our Monte Carlo experiments, it may lead to size distortion in a DGP with dependence between factor loadings and error variances.

S.4.2 Serial correlation

The original CD test of Pesaran (2004, 2015) assumes that idiosyncratic error terms are serially uncorrelated, an assumption that can be difficult to justify for most economic datasets. While this problem is mostly ignored in practice, Baltagi et al. (2016) have very recently proposed a modification of the CD test statistic which ensures that the test statistic is asymptotically standard normal as long as $\varepsilon_{i,t}$ is a stationary short memory process. In particular, under this type of assumption it can be shown that the relevant asymptotic variance for the result in Theorem 3 is given by:

$$
\Omega = \sum_{s=-\infty}^{\infty} \left[\mathrm{E}[(w_i - \mathrm{E}[w_i])^2 \varepsilon_{i,t} \varepsilon_{i,t-s}] \right]^2. \tag{S.10}
$$

The above quantity is non-standard, as it not a function of the long-run variance of $(w_i - \mathrm{E}[w_i])\varepsilon_{i,t}$. In particular, one can use data which is over-differenced in the construction of the CD statistic.

In the context of our testing problem the natural plug-in estimator of Ω is given by

$$
\widehat{\Omega}_N = \frac{2}{TN(N-1)} \sum_{i=2}^{N} \sum_{j<i} (\boldsymbol{l}'_i \boldsymbol{l}_j)^2, \tag{S.11}
$$

where $\boldsymbol{l}_i = (w_i - \overline{w}) \widehat{\boldsymbol{e}}_i$ is constructed using residuals $\widehat{\boldsymbol{e}}_i$. It can be expected that under reasonable regularity conditions on the memory properties of $\varepsilon_{i,t}$ as well as appropriate restrictions on N, T estimator $\widehat{\Omega}_N$ is consistent in our setup. However, we do not attempt to prove this conjecture as it does not add to the main message of this paper.

Note that Baltagi et al. (2016) use mean adjustment variance estimate of Chen and Qin (2010), which is motivated by the need to obtain an unbiased (not only consistent) estimator of Ω. However, as in our setting $\boldsymbol{x}_i$ are correlated by construction, any theoretical justification for including such an adjustment term under null hypothesis is lost. Also note that a factor of 2 is missing in Baltagi et al. (2016).

S.4.3 Optimal weights

Propositions S.1 and S.2 stated that the original CD test statistic diverges at rate $N\sqrt{T}$ if certain conditions on the dependence between error variances and loadings associated with unaccounted common factors hold. A simple generalization of Proposition S.2 to the case of general weights $\{w_1, \dots, w_N\}$ suggest that the same rate of divergence can be achieved under the condition $\text{cov}\left[w_i,\ \boldsymbol{\lambda}_i^{(2)}\right] \neq \mathbf{0}$.[3]

Consequently, a set of weights that generally leads to high power would be given by functions of the data that are estimates of $\boldsymbol{\lambda}_i^{(2)}$ under $\mathbb{H}_1$, such as $w_i = T^{-1}\sum_{t=1}^{T}\widehat{v}_{i,t}$ or $w_{ij} = T^{-1}\sum_{t=1}^{T}\widehat{v}_{i,t}\widehat{v}_{j,t}$ for a more general statistic with index-pair specific weights. However, the latter weight suggestion directly results in a different test statistic, namely one based on *squared* cross-section covariances. This new test statistic, as well as one based on the first suggestion $w_i = T^{-1}\sum_{t=1}^{T}\widehat{v}_{i,t}$, is closer to an existing, separate literature on LM tests for cross-section dependence than it is to the CD test and its extant modifications. The fact that a test statistic based on summing squared cross-section covariances is not centered around zero under its null hypothesis, and that it hence needs additional recentering, further emphasizes its relation to LM tests.

As an alternative to estimates of $\boldsymbol{\lambda}_i^{(2)}$, estimates of other model components may be used to set up a weighted CD test statistic. In this regard, the most sensible choice is an estimate of loadings associated with unaccounted factors in the covariates of our regression model of interest, i.e. the parameter matrix $\boldsymbol{\Lambda}_i^{(2)}$ in

$$\boldsymbol{X}_i = \boldsymbol{F}^{(1)}\boldsymbol{\Lambda}_i^{(1)} + \boldsymbol{F}^{(2)}\boldsymbol{\Lambda}_i^{(2)} + \boldsymbol{E}_i.$$

This choice directs the power of CD_W towards alternatives under which loadings associated with unaccounted factors are correlated in the sense of Westerlund and Urbain (2013) and Kapetanios et al. (2019). Following the theoretical results in these two studies, a CD test statistic that weights cross-section covariances with estimates of $\boldsymbol{\Lambda}_i^{(2)}$ is effectively a test for inconsistency of either the two-way fixed effects or CCE estimator of $\boldsymbol{\beta}$. It is generally possible to direct the CD test statistic towards these more specific hypothesis. However, if one is to test whether an estimate of the slope coefficients $\boldsymbol{\beta}$ on $\boldsymbol{X}_i$ are inconsistent, it is simpler and more sensible to

[3] Analogous results hold for a generalization of Proposition S.1

do this using a test statistic that addresses the moment condition $\mathrm{cov}\left[\boldsymbol{\lambda}_i^{(2)},\ \mathrm{vec}\left(\boldsymbol{\Lambda}_i^{(2)}\right)\right] = \mathbf{O}$ directly.

S.5 Proofs

S.5.1 Notation

Extending the paragraph on notation in introduction in the main text, we will use the following notation in this appendix.

- $\boldsymbol{I}_m$ denotes an $m \times m$ identity matrix and the subscript is sometimes disregarded from for the sake of simplicity. $\mathbf{0}$ denotes a column vector of zeros while $\mathbf{O}$ stands for a matrix of zeros. $\boldsymbol{s}_m$ denotes a selection vector all of whose elements are zero except for element m which is one. $\boldsymbol{\iota}$ is a vector entirely consisting of ones. The dimension of these latter vectors and matrices is generally suppressed for the sake of simplicity and needs to be inferred from context.

- For a generic $m \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{P_A} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ projects onto the space spanned by the columns of $\boldsymbol{A}$ and $\boldsymbol{M_A} = \boldsymbol{I}_m - \boldsymbol{P_A}$. $\dim(\boldsymbol{A}), \mathrm{col}(\boldsymbol{A})$ and $\ker(\boldsymbol{A})$ refer to the dimension, column space and kernel of $\boldsymbol{A}$. Moreover, $\mathrm{rk}(\boldsymbol{A})$ denotes the rank of $\boldsymbol{A}$, $\mathrm{tr}(\boldsymbol{A})$ its trace and $\|\boldsymbol{A}\| = (\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A}))^{1/2}$ the Frobenius norm of $\boldsymbol{A}$.

- For a set of $m \times n$ matrices $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_N\}$, $\overline{\boldsymbol{A}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{A}_i$. Multiple sums are generally abbreviated, so that $\sum_{i,j}^{N}$ is shorthand notation for $\sum_{i=1}^{N}\sum_{j=1}^{N}$.

- δ and M stand for a small and large positive real number, respectively. For two real numbers a and b, $a \vee b = \max\{a, b\}$.

- For some random variable $\varepsilon_{i,t}$, $\kappa_4\left[\varepsilon_{i,t}\right]$ denotes its fourth-order cumulant. Moreover, $\mathcal{O}(\cdot)$ and $\mathrm{o}(\cdot)$ express order of magnitude relations whereas $\mathcal{O}_P(\cdot)$ and $\mathrm{o}_P(\cdot)$ denote stochastic order relations, see e.g. Definitions 2.5 and 2.33 in White (2001).

- Define $k_{N,T} = \sqrt{\frac{1}{2TN(N-1)}}$.

- Notice that we use the following vector and scalar notation for idiosyncratic components: $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \varepsilon_{i,2}, \ldots, \varepsilon_{i,T})'$, $\overline{\boldsymbol{\varepsilon}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i$, and $\overline{\varepsilon} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}$.

Additional notation in the multiplicative model

Proofs that address a model with multifactor error structure which is controlled for via cross-section averages frequently employ the rotation and rescaling matrix

$$\boldsymbol{B} = \begin{bmatrix} 1, & \mathbf{0}' \\ \boldsymbol{\beta}, & \boldsymbol{I}_m \end{bmatrix}$$

which relates the $T \times (m+1)$ matrix $[\boldsymbol{y}_i, \boldsymbol{X}_i]$ (or alternatively the average over all i) to common and idiosyncratic variation affecting each of the observed variables directly. This relation is given by

$$\begin{bmatrix} \boldsymbol{y}_i, & \boldsymbol{X}_i \end{bmatrix} = \left(\boldsymbol{F} \begin{bmatrix} \boldsymbol{\lambda}_i, & \boldsymbol{\Lambda}_i \end{bmatrix} + \begin{bmatrix} \varepsilon_i, & \boldsymbol{E}_i \end{bmatrix} \right) \boldsymbol{B}$$
$$= \boldsymbol{F}\boldsymbol{C}_i + \boldsymbol{U}_i$$

In the following, it is assumed that $\boldsymbol{\beta}$ is bounded so that $\|\boldsymbol{B}\| < \infty$. Furthermore, $\boldsymbol{B}$ is non-singular by construction, implying that $\boldsymbol{B}^{-1}$ exists. Given these properties, it trivially follows from the properties of $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ that $\|\overline{\boldsymbol{C}}\| = \mathcal{O}_P(1)$ and $\|\overline{\boldsymbol{C}}^{-1}\| = \mathcal{O}_P(1)$. The same holds for the matrix $\overline{\boldsymbol{C}}^{(1)}$ which is considered in the proof of Proposition S.2.

S.5.2 Proofs of Theorem 1 and Corollary 1

Proof of Theorem 1. We start the proof by focusing on the individual components of the double sum over cross-sections which constitutes the CD test statistic. Here, we apply a first-order Taylor expansion around the true error variances. Using the mean-value theorem, we can hence set up the equality

$$\frac{\widehat{\varepsilon}_{i,t}\widehat{\varepsilon}_{j,t}}{\sqrt{\widehat{\sigma}_i^2\widehat{\sigma}_j^2}} = \frac{\widehat{\varepsilon}_{i,t}\widehat{\varepsilon}_{j,t}}{\sqrt{\sigma_i^2\sigma_j^2}} - \frac{1}{2}\frac{\widehat{\varepsilon}_{i,t}\widehat{\varepsilon}_{j,t}}{\left(\varsigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) - \frac{1}{2}\frac{\widehat{\varepsilon}_{i,t}\widehat{\varepsilon}_{j,t}}{\left(\sigma_i^2\right)^{1/2}\left(\varsigma_j^2\right)^{3/2}}\left(\widehat{\sigma}_j^2 - \sigma_j^2\right)$$

for some $\left(\varsigma_i^2, \varsigma_j^2\right)$ on the line interval between $\left(\sigma_i^2, \sigma_j^2\right)$, and $\left(\widehat{\sigma}_i^2, \widehat{\sigma}_j^2\right)$. This Taylor approximation has two extra terms. However, both of them can be combined into one single extra term once we sum over all possible combinations of i and j. Hence, we can write

$$CD = 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\sqrt{\sigma_i^2\sigma_j^2}} - k_{N,T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\varsigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right)$$
$$= CD_\sigma + CD_{(\widehat{\sigma}-\sigma)}. \tag{S.12}$$

Consider now CD_σ. Since model residuals can be written out as $\widehat{\varepsilon}_{i,t} = \varepsilon_{i,t} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon}$, we can further decompose this term into

$$CD_\sigma = \sqrt{\frac{2}{TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t=1}^{T}\frac{(\varepsilon_{i,t}-\bar{\varepsilon}_t)(\varepsilon_{j,t}-\bar{\varepsilon}_t)}{\sigma_i\sigma_j} - Tk_{N,T}\left[\left(\sum_{i=1}^{N}\frac{\bar{\varepsilon}_i-\bar{\varepsilon}}{\sigma_i}\right)^2 - \sum_{i=1}^{N}\left(\frac{\bar{\varepsilon}_i-\bar{\varepsilon}}{\sigma_i}\right)^2\right]$$
$$= CD_{\mu,\sigma} - CD_{(\widehat{\mu}-\mu)}. \tag{S.13}$$

As shown formally in Section S.5.6, $CD_{(\widehat{\mu}-\mu)} = \mathcal{O}_P\left(T^{-1/2}\right)$ and $CD_{(\widehat{\sigma}-\sigma)} = \mathcal{O}_P(\sqrt{N}T^{-1}) + \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T})$. Accordingly, terms that capture the effect of estimating fixed effects and error variances have an asymptotically negligible effect if some weak restrictions on the relative rate of divergence of N and T are satisfied. For this reason, $CD_{\mu,\sigma}$ is the leading component of the CD test statistic. It can be expressed in terms of four U statistics, and two additional terms that contribute to the bias. First, let

$$CD_{\varepsilon} = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t},$$

$$CD_{\varepsilon/\sigma} = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \frac{1}{T} \sum_{t=1}^{T} \frac{\varepsilon_{i,t}\varepsilon_{j,t}}{\sigma_i\sigma_j},$$

$$CD_{\varepsilon+} = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t}(\sigma_i^{-1} + \sigma_j^{-1}),$$

$$CD_{\varepsilon++} = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t}(\sigma_i^{-2} + \sigma_j^{-2}).$$

Now observe that the leading term of CD can be decomposed into

$$\begin{aligned}
CD_{\mu,\sigma} &= \sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \frac{(\varepsilon_{i,t} - \bar{\varepsilon}_t)(\varepsilon_{j,t} - \bar{\varepsilon}_t)}{\sigma_i\sigma_j} \\
&= k_{N,T} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \frac{(\varepsilon_{i,t} - \bar{\varepsilon}_t)}{\sigma_i} \right)^2 - k_{N,T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{\varepsilon_{i,t} - \bar{\varepsilon}_t}{\sigma_i} \right)^2 \\
&- k_{N,T} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \frac{(\bar{\varepsilon}_i - \bar{\varepsilon})}{\sigma_i} \right)^2 + k_{N,T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{\bar{\varepsilon}_i - \bar{\varepsilon}}{\sigma_i} \right)^2 \\
&= I - II - III + IV.
\end{aligned}$$

Let us now consider each term separately. First,

$$\begin{aligned}
I &= k_{N,T} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \frac{\varepsilon_{i,t}}{\sigma_i} \right)^2 - 2k_{N,T}\overline{\sigma^{-1}} \sum_{t=1}^{T} \left[\left(\sum_{i=1}^{N} \frac{\varepsilon_{i,t}}{\sigma_i} \right) \left(\sum_{i=1}^{N} \varepsilon_{i,t} \right) \right] + k_{N,T} \left(\overline{\sigma^{-1}} \right)^2 \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \varepsilon_{i,t} \right)^2 \\
&= CD_{\varepsilon/\sigma} - \overline{\sigma^{-1}} CD_{\varepsilon+} + \left(\overline{\sigma^{-1}} \right)^2 CD_{\varepsilon} \\
&+ k_{N,T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{\varepsilon_{i,t}}{\sigma_i} \right)^2 - 2k_{N,T}\overline{\sigma^{-1}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_i \left(\frac{\varepsilon_{i,t}}{\sigma_i} \right)^2 + k_{N,T} \left(\overline{\sigma^{-1}} \right)^2 \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{i,t}^2.
\end{aligned}$$

Similarly, for the second term we have

$$\begin{aligned}
II &= k_{N,T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{\varepsilon_{i,t}}{\sigma_i} \right)^2 - 2k_{N,T}N^{-1} \sum_{t=1}^{T} \left[\left(\sum_{i=1}^{N} \frac{\varepsilon_{i,t}^2}{\sigma_i^2} \right) \left(\sum_{i=1}^{N} \varepsilon_{i,t} \right) \right] + k_{N,T}N^{-1}\overline{\sigma^{-2}} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \varepsilon_{i,t} \right)^2 \\
&= -N^{-1}CD_{\varepsilon++} + N^{-1}\overline{\sigma^{-2}}CD_{\varepsilon} \\
&+ k_{N,T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{\varepsilon_{i,t}}{\sigma_i} \right)^2 - 2k_{N,T}N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{\varepsilon_{i,t}}{\sigma_i} \right)^2 + k_{N,T}N^{-1}\overline{\sigma^{-2}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\varepsilon_{i,t})^2.
\end{aligned}$$

As for the third component III,

$$\begin{aligned}
III &= Tk_{N,T}\left(\sum_{i=1}^{N}\frac{(\bar{\varepsilon}_i-\bar{\varepsilon})}{\sigma_i}\right)^2\\
&= Nk_{N,T}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}\right)^2\\
&= \sqrt{\frac{N}{2T(N-1)}}\left(\mathcal{O}_P(1)-\mathcal{O}_P(1)\right)^2\\
&= \mathcal{O}_P(T^{-1/2}).
\end{aligned}$$

Here the third line follows from Theorem 3 in Phillips and Moon (1999) applied to sequences $\varepsilon_{i,t}$ and $\varepsilon_{i,t}/\sigma_i$.

As for the fourth component IV, notice that:

$$\begin{aligned}
IV &= Tk_{N,T}\sum_{i=1}^{N}\left(\frac{\bar{\varepsilon}_i-\bar{\varepsilon}}{\sigma_i}\right)^2\\
&= k_{N,T}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2-2k_{N,T}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}\right)\\
&+k_{N,T}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)\left(\sqrt{NT}\bar{\varepsilon}\right)^2\\
&= k_{N,T}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2-\mathcal{O}_P(N^{-1}T^{-1/2}).
\end{aligned}$$

Here in the final use we use the same CLT from Theorem 3 in Phillips and Moon (1999). For the final component notice that:

$$\begin{aligned}
k_{N,T}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2 &= Nk_{N,T}+\sqrt{N}k_{N,T}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2-1\right)\\
&= \mathcal{O}(T^{-1/2})+\mathcal{O}_P(N^{-1/2}T^{-1/2}).
\end{aligned}$$

It follows that $IV=\mathcal{O}_P(T^{-1/2})$.

Combining these expressions yields the expression in terms of four U-statistics and two bias terms referred to above:

$$\begin{aligned}
CD_{\mu,\sigma} &= I-II-III+IV\\
&= CD_{\varepsilon/\sigma}-2\overline{\sigma^{-1}}CD_{\varepsilon+}-2N^{-1}CD_{\varepsilon++}+\left(\left(\overline{\sigma^{-1}}\right)^2-N^{-1}\overline{\sigma^{-2}}\right)CD_{\varepsilon}\\
&+k_{N,T}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2\left(\left(\overline{\sigma^{-1}}\right)^2-2\overline{\sigma^{-1}}\sigma_i^{-1}\right)-N^{-1}k_{N,T}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2\left(\left(\overline{\sigma^{-1}}\right)^2-2\sigma_i^{-2}\right)+\mathcal{O}_P(T^{-1/2}).
\end{aligned}\tag{S.14}$$

Concerning the second bias term in the last line above, we note that $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2 = \mathcal{O}_P(1)$ and $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2\sigma_i^{-2} = \mathcal{O}_P(1)$ by application of Markov's inequality. It follows that

$$N^{-1}k_{N,T}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2\left(\left(\overline{\sigma^{-1}}\right)^2 - 2\sigma_i^{-2}\right) = \mathcal{O}_P\left(N^{-1}\sqrt{T}\right).$$

Concerning the four U-statistics, note that under Assumption 1 the error variances σ_i^2 are bounded, implying that all averages involving σ_i are in general of order $\mathcal{O}_P(1)$ by Markov's inequality. Lemma S.1 can hence be used to show that all four U-statistics $CD_\varepsilon, CD_{\varepsilon/\sigma}, CD_{\varepsilon+}$ and $CD_{\varepsilon++}$ are of order $\mathcal{O}_P(1)$.[4] As a consequence, all terms in the second line of Eq. (S.14) involving N^{-1} are $\mathcal{O}_P(N^{-1})$. Additionally, a standard (cross-sectional) Lindeberg-Levi CLT implies that

$$\overline{\sigma^{-1}} = \mathrm{E}[\sigma_i^{-1}] + \mathcal{O}_P(N^{-1/2}).$$

This allows us to write

$$\begin{aligned} &CD_{\varepsilon/\sigma} - 2\overline{\sigma^{-1}}CD_{\varepsilon+} + \left(\overline{\sigma^{-1}}\right)^2 CD_\varepsilon \\ &= 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t=1}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\left(\sigma_i^{-1} - \mathrm{E}[\sigma_i^{-1}]\right)\left(\sigma_j^{-1} - \mathrm{E}[\sigma_i^{-1}]\right) + \mathcal{O}_P(N^{-1/2}). \end{aligned}$$

Combining our results on the components of equation (S.14), we conclude that

$$\begin{aligned} CD_{\mu,\sigma} = &\sqrt{\frac{2}{TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t=1}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\left(\sigma_i^{-1} - \mathrm{E}[\sigma_i^{-1}]\right)\left(\sigma_j^{-1} - \mathrm{E}[\sigma_i^{-1}]\right) \\ &+ \sqrt{T}\Xi + \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(T^{-1/2}), \end{aligned}$$

where

$$\Xi = \sqrt{\frac{N}{2(N-1)}}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2\left(\left(\overline{\sigma^{-1}}\right)^2 - 2\overline{\sigma^{-1}}\sigma_i^{-1}\right).$$

Adding the remaining two components of the decomposition of CD in Eqs. (S.13) and (S.12) merely results in the additional order terms $\mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right)$. It remains to prove weak convergence of the leading stochastic component in CD. For this purpose, set $a_{i,t} = \varepsilon_{i,t}(\sigma_i^{-1} - \mathrm{E}[\sigma_i^{-1}])$ and $q_i = \sigma_i(\sigma_i^{-1} - \mathrm{E}[\sigma_i^{-1}])$. We can then apply Lemma S.1 and conclude that

$$CD - \sqrt{T}\Xi \xrightarrow{d} N(0,\Omega), \tag{S.15}$$

as $N, T \to \infty$ subject to $\sqrt{N}T^{-1} \to 0$ and $\sqrt{T}N^{-1} \to 0$, where $\Omega = (\mathrm{E}[q_i^2])^2 = \left(\mathrm{E}[(1 - \sigma_i\,\mathrm{E}[\sigma_i^{-1}])^2]\right)^2$. □

[4] Strictly speaking, Lemma S.1 does not apply to $CD_{\varepsilon+}$ and $CD_{\varepsilon++}$. However, it can be straightforwardly extended to accommodate the specific structure of the U-statistics in both expressions .

Proof of Corollary 1. Using the decomposition in Eqs. (S.12)- (S.13) for a general value of σ^2, we have

$$CD = CD_{\mu,\sigma} + CD_{(\widehat{\sigma}-\sigma)} - CD_{(\widehat{\mu}-\mu)}, \tag{S.16}$$

where

$$CD_{\mu,\sigma} = \frac{1}{\sigma^2}\sqrt{\frac{1}{2TN(N-1)}}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\widehat{\varepsilon}_{i,t}\right)^2 - \frac{1}{\sigma^2}\sqrt{\frac{1}{2TN(N-1)}}\sum_{t=1}^{T}\sum_{i=1}^{N}\widehat{\varepsilon}_{i,t}^2. \tag{S.17}$$

The first component on the right-hand side above is 0 by construction. The second can be rewritten as

$$\begin{aligned}
CD_{\mu,\sigma} &= \frac{1}{N}CD_{\varepsilon/\sigma} - \frac{1}{\sigma^2}\sqrt{\frac{2}{TN(N-1)}}\frac{N-1}{2N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\varepsilon_{i,t}^2\right) \\
&= \frac{1}{N}CD_{\varepsilon/\sigma} - \sqrt{\frac{N-1}{2N^2}}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\left(\frac{\varepsilon_{i,t}}{\sigma}\right)^2 - 1\right) - \sqrt{\left(T-\frac{T}{N}\right)/2}, \\
&= -\sqrt{\left(T-\frac{T}{N}\right)/2} - \sqrt{\frac{1}{2N}}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\left(\frac{\varepsilon_{i,t}}{\sigma}\right)^2 - 1\right) + \mathcal{O}_P(N^{-1/2}).
\end{aligned}$$

Concerning the first expression in the last line, we can use the restriction $\sqrt{T}N \to 0$

Here, $CD_{\varepsilon/\sigma}$ is the usual CD test statistic based on raw error terms $\varepsilon_{i,t}/\sigma$. Lemma S.1 allows us to conclude that $CD_{\varepsilon/\sigma} = \mathcal{O}_P(1)$. The final result follows after observing that by double-index CLT:

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\left(\frac{\varepsilon_{i,t}}{\sigma}\right)^2 - 1\right) = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\eta_{i,t}^2 - 1\right) = \mathcal{O}_P(1), \tag{S.18}$$

given $\mathrm{E}[|\eta_{i,t}|^8] < \infty$. □

S.5.3 Proofs of Theorem 2 and Proposition 1

Proof of Proposition 1.
Analogous to the proof of Theorem 1, we decompose CD into

$$CD = 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\sqrt{\sigma_i^2\sigma_j^2}} - k_{N,T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\varsigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)$$
$$= CD_\sigma + CD_{(\widehat{\sigma}-\sigma)}. \tag{S.19}$$

Under the assumptions of this proposition, the leading term of the CD test statistic is given by

$$CD_\sigma = \frac{1}{\sigma^2}2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\widehat{\varepsilon}_i'\widehat{\varepsilon}_j = \frac{1}{\sigma^2}k_{N,T}\sum_{i,j}^{N}\widehat{\varepsilon}_i'\widehat{\varepsilon}_j - \frac{1}{\sigma^2}k_{N,T}\sum_{i=1}^{N}\widehat{\varepsilon}_i'\widehat{\varepsilon}_i$$
$$= I - II.$$

We begin by proving that $I = 0$. Note that

$$\sum_{i,j}^{N}\widehat{\varepsilon}_i'\widehat{\varepsilon}_j = \sum_{t=1}^{T}\left(\sum_{i=1}^{N}\widehat{\varepsilon}_{i,t}\right)^2.$$

Furthermore, for each time period t the sum of residuals can be expressed as

$$\sum_{i=1}^{N}\widehat{\varepsilon}_{i,t} = N\left(\overline{y}_t - \overline{\widehat{\boldsymbol{\lambda}}}'\widehat{\boldsymbol{f}}_t\right), \tag{S.20}$$

where the average estimated loading $\overline{\widehat{\boldsymbol{\lambda}}}$ is implicitly defined by projection off the space spanned by $\widehat{\boldsymbol{f}}_t$. That is, it can be written

$$\overline{\widehat{\boldsymbol{\lambda}}} = \left(\sum_{t=1}^{T}\widehat{\boldsymbol{f}}_t\widehat{\boldsymbol{f}}_t'\right)^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{f}}_t\overline{y}_t = \left(\sum_{t=1}^{T}\widehat{\boldsymbol{f}}_t\widehat{\boldsymbol{f}}_t'\right)^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{f}}_t\widehat{\boldsymbol{f}}_t'\left[1,\mathbf{0}_m'\right]' = \left[1,\mathbf{0}_m'\right]', \tag{S.21}$$

such that $\overline{\widehat{\boldsymbol{\lambda}}}'\widehat{\boldsymbol{f}}_t = \overline{y}_t$. Combining these results we conclude:

$$\sum_{i=1}^{N}\widehat{\varepsilon}_{i,t} = N\left(\overline{y}_t - \overline{\widehat{\boldsymbol{\lambda}}}'\widehat{\boldsymbol{f}}_t\right) = 0, \tag{S.22}$$

therefore also $I = 0$. Next, we show that $II = \mathcal{O}_P(\sqrt{T})$. Using corresponding results from the proof of Theorem 2 with $w_i = \sigma^{-1}\ \forall i$, we can write

$$II = k_{N,T}\frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2 + \mathcal{O}_P(N^{-1/2}\vee T^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T})$$
$$= NT\sqrt{\frac{1}{2TN(N-1)}} + \frac{1}{\sqrt{2(N-1)}}\frac{1}{\sqrt{NT}}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{i,t}^2 - 1\right) + \mathcal{O}_P(N^{-1/2}\vee T^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T}),$$
$$= \sqrt{\frac{TN}{2(N-1)}} + \mathcal{O}_P(N^{-1/2}\vee T^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T}),$$
$$= \sqrt{\frac{T-T/N}{2}} + \mathcal{O}_P(N^{-1/2}\vee T^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T}).$$

Here the second line follows from double-index CLT applied to the sequence $\eta_{i,t}^2 - 1$. Combining our results with those on expressions I and II we conclude

$$CD = -\sqrt{\left(T - \frac{T}{N}\right)/2} + \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T}). \tag{S.23}$$

□

Proof of Theorem 2. As in Proposition 1:

$$\begin{aligned} CD &= 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\sqrt{\sigma_i^2\sigma_j^2}} - k_{N,T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\varsigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \\ &= CD_{\sigma} + CD_{(\widehat{\sigma}-\sigma)}. \end{aligned} \tag{S.24}$$

As shown in Section S.5.7, $CD_{(\widehat{\sigma}-\sigma)} = \mathcal{O}_P\left(\sqrt{N}T^{-1}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right)$, implying that the estimation of unit-specific error variances is asymptotically negligible under weak assumptions on the relative rate of divergence of N and T.

Consider now CD_{σ}. To simplify notation and to emphasize the generality of our result, we derive its asymptotic properties for a generic set of random weights $\{w_i\}_{i=1}^N$ which has the same properties as spelled out in Assumption 2. Taking some conflicting notation with respect to Theorem 3 into account, let the equivalent to CD_{σ} based on these generic weights be denoted

$$\begin{aligned} CD_w = 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\widehat{\varepsilon}_i'\widehat{\varepsilon}_j w_i w_j &= k_{N,T}\sum_{i,j}^{N}\widehat{\varepsilon}_i'\widehat{\varepsilon}_j w_i w_j - k_{N,T}\sum_{i=1}^{N}\widehat{\varepsilon}_i'\widehat{\varepsilon}_i w_i^2 \\ &= I - II, \end{aligned}$$

where

$$\begin{aligned} \widehat{\varepsilon}_i &= \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{F}\boldsymbol{\lambda}_i + \varepsilon_i\right) \\ &= \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_i - \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\left(\widehat{\boldsymbol{F}} - \boldsymbol{F}\overline{\boldsymbol{C}}\right)\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i \\ &= \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_i - \boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i. \end{aligned}$$

Now let

$$\begin{aligned} I &= k_{N,T}\sum_{i,j}^{N}\left(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)'\left(\boldsymbol{\varepsilon}_j - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-}\right)\boldsymbol{\lambda}_j\right)w_i w_j \\ &\quad - k_{N,T}\sum_{i,j}^{N}\left(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_j - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j\right)w_i w_j \\ &= I_1 - I_2. \end{aligned}$$

Here, analogous to how we proceeded in the time fixed effects model, we write

$$\sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)w_i=\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i w_i-\sum_{i=1}^{N}\boldsymbol{U}_i\left(\overline{\boldsymbol{C}}^{-1}\right)\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell w_\ell\right).$$

Consequently, letting $\overline{\boldsymbol{\lambda}}_w=N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell w_\ell$ and $\xi_i=\boldsymbol{\varepsilon}_i w_i-\boldsymbol{U}_i\left(\overline{\boldsymbol{C}}^{-}\right)\overline{\boldsymbol{\lambda}}_w$

$$\begin{aligned}I_1&=k_{N,T}\sum_{i,j}^{N}\xi_i'\xi_j\\&=2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\xi_i'\xi_j+k_{N,T}\sum_{i=1}^{N}\xi_i'\xi_i,\end{aligned}\tag{S.25}$$

where

$$\begin{aligned}k_{N,T}\sum_{i=1}^{N}\xi_i'\xi_i=k_{N,T}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i w_i^2&+k_{N,T}\overline{\boldsymbol{\lambda}}_w'\left(\overline{\boldsymbol{C}}^{-1}\right)'\left(\sum_{i=1}^{N}\boldsymbol{U}_i'\boldsymbol{U}_i\right)\left(\overline{\boldsymbol{C}}^{-1}\right)\overline{\boldsymbol{\lambda}}_w\\&-2k_{N,T}\overline{\boldsymbol{\lambda}}_w'\left(\overline{\boldsymbol{C}}^{-1}\right)'\left(\sum_{i=1}^{N}\boldsymbol{U}_i'\boldsymbol{\varepsilon}_i w_i\right)\end{aligned}$$

is a linear function of asymptotically non-negligible terms. Lemma S.3 implies that $I_2=\mathrm{o}_P(1)$ if both $\sqrt{T}/N\to 0$ and $T\to\infty$. Hence, proceed to

$$\begin{aligned}II&=k_{N,T}\sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)'\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)w_i^2\\&\quad-k_{N,T}\sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)w_i^2\\&=II_1-II_2.\end{aligned}$$

As shown in Lemma S.4, II_2 is $\mathrm{o}_P(1)$, if $N^{-1}\sqrt{T}\to 0$ and $T\to\infty$. For this reason, consider the first term above instead. We can write

$$II_1=k_{N,T}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i w_i^2+k_{N,T}\sum_{i=1}^{N}\left\|\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right\|^2 w_i^2-2k_{N,T}\sum_{i=1}^{N}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i w_i^2,$$

where

$$\begin{aligned}&k_{N,T}\sum_{i=1}^{N}\left\|\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right\|^2 w_i^2\\&\leq k_{N,T}\frac{T}{N}\sum_{i=1}^{N}w_i^2\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\,\mathrm{tr}\left((N/T)\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}}\right)\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2\\&=\mathcal{O}_P(N^{-1}\sqrt{T})\end{aligned}$$

by using result (S.41) and Markov's inequality on $\frac{1}{N}\sum_{i=1}^{N} w_i^2 \boldsymbol{\lambda}_i' \boldsymbol{\lambda}_i$. Next,

$$2k_{N,T}\left|\sum_{i=1}^{N} \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1}\right)' \overline{\boldsymbol{U}}' \varepsilon_i w_i^2\right| \leq 2k_{N,T}\sqrt{N}\left(N^{-1}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\right\|^2\right)^{1/2}\left(\sum_{i=1}^{N}\left\|\overline{\boldsymbol{U}}'\varepsilon_i w_i^2\right\|^2\right)^{1/2}.$$

As noted above, $N^{-1}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\right\|^2 = \mathcal{O}_P(1)$. In order to proceed in with the second term above, we use the identity $\overline{\boldsymbol{U}} = \left[\overline{\varepsilon},\ \overline{\boldsymbol{E}}\right]\boldsymbol{B}$ and write

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{N}\left\|\overline{\boldsymbol{U}}'\varepsilon_i w_i^2\right\|^2\right] &\leq \|\boldsymbol{B}\|^2 N^{-2}\sum_{i,i',i''}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\right]\mathrm{E}\left[w_i^4\right] \\
&\quad + \|\boldsymbol{B}\|^2 N^{-2}\sum_{i,i',i''}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{i,t'}\right]\mathrm{E}\left[\boldsymbol{e}_{i',t}'\boldsymbol{e}_{i'',t'}\right]\mathrm{E}\left[w_i^4\right] \\
&= \mathcal{O}(T) + \mathcal{O}(N^{-1}T^2),
\end{aligned}$$

which follows from combining indexes such that only nonzero expectations remain. Together with uniform boundedness of w_i^4 and boundedness of $\|\boldsymbol{B}\|^2$, we have

$$\begin{aligned}
2k_{N,T}\left|\sum_{i=1}^{N}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\varepsilon_i w_i^2\right| &= \sqrt{\frac{2}{TN(N-1)}}\sqrt{N}\,\mathcal{O}_P(1)\left(\mathcal{O}_P(\sqrt{T}) + \mathcal{O}_P(N^{-1/2}T)\right) \\
&= \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T}). \qquad \text{(S.26)}
\end{aligned}$$

Hence, we can conclude that

$$II_1 = k_{N,T}\sum_{i=1}^{N}\varepsilon_i'\varepsilon_i w_i^2 + \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T}) \qquad \text{(S.27)}$$

Combining the results on I_1, I_2, II_1 and II_2, we have

$$\begin{aligned}
CD_w &= I_1 - I_2 - II_1 + II_2 \\
&= 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\xi_i'\xi_j + k_{N,T}\overline{\boldsymbol{\lambda}}_w'\left(\overline{\boldsymbol{C}}^{-1}\right)'\left(\sum_{i=1}^{N}\boldsymbol{U}_i'\boldsymbol{U}_i\right)\left(\overline{\boldsymbol{C}}^{-1}\right)\overline{\boldsymbol{\lambda}}_w \\
&\quad - 2k_{N,T}\overline{\boldsymbol{\lambda}}_w'\left(\overline{\boldsymbol{C}}^{-1}\right)'\left(\sum_{i=1}^{N}\boldsymbol{U}_i'\varepsilon_i w_i\right) + \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(N^{-1}\sqrt{T}).
\end{aligned}$$

This result transforms into one about CD_σ, the leading component of CD in Eq. (S.24), if we set $w_i = \sigma_i^{-1}$, and define the influence function as $\boldsymbol{\psi}_{i,t} = \left(\overline{\boldsymbol{C}}'\right)^{-1}\boldsymbol{u}_{i,t}$. Additionally using the results of Section S.5.7, the main result of this theorem follows. □

S.5.4 Proof of Theorem 3

Additive part.
Consider the CD_W statistic without variance scaling, and denote it by $\ddot{CD}_W$, i.e:

$$\ddot{CD}_W = \sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \widehat{\varepsilon}_i' \widehat{\varepsilon}_j w_i w_j$$
$$= CD_{\mu,w} - CD_{(\widehat{\mu}-\mu)}.$$

From the results in Section S.5.6 we notice that $CD_{(\widehat{\mu}-\mu)} = \mathcal{O}_P(T^{-1/2})$ irrespective if one considers σ_i^{-1} or any w_i as weights. Moreover, the decomposition of CD provided in Theorem 1 is purely algebraic and holds even with general weights w_i. From Theorem 1 :

$$CD_{\mu,w} = 2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \varepsilon_i' \varepsilon_j \left(w_i - \overline{w}\right) \left(w_j - \overline{w}\right) + \sqrt{T} \left(\Xi_{W,1} + \Xi_{W,2}\right),$$

where

$$\Xi_{W,1} = \sqrt{\frac{N}{2(N-1)}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{i,t}^2 \left((\overline{w})^2 - 2\overline{w} w_i \right)$$
$$\Xi_{W,2} = \sqrt{\frac{1}{2N(N-1)}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{i,t}^2 \left((\overline{w})^2 - 2w_i^2 \right).$$

Here, we consider the leading bias term $\Xi_{1,W}$ where $(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{i,t}^2 = \mathcal{O}_P(1)$, $\overline{w} = \mathcal{O}_P\left(N^{-1/2}\right)$ and $N^{-1/2}T^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{i,t}^2 w_i^2 = \mathcal{O}_P\left(1\right)$ straightforwardly hold by Markov's and Chebyshev's inequalities and uniform boundedness of the error variances. It follows that $\Xi_{1,W} = \mathcal{O}_P\left(N^{-1}\right)$. For the second remainder term $\Xi_{W,2}$, we note that $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{i,t}^2 w_i^2 = \mathcal{O}_P(1)$ which together with the previous intermediary results implies $\Xi_{W,2} = \mathcal{O}_P(N^{-2})$.

We continue with the leading stochastic term in CD_W, given by

$$2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t} \left(w_i - \overline{w}\right) \left(w_j - \overline{w}\right)$$
$$= 2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t} w_i w_j - \left(\overline{w} - N^{-1}\right) \left(2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t}(w_i + w_j) \right)$$
$$+ \left(\overline{w}^2 - N^{-1}\overline{w^2}\right) \left(2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t} \right).$$

Given that $\mathrm{E}[w_i] = 0$, the second and third components are of orders $\mathcal{O}_P(N^{-1/2})$ and $\mathcal{O}_P(N^{-1})$, respectively. This follows from the application of Lemma S.1 on corresponding U-statistics. The remaining non-negligible term is of the form:

$$\sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t} w_i w_j.$$

From Lemma S.1 it follows that:

$$\sqrt{\frac{2}{TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t=1}^{T}\varepsilon_{i,t}\varepsilon_{j,t}w_iw_j \xrightarrow{d} N(0,\Omega) \tag{S.28}$$

as $N,T \to \infty$ jointly, where $\Omega = \mathrm{E}[\sigma_i^2 w_i^2]^2$ The last step in this proof is to show consistency of the standard deviation estimator $\widehat{\Omega} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{i,t}^2 w_i^2$ so that convergence in law of CD_W to a standard normal distribution follows by Slutsky's Theorem. Note here that an equivalent expression is given by term II in the proof of Theorem 1. Drawing from results in Section S.5.6, we conclude that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{i,t}^2 w_i^2 = \frac{1}{N}\sum_{i=1}^{N}\widehat{\sigma}_i^2 w_i^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 w_i^2 + \mathcal{O}_P((NT)^{-1/2}) + \mathcal{O}_P(T^{-1}) + \mathcal{O}_P(N^{-1}).$$

Consistency follow from the Chebyshev's inequality. It follows that $CD_W \xrightarrow{d} N(0,1)$.

Multifactor error part.
Proceeding from the decomposition given in Theorem 2, we consider the two bias terms $\Phi_{W,1}$ and $\Phi_{W,2}$. For both terms, it is instructive to note that

$$\mathrm{E}\left[\left\|N^{-1}\sum_{i=1}^{N}w_i\boldsymbol{\lambda}_i\right\|^2\right] = N^{-2}\sum_{i,j}^{N}\mathrm{E}\left[w_iw_j\right]\mathrm{E}\left[\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_j\right] = N^{-2}\sum_{i=1}^{N}\mathrm{var}\left[w_i\right]\mathrm{E}\left[\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right] = \mathcal{O}\left(N^{-1}\right),$$

so that

$$\left\|N^{-1}\sum_{i=1}^{N}w_i\boldsymbol{\lambda}_i\right\| = \mathcal{O}_P\left(N^{-1/2}\right). \tag{S.29}$$

Concerning $\Phi_{W,1}$, we also need to take into account

$$\begin{aligned}\left\|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\psi}_{i,t}\boldsymbol{\psi}_{i,t}'\right\| &\leq (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{u}_{i,t}'\boldsymbol{u}_{i,t}\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2 \\ &= \mathcal{O}_P\left(1\right),\end{aligned}$$

which follows from Markov's inequality. Using this result, we have

$$\begin{aligned}\Phi_{1,W} &\leq \sqrt{\frac{N}{2(N-1)}}\left\|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\psi}_{i,t}\boldsymbol{\psi}_{i,t}'\right\|\left\|N^{-1}\sum_{i=1}^{N}w_i\boldsymbol{\lambda}_i\right\|^2 \\ &= \mathcal{O}_P(N^{-1}).\end{aligned}$$

Next, consider $\widetilde{\Phi}_{2,W}$. By isolating $\overline{\boldsymbol{C}}^{-1}$ from the definition of $\boldsymbol{\psi}_{i,t}$ and by using result (S.43) on the remaining part:

$$\left\|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\psi}_{i,t}w_i\varepsilon_{i,t}\right\| = \mathcal{O}_P\left(N^{-1/2}\right).$$

Consequently,

$$|\Phi_{W,2}| \leq \sqrt{\frac{N}{2(N-1)}} \left\| N^{-1} \sum_{i=1}^{N} w_i \boldsymbol{\lambda}_i \right\| \left\| (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{\psi}_{i,t} w_i \varepsilon_{i,t} \right\|$$
$$= \mathcal{O}_P\left(N^{-1}\right).$$

Thus, the two bias terms affecting CD_W in Theorem 2 are negligible for weights satisfying Assumption 7 as long as $N^{-1}\sqrt{T} \to 0$. We continue with the leading stochastic term of the weighted CD test statistic, given by

$$2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \xi_{iN,t} \xi_{jN,t}$$

with $\xi_{iN,t} = \left[w_i \varepsilon_{i,t} - \left(N^{-1} \sum_{\ell=1}^{N} w_\ell \boldsymbol{\lambda}_\ell' \right) \boldsymbol{\psi}_{i,t} \right]$ and $\boldsymbol{\psi}_{i,t} = \overline{\boldsymbol{C}}^{-1} \boldsymbol{u}_{i,t}$. Analogous to the time fixed effects part of this proof, we can apply Lemma S.1 with $a_{i,t} = w_i \varepsilon_{i,t}$ and $q_i = w_i \sigma_i$ to obtain

$$\sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t} \varepsilon_{j,t} w_i w_j \xrightarrow{d} N(0, (\sigma_w^2 \operatorname{E}\left[\sigma_i^2\right])^2)$$

as $N, T \to \infty$ subject to $N^{-1}\sqrt{T} \to 0$. Now note that by Eq. (S.27) in the proof of Theorem 2,

$$(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{i,t}^2 w_i^2 = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i w_i^2 + \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P(N^{-1}) + \mathcal{O}_P(T^{-1})$$

where $(NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i w_i^2 \xrightarrow{p} \sigma_w^2 \operatorname{E}\left[\sigma_i^2\right]$ as argued in the time fixed effects part of the proof of this theorem. By application of Slutsky's Theorem, it then follows that

$$\left[(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{i,t}^2 w_i^2 \right]^{-1} \sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t} \varepsilon_{j,t} w_i w_j \xrightarrow{d} N(0,1).$$

Next, consider the sum

$$2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \boldsymbol{u}_{i,t} \boldsymbol{u}_{j,t}' = \boldsymbol{B}' \left(2k_{N,T} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} (\varepsilon_{i,t}, \boldsymbol{e}_{i,t}')' (\varepsilon_{j,t}, \boldsymbol{e}_{j,t}') \right) \boldsymbol{B}. \tag{S.30}$$

Given that $\boldsymbol{e}_{i,t}$ has properties similar to $\varepsilon_{i,t}$, it can be shown analogously the reasoning leading to (S.28) that a CLT holds for every element of the $(m+1) \times (m+1)$ matrix in the parentheses of Eq. (S.30) by Lemma S.1. This allows us to conclude that the elements of (S.30) are stochastically bounded. Thus, recalling result (S.29),

$$\left(N^{-1} \sum_{\ell=1}^{N} w_\ell \boldsymbol{\lambda}_\ell' \right) \left(\sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \boldsymbol{\psi}_{i,t} \boldsymbol{\psi}_{j,t}' \right) \left(N^{-1} \sum_{\ell=1}^{N} \boldsymbol{\lambda}_\ell w_\ell \right) = \mathcal{O}_P(N^{-1}),$$
$$\left(\sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \varepsilon_{i,t} \boldsymbol{\psi}_{j,t}' \right) \left(N^{-1} \sum_{\ell=1}^{N} \boldsymbol{\lambda}_\ell w_\ell \right) = \mathcal{O}_P(N^{-1/2}),$$
$$\left(N^{-1} \sum_{\ell=1}^{N} w_\ell \boldsymbol{\lambda}_\ell' \right) \left(\sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} \boldsymbol{\psi}_{i,t} \varepsilon_{i,t} \right) = \mathcal{O}_P(N^{-1/2}),$$

so that

$$\left((NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{i,t}^{2}w_i^2\right)^{-1}\sqrt{\frac{2}{TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t=1}^{T}\xi_{iN,t}\xi_{jN,t}\xrightarrow{d}N(0,1),$$

which leads to the final result of this part of the theorem.

S.5.5 Auxiliary lemmas for Sections S.5.2–S.5.4

Lemma S.1. *Let $\{a_{i,t}\}_{t=1}^{T}$ be a scalar and $\boldsymbol{c}_i$ $L-$dimensional sequences of random variables for $i = 1,\dots,N$ such that:*

- *$a_{i,t}, a_{j,s}$ are independent for all $i \neq j$ and all s,t.*
- *$a_{i,t}, a_{i,s}$ are iid conditionally on $\boldsymbol{c}_i$ for all s,t.*
- *$\mathrm{E}[|a_{i,t}|^8|\boldsymbol{c}_i] < M < \infty$. $\boldsymbol{c}_i$ are iid over i.*
- *$q_i^k \equiv \mathrm{E}[|a_{i,t}|^k|\boldsymbol{c}_i]$ and $\mathrm{E}[a_{i,t}|\boldsymbol{c}_i] = 0$ for $k \leq 8$.*
- *$\mathrm{E}[|q_i^k|] < \infty$ for $k \leq 8$.*

Then:

$$U = \sqrt{\frac{2}{NT(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t=1}^{T}a_{i,t}a_{j,t}\xrightarrow{d}N(0,\mathrm{E}[q_i^2]^2), \tag{S.31}$$

jointly as $N,T \to \infty$.

Proof of Lemma S.1.
The prove this lemma we use Theorem 3.2 of Hall and Heyde (1980). In particular, express U as:

$$U = \sum_{t=1}^{T}\xi_{t,N,T}, \quad \xi_{t,N,T} = \sqrt{\frac{2}{NT(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}a_{i,t}a_{j,t}. \tag{S.32}$$

Denote by $\mathcal{C} = \sigma(\{\boldsymbol{c}_j\}_{j=1}^{N})$ the σ-algebra generated by $\{\boldsymbol{c}_j\}_{j=1}^{N}$ and let $\mathcal{F}_{t-1,N,T} = \sigma(\mathcal{C} \vee \{\xi_{s,N,T}\}_{s=1}^{t-1})$ be the σ-algebra generated by $\mathcal{C}$ and $\{\xi_{s,N,T}\}_{s=1}^{t-1}$. It is easy to see that $\{\xi_{t,N,T}, \mathcal{F}_{t-1,N,T} : t = 1,\dots,T\}$ is a Martingale Difference Array.

At first we establish the limiting variance of this MDS array. From Corollary 3.1 in Hall and Heyde (1980) the variance is determined from:

$$V_T = \sum_{t=1}^{T}\mathrm{E}[\xi_{t,N,T}^2|\mathcal{F}_{t-1,N,T}]\xrightarrow{p}\eta^2 \tag{S.33}$$

By conditional independence of $a_{i,t}$, the above result simplifies:

$$\begin{aligned}
\sum_{t=1}^{T} \mathrm{E}[\xi_{t,N,T}^2 | \mathcal{F}_{t-1,N,T}] &= \sum_{t=1}^{T} \mathrm{E}[\xi_{t,N,T}^2 | \mathcal{C}] \\
&= \frac{2}{N(N-1)} \mathrm{E}\left[\left(\sum_{i=2}^{N} \sum_{j=1}^{i-1} a_{i,t} a_{j,t} \right)^2 | \mathcal{C} \right] \\
&= \frac{2}{N(N-1)} \mathrm{E}\left[\left(\sum_{i=2}^{N} a_{i,t} A_{i-1,t} \right)^2 | \mathcal{C} \right] \\
&= \frac{2}{N(N-1)} \mathrm{E}\left[\sum_{i=2}^{N} a_{i,t}^2 A_{i-1,t}^2 | \mathcal{C} \right] \\
&= \frac{2}{N(N-1)} \sum_{i=2}^{N} q_i^2 \sum_{j=1}^{i-1} q_j^2,
\end{aligned}$$

where in the third line we defined the "integrated" variable $A_{i-1,t} = \sum_{j=1}^{i-1} a_{j,t}$, the fourth line uses the fact that $a_{i,t} A_{i-1,t}$ is a Martingale Difference Sequence. The last line uses the definition of q_i^2. After re-arranging elements:

$$\begin{aligned}
V_T &= \frac{2}{N(N-1)} \sum_{i=2}^{N} q_i^2 \sum_{j=1}^{i-1} q_j^2 \\
&= \frac{1}{N(N-1)} \left(N^2 \left(\frac{1}{N} \sum_{i=1}^{N} q_i^2 \right)^2 - N \left(\frac{1}{N} \sum_{i=1}^{N} q_i^4 \right) \right) \\
&= \left(\frac{1}{N} \sum_{i=1}^{N} q_i^2 \right)^2 + \mathcal{O}_P(N^{-1}) \\
&= \mathrm{E}[q_i^2]^2 + \mathcal{O}_P(N^{-1}).
\end{aligned}$$

Here the third and the fourth lines follow from the application of the Kolmogorov's SLLN to iid sequences q_i^2 and q_i^4. Thus we can expect that

$$U = \sqrt{\frac{2}{NT(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{t=1}^{T} a_{i,t} a_{j,t} \xrightarrow{d} N(0, \mathrm{E}[q_i^2]^2). \tag{S.34}$$

It remains to prove that the sufficient condition for Corollary 3.1 in Hall and Heyde (1980) is satisfied. In particular, it is sufficient to show that $\xi_{t,N,T}$ is a (conditionally) uniformly integrable sequence:

$$\sum_{t=1}^{T} \mathrm{E}[\xi_{t,N,T}^2 I(|\xi_{t,N,T}| > \varepsilon) | \mathcal{F}_{t-1,N,T}] \xrightarrow{p} 0. \tag{S.35}$$

Instead of proving uniform integrability, we borrow the idea from Kao et al. (2012) and instead show that conditional Lyapunov's condition:

$$B_T = \sum_{t=1}^{T} \mathrm{E}[|\xi_{t,N,T}|^{2+\delta}|\mathcal{F}_{t-1,N,T}] \xrightarrow{p} 0, \quad \delta > 0, \tag{S.36}$$

is satisfied under the maintained assumptions. In what follows we prove that the above condition is satisfied for $\delta = 2$. Observe how:

$$B_T = T\,\mathrm{E}[\xi_{t,N,T}^4|\mathcal{C}] = T^{-1}\frac{N}{4(N-1)}\,\mathrm{E}\left[\left(N\overline{a}_t^2 - \overline{a^2}_t\right)^4|\mathcal{C}\right],$$

where $\overline{a}_t = N^{-1}\sum_{i=1}^{N} a_{i,t}$ and similarly for $\overline{a^2}_t$. By Minkowski's inequality:

$$B_T \leq T^{-1}\frac{N}{4(N-1)}\left(\sqrt[4]{\mathrm{E}\left[\left(N\overline{a}_t^2\right)^4|\mathcal{C}\right]} + \sqrt[4]{\mathrm{E}\left[\left(\overline{a^2}_t\right)^4|\mathcal{C}\right]}\right)^4.$$

As by assumption $\mathrm{E}[|a_{i,t}|^8|\boldsymbol{c}_i] < M$ we can use the (conditional) Rosenthal's inequality to conclude that:

$$\mathrm{E}\left[\left(N\overline{a}_t^2\right)^4|\mathcal{C}\right] = \mathcal{O}_P(1). \tag{S.37}$$

Moreover, another application of Minkowski's inequality yields:

$$\sqrt[4]{\mathrm{E}\left[\left(\overline{a^2}_t\right)^4|\mathcal{C}\right]} \leq \sqrt[4]{\mathrm{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}(a_{i,t}^2 - q_i^2)\right)^4|\mathcal{C}\right]} + \sqrt[4]{\left(\frac{1}{N}\sum_{i=1}^{N}(q_i^2)\right)^4}. \tag{S.38}$$

Further application of the (conditional) Rosenthal's inequality yields:

$$\mathrm{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}(a_{i,t}^2 - q_i^2)\right)^4|\mathcal{C}\right] = \mathcal{O}_P(N^{-2}), \tag{S.39}$$

which together with the fact that $N^{-1}\sum_{i=1}^{N}(q_i^2) = \mathcal{O}_P(1)$ implies that:

$$B_T \leq T^{-1}\frac{N}{4(N-1)}\,\mathcal{O}_P(1) = \mathrm{o}_P(1).$$

As required. This completes the proof. □

Lemma S.2. *Let w_i denote any stochastic weights that are independent across i and satisfy $\mathrm{E}\left[w_i^2\right] < M$ for all i. Furthermore, let w_i be independent of $\boldsymbol{f}_t$, $\varepsilon_{i,t}$ and $\boldsymbol{e}_{i,t}$ for all i and t. Under Assumptions 2-6,*

$$\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1} = \mathcal{O}_P(1) \tag{S.40}$$

$$\sqrt{N}T^{-1/2}\left\|N^{-1}\sum_{i=1}^{N} w_i\boldsymbol{U}_i'\right\| = \mathcal{O}_P(1) \tag{S.41}$$

$$\sqrt{N}T^{-1/2}\left\|\left(N^{-1}\sum_{i=1}^{N} w_i\boldsymbol{U}_i'\right)\boldsymbol{F}\right\| = \mathcal{O}_P(1) \tag{S.42}$$

Additionally, under Assumptions 2-6 and given weights w_i that satisfy Assumption 7, we have

$$\left\|(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{U}_i'\boldsymbol{\varepsilon}_i w_i\right\| = \mathcal{O}_P(N^{-1/2}) \tag{S.43}$$

Proof of Lemma S.2.

1. *Result* (S.40): By Pesaran (2006, equation (36)) it holds that

   $$T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}} = \overline{\boldsymbol{C}}'\left(T^{-1}\boldsymbol{F}'\boldsymbol{F}\right)\overline{\boldsymbol{C}} + \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P(N^{-1}),$$

   where the function of true factors and true loadings itself converges to a positive definite matrix by Assumptions 3 and 4. Hence, Theorem 1 in Karabıyık et al. (2017) applies and equation (S.40) follows.

2. *Result* (S.41): Taking the square of the expression of interest, we can write

   $$NT^{-1}\left\|N^{-1}\sum_{i=1}^{N}w_i\boldsymbol{U}_i'\right\|^2 \leq (NT)^{-1}\sum_{i,j}^{N}\sum_{t=1}^{T}w_iw_j\left(\varepsilon_{i,t}\varepsilon_{j,t} + \boldsymbol{e}_{i,t}'\boldsymbol{e}_{j,t}\right)\|\boldsymbol{B}\|^2$$

   where $(NT)^{-1}\sum_{i,j}^{N}\sum_{t=1}^{T}w_iw_j\varepsilon_{i,t}\varepsilon_{j,t} = \mathcal{O}_P(1)$ and $(NT)^{-1}\sum_{i,j}^{N}\sum_{t=1}^{T}w_iw_j\boldsymbol{e}_{i,t}'\boldsymbol{e}_{j,t} = \mathcal{O}_P(1)$ by Markov's inequality and zero correlation of idiosyncratic variation in both $\boldsymbol{y}_i$ and $\boldsymbol{X}_i$ across i. Result (S.41) follows accordingly.

3. *Result* (S.42): Analogous to the proof of (S.41) we take the square of (S.42) and rearrange to arrive at

   $$NT^{-1}\left\|N^{-1}\sum_{i=1}^{N}w_i\boldsymbol{U}_i'\boldsymbol{F}\right\|^2 \leq (NT)^{-1}\sum_{i,j}^{N}\sum_{t,t'}^{T}w_iw_j\boldsymbol{f}_t'\boldsymbol{f}_{t'}\left(\varepsilon_{i,t}\varepsilon_{j,t'} + \boldsymbol{e}_{i,t}'\boldsymbol{e}_{j,t'}\right)\|\boldsymbol{B}\|^2.$$

   Now taking expectations of the non-negative expression $(NT)^{-1}\sum_{i,j}^{N}\sum_{t,t'}^{T}w_iw_j\boldsymbol{f}_t'\boldsymbol{f}_{t'}\varepsilon_{i,t}\varepsilon_{j,t'}$ and using uncorrelatedness of $\varepsilon_{i,t}$ across i and t as well as $\mathrm{E}\left(w_i^2\right) \leq M$, we obtain

   $$\begin{aligned}\mathrm{E}\left[(NT)^{-1}\sum_{i,j}^{N}\sum_{t,t'}^{T}w_iw_j\boldsymbol{f}_t'\boldsymbol{f}_{t'}\varepsilon_{i,t}\varepsilon_{j,t'}\right] &\leq (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}M\ \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{F}})\,\mathrm{E}\left[\varepsilon_{i,t}^2\right]\\ &= \mathcal{O}(1)\end{aligned}$$

   An identical result is given for the term involving $\boldsymbol{e}_{i,t}'\boldsymbol{e}_{j,t'}$. Result (S.42) is then obtained via Markov's inequality.

4. *Result* (S.43): Given the definition of $\boldsymbol{U}_i$, we can write

$$\mathrm{E}\left[\left\|(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{U}_i'\varepsilon_i w_i\right\|^2\right] \leq (NT)^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\left(\varepsilon_{i,t}\varepsilon_{i,t'}+\boldsymbol{e}_{i,t}'\boldsymbol{e}_{i,t'}\right)\varepsilon_{i,t'}\right]\mathrm{var}\left[w_i\right]\|\boldsymbol{B}\|^2$$

where we directly use independence of w_i across i and $\mathrm{E}[w_i]=0$. Additionally,

$$\begin{aligned} i(NT)^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{i,t}\varepsilon_{i,t'}\varepsilon_{i,t'}\right] &= (NT)^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\kappa_4\left[\varepsilon_{i,t}\right]+(NT)^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}^2\right]\mathrm{E}\left[\varepsilon_{i,t'}^2\right] \\ &= \mathcal{O}(N^{-1}) \end{aligned}$$

and

$$\begin{aligned} (NT)^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\boldsymbol{e}_{i,t}'\boldsymbol{e}_{i,t'}\varepsilon_{i,t'}\right] &= (NT)^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left[\varepsilon_{i,t}^2\right]\mathrm{tr}\left(\boldsymbol{\Sigma}\right) \\ &= \mathcal{O}\left[(NT)^{-1}\right], \end{aligned}$$

from which it follows that

$$\left\|(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{U}_i'\varepsilon_i w_i\right\|^2 = \mathcal{O}_P\left(N^{-1}\right)$$

by Markov's inequality. Taking the square root leads to result (S.43).

□

Lemma S.3. *Under Assumptions 2-6 and* $\mathrm{E}[|w_i|^8] < M$*:*

$$\begin{aligned} &\sqrt{\frac{1}{2TN(N-1)}}\sum_{i,j}^{N}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)'\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_j-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)'\boldsymbol{\lambda}_j\right)w_i w_j \\ &= \mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\sqrt{T}\right) \end{aligned}$$

Proof of Lemma S.3.
Let $\overline{\boldsymbol{\lambda}}_w = \left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell w_\ell\right)$ and note that $\sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)'\boldsymbol{\lambda}_i\right)w_i = \sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i w_i-\boldsymbol{U}_i\left(\overline{\boldsymbol{C}}^{-1}\right)\overline{\boldsymbol{\lambda}}_w\right)$.
This allows us to write

$$\begin{aligned} &\sqrt{\frac{1}{2TN(N-1)}}\sum_{i,j}^{N}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)'\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_j-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)'\boldsymbol{\lambda}_j\right)w_i w_j \\ &\leq \sqrt{\frac{1}{2TN(N-1)}}\left(\left\|\sum_{i=1}^{N}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i w_i\right\|+\left\|\sum_{i=1}^{N}\widehat{\boldsymbol{F}}'\boldsymbol{U}_i\right\|\left\|\overline{\boldsymbol{C}}^{-1}\right\|\left\|\overline{\boldsymbol{\lambda}}_w\right\|\right)^2\left\|\left(\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|. \end{aligned}$$

First, note that

$$\left\|\overline{\boldsymbol{\lambda}}_w\right\|^2 = N^{-2}\sum_{i,j}^{N} w_i w_j \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_j = \mathcal{O}_P(1) \tag{S.44}$$

since $\mathrm{E}\left[N^{-2}\sum_{i,j}^{N} w_i w_j \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_j\right] \leq \left(N^{-2}\sum_{i,j}^{N}\mathrm{E}\left[w_i^2 w_j^2\right]\right)^{1/2}\left(N^{-2}\sum_{i,j}^{N}\mathrm{E}\left[\|\boldsymbol{\lambda}_i\|^2\|\boldsymbol{\lambda}_j\|^2\right]\right)^{1/2} = \mathcal{O}(1)$ by boundedness of the fourth moments of all stochastic components involved. Next, recall that $\widehat{\boldsymbol{F}} = \boldsymbol{F}\overline{\boldsymbol{C}} + \overline{\boldsymbol{U}}$. Thus,

$$\begin{aligned}\left\|\sum_{i=1}^{N}\widehat{\boldsymbol{F}}'\boldsymbol{U}_i\right\| &\leq \sqrt{NT}\left\|\overline{\boldsymbol{C}}\right\|\left\|\sqrt{N/T}\boldsymbol{F}'\overline{\boldsymbol{U}}\right\| + T\left\|\sqrt{N/T}\overline{\boldsymbol{U}}\right\|^2\\ &= \mathcal{O}_P\left(\sqrt{NT}\right) + \mathcal{O}_P\left(T\right),\end{aligned}$$

where the last line is a consequence of results (S.41) and (S.42). Furthermore, since $\boldsymbol{\varepsilon}_i = \boldsymbol{U}_i\boldsymbol{B}^{-1}\left[1,\boldsymbol{0}'\right]'$, we can proceed analogously for $\left\|\sum_{i=1}^{N}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i w_i\right\| \leq \left\|\sum_{i=1}^{N}\widehat{\boldsymbol{F}}'\boldsymbol{U}_i w_i\right\|\left\|\boldsymbol{B}^{-1}\right\|$ and obtain the same orders in probability as in the last equation above. Additionally using result (S.40), we arrive at

$$\begin{aligned}&\sqrt{\frac{1}{2TN(N-1)}}\sum_{i,j}^{N}\left(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)'\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_j - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)'\boldsymbol{\lambda}_j\right)w_i w_j\\ &= \sqrt{\frac{1}{2TN(N-1)}}\left[\mathcal{O}_P\left(\sqrt{NT}\right) + \mathcal{O}_P\left(T\right)\right]^2\mathcal{O}_P\left(T^{-1}\right)\\ &= \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right),\end{aligned}$$

which concludes this proof. □

Lemma S.4. *Under Assumptions 2-6 and* $\mathrm{E}[|w_i|^8] < M$*:*

$$\begin{aligned}&\sqrt{\frac{1}{2TN(N-1)}}\sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_i - \boldsymbol{U}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)w_i^2\\ &= \mathcal{O}_P(N^{-2}\sqrt{T}) + \mathcal{O}_P(N^{-3/2}) + \mathcal{O}_P(T^{-1/2})\end{aligned}$$

Proof of Lemma S.4.
Given that $\|\boldsymbol{A}_1 - \boldsymbol{A}_2\|^2 \leq 3\|\boldsymbol{A}_1\|^2 + 3\|\boldsymbol{A}_2\|^2$ for two arbitrary $m \times n$ matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, we have

$$\begin{aligned}&\sqrt{\frac{1}{2TN(N-1)}}\sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_i - \boldsymbol{U}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)w_i^2\\ &\leq 3\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|\\ &+ 3\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\widehat{\boldsymbol{F}}\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|.\end{aligned} \tag{S.45}$$

Here,

$$\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\varepsilon_i'\widehat{\boldsymbol{F}}\right\|^2$$
$$=3\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\varepsilon_i'\boldsymbol{F}\right\|^2\left\|\overline{\boldsymbol{C}}\right\|^2+3\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\varepsilon_i'\overline{\boldsymbol{U}}\right\|^2.$$

Concerning the first term on the right-hand side above, we can use Chebyshev's inequality together with

$$\begin{aligned}\mathrm{E}\left[\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}T^{-2}w_i^2\varepsilon_i'\boldsymbol{F}\boldsymbol{F}'\varepsilon_i\right] &= \sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\sum_{t,t'}^{T}T^{-2}\,\mathrm{E}\left[w_i^2\right]\mathrm{E}\left[\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{i,t'}|\sigma_i\right]\right]\mathrm{E}\left[\boldsymbol{f}_t'\boldsymbol{f}_{t'}\right]\\ &= \sqrt{\frac{1}{2TN(N-1)}}\sum_{i=1}^{N}\mathrm{E}\left[\sigma_i^2\right]\mathrm{E}\left[w_i^2\right]\mathrm{tr}(\boldsymbol{\Sigma_F})\\ &= \mathcal{O}(T^{-1/2})\end{aligned}$$

to arrive at $\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\varepsilon_i'\boldsymbol{F}\right\|^2\left\|\overline{\boldsymbol{C}}\right\|^2=\mathcal{O}_P(T^{-1/2})$. Likewise, recalling that $\boldsymbol{u}_{i,t}=\boldsymbol{B}'\left[\varepsilon_{i,t},\boldsymbol{e}_{i,t}'\right]'$, we have

$$\begin{aligned}&\mathrm{E}\left[\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\varepsilon_i'\overline{\boldsymbol{U}}\right\|^2\right]\\ &\leq\sqrt{\frac{T}{2N(N-1)}}\sum_{i,i',i''}^{N}\sum_{t,t'}^{T}(NT)^{-2}\,\mathrm{E}\left[w_i^2\right]\mathrm{E}\left[\varepsilon_{i,t}\left(\varepsilon_{i',t}\varepsilon_{i'',t'}+\boldsymbol{e}_{i',t}'\boldsymbol{e}_{i'',t'}\right)\varepsilon_{i,t'}\right]\|\boldsymbol{B}\|\\ &=\sqrt{\frac{T}{2N(N-1)}}\left(\mathcal{O}(N^{-1})+\mathcal{O}(T^{-1})\right)\mathcal{O}(1)\\ &=\mathcal{O}(N^{-2}\sqrt{T})+\mathcal{O}(N^{-1}T^{-1/2}).\end{aligned}$$

This result is obtained by noting that

$$\begin{aligned}(NT)^{-2}\sum_{i,i',i''}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\right] &= (NT)^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\kappa_4\left[\varepsilon_{i,t}\right]+(NT)^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}^2\right]\mathrm{E}\left[\varepsilon_{i,t'}^2\right]\\ &\quad+(NT)^{-2}\sum_{i,i'}^{N}\sum_{t=1}^{T}\mathrm{E}\left[\varepsilon_{i',t}^2\right]\mathrm{E}\left[\varepsilon_{i,t}^2\right]\\ &=\mathcal{O}(N^{-1})+\mathcal{O}(T^{-1})\end{aligned}$$

and

$$\begin{aligned}(NT)^{-2}\sum_{i,i',i''}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\boldsymbol{e}_{i',t}'\boldsymbol{e}_{i'',t'}\varepsilon_{i,t'}\right] &= (NT)^{-2}\sum_{i,i'}^{N}\sum_{t=1}^{T}\mathrm{E}\left[\varepsilon_{i,t}^2\right]\mathrm{E}\left[\boldsymbol{e}_{i',t}'\boldsymbol{e}_{i',t}\right]\\ &=\mathcal{O}(T^{-1}).\end{aligned}$$

Combining the results obtained up to this point and additionally using (S.40), we obtain

$$\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|=\mathcal{O}_P(N^{-2}\sqrt{T})+\mathcal{O}_P(T^{-1/2}). \tag{S.46}$$

Next, consider

$$\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}\widehat{\boldsymbol{F}}\right\|^2=\sqrt{\frac{NT}{2(N-1)}}\left(N^{-1}\sum_{i=1}^{N}w_i^2\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right)\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2\left\|T^{-1}\overline{\boldsymbol{U}}\widehat{\boldsymbol{F}}\right\|^2$$

where $N^{-1}\sum_{i=1}^{N}w_i^2\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i=\mathcal{O}_P(1)$ since $\mathrm{E}\left[N^{-1}\sum_{i=1}^{N}w_i^2\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right]\leq\left(N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[w_i^4\right]\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\|\boldsymbol{\lambda}_i\|^4\right]\right)^{1/2}$ Additionally, we can write

$$\begin{aligned}\left\|T^{-1}\overline{\boldsymbol{U}}\widehat{\boldsymbol{F}}\right\|^2&\leq\left(\left\|T^{-1}\boldsymbol{F}'\overline{\boldsymbol{U}}\right\|+\left\|T^{-1}\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}}\right\|\right)^2\\&=\mathcal{O}_P(N^{-2})+\mathcal{O}_P(N^{-3/2}T^{-1/2})+\mathcal{O}_P((NT)^{-1})\end{aligned}$$

with the last step following from (S.41) and (S.42). Consequently, it holds that

$$\begin{aligned}&\sqrt{\frac{T}{2N(N-1)}}\sum_{i=1}^{N}\left\|T^{-1}w_i\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}\widehat{\boldsymbol{F}}\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|\\&=\mathcal{O}_P(N^{-2}\sqrt{T})+\mathcal{O}_P(N^{-3/2})+\mathcal{O}_P(N^{-1}T^{-1/2}).\end{aligned}$$

Summarizing the results derived so far, we have

$$\begin{aligned}&\sqrt{\frac{1}{2TN(N-1)}}\sum_{i=1}^{N}\left(\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_i-\boldsymbol{U}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)w_i^2\\&=\mathcal{O}_P(N^{-2}\sqrt{T})+\mathcal{O}_P(N^{-3/2})+\mathcal{O}_P(T^{-1/2})\end{aligned}$$

which concludes the proof. □

S.5.6 Additive model. The effect of estimating fixed effects and error variances

Estimating fixed effects We start at the decomposition of CD_σ in Eq. (S.13) and focus on its second term

$$CD_{(\widehat{\mu}-\mu)} = Tk_{N,T}\left(\sum_{i=1}^{N}\frac{\overline{\varepsilon}_i-\overline{\varepsilon}}{\sigma_i}\right)^2 - Tk_{N,T}\sum_{i=1}^{N}\left(\frac{\overline{\varepsilon}_i-\overline{\varepsilon}}{\sigma_i}\right)^2$$

Consider now the first term on the right-hand side above.

$$\begin{aligned} Tk_{N,T}\left(\sum_{i=1}^{N}\frac{\overline{\varepsilon}_i-\overline{\varepsilon}}{\sigma_i}\right)^2 &= Nk_{N,T}\left[(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i} - \left(N^{-1}\sum_{i=1}^{N}\sigma_i^{-1}\right)(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}\right]^2 \\ &= \sqrt{\frac{N}{2T(N-1)}}\,\mathcal{O}_P(1) \\ &= \mathcal{O}_P\left(T^{-1/2}\right). \end{aligned}$$

Here, the second line line follows from Theorem 3 in Phillips and Moon (1999) applied to sequences $\varepsilon_{i,t}$ and $\varepsilon_{i,t}/\sigma_i$ as well as the lower bound of σ_i at a value above zero. As for the second component above, note that

$$\begin{aligned} &Tk_{N,T}\sum_{i=1}^{N}\left(\frac{\overline{\varepsilon}_i-\overline{\varepsilon}}{\sigma_i}\right)^2 \\ &= k_{N,T}\sum_{i=1}^{N}\left(T^{-1/2}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2 - 2k_{N,T}\left((NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)\left((NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}\right) \\ &+ k_{N,t}\left(N^{-1}\sum_{i=1}^{N}\sigma_i^{-1}\right)\left(\sqrt{NT}\overline{\varepsilon}\right)^2 \\ &= k_{N,T}\sum_{i=1}^{N}\left(T^{-1/2}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2 + \mathcal{O}_P\left(N^{-1}T^{-1/2}\right) \end{aligned}$$

Again, we use Theorem 3 in Phillips and Moon (1999) to arrive at the last line. In order to derive a result for the remaining explicit term, we note that

$$\begin{aligned} k_{N,T}\sum_{i=1}^{N}\left(T^{-1/2}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2 &= Nk_{N,T} + \sqrt{N}k_{N,T}N^{-1/2}\sum_{i=1}^{N}\left[\left(T^{-1/2}\sum_{t=1}^{T}\frac{\varepsilon_{i,t}}{\sigma_i}\right)^2 - 1\right] \\ &= \mathcal{O}\left(T^{-1/2}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right]. \end{aligned}$$

From this result and the previous one it follows that

$$CD_{(\widehat{\mu}-\mu)} = \mathcal{O}_P\left(T^{-1/2}\right).$$

Estimating error variances We proceed from equation (S.12) in the proof of Theorem 1, and show below that

$$
\begin{aligned}
CD_{(\widehat{\sigma}-\sigma)} &= k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\varsigma_i^2\right)^{3/2} \left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \\
&= \mathcal{O}_P\left(N^{-1}\sqrt{T}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(T^{-1}\sqrt{N}\right).
\end{aligned}
$$

The expression in the first line above depends on an unknown component ς_i^2 which is bounded by σ_i^2 and $\widehat{\sigma}_i^2$. A simple expansion allows us to write

$$
\begin{aligned}
k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\varsigma_i^2\right)^{3/2} \left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) &= k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2} \left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \\
&\quad + k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \left(\frac{1}{\left(\varsigma_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right) \\
&= a_1 + a_2. \qquad \text{(S.47)}
\end{aligned}
$$

The two expressions above are: 1.) A first-order expansion *at* the true error variances and 2.) the approximation error of this first-order approximation.

Consider now the first term on the right-hand side above, which is the first-order expansion *at* the true error variances. It can be written

$$
\begin{aligned}
a_1 &= k_{N,T} \sum_{i,j}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2} \left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) - k_{N,T} \sum_{i=1}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_i}{\left(\sigma_i^2\right)^{2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \qquad \text{(S.48)} \\
&= a_{11} - a_{12}
\end{aligned}
$$

The order of the two expressions above will be derived by expanding

$$
\widehat{\sigma}_i^2 = \frac{\varepsilon_i' \varepsilon_i}{T} + \frac{\overline{\varepsilon}' \overline{\varepsilon}}{T} + \overline{\varepsilon}_i^2 + \overline{\varepsilon}^2 - 2 \frac{\overline{\varepsilon}' \varepsilon_i}{T} - 2 \overline{\varepsilon}_i \overline{\varepsilon} \qquad \text{(S.49)}
$$

and

$$
\widehat{\varepsilon}_i' \widehat{\varepsilon}_j = \varepsilon_i' \varepsilon_j + \overline{\varepsilon}' \overline{\varepsilon} + T \overline{\varepsilon}_i \overline{\varepsilon}_j + T \overline{\varepsilon}^2 - \overline{\varepsilon}' \varepsilon_i - \overline{\varepsilon}' \varepsilon_j - T \overline{\varepsilon}_i \overline{\varepsilon} - T \overline{\varepsilon}_j \overline{\varepsilon}. \qquad \text{(S.50)}
$$

and by looking at the resulting terms.

Consider the first expression, a_{11}, which we write

$$\begin{aligned}
a_{11} &= k_{N,T}\sum_{i,j}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
&= k_{N,T}\sum_{i,j}^{N}\frac{\varepsilon_i'\varepsilon_j}{\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)+k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\left[N^{-1}\sum_{\ell=1}^{N}\frac{\left(\widehat{\sigma}_\ell^2-\sigma_\ell^2\right)}{\left(\sigma_\ell^2\right)^{3/2}}\right]\left[N^{-1}\sum_{\ell=1}^{N}\sigma_\ell^{-1}\right]\\
&\quad - k_{N,T}\sum_{i,j}^{N}\frac{\varepsilon_i'\varepsilon_j}{\left(\sigma_j^2\right)^{1/2}}\left[N^{-1}\sum_{\ell=1}^{N}\frac{\left(\widehat{\sigma}_\ell^2-\sigma_\ell^2\right)}{\left(\sigma_\ell^2\right)^{3/2}}\right]-k_{N,T}\sum_{i,j}^{N}\frac{\varepsilon_i'\varepsilon_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)}{\left(\sigma_i^2\right)^{3/2}}\left[N^{-1}\sum_{\ell=1}^{N}\sigma_\ell^{-1}\right]\\
&\quad + Tk_{N,T}\sum_{i,j}^{N}\frac{\bar{\varepsilon}_i\bar{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)+Tk_{N,T}\sum_{i,j}^{N}\bar{\varepsilon}_i\bar{\varepsilon}_j\left[N^{-1}\sum_{\ell=1}^{N}\frac{\left(\widehat{\sigma}_\ell^2-\sigma_\ell^2\right)}{\left(\sigma_\ell^2\right)^{3/2}}\right]\left[N^{-1}\sum_{\ell=1}^{N}\sigma_\ell^{-1}\right]\\
&\quad - Tk_{N,T}\sum_{i,j}^{N}\frac{\bar{\varepsilon}_i\bar{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}}\left[N^{-1}\sum_{\ell=1}^{N}\frac{\left(\widehat{\sigma}_\ell^2-\sigma_\ell^2\right)}{\left(\sigma_\ell^2\right)^{3/2}}\right]-Tk_{N,T}\sum_{i,j}^{N}\frac{\bar{\varepsilon}_i\bar{\varepsilon}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)}{\left(\sigma_i^2\right)^{3/2}}\left[N^{-1}\sum_{\ell=1}^{N}\sigma_\ell^{-1}\right]\\
&= a_{111}+a_{112}-a_{113}-a_{114}+a_{115}+a_{116}-a_{117}-a_{118}. \qquad \text{(S.51)}
\end{aligned}$$

The order in probability of all terms above will now be established by implicit application of either Markov's or Chebyshev's inequality. Handling σ_ℓ^k for $k<0$ is straightforward since σ_ℓ is bounded from below at a value above zero by Assumption 1in the main paper. For this reason, and in order to limit the notational burden, we disregard from powers of σ_j^2 that appear in the denominator of fractions involving $\varepsilon_{i,t}$. For the same reason, we disregard from the average $N^{-1}\sum_{\ell=1}^{N}\sigma_\ell^{-1}$. The second and third terms in (S.51), a_{112} and a_{113}, are differently weighted versions of

$$\begin{aligned}
&2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varepsilon_i'\varepsilon_j\left[N^{-1}\sum_{\ell=1}^{N}\left(\widehat{\sigma}_\ell^2-\sigma_\ell^2\right)\right]+k_{N,T}\sum_{i=1}^{N}\varepsilon_i'\varepsilon_i\left[N^{-1}\sum_{\ell=1}^{N}\left(\widehat{\sigma}_\ell^2-\sigma_\ell^2\right)\right]\\
=&a_{1121}+a_{1122}.
\end{aligned}$$

The order of the middle part appearing in both terms is $\mathcal{O}_P\left[(NT)^{-1/2}\right]+\mathcal{O}_P\left(N^{-1}\right)$, as given by result (S.75). Accordingly, we get

$$\begin{aligned}
a_{1121} &= \mathcal{O}_P\left[(NT)^{-1/2}\right]+\mathcal{O}_P\left(N^{-1}\right)\\
a_{1122} &= \mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\sqrt{T}\right),
\end{aligned}$$

since the first term in a_{1121} is $\mathcal{O}_P(1)$, whereas the corresponding term in a_{1122} is $\mathcal{O}_P\left(\sqrt{T}\right)$. It follows that $a_{112}=\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\sqrt{T}\right)$ and $a_{113}=\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\sqrt{T}\right)$. We continue with a_{111} and a_{114} which both are differently weighted versions of $k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)$. Hence, Lemma S.7 implies

$$a_{111}=\mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\sqrt{T}\right),$$

with an identical result for a_{114}. Concerning a_{111}, we additionally decompose

$$
\begin{aligned}
a_{111} &= k_{N,T} \sum_{i,j}^{N} \varepsilon_i' \varepsilon_j \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \\
&= k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \varepsilon_i' \varepsilon_j \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) + k_{N,T} \sum_{i=1}^{N} \varepsilon_i' \varepsilon_i \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right). \qquad \text{(S.52)}
\end{aligned}
$$

where the bound that we stated above holds for both expressions on the right-hand side above. We emphasize this point because the second term in the decomposition of a_{111} cancels out with an identical term in the decomposition of a_{12} in (S.48). To appreciate this point, see the discussion of expression (S.54) below.

Terms a_{115} and a_{118} are almost identical to a_{111} and a_{114}, being weighted versions of are functions of $(N\bar{\varepsilon})\, Tk_{N,T} \sum_{i=1}^{N} \bar{\varepsilon}_i \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right)$. Noting that $\bar{\varepsilon} = \mathcal{O}_P \left[(NT)^{-1/2} \right]$, we can use Lemma S.8 to establish

$$
\begin{aligned}
a_{115} &= \mathcal{O}_P \left(\sqrt{N} T^{-1/2} \right) \left[\mathcal{O}_P \left(T^{-1/2} \right) + \mathcal{O}_P \left(N^{-1/2} T^{-1} \right) + \mathcal{O}_P \left(N^{-3/2} \right) \right] \\
&= \mathcal{O}_P \left(T^{-1} \sqrt{N} \right) + \mathcal{O}_P \left(T^{-3/2} \right) + \mathcal{O}_P \left(N^{-1} T^{-1/2} \right),
\end{aligned}
$$

with a_{118} being of the same order. We emphasize the link between a_{115} and a_{118} as well as a_{111} and a_{114} since one part of a_{115} cancels out with an identical expression in a_{12} below, precisely as a component of a_{111} cancels out.[5] To be more precise, the second term in

$$
a_{115} = \frac{Tk_{N,T}}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\bar{\varepsilon}_i \bar{\varepsilon}_j}{\left(\sigma_i^2 \right)^{3/2} \left(\sigma_j^2 \right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) + \frac{Tk_{N,T}}{2} \sum_{i=1}^{N} \frac{\bar{\varepsilon}_i^2}{\left(\sigma_i^2 \right)^2} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right)
$$

disappears if we combine all terms in a_{11} and a_{12}. The last two terms in our decomposition of a_{11}, a_{116} and a_{117}, both depend on the term $Tk_{N,T} (N\bar{\varepsilon})^2 \left[N^{-1} \sum_{\ell=1}^{N} \left(\widehat{\sigma}_\ell^2 - \sigma_\ell^2 \right) \right]$. Using result (S.75) here, we can conclude that

$$
\begin{aligned}
a_{116} &= k_{N,T} N^2 T\, \mathcal{O}_P \left[(NT)^{-1} \right] \left\{ \mathcal{O}_P \left[(NT)^{-1/2} \right] + \mathcal{O}_P \left(N^{-1} \right) \right\} \\
&= N\sqrt{T} \left\{ \mathcal{O}_P \left[(NT)^{-3/2} \right] + \mathcal{O}_P \left(N^{-2} T^{-1} \right) \right\} \\
&= \mathcal{O}_P \left(N^{-1/2} T^{-1} \right) + \mathcal{O}_P \left(N^{-1} T^{-1/2} \right),
\end{aligned}
$$

and that the same order holds for a_{117}.

Taking together all eight terms of a_{11}, as defined in expression (S.51), we can conclude that

$$
a_{11} = \mathcal{O}_P \left(N^{-1} \sqrt{T} \right) + \mathcal{O}_P \left(N^{-1/2} \right) + \mathcal{O}_P \left(T^{-1/2} \right) + \mathcal{O}_P \left(T^{-1} \sqrt{N} \right). \qquad \text{(S.53)}
$$

[5] This point continues to hold even if we explicitly acknowledge the scaling with functions of σ_i and σ_j which we largely disregard from in this proof.

Next, we write out the second term in our expansion (S.48), which is given by

$$
\begin{aligned}
a_{12} &= k_{N,T}\sum_{t=1}^{T}\sum_{i=1}^{N}\widehat{\varepsilon}_{i,t}^{2}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
& k_{N,T}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}^{2}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)+Tk_{N,T}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\left(T^{-1}\overline{\varepsilon}'\overline{\varepsilon}\right)+Tk_{N,T}\sum_{i=1}^{N}\overline{\varepsilon}_i^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
&+\overline{\varepsilon}^2Tk_{N,T}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)-2k_{N,T}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\overline{\varepsilon}_t\left(\widehat{\sigma}_i^2-\sigma_i^2\right)-\overline{\varepsilon}\,2Tk_{N,T}\sum_{i=1}^{N}\overline{\varepsilon}_i\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
&= a_{121}+a_{122}+a_{123}+a_{124}-a_{125}-a_{126}. \qquad \text{(S.54)}
\end{aligned}
$$

A term identical to a_{121} is given by the second term of expression (S.52), a component of the term a_{11}. Since $a_1 = a_{11} - a_{21}$, term a_{121} and its equivalent in expression (S.52) cancel out. The same holds for a_{123} where an identical term is contained in expression a_{115}. Among the four remaining terms,

$$a_{122} = \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left(N^{-2}\sqrt{T}\right).$$

and analogously

$$a_{124} = \mathcal{O}_P\left(N^{-3/2}T^{-1}\right) + \mathcal{O}_P\left(N^{-2}T^{-1/2}\right)$$

are a consequence of result (S.75) together with $T^{-1}\overline{\varepsilon}'\overline{\varepsilon} = \mathcal{O}_P\left(N^{-1}\right)$. Next, we consider a_{125} which, up to a different weighting, corresponds to either of expressions a_{111} or a_{114} times an additional factor of N^{-1}. Hence, by results (S.78) and (S.79), we arrive at

$$a_{125} = \mathcal{O}_P\left(N^{-2}\sqrt{T}\right) + \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left(N^{-1}T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}T^{-1}\right).$$

The same observation as for term a_{125} can be made for a_{126}which is equivalent to a_{115} or a_{118} times an additional factor of N^{-1}. It follows that

$$a_{126} = \mathcal{O}_P\left(N^{-1/2}T^{-3}\right) + \mathcal{O}_P\left(N^{-1}T^{-5/2}\right) + \mathcal{O}_P\left(N^{-3/2}T^{-2}\right) + \mathcal{O}_P\left(N^{-2}T^{-3/2}\right) + \mathcal{O}_P\left(N^{-3}\sqrt{T}\right).$$

This specifies the properties of all six terms in the decomposition of a_{12}. Combining the resulting orders, we arrive at

$$a_{12} = \mathcal{O}_P\left(N^{-2}\sqrt{T}\right) + \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right]$$

where we disregard from terms a_{121} and a_{123} which are eliminated completely in the difference $a_{11} - a_{12}$. Using the results obtained for both components of this difference, we conclude that

$$
\begin{aligned}
a_1 &= 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
&= \mathcal{O}_P\left(N^{-1}\sqrt{T}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(T^{-1}\sqrt{N}\right), \qquad \text{(S.55)}
\end{aligned}
$$

where we include the bounded random variables $\left(\sigma_i^2\right)^{-3/2}$ and $\left(\sigma_j^2\right)^{-1/2}$ for the sake of completeness.

Up to this point, we have investigated the first-order approximation in equation (S.47) *at* the true error variances. We proceed with results on term a_2, the error of our linear approximation. We take absolute values and use the Cauchy-Schwarz inequality to arrive at

$$
\begin{aligned}
|a_2| &= \left| k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \left[\frac{1}{\left(\varsigma_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right] \right| \\
&\leq k_{N,T} \left(\sum_{i=1}^{N} \left[\sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right]^2 \right)^{1/2} \left(\sum_{i=1}^{N} \left[\frac{1}{\left(\varsigma_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right]^2 \right)^{1/2} \\
&\leq k_{N,T} \left(\sum_{i=1}^{N} \left[\sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right]^2 \right)^{1/2} \left(\sum_{i=1}^{N} \left[\frac{1}{\left(\widehat{\sigma}_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right]^2 \right)^{1/2} \\
&= k_{N,T} \sqrt{a_{21} a_{22}}.
\end{aligned}
\tag{S.56}
$$

A further first-order Taylor approximation of $\left(\widehat{\sigma}_i^2\right)^{-3/2}$ around $\left(\sigma_i^2\right)^{-3/2}$ results in

$$
\frac{1}{\left(\widehat{\sigma}_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} = -3/2 \frac{1}{\left(\varsigma_i^2\right)^{5/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)
$$

where we take into account a slight misuse of notation by letting ς_i^2 be a point in between $\widehat{\sigma}_i^2$ and σ_i^2. Substituting this expression into a_{22} leads to

$$
\begin{aligned}
a_{22} &= \sum_{i=1}^{N} \left[-3/2 \frac{1}{\left(\varsigma_i^2\right)^{5/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right]^2 \\
&\leq 9/4 \left(\sup_i \frac{1}{\left(\varsigma_i^2\right)^{5/2}} \right) \sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)^2 \\
&= 9/4 \left(\frac{1}{\left(\inf_i \varsigma_i^2\right)^{5/2}} \right) \left[\mathcal{O}_P\left(N/T\right) + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right) \right] \\
&= \mathcal{O}_P\left(N/T\right) + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right).
\end{aligned}
\tag{S.57}
$$

where Lemma S.9 is used to ensure that the last line holds as long as $T^{-1}\sqrt{N} \to 0$ is satisfied. To arrive this result, we use result (S.76), as well as Lemma S.9 to establish conditions on the relative expansion rate of N and T under which $\inf_i \varsigma_i^2$ is stochastically bounded. The lemma is merely a generalization of Lemma 5 in Moon et al. (2007) and the resulting assumptions on the divergence rates of N and T can be relaxed even further if one is willing to assume the existence of higher-order moments of $\varepsilon_{i,t}$ beyond $\mathrm{E}\left[\varepsilon_{i,t}^{8}\right]$. In addition to result (S.57), which bounds a_{22}, the order in probability of a_2 is determined by that of a_{21}. Lemma S.10 establishes the corresponding result on a_{21} so that the order in probability of (S.56) is given by

$$a_2 = \left\{ \frac{2}{TN\left(N-1\right)} \left[\mathcal{O}_P\left(N^2\right) + \mathcal{O}_P\left(NT\right) + \mathcal{O}_P\left(N^{-1}T^2\right)\right] \left[\mathcal{O}_P\left(N/T\right) + \mathcal{O}_P\left(N^{-1}\right)\right] \right\}^{1/2}$$
$$= \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-3/2}\sqrt{T}\right). \tag{S.58}$$

This result marks the last one needed to characterize the impact of estimating error variances rather than knowing their true value. We return to expression (S.47) and use our results (S.55) and (S.58) on its two components a_1 and a_2 to arrive at

$$CD_{(\widehat{\sigma}-\sigma)} = \mathcal{O}_P\left(N^{-1}\sqrt{T}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(T^{-1}\sqrt{N}\right). \tag{S.59}$$

S.5.7 Multifactor model. The effect of error variance estimation

Similarly to Section S.5.6, we proceed from equation (S.24) in the proof of Theorem 2, and show below that

$$CD_{(\widehat{\sigma}-\sigma)} = k_{N,T} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\varsigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)$$
$$= \mathcal{O}_P\left(N^{-1}\sqrt{T}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(T^{-1}\sqrt{N}\right).$$

The expression depends on an unknown component ς_i^2 which is bounded by σ_i^2 and $\widehat{\sigma}_i^2$. A simple expansion allows us to write

$$k_{N,T} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\varsigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) = k_{N,T} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)$$
$$+ k_{N,T} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \left(\frac{1}{\left(\varsigma_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right)$$
$$= a_1^* + a_2^*. \tag{S.60}$$

The two expressions above are: 1.) A first-order expansion *at* the true error variances and 2.) the approximation error of this first-order approximation.

Consider now the first term on the right-hand side above, which is the first-order expansion *at* the true error variances. It can be written

$$a_1^* = k_{N,T} \sum_{i,j}^{N} \frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) - k_{N,T} \sum_{i=1}^{N} \frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_i}{\left(\sigma_i^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \tag{S.61}$$
$$= a_{11}^* - a_{12}^*$$

Note that the scalings $\left[\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}\right]^{-1}$ and $\left(\sigma_i^2\right)^{-1/2}$ in both expressions are immaterial since all orders in probability of subterms arising from either a_{11}^* or a_{12}^* will be derived via either Markov's or Chebyshev's inequality. Since error variances have a lower bound above zero, we can always bound their inverse from above. For this reason, and to simplify notation in the following proofs, we disregard from the denominators in the fractions that constitute a_{11}^* and a_{12}^*. The order of the two expressions above will further be derived by expanding

$$\widehat{\sigma}_i^2 = \frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i}{T} + \frac{\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{F}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i}{T} - \frac{\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i}{T} - 2\frac{\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i}{T}. \tag{S.62}$$

and

$$\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_j = \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j + \boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j - \boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_j - \boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j \tag{S.63}$$

and by looking at the resulting terms.

Consider the first expression, a_{11}^*, where we use (S.63) to write

$$\begin{aligned}
a_{11}^* =& k_{N,T}\sum_{i,j}^{N}\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
=& k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right) + k_{N,T}\sum_{i,j}^{N}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
&- k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right) - k_{N,T}\sum_{i,j}^{N}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
&- k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\\
&= a_{111}^* + a_{112}^* - a_{113}^* - a_{114}^* - a_{115}^*.
\end{aligned} \tag{S.64}$$

Concerning the second term in (S.64), a_{112}^*, we can write

$$\begin{aligned}
\left|a_{112}^*\right| &= \left|k_{N,T}\sum_{i,j}^{N}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right|\\
&\le Tk_{N,T}\left[\sum_{i=1}^{N}\left\|\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\boldsymbol{\lambda}_i'\right\|^2\right]^{1/2}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|\left\|\left(\overline{\boldsymbol{C}}^{-1}\right)\right\|^2\left\|\sum_{j=1}^{N}\boldsymbol{\lambda}_j\right\|\\
&= \sqrt{\frac{T}{2N\left(N-1\right)}}\left[\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right)+\mathcal{O}_P\left(1\right)\right]\left\{\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left[\left(NT\right)^{-1/2}\right]\right\}\mathcal{O}_P\left(N\right)\\
&= \mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\sqrt{T}\right).
\end{aligned} \tag{S.65}$$

Here, the third line results from Lemma S.14 as well as result (S.92). The third term in (S.64), a^*_{113}, can be bounded from above in absolute value by

$$
\begin{aligned}
|a^*_{113}| &= \left| k_{N,T} \sum_{i,j}^{N} \varepsilon_i' \boldsymbol{P}_{\widehat{\boldsymbol{F}}} \varepsilon_j \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right| \\
&\leq T k_{N,T} \left\| T^{-1} \sum_{i=1}^{N} \varepsilon_i' \widehat{\boldsymbol{F}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right\| \left\| T^{-1} \sum_{j=1}^{N} \widehat{\boldsymbol{F}}' \varepsilon_j \right\| \left\| \left(T^{-1} \widehat{\boldsymbol{F}}' \widehat{\boldsymbol{F}}\right)^{-1} \right\| \\
&= T k_{N,T} \left[\mathcal{O}_P\left(NT^{-3/2}\right) + \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\right) \right] \\
&\times \left[\mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right) \right] \\
&= \sqrt{\frac{T}{2N(N-1)}} \Big[\mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P\left(N^{3/2}T^{-2}\right) \\
&+ \mathcal{O}_P\left(NT^{-3/2}\right) + \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) \Big] \\
&= \mathcal{O}_P\left(\sqrt{T}N^{-5/2}\right) + \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left(\sqrt{N}T^{-3/2}\right) + \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right].
\end{aligned}
\tag{S.66}
$$

Arriving at the order in probability stated here requires us to use Lemma S.12 as well as results (S.74) and (S.96). We proceed to the fourth term, where we write

$$
\begin{aligned}
|a^*_{114}| &\leq T k_{N,T} \left[\sum_{i=1}^{N} \left\| \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \boldsymbol{\lambda}_i' \right\|^2 \right]^{1/2} \left\| \overline{\boldsymbol{C}}^{-1} \right\| \left\| T^{-1} \sum_{j=1}^{N} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \varepsilon_j \right\| \\
&= \sqrt{\frac{T}{2N(N-1)}} \left[\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right) + \mathcal{O}_P(1) \right] \left[\mathcal{O}_P(1) + \mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right) \right] \\
&= \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right).
\end{aligned}
\tag{S.67}
$$

In addition to Lemma S.14, we can use result (S.92) to derive the upper bound for $|a^*_{114}|$ stated above. The relevance of result (S.92) can be seen by noting that $\left\| T^{-1} \sum_{j=1}^{N} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \varepsilon_j \right\| = N \left\| T^{-1} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\varepsilon} \right\|$ and by recalling that $\left\| T^{-1} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \right\|$ includes all terms in $\left\| T^{-1} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\varepsilon} \right\|$.

We proceed with the fifth term in (S.64), a^*_{115}. After having taken absolute values, we can write

$$
\begin{aligned}
|a^*_{115}| &\leq T k_{N,T} \left\| T^{-1} \sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \varepsilon_i' \overline{\boldsymbol{U}} \right\| \left\| \overline{\boldsymbol{C}}^{-1} \right\| \left\| \sum_{j=1}^{N} \boldsymbol{\lambda}_j \right\| \\
&+ T k_{N,T} \left\| T^{-1} \sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \varepsilon_i' \boldsymbol{P}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \right\| \left\| \overline{\boldsymbol{C}}^{-1} \right\| \left\| \sum_{j=1}^{N} \boldsymbol{\lambda}_j \right\| \\
&= \left(a^*_{1151} + a^*_{1152}\right) \left\| \overline{\boldsymbol{C}}^{-1} \right\| \left\| \sum_{j=1}^{N} \boldsymbol{\lambda}_j \right\|
\end{aligned}
$$

Consider now a^*_{1152}. The order in probability of this expression is bounded by Lemma S.12 as well as results (S.74) and (S.102). More specifically, we can write

$$a^*_{1152} \leq Tk_{N,T} \left\| T^{-1} \sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \varepsilon_i' \widehat{\boldsymbol{F}} \right\| \left\| \left(T^{-1} \widehat{\boldsymbol{F}}' \widehat{\boldsymbol{F}}\right)^{-1} \right\| \left\| T^{-1} \widehat{\boldsymbol{F}}' \overline{\boldsymbol{U}} \right\|$$
$$= \sqrt{\frac{T}{2N(N-1)}} \left[\mathcal{O}_P\left(NT^{-3/2}\right) + \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1/2}\right)\right]$$
$$\times \left\{\mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P\left(N^{-1}\right)\right\}$$
$$= \mathcal{O}_P\left(N^{-1/2}T^{-3/2}\right) + \mathcal{O}_P\left[(NT)^{-1}\right] + \mathcal{O}_P\left(N^{-3/2}T^{-1/2}\right) + \mathcal{O}_P\left(N^{-2}\right) + \mathcal{O}_P\left(N^{-3}\sqrt{T}\right).$$

The a statement for the remaining term a^*_{1151} can be made by invoking Lemma S.12. This is possible since a^*_{1151} is completely contained in the expression $T^{-1}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right)\varepsilon_i'\widehat{\boldsymbol{F}}$, which is the expression handled by this lemma. Accordingly, we have

$$a^*_{1151} = \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(\sqrt{T}N^{-2}\right).$$

Using this result on a^*_{1151} together with the previous one on a^*_{1152} we arrive at

$$a^*_{115} = \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(\sqrt{T}N^{-2}\right). \tag{S.68}$$

Lastly, we consider term a^*_{111} which is decomposed into

$$a^*_{111} = k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \varepsilon_i' \varepsilon_j \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) + k_{N,T} \sum_{i=1}^{N} \varepsilon_i' \varepsilon_i \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)$$
$$= a^*_{1111} + a^*_{1112}.$$

By Lemma S.13, it holds that

$$a^*_{1111} = \mathcal{O}_p\left(\sqrt{N}T^{-1}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right).$$

As for expression a^*_{1112}, we will emphasize below that an identical expression exists in the decomposition of a^*_{12} which eliminates this term. However, for now, we simply conclude from results (S.65)-(S.68) and our discussion here that

$$a^*_{11} = k_{N,T} \sum_{i=1}^{N} \varepsilon_i' \varepsilon_i \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) + \mathcal{O}_P\left(\sqrt{N}T^{-1}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right) \tag{S.69}$$

Having investigated a^*_{11} we proceed with the second major component of the linear approximation term at the true error variances. We can decompose this second term as

$$a^*_{12} = k_{N,T} \sum_{i=1}^{N} \varepsilon_i' \varepsilon_i \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) + k_{N,T} \sum_{i=1}^{N} \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1}\right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{-1}\right) \boldsymbol{\lambda}_i \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)$$
$$- k_{N,T} \sum_{i=1}^{N} \varepsilon_i' \boldsymbol{P}_{\widehat{\boldsymbol{F}}} \varepsilon_i \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) - 2k_{N,T} \sum_{i=1}^{N} \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1}\right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \varepsilon_i \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)$$
$$= a^*_{121} + a^*_{122} - a^*_{123} - a^*_{124}$$

Term a^*_{121} is identical to term a^*_{1112} in a^*_{11} and hence cancels out. Concerning a^*_{122}, we write

$$\begin{aligned}
&\left| k_{N,T} \sum_{i=1}^{N} \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1}\right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{-1}\right) \boldsymbol{\lambda}_i \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right| \\
\leq & T k_{N,T} \left(\sum_{i=1}^{N} \|\boldsymbol{\lambda}_i\|^2 \right)^{1/2} \left\| \overline{\boldsymbol{C}}^{-1} \right\|^2 \left\| T^{-1} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \right\| \left[\sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)^2 \right]^{1/2} \\
= & \sqrt{\frac{T}{2N(N-1)}} \sqrt{N} \left\{ \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right] \right\} \left[\mathcal{O}_P\left(\sqrt{N} T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) \right] \\
= & \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(N^{-2}\sqrt{T}\right)
\end{aligned}$$

where we use result (S.92) and Lemma S.14. The same lemma, together with result (S.97) is employed to show that the next term, a^*_{123}, is

$$\begin{aligned}
a^*_{123} \leq & T k_{N,T} \left[T^{-1} \sum_{i=1}^{N} \left(\boldsymbol{\varepsilon}_i' \boldsymbol{P}_{\widehat{\boldsymbol{F}}} \boldsymbol{\varepsilon}_i\right)^2 \right]^{1/2} \left[\sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)^2 \right]^{1/2} \\
= & \sqrt{\frac{T}{2N(N-1)}} \left[\mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(N^{-1}\right) \right] \left[\mathcal{O}_P\left(\sqrt{N} T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) \right] \\
= & \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left(N^{-1} T^{-1/2}\right) + \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left(N^{-5/2}\sqrt{T}\right)
\end{aligned}$$

Lastly, we use result (S.99) to obtain

$$\begin{aligned}
|a^*_{124}| \leq & 2T k_{N,T} \left[\sum_{i=1}^{N} \left(T^{-1} \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1}\right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \boldsymbol{\varepsilon}_i \right)^2 \right]^{1/2} \left[\sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)^2 \right]^{1/2} \\
= & \sqrt{\frac{2T}{N(N-1)}} \left[\mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right) \right] \left[\mathcal{O}_P\left(\sqrt{N} T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) \right] \\
= & \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(N^{-2}\sqrt{T}\right).
\end{aligned}$$

We can now summarize the four last intermediary results by concluding that

$$a^*_{12} = k_{N,T} \sum_{i=1}^{N} \frac{\boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i}{\left(\sigma_i^2\right)^2} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) + \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right] + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(N^{-2}\sqrt{T}\right),$$

which together with (S.69) leads to

$$\begin{aligned}
a^*_1 = & a^*_{11} - a^*_{12} \\
= & \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right). \qquad \text{(S.70)}
\end{aligned}$$

So far we, our focus has been on a linear approximation of estimated error variances, evaluated at their true values. In the remainder of this proof, we focus on its approximation error which we denoted a_2^* in expression (S.60). We take absolute values and use the Cauchy-Schwarz inequality to arrive at

$$\begin{aligned}
|a_2^*| &= \left| k_{N,T} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \left[\frac{1}{\left(\varsigma_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right] \right| \\
&\leq T k_{N,T} \left(\sum_{i=1}^{N} \left[T^{-1} \sum_{j\neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right]^2 \right)^{1/2} \left(\sum_{i=1}^{N} \left[\frac{1}{\left(\varsigma_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right]^2 \right)^{1/2} \\
&\leq T k_{N,T} \left(\sum_{i=1}^{N} \left[T^{-1} \sum_{j\neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right]^2 \right)^{1/2} \left(\sum_{i=1}^{N} \left[\frac{1}{\left(\widehat{\sigma}_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} \right]^2 \right)^{1/2} \\
&= \sqrt{\frac{T}{2N(N-1)}} \sqrt{a_{21}^* a_{22}^*}. \qquad \text{(S.71)}
\end{aligned}$$

A further first-order Taylor approximation of $\left(\widehat{\sigma}_i^2\right)^{-3/2}$ around $\left(\sigma_i^2\right)^{-3/2}$ results in

$$\frac{1}{\left(\widehat{\sigma}_i^2\right)^{3/2}} - \frac{1}{\left(\sigma_i^2\right)^{3/2}} = -3/2 \frac{1}{\left(\varsigma_i^2\right)^{5/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)$$

Substituting this expression into a_{22}^* leads to

$$\begin{aligned}
a_{22}^* &= \sum_{i=1}^{N} \left[-3/2 \frac{1}{\left(\varsigma_i^2\right)^{5/2}} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \right]^2 \\
&\leq 9/4 \left(\sup_i \frac{1}{\left(\varsigma_i^2\right)^{5/2}} \right) \sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)^2 \\
&= 9/4 \left(\frac{1}{\left(\inf_i \varsigma_i^2\right)^{5/2}} \right) \left[\mathcal{O}_P\left(N/T\right) + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right) \right] \\
&= \mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(N^{-1}\right). \qquad \text{(S.72)}
\end{aligned}$$

where result (S.14) is used in line three and Lemma S.9 in the last line to ensure that its result holds as long as $T^{-1}\sqrt{N} \to 0$ is satisfied. In addition to result (S.72), which bounds a_{22}^*, the order in probability of a_2^* is determined by that of a_{21}^*. Lemma S.17 establishes the corresponding result on a_{21}^*.

Accordingly, we have

$$\begin{aligned}
|a_2^*| &= \left[\mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right) \right] \left[\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) \right] \\
&= \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-3/2}\sqrt{T}\right). \qquad \text{(S.73)}
\end{aligned}$$

This result marks the last one needed to characterize the impact of estimating error variances rather than knowing their true value. We return to expression (S.60) and use our results (S.70) and (S.73) on its two components a_1^* and a_2^* to arrive at

$$k_{N,T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_i}{\left(\varsigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)=\mathcal{O}_P\left(N^{-1}\sqrt{T}\right)+\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(T^{-1}\sqrt{N}\right).$$

S.5.8 Auxiliary lemmas for Section S.5.6

Lemma S.5. *Suppose that Assumption 1 holds. Then,*

$$\widehat{\sigma}_i^2-\sigma_i^2=\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\right) \tag{S.74}$$

$$N^{-1}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)=\mathcal{O}_P\left[(NT)^{-1/2}\right]+\mathcal{O}_P\left(N^{-1}\right) \tag{S.75}$$

$$\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2=\mathcal{O}_P\left(N/T\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right) \tag{S.76}$$

Proof of Lemma S.5. All orders in probability in this proof are derived via application of Markov's or Chebyshev's inequality. Details are left out for the sake of brevity. For the first result, we can explicitly write out all sums in equation (S.49) in order to arrive at

$$\begin{aligned}\widehat{\sigma}_i^2-\sigma_i^2&=T^{-1}\sum_{t=1}^{T}\left(\varepsilon_{i,t}^2-\sigma_i^2\right)+T^{-1}\sum_{t=1}^{T}\left(\bar{\varepsilon}_t\right)^2+\left(T^{-1}\sum_{t=1}^{T}\varepsilon_{i,t}\right)^2+\left[(NT)^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{j,t}\right]^2\\&\quad-2\,(NT)^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}\varepsilon_{j,t}-2\left(T^{-1}\sum_{t=1}^{T}\varepsilon_{i,t}\right)\left[(NT)^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{j,t}\right]\\&=\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left[(NT)^{-1/2}\right].\end{aligned}$$

For the second result, note that summation over N simplifies the decomposition in equation (S.49) to

$$\begin{aligned}N^{-1}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)&=(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{i,t}^2-\sigma_i^2\right)-N^{-2}T^{-1}\sum_{i,j}\sum_{t=1}^{T}\varepsilon_{i,t}\varepsilon_{j,t}+N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}\varepsilon_{i,t'}-\bar{\varepsilon}^2\\&=\mathcal{O}_P\left[(NT)^{-1/2}\right]+\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right).\end{aligned}$$

For the third result, we borrow from the explicit expression of $\widehat{\sigma}_i^2-\sigma_i^2$ from the first result above and describe an upper bound as

$$
\begin{aligned}
\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2 &\leq M\sum_{i=1}^{N}\left[T^{-1}\sum_{t=1}^{T}\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\right]^2+MN\left[T^{-1}\sum_{t=1}^{T}\left(\overline{\varepsilon}_t\right)^2\right]^2+M\sum_{i=1}^{N}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_{i,t}\right)^4\\
&+MN\overline{\varepsilon}^4+M\sum_{i=1}^{N}\left[(NT)^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\right]^2+2\sum_{i=1}^{N}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_{i,t}\right)^2\overline{\varepsilon}^2\\
&=\frac{M}{T^2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)+\frac{MN}{N^4T^2}\sum_{i,i',i'',i'''}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i''',t'}\\
&+\frac{MN}{T^4}\sum_{i=1}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}+MN\overline{\varepsilon}^4\\
&+\frac{M}{(NT)^2}\sum_{i,j,j'}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}+\overline{\varepsilon}^2\frac{M}{T^2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\varepsilon_{i,t'}\\
&=\mathcal{O}_P\left(N/T\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right).
\end{aligned}
$$

□

Lemma S.6. *Suppose that Assumption 1 holds. Then,*

$$\sum_{i=1}^{N}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_{i,t}\overline{\varepsilon}_t\right)^2=\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right) \tag{S.77}$$

$$k_{N,T}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\left(\frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i}{T}-\sigma_i^2\right)=\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(T^{-1/2}\right) \tag{S.78}$$

$$k_{N,T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\left(\frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i}{T}-\sigma_i^2\right)=\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(T^{-1}\sqrt{N}\right) \tag{S.79}$$

$$\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\right)^2\left(\frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i}{T}-\sigma_i^2\right)^2=\mathcal{O}_P\left(N^2\right) \tag{S.80}$$

$$\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\right)^2\left(\frac{\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{\varepsilon}}}{T}\right)^2=\mathcal{O}_P\left(N\right)+\mathcal{O}_P\left(T\right) \tag{S.81}$$

$$\sum_{i=1}^{N}\overline{\varepsilon}_i\left(\frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i}{T}-\sigma_i^2\right)=\mathcal{O}_P\left(NT^{-1}\right) \tag{S.82}$$

$$\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\overline{\varepsilon}_i\overline{\varepsilon}_j\right)^2\left(\frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i}{T}-\sigma_i^2\right)^2=\mathcal{O}_P\left(N^2T^{-1}\right) \tag{S.83}$$

$$\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\overline{\varepsilon}_i\overline{\varepsilon}_j\right)^2\left(\frac{\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{\varepsilon}}}{T}\right)^2=\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(1\right) \tag{S.84}$$

Proof of Lemma S.6. Non-negativity of the first expression allows us to look at

$$\mathrm{E}\left[T^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\bar{\varepsilon}_t\bar{\varepsilon}_{t'}\varepsilon_{i,t'}\right] = (NT)^{-2}\sum_{i,i',i''}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\right]$$
$$= \mathcal{O}\left(N^{-1}\right) + \mathcal{O}\left(T^{-1}\right),$$

where the last line results from eliminating terms where indices cannot be matched to arrive at a non-zero expectation term.

Consider the second result now. We can write

$$k_{N,T}\sum_{i=1}^{N}\varepsilon_i'\varepsilon_i\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right) = Tk_{N,T}\sum_{i=1}^{N}\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)^2 + Tk_{N,T}\sum_{i=1}^{N}\sigma_i^2\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right).$$
$$= N^{-1}\sqrt{T}\left[\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(\sqrt{NT}\right)\right]$$
$$= \mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right).$$

The first order term in the second line above is given by results in the proof of (S.76). For the second term, we use Chebyshev's inequality and boundedness of σ_i^2.

We proceed with the third term. For notational simplicity we will abbreviate $\sum_{i=1}^{N}\sum_{j\neq i}^{N}$ by $\sum_{i\neq j}^{N}$ in this proof. Given that $i \neq j$, the expression

$$\frac{k_{N,T}}{T}\sum_{i\neq j}^{N}\sum_{t,t'}^{T}\varepsilon_{j,t}\varepsilon_{i,t}\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)$$

has a zero expected value, we derive its order in probability from its variance. In the expression

$$E\left\{\left[\frac{k_{N,T}}{T}\sum_{i\neq j}^{N}\sum_{t,t'}^{T}\varepsilon_{j,t}\varepsilon_{i,t}\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)\right]^2\right\} = \frac{k_{N,T}^2}{T^2}\sum_{i\neq j}^{N}\sum_{i'\neq j'}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{j,t}\varepsilon_{i,t}\left(\varepsilon_{it'}^2-\sigma_i^2\right)\varepsilon_{j',t''}\varepsilon_{i',t''}\left(\varepsilon_{i',t'''}^2-\sigma_{i'}^2\right)$$

a number of terms cancels out because independence of errors across cross-sections and across time leads to a number of terms that depend on the zero expected value of ε_{it}. While a detailed account of all individual cases is omitted here for the sake of brevity, we can state that the variance of $k_{N,T}T^{-1}\sum_{i\neq j}^{N}\sum_{t,t'}^{T}\varepsilon_{j,t}\varepsilon_{i,t}\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)$ reduces to

$$E\left\{\left[\frac{k_{N,T}}{T}\sum_{i\neq j}^{N}\sum_{t,t'}^{T}\varepsilon_{j,t}\varepsilon_{i,t}\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)\right]^2\right\}$$
$$= \frac{1}{4N^2T^3}\sum_{i\neq j}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left[\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)^2\right]$$
$$+ \frac{1}{4N^2T^3}\sum_{i\neq i'\neq j}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t}^3\right)\mathrm{E}\left(\varepsilon_{i',t}^3\right)+\mathrm{o}\left(T^{-1}\right)+\mathrm{o}\left(T^{-2}N\right)$$
$$= \mathcal{O}\left(T^{-1}\right)+\mathcal{O}\left(T^{-2}N\right).$$

Result (S.79) follows from the equation above by the application of Chebyshev's inequality.

Before applying Markov's inequality to establish result number four, note that all elements in the sum

$$\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\left(\frac{\varepsilon_{i,t''}^{2}}{T}-\sigma_i^2\right)\left(\frac{\varepsilon_{i,t'''}^{2}}{T}-\sigma_i^2\right)$$

whose indexes violate either $j=j'$ or $t=t'$ will have an expected value of zero. This is a direct consequence of the fact that $i\neq j$ and $i\neq j'$ generally holds. If $j=j'$ and $t=t'$ are satisfied we also need the index restriction $t''=t'''$ to ensure a zero expected value. Hence,

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)^2\right] &= T^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t''}^{T}\mathrm{E}\left[\varepsilon_{i,t}^2\varepsilon_{j,t}^2\left(\varepsilon_{i,t''}^2-\sigma_i^2\right)^2\right]+\mathrm{o}\left(N^2\right)\\ &=\mathcal{O}\left(N^2\right),\end{aligned}$$

which leads to result (S.80).

Moving to result (S.81), we continue to look for restrictions on the indexes which would result in nonzero expected values. However, since the term under consideration involves more sums over cross-section indexes, we limit our attention to restrictions that would lead to the maximal number of sums rather than documenting all possible index restrictions which result in nonzero expectations. Taking expectations, we have

$$\mathrm{E}\left[\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right]^2\left(\frac{\varepsilon_i'\bar{\varepsilon}}{T}\right)^2\right]=(NT)^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t'''}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\varepsilon_{i,t'}\varepsilon_{i',t''}\varepsilon_{i'',t'''}\varepsilon_{i,t'''}\right].$$

The set of index restrictions maximizing the number of sums over cross-sections is $j=j'$, $i'=i''$, $t=t'$ and $t''=t'''$. This leads to three sums over cross-sections as well as two sums over time, resulting in a term of order $\mathcal{O}(N)$. The alternative set of index restrictions $j=j', i=i'=i''$ and $t=t'$ maximizes number of sums over time and leads to a term of order $\mathcal{O}(T)$. Summarizing these two results, we can state that application of Markov's inequality establishes that

$$\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right]^2\left(\frac{\varepsilon_i'\bar{\varepsilon}}{T}\right)^2=\mathcal{O}_P(N)+\mathcal{O}_P(T).$$

The last three results in Lemma S.6 are very similar to the previous three expressions targeted above. For this reason, the following proofs will leave out unnecessary details. For result (S.82), we write out the expression of interest as $T^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)$. Taking first squares and then expectations, we get

$$
\begin{aligned}
&E\left\{\left[T^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\left(\varepsilon_{i,t'}^2-\sigma_i^2\right)\right]^2\right\}\\
&=T^{-4}\sum_{i,j}^{N}\sum_{t,t',t'',t'''}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{j,t'}\left(\varepsilon_{i,t''}^2-\sigma_i^2\right)\left(\varepsilon_{j,t'''}^2-\sigma_i^2\right)\right]\\
&=T^{-4}\sum_{i,j}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}^3\right)\mathrm{E}\left(\varepsilon_{i',t''}^3\right)+\mathrm{o}\left(N^2T^{-2}\right)\\
&=\mathcal{O}\left(N^2T^{-2}\right),
\end{aligned}
$$

from which the desired result follows.

Concerning result (S.83), we write

$$
\begin{aligned}
&E\left\{\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}T\bar{\varepsilon}_i\bar{\varepsilon}_j\right]^2\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)^2\right\}\\
=&T^{-4}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t'}^2\right)\mathrm{E}\left[\left(\varepsilon_{i,t''}^2-\sigma_i^2\right)^2\right]+T^{-4}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^3\right)\mathrm{E}\left(\varepsilon_{j,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^3\right)\\
&+\mathrm{o}\left(N^2T^{-1}\right)\\
=&\mathcal{O}\left(N^2T^{-1}\right).
\end{aligned}
$$

which leads to result (S.83).

Moving to result (S.84), we can state that

$$
\begin{aligned}
&E\left\{\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}T\bar{\varepsilon}_i\bar{\varepsilon}_j\right]^2\left(\frac{\varepsilon_i'\bar{\varepsilon}}{T}\right)^2\right\}\\
=&N^{-2}T^{-4}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t',t'',t'''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i,t'''}^2\right)\\
&+N^{-2}T^{-4}\sum_{i,i'}^{N}\sum_{j\neq i}^{N}\sum_{t,t',t'''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i',t''}^2\right)+\mathrm{o}\left(1\right)+\mathrm{o}\left(NT^{-1}\right)\\
=&\mathcal{O}\left(1\right)+\mathcal{O}\left(NT^{-1}\right),
\end{aligned}
$$

so that we arrive at result (S.84). □

Lemma S.7. *Suppose that Assumption 1 holds. Then,*

$$
k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)=\mathbf{1}\left[\mathrm{E}\left(\varepsilon_{it}^3\right)\neq 0\right]\mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\sqrt{T}\right)
$$

Proof of Lemma S.7. We use the decomposition of $\widehat{\sigma}_i^2$ in equation (S.49) to write

$$k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)=k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)+k_{N,T}\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}\left(\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\right)+k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\overline{\varepsilon}_i^2$$
$$+k_{N,T}\overline{\varepsilon}^2\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j-2k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\frac{\overline{\varepsilon}'\varepsilon_i}{T}-2k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\overline{\varepsilon}_i\overline{\varepsilon}$$
$$=a_{1111}+a_{1112}+a_{1113}+a_{1114}-a_{1115}-a_{1116}.$$

The leading term in the decomposition above is a_{1111} for which results (S.79) and (S.78) can be used to arrive at

$$a_{1111}=k_{N,T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\left(T^{-1}\varepsilon_i'\varepsilon_i-\sigma_i^2\right)+k_{N,T}\sum_{i=1}^{N}\varepsilon_i'\varepsilon_i\left(T^{-1}\varepsilon_i'\varepsilon_i-\sigma_i^2\right)$$
$$=\mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right).$$

Next, terms a_{1112} and a_{1114} can be combined and rewritten in order to yield

$$a_{1112}+a_{1114}=k_{N,T}\left(\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}+\overline{\varepsilon}^2\right)N^2T\left(\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}\right)$$
$$=k_{N,T}\left\{\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left[(NT)^{-1}\right]\right\}\mathcal{O}_P\left(NT\right)$$
$$=\mathcal{O}_P\left(N^{-1}\sqrt{T}\right)$$

Expression a_{1113} is non-negative since $\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j$ is and because $\overline{\varepsilon}_i^2$ does not change the sign of ε_i. We can hence directly apply expectations without taking squares in beforehand. Hence, we get

$$\mathrm{E}\left(a_{1113}\right)=k_{N,T}T^{-2}\,\mathrm{E}\left(\sum_{i,j}^{N}\sum_{t,t',t''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{i,t'}\varepsilon_{i,t''}\right)$$
$$=k_{N,T}T^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}\right)\mathrm{E}\left(\varepsilon_{i,t'}\right)+\mathrm{o}\left(T^{-1/2}\right)$$
$$=\mathcal{O}\left(T^{-1/2}\right),$$

which leads to $a_{1113}=\mathcal{O}_P\left(T^{-1/2}\right)$. The non-negativity holds as well for $a_{1115}=2k_{N,T}NT^{-1}\sum_{i=1}^{N}\overline{\varepsilon}'\varepsilon_i\varepsilon_i'\overline{\varepsilon}$, so that

$$\mathrm{E}\left(a_{1115}\right)=\frac{k_{N,T}}{NT}\sum_{i,j}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)+\frac{k_{N,T}}{NT}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)$$
$$+\mathrm{o}\left(T^{-1/2}\right)+\mathrm{o}\left(N^{-1}\sqrt{T}\right)$$
$$=\mathcal{O}\left(T^{-1/2}\right)+\mathcal{O}\left(N^{-1}\sqrt{T}\right)$$

implies $a_{1115} = \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right)$. Only in the case of a_{1116} we have to take squares before applying expectations. Here,

$$\begin{aligned}
\mathrm{E}\left[\left(k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\bar{\varepsilon}_i\right)^2\right] &= \frac{k_{N,T}^2}{N^2}\sum_{i,i',j}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i',t}^2\right) + \frac{k_{N,T}^2}{N^2}\sum_{i=1}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right) \\
&\quad + 2\frac{k_{N,T}^2}{N^2}\sum_{i,i'}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\varepsilon_{i',t'}^2\right) + \mathrm{o}\left(N^{-1}\right) + \mathrm{o}\left(N^{-3}T^2\right) + \mathrm{o}\left(N^{-2}T\right) \\
&= \mathcal{O}\left(N^{-1}\right) + \mathcal{O}\left(N^{-3}T^2\right) + \mathcal{O}\left(N^{-2}T\right),
\end{aligned}$$

so that we arrive at

$$\begin{aligned}
a_{1116} &= \mathcal{O}_P\left[(NT)^{-1/2}\right]\left[\mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-3/2}T\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right)\right] \\
&= \mathcal{O}_P\left(N^{-1}T^{-1/2}\right) + \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left(N^{-2}\sqrt{T}\right).
\end{aligned}$$

Combining our results on terms a_{1111} to a_{1116}, we arrive at

$$k_{N,T}\sum_{i,j}^{N}\varepsilon_i'\varepsilon_j\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) = \mathbf{1}\left[\mathrm{E}\left(\varepsilon_{it}^3\right) \neq 0\right]\mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(N^{-1}\sqrt{T}\right).$$

□

Lemma S.8. *Suppose that Assumption 1 holds. Then,*

$$Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) = \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}T^{-1}\right) + \mathcal{O}_P\left(N^{-3/2}\right)$$

Proof of Lemma S.8. We use the decomposition of $\widehat{\sigma}_i^2$ in equation (S.49) to write

$$\begin{aligned}
Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) &= Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i\left(\frac{\varepsilon_i'\varepsilon_i}{T} - \sigma_i^2\right) + Tk_{N,T}\left(\frac{\bar{\varepsilon}'\bar{\varepsilon}}{T} + \bar{\varepsilon}^2\right)\sum_{i=1}^{N}\bar{\varepsilon}_i + Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i^3 \\
&\quad - 2Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i\frac{\bar{\varepsilon}'\varepsilon_i}{T} - 2Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i^2\bar{\varepsilon} \\
&= a_{1151} + a_{1152} + a_{1153} + a_{1154} - a_{1155}.
\end{aligned}$$

Here, using result (S.82), we directly get

$$\begin{aligned}
a_{1151} &= Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i\left(\frac{\varepsilon_i'\varepsilon_i}{T} - \sigma_i^2\right) \\
&= N^{-1}\sqrt{T}\,\mathcal{O}_P\left(NT^{-1}\right) \\
&= \mathcal{O}_P\left(T^{-1/2}\right)
\end{aligned}$$

Furthermore, since $T^{-1}\overline{\varepsilon}'\overline{\varepsilon} = \mathcal{O}_P\left(N^{-1}\right)$ and $\overline{\varepsilon} = \mathcal{O}_P\left[(NT)^{-1/2}\right]$, we have

$$\begin{aligned}
a_{1152} &= Tk_{N,T}\left(\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T} + \overline{\varepsilon}^2\right)(N\overline{\varepsilon}) \\
&= N^{-1}\sqrt{T}\,\mathcal{O}_P\left(N^{-1}\right)\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right) \\
&= \mathcal{O}_P\left(N^{-3/2}\right)
\end{aligned}$$

For a_{1153}, we acknowledge that

$$\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{N}\overline{\varepsilon}_i^3\right)^2\right] &= T^{-6}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^3\right)\right]^2 + T^{-6}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\right]^3 + \mathrm{o}\left(N^2T^{-4}\right) + \mathrm{o}\left(NT^{-3}\right) \\
&= \mathcal{O}\left(N^2T^{-4}\right) + \mathcal{O}\left(NT^{-3}\right)
\end{aligned}$$

in order to arrive at

$$\begin{aligned}
a_{1153} &= Tk_{N,T}\left[\mathcal{O}_P\left(NT^{-2}\right) + \mathcal{O}_P\left(\sqrt{N}T^{-3/2}\right)\right] \\
&= \mathcal{O}_P\left(T^{-3/2}\right) + \mathcal{O}_P\left(N^{-1/2}T^{-1}\right).
\end{aligned}$$

Next, we move on to a_{1154} which we express as

$$a_{1155} = 2k_{N,T}\left(NT\right)^{-1}\sum_{i,i'}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\varepsilon_{i',t'}\varepsilon_{i,t'}.$$

Taking squares of the double sum over cross-sections and time and applying expectations, we get

$$\begin{aligned}
\mathrm{E}\left[\left(\sum_{i,i'}^{N}\sum_{t,t'}^{T}\varepsilon_{i,t}\varepsilon_{i',t'}\varepsilon_{i,t'}\right)^2\right] &= \sum_{i,i',i''}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i',t}^2\right)\mathrm{E}\left(\varepsilon_{i'',t}^2\right) + \sum_{i=1}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right) \\
&\quad + \sum_{i,i'}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}^3\right)\mathrm{E}\left(\varepsilon_{i',t'}^3\right) + \mathrm{o}\left(N^3T\right) + \mathrm{o}\left(NT^3\right) + \mathrm{o}\left[(NT)^2\right] \\
&= \mathcal{O}\left(N^3T\right) + \mathcal{O}\left(NT^3\right) + \mathcal{O}\left[(NT)^2\right],
\end{aligned}$$

implying that

$$\begin{aligned}
a_{1154} &= k_{N,T}\left(NT\right)^{-1}\left\{\mathcal{O}_P\left(N^{3/2}\sqrt{T}\right) + \mathcal{O}_P\left(\sqrt{N}T^{3/2}\right) + \mathcal{O}_P\left(NT\right)\right\}. \\
&= k_{N,T}\left\{\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\sqrt{T}\right) + \mathcal{O}_P\left(1\right)\right\} \\
&= \mathcal{O}_P\left(N^{-1/2}T^{-1}\right) + \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left(N^{-1}T^{-1/2}\right)
\end{aligned}$$

Lastly, we have $a_{1155} = Tk_{N,T}\overline{\varepsilon}\sum_{i=1}^{N}\overline{\varepsilon}_i^2$ where $\sum_{i=1}^{N}\overline{\varepsilon}_i^2 = \mathcal{O}_P\left(NT^{-1}\right)$ by Markov's and Rosenthal's inequalities, leading us to $a_{1155} = \mathcal{O}_P\left(N^{-1/2}T^{-1}\right)$. Combining our results on terms a_{1151}

to a_{1155}, we can state that

$$Tk_{N,T}\sum_{i=1}^{N}\bar{\varepsilon}_i\left(\widehat{\sigma}_i^2-\sigma_i^2\right)=\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}T^{-1}\right)+\mathcal{O}_P\left(N^{-3/2}\right)$$

□

Lemma S.9. *Suppose that Assumption 1 holds and assume that $NT^{-2}\to 0$. Then there exists a constant $M>0$ such that $\inf_i \varsigma_i^2\geq M$ with probability approaching one.*

Proof of Lemma S.9. First, note that $\inf_i\varsigma_i^2\geq\inf_i\sigma_i^2-\sup_i\left|\widehat{\sigma}_i^2-\sigma_i^2\right|$. Here, σ_i^2 is bounded from below by Assumption 1 . Concerning $\sup_i\left|\widehat{\sigma}_i^2-\sigma_i^2\right|$, we consider the tail probability

$$\begin{aligned}\Pr\left(\sup_i\left|\widehat{\sigma}_i^2-\sigma_i^2\right|>\epsilon\right)&\leq\sum_{i=1}^{N}\Pr\left(\left|\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_i}{T}-\sigma_i^2\right|>\epsilon\right)\\&\leq\sum_{i=1}^{N}\Pr\left(\left|\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right|>\frac{\epsilon}{6}\right)+\sum_{i=1}^{N}\Pr\left(\frac{\bar{\varepsilon}'\bar{\varepsilon}}{T}>\frac{\epsilon}{6}\right)+\sum_{i=1}^{N}\Pr\left(\bar{\varepsilon}_i^2>\frac{\epsilon}{6}\right)\\&\quad+\sum_{i=1}^{N}\Pr\left(\bar{\varepsilon}^2>\frac{\epsilon}{6}\right)+\sum_{i=1}^{N}\Pr\left(\left|\frac{\varepsilon_i'\bar{\varepsilon}}{T}\right|>\frac{\epsilon}{6}\right)+\sum_{i=1}^{N}\Pr\left(\left|\bar{\varepsilon}_i\bar{\varepsilon}\right|>\frac{\epsilon}{6}\right)\end{aligned}$$

Sums over probabilities involving $\frac{\bar{\varepsilon}'\bar{\varepsilon}}{T}$, $\left|\frac{\varepsilon_i'\bar{\varepsilon}}{T}\right|$,$\bar{\varepsilon}^2$, $\bar{\varepsilon}_i^2$ and $|\bar{\varepsilon}_i\bar{\varepsilon}|$ can be shown to converge to zero via application of Chebyshev's inequality. For the remaining probability in terms of $\left|\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right|$, we choose the power of four instead. Given that the random variable $\left(\varepsilon_{i,t}^2-\sigma_i^2\right)$ has zero expected value and is independent across cross-sections and (conditionally) time, by Rosenthals' inequality:

$$\mathrm{E}\left[T^{-1}\sum_{t=1}^{T}\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\right]^4=\mathrm{E}(\sigma_i^8)\,\mathrm{E}\left[T^{-1}\sum_{t=1}^{T}\left(\eta_{i,t}^2-1\right)\right]^4=\mathcal{O}(T^{-2})$$

so that it holds that $\sum_{i=1}^{N}\left\{\mathrm{E}\left[T^{-1}\sum_{t=1}^{T}\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\right]^4\right\}=\mathcal{O}\left(NT^{-2}\right)$. Accordingly,

$$\begin{aligned}\sum_{i=1}^{N}\Pr\left(\left|\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right|>\epsilon\right)&\leq\sum_{i=1}^{N}\frac{\mathrm{E}\left[\left(T^{-1}\varepsilon_i'\varepsilon_i-\sigma_i^2\right)^4\right]}{\epsilon^4}\\&=\frac{N}{T^2}\frac{1}{\epsilon^4}\frac{T^2}{N}\sum_{i=1}^{N}E\left\{\left[T^{-1}\sum_{t=1}^{T}\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\right]^4\right\}\\&=\mathcal{O}_P\left(NT^{-2}\right).\end{aligned}$$

Hence under the assumption that $NT^{-2}\to 0$, the above expression is $o_P(1)$. This result implies that the statement made in Lemma S.9 holds. □

Lemma S.10. *Under Assumption 1 ,*

$$\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\sigma_j}\right)^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2=\mathcal{O}_P\left(N^2\right)+\mathcal{O}_P\left(NT\right)+\mathcal{O}_P\left(N^{-1}T^2\right)$$

Proof of Lemma S.10. To begin with, note that σ_j is bounded from below at a value greater than zero by Assumption 1 . Hence, σ_j^{-1} is bounded from above by some finite number M. Accordingly, we can write the term

$$\begin{aligned}
&\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\sigma_j}\right)^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2\\
\leq &M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2+M\sum_{i=1}^{N}\left[(N-1)\left(\overline{\varepsilon}'\overline{\varepsilon}+T\overline{\varepsilon}^2\right)\right]^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2\\
+&M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\overline{\varepsilon}_i\overline{\varepsilon}_j\right)^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2+M\sum_{i=1}^{N}\left[(N-1)\,\varepsilon_i'\overline{\varepsilon}\right]^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2\\
+&M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\overline{\varepsilon}'\varepsilon_j\right)^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2+M\sum_{i=1}^{N}\left[T\left(N-1\right)\overline{\varepsilon}_i\overline{\varepsilon}\right]^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2+M\sum_{i=1}^{N}\left(T\sum_{j\neq i}^{N}\overline{\varepsilon}_j\overline{\varepsilon}\right)^2\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2\\
=&a_{211}+a_{212}+a_{213}+a_{214}+a_{215}+a_{216}+a_{217}. \qquad \text{(S.85)}
\end{aligned}$$

The first out of seven terms in (S.85) can be decomposed into

$$\begin{aligned}
a_{211}\leq M&\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)^2+\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\left(\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}+\overline{\varepsilon}^2\right)^2+\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\overline{\varepsilon}_i^4\\
&+M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\left(\frac{\varepsilon_i'\overline{\varepsilon}}{T}\right)^2+M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\left(\overline{\varepsilon}_i\overline{\varepsilon}\right)^2 \qquad \text{(S.86)}
\end{aligned}$$

For the first and fourth expression on the right-hand side above we use results (S.80) and (S.81). Concerning the second expression, we only need to find an upper bound for the non-negative term $\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2$. Taking expectations here, we get

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\right]&=\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\varepsilon_{j,t}^2\right)\\
&=\mathcal{O}\left(N^2T\right).
\end{aligned}$$

Given that $\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}=\mathcal{O}_P\left(N^{-1}\right)$ and $\overline{\varepsilon}^2=\mathcal{O}_P\left[(NT)^{-1}\right]$, we get $\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\left(\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}\right)^2=\mathcal{O}_P\left(T\right)$. Next, we consider

$$\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\overline{\varepsilon}_i^4=T^{-4}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t''',t'''',t'''''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}\varepsilon_{i,t''''}\varepsilon_{i,t'''''},$$

whose expected value is

$$\mathrm{E}\left[\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\bar{\varepsilon}_i^4\right] = T^{-4}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t'',t'''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i,t'''}^2\right) + \mathrm{o}\left(N^2T^{-1}\right)$$
$$= \mathcal{O}\left(N^2T^{-1}\right),$$

which implies that $\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\bar{\varepsilon}_i^4 = \mathcal{O}_P\left(N^2T^{-1}\right)$. Lastly, the fifth term in (S.86) requires us to determine the order of

$$\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\bar{\varepsilon}_i^2 = T^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}.$$

Taking expectations, we get

$$\mathrm{E}\left[\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2\bar{\varepsilon}_i^2\right] = T^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right) + \mathrm{o}\left(N^2\right)$$
$$= \mathcal{O}\left(N^2\right).$$

Additionally, recalling that $\bar{\varepsilon}^2 = \mathcal{O}_P\left[(NT)^{-1}\right]$ leads to $\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\right)^2(\bar{\varepsilon}_i\bar{\varepsilon})^2 = \mathcal{O}_P\left(NT^{-1}\right)$.

Summarizing our results, we arrive at

$$a_{211} = \mathcal{O}_P\left(N^2\right) + \mathcal{O}_P\left(T\right) \tag{S.87}$$

The second out of seven terms in (S.85) is bounded from above by application of result (S.76). This allows us to write the term as

$$a_{212} \leq MN^2\left(\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}} + T\bar{\varepsilon}^2\right)^2\sum_{i=1}^{N}\left(\hat{\sigma}_i^2 - \sigma_i^2\right)^2$$
$$= \mathcal{O}_P\left(NT\right) + \mathcal{O}_P\left(N^{-1}T^2\right). \tag{S.88}$$

We continue with the third out of seven terms in (S.85) which is written

$$a_{213} \leq M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\bar{\varepsilon}_i\bar{\varepsilon}_j\right)^2\left(\frac{\varepsilon_i'\varepsilon_i}{T} - \sigma_i^2\right)^2 + \sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\bar{\varepsilon}_i\bar{\varepsilon}_j\right)^2\left(\frac{\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}}}{T} + \bar{\varepsilon}^2\right)^2 + \sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\bar{\varepsilon}_i\bar{\varepsilon}_j\right)^2\bar{\varepsilon}_i^4$$
$$+ M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\bar{\varepsilon}_i\bar{\varepsilon}_j\right)^2\left(\frac{\varepsilon_i'\bar{\boldsymbol{\varepsilon}}}{T}\right)^2 + M\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}T\bar{\varepsilon}_i\bar{\varepsilon}_j\right)^2\left(\bar{\varepsilon}_i\bar{\varepsilon}\right)^2$$
$$= a_{2131} + a_{2132} + a_{2133} + a_{2134} + a_{2135}.$$

Application of results (S.83) and (S.84) leads to $a_{2131} = \mathcal{O}_P\left(N^2T^{-1}\right)$ and $a_{2134} = \mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(1\right)$. In order to derive orders for the remaining terms in a_{213}, we implicitly use the fact that, for $r > 2$, $\bar{\varepsilon}_i^r = \mathcal{O}\left(T^{-r/2}\right)$ by Rosenthal's inequality for independent random variables. For term a_{2132}, we note that independence of errors across cross-sections implies

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N} T\bar{\varepsilon}_i\bar{\varepsilon}_j\right)^2\right] &= \sum_{i=1}^{N}\mathrm{E}\left(\bar{\varepsilon}_i^2\right)\mathrm{E}\left[\left(\sum_{j\neq i}^{N}\sum_{t''=1}^{T}\varepsilon_{j,t''}\right)^2\right] \\
&= \mathcal{O}_P\left(NT^{-1}\right)\mathcal{O}_P\left(NT\right) \\
&= \mathcal{O}_P\left(N^2\right),
\end{aligned}$$

from which it follows that

$$\begin{aligned}
a_{2132} &= \mathcal{O}_P\left(N^2\right)\left\{\mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1}\right]\right\} \\
&= \mathcal{O}_P\left(N\right).
\end{aligned}$$

By the same reasoning, we have

$$\begin{aligned}
\mathrm{E}\left(a_{2133}\right) &= \sum_{i=1}^{N}\mathrm{E}\left(\bar{\varepsilon}_i^6\right)\mathrm{E}\left[\left(\sum_{j\neq i}^{N}\sum_{t=1}^{T}\varepsilon_{j,t}\right)^2\right] \\
&= \mathcal{O}\left(NT^{-3}\right)\mathcal{O}\left(NT\right) \\
&= \mathcal{O}\left(N^2T^{-2}\right),
\end{aligned}$$

which leads to $a_{2133} = \mathcal{O}_P\left(N^2T^{-2}\right)$. Finally, concerning term a_{2135} we can write

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{N}\sum_{j\neq i}^{N}\bar{\varepsilon}_i^4\bar{\varepsilon}_j^2\right) &= T^2\sum_{i=1}^{N}\mathrm{E}\left(\bar{\varepsilon}_i^4\right)\mathrm{E}\left[\left(\sum_{j\neq i}^{N}\sum_{t=1}^{T}\varepsilon_{j,t}\right)^2\right] \\
&= T^2\,\mathcal{O}\left(NT^{-2}\right)\mathcal{O}\left(NT\right) \\
&= \mathcal{O}\left(N^2T^{-3}\right),
\end{aligned}$$

so that

$$\begin{aligned}
a_{2135} &= MT^2\bar{\varepsilon}^2\sum_{i=1}^{N}\bar{\varepsilon}_i^4\left(\sum_{j\neq i}^{N}\bar{\varepsilon}_j^2\right) \\
&= T^2\,\mathcal{O}_P\left[(NT)^{-1}\right]\mathcal{O}_p\left(N^2T^{-3}\right) \\
&= \mathcal{O}_P\left(NT^{-2}\right).
\end{aligned}$$

This result, together with the four previous ones, establishes

$$a_{213} = \mathcal{O}_P\left(N^2T^{-1}\right) + \mathcal{O}_P\left(N\right)$$

Concerning the fourth out of seven terms in (S.85), we write

$$
\begin{aligned}
a_{214} \leq & MN^2T^{-2}\sum_{i=1}^{N}\left(\varepsilon_i'\overline{\varepsilon}\right)^2\left(\varepsilon_i'\varepsilon_i-\sigma_i^2\right)^2+MN^2\sum_{i=1}^{N}\left(\varepsilon_i'\overline{\varepsilon}\right)^2\left(\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}+\overline{\varepsilon}^2\right)^2 \\
& +MN^2\sum_{i=1}^{N}\left(\varepsilon_i'\overline{\varepsilon}\right)^2\overline{\varepsilon}_i^4+MN^2T^{-2}\sum_{i=1}^{N}\left(\varepsilon_i'\overline{\varepsilon}\right)^4+N^2\sum_{i=1}^{N}\left(\varepsilon_i'\overline{\varepsilon}\right)^2\overline{\varepsilon}_i^2\overline{\varepsilon}^2 \\
= & a_{2141}+a_{2142}+a_{2143}+a_{2144}+a_{2145}.
\end{aligned}
$$

Concerning a_{2142}, we use result (S.77) which allows us to write

$$
\begin{aligned}
a_{2142} &= N^2\left[\mathcal{O}_P\left(N^{-1}T^2\right)+\mathcal{O}_P\left(T\right)\right]\mathcal{O}_P\left(N^{-2}\right) \\
&= \mathcal{O}_P\left(N^{-1}T^2\right)+\mathcal{O}_P\left(T\right).
\end{aligned}
$$

With regards to

$$
a_{2141} = MT^{-2}\sum_{i,i',i''}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\left(\varepsilon_{i,t''}^2-\sigma_i^2\right)\left(\varepsilon_{i,t'''}^2-\sigma_i^2\right),
$$

we proceed by looking for index restrictions that result in the maximum number of sums over cross-sections or time, respectively. The restrictions $i'=i'', t=t'$ and $t''=t'''$ maximize the number of sums over N, leading to an expression of order $T^{-2}\,\mathcal{O}_P\left[(NT)^2\right]$. Alternatively, the restrictions $i=i'=i''$ and $t''=t'''$ maximize the number of sums over T and result in a term of order $T^{-2}\,\mathcal{O}_P\left(NT^3\right)$. Overall, we can hence state that.

$$
a_{2141} = \mathcal{O}_P\left(N^2\right)+\mathcal{O}_P\left(NT\right).
$$

Next, a statement concerning

$$
a_{2144} = M\left(NT\right)^{-2}\sum_{i,i',i'',i''',i''''}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i''',t''}\varepsilon_{i,t'''}\varepsilon_{i'''',t''}
$$

can be made by noting that the index restrictions $i'=i''$, $i'''=i''''$, $t=t'$ and $t''=t'''$ lead to one component with non-zero expected value consisting of three sums over cross-sections and two sums over time. The corresponding component is hence $(NT)^{-2}\,\mathcal{O}_P\left(N^3T^2\right)$. Additionally, the index restrictions $i=i'=i''=i'''=i''''$ lead to another component of order $(NT)^{-2}\,\mathcal{O}_P\left(NT^4\right)$. Moreover, a third component of order $(NT)^{-2}\,\mathcal{O}_P\left(N^2T^3\right)$ results from the index restrictions $i=i'''=i''''$, $i'=i''$ and $t=t'$. Hence, we can state that

$$
a_{2144} = \mathcal{O}_P\left(N\right)+\mathcal{O}_P\left(T\right)+\mathcal{O}_P\left(N^{-1}T^2\right).
$$

We continue with the similar expression

$$
a_{2143} = MT^{-4}\sum_{i,i',i''}^{N}\sum_{t,t',t'',t''',t'''',t'''''}^{T}\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}\varepsilon_{i,t''''}\varepsilon_{i,t'''''}.
$$

Applying expectations and looking for index restrictions which imply nonzero expected values, we obtain

$$\begin{aligned}\mathrm{E}\left(a_{2143}\right) &= MT^{-4}\sum_{i,i'}^{N}\sum_{t,t'',t'''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^{2}\right)\mathrm{E}\left(\varepsilon_{i',t}^{2}\right)\mathrm{E}\left(\varepsilon_{i,t''}^{2}\right)\mathrm{E}\left(\varepsilon_{i,t'''}^{2}\right)+T^{-4}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^{2}\right)\right]^{4}\\ &\quad+\mathrm{o}\left(N^{2}T^{-1}\right)+\mathrm{o}\left(N\right)\\ &=\mathcal{O}\left(N^{2}T^{-1}\right)+\mathcal{O}\left(N\right),\end{aligned}$$

from which it follows that $a_{2143}=\mathcal{O}_P\left(N^{2}T^{-1}\right)+\mathcal{O}_P\left(N\right)$. Lastly, we focus on

$$a_{2145}=\bar{\varepsilon}^{2}T^{-2}\sum_{i,i',i''}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}$$

where

$$\begin{aligned}&\mathrm{E}\left(T^{-2}\sum_{i,i',i''}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}\right)\\ =&T^{-2}\sum_{i,i'}^{N}\sum_{t,t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^{2}\right)\mathrm{E}\left(\varepsilon_{i',t}^{2}\right)\mathrm{E}\left(\varepsilon_{i,t''}^{2}\right)\\ +&T^{-2}\sum_{i=1}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^{2}\right)\mathrm{E}\left(\varepsilon_{i,t'}^{2}\right)\mathrm{E}\left(\varepsilon_{i,t''}^{2}\right)+\mathrm{o}\left(N^{2}\right)+\mathrm{o}\left(NT\right)\\ =&\,\mathcal{O}\left(N^{2}\right)+\mathcal{O}\left(NT\right).\end{aligned}$$

Accordingly, it holds that

$$\begin{aligned}a_{2145} &= \mathcal{O}_P\left[\left(NT\right)^{-1}\right]\left[\mathcal{O}_P\left(N^{2}\right)+\mathcal{O}_P\left(NT\right)\right]\\ &=\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(1\right).\end{aligned}$$

Using this result, we can now state that the fourth out of seven terms in (S.85) is

$$a_{214}=\mathcal{O}_P\left(N^{2}\right)+\mathcal{O}_P\left(NT\right)+\mathcal{O}_P\left(N^{-1}T^{2}\right) \tag{S.89}$$

We continue with the fifth out of seven terms in (S.85) which we bound from above by extending the sums over j and j' with the cases $j=i$ and $j'=i$. Accordingly, we can write

$$a_{215}\leq M\left(NT\right)^{2}\left(\frac{\overline{\varepsilon}'\overline{\varepsilon}}{T}\right)^{2}\sum_{i=1}^{N}\left(\widehat{\sigma}_{i}^{2}-\sigma_{i}^{2}\right)^{2}$$

Using result (S.76) on the term $\sum_{i=1}^{N}\left(\widehat{\sigma}_{i}^{2}-\sigma_{i}^{2}\right)^{2}$, we get

$$\begin{aligned}a_{215} &= \left(NT\right)^{2}\mathcal{O}_P\left(N^{-2}\right)\left[\mathcal{O}_P\left(N/T\right)+\mathcal{O}_P\left(N^{-1}\right)\right]\\ &=\mathcal{O}_P\left(NT\right)+\mathcal{O}_P\left(N^{-1}T^{2}\right).\end{aligned} \tag{S.90}$$

The sixth out of seven terms in (S.85) is quite similar to the fourth term seen above. We can write

$$\begin{aligned}
a_{216} &\leq M\left(NT\right)^2 \bar{\varepsilon}^2 \sum_{i=1}^{N} \bar{\varepsilon}_i^2 \left(\frac{\varepsilon_i'\varepsilon_i}{T} - \sigma_i^2\right)^2 + M\left(NT\right)^2 \left[\bar{\varepsilon}^2 \left(\frac{\bar{\varepsilon}'\bar{\varepsilon}}{T}\right)^2 + \bar{\varepsilon}^6\right] \sum_{i=1}^{N} \bar{\varepsilon}_i^2 \\
&+ M\left(NT\right)^2 \bar{\varepsilon}^2 \sum_{i=1}^{N} \bar{\varepsilon}_i^2 \left(\frac{\bar{\varepsilon}'\varepsilon_i}{T}\right)^2 + M\left(NT\right)^2 \bar{\varepsilon}^2 \sum_{i=1}^{N} \bar{\varepsilon}_i^6 \\
&+ M\left(NT\right)^2 \bar{\varepsilon}^4 \sum_{i=1}^{N} \bar{\varepsilon}_i^4 \\
&= a_{2161} + a_{2162} + a_{2163} + a_{2164} + a_{2165}.
\end{aligned}$$

Concerning a_{2161}, we need to have a look at the expected value of $\sum_{i=1}^{N} \bar{\varepsilon}_i^2 \left(\frac{\varepsilon_i'\varepsilon_i}{T} - \sigma_i^2\right)^2$. Here,

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{N} \bar{\varepsilon}_i^2 \left(\frac{\varepsilon_i'\varepsilon_i}{T} - \sigma_i^2\right)^2\right] &= 2T^{-4} \sum_{i=1}^{N} \sum_{t,t'}^{T} \varepsilon_{i,t}^2 \left(\varepsilon_{i,t'}^2 - \sigma_i^2\right)^2 + 2T^{-4} \sum_{i=1}^{N} \sum_{t,t'}^{T} \varepsilon_{i,t}^3 \varepsilon_{i,t'}^3 + \mathcal{O}\left(NT^{-2}\right) \\
&= \mathcal{O}\left(NT^{-2}\right),
\end{aligned}$$

which implies $a_{2161} = \left(NT\right)^2 \mathcal{O}_P\left[\left(NT\right)^{-1}\right] \mathcal{O}_P\left(NT^{-2}\right) = \mathcal{O}_P\left(N^2T^{-1}\right)$. For the next term, knowledge of $\sum_{i=1}^{N} \bar{\varepsilon}_i^2 = \mathcal{O}_P\left(NT^{-1}\right)$ allows us to arrive at $a_{2162} = \mathcal{O}_P\left(1\right)$. For term a_{2163}, we look at $\sum_{i=1}^{N} \bar{\varepsilon}_i^2 \left(\frac{\bar{\varepsilon}'\varepsilon_i}{T}\right)^2$, which has expected value

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{N} \bar{\varepsilon}_i^2 \left(\frac{\bar{\varepsilon}'\varepsilon_i}{T}\right)^2\right] &= N^{-2}T^{-4} \sum_{i=1}^{N} \sum_{t,t',t''}^{T} \mathrm{E}\left(\varepsilon_{i,t}^2\right) \mathrm{E}\left(\varepsilon_{i,t'}^2\right) \mathrm{E}\left(\varepsilon_{i,t''}^2\right) + N^{-2}T^{-4} \sum_{i,i'}^{N} \sum_{t,t'}^{T} \mathrm{E}\left(\varepsilon_{i,t}^2\right) \mathrm{E}\left(\varepsilon_{i,t'}^2\right) \mathrm{E}\left(\varepsilon_{i',t'}^2\right) \\
&+ \mathrm{o}\left[\left(NT\right)^{-1}\right] + \mathrm{o}\left(T^{-2}\right) \\
&= \mathcal{O}\left[\left(NT\right)^{-1}\right] + \mathcal{O}\left(T^{-2}\right).
\end{aligned}$$

Accordingly, we have

$$\begin{aligned}
a_{2163} &= \left(NT\right)^2 \mathcal{O}_P\left[\left(NT\right)^{-1}\right] \left\{\mathcal{O}_P\left[\left(NT\right)^{-1}\right] + \mathcal{O}_P\left(T^{-2}\right)\right\} \\
&= \mathcal{O}_P\left(1\right) + \mathcal{O}_P\left(NT^{-1}\right).
\end{aligned}$$

The last two terms are very similar. Using Markov's inequality, it is straight forward to show that $\sum_{i=1}^{N} \bar{\varepsilon}_i^4 = \mathcal{O}_P\left(NT^{-2}\right)$, and using the same reasoning we arrive at $\sum_{i=1}^{N} \bar{\varepsilon}_i^6 = \mathcal{O}_P\left(NT^{-3}\right)$. It follows that $a_{2164} = \mathcal{O}_P\left(N^2T^{-2}\right)$ and $a_{2165} = \mathcal{O}_P\left(NT^{-2}\right)$. For the entire term a_{216}, we can hence state that

$$a_{216} = \mathcal{O}_P\left(N^2T^{-1}\right) + \mathcal{O}_P\left(1\right).$$

The last building block needed for the central result of this lemma is term a_{217}. Similar to our treatment of term a_{215},we write

$$\begin{aligned} a_{217} &\leq (NT)^2 M\bar{\varepsilon}^4 \sum_{i=1}^{N} \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)^2 \\ &= \mathcal{O}_P\left(N/T\right) + \mathcal{O}_P\left(N^{-1}\right), \end{aligned} \tag{S.91}$$

where the last line follows from result (S.76). Combining the results on terms (S.87) to (S.91), we arrive at the final result of this lemma, namely

$$\sum_{i=1}^{N} \left(\sum_{j \neq i}^{N} \frac{\widehat{\varepsilon}_i' \widehat{\varepsilon}_j}{\sigma_j} \right)^2 \left(\widehat{\sigma}_i^2 - \sigma_i^2\right)^2 = \mathcal{O}_P\left(N^2\right) + \mathcal{O}_P\left(NT\right) + \mathcal{O}_P\left(N^{-1}T^2\right).$$

□

S.5.9 Auxiliary lemmas for Section S.5.7

Lemma S.11. *Suppose that Assumptions 2-6 hold. Then,*

$$\left\| T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}} \right\| = \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right] \tag{S.92}$$

$$\sum_{i=1}^{N} \left\| T^{-1}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}} \right\|^2 = \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right) \tag{S.93}$$

$$\sum_{i=1}^{N} \left\| T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}} \right\|^2 = \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right) \tag{S.94}$$

$$\sum_{i=1}^{N} \left\| T^{-1}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}} \right\|^2 = \mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(N^{-1}\right) \tag{S.95}$$

$$\sum_{i=1}^{N} \left\| T^{-1}\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}} \right\|^2 = \mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(N^{-1}\right) \tag{S.96}$$

$$\sum_{i=1}^{N} \left\| T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i \right\|^2 = \mathcal{O}_P\left(NT^{-2}\right) + \mathcal{O}_P\left(N^{-2}\right) \tag{S.97}$$

$$\sum_{i=1}^{N} \left\| T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i \right\|^2 = \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right) \tag{S.98}$$

$$\sum_{i=1}^{N} \left(T^{-1}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i \right)^2 = \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right) \tag{S.99}$$

$$T^{-2}\sum_{i,j}^{N} \left(\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_i\right)^2 \boldsymbol{\varepsilon}_j\overline{\boldsymbol{U}}\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i. = \mathcal{O}_P\left(1\right) \tag{S.100}$$

$$T^{-2}\sum_{i,j}^{N} \left(\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_i\right)^2 \boldsymbol{\varepsilon}_j'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i = \mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(1\right) \tag{S.101}$$

$$\left\| T^{-1}\overline{\boldsymbol{U}}'\widehat{\boldsymbol{F}} \right\|^2 = \mathcal{O}_P\left[(NT)^{-1}\right] + \mathcal{O}_P\left(N^{-2}\right) \tag{S.102}$$

$$\sum_{i=1}^{N} \left(\left\| T^{-1}\sum_{j\neq i}^{N} \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\boldsymbol{F} \right\|^2 \right) = \mathcal{O}_P\left(N^2\right) \tag{S.103}$$

$$\sum_{i=1}^{N} \left(\left\| T^{-1}\sum_{j\neq i}^{N} \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}} \right\|^2 \right) = \mathcal{O}_P\left(N\right) + \mathcal{O}_P\left(T\right) \tag{S.104}$$

Proof of Lemma S.11. For the first result, note that

$$\begin{aligned} \left\| T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}} \right\| &\leq \left\| T^{-1}\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}} \right\| + \left\| T^{-1}\overline{\boldsymbol{U}}'\widehat{\boldsymbol{F}} \right\|^2 \left\| \left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1} \right\| \\ &= \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right] \end{aligned}$$

by the results in Lemma S.2.

In order to arrive at (S.93), we write

$$\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2 \leq T^{-2}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\left\|\boldsymbol{\lambda}_i\right\|^2$$

Taking expectations of this non-negative term, we obtain

$$\begin{aligned}
\mathrm{E}\left(T^{-2}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\left\|\boldsymbol{\lambda}_i\right\|^2\right) &= (NT)^{-2}\sum_{i,i'}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i',t}^2\right)\mathrm{E}\left(\left\|\boldsymbol{\lambda}_i\right\|^2\right)\\
&+ (NT)^{-2}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\left\|\boldsymbol{\lambda}_i\right\|^2\right)\\
&+ (NT)^{-2}\sum_{i,i'}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\boldsymbol{e}_{i',t}'\boldsymbol{e}_{i',t}\right)\mathrm{E}\left(\left\|\boldsymbol{\lambda}_i\right\|^2\right) + \mathcal{O}\left[(NT)^{-1}\right]\\
&= \mathcal{O}\left(N^{-1}\right) + \mathcal{O}\left(T^{-1}\right),
\end{aligned}$$

so that the required result follows by the Markov's inequality. The proof of (S.94) uses the same reasoning and is therefore omitted.

For result (S.95), we can expand the expression into

$$\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}\right\|^2 \leq 3\sum_{i=1}^{N}\left\|T^{-1}\frac{\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\boldsymbol{F}}{\left(\sigma_i^2\right)^{3/2}}\right\|^2\left\|\overline{\boldsymbol{C}}\right\|^2 + 3\sum_{i=1}^{N}\left\|T^{-1}\frac{\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}}{\left(\sigma_i^2\right)^{3/2}}\right\|^2.$$

If we take expectations of the first part on the right-hand side above, we get

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{N}\left\|T^{-1}\frac{\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\boldsymbol{F}}{\left(\sigma_i^2\right)^{3/2}}\right\|^2\right) &\leq T^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\left\|\boldsymbol{\lambda}_i\right\|^2\right)\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\boldsymbol{F}_t'\boldsymbol{F}_t\right)\\
&= \mathcal{O}\left(NT^{-1}\right).
\end{aligned}$$

Using this result together with Markov's inequality and result (S.93), we arrive at

$$\sum_{i=1}^{N}\left\|T^{-1}\frac{\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}}{\left(\sigma_i^2\right)^{3/2}}\right\|^2 = \mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(N^{-1}\right).$$

The proof of result (S.96) is almost identical to that of (S.95) and therefore omitted.

We proceed with result (S.97), where we write

$$\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\right\|^2 \leq \sum_{i=1}^{N}\left\|T^{-2}\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|.$$

Furthermore, from Lemma S.2 $\left\|\left(\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\| = \mathcal{O}_P(1)$. It remains to consider

$$\sum_{i=1}^{N}\left\|T^{-2}\varepsilon_i'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\varepsilon_i\right\|^2 \leq 3\sum_{i=1}^{N}\left\|T^{-2}\varepsilon_i'\boldsymbol{F}\boldsymbol{F}'\varepsilon_i\right\|^2\left\|\overline{\boldsymbol{C}}\right\|^4 + 3\sum_{i=1}^{N}\left\|T^{-2}\varepsilon_i'\overline{\boldsymbol{U}}\overline{\boldsymbol{U}}'\varepsilon_i\right\|^2.$$

Here,

$$\begin{aligned}\sum_{i=1}^{N}\left\|T^{-2}\varepsilon_i'\boldsymbol{F}\boldsymbol{F}'\varepsilon_i\right\|^2 &= T^{-4}\sum_{i=1}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}\boldsymbol{f}_t'\boldsymbol{f}_{t'}\boldsymbol{f}_{t''}'\boldsymbol{f}_{t'''}\\ &= \mathcal{O}_P\left(NT^{-2}\right)\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{N}\left\|T^{-2}\varepsilon_i'\overline{\boldsymbol{U}}\overline{\boldsymbol{U}}'\varepsilon_i\right\|^2 &\leq \sum_{i=1}^{N}\left\|T^{-1/2}\varepsilon_i\right\|^4\left\|T^{-1/2}\overline{\boldsymbol{U}}\right\|^4\\ &= T^{-4}\sum_{i=1}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}^2\varepsilon_{i,t'}^2\varepsilon_{i,t''}^2\varepsilon_{i,t'''}^2\left\|T^{-1/2}\overline{\boldsymbol{U}}\right\|^4\\ &= \mathcal{O}_P\left(N^{-2}\right)\end{aligned}$$

follow from application of Markov's inequality and the elimination of terms with zero expected value. Accordingly, we have

$$\sum_{i=1}^{N}\left\|T^{-2}\varepsilon_i'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\varepsilon_i\right\|^2 = \mathcal{O}_P\left(NT^{-2}\right) + \mathcal{O}_P\left(N^{-2}\right),$$

which establishes result (S.97).

Next, consider result (S.98). Here we use results (S.74), (S.94), (S.96) and (S.102) to arrive at

$$\begin{aligned}\sum_{i=1}^{N}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right\|^2 &\leq 3\sum_{i=1}^{N}\left\|T^{-1}\overline{\boldsymbol{U}}'\varepsilon_i\right\|^2 + 3\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2\left\|T^{-1}\overline{\boldsymbol{U}}'\widehat{\boldsymbol{F}}\right\|^2\sum_{i=1}^{N}\left\|T^{-1}\widehat{\boldsymbol{F}}'\varepsilon_i\right\|^2\\ &= \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right)\\ &\quad + \left\{\mathcal{O}_P\left[(NT)^{-1}\right] + \mathcal{O}_P\left(N^{-2}\right)\right\}\left[\mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(N^{-1}\right)\right]\\ &= \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right).\end{aligned}$$

The proof of result (S.99) is identical to that seen immediately above except for the use of results (S.93) and (S.95) instead of (S.94) and (S.96).

We move on with results that involve a double sum over cross-sections. To derive result

(S.100), we take expectations in order to obtain

$$
\begin{aligned}
& T^{-2}\sum_{i,j}^{N}\mathrm{E}\left[\left(\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_i\right)^2\right]\mathrm{E}\left[\boldsymbol{\varepsilon}_j'\overline{\boldsymbol{U}\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\right] \\
\leq & MT^{-2}\sum_{i,j}^{N}\mathrm{E}\left[\boldsymbol{\varepsilon}_j'\overline{\boldsymbol{U}\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\right] \\
= & (NT)^{-2}\sum_{i,i',j,j'}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{j,t}\varepsilon_{j',t}\varepsilon_{i',t'}\varepsilon_{i,t'}\right)\|\boldsymbol{B}\|^2+(NT)^{-2}\sum_{i,i',j,j'}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{j,t}\varepsilon_{i,t'}\right)\mathrm{E}\left(\boldsymbol{e}_{j',t}'\boldsymbol{e}_{i',t'}\right)\|\boldsymbol{B}\|^2 \\
= & (NT)^{-2}\sum_{i,j}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{j,t}^2\varepsilon_{i,t'}^2\right)\|\boldsymbol{B}\|^2+(NT)^{-2}\sum_{i,i'}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\boldsymbol{e}_{i',t}'\boldsymbol{e}_{i',t}\right)\|\boldsymbol{B}\|^2, \\
= & \mathcal{O}(1)+\mathcal{O}\left(T^{-1}\right),
\end{aligned}
$$

from which result (S.100) directly follows. The result also constitutes one half of the required steps to prove result (S.101). The corresponding expression can be written

$$
T^{-2}\sum_{i,j}^{N}\left(\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_i\right)^2\boldsymbol{\varepsilon}_j'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i\leq\frac{3}{T^2}\sum_{i,j}^{N}\left(\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_i\right)^2\boldsymbol{\varepsilon}_j'\boldsymbol{F}\boldsymbol{F}'\boldsymbol{\varepsilon}_i\left\|\overline{\boldsymbol{C}}\right\|^2+\frac{3}{T^2}\sum_{i,j}^{N}\left(\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_i\right)^2\boldsymbol{\varepsilon}_j'\overline{\boldsymbol{U}\boldsymbol{U}}'\boldsymbol{\varepsilon}_i,
$$

where we have already shown that the second term on the right-hand side is stochastically bounded. Taking expectations of the first term on the right-hand side above results in

$$
\begin{aligned}
\mathrm{E}\left[T^{-2}\sum_{i,j}^{N}\sum_{t,t'}^{T}\left(\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_i\right)^2\varepsilon_{j,t}\boldsymbol{F}_t'\boldsymbol{F}_{t'}\varepsilon_{i,t'}\right] &= T^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\|\boldsymbol{\lambda}_i\|^4\right)\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\boldsymbol{F}_t'\boldsymbol{F}_{t'}\right) \\
&= \mathcal{O}\left(NT^{-1}\right),
\end{aligned}
$$

resulting in a corresponding order in probability for the term without expectations. Combining both result allows us to arrive at (S.101).

Result (S.102) is proven in the same way as result (S.101)and a detailed proof is therefore omitted. To see the link between both terms, note that

$$
\left\|T^{-1}\overline{\boldsymbol{U}}'\widehat{\boldsymbol{F}}\right\|^2=(NT)^{-2}\sum_{i,j}^{N}\mathrm{tr}\left(\boldsymbol{u}_i\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\boldsymbol{u}_j\right),
$$

where $\boldsymbol{u}_i=\boldsymbol{B}'[\varepsilon_{i,t},\boldsymbol{e}_{i,t}]'$.

Concerning result (S.103), we write out the expectation of the norm as

$$
\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\boldsymbol{F}\right\|^2\right) &\leq MT^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t'''}^{T}\mathrm{E}\left(\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{i,t'}\varepsilon_{j',t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}\right)\mathrm{E}\left(\boldsymbol{f}_{t''}'\boldsymbol{f}_{t'''}\right) \\
{}'f &= MT^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\|\boldsymbol{f}_{t''}\|^2\right)+\mathrm{o}\left(N^2\right) \\
&= \mathcal{O}\left(N^2\right),
\end{aligned}
$$

from which result (S.103) directly follows.

We proceed with result (S.104), where we can expend the term under investigation into

$$\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\varepsilon_i'\overline{U}\right\|^2\leq M\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\varepsilon_i'\overline{\varepsilon}\right\|^2+M\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\varepsilon_i'\overline{e}\right\|^2.$$

Here the order of the first result is determined by result (S.81). Concerning the second result, an proof analogous to that of (S.103) allows us to show that $\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\left(\sigma_j^2\right)^{1/2}\varepsilon_i'\varepsilon_j\varepsilon_i'\overline{e}\right\|^2=\mathcal{O}_P\left(N\right)$. The combination of these two results establishes (S.104). □

Lemma S.12. *Suppose that Assumptions 2-6 hold. Then,*

$$\left\|T^{-1}\sum_{i=1}^{N}\varepsilon_i'\widehat{F}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right\|=\mathcal{O}_P\left(NT^{-3/2}\right)+\mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1/2}\right)$$

Proof of Lemma S.12. We decompose the term under consideration into

$$\begin{aligned}\left\|T^{-1}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\varepsilon_i'\widehat{F}\right\|\leq&\left\|T^{-2}\sum_{i=1}^{N}\left(\varepsilon_i'\varepsilon_i-\sigma_i^2\right)\varepsilon_i'\widehat{F}\right\|+\left\|T^{-2}\sum_{i=1}^{N}\lambda_i'\left(\overline{C}^{-1}\right)'\overline{U}'M_{\widehat{F}}\overline{U}\left(\overline{C}^{-1}\right)\lambda_i\varepsilon_i'\widehat{F}\right\|\\&+\left\|T^{-2}\sum_{i=1}^{N}\varepsilon_i'P_{\widehat{F}}\varepsilon_i\varepsilon_i'\widehat{F}\right\|+\left\|2T^{-2}\sum_{i=1}^{N}\lambda_i'\left(\overline{C}^{-1}\right)'\overline{U}'M_{\widehat{F}}\varepsilon_i\varepsilon_i'\widehat{F}\right\|\\&=g_{21}+g_{22}+g_{23}+g_{24}.\end{aligned}$$

Consider now

$$\begin{aligned}g_{21}\leq&\left\|T^{-2}\sum_{i=1}^{N}\left(\varepsilon_i'\varepsilon_i-\sigma_i^2\right)\varepsilon_i'F\right\|\left\|\overline{C}\right\|+\left\|T^{-2}\sum_{i=1}^{N}\left(\varepsilon_i'\varepsilon_i-\sigma_i^2\right)\varepsilon_i'\overline{U}\right\|\\&=g_{211}\left\|\overline{C}\right\|+g_{212}.\end{aligned}$$

We take squares of the first term in this decomposition and take expectations. This allows us to get

$$\begin{aligned}\mathrm{E}\left(g_{211}^2\right)\leq&\, T^{-4}\sum_{i,j}^{N}\sum_{t,t',t'',t'''}^{T}\mathrm{E}\left[\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\varepsilon_{i,t'}\varepsilon_{j,t''}\left(\varepsilon_{j,t'''}^2-\sigma_j^2\right)\right]\mathrm{E}\left(f_{t'}'f_{t''}\right),\\&=T^{-4}\sum_{i,j}^{N}\sum_{t,t''}^{T}\mathrm{E}\left[\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\varepsilon_{i,t}\right]\mathrm{E}\left[\varepsilon_{j,t''}\left(\varepsilon_{j,t''}^2-\sigma_j^2\right)\right]\mathrm{E}\left(f_t'f_{t''}\right)\\&+T^{-4}\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\left(\varepsilon_{i,t}^2-\sigma_i^2\right)\left(\varepsilon_{j,t}^2-\sigma_j^2\right)\right]\mathrm{E}\left[\varepsilon_{i,t'}^2\right]\mathrm{E}\left(f_{t'}'f_{t''}\right)+\mathcal{O}\left(NT^{-2}\right)\\&=\mathcal{O}\left(NT^{-2}\right).\end{aligned}$$

Accordingly, $g_{211} = \mathcal{O}_P\left(T^{-1}\sqrt{N}\right)$ results from application of Chebyshev's inequality. When dealing with g_{212},we first split the entire expression using the Cauchy-Schwarz inequality and establish somewhat more conservative upper bounds on the resulting terms:

$$
\begin{aligned}
g_{212} &\leq \left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2\right)^{1/2}\left(\sum_{i=1}^{N}\left\|T^{-1}\left(\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i-\sigma_i^2\right)\right\|^2\right)^{1/2} \\
&= \left[\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(T^{-1/2}\right)\right]\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right) \\
&= \mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(T^{-1/2}\right).
\end{aligned}
$$

Here, the second line result from (S.94) as well as a straightforward application of Markov's inequality. Given this result, we can state that

$$
g_{21} = \mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(T^{-1/2}\right).
$$

We continue with the use of the Cauchy-Schwarz inequality to split up lengthy terms when handling the remaining terms g_{22}, g_{23} and g_{24}. First, we write

$$
\begin{aligned}
g_{22} &\leq \left(\sum_{i=1}^{N}\left\|T^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}\right\|^2\right)^{1/2}\left(\sum_{i=1}^{N}\|\boldsymbol{\lambda}_i\|^2\right)^{1/2}\left\|\left(\overline{\boldsymbol{C}}^{-1}\right)\right\|^2\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\| \\
&= \left[\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)\right]\mathcal{O}_P\left(\sqrt{N}\right)\left\{\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left[(NT)^{-1/2}\right]\right\} \\
&= \mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\right),
\end{aligned}
$$

which follows from results (S.92) and (S.95). Next, we have

$$
\begin{aligned}
g_{23} &\leq \left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}\right\|^2\right)^{1/2}\left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\right\|^2\right)^{1/2} \\
&= \left[\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)\right]\left[\mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(N^{-1}\right)\right] \\
&= \mathcal{O}_P\left(NT^{-3/2}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left[(NT)^{-1/2}\right]+\mathcal{O}_P\left(N^{-3/2}\right),
\end{aligned}
$$

which relies on results (S.96) and (S.97). Lastly, we consider

$$
g_{24} = \left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\widehat{\boldsymbol{F}}\right\|^2\right)^{1/2}\left\|\overline{\boldsymbol{C}}^{-1}\right\|\left(\sum_{i=1}^{N}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\right\|^2\right)^{1/2}.
$$

The first term on the right-hand side above is $\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)$ by result (S.95). For the last term above we use result (S.98) to show that its it is $\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(T^{-1/2}\right)$. Given these two intermediary results, we can conclude that

$$
\begin{aligned}
g_{24} &= \left[\mathcal{O}_P\left(\sqrt{N}T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)\right]\left[\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left(T^{-1/2}\right)\right] \\
&= \mathcal{O}_P\left(T^{-1}\sqrt{N}\right)+\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1}\right),
\end{aligned}
$$

which is the last building block that we need to establish the main result of this lemma. Combining it with our results on g_{21}, g_{22}, and g_{23}, we arrive at

$$\left\| T^{-1} \sum_{i=1}^{N} \varepsilon_i' \widehat{\boldsymbol{F}} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \right\| = \mathcal{O}_P \left(NT^{-3/2} \right) + \mathcal{O}_P \left(T^{-1}\sqrt{N} \right) + \mathcal{O}_P \left(N^{-1} \right) + \mathcal{O}_P \left(T^{-1/2} \right)$$

□

Lemma S.13. *Suppose that Assumptions 2-6 hold. Then,*

$$k_{N,T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \varepsilon_i' \varepsilon_j \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) = \mathcal{O}_P \left(\sqrt{N} T^{-1} \right) + \mathcal{O}_P \left(T^{-1/2} \right) + \mathcal{O}_P \left(N^{-1/2} \right)$$

Proof of Lemma S.13. For notational simplicity, we abbreviate $\sum_{i=1}^{N} \sum_{j \neq i}^{N}$ by $\sum_{i \neq j}^{N}$ in this proof. We can expand the expression into

$$\begin{aligned}
&\left\| k_{N,T} \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \right\| \\
\leq &\left\| \frac{k_{N,T}}{T} \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \left(\varepsilon_i' \varepsilon_i - \sigma_i^2 \right) \right\| + \left\| \frac{k_{N,T}}{T} \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1} \right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{-1} \right) \boldsymbol{\lambda}_i \right\|. \\
+ &\left\| \frac{k_{N,T}}{T} \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \varepsilon_i' \boldsymbol{P}_{\widehat{\boldsymbol{F}}} \varepsilon_i \right\| + \left\| \frac{2k_{N,T}}{T} \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1} \right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \varepsilon_i \right\| \\
= &g_{31} + g_{32} + g_{33} + g_{34}
\end{aligned}$$

In order to determine the order in probability of g_{31} we can use result (S.79) in Lemma S.6 which was established for the two-way fixed effects model but continues to apply. Hence

$$g_{31} = \mathcal{O}_P \left(T^{-1/2} \right) + \mathcal{O}_P \left(T^{-1}\sqrt{N} \right).$$

We continue with g_{32} where we isolate the inner term $\left(\overline{\boldsymbol{C}}^{-1} \right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{-1} \right)$ by the use of selection vectors for element ℓ out of r element, formally denoted $\boldsymbol{s}_{\ell_r}$:

$$\begin{aligned}
|g_{32}| \leq &\frac{k_{N,T}}{T} \left(\sum_{\ell_r, \ell_r'}^{r} \left\| \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \boldsymbol{\lambda}_i' \boldsymbol{s}_{\ell_r} \boldsymbol{s}_{\ell_r'}' \boldsymbol{\lambda}_i \right\|^2 \right)^{1/2} \left(\sum_{\ell_r, \ell_r'}^{r} \left\| \boldsymbol{s}_{\ell_r}' \left(\overline{\boldsymbol{C}}^{-1} \right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{-1} \right) \boldsymbol{s}_{\ell_r'} \right\|^2 \right)^{1/2} \\
\leq &k_{N,T} \left| \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \boldsymbol{\lambda}_i' \boldsymbol{\lambda}_i \right| \left\| \left(\overline{\boldsymbol{C}}^{-1} \right) \right\|^2 \left\| T^{-1} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \right\|.
\end{aligned}$$

Here we square the term in absolute values and take expectations, resulting in

$$\begin{aligned}
\mathrm{E} \left(\left| \sum_{i \neq j}^{N} \varepsilon_i' \varepsilon_j \boldsymbol{\lambda}_i' \boldsymbol{\lambda}_i \right|^2 \right) &\leq \sum_{i,i'}^{N} \sum_{j \neq i}^{N} \sum_{j' \neq i'}^{N} \sum_{t,t'}^{T} \mathrm{E} \left(\varepsilon_{i,t} \varepsilon_{j,t} \varepsilon_{i't'} \varepsilon_{j',t'} \right) \mathrm{E} \left(\| \boldsymbol{\lambda}_i \|^4 \right) \\
&= \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sum_{t=1}^{T} \mathrm{E} \left(\varepsilon_{i,t}^2 \right) \mathrm{E} \left(\varepsilon_{j,t}^2 \right) \mathrm{E} \left(\| \boldsymbol{\lambda}_i \|^4 \right) \\
&= \mathcal{O} \left(N^2 T \right),
\end{aligned}$$

so that $k_{N,T} \left|\sum_{i\neq j}^{N} \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_j \boldsymbol{\lambda}_i' \boldsymbol{\lambda}_i\right|$ is stochastically bounded. Further using result (S.92), we arrive at

$$g_{32} = \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left[(NT)^{-1/2}\right].$$

Next, consider

$$|g_{33}| \leq T k_{N,T} \left(\sum_{i=1}^{N}\left\|\sum_{j\neq i}^{N} T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\right\|^2\right)^{1/2} \left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\right\|^2\right)^{1/2}.$$

Here we can apply result (S.97) to state that $\left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\right\|^2\right)^{1/2} = \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(N^{-1}\right)$. Additionally, we have

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{N}\left\|\sum_{j\neq i}^{N} T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\right\|^2\right] &\leq T^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right) + 0 \\
&= \mathcal{O}\left(N^2T^{-1}\right).
\end{aligned}$$

This result allows us to arrive at

$$\begin{aligned}
|g_{33}| &= N^{-1}\sqrt{T}\,\mathcal{O}_P\left(NT^{-1/2}\right)\left[\mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(N^{-1}\right)\right] \\
&= \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(N^{-1}\right).
\end{aligned}$$

Lastly, we have the term

$$g_{34} \leq 2Tk_{N,T}\left[\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\lambda}_i'\right\|^2\right]^{1/2}\left\|\overline{\boldsymbol{C}}^{-1}\right\|\left(\sum_{i=1}^{N}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\right\|^2\right)^{1/2}.$$

Concerning the right-hand side of the expression above, $\sum_{i=1}^{N}\left\|\sum_{j\neq i}^{N} T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\lambda}_i'\right\|^2 = \mathcal{O}_P\left(N^2T^{-1}\right)$ can be shown exactly as in the case of g_{33} via application of Markov's inequality. The only extra requirement is $\mathrm{E}\left(\|\boldsymbol{\lambda}_i\|^2\right) < \infty$, which is ensured by Assumption 4. Additionally, we use result (S.98) to make a statement about the sum involving $\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\right\|^2$. Together with our previous result and the factor $k_{N,T}$ we can now state that

$$\begin{aligned}
g_{34} &= N^{-1}\sqrt{T}\,\mathcal{O}_P\left(NT^{-1/2}\right)\left[\mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right)\right]. \\
&= \mathcal{O}_P\left(N^{-1/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right)
\end{aligned}$$

The result on g_{34} is the last one that we require to establish an order in probability for the expression at the center of this lemma. Using the upper bounds derived for g_{31}, g_{32}, g_{33} and g_{34} we conclude that

$$k_{N,T}\sum_{i\neq j}^{N}\frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j}{\left(\sigma_i^2\right)^{3/2}\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) = \mathcal{O}_P\left(T^{-1}\sqrt{N}\right) + \mathcal{O}_P\left(T^{-1/2}\right) + \mathcal{O}_P\left(N^{-1/2}\right)$$

which was to be shown. □

Lemma S.14. *Suppose that Assumptions 2-6. Then,*

$$\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2=\mathcal{O}_P\left(N/T\right)+\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right).$$

$$\sum_{i=1}^{N}\left\|\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\boldsymbol{\lambda}_i\right\|^2=\mathcal{O}_P\left(N/T\right)+\mathcal{O}_P\left(1\right)$$

Proof of Lemma S.14. Using expression (S.62), the left-hand side of the first result can be decomposed into

$$\begin{aligned}\sum_{i=1}^{N}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2 &\leq 5\sum_{i=1}^{N}\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)^2+5T^{-2}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)^2\\ &\quad+5T^{-2}\sum_{i=1}^{N}\left(\varepsilon_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right)^2+10T^{-2}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right)^2\\ &=g_{41}+g_{42}+g_{43}+g_{44}.\end{aligned}$$

We begin with noting that $g_{41}=\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)$ by Lemma S.5 for the two-way fixed effects model. Next, we write term g_{42} as

$$\begin{aligned}g_{42} &\leq \left\|\left(\overline{\boldsymbol{C}}^{-1}\right)\right\|^2\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|^2\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_i\right\|^4\\ &=\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right),\end{aligned}$$

where the order in probability holds due to result (S.92) and Assumption 4. Lastly, results for the terms g_{43} and g_{44} are given by (S.97) and (S.99), respectively. Combining the results on g_{41}, g_{42}, g_{43} and g_{44}, we arrive at the first result of this lemma.

Concerning the second result, we write

$$\begin{aligned}\sum_{i=1}^{N}\left\|\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\boldsymbol{\lambda}_i\right\|^2 &\leq 5\sum_{i=1}^{N}\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)^2\left\|\boldsymbol{\lambda}_i\right\|^2+5T^{-2}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right)^2\left\|\boldsymbol{\lambda}_i\right\|^2\\ &\quad+5T^{-2}\sum_{i=1}^{N}\left(\varepsilon_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right)^2\left\|\boldsymbol{\lambda}_i\right\|^2+10T^{-2}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right)^2\left\|\boldsymbol{\lambda}_i\right\|^2\\ &=g_{45}+g_{46}+g_{47}+g_{48}.\end{aligned}$$

The order of g_{45} is the same as that of g_{41}. This can be seen by noting that

$$\mathrm{E}\left(g_{45}\right)=\sum_{i=1}^{N}\mathrm{E}\left[\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)^2\right]\left(\mathrm{E}\left\|\boldsymbol{\lambda}_i\right\|^2\right),$$

which extends the expectation that is used in the proof of Lemma S.5 for the two-way fixed effects model simply by $\mathrm{E}[\|\boldsymbol{\lambda}_i\|^2]$. Since the second moments of factor loadings are assumed to be finite by assumption, their presence does not affect the order in probability. Hence,

$g_{45} = \mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right)$ holds. The same conclusion can be drawn for terms g_{47} and g_{48}, where the proofs of terms (S.97) and (S.99) are merely extended with an additional term $\boldsymbol{\lambda}_i$ in the sum over cross-sections. For term g_{46},proceeding in the same way would require us to assume finiteness of $\mathrm{E}\left(\|\boldsymbol{\lambda}_i\|^6\right)$, so we choose a more conservative bound that is valid with finite fourth moments of $\boldsymbol{\lambda}_i$. Since $\sum_{i=1}^N \|\boldsymbol{\lambda}_i\|^4 \|\boldsymbol{\lambda}_i\|^2 \leq \sum_{i=1}^N \|\boldsymbol{\lambda}_i\|^4 \sup_i \|\boldsymbol{\lambda}_i\|^2 \leq \left(\sum_{i=1}^N \|\boldsymbol{\lambda}_i\|^4\right)\left(\sum_{i=1}^N \|\boldsymbol{\lambda}_i\|^2\right)$, we get

$$\begin{aligned} g_{46} &\leq \left\|\left(\overline{\boldsymbol{C}}^{-1}\right)\right\| \left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|^2 \sum_{i=1}^N \|\boldsymbol{\lambda}_i\|^4 \sum_{i=1}^N \|\boldsymbol{\lambda}_i\|^2 \\ &= \mathcal{O}_P(1) + \mathcal{O}_P\left(NT^{-1}\right). \end{aligned}$$

Combining this result with the previous three leads to

$$\sum_{i=1}^N \left\|\left(\widehat{\sigma}_i^2 - \sigma_i^2\right)\boldsymbol{\lambda}_i\right\|^2 = \mathcal{O}_P(1) + \mathcal{O}_P\left(NT^{-1}\right).$$

□

Lemma S.15. *Suppose that Assumptions 2-6 hold. Then,*

$$\sum_{i=1}^N \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right)\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2 = \mathcal{O}_P\left(NT^{-3}\right) + \mathcal{O}_P\left(T^{-2}\right) + \mathcal{O}_P\left[(NT)^{-1}\right] + \mathcal{O}_P\left(N^{-2}\right)$$

Proof of Lemma S.15. This term is given by

$$\begin{aligned} \sum_{i=1}^N \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right)\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2 &\leq \sum_{i=1}^N \left\|T^{-1}\left(\frac{\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i}{T} - \sigma_i^2\right)\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2 + \sum_{i=1}^N \left\|T^{-2}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2 \\ &\quad + 4\sum_{i=1}^N \left\|T^{-2}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2 + \sum_{i=1}^N \left\|T^{-2}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2 \\ &= g_{51} + g_{52} + g_{53} + g_{54} \end{aligned}$$

Consider the first term which we can decompose into

$$\begin{aligned} g_{51} &= N^{-2}T^{-4}\sum_{i,j,j'}^N \sum_{t,t',t'',t'''}^T \left(\varepsilon_{i,t}^2 - \sigma_i^2\right)\varepsilon_{i,t'}\varepsilon_{j,t'}\varepsilon_{j,t''}\varepsilon_{i,t''}\left(\varepsilon_{i,t'''}^2 - \sigma_i^2\right) \\ &\quad + N^{-2}T^{-4}\sum_{i,j,j'}^N \sum_{t,t',t'',t'''}^T \left(\varepsilon_{i,t}^2 - \sigma_i^2\right)\varepsilon_{i,t'}\varepsilon_{i,t''}\left(\varepsilon_{i,t'''}^2 - \sigma_i^2\right)\boldsymbol{e}_{j,t'}'\boldsymbol{e}_{j',t''} \\ &= g_{511} + g_{512}. \end{aligned}$$

Taking expectations of g_{511}, we can characterize it as follows in terms of its leading components:

$$\begin{aligned}
\mathrm{E}\left(g_{511}\right) &= \mathrm{E}\left[N^{-2}T^{-4}\sum_{i,j,j'}^{N}\sum_{t,t',t'',t'''}^{T}\left(\varepsilon_{i,t}^{2}-\sigma_{i}^{2}\right)\varepsilon_{i,t'}\varepsilon_{j,t'}\varepsilon_{j',t''}\varepsilon_{i,t''}\left(\varepsilon_{i,t'''}^{2}-\sigma_{i}^{2}\right)\right] \\
&= N^{-2}T^{-4}\sum_{i=1}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left[\left(\varepsilon_{i,t}^{2}-\sigma_{i}^{2}\right)^{2}\right]\mathrm{E}\left(\varepsilon_{i,t'}^{2}\right)\mathrm{E}\left(\varepsilon_{i,t''}^{2}\right) \\
&+ N^{-2}T^{-4}\sum_{i,j}^{N}\sum_{t,t'}^{T}\mathrm{E}\left[\left(\varepsilon_{i,t}^{2}-\sigma_{i}^{2}\right)^{2}\right]\mathrm{E}\left(\varepsilon_{j,t'}^{2}\right)\mathrm{E}\left(\varepsilon_{i,t'}^{2}\right)+\mathrm{o}\left[(NT)^{-1}\right]+\mathrm{o}\left(T^{-2}\right) \\
&= \mathcal{O}\left[(NT)^{-1}\right]+\mathcal{O}\left(T^{-2}\right),
\end{aligned}$$

so that $g_{511}=\mathcal{O}_P\left[(NT)^{-1}\right]+\mathcal{O}_P\left(T^{-2}\right)$. Using an analogous argument, it can be shown that $g_{512}=\mathcal{O}_P\left(T^{-2}\right)$. Accordingly, it holds that

$$g_{51}=\mathcal{O}_P\left[(NT)^{-1}\right]+\mathcal{O}_P\left(T^{-2}\right).$$

We continue with g_{52} to which we apply results (S.92) and (S.93) to arrive at

$$\begin{aligned}
g_{52} &\leq \left\|\overline{\boldsymbol{C}}^{-1}\right\|^{4}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|^{2}\left(\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_{i}'\right\|^{2}\right)\left(T^{-2}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_{i}\boldsymbol{\varepsilon}_{i}'\overline{\boldsymbol{U}}\right\|^{2}\right) \\
&= \left\{\mathcal{O}_P\left(N^{-2}\right)+\mathcal{O}_P\left[(NT)^{-1}\right]\right\}\mathcal{O}_P\left(N\right)\left[\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)\right] \\
&= \mathcal{O}_P\left(N^{-2}\right)+\mathcal{O}_P\left[(NT)^{-1}\right]+\mathcal{O}_P\left(T^{-2}\right).
\end{aligned}$$

Concerning term g_{53}, we write

$$\begin{aligned}
g_{53} &\leq 4\left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\varepsilon}_{i}'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_{i}\right\|^{2}\right)\left(T^{-2}\sum_{i=1}^{N}\left\|\boldsymbol{\varepsilon}_{i}'\overline{\boldsymbol{U}}\right\|^{2}\right) \\
&= \left[\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(N^{-2}\right)\right]\left[\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)\right] \\
&= \mathcal{O}_P\left(NT^{-3}\right)+\mathcal{O}_P\left(T^{-2}\right)+\mathcal{O}_P\left(N^{-2}T^{-1}\right)+\mathcal{O}_P\left(N^{-3}\right),
\end{aligned}$$

where results (S.94) and (S.97) are used to arrive at the second line. Lastly, via results (S.93) and (S.98) we arrive at

$$\begin{aligned}
g_{54} &= \left\|\overline{\boldsymbol{C}}^{-1}\right\|^{2}\left(T^{-2}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_{i}\boldsymbol{\varepsilon}_{i}'\overline{\boldsymbol{U}}\right\|^{2}\right)\left(\sum_{i=1}^{N}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varepsilon}_{i}\right\|^{2}\right) \\
&= \left[\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)\right]\left[\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)\right] \\
&= \mathcal{O}_P\left(N^{-2}\right)+\mathcal{O}_P\left[(NT)^{-1}\right]+\mathcal{O}_P\left(T^{-2}\right).
\end{aligned}$$

Combining the last four intermediary results, we hence arrive at

$$\sum_{i=1}^{N}\left\|T^{-1}\left(\widehat{\sigma}_{i}^{2}-\sigma_{i}^{2}\right)\boldsymbol{\varepsilon}_{i}'\overline{\boldsymbol{U}}\right\|^{2}=\mathcal{O}_P\left(NT^{-3}\right)+\mathcal{O}_P\left(T^{-2}\right)+\mathcal{O}_P\left[(NT)^{-1}\right]+\mathcal{O}_P\left(N^{-2}\right).$$

□

Lemma S.16. *Suppose that Assumptions 2-6 hold. Then,*

$$\sum_{i=1}^{N} \left\| T^{-1} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \varepsilon_i' \widehat{\boldsymbol{F}} \right\|^2 = \mathcal{O}_P \left(NT^{-2} \right) + \mathcal{O}_P \left(T^{-1} \right) + \mathcal{O}_P \left(N^{-2} \right).$$

Proof of Lemma S.16. The expression of interest can bounded from above by

$$\sum_{i=1}^{N} \left\| T^{-1} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \varepsilon_i' \widehat{\boldsymbol{F}} \right\|^2 \leq \sum_{i=1}^{N} \left\| T^{-1} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \varepsilon_i' \boldsymbol{F} \right\|^2 \left\| \overline{\boldsymbol{C}} \right\|^2 + \sum_{i=1}^{N} \left\| T^{-1} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \varepsilon_i' \overline{\boldsymbol{U}} \right\|^2,$$

where we can use Lemma S.15 for the second term on the right-hand side above. It remains to establish a corresponding order in probability for

$$\begin{aligned}
&\sum_{i=1}^{N} \left\| T^{-1} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right) \varepsilon_i' \boldsymbol{F} \right\|^2 \\
\leq & T^{-2} \sum_{i=1}^{N} \left\| \left(\frac{\boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i}{T} - \sigma_i^2 \right) \varepsilon_i' \boldsymbol{F} \right\|^2 + T^{-2} \sum_{i=1}^{N} \left\| T^{-1} \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1} \right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{-1} \right) \boldsymbol{\lambda}_i \varepsilon_i' \boldsymbol{F} \right\|^2 \\
& + 4T^{-2} \sum_{i=1}^{N} \left\| T^{-1} \varepsilon_i' \boldsymbol{P}_{\widehat{\boldsymbol{F}}} \varepsilon_i \varepsilon_i' \boldsymbol{F} \right\|^2 + T^{-2} \sum_{i=1}^{N} \left\| T^{-1} \boldsymbol{\lambda}_i' \left(\overline{\boldsymbol{C}}^{-1} \right)' \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \varepsilon_i \varepsilon_i' \boldsymbol{F} \right\|^2 \\
= & g_{55} + g_{56} + g_{57} + g_{58}.
\end{aligned}$$

Consider now g_{55}, whose expected value is given by

$$\begin{aligned}
\mathrm{E}\left(g_{55} \right) &= T^{-4} \sum_{i=1}^{N} \sum_{t,t''}^{T} \mathrm{E}\left[\left(\varepsilon_{i,t}^2 - \sigma_i^2 \right)^2 \right] \mathrm{E}\left(\varepsilon_{i,t''}^2 \right) \mathrm{E}\left(\boldsymbol{f}_{t''}' \boldsymbol{f}_{t''} \right) \\
&\quad + T^{-4} \sum_{i=1}^{N} \sum_{t,t'}^{T} \mathrm{E}\left(\varepsilon_{i,t}^3 \right) \mathrm{E}\left(\varepsilon_{i,t'}^3 \right) \mathrm{E}\left(\boldsymbol{f}_t' \boldsymbol{f}_{t'} \right) + \mathcal{O}\left(NT^{-2} \right) \\
&= \mathcal{O}\left(NT^{-2} \right),
\end{aligned}$$

implying $g_{55} = \mathcal{O}_P\left(NT^{-2} \right)$. We continue with g_{56} where we can write

$$\begin{aligned}
g_{56} &\leq \left\| \overline{\boldsymbol{C}}^{-1} \right\|^4 \left\| T^{-1} \overline{\boldsymbol{U}}' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}} \right\|^2 \left(\sum_{i=1}^{N} \left\| \boldsymbol{\lambda}_i' \right\|^2 \right) \left(\sum_{i=1}^{N} \left\| \frac{\boldsymbol{\lambda}_i \varepsilon_i' \boldsymbol{F}}{T} \right\|^2 \right) \\
&= \left\{ \mathcal{O}_P\left(N^{-2} \right) + \mathcal{O}_P\left[(NT)^{-1} \right] \right\} \mathcal{O}_P(N) \left[\mathcal{O}_P\left(NT^{-1} \right) + \mathcal{O}_P\left(N^{-1} \right) \right] \\
&= \mathcal{O}_P\left(NT^{-2} \right) + \mathcal{O}_P\left(T^{-1} \right) + \mathcal{O}_P\left(N^{-2} \right).
\end{aligned}$$

The second line above is the consequence of applying results(S.95) and (S.92). It is valid to use the former in order to arrive at an order in probability for $\sum_{i=1}^{N} \| \boldsymbol{\lambda}_i \varepsilon_i' \boldsymbol{F} \|^2$ since $\widehat{\boldsymbol{F}} = \boldsymbol{F} \overline{\boldsymbol{C}} + \overline{\boldsymbol{U}}$. We continue with the third term, g_{57}, which is written out as

$$\begin{aligned}
g_{57} &\leq \sum_{i=1}^{N} \left\| T^{-1} \varepsilon_i' \widehat{\boldsymbol{F}} \right\|^4 \left\| \left(T^{-1} \widehat{\boldsymbol{F}}' \widehat{\boldsymbol{F}} \right) \right\|^2 \left\| T^{-1} \boldsymbol{F}' \varepsilon_i \right\|^2 \\
&\leq M \left\| \left(T^{-1} \widehat{\boldsymbol{F}}' \widehat{\boldsymbol{F}} \right) \right\|^2 \sum_{i=1}^{N} \left\| T^{-1} \varepsilon_i' \boldsymbol{F} \right\|^6 + M \left\| \left(T^{-1} \widehat{\boldsymbol{F}}' \widehat{\boldsymbol{F}} \right) \right\|^2 \left(\sum_{i=1}^{N} \left\| T^{-1} \varepsilon_i' \overline{\boldsymbol{U}} \right\|^4 \right) \left(\sum_{i=1}^{N} \left\| T^{-1} \boldsymbol{F}' \varepsilon_i \right\|^2 \right) \\
&= g_{571} + g_{572}.
\end{aligned}$$

Here, note that

$$\sum_{i=1}^{N}\left\|T^{-1}\varepsilon_i'\boldsymbol{F}\right\|^6=\sum_{i=1}^{N}\sum_{t,t',t'',t''',t'''',t'''''}^{T}\varepsilon_{i,t}\boldsymbol{f}_t'\boldsymbol{f}_{t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\boldsymbol{f}_{t''}'\boldsymbol{f}_{t'''}\varepsilon_{i,t'''}\varepsilon_{i,t''''}\boldsymbol{f}_{t''''}'\boldsymbol{f}_{t'''''}\varepsilon_{i,t'''''}.$$

The expectation of this term is given by

$$\begin{aligned}\mathrm{E}\left(\sum_{i=1}^{N}\left\|T^{-1}\varepsilon_i'\boldsymbol{F}\right\|^6\right)&=T^{-6}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\|\boldsymbol{f}_t\|^2\right)\right]^3+\mathrm{o}\left(NT^{-3}\right)\\&=\mathcal{O}\left(NT^{-3}\right),\end{aligned}$$

from which $g_{571}=\mathcal{O}_P\left(NT^{-3}\right)$ follows. In order to get a corresponding result for g_{572}, we need to investigate the expression

$$\begin{aligned}\sum_{i=1}^{N}\left\|T^{-1}\varepsilon_i'\overline{\boldsymbol{U}}\right\|^4&\leq 3T^{-4}\sum_{i=1}^{N}\left(\varepsilon_i'\overline{\varepsilon}\overline{\varepsilon}'\varepsilon_i\right)^2+3T^{-4}\sum_{i=1}^{N}\left(\varepsilon_i'\overline{e}\overline{e}'\varepsilon_i\right)^2\\&=(NT)^{-4}\sum_{i,i',i'',i''',i''''}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{i',t}\varepsilon_{i'',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i''',t''}\varepsilon_{i'''',t'''}\varepsilon_{i,t'''}\\&\quad+(NT)^{-4}\sum_{i,i',i'',i''',i''''}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\boldsymbol{e}_{i',t}'\boldsymbol{e}_{i'',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\boldsymbol{e}_{i''',t''}'\boldsymbol{e}_{i'''',t'''}\varepsilon_{i,t'''}.\end{aligned}$$

The first term in the decomposition above has expected value

$$\begin{aligned}&\mathrm{E}\left(T^{-4}\sum_{i=1}^{N}\left(\varepsilon_i'\overline{\varepsilon}\overline{\varepsilon}'\varepsilon_i\right)^2\right)\\&=(NT)^{-4}\sum_{i=1}^{N}\sum_{t,t',t'',t'''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i,t'''}^2\right)+3(NT)^{-4}\sum_{i,i',i'''}^{N}\sum_{t,t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i',t}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i''',t''}^2\right)\\&+2(NT)^{-4}\sum_{i,i'''}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i''',t''}^2\right)+\mathrm{o}\left(N^{-3}\right)+\mathrm{o}\left(N^{-1}T^{-2}\right)+\mathrm{o}\left(N^{-2}T^{-1}\right)\\&=\mathcal{O}\left(N^{-1}T^{-2}\right)+\mathcal{O}\left(N^{-3}\right)+\mathcal{O}\left(N^{-2}T^{-1}\right).\end{aligned}$$

Analogously, we can write the expectation of the second term as

$$\begin{aligned}&\mathrm{E}\left(T^{-4}\sum_{i=1}^{N}\left(\varepsilon_i'\overline{e}\overline{e}'\varepsilon_i\right)^2\right)\\&=N^{-2}T^{-4}\sum_{i,i',i''}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\|\boldsymbol{e}_{i',t}\|^2\right)\mathrm{E}\left(\|\boldsymbol{e}_{i'',t'}'\|^2\right)+\mathrm{o}\left(N^{-1}T^{-2}\right)\\&=\mathcal{O}\left(N^{-1}T^{-2}\right).\end{aligned}$$

Consequently, $\sum_{i=1}^{N}\left\|T^{-1}\varepsilon_i'\overline{\boldsymbol{U}}\right\|^4=\mathcal{O}_P\left(N^{-1}T^{-2}\right)+\mathcal{O}_P\left(N^{-2}T^{-1}\right)+\mathcal{O}_P\left(N^{-3}\right)$. In order to arrive at an order in probability for g_{572}, we also need a result for $\sum_{i=1}^{N}\|\boldsymbol{F}'\varepsilon_i\|^2$, a term that is completely contained in $\sum_{i=1}^{N}\left\|\widehat{\boldsymbol{F}}'\varepsilon_i\right\|^2$. Accordingly, we can use result (S.96) to arrive at the upper bound $\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{F}'\varepsilon_i\right\|^2 = \mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right)$. Using this result as well as the former, we can state that

$$\begin{aligned} g_{572} &= \left[\mathcal{O}_P\left(N^{-1}T^{-2}\right)+\mathcal{O}_P\left(N^{-2}T^{-1}\right)+\mathcal{O}_P\left(N^{-3}\right)\right]\left[\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right)\right] \\ &= \mathcal{O}_P\left(T^{-3}\right)+\mathcal{O}_P\left(N^{-2}T^{-1}\right)+\mathcal{O}_P\left(N^{-1}T^{-2}\right)+\mathcal{O}_P\left(N^{-4}\right). \end{aligned}$$

If we now add our result on g_{571}, we arrive at

$$g_{57} = \mathcal{O}_P\left(NT^{-3}\right)+\mathcal{O}_P\left(N^{-1}T^{-2}\right)+\mathcal{O}_P\left(N^{-2}T^{-1}\right)+\mathcal{O}_P\left(N^{-4}\right)$$

Lastly, application of results (S.95) and (S.98) leads to

$$\begin{aligned} g_{58} &= \left\|\overline{\boldsymbol{C}}^{-1}\right\|^2\left(\sum_{i=1}^{N}\left\|T^{-1}\boldsymbol{\lambda}_i\varepsilon_i'\boldsymbol{F}\right\|^2\right)\left(\sum_{i=1}^{N}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right\|^2\right) \\ &= \left[\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right)\right]\left[\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)\right] \\ &= \mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(N^{-2}\right). \end{aligned}$$

Combining our results on terms g_{55}, g_{56}, g_{57} and g_{58}, we arrive at

$$\sum_{i=1}^{N}\left\|T^{-1}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\varepsilon_i'\boldsymbol{F}\right\|^2 = \mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(N^{-2}\right).$$

Together with Lemma S.15, this result allows us to conclude that

$$\sum_{i=1}^{N}\left\|T^{-1}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\varepsilon_i'\widehat{\boldsymbol{F}}\right\|^2 = \mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(N^{-2}\right).$$

□

Lemma S.17. *Suppose that Assumptions 2-6 hold. Then,*

$$\sum_{i=1}^{N}\left[T^{-1}\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\sigma_j}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2 = \mathcal{O}_P\left(N^2T^{-2}\right)+\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(1\right)+\mathcal{O}_P\left(N^{-1}T\right)+\mathcal{O}_P\left(N^{-2}T^2\right)$$

Proof. Analogous to our treatment of term a_1^* we disregard from the scaling σ_i^{-1} since its values have a lower bound above zero. When using a form of Markov's inequality to establish orders in probability, we can find an upper bound for any component of the expression of interest for this lemma.

We decompose the expression of interest into

$$
\begin{aligned}
&\sum_{i=1}^{N}\left[T^{-1}\sum_{j\neq i}^{N}\widehat{\varepsilon}_i'\widehat{\varepsilon}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2\\
=&T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2+T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2\\
&+T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\varepsilon_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2+T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2\\
+&T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_j\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2\\
=&g_{61}+g_{62}+g_{63}+g_{64}+g_{65}.
\end{aligned}
$$

Consider expression g_{62} which can be written

$$
g_{62}\leq\sum_{i=1}^{N}\left(\left\|\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\boldsymbol{\lambda}_i'\right\|^2\left\|\sum_{j\neq i}^{N}\boldsymbol{\lambda}_j\right\|^2\right)\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|^2\left\|\overline{\boldsymbol{C}}^{-1}\right\|^4.
$$

The order in probability of $\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|^2$ is determined by result (S.92). For the sum over i, we can use Lemma S.14 despite the presence of the additional term $\left\|\sum_{j\neq i}^{N}\boldsymbol{\lambda}_j\right\|^2$. To appreciate that this additional term does not change the results of Lemma S.14, note that the proof of Lemma (S.92) mainly relies on Markov's inequality. Given independence of errors and loadings as well as independence of factor loadings across cross-sections, the expectation of $\left\|\sum_{j\neq i}^{N}\boldsymbol{\lambda}_j\right\|^2$ can hence always be separated from the expectation of terms that are indexed by i. For this reason, we can conclude that

$$
\begin{aligned}
g_{62}&=\left[\mathcal{O}_P(1)+\mathcal{O}_P\left(NT^{-1}\right)\right]\left\{\mathcal{O}_P\left(N^{-2}\right)+\mathcal{O}_P\left[(NT)^{-1}\right]\right\}\mathcal{O}_P(N)\\
&=\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(NT^{-2}\right).
\end{aligned}\tag{S.105}
$$

Next, we focus on expression g_{64} which can be bounded in a very similar way. We write

$$
\begin{aligned}
g_{64}&\leq\sum_{i=1}^{N}\left\|\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\boldsymbol{\lambda}_i'\right\|^2\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2\left\|T^{-1}\sum_{j\neq i}^{N}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_j\right\|^2\\
&=\left[\mathcal{O}_P(1)+\mathcal{O}_P\left(NT^{-1}\right)\right]\left[\mathcal{O}_P(1)+\mathcal{O}_P\left(NT^{-1}\right)\right]\\
&=\mathcal{O}_P(1)+\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^2T^{-2}\right),
\end{aligned}\tag{S.106}
$$

where the result $\left\|T^{-1}\sum_{j\neq i}^{N}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_j\right\|^2=\mathcal{O}_P(1)+\mathcal{O}_P\left(NT^{-1}\right)$ follows from (S.92) since the former term is entirely contained in $N^2\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|^2$ and since the error variances σ_j^2 have lower bound above zero.

We continue with expression

$$g_{61} \leq 5T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\left(\frac{\varepsilon_i'\varepsilon_i}{T}-\sigma_i^2\right)\right]^2 + 5T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\left(T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right)\left(\overline{\boldsymbol{C}}^{-1}\right)\boldsymbol{\lambda}_i\right]^2$$

$$5T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\left(T^{-1}\varepsilon_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right)\right]^2 + 5T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\boldsymbol{\lambda}_i'\left(\overline{\boldsymbol{C}}^{-1}\right)'\left(T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\varepsilon_i\right)\right]^2$$

$$= g_{611} + g_{612} + g_{613} + g_{614}.$$

Here, $g_{611} = \mathcal{O}_P\left(N^2T^{-2}\right)$ follows from result (S.80) in the time fixed effects case. When considering g_{612}, we first note that

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right)^2\right] &\leq M\sum_{i=1}^{N}\sum_{t,t'}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\\
&\quad + M\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right) + \mathrm{o}\left(NT^2\right) + \mathrm{o}\left(N^2T\right)\\
&= \mathcal{O}\left(NT^2\right) + \mathcal{O}\left(N^2T\right),
\end{aligned}$$

from which it follows that

$$T^{-2}\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right)^2 = \mathcal{O}_P\left(N\right) + \mathcal{O}_P\left(N^2T^{-1}\right). \tag{S.107}$$

This result, together with result (S.92), is then used to arrive at

$$\begin{aligned}
g_{612} &\leq MT^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\left(\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right)\right]^2\left\|\overline{\boldsymbol{C}}^{-1}\right\|^4\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\overline{\boldsymbol{U}}\right\|^2\\
&= \left[\mathcal{O}_P\left(N\right) + \mathcal{O}_P\left(N^2T^{-1}\right)\right]\left\{\mathcal{O}_P\left(N^{-2}\right) + \mathcal{O}_P\left[\left(NT\right)^{-1}\right]\right\}\\
&= \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left(NT^{-2}\right)
\end{aligned}$$

We continue with

$$\begin{aligned}
g_{613} &\leq T^{-2}\left\|T^{-1}\left(\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2\left\|\overline{\boldsymbol{C}}\right\|^4\sum_{i=1}^{N}\left(\left\|T^{-1}\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\varepsilon_i'\boldsymbol{F}\right\|^2\left\|T^{-1}\boldsymbol{F}'\varepsilon_i\right\|^2\right)\\
&\quad + T^{-2}\left\|T^{-1}\left(\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2\sum_{i=1}^{N}\left(\left\|T^{-1}\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\varepsilon_i'\overline{\boldsymbol{U}}\right\|^2\left\|T^{-1}\widehat{\boldsymbol{F}}'\varepsilon_i\right\|^2\right)\\
&\quad + T^{-2}\left\|T^{-1}\left(\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2\left\|\overline{\boldsymbol{C}}\right\|^2\sum_{i=1}^{N}\left(\left[\left\|T^{-1}\sum_{j\neq i}^{N}\varepsilon_i'\varepsilon_j\varepsilon_i'\boldsymbol{F}\right\|^2\right]\left\|T^{-1}\overline{\boldsymbol{U}}'\varepsilon_i\right\|^2\right)\\
&g_{6131} + g_{6132} + g_{6133}
\end{aligned}$$

Here,

$$
\begin{aligned}
&\mathrm{E}\left(T^{-2}\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\boldsymbol{F}\right\|^2\left\|T^{-1}\boldsymbol{F}'\boldsymbol{\varepsilon}_i\right\|^2\right)\\
&=T^{-6}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t''',t'''',t'''''}^{T}\mathrm{E}\left[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{i,t'}\varepsilon_{j',t'}\varepsilon_{i,t''}\varepsilon_{i,t'''}\varepsilon_{i,t''''}\varepsilon_{i,t'''''}\right]\mathrm{E}\left[\boldsymbol{f}_{t''}'\boldsymbol{f}_{t'''}\boldsymbol{f}_{t''''}'\boldsymbol{f}_{t'''''}\right]\\
&=T^{-6}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t,t'',t''''}^{T}\mathrm{E}\left[\varepsilon_{i,t}^2\right]\mathrm{E}\left[\varepsilon_{j,t}^2\right]\mathrm{E}\left[\varepsilon_{i,t''}^2\right]\mathrm{E}\left[\varepsilon_{i,t''''}^2\right]\mathrm{E}\left[\left\|\boldsymbol{f}_{t''}\right\|^2\left\|\boldsymbol{f}_{t''''}\right\|^2\right]+\mathrm{o}\left(N^2T^{-3}\right)\\
&=\mathcal{O}\left(N^2T^{-3}\right),
\end{aligned}
$$

which implies that $g_{6131}=\mathcal{O}_P\left(N^2T^{-3}\right)$. Concerning the next term, g_{6132}, we note that

$$
\begin{aligned}
T^{-2}\sum_{i=1}^{N}\left(\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2\left\|T^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i\right\|^2\right)&\leq T^{-2}\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\right\|^2\sum_{i=1}^{N}\left\|T^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i\right\|^2\\
&=T^{-2}\left[\mathcal{O}_P\left(N\right)+\mathcal{O}_P\left(T\right)\right]\left[\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right)\right]\\
&=\mathcal{O}_P\left(N^2T^{-3}\right)+\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left[\left(NT\right)^{-1}\right],
\end{aligned}
$$

which follows from results (S.96) and (S.104). Accordingly, $g_{6132}=\mathcal{O}_P\left(N^2T^{-3}\right)+\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left[\left(NT\right)^{-1}\right]$. In the last component of g_{613},term g_{6133}, we can use results (S.94) and (S.103) to write

$$
\begin{aligned}
T^{-2}\sum_{i=1}^{N}\left(\left[\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\boldsymbol{F}\right\|^2\right]\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\right\|^2\right)&\leq T^{-2}\sum_{i=1}^{N}\left[\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_i'\boldsymbol{F}\right\|^2\right]\sum_{i=1}^{N}\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\right\|^2\\
&=T^{-2}\,\mathcal{O}_P\left(N^2\right)\left[\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)\right]\\
&=\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(N^2T^{-3}\right),
\end{aligned}
$$

which also determines the overall order in probability of g_{6133}. Combining our results on terms g_{6131}, g_{6132} and g_{6133}, we arrive at

$$
g_{613}=\mathcal{O}_P\left(N^2T^{-3}\right)+\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left[\left(NT\right)^{-1}\right].
$$

Lastly, we investigate

$$
\begin{aligned}
g_{614}&\leq 5\sum_{i=1}^{N}\left(\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\lambda}_i'\right\|\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\right\|\right)^2\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2\\
&\quad+5\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2\sum_{i=1}^{N}\left\|T^{-1}\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\lambda}_i'\right\|^2\left\|T^{-1}\overline{\boldsymbol{U}}'\widehat{\boldsymbol{F}}\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2\sum_{i=1}^{N}\left\|T^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i\right\|^2\\
&=g_{6141}+g_{6142}
\end{aligned}
$$

Here, the order of g_{6141} is determined by the expression

$$\sum_{i=1}^{N}\left[T^{-1}\left\|\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\lambda}_i'\right\|\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i\right\|\right]^2$$
$$=N^{-2}T^{-4}\sum_{i,i',i''}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i',t''}\varepsilon_{i'',t'''}\varepsilon_{i,t'''}\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i$$
$$+N^{-2}T^{-4}\sum_{i,i',i''}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\boldsymbol{e}_{i',t''}'\boldsymbol{e}_{i'',t'''}\varepsilon_{i,t'''}\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i.$$

Now take the first term in the decomposition above. Its expected value can be expressed as

$$\mathrm{E}\left(N^{-2}T^{-4}\sum_{i,i',i'',j,j'}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\varepsilon_{i',t''}\varepsilon_{i'',t'''}\varepsilon_{i,t'''}\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right)$$
$$\leq\frac{M}{N^2T^4}\sum_{i,i',j}^{N}\sum_{t,t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i',t''}^2\right)$$
$$+\frac{M}{N^2T^4}\sum_{i,i'}^{N}\sum_{t,t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t'}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\varepsilon_{i',t''}^2\right)+\mathrm{o}\left(NT^{-2}\right)+\mathrm{o}\left(T^{-1}\right)$$
$$=\mathcal{O}\left(NT^{-2}\right)+\mathcal{O}\left(T^{-1}\right).$$

Analogously, the expectation of the second term is given by

$$\mathrm{E}\left(N^{-2}T^{-4}\sum_{i,i',i''}^{N}\sum_{j\neq i}^{N}\sum_{j'\neq i}^{N}\sum_{t,t',t'',t'''}^{T}\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{j',t'}\varepsilon_{i,t'}\varepsilon_{i,t''}\boldsymbol{e}_{i',t''}'\boldsymbol{e}_{i'',t'''}\varepsilon_{i,t'''}\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right)$$
$$\leq\frac{M}{N^2T^4}\sum_{i,i''}^{N}\sum_{j\neq i}^{N}\sum_{t',t''}^{T}\mathrm{E}\left(\varepsilon_{i,t}^2\right)\mathrm{E}\left(\varepsilon_{j,t}^2\right)\mathrm{E}\left(\varepsilon_{i,t''}^2\right)\mathrm{E}\left(\boldsymbol{e}_{i',t''}'\boldsymbol{e}_{i',t''}\right)+\mathrm{o}\left(NT^{-2}\right)$$
$$=\mathcal{O}\left(NT^{-2}\right).$$

These results suggest that $g_{6141}=\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(T^{-1}\right)$. We continue with g_{6142} where results (S.96), (S.102) and (S.107) lead to

$$g_{6142}=5\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2\sum_{i=1}^{N}\left\|\sum_{j\neq i}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\lambda}_i'\right\|^2\left\|T^{-1}\overline{\boldsymbol{U}}'\widehat{\boldsymbol{F}}\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2\sum_{i=1}^{N}\left\|T^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i\right\|^2$$
$$=\left[\mathcal{O}_P\left(N\right)+\mathcal{O}_P\left(N^2T^{-1}\right)\right]\left\{\mathcal{O}_P\left(N^{-2}\right)+\mathcal{O}_P\left[(NT)^{-1}\right]\right\}\left[\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right)\right]$$
$$=\left[\mathcal{O}_P\left(N^{-1}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(NT^{-2}\right)\right]\left[\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(N^{-1}\right)\right]$$
$$=\mathcal{O}_P\left(N^2T^{-3}\right)+\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(N^{-2}\right).$$

Combining our results on g_{6141} and g_{6142}, we get

$$g_{614}=\mathcal{O}_P\left(N^2T^{-3}\right)+\mathcal{O}_P\left(NT^{-2}\right)+\mathcal{O}_P\left(T^{-1}\right)+\mathcal{O}_P\left(N^{-2}\right)$$

Our result on g_{614} is the last building block required to make a statement about g_{61}. Additionally using our results on g_{611}, g_{612} and g_{613}, we arrive at

$$g_{61} = \mathcal{O}_P\left(N^2T^{-2}\right) + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left(T^{-1}\right). \tag{S.108}$$

Now consider the term

$$\begin{aligned}
g_{65} &\leq \left\|\overline{\boldsymbol{C}}^{-1}\right\|^2 \left\|\sum_{j\neq i}^{N} \boldsymbol{\lambda}_j\right\|^2 \sum_{i=1}^{N} \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \boldsymbol{\varepsilon}_i' \boldsymbol{M}_{\widehat{\boldsymbol{F}}} \overline{\boldsymbol{U}}\right\|^2 \\
&\leq 3\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2 \left\|\sum_{j\neq i}^{N} \boldsymbol{\lambda}_j\right\|^2 \sum_{i=1}^{N} \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \boldsymbol{\varepsilon}_i' \overline{\boldsymbol{U}}\right\|^2 \\
&+ 3\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2 \left\|\sum_{j\neq i}^{N} \boldsymbol{\lambda}_j\right\|^2 \sum_{i=1}^{N} \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \boldsymbol{\varepsilon}_i' \widehat{\boldsymbol{F}}\right\|^2 \left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2 + \mathcal{O}_P\left(NT^{-1}\right)\left\|T^{-1}\widehat{\boldsymbol{F}}\overline{\boldsymbol{U}}\right\|^2 \\
&= 3\left\|\overline{\boldsymbol{C}}^{-1}\right\|^2 \left\|\sum_{j\neq i}^{N} \boldsymbol{\lambda}_j\right\|^2 \left(g_{651} + g_{652}\right).
\end{aligned}$$

The two resulting expressions g_{651} and g_{652} require a further investigation of the terms $\sum_{i=1}^{N} \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \boldsymbol{\varepsilon}_i' \overline{\boldsymbol{U}}\right.$ and $\sum_{i=1}^{N} \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \boldsymbol{\varepsilon}_i' \widehat{\boldsymbol{F}}\right\|^2$, a task that is carried out in Lemmas S.15 and S.16 as well as their respective proofs. Using the results of these two lemmas, as well as result (S.102), we arrive at

$$\begin{aligned}
g_{651} &= \mathcal{O}_P\left(NT^{-3}\right) + \mathcal{O}_P\left(T^{-2}\right) + \mathcal{O}_P\left[(NT)^{-1}\right] + \mathcal{O}_P\left(N^{-2}\right) \\
g_{652} &= \left[\mathcal{O}_P\left(NT^{-2}\right) + \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left(N^{-2}\right)\right]\left[\mathcal{O}_P\left[(NT)^{-1}\right] + \mathcal{O}_P\left(N^{-2}\right)\right] \\
&= \mathcal{O}_P\left(T^{-3}\right) + \mathcal{O}_P\left(N^{-1}T^{-2}\right) + \mathcal{O}_P\left(N^{-2}T^{-1}\right) + \mathcal{O}_P\left(N^{-4}\right),
\end{aligned}$$

from which it follows that

$$g_{65} = \mathcal{O}_P\left(NT^{-3}\right) + \mathcal{O}_P\left(T^{-2}\right) + \mathcal{O}_P\left[(NT)^{-1}\right] + \mathcal{O}_P\left(N^{-2}\right). \tag{S.109}$$

The last term to consider is g_{53} where we write

$$\begin{aligned}
g_{63} &\leq \sum_{i=1}^{N} \left\|T^{-1}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) \boldsymbol{\varepsilon}_i' \widehat{\boldsymbol{F}}\right\|^2 \left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|^2 \left\|T^{-1}\widehat{\boldsymbol{F}}'\left(\sum_{j\neq i}^{N} \boldsymbol{\varepsilon}_j\right)\right\|^2 \\
&= \left[\mathcal{O}_P\left(NT^{-2}\right) + \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left(N^{-2}\right)\right]\left[\mathcal{O}_P\left(NT^{-1}\right) + \mathcal{O}_P(1)\right] \\
&= \mathcal{O}_P\left(N^2T^{-3}\right) + \mathcal{O}_P\left(NT^{-2}\right) + \mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left(N^{-2}\right)
\end{aligned} \tag{S.110}$$

In order to arrive at the second line above, we use Lemma S.16 together with result (S.102). Application of the second result is valid since σ_j^2 has a lower bound above zero and since $\left\|T^{-1}\widehat{\boldsymbol{F}}' \sum_{j\neq i}^{N} \boldsymbol{\varepsilon}_j\right\|^2$ is completely contained in $N^2\left\|T^{-1}\widehat{\boldsymbol{F}}\overline{\boldsymbol{U}}\right\|^2$.

Results (S.105)–(S.110) address all five required terms to arrive at the central result of this lemma:

$$T^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\frac{\widehat{\varepsilon}_i'\widehat{\varepsilon}_j}{\left(\sigma_j^2\right)^{1/2}}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\right]^2=\mathcal{O}_P\left(N^2T^{-2}\right)+\mathcal{O}_P\left(NT^{-1}\right)+\mathcal{O}_P\left(1\right)$$

□

S.6 Proofs for additional theoretical contributions

S.6.1 Proofs of Propositions S.1 and S.2

Proof of Proposition S.1.
The same steps as in the proof of Theorem 1 allow us to arrive at the decomposition

$$\begin{aligned}CD_{\mathbb{H}_1}=&\sqrt{\frac{1}{2TN\left(N-1\right)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\left(\varsigma_{\nu,i}^{-1}-\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right]\right)\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\left(\varsigma_{\nu,j}^{-1}-\mathrm{E}\left[\varsigma_{\nu,j}^{-1}\right]\right)+\sqrt{T}\Xi_{\mathbb{H}_1}\\&+\frac{\sqrt{T}}{N}\sqrt{\frac{N}{2\left(N-1\right)}}\left(\left(NT\right)^{-1}\sum_{i,j}^{N}\frac{\boldsymbol{\nu}_i'\boldsymbol{\nu}_j}{\varsigma_{\nu,i}^2}-\overline{\varsigma_\nu^{-2}}\left(NT\right)^{-1}\sum_{i,j}^{N}\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\right),\end{aligned}$$

where

$$\Xi_{\mathbb{H}_1}=\sqrt{\frac{N}{2\left(N-1\right)}}\left(NT\right)^{-1}\sum_{i=1}^{N}\boldsymbol{\nu}_i'\boldsymbol{\nu}_i\left[\left(\overline{\varsigma_\nu^{-1}}\right)^2-2\overline{\varsigma_\nu^{-1}}\varsigma_{\nu,i}\right].$$

The terms in parentheses in the second line are stochastically bounded, entailing that the entire second line is $\mathcal{O}_P\left(N^{-1}\sqrt{T}\right)$. Next, consider the leading stochastic component of $CD_{\mathbb{H}_1}$, namely $\sqrt{\frac{1}{2TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\left(\varsigma_{\nu,i}^{-1}-\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right]\right)\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\left(\varsigma_{\nu,j}^{-1}-\mathrm{E}\left[\varsigma_{\nu,j}^{-1}\right]\right)$. Here, we have

$$\begin{aligned}&\mathrm{E}\left[\left(\varsigma_{\nu,i}^{-1}-\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right]\right)\nu_{i,t}\nu_{j,t}\left(\varsigma_{\nu,j}^{-1}-\mathrm{E}\left[\varsigma_{\nu,j}^{-1}\right]\right)\right]\\&=\mathrm{E}\left[\left(\varsigma_{\nu,i}^{-1}-\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\right]\right)\left(\boldsymbol{\lambda}_i-\mathrm{E}\left[\boldsymbol{\lambda}_i\right]\right)'\right]\boldsymbol{\Sigma}_F\,\mathrm{E}\left[\left(\boldsymbol{\lambda}_j-\mathrm{E}\left[\boldsymbol{\lambda}_j\right]\right)\left(\varsigma_{\nu,j}^{-1}-\mathrm{E}\left[\varsigma_{\nu,j}^{-1}\right]\right)\right]+0\\&=\mathrm{cov}\left[\varsigma_{\nu,i}^{-1},\boldsymbol{\lambda}_i'\right]\boldsymbol{\Sigma}_F\,\mathrm{cov}\left[\boldsymbol{\lambda}_j,\varsigma_{\nu,j}^{-1}\right]\\&=\mathrm{cov}\left[\varsigma_{\nu,1}^{-1},\boldsymbol{\lambda}_1'\right]\boldsymbol{\Sigma}_F\,\mathrm{cov}\left[\boldsymbol{\lambda}_1,\varsigma_{\nu,1}^{-1}\right],\end{aligned}$$

where we use *i.i.d.*-ness of $\boldsymbol{\lambda}_i$ and σ_i^2 to conclude that the covariance $\mathrm{cov}\left[\varsigma_{\nu,i}^{-1},\boldsymbol{\lambda}_i'\right]$ is the same across cross-sections.

Lastly, consider the variance of $\sqrt{\frac{1}{2TN(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\left(\varsigma_{\nu,i}^{-1}-\mathrm{E}\left[\varsigma_v^{-1}\right]\right)\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\left(\varsigma_{\nu,j}^{-1}-\mathrm{E}\left[\varsigma_v^{-1}\right]\right)$

in the case where $\text{cov}\left[\varsigma_{v,i}^{-1}, \boldsymbol{\lambda}_i'\right] = 0$. We have

$$\begin{aligned}
&\text{var}\left[\sum_{i=2}^{N}\sum_{j=1}^{i-1}\left(\varsigma_{v,i}^{-1} - \text{E}\left[\varsigma_{v,i}^{-1}\right]\right)\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\left(\varsigma_{v,j}^{-1} - \text{E}\left[\varsigma_{v,j}^{-1}\right]\right)\right] \\
&= \text{E}\left\{\left[\sum_{i=2}^{N}\sum_{j=1}^{i-1}\left(\varsigma_{v,i}^{-1} - \text{E}\left[\varsigma_{v,i}^{-1}\right]\right)\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\left(\varsigma_{v,j}^{-1} - \text{E}\left[\varsigma_{v,j}^{-1}\right]\right)\right]^2\right\} - 0 \\
&= \sum_{i=2}^{N}\sum_{j=1}^{i-1}\text{E}\left\{\left[\left(\varsigma_{v,i}^{-1} - \text{E}\left[\varsigma_{v,i}^{-1}\right]\right)\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\left(\varsigma_{v,j}^{-1} - \text{E}\left[\varsigma_{v,j}^{-1}\right]\right)\right]^2\right\}.
\end{aligned}$$

Here, $\left(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\right)^2 = \left[\left(\boldsymbol{\lambda}_i'\boldsymbol{F}' + \boldsymbol{\varepsilon}_i'\right)\left(\boldsymbol{F}\boldsymbol{\lambda}_j + \boldsymbol{\varepsilon}_j\right)\right]^2$, where we only consider the leading term $\text{tr}\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j'\boldsymbol{F}'\boldsymbol{F}\right)$ further. Here, it is possible to show that

$$\text{E}\left[tr\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j'\boldsymbol{F}'\boldsymbol{F}\right) \mid \boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j, \sigma_i^2, \sigma_j^2\right] \leq T^2 M \ \text{tr}\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j'\right),$$

since the fourth-order moments of $\boldsymbol{f}_t$ are bounded. By application of the Law of Iterated expectations (LIE), it then follows that

$$\begin{aligned}
&\sum_{i=2}^{N}\sum_{j=1}^{i-1}\text{E}\left\{\left(\varsigma_{v,i}^{-1} - \text{E}\left[\varsigma_{v,i}^{-1}\right]\right)^2 tr\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j'\boldsymbol{F}'\boldsymbol{F}\right)\left(\varsigma_{v,j}^{-1} - \text{E}\left[\varsigma_{v,j}^{-1}\right]\right)^2\right\} \\
&= M\,T^2\sum_{i=2}^{N}\sum_{j=1}^{i-1}\text{E}\left\{\left(\varsigma_{v,i}^{-1} - \text{E}\left[\varsigma_{v,i}^{-1}\right]\right)^2 tr\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j'\right)\left(\varsigma_{v,j}^{-1} - \text{E}\left[\varsigma_{v,j}^{-1}\right]\right)^2\right\} \\
&= \mathcal{O}\left(N^2T^2\right),
\end{aligned}$$

assuming that the higher-order moment $\text{E}\left\{\left(\varsigma_{v,i}^{-1} - \text{E}\left[\varsigma_{v,i}^{-1}\right]\right)^2 tr\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\right)\right\}$ exists. Involving the square of the scaling $\sqrt{\frac{1}{2TN(N-1)}}$, we can conclude that the variance of

$$\sqrt{\frac{1}{2TN\left(N-1\right)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\left(\varsigma_{v,i}^{-1} - \text{E}\left[\varsigma_{v}^{-1}\right]\right)\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\left(\varsigma_{v,j}^{-1} - \text{E}\left[\varsigma_{v}^{-1}\right]\right) \tag{S.111}$$

diverges at rate T. □

Proof of Proposition S.2.
Let $k_{N,T} = \sqrt{\frac{1}{2TN(N-1)}}$ and $\boldsymbol{\varpi}_i = \left[\varsigma_{v,i}^{-1}\boldsymbol{\nu}_i - \boldsymbol{D}_i\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{v,\ell}^{-1}\right)\right]$, where $\boldsymbol{\nu}_i$ and $\varsigma_{v,i}^2$ are defined in the discussion following equation (S.6) and where

$$\boldsymbol{D}_i = \boldsymbol{F}\left[\boldsymbol{\lambda}_i^{(2)}, \quad \boldsymbol{\Lambda}_i^{(2)}\right] + \left[\boldsymbol{\varepsilon}_i, \quad \boldsymbol{e}_i\right].$$

Using the same steps as in the Proof of Theorem 2, we can decompose the CD test statistic as

$$CD_{\mathbb{H}_1} = I - II$$

where

$$\begin{aligned} I &= 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j + k_{N,T}\sum_{i=1}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_i - k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varpi}_j \\ II &= k_{N,T}\sum_{i=1}^{N}\varsigma_{v,i}^{-2}\left(\boldsymbol{\nu}_i' - \boldsymbol{\lambda}_i^{(1)\prime}\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\overline{\boldsymbol{D}}'\right)\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\nu}_i - \overline{\boldsymbol{D}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right). \end{aligned}$$

As shown in Lemmas S.19 and S.20, it holds that

$$k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varpi}_j = \mathcal{O}_P\left[\max\left(N^{-1}, T^{-1}\right)\left(k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j\right)\right]$$

and

$$II = k_{N,T}\sum_{i=1}^{N}\varsigma_v^{-2}\boldsymbol{\nu}_i'\boldsymbol{\nu}_i + \mathcal{O}_P\left(N^{-1/2}\sqrt{T}\right) + \mathcal{O}_P\left(T^{-1/2}\right).$$

Furthermore, the expression $k_{N,T}\sum_{i=1}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_i$ is written out

$$\begin{aligned} k_{N,T}\sum_{i=1}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_i, \quad = \quad & k_{N,T}\sum_{i=1}^{N}\varsigma_{v,i}^{-2}\boldsymbol{\nu}_i'\boldsymbol{\nu}_i \\ & +\sqrt{\frac{TN}{(N-1)}}\left(N^{-1}\sum_{\ell=1}^{N}\varsigma_{v,\ell}^{-1}\boldsymbol{\lambda}_\ell^{(1)\prime}\right)\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{D}_i'\boldsymbol{D}_i\right) \\ & \times\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{v,\ell}^{-1}\right) \\ & -2\sqrt{\frac{TN}{(N-1)}}(NT)^{-1}\sum_{i=1}^{N}\varsigma_{v,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{D}_j\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{v,\ell}^{-1}\right). \end{aligned}$$

The first term on the right-hand side above is cancelled out by the leading term in II. The two remaining terms constitute expressions $\Phi_{1,\mathbb{H}_1}$ and $\Phi_{2,\mathbb{H}_1}$, respectively. Next, consider the properties of $2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j$, which we write as

$$\begin{aligned} & 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j \\ =& 2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{v,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\varsigma_{v,j}^{-1} - 4k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{v,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{D}_j\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{v,\ell}^{-1}\right) \\ & + \left(N^{-1}\sum_{\ell=1}^{N}\varsigma_{v,\ell}^{-1}\boldsymbol{\lambda}_\ell^{(1)\prime}\right)\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\left(2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{D}_i'\boldsymbol{D}_j\right)\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{v,\ell}^{-1}\right). \end{aligned} \tag{S.112}$$

The order in probability of all three terms above is determined by the expected values of the elements within the double sum $\sum_{i=2}^{N}\sum_{j=1}^{i-1}$. Conditioning on the values taken by factor loadings and error variances, we have

$$\begin{aligned}
\mathrm{E}\left[\nu_{i,t}\nu_{j,t} \mid \boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j, \sigma_i^2, \sigma_j^2\right] &= \boldsymbol{\lambda}_i^{(2)\prime}\boldsymbol{\Sigma}_{F,22}\boldsymbol{\lambda}_j^{(2)}, \\
\mathrm{E}\left[\boldsymbol{D}_{i,t}\boldsymbol{D}_{j,t}^{\prime} \mid \boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j, \boldsymbol{\Lambda}_i, \boldsymbol{\Lambda}_j, \sigma_i^2, \sigma_j^2\right] &= \begin{bmatrix} \boldsymbol{\lambda}_i^{(2)\prime} \\ \boldsymbol{\Lambda}_i^{(2)\prime} \end{bmatrix} \boldsymbol{\Sigma}_{F,22}\left[\boldsymbol{\lambda}_j^{(2)},\, \boldsymbol{\Lambda}_j^{(2)}\right], \\
\mathrm{E}\left[\nu_{i,t}\boldsymbol{D}_{jt}^{\prime} \mid \boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j, \boldsymbol{\Lambda}_j, \sigma_i^2, \sigma_j^2\right] &= \boldsymbol{\lambda}_i^{(2)\prime}\boldsymbol{\Sigma}_{F,22}\left[\boldsymbol{\lambda}_j^{(2)},\, \boldsymbol{\Lambda}_j^{(2)}\right],
\end{aligned}$$

for $i \neq j$. The zero mean property of $\boldsymbol{\lambda}_i^{(2)}$ and $\boldsymbol{\Lambda}_i^{(2)}$ implies that $\mathrm{E}\left[\boldsymbol{D}_{i,t}\boldsymbol{D}_{j,t}^{\prime}\right] = \mathbf{O}$. This result carries over to $\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i^{\prime}\boldsymbol{D}_j\right]$ since independence of $\boldsymbol{\lambda}_i$ and σ_i^2 over cross-sections implies that we can write $\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i^{\prime}\boldsymbol{D}_j\right] = \mathrm{E}\left[\varsigma_{\nu,i}^{-1}\boldsymbol{\lambda}_i^{(2)\prime}\right]\boldsymbol{\Sigma}_{F,22}\,\mathrm{E}\left\{\left[\boldsymbol{\lambda}_j^{(2)},\, \boldsymbol{\Lambda}_j^{(2)}\right]\right\} = \mathbf{0}$. Consequently, the last two terms on the right-hand side of (S.112) are asymptotically centered around zero since their components $\overline{\boldsymbol{C}}^{(1)\prime}$ and $N^{-1}\sum_{\ell=1}^{N}\varsigma_{\nu,\ell}^{-1}\boldsymbol{\lambda}_{\ell}^{(1)\prime}$ converge to nonstochastic limiting expressions. Lastly, by the LIE, we have

$$\mathrm{E}\left[\varsigma_{\nu,i}^{-1}\nu_{i,t}\nu_{j,t}\varsigma_{\nu,j}^{-1}\right] = \mathrm{cov}\left[\varsigma_{\nu,i}^{-1},\, \boldsymbol{\lambda}_i^{(2)\prime}\right]\boldsymbol{\Sigma}_{F,22}\,\mathrm{cov}\left[\boldsymbol{\lambda}_j^{(2)},\, \varsigma_{\nu,j}^{-1}\right],$$

and consequently

$$\mathrm{E}\left[2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i^{\prime}\boldsymbol{\nu}_j\varsigma_{\nu,j}^{-1}\right] = N\sqrt{T}\,\mathrm{cov}\left[\varsigma_{\nu,1}^{-1},\, \boldsymbol{\lambda}_1^{(2)\prime}\right]\boldsymbol{\Sigma}_{F,22}\,\mathrm{cov}\left[\boldsymbol{\lambda}_1^{(2)},\, \varsigma_{\nu,1}^{-1}\right],$$

which parallels results obtained in the proof of Proposition S.1. It will be shown below that variation around this expected value is of lower order in probability. Assuming this for now, we state that

$$\operatorname*{plim}_{N,T\to\infty} N^{-1}T^{-1/2}\sqrt{\frac{2}{TN\left(N-1\right)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i^{\prime}\boldsymbol{\varpi}_j = \mathrm{cov}\left[\varsigma_{\nu,i}^{-1},\, \boldsymbol{\lambda}_i^{(2)\prime}\right]\boldsymbol{\Sigma}_{F,22}\,\mathrm{cov}\left[\boldsymbol{\lambda}_j^{(2)},\, \varsigma_{\nu,j}^{-1}\right].$$

A consequence of this result, as well as those on the second and third terms in Eq. (S.112) is that $2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i^{\prime}\boldsymbol{\varpi}_j$ is asymptotically centered around 0 if $\varsigma_{\nu,i}^{-1}$ and $\boldsymbol{\lambda}_i^{(2)\prime}$ are uncorrelated. In this special case, it is required to investigate the rate at which the variance of $2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{\varpi}_i^{\prime}\boldsymbol{\varpi}_j$ diverges. First, consider

$$\mathrm{var}\left[2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i^{\prime}\boldsymbol{\nu}_j\varsigma_{\nu,j}^{-1}\right] = \mathrm{E}\left[4k_{N,T}^2\left(\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i^{\prime}\boldsymbol{\nu}_j\varsigma_{\nu,j}^{-1}\right)^2\right].$$

Apart from a superscript $^{(2)}$ on factors and factor loadings as well as the absence of the expected value of $\varsigma_{\nu,i}^{-1}$, the term $\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i^{\prime}\boldsymbol{\nu}_j\varsigma_{\nu,j}^{-1}$ is identical to the one occurred in the proof of Proposition S.1. Hence, we can follow the same reasoning to conclude that

$$\mathrm{var}\left[\sqrt{\frac{2}{TN\left(N-1\right)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i^{\prime}\boldsymbol{\nu}_j\varsigma_{\nu,j}^{-1}\right] = \mathcal{O}\left(T\right).$$

Next, consider the middle term of the third expression in Eq. (S.112). After vectorizing this matrix, we can write

$$\begin{aligned}\operatorname{var}\left[\operatorname{vec}\left(\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{D}_i'\boldsymbol{D}_j\right)\right] &= \mathrm{E}\left[\operatorname{vec}\left(\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{D}_i'\boldsymbol{D}_j\right)\operatorname{vec}\left(\sum_{i'=2}^{N}\sum_{j'=1}^{i'-1}\boldsymbol{D}_{i'}'\boldsymbol{D}_{j'}\right)'\right]\\ &= \sum_{i=2}^{N}\sum_{j=1}^{i-1}\mathrm{E}\left[\operatorname{vec}\left(\boldsymbol{D}_i'\boldsymbol{D}_j\right)\operatorname{vec}\left(\boldsymbol{D}_i'\boldsymbol{D}_j\right)'\right].\end{aligned}$$

Recall here that

$$\begin{aligned}\boldsymbol{D}_i'\boldsymbol{D}_j &= \begin{bmatrix}\boldsymbol{\lambda}_i^{(2)}\\ \boldsymbol{\Lambda}_i^{(2)\prime}\end{bmatrix}\boldsymbol{F}^{(2)'}\boldsymbol{F}^{(2)}\left[\boldsymbol{\lambda}_j^{(2)},\boldsymbol{\Lambda}_j^{(2)}\right]+\begin{bmatrix}\varepsilon_i'\\ \boldsymbol{E}_i'\end{bmatrix}\left[\varepsilon_j,\ \boldsymbol{E}_j\right]\\ &\quad+\begin{bmatrix}\boldsymbol{\lambda}_i^{(2)}\\ \boldsymbol{\Lambda}_i^{(2)\prime}\end{bmatrix}\boldsymbol{F}^{(2)'}\left[\varepsilon_j,\ \boldsymbol{E}_j\right]+\begin{bmatrix}\varepsilon_i'\\ \boldsymbol{E}_i'\end{bmatrix}\boldsymbol{F}^{(2)}\left[\boldsymbol{\lambda}_j^{(2)},\ \boldsymbol{\Lambda}_j^{(2)}\right],\end{aligned}$$

whose leading term is the first on the right-hand side of the first line and where we recall that $\boldsymbol{C}_i^{(2)}=\left[\boldsymbol{\lambda}_j^{(2)},\ \boldsymbol{\Lambda}_j^{(2)}\right]$. Hence, the leading term of $\operatorname{vec}\left(\boldsymbol{D}_i'\boldsymbol{D}_j\right)\operatorname{vec}\left(\boldsymbol{D}_i'\boldsymbol{D}_j\right)'$ is written

$$\operatorname{vec}\left(\boldsymbol{D}_i'\boldsymbol{D}_j\right)\operatorname{vec}\left(\boldsymbol{D}_i'\boldsymbol{D}_j\right)'=\left(\boldsymbol{C}_i^{(2)}\otimes\boldsymbol{C}_j^{(2)}\right)'\operatorname{vec}\left(\boldsymbol{F}^{(2)'}\boldsymbol{F}^{(2)}\right)\operatorname{vec}\left(\boldsymbol{F}^{(2)'}\boldsymbol{F}^{(2)}\right)\left(\boldsymbol{C}_i^{(2)}\otimes\boldsymbol{C}_j^{(2)}\right),$$

which is nothing more than a multivariate version of the leading term in $\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\varsigma_{\nu,j}^{-1}$. The current case is even simpler due to the absence of the inverse error standard deviation $\varsigma_{\nu,i}^{-1}$. Consequently, finite fourth moments for factors and finite second moments for loadings together with the reasoning applied for the variance of $2k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{\nu}_j\varsigma_{\nu,j}^{-1}$ allow us to conclude that

$$\operatorname{var}\left[\sqrt{\frac{2}{TN\left(N-1\right)}}\operatorname{vec}\left(\sum_{i=2}^{N}\sum_{j=1}^{i-1}\boldsymbol{D}_i'\boldsymbol{D}_j\right)\right]=\mathcal{O}_P\left(T\right).$$

Since the outer terms $\left(N^{-1}\sum_{\ell=1}^{N}\varsigma_{\ell,v}^{-1}\boldsymbol{\lambda}_\ell^{(1)\prime}\right)$ and $\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}$ converge to finite, non-stochastic components, this rate of divergence carries over to the variance of the entire term $\left(N^{-1}\sum_{\ell=1}^{N}\varsigma_{\nu,\ell}^{-1}\boldsymbol{\lambda}_\ell^{(1)\prime}\right)\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)$

Lastly, consider the second term in (S.112), whose middle term is of particular interest. The variance of this term is given by

$$\begin{aligned}\operatorname{var}\left[\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{D}_j\right] &= \mathrm{E}\left[\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{i'=2}^{N}\sum_{j'=1}^{i'-1}\boldsymbol{D}_j'\boldsymbol{\nu}_i\boldsymbol{\nu}_{i'}'\boldsymbol{D}_{j'}\varsigma_{\nu,i}^{-1}\varsigma_{i',v}^{-1}\right]\\ &= \sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t,t'}^{T}\mathrm{E}\left[\boldsymbol{D}_{j,t}\nu_{i,t}\nu_{i,t'}\boldsymbol{D}_{j,t'}'\varsigma_{\nu,i}^{-2}\right].\end{aligned}$$

Similar to the previous two terms, the leading expression in the term above is given by

$$\mathrm{E}\left[\boldsymbol{C}_j^{(2)\prime}\boldsymbol{F}_t^{(2)}\boldsymbol{F}_t^{(2)\prime}\boldsymbol{\lambda}_i^{(2)}\boldsymbol{\lambda}i^{(2)\prime}\boldsymbol{F}_t^{(2)}\boldsymbol{F}_t^{(2)\prime}\boldsymbol{C}_j^{(2)}\varsigma_{\nu,i}^{-2}\right]\leq M\ \mathrm{E}\left[\boldsymbol{C}_j^{(2)\prime}\boldsymbol{\lambda}_i^{(2)}\varsigma_{\nu,i}^{-2}\boldsymbol{\lambda}_i^{(2)\prime}\boldsymbol{C}_j^{(2)}\right]$$

where the upper bound on the right-hand side above is established by the LIE and boundedness of the fourth-order moments of $\boldsymbol{F}_t$. Additionally assuming that $\mathrm{E}\left[\boldsymbol{\lambda}_i^{(2)}\varsigma_{\nu,i}^{-2}\boldsymbol{\lambda}_i^{(2)\prime}\right]$ is bounded and conditioning on $\boldsymbol{C}_j^{(2)}$, we arrive at

$$\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{t,t'}^{T}\mathrm{E}\left[\boldsymbol{D}_{j,t}\nu_{i,t}\nu_{i,t'}\boldsymbol{D}_{j,t'}'\varsigma_{\nu,i}^{-2}\right]=\mathcal{O}\left[(NT)^2\right],$$

implying that

$$\mathrm{var}\left[\sqrt{\frac{2}{TN\,(N-1)}}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{D}_j\right]=\mathcal{O}\left(T\right),$$

which carries over to the variance of $k_{N,T}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i'\boldsymbol{D}_j\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)$. $\square$

S.6.2 Auxiliary lemmas for Section S.6.1

Lemma S.18. *Suppose that the true model is* (16) *and that Assumptions 2–5 and S.1 hold. Also, assume that the factor estimator* $\widehat{\boldsymbol{F}}$ *is given by* $\widehat{\boldsymbol{F}} = [\overline{\boldsymbol{y}}, \overline{\boldsymbol{X}}]$. *Then we can without loss of generality assume that the true factors and loadings can be partitioned into*

$$\boldsymbol{f}_t = \begin{bmatrix} \boldsymbol{f}_t^{(1)} \\ \boldsymbol{f}_t^{(2)} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\lambda}_i = \begin{bmatrix} \boldsymbol{\lambda}_i^{(1)} \\ \boldsymbol{\lambda}_i^{(2)} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Lambda}_i = \begin{bmatrix} \boldsymbol{\Lambda}_i^{(1)} \\ \boldsymbol{\Lambda}_i^{(2)} \end{bmatrix}$$

such that $\mathrm{E}\left[\boldsymbol{f}_t^{(1)}\boldsymbol{f}_t^{(2)\prime}\right] = \mathbf{O}$, $\mathrm{E}\left[\boldsymbol{\lambda}_i^{(2)}\right] = \mathbf{0}$ *and* $\mathrm{E}\left[\boldsymbol{\Lambda}_i^{(2)}\right] = \mathbf{O}$. *It holds that* $\boldsymbol{f}^{(1)}$ *and* $\boldsymbol{f}^{(2)}$ *are* $(m+1)\times 1$ *and* $(r-m-1)\times 1$, *respectively.* $\boldsymbol{\lambda}_i^{(1)}$ *and* $\boldsymbol{\lambda}_i^{(2)}$ *are* $(m+1)\times 1$ *and* $(r-m-1)\times 1$. *Lastly,* $\boldsymbol{\Lambda}_i^{(1)}$ *and* $\boldsymbol{\Lambda}_i^{(2)}$ *are* $(m+1)\times m$ *and* $(r-m-1)\times m$.

Proof of Lemma S.18.
Under a violation of the rank condition, cross-section averages of all variables will only identify the $m+1$-dimensional linear combination $\boldsymbol{f}_t^{(1)} = \boldsymbol{C}_0'\boldsymbol{f}_t$ of the r true common factors. The $r\times(m+1)$ matrix $\boldsymbol{C}_0$ is given by

$$\boldsymbol{C}_0 = \left[\mathrm{E}\left[\boldsymbol{\lambda}_i\right], \mathrm{E}\left[\boldsymbol{\Lambda}_i\right]\right]\boldsymbol{B}.$$

Now define the $r\times(m+1)$ matrix $\boldsymbol{\Phi}_1 = \mathrm{E}\left[\boldsymbol{f}_t\boldsymbol{f}_t^{(1)\prime}\right]\left(\mathrm{E}\left[\boldsymbol{f}_t^{(1)}\boldsymbol{f}_t^{(1)\prime}\right]\right)^{-1}$ and the $r\times 1$ vector $\boldsymbol{f}_t^{\perp} = \boldsymbol{f}_t - \boldsymbol{\Phi}_1\boldsymbol{f}_t^{(1)}$ and note that by construction it holds that $\mathrm{E}\left[\boldsymbol{f}_t^{\perp}\boldsymbol{f}_t^{(1)\prime}\right] = \mathbf{O}$. Using these two expressions, we can decompose

$$\boldsymbol{\lambda}_i'\boldsymbol{f}_t = \boldsymbol{\lambda}_i'\boldsymbol{\Phi}_1\boldsymbol{f}_t^{(1)} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t^{\perp}. \tag{S.113}$$

The expression above decomposes the common component of a random variable that has a factor structure into the variation due to the $(m+1)$-dimensional linear combination of the r true factors that is identified by a cross-section average-based factor estimator and residual common variation that is not picked up. This residual source of common variation is bound to be part of the error term in a misspecified factor model.

Next, consider the product of $\boldsymbol{C}_0'$ and $\boldsymbol{f}_t^{\perp}$

$$\begin{aligned} \boldsymbol{C}_0'\boldsymbol{f}_t^{\perp} &= \boldsymbol{C}_0'\boldsymbol{f}_t - \boldsymbol{C}_0'\mathrm{E}\left[\boldsymbol{f}_t\boldsymbol{f}_t^{(1)\prime}\right]\left(\mathrm{E}\left[\boldsymbol{f}_t^{(1)}\boldsymbol{f}_t^{(1)\prime}\right]\right)^{-1}\boldsymbol{f}_t^{(1)} \\ &= \boldsymbol{f}_t^{(1)} - \mathrm{E}\left[\boldsymbol{C}_0'\boldsymbol{f}_t\boldsymbol{f}_t^{(1)\prime}\right]\left(\mathrm{E}\left[\boldsymbol{f}_t^{(1)}\boldsymbol{f}_t^{(1)\prime}\right]\right)^{-1}\boldsymbol{f}_t^{(1)} \\ &= \boldsymbol{f}_t^{(1)} - \boldsymbol{f}_t^{(1)} \\ &= \mathbf{0}, \end{aligned}$$

In order to appreciate the implications of this result, let $\mathrm{col}\left(\boldsymbol{C}_0\right)$ denote the column space of $\boldsymbol{C}_0$ and $\ker\left(\boldsymbol{C}_0'\right)$ the kernel of $\boldsymbol{C}_0'$. The result reported in the last sequence of equations above states that $\boldsymbol{f}_t^{\perp} \in \ker(\boldsymbol{C}_0')$. Note now that $\dim(\mathrm{col}(\boldsymbol{C}_0)) + \dim(\ker(\boldsymbol{C}_0')) = r$ (see e.g. Abadir and Magnus 2005, Exercise 4.3(c)) and that $\dim(\mathrm{col}(\boldsymbol{C}_0)) = \dim(\mathrm{col}(\boldsymbol{C}_0'))$ (Abadir and Magnus 2005, Exercise 4.5(a)). As a result of the rank condition on $\boldsymbol{C}_0$, we have $\dim(\mathrm{col}(\boldsymbol{C}_0')) = m+1$ and hence $\dim(\ker(\boldsymbol{C}_0')) = r-m-1$. Consequently, $\boldsymbol{f}_t^{\perp}$ is an element in the $s = r-m-1$-dimensional vector space orthogonal to $\boldsymbol{C}_0$. This implies that, given an arbitrary basis for $\ker(\boldsymbol{C}_0')$, here denoted by the $r \times s$-dimensional matrix $\boldsymbol{\Phi}_2$, we can express $\boldsymbol{f}_t^{\perp}$ as $\boldsymbol{f}_t^{\perp} = \boldsymbol{\Phi}_2 \boldsymbol{f}_t^{(2)}$ where $\boldsymbol{f}_t^{(2)}$ is some $s \times 1$ vector of coordinates. Plugging this expression into (S.113) results in

$$\begin{aligned}
\boldsymbol{\lambda}_i'\boldsymbol{f}_t &= \boldsymbol{\lambda}_i'\boldsymbol{\Phi}_1\boldsymbol{f}_t^{(1)} + \boldsymbol{\lambda}_i'\boldsymbol{\Phi}_2\boldsymbol{f}_t^{(2)} \\
&= \boldsymbol{\lambda}_i' \begin{bmatrix} \boldsymbol{\Phi}_1, & \boldsymbol{\Phi}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_t^{(1)} \\ \boldsymbol{f}_t^{(2)} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\lambda}_i^{(1)\prime}, & \boldsymbol{\lambda}_i^{(2)\prime} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_t^{(1)} \\ \boldsymbol{f}_t^{(2)} \end{bmatrix}
\end{aligned}$$

From the previously mentioned result $\mathrm{E}\left[\boldsymbol{f}_t^{\perp}\boldsymbol{f}_t^{(1)\prime}\right] = \mathbf{O}$ it follows that $\mathrm{E}\left[\boldsymbol{f}_t^{(2)}\boldsymbol{f}_t^{(1)\prime}\right] = \mathbf{O}$. Furthermore, given that $\mathrm{E}[\boldsymbol{\lambda}_i] = \boldsymbol{C}_0\boldsymbol{B}^{-1}[1,\mathbf{0}']'$ and that $\boldsymbol{\lambda}_i^{(2)\prime} = \boldsymbol{\lambda}_i'\boldsymbol{\Phi}_2$ where $\boldsymbol{\Phi}_2$ is a basis for $\ker(\boldsymbol{C}_0')$,

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{\lambda}_i^{(2)}\right] &= \boldsymbol{\Phi}_2'\boldsymbol{C}_0\boldsymbol{B}^{-1}\left[1,\mathbf{0}'\right]' \\
&= \mathbf{0}.
\end{aligned}$$

Analogously, we can define $\left[\boldsymbol{\Lambda}_i^{(1)\prime}, \boldsymbol{\Lambda}_i^{(2)\prime}\right] = \boldsymbol{\Lambda}_i'[\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2]$ where $\mathrm{E}\left[\boldsymbol{\Lambda}_i^{(2)}\right] = \mathbf{O}$ by the same reasoning as above.

Consequently, whenever the factor estimates include cross-section averages of the dependent variable $\boldsymbol{Y}$, we can rewrite the common component as being generated by a linear combination of the true factors and the true loadings such that the two moment conditions above are satisfied. The matrix $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$ which links the loadings $\boldsymbol{\lambda}_i'$ ($\boldsymbol{\Lambda}_i'$) and $\left[\boldsymbol{\lambda}_i^{(1)\prime}, \boldsymbol{\lambda}_i^{(2)\prime}\right]$ ($\left[\boldsymbol{\Lambda}_i^{(1)\prime}, \boldsymbol{\Lambda}_i^{(2)\prime}\right]$) is of full rank by the rank assumptions on $\mathrm{E}[\boldsymbol{f}_t\boldsymbol{f}_t']$ as well as $\boldsymbol{C}_0$. Hence, $\boldsymbol{\Phi}$ is an invertible rotation and we can write $\left[\boldsymbol{\lambda}_i^{(1)\prime}, \boldsymbol{\lambda}_i^{(2)\prime}\right]\left[\boldsymbol{f}_t^{(1)\prime}, \boldsymbol{f}_t^{(2)\prime}\right]' = \boldsymbol{\lambda}_i'\boldsymbol{\Phi}\boldsymbol{\Phi}^{-1}\boldsymbol{f}_t$. Now note that in factor models it is only the common component $\boldsymbol{\lambda}_i'\boldsymbol{f}_t$ that is identified whereas the loadings $\boldsymbol{\lambda}_i$ and factors $\boldsymbol{f}_t$ are identified only up to an invertible rotation. As argued above, the matrix $\boldsymbol{\Phi}$ is invertible. Hence, irrespective of the true values of $\boldsymbol{\lambda}_i$, $\boldsymbol{\Lambda}_i$ and $\boldsymbol{f}_t$ (subject to the assumptions we make on them), we can assume that the data are generated by $\boldsymbol{\Phi}'\boldsymbol{\lambda}_i$, $\boldsymbol{\Phi}'\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Phi}^{-1}\boldsymbol{f}_t$ instead. □

Lemma S.19. *Under Assumptions 2-6 and* $\mathrm{E}[|\varsigma_{v,i}^{-1}|^4] < M$,

$$k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varpi}_j = \mathcal{O}_P\left[\max\left(N^{-1}, T^{-1}\right)\left(k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j\right)\right] \tag{S.114}$$

Proof of Lemma S.19. Note that

$$\boldsymbol{\varpi}_i = \left[\varsigma_{\nu,i}^{-1}\boldsymbol{\nu}_i - \boldsymbol{D}_i\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right]$$

where

$$\begin{aligned}\boldsymbol{\nu}_i &= \boldsymbol{\varepsilon}_i + \boldsymbol{F}^{(2)}\boldsymbol{\lambda}_i^{(2)},\\ \boldsymbol{D}_i &= \boldsymbol{U}_i + \boldsymbol{F}^{(2)}\boldsymbol{C}_i^{(2)}.\end{aligned}$$

Then, we can rewrite

$$\begin{aligned}\boldsymbol{\varpi}_i = &\left[\boldsymbol{\varepsilon}_i\varsigma_{\nu,i}^{-1} - \boldsymbol{U}_i\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right]\\ &+\boldsymbol{F}^{(2)}\left[\boldsymbol{\lambda}_i^{(2)}\varsigma_{\nu,i}^{-1} - \boldsymbol{C}_i^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right].\end{aligned}$$

Consequently, we can write out

$$\begin{aligned}&k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varpi}_j\\ &\le 2\sqrt{\frac{T}{2N(N-1)}}\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|\left\|T^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{F}^{(2)}\right\|^2\\ &\times\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{\nu,i}^{-1} - \sum_{i=1}^{N}\boldsymbol{C}_i^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right\|^2\\ &+2k_{N,T}\sum_{i,j}^{N}\left(\boldsymbol{\varepsilon}_i\varsigma_{\nu,i}^{-1} - \boldsymbol{U}_i\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\\ &\times\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_j\varsigma_{\nu,j}^{-1} - \boldsymbol{U}_j\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right),\end{aligned}\tag{S.115}$$

where the second line is of order $\mathcal{O}_P\left(T^{-1/2}\right)+\mathcal{O}_P\left(N^{-1/2}\right)+\mathcal{O}_P\left[N^{-1}\sqrt{T}\right]$ by Lemma S.3 with $w_i=\varsigma_{\nu,i}^{-1}$. Concerning the first line:

$$\sqrt{\frac{T}{2N(N-1)}}\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{\nu,i}^{-1} - \sum_{i=1}^{N}\boldsymbol{C}_i^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right\|^2 = \mathcal{O}_P\left(k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j\right).\tag{S.116}$$

To appreciate this result, we write

$$\begin{aligned}&\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{\nu,i}^{-1} - \sum_{i=1}^{N}\boldsymbol{C}_i^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\left(N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right)\right\|^2\\ =&3\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{\nu,i}^{-1}\right\|^2+3\left\|\sum_{i=1}^{N}\boldsymbol{C}_i^{(2)}\right\|^2\left\|\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\right\|^2\left\|N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{\nu,\ell}^{-1}\right\|^2\end{aligned}$$

and note that $N^{-2}\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(1)}\varsigma_{v,i}^{-1}\right\|^2 = \mathcal{O}_P(1)$ by the same reasoning leading to (S.44). Furthermore, $\left\|\sum_{i=1}^{N}\boldsymbol{C}_i^{(2)}\right\|^2 = \mathcal{O}_P(N)$ holds by the zero mean property of $\boldsymbol{\lambda}_i^{(2)}$ and $\boldsymbol{\Lambda}_i^{(2)}$. Consequently,

$$\left\|\sum_{i=1}^{N}\boldsymbol{C}_i^{(2)}\right\|^2\left\|\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\right\|^2\left\|N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\lambda}_\ell^{(1)}\varsigma_{v,\ell}^{-1}\right\|^2 = \mathcal{O}_P(N),$$

which never has a higher order in probability than $\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{v,i}^{-1}\right\|^2$. We prove this claim by noting that the expected value of $\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{v,i}^{-1}\right\|^2$ is given by

$$\mathrm{E}\left[\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{v,i}^{-1}\right\|^2\right] = \sum_{i=1}^{N}\mathrm{E}\left[\varsigma_{v,i}^{-2}\boldsymbol{\lambda}_i^{(2)\prime}\boldsymbol{\lambda}_i^{(2)}\right] + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathrm{E}\left[\varsigma_{v,i}^{-1}\boldsymbol{\lambda}_i^{(2)\prime}\right]\mathrm{E}\left[\boldsymbol{\lambda}_j^{(2)}\varsigma_{v,j}^{-1}\right].$$

It follows straightforwardly from Markov's inequality that

$$\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{v,i}^{-1}\right\|^2 = \begin{cases}\mathcal{O}_P(N) & \text{if } \mathrm{cov}\left[\boldsymbol{\lambda}_j^{(2)};\varsigma_{v,j}^{-1}\right] = \mathbf{0} \\ \mathcal{O}_P\left(N^2\right) & \textit{otherwise}\end{cases}.$$

The reduction in the order in probability due to a zero correlation between $\boldsymbol{\lambda}_i^{(2)}$ and $\varsigma_{v,i}^{-1}$ is the same as observed in the case of $k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j$. Additionally using the scaling $\sqrt{\frac{T}{2N(N-1)}}$ and using our result on the order of $k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j$ in the proof of Proposition 3, we arrive at

$$\sqrt{\frac{T}{2N(N-1)}}\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^{(2)}\varsigma_{v,i}^{-1}\right\|^2 = \mathcal{O}_P\left(k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j\right),$$

which leads to result (S.116). Now returning to the first line of expression (S.115), we consider

$$\begin{aligned}&\left\|T^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{F}^{(2)}\right\|^2\\ \leq&3\left\|\overline{\boldsymbol{C}}^{(1)\prime}\right\|^2\left\|T^{-1}\boldsymbol{F}^{(1)\prime}\boldsymbol{F}^{(2)}\right\|^2 + 9\left\|\overline{\boldsymbol{C}}^{(2)\prime}\right\|^2\left\|T^{-1}\boldsymbol{F}^{(2)\prime}\boldsymbol{F}^{(2)}\right\|^2 + k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j\left\|T^{-1}\overline{\boldsymbol{U}}'\boldsymbol{F}^{(2)}\right\|^2\\ =&\mathcal{O}_P\left(T^{-1}\right) + \mathcal{O}_P\left(N^{-1}\right) + \mathcal{O}_P\left((NT)^{-1}\right),\end{aligned} \tag{S.117}$$

which holds by zero correlation between $\boldsymbol{F}^{(1)}$ and $\boldsymbol{F}^{(2)}$, the zero expected value of $\overline{\boldsymbol{C}}^{(2)}$ and result (S.42). Consequently, an upper bound for the order in probability of (S.115) is given by

$$k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{\varpi}_j = \mathcal{O}_P\left[\max\left[N^{-1}, T^{-1}\right]\left(k_{N,T}\sum_{i,j}^{N}\boldsymbol{\varpi}_i'\boldsymbol{\varpi}_j\right)\right].$$

□

Lemma S.20. *Under Assumptions 2-6 and* $\mathrm{E}[|\varsigma_{\nu,i}^{-1}|^4] < M$,

$$k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left(\boldsymbol{\nu}_i' - \boldsymbol{\lambda}_i^{(1)\prime}\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\overline{\boldsymbol{D}}'\right)\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\nu}_i - \overline{\boldsymbol{D}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)$$
$$= k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\nu}_i'\boldsymbol{\nu}_i + \mathcal{O}_P\left(N^{-1/2}\sqrt{T}\right) + \mathcal{O}_P\left(T^{-1/2}\right)$$

Proof of Lemma S.20. The term that we are interested in can be expanded into

$$k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left(\boldsymbol{\nu}_i' - \boldsymbol{\lambda}_i^{(1)\prime}\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\overline{\boldsymbol{D}}'\right)\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\nu}_i - \overline{\boldsymbol{D}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)$$
$$= k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\nu}_i'\boldsymbol{\nu}_i - 2k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\nu}_i'\overline{\boldsymbol{D}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}$$
$$+ k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\lambda}_i^{(1)\prime}\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\overline{\boldsymbol{D}}'\overline{\boldsymbol{D}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}$$
$$+ k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left(\boldsymbol{\nu}_i' - \boldsymbol{\lambda}_i^{(1)\prime}\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\overline{\boldsymbol{D}}'\right)\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\nu}_i - \overline{\boldsymbol{D}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)$$
$$= k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\nu}_i'\boldsymbol{\nu}_i - 2a_1 + a_2 + a_3,$$

where we recall that $\boldsymbol{\nu}_i = \boldsymbol{\varepsilon}_i + \boldsymbol{F}^{(2)}\boldsymbol{\lambda}_i^{(2)}$ and $\boldsymbol{D}_i = \boldsymbol{U}_i + \boldsymbol{F}^{(2)}\boldsymbol{C}_i^{(2)}$. Now consider a_3. Similar to the proof of Lemma S.19, we can write

$$a_3 \leq 2k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left(\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)'\boldsymbol{F}^{(2)\prime}\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{F}^{(2)}\left(\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)$$
$$+ 2k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)'\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{U}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right).$$

By Lemma S.4, the last line in the expression above is $\mathcal{O}_P\left(N^{-2}\sqrt{T}\right) + \mathcal{O}_P\left(N^{-3/2}\right) + \mathcal{O}_P\left(T^{-1/2}\right)$. Concerning the remaining term, we can write

$$k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left(\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)'\boldsymbol{F}^{(2)\prime}\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{F}^{(2)}\left(\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)$$
$$\leq Tk_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left\|\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right\|^2\left\|T^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{F}^{(2)}\right\|^2\left\|\left(T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}\right)^{-1}\right\|$$

where we focus on the term

$$Tk_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left\|\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right\|^2$$
$$\leq \left(Tk_{N,T}3\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\lambda}_i^{(2)\prime}\boldsymbol{\lambda}_i^{(2)}\right) + \left(Tk_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\lambda}_i^{(1)\prime}\boldsymbol{\lambda}_i^{(1)}\right)\left\|\overline{\boldsymbol{C}}^{(2)}\right\|^2\left\|\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\right\|^2$$
$$= \mathcal{O}_P\left(\sqrt{T}\right).$$

This result follows from applying the reasoning given below equation (S.46) to both weighted sums over i in the expression above. Combining it with result (S.117), we arrive at

$$
\begin{aligned}
& k_{N,T} \sum_{i=1}^{N} \varsigma_{v,i}^{-2} \left(\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)} \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \boldsymbol{\lambda}_i^{(1)} \right)' \boldsymbol{F}^{(2)\prime} \boldsymbol{P}_{\widehat{\boldsymbol{F}}} \boldsymbol{F}^{(2)} \left(\boldsymbol{\lambda}_i^{(2)} - \overline{\boldsymbol{C}}^{(2)} \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \boldsymbol{\lambda}_i^{(1)} \right) \\
& = \mathcal{O}_P \left(N^{-1} \sqrt{T} \right) + \mathcal{O}_P \left(T^{-1/2} \right).
\end{aligned}
$$

which, together with Lemma S.4, leads to

$$
a_3 = \mathcal{O}_P \left(N^{-1} \sqrt{T} \right) + \mathcal{O}_P \left(T^{-1/2} \right). \tag{S.118}
$$

Next, we investigate a_2 and write

$$
\begin{aligned}
a_2 & \leq k_{N,T} \sum_{i=1}^{N} \left\| \varsigma_{v,i}^{-1} \boldsymbol{\lambda}_i^{(1)\prime} \right\|^2 \left\| \left(\overline{\boldsymbol{C}}^{(1)\prime} \right)^{-1} \right\|^2 3 \left(\left\| \overline{\boldsymbol{C}}^{(2)} \right\|^2 \| \boldsymbol{F} \|^2 + \| \overline{\boldsymbol{U}} \|^2 \right) \\
& = \mathcal{O}_P \left(N^{-1} \sqrt{T} \right),
\end{aligned} \tag{S.119}
$$

where the last line follows from result (S.41), the zero mean of $\overline{\boldsymbol{C}}^{(2)}$ and the result on $\sum_{i=1}^{N} \varsigma_{v,i}^{-2} \boldsymbol{\lambda}_i^{(1)\prime} \boldsymbol{\lambda}_i^{(1)}$ referred to above. The last term to consider is given by

$$
\begin{aligned}
a_1 & = k_{N,T} \sum_{i=1}^{N} \varsigma_{v,i}^{-2} \boldsymbol{\varepsilon}' \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \boldsymbol{\lambda}_i^{(1)} + k_{N,T} \sum_{i=1}^{N} \varsigma_{v,i}^{-2} \boldsymbol{\lambda}_i^{(2)\prime} \boldsymbol{F}^{(2)\prime} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \boldsymbol{\lambda}_i^{(1)} \\
& + k_{N,T} \sum_{i=1}^{N} \varsigma_{v,i}^{-2} \boldsymbol{\varepsilon}_i' \boldsymbol{F}^{(2)} \overline{\boldsymbol{C}}^{(2)} \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \boldsymbol{\lambda}_i^{(1)} \\
& + k_{N,T} \sum_{i=1}^{N} \varsigma_{v,i}^{-2} \boldsymbol{\lambda}_i^{(2)\prime} \boldsymbol{F}^{(2)\prime} \boldsymbol{F}^{(2)} \overline{\boldsymbol{C}}^{(2)} \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \boldsymbol{\lambda}_i^{(1)} \\
& = a_{11} + a_{12} + a_{13} + a_{14}.
\end{aligned}
$$

Here, $a_{11} = \mathcal{O}_P \left(N^{-1/2} \right) + \mathcal{O}_P \left(N^{-1} \sqrt{T} \right)$ results from proceeding almost exactly as in the steps leading to equation (S.26) in the proof of Theorem 2. Concerning the second term, we can write

$$
\begin{aligned}
|a_{12}| & \leq k_{N,T} \sum_{i=1}^{N} \left\| \varsigma_{v,i}^{-1} \boldsymbol{\lambda}_i^{(2)} \right\| \left\| \boldsymbol{F}^{(2)\prime} \overline{\boldsymbol{U}} \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \right\| \left\| \varsigma_{v,i}^{-1} \boldsymbol{\lambda}_i^{(1)} \right\| \\
& \leq k_{N,T} N \sqrt{\frac{T}{N}} \left(N^{-1} \sum_{i=1}^{N} \left\| \varsigma_{v,i}^{-1} \boldsymbol{\lambda}_i^{(2)} \right\|^2 \right)^{1/2} \left(N^{-1} \sum_{i=1}^{N} \left\| \varsigma_{v,i}^{-1} \boldsymbol{\lambda}_i^{(1)} \right\|^2 \right)^{1/2} \left\| \sqrt{\frac{N}{T}} \boldsymbol{F}^{(2)\prime} \overline{\boldsymbol{U}} \right\| \left\| \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \right\| \\
& = \mathcal{O}_P \left(N^{-1/2} \right),
\end{aligned}
$$

where we make use of (S.42) with $w_i = 1$.With regards to the third term, we have

$$
\begin{aligned}
|a_{13}| & \leq k_{N,T} \sqrt{NT} \left((NT)^{-1} \sum_{i=1}^{N} \left\| \boldsymbol{\varepsilon}_i' \boldsymbol{F}^{(2)} \right\|^2 \right)^{1/2} \left(N^{-1} \sum_{i=1}^{N} \left\| \boldsymbol{\lambda}_i^{(1)} \varsigma_{v,i}^{-2} \right\|^2 \right)^{1/2} \left\| \sqrt{N} \overline{\boldsymbol{C}}^{(2)} \right\| \left\| \left(\overline{\boldsymbol{C}}^{(1)} \right)^{-1} \right\| \\
& = \mathcal{O}_P \left(N^{-1/2} \right).
\end{aligned}
$$

Here, we can use results from the proof of Lemma S.4 to obtain $(NT)^{-1}\sum_{i=1}^{N}\left\|\boldsymbol{\varepsilon}_i'\boldsymbol{F}^{(2)}\right\|^2 = \mathcal{O}_P(1)$. Lastly, we can apply the Cauchy-Schwarz inequality to $|a_{14}|$ in order to arrive at

$$\begin{aligned}|a_{14}| &= k_{N,T}\sqrt{N}T\left(N^{-1}\sum_{i=1}^{N}\left\|\varsigma_{\nu,i}^{-1}\boldsymbol{\lambda}_i^{(2)}\right\|^2\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\left\|\varsigma_{\nu,i}^{-1}\boldsymbol{\lambda}_i^{(1)}\right\|^2\right)^{1/2}\\ &\quad\times\left\|T^{-1/2}\boldsymbol{F}^{(2)}\right\|^2\left\|\sqrt{N}\overline{\boldsymbol{C}}^{(2)}\right\|\left\|\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\right\|\\ &= \mathcal{O}_P\left(N^{-1/2}\sqrt{T}\right).\end{aligned}$$

Using the last four intermediary results allows us to conclude that

$$a_1 = \mathcal{O}_P\left(N^{-1/2}\sqrt{T}\right).$$

This result, together with (S.118) and (S.119) establishes

$$\begin{aligned}&k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\left(\boldsymbol{\nu}_i'-\boldsymbol{\lambda}_i^{(1)\prime}\left(\overline{\boldsymbol{C}}^{(1)\prime}\right)^{-1}\overline{\boldsymbol{D}}'\right)\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\left(\boldsymbol{\nu}_i'-\overline{\boldsymbol{D}}\left(\overline{\boldsymbol{C}}^{(1)}\right)^{-1}\boldsymbol{\lambda}_i^{(1)}\right)\\ &= k_{N,T}\sum_{i=1}^{N}\varsigma_{\nu,i}^{-2}\boldsymbol{\nu}_i'\boldsymbol{\nu}_i+\mathcal{O}_P\left(N^{-1/2}\sqrt{T}\right)+\mathcal{O}_P\left(T^{-1/2}\right),\end{aligned}$$

which was to be shown. □

References

ABADIR, K. M. AND J. R. MAGNUS (2005): *Matrix algebra*, vol. 1 of *Econometric Exercises*, Cambridge University Press.

BALTAGI, B. H., Q. FENG, AND C. KAO (2012): "A Lagrange Multiplier Test for Cross-sectional Dependence in a Fixed Effects Panel Data Model," *Journal of Econometrics*, 170, 164 – 177.

BALTAGI, B., C. KAO, AND B. PENG (2016): "Testing Cross-Sectional Correlation in Large Panel Data Models with Serial Correlation," *Econometrics*, 4, –.

CHEN, S. X. AND Y.-L. QIN (2010): "A Two-sample Test for High-dimensional Data with Applications to Gene-set Testing," *The Annals of Statistics*, 38, 808–835.

HALL, P. AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*, Probability and Mathematical Statistics, Academic Press.

KAO, C., L. TRAPANI, AND G. URGA (2012): "Asymptotics for Panel Models with Common Shocks," *Econometric Reviews*, 31, 390–439.

KAPETANIOS, G., L. SERLENGA, AND Y. SHIN (2019): "Testing for Correlated Factor Loadings in Cross Sectionally Dependent Panels," SERIES Working papers N. 02/2019. Available at SSRN: `https://ssrn.com/abstract=3401745`

KARABIYIK, H., S. REESE, AND J. WESTERLUND (2017): "On The Role of The Rank Condition in CCE Estimation of Factor-augmented Panel Regressions," *Journal of Econometrics*, 197, 60–64.

MOON, H.R., B. PERRON, AND P.C.B. PHILLIPS (2007): "Incidental Trends and the Power of Panel Unit Root Tests," *Journal of Econometrics*, 141, 416–459.

PARKER, J. AND D. SUL (2016): "Identification of Unknown Common Factors: Leaders and Followers," *Journal of Business & Economic Statistics*, 34, 227–239.

PESARAN, M. H. (2004): "General Diagnostic Tests for Cross Section Dependence in Panels," CESifo Working Paper No. 1229.

——— (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

——— (2015): "Testing Weak Cross-Sectional Dependence in Large Panels," *Econometric Reviews*, 34, 1089–1117.

PESARAN, M. H., A. ULLAH, AND T. YAMAGATA (2008): "A Bias-adjusted LM Test of Error Cross-section Independence," *The Econometrics Journal*, 11(1), 105–127.

PHILLIPS, P. C. B. AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

SARAFIDIS, V., T. YAMAGATA, AND D. ROBERTSON (2009): "A Test of Cross Section Dependence for a Linear Dynamic Panel Model with Regressors," *Journal of Econometrics*, 148, 149–161.

WESTERLUND, J. AND J.-P. URBAIN (2013): "On the Estimation and Inference in Factor-augmented Panel Regressions with Correlated Loadings," *Economics Letters*, 119, 247 – 250.

WHITE, H. (2001): *Asymptotic Theory for Econometricians*, Academic Press.