Internet Appendix for "Taming the Factor Zoo: a Test of New Factors"

Guanhao Feng[*]
College of Business
City University of Hong Kong

Stefano Giglio[†]
Yale School of Management
NBER and CEPR

Dacheng Xiu[‡]
Booth School of Business
University of Chicago

This Version: January 2, 2019

Abstract

This appendix contains Monte Carlo simulations, mathematical proofs, and robustness for Table 2 in the paper.

Appendix A Simulation Evidence

One of the central advantages of our double-selection method is that it produces proper inference on the SDF loading λ_g of a factor, taking explicitly into account the possibility that the model-selection step (based on LASSO) may mistakenly include some irrelevant factors or exclude useful factors in any finite sample.

In this section, we therefore study the finite-sample performance of our inference procedure using Monte Carlo simulations. In particular, we show that if one were to make inference on λ_g by selecting the control factors via standard LASSO (and ignoring potential mistakes in model selection), the omitted variable bias resulting from selection mistakes would yield incorrect inference about λ_g. Instead, our double-selection procedure fully corrects for this problem in a finite sample and produces valid inference. In what follows, we first give details of the simulation procedure and then show the results of the Monte Carlo experiment.

A.1 Simulating the Data-Generating Process

We are interested in making inference on λ_g, the vector of SDF loadings of three factors in g_t. g_t includes a useful factor (denoted as g_{1t}) as well as a useless factor and a redundant factor (denoted

[*]Address: 83 Tat Chee Avenue, Kowloon Tong, Hong Kong. E-mail address: gavin.feng@cityu.edu.hk.
[†]Address: 165 Whitney Avenue, New Haven, CT 06520, USA. E-mail address: stefano.giglio@yale.edu.
[‡]Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: dacheng.xiu@chicagobooth.edu.

together as a 2×1 vector g_{2t}). g_{2t} has a zero SDF loading, that is, $\lambda_{g_2} = 0$, but the covariance of the redundant factor is correlated with the cross section of expected returns. In our simulation, h_t is a large set of factors that includes 4 useful factors h_{1t}, and $p - 4$ useless and redundant factors collected in h_{2t} (so the total dimension of h_t is p).

We simulate returns of test assets and factors according to the following steps:

(1) Simulate C_e ($n \times d$) and C_{h_1} ($n \times 4$) independently from multivariate normal distributions.

(2) Calculate $C_{h_2} = \iota_n \theta_0^\intercal + C_{h_1} \theta_1^\intercal + C_\epsilon$, where C_ϵ is simulated independently from an $n \times (p-4)$ multivariate normal distribution, θ_0 is a $(p-4) \times 1$ vector, and θ_1 is a $(p-4) \times 4$ matrix.

(3) Calculate C_g from C_e and $C_h = (C_{h_1} : C_{h_2})$ using $C_g = \iota_n \xi + C_h \chi^\intercal + C_e$, where χ is a $d \times p$ matrix.

(4) Calculate C_z using $C_z = C_g - C_h \eta^\intercal$, as implied from the DGP $g_t = \eta h_t + z_t$ we aim to simulate, where η is a $d \times p$ matrix.

(5) Calculate $\mathrm{E}(r_t)$ using $\mathrm{E}(r_t) = \iota_n \gamma_0 + C_g \lambda_g + C_h \lambda_h$, where λ_g is a $d \times 1$ vector and λ_h is a $p \times 1$ vector.

(6) Calculate $\beta_g = C_z \Sigma_z^{-1}$ and $\beta_h = C_h \Sigma_h^{-1} - \beta_g \eta$, as implied from the DGP of r_t we aim to simulate: $r_t = \mathrm{E}(r_t) + \beta_g g_t + \beta_h h_t + u_t$.

(7) For each Monte Carlo trial, generate u_t from a Student's t distribution with 5 degrees of freedom and a covariance matrix Σ_u. Generate $h_t \sim \mathcal{N}_p(0, \Sigma_h)$, $z_t \sim \mathcal{N}_d(0, \Sigma_z)$, and calculate g_t and then r_t using the DGPs specified in Steps (4) and (6), respectively.

The total number of Monte Carlo trials is 2,000. Because we assume non-random selection of assets and that the randomness in the selection of test assets does not affect the inference to the first order, we simulate only once C_g, C_h, and hence β_g, β_h, in Steps (1) - (6), so that they are constant throughout the Monte Carlo trials in Step (7).

We calibrate our DGP to mimic the actual Fama-French 5-factor model. In particular, we calibrate χ, η, λ, Σ_z, the mean and covariance matrices of C_e, C_{h_1}, as well as Σ_h to match the summary statistics (times series and cross-sectional R^2, factor-return covariances, etc.) of the Fama-French five factors estimated using characteristic-sorted portfolios, described in detail in Section 3. We calibrate a diagonal Σ_u to match the average time series R^2 for this 5-factor model. For redundant and useless factors, we calibrate their parameters using all the other factors in our data library, again described in detail in Section 3. We maintain the sparsity requirement on χ, η, and λ, by restricting the loadings of C_g, $\mathrm{E}(r_t)$ and g on C_{h_2} and h_2 to be zero. We set to zero the loading of C_g on C_h for the useless factor in g_2. Moreover, we randomly simulate θ_1 from normal distribution so that factors in h_2 are either redundant or (rather close to be) useless. We allow non-zero loading of g_2 on h_1, and the covariance matrix Σ_h to be non-diagonal, so that both useless and redundant factors in g_2 and h_2 can be correlated with the true factors in g_1 and h_1: so they will command risk *premia* simply due to this correlation, even though they have zero SDF loadings because they do not affect marginal utility once the true factors are controlled for.

A.2 Simulation Results

We report here the results of various simulations from the model. We consider various settings with number of total factors $p = 25, 50, 100, 200$, number of assets $n = 100, 200, 300$, and length of time series $T = 240, 360, 480$.

Figure A1 compares the asymptotic distributions of the proposed double-selection estimator with that of the single-selection estimator for the case $p = 100$, $n = 300$, and $T = 480$. The right side of the figure shows the distribution of the t-test for λ_g of the three factors (useful in the first row, redundant in the second row, and useless in the third row) when using the controls selected by standard LASSO (i.e., a single-selection-based estimator). The panels show that inference without double-selection adjustment displays substantial biases for useful and redundant factors and distortion from normality for all factors. The left side of the figure shows instead that our double-selection procedure produces an unbiased and asymptotically normal test, as predicted by Theorem 1.

Figure A2 plots the frequency with which each of the simulated factors is selected across simulations (with each bar corresponding to a different simulated factor, identified by its ID from 1 to 100). The top panel corresponds to the factors selected in the first LASSO selection, the second panel corresponds to the factors selected in the second selection, and the last panel corresponds to the union of the two.

Note that by construction, the true factors in h_t are the first 4 (the fifth true factor is part of g_t). So if model selection were able to identify the right control factors in all samples perfectly, the first 4 bars should read 100%, while all other bars (corresponding to factors 5-100) should read 0%.

That is not the case in the simulations. While some factors are often selected by LASSO (top panel), not all are: factor 1 is selected in about 70% of the samples, and factor 3 about 40% of the samples. Therefore, in a large fraction of samples, the control model would be missing some true factors, generating the omitted variable bias displayed in Figure A1. At the same time, LASSO often includes erroneously spurious factors – as shown in Table A5. The key to correct inference that our procedure achieves is that the two-step selection procedure minimizes the potential omitted factor bias.

Tables A2, A3, and A4 compare the biases and root-mean-squared errors (RMSEs) for double-selection (DS), single-selection (SS), and the OLS estimators of each entry of λ_g, respectively. All regularization parameters are selected based on 10-fold cross-validation.

Not surprisingly, the bias of the SS is clearly visible when compared to DS and OLS for useful and redundant factors. In addition, DS outperforms SS and OLS in terms of their RMSEs in these scenarios. The efficiency gain of DS over OLS is particularly substantial when p is large relative to n. When p is equal to n, OLS becomes infeasible (because the number of regressors is $p + d$). For the useless factor, because SS does not suffer from a bias, its RMSE is the smallest among all. This result confirms the efficiency benefits of machine learning techniques over OLS. Although DS is in general less biased than SS, its main advantage relative to SS is in removing the distortions to inference, visible from the distribution of standardized statistics in Figure A1.

Overall, the simulation results confirm our econometric analysis: the DS estimator outperforms the benchmarks.

Table A1: Testing Factors Recursively by Year of Sample

	(1)	(2)	(3)										
Year	# Assets	# Controls	New factors (IDs)										
1995	240	47	42	43	46	52	53	54	55	56	57	69	71
1996	306	58	47										
1997	306	59	58	59	60	61	63	67					
1998	342	65	48	49	50	51							
1999	360	69	65	92	95								
2000	378	72	62	74	93	97	98	99	104				
2001	408	79	68	70	75	76	77	78	79	80	81	82	83
			84	85	86	87	88	89	90	101	102	108	
2002	504	100	72	73	112	116	117	118	119	120			
2003	546	108	94	96	100	105	106	114					
2004	582	114	111										
2005	588	115	103	113	115	123	127	129	131				
2006	630	122	91	110	122	126	144						
2007	654	127											
2008	654	127	124	125	128	130	132	134	139				
2009	696	134	121	133	135	136							
2010	720	138	138	142	143								
2011	738	141	137										
2012	738	142	140	141	147	148	149	150					
2013	738	148	145	146									

Note. Same as Table 2, but the date used to order the factors is the last date of the sample used in each paper.

Table A2: Asymptotic Approximation Performance for λ_{useful}

		$p=25$			$p=50$			$p=100$			$p=200$		
T	n	DS	SS	OLS	DS	SS	OLS	DS	SS	OLS	DS	SS	OLS
							Panel A: Bias						
240	100	-0.71	-9.23	-0.19	-0.96	-9.32	-0.13	-2.06	-11.26	-	-3.37	-9.88	-
240	200	-0.82	-9.53	-0.13	-0.95	-9.11	-0.14	-1.80	-9.01	-0.43	-3.14	-9.65	-
240	300	-0.26	-7.87	0.06	-1.06	-10.39	-0.50	-1.41	-8.43	-0.24	-2.81	-9.93	0.08
360	100	-0.31	-8.33	-0.14	-0.40	-8.71	0.08	-1.60	-10.66	-	-2.27	-9.07	-
360	200	-0.32	-8.48	0.00	-0.43	-8.44	-0.08	-1.33	-8.31	-0.28	-2.23	-8.79	-
360	300	-0.05	-7.07	0.18	-0.51	-9.44	-0.16	-1.09	-7.41	-0.13	-1.99	-8.79	-0.31
480	100	-0.21	-7.87	0.03	-0.12	-8.22	0.39	-1.02	-10.06	-	-1.83	-8.71	-
480	200	-0.14	-7.86	0.13	-0.19	-7.80	0.06	-0.87	-7.89	-0.09	-1.57	-8.57	-
480	300	-0.01	-6.76	0.15	-0.25	-8.74	0.05	-0.55	-7.18	-0.07	-1.33	-8.53	-0.11
							Panel B: RMSE						
240	100	5.80	11.60	6.46	6.14	11.57	8.19	7.52	13.93	-	8.98	12.23	-
240	200	5.78	12.05	5.84	5.94	11.56	6.55	6.73	11.33	9.07	7.88	11.82	-
240	300	5.54	10.33	5.66	5.83	13.05	5.98	6.46	11.09	7.21	7.54	11.94	19.76
360	100	4.62	10.90	5.07	4.88	10.94	6.68	5.73	13.12	-	6.88	11.23	-
360	200	4.53	11.23	4.63	4.66	10.80	5.23	5.22	10.57	6.55	6.54	10.84	-
360	300	4.40	9.66	4.49	4.66	12.23	4.84	5.03	9.99	5.55	6.13	10.89	10.11
480	100	4.10	10.31	4.44	4.16	10.60	5.53	5.01	12.77	-	5.92	10.83	-
480	200	3.99	10.63	4.12	4.00	10.21	4.45	4.47	10.17	5.64	5.41	10.52	-
480	300	3.88	9.23	4.01	3.92	11.51	4.15	4.22	9.77	4.71	4.85	10.56	7.88

Note. This table provides the biases and root-mean-squared errors (RMSE) of the estimates of the SDF loading λ of the useful factor from Monte Carlo simulations. DS is the double-selection estimator, SS is the single-selection estimator, and OLS is the ordinary least squares without selection. The regularization parameters in the LASSO are selected using 10-fold cross-validation, where we partition the cross-validation subsamples in the time series dimension. The true value λ_{useful} is 16.76. Note that in cases of $n \geq p$, OLS is infeasible.

Table A3: Asymptotic Approximation Performance for $\lambda_{\text{redundant}}$

		$p = 25$			$p = 50$			$p = 100$			$p = 200$		
T	n	DS	SS	OLS	DS	SS	OLS	DS	SS	OLS	DS	SS	OLS
						Panel A: Bias							
240	100	0.24	6.34	0.10	0.29	6.24	-0.22	0.64	7.23	-	1.58	5.95	-
240	200	0.39	6.78	0.14	0.11	7.25	0.08	0.33	6.74	0.06	1.22	5.76	-
240	300	0.17	5.98	0.07	0.15	6.92	0.06	0.63	6.39	-0.04	0.84	6.39	-0.46
360	100	0.09	5.20	0.09	0.04	5.36	0.06	0.06	6.60	-	0.74	6.07	-
360	200	0.08	5.63	0.02	0.06	6.38	-0.02	0.00	6.02	-0.05	0.50	5.35	-
360	300	0.08	4.86	0.08	0.10	5.95	0.04	0.12	5.87	0.04	0.33	6.37	0.08
480	100	0.04	4.64	0.08	0.00	4.80	-0.15	0.05	6.10	-	0.22	5.89	-
480	200	-0.03	5.12	-0.06	-0.01	5.53	0.04	0.01	5.84	0.08	0.11	5.51	-
480	300	0.02	4.56	0.02	-0.01	4.98	-0.03	0.07	5.45	-0.07	0.08	6.55	0.19
						Panel B: RMSE							
240	100	5.58	9.96	6.40	5.76	10.01	8.15	6.12	11.67	-	7.69	9.98	-
240	200	5.65	10.40	5.78	5.47	11.15	6.19	5.62	10.76	9.07	6.56	9.98	-
240	300	5.43	9.71	5.55	5.42	10.85	6.01	5.69	10.66	7.20	5.83	10.61	20.00
360	100	4.34	8.43	5.01	4.70	8.86	6.60	4.62	10.83	-	5.21	9.53	-
360	200	4.30	8.99	4.49	4.53	9.95	5.26	4.32	9.70	6.64	4.72	9.12	-
360	300	4.26	8.27	4.38	4.42	9.78	4.76	4.50	9.73	5.54	4.44	10.33	9.71
480	100	3.80	7.75	4.23	3.86	8.11	5.62	3.88	10.03	-	3.99	8.97	-
480	200	3.70	8.41	3.88	3.83	9.07	4.37	3.73	9.28	5.44	3.75	8.99	-
480	300	3.66	7.85	3.80	3.75	8.64	4.09	3.77	9.02	4.71	3.61	10.17	7.82

Note. This table provides the biases and root-mean-squared errors (RMSE) of the estimates of the SDF loading λ of the redundant factor from Monte Carlo simulations. DS is the double-selection estimator, SS is the single-selection estimator, and OLS is the ordinary least squares without selection. The regularization parameters in the LASSO are selected using 10-fold cross-validation, where we partition the cross-validation subsamples in the time series dimension. The true value $\lambda_{\text{redundant}}$ is 0. Note that in cases of $n \geq p$, OLS is infeasible.

Table A4: Asymptotic Approximation Performance for λ_{useless}

		$p=25$			$p=50$			$p=100$			$p=200$		
T	n	DS	SS	OLS	DS	SS	OLS	DS	SS	OLS	DS	SS	OLS
						Panel A: Bias							
240	100	-0.37	-1.04	-0.22	-0.19	-2.26	-0.11	-0.03	-0.85	-	-0.08	-0.37	-
240	200	0.03	1.86	-0.06	-0.20	-2.03	-0.32	-0.05	-0.66	-0.04	-0.23	-0.20	-
240	300	-0.35	-0.29	-0.28	-0.02	-0.09	-0.01	-0.08	-1.04	0.05	-0.05	0.03	0.43
360	100	-0.10	-0.71	-0.02	-0.18	-2.13	-0.06	0.13	-0.58	-	-0.03	-0.18	-
360	200	0.17	2.10	0.17	-0.23	-1.89	-0.31	0.06	-0.44	0.01	-0.06	-0.01	-
360	300	-0.11	-0.01	-0.12	-0.12	0.00	-0.19	0.02	-0.87	0.05	0.04	0.29	-0.24
480	100	0.01	-0.55	0.13	0.01	-1.89	0.09	-0.10	-0.78	-	0.07	-0.03	-
480	200	0.14	1.88	0.08	0.04	-1.53	0.03	-0.10	-0.65	-0.04	0.09	-0.06	-
480	300	0.03	0.07	0.05	0.14	0.16	0.06	-0.06	-0.96	0.09	0.13	0.30	-0.05
						Panel B: RMSE							
240	100	5.37	5.56	6.17	5.40	6.72	8.24	5.51	6.13	-	5.87	5.74	-
240	200	5.17	5.61	5.47	5.22	6.01	6.36	5.19	5.24	8.90	5.46	5.73	-
240	300	5.16	5.09	5.41	5.29	5.33	5.92	5.19	5.72	7.14	5.23	5.40	19.69
360	100	4.40	4.46	5.01	4.40	5.47	6.53	4.46	5.00	-	4.41	4.60	-
360	200	4.32	5.08	4.51	4.27	4.97	5.02	4.28	4.54	6.85	4.37	4.59	-
360	300	4.25	4.18	4.42	4.24	4.30	4.65	4.27	4.75	5.63	4.18	4.38	10.31
480	100	3.80	3.90	4.32	3.84	5.01	5.58	3.73	4.28	-	3.64	4.02	-
480	200	3.74	4.50	3.96	3.68	4.38	4.33	3.65	3.79	5.50	3.57	3.84	-
480	300	3.67	3.68	3.79	3.66	3.79	3.96	3.66	4.05	4.54	3.50	3.77	7.73

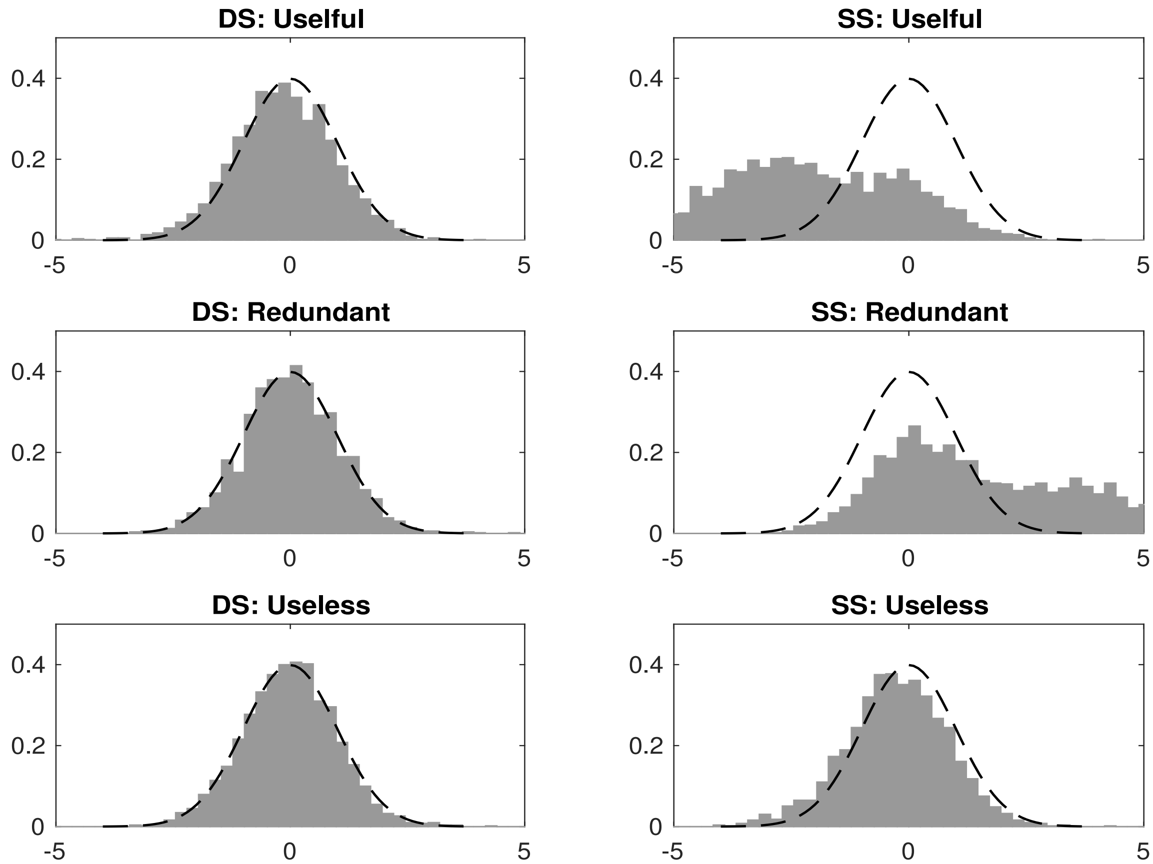
Note. This table provides the biases and root-mean-squared errors (RMSE) of the estimates of the SDF loading λ of the useless factor from Monte Carlo simulations. DS is the double-selection estimator, SS is the single-selection estimator, and OLS is the ordinary least squares without selection. The regularization parameters in the LASSO are selected using 10-fold cross-validation, where we partition the cross-validation subsamples in the time series dimension. The true value λ_{useless} is 0. Note that in cases of $n \geq p$, OLS is infeasible.

Table A5: Table of the Variable Selection in Simulations

		$p = 25$			$p = 50$			$p = 100$			$p = 200$		
T	n	1st	2nd	Total	1st	2nd	Total	1st	2nd	Total	1st	2nd	Total
						Panel A: Useful Factors							
240	100	45.5	98.5	99.2	45.6	97.3	98.4	46.8	94.6	96.6	46	86.9	92.1
240	200	46.5	97.3	98.5	47.4	98.4	99.1	47	96.6	97.9	45.7	86.1	90.9
240	300	48	98.7	99.2	48.1	99	99.3	50.4	94.2	96.7	48.5	89.7	93.7
360	100	52.2	99.7	99.9	50.4	99.2	99.6	50.9	98.2	98.9	48	94.3	96.9
360	200	54.5	99.2	99.6	52.9	99.6	99.8	51.6	99.1	99.4	50.2	94.2	96.6
360	300	54.6	99.8	99.9	53.6	99.7	99.8	54	97.3	98.5	51.9	96.5	98.1
480	100	56.1	99.9	100	54	99.8	99.9	53.7	99.3	99.6	49.9	97.6	98.7
480	200	57.9	99.7	99.9	57.4	99.9	99.9	53.1	99.7	99.8	50.3	97.8	98.7
480	300	57.1	100	100	58.5	99.9	100	56.2	99.2	99.7	51.9	98.5	99.2
						Panel B: Redundant and Useless Factors							
240	100	5.6	2.5	7.9	4.4	1.7	5.9	3.4	2.2	5.4	2.5	3	5.2
240	200	6.3	3.4	9.2	5.4	1.8	7	4.5	2.5	6.8	3.3	4.4	7.3
240	300	6.4	3.1	9.1	6.4	2	8.1	5.8	4.7	9.8	3.9	4.5	7.9
360	100	5	1.6	6.4	4.1	1	4.9	3.3	0.8	4	2.1	1	2.9
360	200	6.5	2.4	8.6	5.3	1.1	6.2	4.8	0.9	5.6	3.2	1.8	4.8
360	300	6.1	2.2	8	6.3	0.9	7.1	5.3	1.5	6.6	3.8	1.6	5.2
480	100	4.7	1.1	5.7	3.4	0.7	4	2.7	0.5	3.1	2	0.4	2.3
480	200	5.1	1.6	6.5	5.1	0.7	5.7	4.3	0.4	4.7	2.7	0.8	3.4
480	300	4.9	1.5	6.2	5.4	0.7	6	4.3	0.8	5	3	0.7	3.6

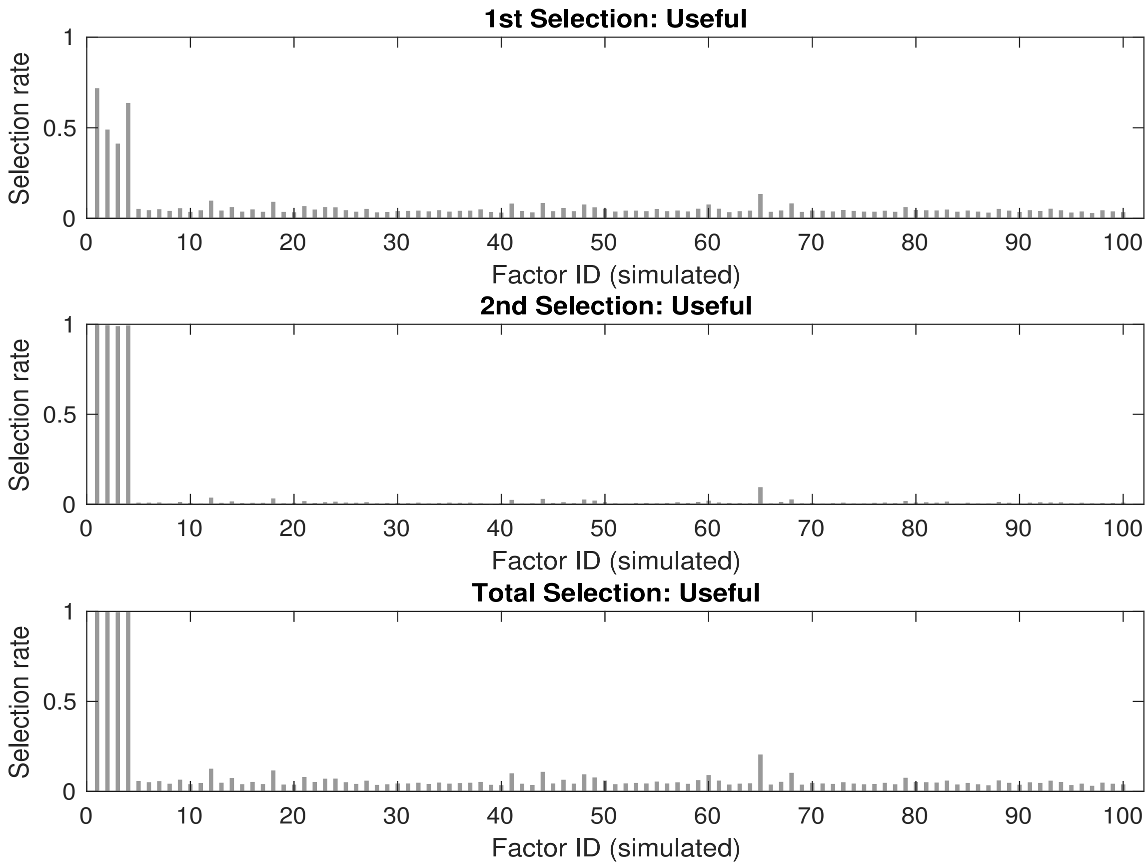
Note. The table reports how often useful, redundant and useless factors are selected in each step of our double selection procedure (first and second columns corresponding to the first and second step, and their union in the third column), in Monte Carlo simulations. Panel A reports the average selection percentages for useful factors, and Panel B reports the average selection percentages for redundant or useless factors. The regularization parameters in the LASSO are selected using 10-fold cross-validation, where we partition the cross-validation subsamples in the time series dimension.

Figure A1: Histograms of the Standardized Estimates in Simulations



Note. The figure presents the histograms of the standardized double-selection and single-selection estimates using estimated standard errors, compared with the standard normal density in solid dashed lines. The left panel reports the double-selection histograms, and the right panel the single-selection histograms. The top row reports the distribution of standardized estimates for a useful factor; the middle row for a redundant factor; the last row for a useless factor. In the simulation, we set $T = 480$, $n = 300$, and $p = 100$. The regularization parameters in the LASSO are selected using 10-fold cross-validation, where we partition the cross-validation subsamples in the time series dimension.

Figure A2: Histograms of the Selected Variables



Note. The figure reports how often each factor is selected in each step of our double selection procedure (first and second panels corresponding to the first and second step, and their union in the bottom panel), in Monte Carlo simulations. Each factor corresponds to a number on the X axis. Factors 1 - 4 are part of the true factors in the DGP. Factors 5 - 100 are either redundant or close to be useless. We set $T = 480$, $n = 300$, and $p = 100$. The regularization parameters in the LASSO are selected using 10-fold cross-validation, where we partition the cross-validation subsamples in the time series dimension.

Appendix B Technical Details and Proofs

B.1 Notation

We summarize the notation used throughout. Let e_i be a vector with 1 in the ith entry and 0 elsewhere, whose dimension depends on the context. Let ι_k denote a k-dimensional vector with all entries being 1. We use $a \vee b$ to denote the max of a and b, and $a \wedge b$ as their min for any scalars a and b. We also use the notation $a \lesssim b$ to denote $a \leq Kb$ for some constant $K > 0$; and $a \lesssim_p b$ to denote $a = O_p(b)$. For any time series of vectors $\{a_t\}_{t=1}^T$, we denote $\bar{a} = T^{-1} \sum_{t=1}^T a_t$. In addition, we write $\bar{a}_t = a_t - \bar{a}$. We use the capital letter A to denote the matrix $(a_1 : a_2 : \ldots : a_T)$, and write $\bar{A} = A - \iota_T^\intercal \bar{a}$ correspondingly. We use $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote the minimum and maximum eigenvalues of A. We use $\|A\|_1$, $\|A\|_\infty$, $\|A\|$, and $\|A\|_{\mathrm{F}}$ to denote the $\mathbb{L}_1$ norm, the $\mathbb{L}_\infty$ norm, the operator norm (or $\mathbb{L}_2$ norm), and the Frobenius norm of a matrix $A = (a_{ij})$, that is, $\max_j \sum_i |a_{ij}|$, $\max_i \sum_j |a_{ij}|$, $\sqrt{\lambda_{\max}(A^\intercal A)}$, and $\sqrt{\mathrm{Tr}(A^\intercal A)}$, respectively. We also use $\|A\|_{\mathrm{MAX}} = \max_{i,j} |a_{ij}|$ to denote the $\mathbb{L}_\infty$ norm of A on the vector space. When a is a vector, both $\|a\|$ and $\|a\|_{\mathrm{F}}$ are equal to its Euclidean norm. We use $\|a\|_0$ to denote $\sum_i 1_{\{a_i \neq 0\}}$. We also denote $Supp(a) = \{i : a_i \neq 0\}$. We write the projection operator with respect to a matrix A as $\mathbb{P}_A = A(A^\intercal A)^{-1} A^\intercal$, and the corresponding annihilator as $\mathbb{M}_A = \mathbb{I} - \mathbb{P}_A$, where $\mathbb{I}$ is the identity matrix whose size depends on the context. For a set of indices I, let $A[I]$ denote a sub-matrix of A, which contains all columns indexed in I.

B.2 Technical Assumptions

Assumption B.1 (Sparsity)**.** *$\|\lambda_h\|_0 \leq s$, $\|\chi_{j,\cdot}\|_0 \leq s$, $\|\eta_{j,\cdot}\|_0 \leq s$, $1 \leq j \leq d$, for some s such that $sn^{-1} \to 0$.*

Definition 1 (LASSO and Post-LASSO Estimators)**.** We consider a generic linear regression problem with sparse coefficients:

$$Y = X\beta + \varepsilon, \quad \text{subject to} \quad \|\beta\|_0 \leq s,$$

where Y is a $n \times 1$ vector, X is a $n \times p$ matrix, β is $p \times 1$ vector of parameters. We define the LASSO estimator as

$$\overline{\beta} = \arg\min_{\beta} \left\{ n^{-1} \|Y - X\beta\|^2 + n^{-1}\tau \|\beta\|_1 \right\}.$$

We define the Post-LASSO estimator $\widetilde{\beta}_{\widehat{I}}$ as

$$\widetilde{\beta}_{\widehat{I}} = \arg\min_{\beta} \left\{ n^{-1} \|Y - X\beta\|^2 : \quad \beta_j = 0, \quad j \notin \widehat{I} \right\},$$

where $\widehat{I}$ is the set of indices of variables selected by a first-step LASSO, that is, $\widehat{I} = Supp(\overline{\beta})$.

We adopt a high-level assumption on the model selection properties of LASSO and the prediction error bounds of the Post-LASSO estimators in (7) and (8). Belloni and Chernozhukov (2013) provide more primitive conditions for these bounds to hold.

Assumption B.2 (Properties of Post-LASSO Estimators). *The Post-LASSO estimators in (7) and (8) satisfy the following properties:*

1. $\widehat{s} = |\widehat{I}_1 \bigcup \widehat{I}_2| \lesssim_p s$.
2. *Moreover, if* $\tau_0 \geq 2c \left\| \lambda_g^\intercal C_e^\intercal (\iota_n : \widehat{C}_h) \right\|_1$, *for some* $c > 1$, *then*

$$n^{-1/2} \left\| \iota_n (\widetilde{\gamma}_{\widehat{I}_1} - \breve{\gamma}_0) + \widehat{C}_h (\widetilde{\lambda}_{\widehat{I}_1} - \breve{\lambda}_h) \right\| \lesssim_p sT^{-1/2} (\log(n \vee p \vee T))^{1/2} + \tau_0 s^{1/2} n^{-1}, \tag{B.1}$$

where $\breve{\gamma}_0 = \gamma_0 + \xi^\intercal \lambda_g$ *and* $\breve{\lambda}_h = \chi^\intercal \lambda_g + \lambda_h$ *are the true parameter values given in (2) and (6). If* $\tau_j \geq 2c_j \left\| e_j^\intercal C_e^\intercal (\iota_n : \widehat{C}_h) \right\|_1$, *for some* $c_j > 1$ *and* $j = 1, 2, \ldots, d$, *then*

$$n^{-1/2} \left\| \iota_n (\widetilde{\xi}_{\widehat{I}_2} - \xi)^\intercal + \widehat{C}_h (\widetilde{\chi}_{\widehat{I}_2} - \chi)^\intercal \right\| \lesssim_p sT^{-1/2} (\log(n \vee p \vee T))^{1/2} + \|\tau\|_{\text{MAX}} s^{1/2} n^{-1}, \tag{B.2}$$

where $\tau = (\tau_1, \tau_2, \ldots, \tau_d)^\intercal$, ξ *and* χ *are the true parameter values given in (6).*

Assumption B.2 gives a probabilistic upper bound on $\widehat{s}$. The prediction error bounds in (B.1) and (B.2) are more conservative than the standard results, because the regressors here are estimated. We provide a sketch of the proof for (B.1) in Appendix B.4, for which we need the following sparse eigenvalues assumption. The proof of (B.2) is similar and simpler. Our theoretical result below would also hold if other model selection procedures are employed, provided that they obey similar properties in Assumption B.2.

Assumption B.3 (Sparse Eigenvalues). *There exist* $K_1, K_2 > 0$ *and a sequence* $l_n \to \infty$, *such that with probability approaching 1,*

$$K_1 \leq \phi_{\min}(l_n s) \left[n^{-1} (\iota_n : \widehat{C}_h)^\intercal (\iota_n : \widehat{C}_h) \right] \leq \phi_{\max}(l_n s) \left[n^{-1} (\iota_n : \widehat{C}_h)^\intercal (\iota_n : \widehat{C}_h) \right] \leq K_2,$$

where we denote

$$\phi_{\min}(k)[A] = \min_{1 \leq \|v\|_0 \leq k} \frac{v^\intercal A v}{\|v\|^2}, \quad \text{and} \quad \phi_{\max}(k)[A] = \max_{1 \leq \|v\|_0 \leq k} \frac{v^\intercal A v}{\|v\|^2}.$$

Assumption B.3 resembles one of the sufficient conditions that lead to desirable statistical properties of LASSO, which has been adopted by, e.g., Belloni et al. (2014). It implies the restricted eigenvalue condition proposed by Bickel et al. (2009).

Assumption B.4 (Large Deviation Bounds). *The stochastic discount factor, the returns, and the factors satisfy*

$$\|\bar{a}\|_{\text{MAX}} \lesssim_p T^{-1/2} (\log(n \vee p \vee T))^{1/2}, \ \text{where } a \in \{m, v, z, u\}. \tag{B.3}$$

$$\left\| T^{-1} \bar{A} \bar{B}^\intercal - \text{Cov}(a_t, b_t) \right\|_{\text{MAX}} \lesssim_p T^{-1/2} (\log(n \vee p \vee T))^{1/2}, \ \text{where } A, B \in \{M, V, Z, U\}. \tag{B.4}$$

Assumption B.4 imposes high-level assumptions on the large deviation type bounds, which can be verified using the same arguments as in Fan et al. (2011) under stationarity, ergodicity, strong mixing, and exponential-type tail conditions.

Next, we impose additional uniform bounds that impose restrictions on the cross-sectional dependence of the "residuals" in the covariance projection (6). Similar assumptions on factor loadings are employed by Giglio and Xiu (2016).

For the sake of clarity and simplicity, we assume the set of testing assets used is not sampled randomly but deterministically, so that the covariances and loadings are treated as non-random. This is without loss of generality, because their sampling variation does not affect the first-order asymptotic inference. By contrast, Gagliardini et al. (2016) consider random loadings as a result of a random sampling scheme from a continuum of assets.

Assumption B.5 ("Moment" Conditions)**.** *The following restrictions hold:*

$$\|C_e\|_{\mathrm{MAX}} \lesssim 1, \quad \|C_e^\intercal \iota_n\|_{\mathrm{MAX}} \lesssim n^{1/2}, \quad \|C_e^\intercal C_h\|_{\mathrm{MAX}} \lesssim n^{1/2}, \tag{B.5}$$

$$\|C_e^\intercal \bar{u}\|_{\mathrm{MAX}} \lesssim_p n^{1/2}T^{-1/2}, \quad \|C_e^\intercal \bar{U}\bar{V}^\intercal\|_{\mathrm{MAX}} \lesssim_p n^{1/2}T^{1/2}, \tag{B.6}$$

$$\lambda_{\min}(n^{-1}C_e^\intercal C_e) \geq K, \quad \|C_e^\intercal(\beta_g\eta + \beta_h)\|_\infty \lesssim sn^{1/2}, \quad \|\beta_h\|_\infty \lesssim s. \tag{B.7}$$

In addition, for $a \in \{m, v, z, u\}$, it holds that

$$\|\Sigma_a\|_{\mathrm{MAX}} \lesssim 1, \quad \|C_a\|_{\mathrm{MAX}} \lesssim 1. \tag{B.8}$$

Finally, we impose a joint central limit theorem for $(z_t, \lambda^\intercal v_t z_t) = (z_t, (1-\gamma_0 m_t)z_t)$. This can be verified by the standard central limit theory for dependent stochastic processes, if more primitive assumptions are satisfied, see, e.g., White (2000).

Assumption B.6 (CLT)**.** *The following results hold as $T \to \infty$:*

$$T^{1/2}\begin{pmatrix} \bar{z} \\ -T^{-1}\gamma_0\bar{Z}\bar{M}^\intercal - \Sigma_z\lambda_g \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{12}^\intercal & \Pi_{22} \end{pmatrix}\right),$$

where Π_{11}, Π_{12}, and Π_{22} are given by

$$\Pi_{11} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \mathrm{E}(z_s z_t^\intercal),$$

$$\Pi_{12} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \mathrm{E}\left(\lambda^\intercal v_s z_s z_t^\intercal\right),$$

$$\Pi_{22} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \mathrm{E}\left(\lambda^\intercal v_s \lambda^\intercal v_t z_s z_t^\intercal\right).$$

Assumption B.7 (Selection for the Asymptotic Variance Estimator)**.** *The Post-LASSO estimator $\widetilde{\eta}_{\widetilde{I}}$ satisfies the usual bounds. That is, if $\bar{\tau}_j \geq 2\bar{c}_j \left\|HZ^\intercal\right\|_\infty$, for some $\bar{c}_j > 1$, $j = 1, 2, \ldots, d$, then we have*

$$\left\|(\widetilde{\eta}_{\widetilde{I}} - \eta)H\right\| \lesssim_{\mathrm{P}} s^{1/2}(\log(p \vee T))^{1/2}, \quad \textit{and} \quad \left\|\widetilde{\eta}_{\widetilde{I}} - \eta\right\| \lesssim_{\mathrm{P}} s^{1/2}T^{-1/2}(\log(p \vee T))^{1/2}.$$

B.3 Proof of Main Theorems

Proof of Theorem 1. The estimator of λ_g can be written in closed-form as

$$\widehat{\lambda}_g = \left(\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])} \widehat{C}_g\right)^{-1} \left(\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])} \bar{r}\right). \tag{B.9}$$

Moreover, by (2) and (5), we can relate C_g and C_h to β_g and β_h:

$$C_g = C_h\eta^\intercal + C_z, \quad \text{where} \quad C_h = (\beta_g\eta + \beta_h)\Sigma_h, \quad C_z = \beta_g\Sigma_z. \tag{B.10}$$

Using (3), (5), (B.10), and the fact that

$$\begin{aligned}
&\widehat{C}_g - C_g = (\widehat{C}_h - C_h)\eta^\intercal + (\widehat{C}_z - C_z),\\
&\widehat{C}_z - C_z = \beta_g\left(T^{-1}\bar{Z}\bar{Z}^\intercal - \Sigma_z\right) + T^{-1}\bar{U}\bar{Z}^\intercal + T^{-1}\left(\beta_g\eta + \beta_h\right)\bar{H}\bar{Z}^\intercal,\\
&\widehat{C}_h - C_h = (\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{H}^\intercal - \Sigma_h\right) + T^{-1}\bar{U}\bar{H}^\intercal + T^{-1}\beta_g\bar{Z}\bar{H}^\intercal,
\end{aligned}$$

we obtain the following decomposition:

$$\begin{aligned}
&T^{1/2}(\widehat{\lambda}_g - \lambda_g)\\
=&\left(n^{-1}\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_g\right)^{-1} n^{-1}T^{1/2}\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left((C_g - \widehat{C}_g)\lambda_g + C_h\lambda_h + \beta_g\bar{z} + ((\beta_g\eta + \beta_h)\bar{h} + \bar{u})\right)\\
=&T^{1/2}\Sigma_z^{-1}\left(\bar{z} - \left(T^{-1}\bar{Z}\bar{V}^\intercal\lambda - \Sigma_z\lambda_g\right)\right)\\
&+\left(n^{-1}\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_g\right)^{-1}\left(n^{-1}T^{1/2}\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u} - T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right.\\
&+n^{-1}T^{1/2}\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g - \widehat{C}_g\Sigma_z^{-1}) \times \left(\bar{z} - \left(T^{-1}\bar{Z}\bar{V}^\intercal\lambda - \Sigma_z\lambda_g\right)\right)\\
&-n^{-1}T^{1/2}\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\\
&\left.+n^{-1}T^{1/2}\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right).
\end{aligned}$$

We first analyze the leading term. Note that $\gamma_0\bar{M}^\intercal = -\bar{V}^\intercal\lambda$, by Assumption B.6 and applying the Delta method, we have

$$\begin{aligned}
&T^{1/2}\left(\Sigma_z^{-1}\bar{z} - \Sigma_z^{-1}\left(-T^{-1}\gamma_0\bar{Z}\bar{M}^\intercal - \Sigma_z\lambda_g\right)\right)\\
&\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \mathrm{E}\left((1 - \lambda^\intercal v_t)(1 - \lambda^\intercal v_s)\Sigma_z^{-1}z_t z_s^\intercal\Sigma_z^{-1}\right)\right).
\end{aligned} \tag{B.11}$$

Next, we show that the reminder terms are of a smaller order. By (B.42), we have

$$n^{-1}T^{1/2}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\| \lesssim_p s(n^{-1/2}+T^{-1/2})\log(n\vee p\vee T).$$

By (B.27), we have

$$n^{-1}T^{1/2}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \lesssim_p s^2(n^{-1}T^{1/2}+T^{-1/2})\log(n\vee p\vee T).$$

By (B.40), we have

$$\begin{aligned}&n^{-1}T^{1/2}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta+\beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda-\Sigma_h(\eta^\intercal\lambda_g+\lambda_h)-\bar{h}\right)\right\|\\ \lesssim_p & s^2(n^{-1/2}+T^{-1/2})\log(n\vee p\vee T).\end{aligned}$$

By Assumption B.4, (B.11), and (B.35), we have

$$\begin{aligned}&n^{-1}T^{1/2}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g-\widehat{C}_g\Sigma_z^{-1})\left(\bar{z}-\left(T^{-1}\bar{Z}\bar{V}^\intercal\lambda-\Sigma_z\lambda_g\right)\right)\right\|\\ \le & n^{-1}T^{1/2}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g-\widehat{C}_g\Sigma_z^{-1})\right\|\left\|\bar{z}-\left(T^{-1}\bar{Z}\bar{V}^\intercal\lambda-\Sigma_z\lambda_g\right)\right\|\\ \lesssim_p & s(n^{-1/2}+T^{-1/2})\log(n\vee p\vee T).\end{aligned}$$

This concludes the proof. □

Proof of Theorem 2. By the identical argument in the proof of Theorem 2 of Newey and West (1987), we have

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T}Q_{tr}(1-\lambda^\intercal v_t)(1-\lambda^\intercal v_r)\left(z_t z_r^\intercal+z_r z_t^\intercal\right)\xrightarrow{p}\Sigma_z\Pi\Sigma_z.$$

So applying the continuous mapping theorem, it is sufficient to show that

$$\widehat{\Sigma}_z\xrightarrow{p}\Sigma_z, \tag{B.12}$$

$$\widetilde{\Pi}-\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T}Q_{tr}(1-\lambda^\intercal v_t)(1-\lambda^\intercal v_r)\left(z_t z_r^\intercal+z_r z_t^\intercal\right)\xrightarrow{p}0, \tag{B.13}$$

where

$$Q_{tr}=\left(1-\frac{|r-t|}{q+1}\right)1_{\{|t-r|\le q\}},\quad \widetilde{\Pi}=\widehat{\Sigma}_z\widehat{\Pi}\widehat{\Sigma}_z.$$

To prove (B.12), we note that by Assumptions B.4 and B.7, we have

$$\begin{aligned}&\left\|\widehat{\Sigma}_z-\Sigma_z\right\|_{\text{MAX}}\\ \lesssim & T^{-1/2}\left\|(\widetilde{\eta}_{\widetilde{I}}-\eta)H\right\|\|Z\|_{\text{MAX}}+T^{-1}\left\|(\widetilde{\eta}_{\widetilde{I}}-\eta)H\right\|^2+\left\|T^{-1}ZZ^\intercal-\Sigma_z\right\|_{\text{MAX}}\\ \lesssim_p & s^{1/2}T^{-1/2}(\log(p\vee T))^{1/2}\|Z\|_{\text{MAX}}+sT^{-1}\log(p\vee T)+T^{-1/2}(\log(n\vee p\vee T))^{1/2}\end{aligned}$$

$$=o_p(1). \tag{B.14}$$

As to (B.13), we can decompose its left-hand side as

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T} Q_{tr}(\widehat{\lambda}-\lambda)^\intercal v_t(1-\widehat{\lambda}^\intercal v_r)\left(\widehat{z}_t\widehat{z}_r^\intercal+\widehat{z}_r\widehat{z}_t^\intercal\right) \tag{B.15}$$

$$+\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T} Q_{tr}(1-\lambda^\intercal v_t)(\widehat{\lambda}-\lambda)^\intercal v_r\left(\widehat{z}_t\widehat{z}_r^\intercal+\widehat{z}_r\widehat{z}_t^\intercal\right) \tag{B.16}$$

$$+\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T} Q_{tr}(1-\lambda^\intercal v_t)(1-\lambda^\intercal v_r)\left(\left(\widehat{z}_t-z_t\right)\widehat{z}_r^\intercal+\left(\widehat{z}_r-z_r\right)\widehat{z}_t^\intercal\right) \tag{B.17}$$

$$+\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T} Q_{tr}(1-\lambda^\intercal v_t)(1-\lambda^\intercal v_r)\left(z_t\left(\widehat{z}_r-z_r\right)^\intercal+z_r\left(\widehat{z}_t-z_t\right)^\intercal\right). \tag{B.18}$$

Analyzing each of these terms, we can obtain that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T} Q_{tr}(\widehat{\lambda}-\lambda)^\intercal v_t(1-\widehat{\lambda}^\intercal v_r)\left(\widehat{z}_t\widehat{z}_r^\intercal+\widehat{z}_r\widehat{z}_t^\intercal\right)\right\|_{\mathrm{MAX}}$$
$$\lesssim qT^{-1}\left\|\widehat{Z}\right\|\left\|\iota_T^\intercal-\widehat{\lambda}^\intercal V\right\|\left\|(\widehat{\lambda}-\lambda)^\intercal V\right\|_{\mathrm{MAX}}\left\|\widehat{Z}\right\|_{\mathrm{MAX}}\lesssim_p qs^{1/2}(T^{-1/2}+n^{-1/2})\left\|V\right\|_{\mathrm{MAX}}\left\|Z\right\|_{\mathrm{MAX}},$$
$$\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T} Q_{tr}(1-\lambda^\intercal v_t)(\widehat{\lambda}-\lambda)^\intercal v_r\left(\widehat{z}_t\widehat{z}_r^\intercal+\widehat{z}_r\widehat{z}_t^\intercal\right)\right\|_{\mathrm{MAX}}$$
$$\lesssim qT^{-1}\left\|\iota_T^\intercal-\lambda^\intercal V\right\|\left\|\widehat{Z}\right\|\left\|(\widehat{\lambda}-\lambda)^\intercal V\right\|_{\mathrm{MAX}}\left\|\widehat{Z}\right\|_{\mathrm{MAX}}\lesssim_p qs^{1/2}(T^{-1/2}+n^{-1/2})\left\|V\right\|_{\mathrm{MAX}}\left\|Z\right\|_{\mathrm{MAX}},$$
$$\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{T} Q_{tr}(1-\lambda^\intercal v_t)(1-\lambda^\intercal v_r)\left(\left(\widehat{z}_t-z_t\right)\widehat{z}_r^\intercal+\left(\widehat{z}_r-z_r\right)\widehat{z}_t^\intercal\right)\right\|_{\mathrm{MAX}}$$
$$\lesssim qT^{-1}\left\|\iota_T^\intercal-\lambda^\intercal V\right\|\left\|(\widehat{\eta}-\eta)H\right\|\left\|\widehat{Z}\right\|_{\mathrm{MAX}}\left\|\iota_T^\intercal-\lambda^\intercal V\right\|_{\mathrm{MAX}}$$
$$\lesssim_p qs^{3/2}(T^{-1/2}+n^{-1/2})\left\|V\right\|_{\mathrm{MAX}}\left\|Z\right\|_{\mathrm{MAX}},$$

where we use

$$\left\|\iota_T^\intercal-\lambda^\intercal V\right\|\lesssim T^{1/2}+\left\|\bar{M}\right\|+\left\|\lambda^\intercal\bar{v}\right\|\lesssim_p T^{1/2},$$
$$\left\|\iota_T^\intercal-\lambda^\intercal V\right\|_{\mathrm{MAX}}\lesssim 1+\left\|\lambda^\intercal V\right\|_{\mathrm{MAX}}\lesssim s\left\|V\right\|_{\mathrm{MAX}},$$
$$\left\|\iota_T^\intercal-\widehat{\lambda}^\intercal V\right\|\leq\left\|\iota_T^\intercal-\lambda^\intercal V\right\|+\left\|(\widehat{\lambda}-\lambda)^\intercal V\right\|\lesssim_p T^{1/2}+\left\|\widehat{\lambda}-\lambda\right\|\left\|V\right\|\lesssim_p T^{1/2},$$
$$\left\|\widehat{Z}\right\|\lesssim T^{1/2}\left\|\widehat{\Sigma}_z\right\|^{1/2}\lesssim_p T^{1/2}\left\|\Sigma_z\right\|^{1/2}\lesssim T^{1/2},$$
$$\left\|(\widehat{\lambda}-\lambda)^\intercal V\right\|_{\mathrm{MAX}}\leq\left\|\widehat{\lambda}-\lambda\right\|_{\infty}\left\|V\right\|_{\mathrm{MAX}}\leq\left\|\widehat{\lambda}-\lambda\right\|\left\|V\right\|_{\mathrm{MAX}}\lesssim_p s^{1/2}(T^{-1/2}+n^{-1/2})\left\|V\right\|_{\mathrm{MAX}},$$
$$\left\|\widehat{Z}\right\|_{\mathrm{MAX}}\leq\left\|(\widehat{\eta}-\eta)H\right\|+\left\|Z\right\|_{\mathrm{MAX}}\lesssim_p\left\|Z\right\|_{\mathrm{MAX}},$$

which hold by (B.14), Assumption B.4, and Lemma 7. This concludes the proof. □

B.4 Proof of Lemmas

Proof of (B.1). We provide a sketch of the proof, as it is very similar to Belloni and Chernozhukov (2013). With respect to the optimization problem (7), we define

$$Q(\gamma,\lambda)=n^{-1}\left\|\bar r-\iota_n\gamma-\widehat C_h\lambda\right\|^2.$$

We denote the solution to this problem as $\widetilde\gamma$ and $\widetilde\lambda$. Let $\delta=\widetilde\lambda-\breve\lambda_h$. Note by (5) and (2), we have

$$\mathrm{E}(r_t)=\iota_n\breve\gamma_0+C_h\breve\lambda_h+C_e\lambda_g,\quad\text{and}\quad \bar r=\mathrm{E}(r_t)+\beta_g\bar g+\beta_h\bar h+\bar u.$$

By direct calculations, we have

$$\begin{aligned}
&Q(\widetilde\gamma,\widetilde\lambda)-Q(\breve\gamma_0,\breve\lambda_h)-n^{-1}\left\|\iota_n(\widetilde\gamma-\breve\gamma_0)+\widehat C_h\delta\right\|^2\\
=&-2n^{-1}\left(\bar r-\iota_n\breve\gamma_0-\widehat C_h\breve\lambda_h\right)^\intercal\left(\iota_n(\widetilde\gamma-\breve\gamma_0)+\widehat C_h\delta\right)\\
=&-2n^{-1}\left(\beta_g\bar g+\beta_h\bar h+\bar u+(C_h-\widehat C_h)\breve\lambda_h+C_e\lambda_g\right)^\intercal\left(\iota_n(\widetilde\gamma-\breve\gamma_0)+\widehat C_h\delta\right)\\
\geq&-2n^{-1}\left\|\beta_g\bar g+\beta_h\bar h+\bar u+(C_h-\widehat C_h)\breve\lambda_h\right\|\left\|\iota_n(\widetilde\gamma-\breve\gamma_0)+\widehat C_h\delta\right\|\\
&-2n^{-1}\left\|(C_e\lambda_g)^\intercal(\iota_n:\widehat C_h)\right\|_1\|(\widetilde\gamma-\breve\gamma_0:\delta^\intercal)^\intercal\|_1\\
\geq&-2n^{-1}\left\|\beta_g\bar g+\beta_h\bar h+\bar u+(C_h-\widehat C_h)\breve\lambda_h\right\|\left\|\iota_n(\widetilde\gamma-\breve\gamma_0)+\widehat C_h\delta\right\|\\
&-\tau_0K^{-1}n^{-1}\left(|\widetilde\gamma-\breve\gamma_0|+\|\delta_I\|_1+\|\delta_{I^c}\|_1\right),
\end{aligned}$$

where I is the set of non-zeros in $\breve\lambda_h$, I^c is its complement, and δ_I is a sub-vector of δ with all entries taken from I.

On the other hand, by definition of $\widetilde\gamma$ and $\widetilde\lambda$, we have

$$\begin{aligned}
Q(\widetilde\gamma,\widetilde\lambda)-Q(\breve\gamma_0,\breve\lambda_h)\leq&\tau_0n^{-1}\left(\left\|(\breve\gamma_0:\breve\lambda_h^\intercal)^\intercal\right\|_1-\left\|(\widetilde\gamma:\widetilde\lambda^\intercal)^\intercal\right\|_1\right)\\
\leq&\tau_0n^{-1}(|\widetilde\gamma-\breve\gamma_0|+\|\delta_I\|_1-\|\delta_{I^c}\|_1).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
n^{-1}\left\|\iota_n(\widetilde\gamma-\breve\gamma_0)+\widehat C_h\delta\right\|^2&-\tau_0c^{-1}n^{-1}\left(|\widetilde\gamma-\breve\gamma_0|+\|\delta_I\|_1+\|\delta_{I^c}\|_1\right)\\
&-2n^{-1}\left\|\beta_g\bar g+\beta_h\bar h+\bar u+(C_h-\widehat C_h)\breve\lambda_h\right\|\left\|\iota_n(\widetilde\gamma-\breve\gamma_0)+\widehat C_h\delta\right\|\\
\leq&\tau_0n^{-1}(|\widetilde\gamma-\breve\gamma_0|+\|\delta_I\|_1-\|\delta_{I^c}\|_1),
\end{aligned}\tag{B.19}$$

where we use the fact that

$$\tau_0\geq 2c\left\|\lambda_g^\intercal C_e^\intercal(\iota_n:\widehat C_h)\right\|_1.$$

If it holds that

$$n^{-1}\left\|\iota_n(\widetilde{\gamma}-\check{\gamma}_0)+\widehat{C}_h\delta\right\|-2n^{-1}\left\|\beta_g\bar{g}+\beta_h\bar{h}+\bar{u}+(C_h-\widehat{C}_h)\check{\lambda}_h\right\|<0,$$

we can establish that

$$n^{-1/2}\left\|\iota_n(\widetilde{\gamma}-\check{\gamma}_0)+\widehat{C}_h\delta\right\|\lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2},$$

where we use the fact that

$$n^{-1/2}\left\|\beta_g\bar{g}\right\|\lesssim\left\|\beta_g\right\|_{\text{MAX}}\left\|\bar{g}\right\|_{\text{MAX}}\lesssim_p T^{-1/2}, \tag{B.20}$$

$$n^{-1/2}\left\|\bar{u}\right\|\lesssim\left\|\bar{u}\right\|_{\text{MAX}}\lesssim_p T^{-1/2}(\log(n\vee p\vee T))^{1/2}, \tag{B.21}$$

$$n^{-1/2}\left\|\beta_h\bar{h}\right\|\leq\left\|\beta_h\right\|_{\infty}\left\|\bar{h}\right\|_{\text{MAX}}\lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2}, \tag{B.22}$$

$$n^{-1/2}\left\|(C_h-\widehat{C}_h)\check{\lambda}_h\right\|\lesssim\left\|C_h-\widehat{C}_h\right\|_{\text{MAX}}\left\|\check{\lambda}_h\right\|_1\lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2}. \tag{B.23}$$

Otherwise, from (B.19) it follows that

$$-c^{-1}\left(|\widetilde{\gamma}-\check{\gamma}_0|+\left\|\delta_I\right\|_1+\left\|\delta_{I^c}\right\|_1\right)\leq|\widetilde{\gamma}-\check{\gamma}_0|+\left\|\delta_I\right\|_1-\left\|\delta_{I^c}\right\|_1,$$

which leads to, writing $\bar{c}=(c+1)(c-1)^{-1}$,

$$\left\|\delta_{I^c}\right\|\leq\bar{c}(|\widetilde{\gamma}-\check{\gamma}_0|+\left\|\delta_I\right\|_1).$$

Then by (B.19) again as well as the restricted eigenvalue condition in Belloni and Chernozhukov (2013), we obtain

$$\begin{aligned}&\left\|\iota_n(\widetilde{\gamma}-\check{\gamma}_0)+\widehat{C}_h\delta\right\|^2-2\left\|\beta_g\bar{g}+\beta_h\bar{h}+\bar{u}+(C_h-\widehat{C}_h)\check{\lambda}_h\right\|\left\|\iota_n(\widetilde{\gamma}-\check{\gamma}_0)+\widehat{C}_h\delta\right\|\\&\qquad\leq(1+c^{-1})\tau_0(|\widetilde{\gamma}-\check{\gamma}_0|+\left\|\delta_I\right\|_1)\lesssim\tau_0 s^{1/2}n^{-1/2}\left\|\iota_n(\widetilde{\gamma}-\check{\gamma}_0)+\widehat{C}_h\delta\right\|.\end{aligned}$$

Therefore, we have

$$\begin{aligned}n^{-1/2}\left\|\iota_n(\widetilde{\gamma}-\check{\gamma}_0)+\widehat{C}_h\delta\right\|&\lesssim n^{-1/2}\left\|\beta_g\bar{g}+\beta_h\bar{h}+\bar{u}+(C_h-\widehat{C}_h)\check{\lambda}_h\right\|+\tau_0 s^{1/2}n^{-1}\\&\lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2}+\tau_0 s^{1/2}n^{-1}.\end{aligned}$$

The Post-LASSO estimator converges at the same rate following the same arguments as in Belloni and Chernozhukov (2013).

□

Lemma 1. *Under Assumptions B.1, B.2, B.4, B.5, we have*

$$n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^{\intercal}\right\|\lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \tag{B.24}$$

$$n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\|\lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \tag{B.25}$$

Proof of Lemma 1. Using the fact that $\widehat{I}_2 \subseteq \widehat{I}$ and by (B.2), we have

$$\begin{aligned} n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\right\| =& n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\widehat{C}_h\chi^\intercal+\iota_n\xi^\intercal)\right\| \leq n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_2])}(\widehat{C}_h\chi^\intercal+\iota_n\xi^\intercal))\right\| \\ \leq& n^{-1/2}\left\|\iota_n(\xi-\widetilde{\xi}_{\widehat{I}_2})^\intercal+\widehat{C}_h\chi^\intercal-\widehat{C}_h\widetilde{\chi}_{\widehat{I}_2}^\intercal\right\| \\ \lesssim_p& sT^{-1/2}(\log(n\vee p\vee T))^{1/2}+\|\tau\|_{\text{MAX}}\, s^{1/2}n^{-1}. \end{aligned}$$

Since by Assumptions B.4 and B.5, our choice of τ satisfies:

$$\begin{aligned} n^{-1}\|\tau\|_{\text{MAX}} \lesssim& n^{-1}\max_{1\leq j\leq d}\left\|e_j^\intercal C_e^\intercal\widehat{C}_h\right\|_1 \lesssim n^{-1}\|C_e^\intercal C_h\|_{\text{MAX}}+n^{-1}\left\|C_e^\intercal(\widehat{C}_h-C_h)\right\|_{\text{MAX}} \\ \lesssim_p& (n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \end{aligned} \tag{B.26}$$

This concludes the proof of (B.24).

Similarly, to prove (B.25), by (B.1) we have

$$\begin{aligned} & n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_1])}\left(\widehat{C}_h\check{\lambda}_h+\iota_n\check{\gamma}_0\right)\right\| \\ \leq& n^{-1/2}\left\|(\iota_n:\widehat{C}_h)(\widetilde{\gamma}_{\widehat{I}_1}-\check{\gamma}_0:(\widetilde{\lambda}_{\widehat{I}_1}-\check{\lambda}_h)^\intercal)^\intercal\right\| \lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2}+\tau_0 s^{1/2}n^{-1}. \end{aligned}$$

Because we can select τ_0 that satisfies

$$\begin{aligned} n^{-1}\tau_0 \leq& n^{-1}\left\|\lambda_g^\intercal C_e^\intercal(\iota_n:\widehat{C}_h)\right\|_1 \leq n^{-1}\,|\lambda_g^\intercal C_e^\intercal\iota_n|+n^{-1}\left\|\lambda_g^\intercal C_e^\intercal\widehat{C}_h\right\|_{\text{MAX}} \\ \lesssim& n^{-1}\|C_e\iota_n\|_{\text{MAX}}+\|C_e\|_{\text{MAX}}\left\|\widehat{C}_h-C_h\right\|_{\text{MAX}}+n^{-1}\|C_e^\intercal C_h\|_{\text{MAX}} \\ \lesssim_p& (n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}, \end{aligned}$$

hence it follows that

$$n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_1])}\left(\widehat{C}_h(\lambda_h+\chi^\intercal\lambda_g)+\iota_n\gamma_0\right)\right\| \lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}.$$

By the triangle inequality and $\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_1])}\iota_n=0$, we have

$$\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_1])}\widehat{C}_h\lambda_h\right\| \leq \left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_1])}\left(\widehat{C}_h(\lambda_h+\chi^\intercal\lambda_g)+\iota_n\gamma_0\right)\right\|+\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_1])}\widehat{C}_h\chi^\intercal\right\|\|\lambda_g\|,$$

which, combined with (B.24) and $\|\lambda_g\| \lesssim 1$, lead to the conclusion. □

Lemma 2. *Under Assumptions B.1, B.2, B.3, B.4, B.5, we have*

$$n^{-1}\left\|\widehat{C}_g^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \lesssim_p s^2(n^{-1}+T^{-1})\log(n\vee p\vee T). \tag{B.27}$$

Proof of Lemma 2 . We note by (6) that

$$\widehat{C}_g=\widehat{C}_h\chi^\intercal+\widehat{C}_g-C_g+\iota_n\xi^\intercal+(C_h-\widehat{C}_h)\chi^\intercal+C_e, \tag{B.28}$$

thereby it follows

$$n^{-1}\left\|\widehat{C}_g^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \leq n^{-1}\left\|\chi\widehat{C}_h^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| + n^{-1}\left\|C_e^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| + n^{-1}\left\|(\widehat{C}_g - C_g + (C_h - \widehat{C}_h)\chi^{\intercal})^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\|.$$

On the one hand, by Lemma 1, we have

$$n^{-1}\left\|\chi\widehat{C}_h^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \leq n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^{\intercal}\right\| n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \lesssim_p s^2(n^{-1}+T^{-1})\log(n\vee p\vee T). \tag{B.29}$$

On the other hand, note that

$$\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h = (\iota_n\gamma_0 + \widehat{C}_h\lambda_h) - (\iota_n:\widehat{C}_h)(\hat{\gamma}_0:\hat{\lambda}_h^{\intercal})^{\intercal} = (\iota_n:\widehat{C}_h)(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})^{\intercal},$$

where $(\hat{\gamma}_0:\hat{\lambda}_h^{\intercal})^{\intercal} = \arg\min_{\gamma,\lambda}\{\iota_n\gamma_0 + \widehat{C}_h\lambda_h - \iota_n\gamma - \widehat{C}_h\lambda : \lambda_j = 0, j\in\widehat{I}^c\}$. By Assumption B.3, we have

$$\begin{aligned} n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| &= n^{-1/2}\left\|(\iota_n:\widehat{C}_h)(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})^{\intercal}\right\| \\ &\geq \phi_{\min}^{1/2}(s+\widehat{s}+1)\left[n^{-1}(\iota_n:\widehat{C}_h)^{\intercal}(\iota_n:\widehat{C}_h)\right]\left\|(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})\right\| \\ &\gtrsim \left\|(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})\right\|, \end{aligned}$$

hence it follows from (B.25) that

$$\left\|(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})\right\| \lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \tag{B.30}$$

Using this, we have

$$\begin{aligned} n^{-1}\left\|C_e^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| &= n^{-1}\left\|C_e^{\intercal}(\iota_n:\widehat{C}_h)(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})^{\intercal}\right\| \\ &\lesssim n^{-1}\left\|C_e^{\intercal}(\iota_n:\widehat{C}_h)\right\|_{\text{MAX}}\left\|(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})^{\intercal}\right\|_1. \end{aligned} \tag{B.31}$$

Using (B.5) and Assumption B.4, it follows that

$$\begin{aligned} n^{-1}\left\|C_e^{\intercal}(\iota_n:\widehat{C}_h)\right\|_{\text{MAX}} &\leq n^{-1}\left\|C_e^{\intercal}(\widehat{C}_h-C_h)\right\|_{\text{MAX}} + n^{-1}\left\|C_e^{\intercal}C_h\right\|_{\text{MAX}} + n^{-1}\left\|C_e^{\intercal}\iota_n\right\|_{\text{MAX}} \\ &\lesssim \left\|C_e\right\|_{\text{MAX}}\left\|\widehat{C}_h-C_h\right\|_{\text{MAX}} + n^{-1}\left\|C_e^{\intercal}C_h\right\|_{\text{MAX}} + n^{-1}\left\|C_e^{\intercal}\iota_n\right\|_{\text{MAX}} \\ &\lesssim_p (n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \end{aligned} \tag{B.32}$$

Moreover, since by sparsity of λ_h and $\hat{\lambda}_h$, we have

$$\left\|(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})^{\intercal}\right\|_1 \leq (s+\widehat{s}+1)^{1/2}\left\|(\gamma_0-\hat{\gamma}_0:\lambda_h^{\intercal}-\hat{\lambda}_h^{\intercal})^{\intercal}\right\|.$$

Combining (B.30), (B.31), and (B.32), we obtain

$$n^{-1}\left\|C_e^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \lesssim_p s^{3/2}(n^{-1}+T^{-1})\log(n\vee p\vee T). \tag{B.33}$$

Finally, by (B.25) we have

$$\begin{aligned} & n^{-1}\left\|(\widehat{C}_g - C_g + (C_h-\widehat{C}_h)\chi^{\intercal})^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \\ \lesssim & \left\|\widehat{C}_g - C_g + (C_h-\widehat{C}_h)\chi^{\intercal}\right\|_{\text{MAX}} n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\lambda_h\right\| \\ \lesssim_p & s^2(n^{-1/2}T^{-1/2}+T^{-1})(\log(n\vee p\vee T))^{1/2}. \end{aligned}$$

The above estimate, along with (B.33) and (B.29), conclude the proof of (B.27).

□

Lemma 3. *Under Assumptions B.1, B.2, B.3, B.4, B.5, we have*

$$n^{-1}\left\|\widehat{C}_g^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\| \lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \tag{B.34}$$

$$n^{-1}\left\|\widehat{C}_g^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g - \widehat{C}_g\Sigma_z^{-1})\right\| \lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \tag{B.35}$$

Proof of Lemma 3. (i) By (6), we have

$$\begin{aligned} n^{-1}\left\|\widehat{C}_g^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\| \leq & n^{-1}\left\|C_e^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\| + n^{-1}\left\|\chi\widehat{C}_h^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\| \\ & + n^{-1}\left\|\left((\widehat{C}_g - C_g)^{\intercal}+\chi(C_h-\widehat{C}_h)^{\intercal}\right)\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\|. \end{aligned}$$

Moreover, by (B.24), we obtain

$$\begin{aligned} n^{-1}\left\|\chi\widehat{C}_h^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\| \leq & n^{-1/2}\left\|\chi\widehat{C}_h^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\right\| n^{-1/2}\left\|C_h\eta^{\intercal}\right\| \\ \lesssim_p & s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}, \end{aligned} \tag{B.36}$$

where we use the fact that $C_g = C_h\eta^{\intercal} + C_z$, and that

$$n^{-1/2}\left\|C_h\eta^{\intercal}\right\| \lesssim \left\|C_h\eta^{\intercal}\right\|_{\text{MAX}} \lesssim \left\|C_g\right\|_{\text{MAX}} + \left\|C_z\right\|_{\text{MAX}} \lesssim 1.$$

In addition, we have

$$n^{-1}\left\|C_e^{\intercal}\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\| \leq n^{-1}\left\|C_e^{\intercal}C_h\eta^{\intercal}\right\| + n^{-1}\left\|C_e^{\intercal}\mathbb{P}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\|.$$

To bound the first term, we have

$$n^{-1}\left\|C_e^{\intercal}C_h\eta^{\intercal}\right\| \lesssim n^{-1}\left\|C_e^{\intercal}C_h\right\|_{\text{MAX}}\left\|\eta\right\|_{\infty} \lesssim_p sn^{-1/2}(\log(n\vee p\vee T))^{1/2}.$$

As to the second term, using (B.32) we obtain

$$n^{-1}\left\|C_e^{\intercal}\mathbb{P}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_h\eta^{\intercal}\right\|$$

$$
\begin{aligned}
&= n^{-1}\left\| C_e^\intercal(\iota_n : \widehat{C}_h[\widehat{I}])\left((\iota_n : \widehat{C}_h[\widehat{I}])^\intercal(\iota_n : \widehat{C}_h[\widehat{I}])\right)^{-1}(\iota_n : \widehat{C}_h[\widehat{I}])^\intercal C_h\eta^\intercal\right\| \\
&\le n^{-1}\left\| C_e^\intercal(\iota_n : \widehat{C}_h[\widehat{I}])\right\| \left\|\left((\iota_n : \widehat{C}_h[\widehat{I}])^\intercal(\iota_n : \widehat{C}_h[\widehat{I}])\right)^{-1}\right\| \left\|(\iota_n : \widehat{C}_h[\widehat{I}])^\intercal C_h\eta^\intercal\right\| \\
&\lesssim (1+\widehat{s})\phi_{\min}^{-1}(\widehat{s}+1)\left[n^{-1}(\iota_n : \widehat{C}_h)^\intercal(\iota_n : \widehat{C}_h)\right] n^{-1}\left\| C_e^\intercal(\iota_n : \widehat{C}_h[\widehat{I}])\right\|_{\text{MAX}} n^{-1}\left\|(\iota_n : \widehat{C}_h[\widehat{I}])^\intercal C_h\eta^\intercal\right\|_{\text{MAX}} \\
&\lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2},
\end{aligned}
$$

where we also use $\|C_h\eta\|_{\text{MAX}} \le \|C_g\|_{\text{MAX}} + \|C_z\|_{\text{MAX}} \lesssim 1$, and

$$
\begin{aligned}
n^{-1}\left\|(\iota_n : \widehat{C}_h[\widehat{I}])^\intercal C_h\eta\right\|_{\text{MAX}} \le & n^{-1}\left\|(\iota_n : \widehat{C}_h)^\intercal C_h\eta\right\|_{\text{MAX}} \lesssim \left\|(\iota_n : \widehat{C}_h)\right\|_{\text{MAX}} \|C_h\eta\|_{\text{MAX}} \\
&\lesssim \left(\|(\iota_n : C_h)\|_{\text{MAX}} + \left\|\widehat{C}_h - C_h\right\|_{\text{MAX}}\right)\|C_h\eta\|_{\text{MAX}} \lesssim_p 1.
\end{aligned}
$$

Therefore, we have

$$
n^{-1}\left\| C_e^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])} C_h\eta^\intercal\right\| \lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}. \tag{B.37}
$$

Similarly, because we have

$$
\begin{aligned}
& n^{-1}\left\|\left((\widehat{C}_g - C_g)^\intercal + \chi(C_h - \widehat{C}_h)^\intercal\right) C_h\eta^\intercal\right\| \\
\lesssim & \left\|(\widehat{C}_g - C_g)^\intercal + \chi(C_h - \widehat{C}_h)^\intercal\right\|_{\text{MAX}} \|C_h\eta^\intercal\|_{\text{MAX}} \lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2}. \\
& n^{-1}\left\|\left((\widehat{C}_g - C_g)^\intercal + \chi(C_h - \widehat{C}_h)^\intercal\right)(\iota_n : \widehat{C}_h[\widehat{I}])\right\|_{\text{MAX}} \\
\le & K\left\|(\widehat{C}_g - C_g)^\intercal + \chi(C_h - \widehat{C}_h)^\intercal\right\|_{\text{MAX}} \left\|(\iota_n : \widehat{C}_h[\widehat{I}])\right\|_{\text{MAX}} \lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2},
\end{aligned}
$$

it follows that

$$
n^{-1}\left\|\left((\widehat{C}_g - C_g)^\intercal + \chi(C_h - \widehat{C}_h)^\intercal\right)\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])} C_h\eta^\intercal\right\| \lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2},
$$

which, along with (B.36) and (B.37), establish the first claim.

(ii) Next, by (5) we have

$$
\widehat{C}_g = \widehat{C}_h\eta^\intercal + \widehat{C}_z.
$$

And recall that $\beta_g = C_z\Sigma_z^{-1}$, so we have

$$
\begin{aligned}
& n^{-1}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g - \widehat{C}_g\Sigma_z^{-1})\right\| \\
\le & n^{-1}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(C_z - \widehat{C}_z)\Sigma_z^{-1}\right\| + n^{-1}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\widehat{C}_h - C_h)\eta^\intercal\Sigma_z^{-1}\right\| \\
& + n^{-1}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])} C_h\eta^\intercal\Sigma_z^{-1}\right\|.
\end{aligned}
$$

Using Assumption B.4 and $\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\right\| \le 1$, we have

$$
n^{-1}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(C_z - \widehat{C}_z)\Sigma_z^{-1}\right\|
$$

$$\lesssim \left\|\widehat{C}_g\right\|_{\text{MAX}} \left\|C_z - \widehat{C}_z\right\|_{\text{MAX}} \|\Sigma_z^{-1}\| \lesssim_p T^{-1/2}(\log(n \vee p \vee T))^{1/2}, \tag{B.38}$$

where we also use the fact that

$$\|\Sigma_z^{-1}\| \le \lambda_{\min}^{-1}(\Sigma_z) \lesssim 1, \quad \left\|\widehat{C}_g\right\|_{\text{MAX}} \le \left\|\widehat{C}_g - C_g\right\|_{\text{MAX}} + \|C_g\|_{\text{MAX}} \lesssim 1.$$

Similarly, we obtain

$$\begin{aligned} n^{-1}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\widehat{C}_h - C_h)\eta^\intercal \Sigma_z^{-1}\right\| &\lesssim \left\|\widehat{C}_g\right\|_{\text{MAX}} \left\|\widehat{C}_h - C_h\right\|_{\text{MAX}} \|\eta\|_\infty \|\Sigma_z^{-1}\| \\ &\lesssim_p sT^{-1/2}(\log(n \vee p \vee T))^{1/2}. \end{aligned} \tag{B.39}$$

Combining (B.38), (B.39), and (B.34) concludes the proof.

□

Lemma 4. *Under Assumptions B.1, B.2, B.3, B.4, B.5, we have*

$$\begin{aligned} &n^{-1}\left\|\widehat{C}_g^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\| \\ \lesssim_p &s^2(n^{-1/2}T^{-1/2} + T^{-1})\log(n \vee p \vee T). \end{aligned} \tag{B.40}$$

Proof of Lemma 4. From (B.24) and Assumption B.4, it follows that

$$\begin{aligned} &n^{-1}\left\|\chi\widehat{C}_h^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\| \\ \le &n^{-1/2}\left\|\chi\widehat{C}_h^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\right\| \|\beta_g\eta + \beta_h\|_\infty \left(\left\|T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h)\right\|_{\text{MAX}} + \left\|\bar{h}\right\|_{\text{MAX}}\right) \\ \lesssim_p &s^2(n^{-1/2}T^{-1/2} + T^{-1})\log(n \vee p \vee T). \end{aligned} \tag{B.41}$$

Next, by triangle inequality, we have

$$\begin{aligned} &n^{-1}\left\|C_e^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{Z}^\intercal\lambda_g + \left(T^{-1}\bar{H}\bar{H}^\intercal - \Sigma_h\right)(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\| \\ \le &n^{-1}\left\|C_e^\intercal(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\| \\ &+ n^{-1}\left\|C_e^\intercal \mathbb{P}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\|. \end{aligned}$$

For the first term, by Assumption B.5 we have

$$\begin{aligned} &n^{-1}\left\|C_e^\intercal(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\| \\ \le &n^{-1}\left\|C_e^\intercal(\beta_g\eta + \beta_h)\right\|_\infty \left\|\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\|_{\text{MAX}} \\ \lesssim_p &sn^{-1/2}T^{-1/2}(\log(n \vee p \vee T))^{1/2}. \end{aligned}$$

For the second term, we use Assumptions B.1, B.3, B.4, and (B.32),

$$n^{-1}\left\|C_e^\intercal \mathbb{P}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta + \beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda - \Sigma_h(\eta^\intercal\lambda_g + \lambda_h) - \bar{h}\right)\right\|$$

$$
\begin{aligned}
\lesssim &(1+\widehat{s})\phi_{\min}^{-1}(\widehat{s}+1)\left[n^{-1}(\iota_n:\widehat{C}_h)^\intercal(\iota_n:\widehat{C}_h)\right]n^{-1}\left\|C_e^\intercal(\iota_n:\widehat{C}_h[\widehat{I}])\right\|_{\mathrm{MAX}}\\
&\times\left\|(\iota_n:\widehat{C}_h[\widehat{I}])^\intercal\right\|_{\mathrm{MAX}}\|\beta_g\eta+\beta_h\|_\infty\left\|T^{-1}\bar{H}\bar{V}^\intercal\lambda-\Sigma_h(\eta^\intercal\lambda_g+\lambda_h)-\bar{h}\right\|_{\mathrm{MAX}}\\
\lesssim_p &s^2(n^{-1/2}T^{-1/2}+T^{-1})\log(n\vee p\vee T).
\end{aligned}
$$

Finally, by Assumptions B.1 and B.4, we have

$$
\begin{aligned}
&n^{-1}\left\|(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal)^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}(\beta_g\eta+\beta_h)\left(T^{-1}\bar{H}\bar{V}^\intercal\lambda-\Sigma_h(\eta^\intercal\lambda_g+\lambda_h)-\bar{h}\right)\right\|\\
\lesssim &\left\|(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal)^\intercal\right\|_{\mathrm{MAX}}\|\beta_g\eta+\beta_h\|_\infty\left\|T^{-1}\bar{H}\bar{V}^\intercal\lambda-\Sigma_h(\eta^\intercal\lambda_g+\lambda_h)-\bar{h}\right\|_{\mathrm{MAX}}\\
\lesssim_p &s^2T^{-1}\log(n\vee p\vee T).
\end{aligned}
$$

The conclusion then follows from (B.28). ☐

Lemma 5. *Under Assumptions B.1, B.2, B.3, B.4, we have*

$$
n^{-1}\left\|\widehat{C}_g^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|\lesssim_p s(n^{-1/2}T^{-1/2}+T^{-1})\log(n\vee p\vee T). \tag{B.42}
$$

Proof of Lemma 5. Note that by (B.24), we have

$$
\begin{aligned}
n^{-1}\left\|\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|\leq &n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\right\|n^{-1/2}\left\|\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right\|\\
\lesssim_p &s(n^{-1/2}T^{-1/2}+T^{-1})\log(n\vee p\vee T),
\end{aligned}
$$

where we use the following estimates as a result of Assumptions B.1 and B.4:

$$
\begin{aligned}
n^{-1/2}\|\bar{u}\|\lesssim\|\bar{u}\|_{\mathrm{MAX}}&\lesssim_p T^{-1/2}(\log n\vee p\vee T)^{1/2},\\
n^{-1/2}\left\|T^{-1}\bar{U}\bar{V}^\intercal\lambda\right\|\lesssim\left\|T^{-1}\bar{U}\bar{M}^\intercal\gamma_0\right\|_{\mathrm{MAX}}&\lesssim_p T^{-1/2}(\log(n\vee p\vee T))^{1/2}.
\end{aligned}
$$

Moreover, by triangle inequality, we have

$$
\begin{aligned}
&n^{-1}\left\|C_e^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|\\
\leq &n^{-1}\left\|C_e^\intercal\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|+n^{-1}\left\|C_e^\intercal\mathbb{P}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|
\end{aligned}
$$

For the first term, we have

$$
n^{-1}\left\|C_e^\intercal\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|\leq n^{-1}\left\|C_e^\intercal\bar{u}\right\|+T^{-1}n^{-1}\left\|C_e^\intercal\bar{U}\bar{V}^\intercal\lambda\right\|\lesssim_p sn^{-1/2}T^{-1/2}.
$$

As to the second term, using Assumption B.3 and (B.32) we have

$$
\begin{aligned}
&n^{-1}\left\|C_e^\intercal\mathbb{P}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|\\
=&n^{-1}\left\|C_e^\intercal(\iota_n:\widehat{C}_h[\widehat{I}])\left((\iota_n:\widehat{C}_h[\widehat{I}])^\intercal(\iota_n:\widehat{C}_h[\widehat{I}])\right)^{-1}(\iota_n:\widehat{C}_h[\widehat{I}])^\intercal\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|
\end{aligned}
$$

$$\lesssim sn^{-1}\left\|C_e^\intercal(\iota_n:\widehat{C}_h[\widehat{I}])\right\|_{\text{MAX}} n^{-1}\left\|(\iota_n:\widehat{C}_h[\widehat{I}])^\intercal\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|_{\text{MAX}}$$
$$\lesssim_p s(n^{-1/2}T^{-1/2}+T^{-1})\log(n\vee p\vee T),$$

where we also use the following

$$n^{-1}\left\|(\iota_n:\widehat{C}_h)^\intercal\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|_{\text{MAX}} \leq \left(\left\|\widehat{C}_h-C_h\right\|_{\text{MAX}}+\|(\iota_n:C_h)\|_{\text{MAX}}\right)\left\|\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right\|_{\text{MAX}}$$
$$\lesssim_p T^{-1/2}(\log(n\vee p\vee T))^{1/2}.$$

Finally, we note that

$$n^{-1}\left\|\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)^\intercal \mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\bar{u}-T^{-1}\bar{U}\bar{V}^\intercal\lambda\right)\right\|$$
$$\lesssim\left\|\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right\|_{\text{MAX}}\left\|\bar{u}-T^{-1}\bar{U}\bar{M}^\intercal\right\|_{\text{MAX}}\lesssim_p sT^{-1}\log(n\vee p\vee T).$$

This concludes the proof. □

Lemma 6. *Under Assumptions B.1, B.2, B.3, B.4, B.5, we have*

$$\left\|n\left(\widehat{C}_g^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_g\right)^{-1}\right\|\lesssim_p 1.$$

Proof of Lemma 6. Note that by (B.28), we have

$$\begin{aligned}
&\widehat{C}_g^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_g\\
=&C_e^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_e+C_e^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal+\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_e+\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\\
&+C_e^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)+\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_e\\
&+\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)+\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\\
&+\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)
\end{aligned}$$

There are 9 terms in total on the right-hand side. By (B.24), we have

$$\begin{aligned}
n^{-1}\left\|\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_e\right\|=&n^{-1}\left\|C_e^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\right\|\lesssim\|C_e\|_{\text{MAX}}\,n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\right\|\\
\lesssim_p&s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2},\\
n^{-1}\left\|\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\right\|\leq&n^{-1}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\right\|^2\lesssim_p s^2(n^{-1}+T^{-1})\log(n\vee p\vee T).
\end{aligned}$$

Also, we have

$$n^{-1}\left\|C_e^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)\right\|=n^{-1}\left\|\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_e\right\|$$
$$\lesssim\|C_e\|_{\text{MAX}}\left\|\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right\|_{\text{MAX}}\lesssim_p sT^{-1/2}(\log(n\vee p\vee T))^{1/2},$$
$$n^{-1}\left\|\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)\right\|=n^{-1}\left\|\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_h\chi^\intercal\right\|$$

$$
\begin{aligned}
&\lesssim n^{-1/2}\left\|\chi\widehat{C}_h^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\right\|\left\|\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right\|_{\mathrm{MAX}}\\
&\lesssim_p s^2(n^{-1/2}T^{-1/2}+T^{-1})\log(n\vee p\vee T),\\
&\quad n^{-1}\left\|\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\left(\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right)\right\|\\
&\lesssim\left\|\widehat{C}_g-C_g+(C_h-\widehat{C}_h)\chi^\intercal\right\|_{\mathrm{MAX}}^2\lesssim_p s^2T^{-1}\log(n\vee p\vee T).
\end{aligned}
$$

Finally, by (B.32) and Assumptions B.2 and B.3, we have

$$
\begin{aligned}
n^{-1}\left\|C_e^\intercal\mathbb{P}_{(\iota_n:\widehat{C}_h[\widehat{I}])}C_e\right\| =&n^{-1}\left\|C_e^\intercal(\iota_n:\widehat{C}_h[\widehat{I}])\left((\iota_n:\widehat{C}_h[\widehat{I}])^\intercal(\iota_n:\widehat{C}_h[\widehat{I}])\right)^{-1}(\iota_n:\widehat{C}_h[\widehat{I}])^\intercal C_e\right\|\\
\lesssim& sn^{-2}\left\|C_e^\intercal(\iota_n:\widehat{C}_h[\widehat{I}])\right\|_{\mathrm{MAX}}^2\lesssim_p s(n^{-1}+T^{-1})\log(n\vee p\vee T).
\end{aligned}
$$

Hence, we obtain

$$
n^{-1}\widehat{C}_g^\intercal\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\widehat{C}_g=n^{-1}C_e^\intercal C_e+o_p(1).
$$

The conclusion follows from (B.5) and Weyl inequalities. □

Lemma 7. *Under Assumptions B.1, B.2, B.3, B.4, B.5, B.6, we have*

$$
\left\|(\widehat{\gamma}_0:\widehat{\lambda}_h^\intercal)-(\gamma_0:\lambda_h^\intercal)\right\|\lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}.
$$

Proof. It follows from (9) that

$$
(\widehat{\gamma}_0:\widehat{\lambda}_h[\widehat{I}]^\intercal)^\intercal=\left(\left(\iota_n:\widehat{C}_h[\widehat{I}]\right)^\intercal\left(\iota_n:\widehat{C}_h[\widehat{I}]\right)\right)^{-1}\left(\iota_n:\widehat{C}_h[\widehat{I}]\right)^\intercal\left(\bar{r}-\widehat{C}_g\widehat{\lambda}_g\right),
$$

which implies that

$$
\left\|(\widehat{\gamma}_0:\widehat{\lambda}_h^\intercal)^\intercal-(\gamma_0:\lambda_h^\intercal)^\intercal\right\|\le\left\|(\widetilde{\gamma}_0:\widetilde{\lambda}_h^\intercal)^\intercal-(\breve{\gamma}_0:\breve{\lambda}_h^\intercal)^\intercal\right\|+\left\|(\widetilde{\xi}:\widetilde{\chi})^\intercal\widehat{\lambda}_g-(\xi:\chi)^\intercal\lambda_g\right\|,
$$

where

$$
\begin{aligned}
&(\widetilde{\gamma}_0:\widetilde{\lambda}_h^\intercal)^\intercal=\arg\min_{\gamma,\lambda}\left\{\left\|\bar{r}-\iota_n\gamma-\widehat{C}_h\lambda\right\|:\quad\lambda_j=0,\quad j\notin\widehat{I}\right\},\\
&(\widetilde{\xi}_j:\widetilde{\chi}_{j,\cdot})^\intercal=\arg\min_{\xi_j,\chi_{j,\cdot}}\left\{\left\|\widehat{C}_{g,\cdot,j}-\iota_n\xi_j-\widehat{C}_h\chi_{j,\cdot}^\intercal\right\|:\quad\chi_{j,k}=0,\quad k\notin\widehat{I}\right\},\quad j=1,2,\ldots,d.
\end{aligned}
$$

Moreover, because

$$
\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}])}\bar{r}=\iota_n\breve{\gamma}_0+\widehat{C}_h\breve{\lambda}_h-\iota_n\widetilde{\gamma}_0-\widehat{C}_h\widetilde{\lambda}_h+(C_h-\widehat{C}_h)\breve{\lambda}_h+C_e\lambda_g+\beta_g\bar{g}+\beta_h\bar{h}+\bar{u}
$$

we obtain, using $\widehat{I}_1\subseteq\widehat{I}$, (B.1), (B.5), (B.20) - (B.23), (B.26),

$$
\begin{aligned}
&n^{-1/2}\left\|\left(\iota_n:\widehat{C}_h\right)\left(\widetilde{\gamma}_0-\breve{\gamma}_0:(\widetilde{\lambda}_h-\breve{\lambda}_h)^\intercal\right)^\intercal\right\|\\
\le& n^{-1/2}\left\|\mathbb{M}_{(\iota_n:\widehat{C}_h[\widehat{I}_1])}\bar{r}\right\|+n^{-1/2}\left\|(C_h-\widehat{C}_h)\breve{\lambda}_h+C_e\lambda_g+\beta_g\bar{g}+\beta_h\bar{h}+\bar{u}\right\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq n^{-1/2}\left\|\left(\iota_n:\widehat{C}_h\right)\left(\widetilde{\gamma}_{\widehat{I}_1}-\breve{\gamma}_0:(\widetilde{\lambda}_{\widehat{I}_1}-\breve{\lambda}_h)^\intercal\right)^\intercal\right\|+2n^{-1/2}\left\|(C_h-\widehat{C}_h)\breve{\lambda}_h+C_e\lambda_g+\beta_g\bar{g}+\beta_h\bar{h}+\bar{u}\right\|\\
&\lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}.
\end{aligned}
$$

Since we have

$$
\begin{aligned}
&n^{-1/2}\left\|\left(\iota_n:\widehat{C}_h\right)\left(\widetilde{\gamma}_0-\breve{\gamma}_0:(\widetilde{\lambda}_h-\breve{\lambda}_h)^\intercal\right)^\intercal\right\|\\
\geq&\phi_{\min}^{1/2}(1+\widehat{s})\left[n^{-1}(\iota_n:\widehat{C}_h)^\intercal(\iota_n:\widehat{C}_h)\right]\left\|\left(\widetilde{\gamma}_0-\breve{\gamma}_0:(\widetilde{\lambda}_h-\breve{\lambda}_h)^\intercal\right)\right\|,
\end{aligned}
$$

it follows that

$$
\left\|\left(\widetilde{\gamma}_0-\breve{\gamma}_0:(\widetilde{\lambda}_h-\breve{\lambda}_h)^\intercal\right)\right\|\lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}.
$$

Similarly, we can obtain

$$
\left\|\left(\widetilde{\xi}-\xi:\widetilde{\chi}-\chi\right)\right\|\lesssim_p s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}.
$$

Therefore, using this, as well as Assumption B.1 and Theorem 1, we obtain

$$
\begin{aligned}
\left\|(\widetilde{\xi}:\widetilde{\chi})^\intercal\widehat{\lambda}_g-(\xi:\chi)^\intercal\lambda_g\right\|\leq&\left\|(\widetilde{\xi}-\xi:\widetilde{\chi}-\chi)\right\|\left\|\widehat{\lambda}_g\right\|+\|(\xi:\chi)\|\left\|\widehat{\lambda}_g-\lambda_g\right\|\\
\lesssim_p& s(n^{-1/2}+T^{-1/2})(\log(n\vee p\vee T))^{1/2}.
\end{aligned}
$$

This concludes the proof. □

References

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics*, 39:3320–3356.

Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, 84(3):985–1046.

Giglio, S. W. and Xiu, D. (2016). Asset pricing with omitted factors. Technical report, University of Chicago.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

White, H. (2000). *Asymptotic Theory for Econometricians: Revised Edition*. Emerald Group Publishing Limited.