Supplement to “Most powerful test against a sequence of high dimensional local alternatives”

Yi He[a,b,c,*], Sombut Jaidee[b], Jiti Gao[b]

[a] *Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands*
[b] *Department of Econometrics and Business Statistics, Monash University, 900 Dandenong Road, Caulfield East, Victoria 3145, Australia*
[c] *Tinbergen Institute, Gustav Mahlerplein 117, 1082MS Amsterdam, The Netherlands*

Abstract

This supplement consists of six appendixes. In Appendix C we report more simulation and empirical analysis results to support the findings in Section 3 and 4. In Appendix D we check the generic conditions of the theorems for some example models in Section 3. In Appendix E we discuss the asymptotic theory for the time-variation adjusted data (see Remark 2 in the main document) and how to relax the additional condition (17). In Appendix F we prove all the lemmas used in the proof of main theorems. In Appendix G we prove Propositions 1 and 2 in Section 2. Finally, in Appendix H, we provide the complete proof of Corollaries 1–6.

Appendix C. More simulation and empirical analysis results

Appendix C.1. Simulation results for $\sqrt{p}/n = 0.1$

We repeat the simulation study, for a larger order of $\sqrt{p}/n = 0.1$. We observe similar patterns from Tables C.1 and C.2 as that in Section 3. The feasible and oracle tests have similar sizes over all scenarios, and require robust corrections for the time-series predictors with large concentration ratio p/n. They show similar power performances for small departures, but more different power for larger departures. This is because the error variance estimator contains a larger finite-sample upward bias under the alternatives.

Appendix C.2. Simulations results for the non-free dense alternatives in Goeman et al. (2006)

In this section, we revisit the simulations in Goeman et al. (2006). We use the same setup in our simulation study in Section 3, but now generate the direction of regression coefficient, that is, ξ adaptively as follows:

*Corresponding author.
Email addresses: yi.he@uva.nl (Yi He), sombut.jaidee@monash.edu (Sombut Jaidee), jiti.gao@monash.edu (Jiti Gao)

Preprint submitted to Elsevier *October 17, 2021*

Table C.1: Size and power (%) of the tests against uniform stochastic coefficient (i) at level $\alpha = 5\%$ with $p/n = \frac{1}{4}, \frac{1}{2}, 1, 2, 4$ and $\sqrt{p}/n = 0.1$. The columns are for: (i) the feasible test using $\widehat{\sigma}_n^2$ and assuming $\rho_n^2 = 0$, (i^o) the oracle test using the true variance σ_n^2 and assuming $\rho_n^2 = 0$, (i*) the robust test using $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n^2$.

	IID			CSD			MA1			AR1		
p/n	(i)	(i^o)	(i*)	(i)	(i^o)	(i*)	(i)	(i^o)	(i*)	(i)	(i^o)	(i*)
$H_0 : \|\beta\|^2 = 0$												
6/25	4.6	5.1	5.7	5.3	5.4	6.4	4.6	4.9	6.0	4.8	4.6	6.0
25/50	5.2	4.9	5.6	5.4	5.6	5.9	5.2	5.5	6.0	5.5	5.5	6.6
100/100	5.6	5.6	6.0	5.8	6.0	6.1	5.0	5.1	6.2	6.2	6.0	7.3
400/200	6.2	6.3	6.3	5.6	5.8	5.8	4.1	4.3	5.8	4.8	4.9	7.1
1600/400	5.7	5.7	5.8	5.5	5.6	5.6	3.2	3.5	5.7	3.2	3.4	6.4
$H_a^1 : \|\beta\|^2 = 1 \times \frac{\sqrt{p}}{n}$												
6/25	17.4	19.8	20.1	17.3	19.6	19.6	17.4	19.2	20.0	17.9	19.2	20.6
25/50	19.6	23.8	21.1	21.8	23.8	23.1	21.5	23.4	23.6	21.8	23.7	24.1
100/100	22.1	27.5	22.9	24.2	27.2	24.9	20.5	23.8	23.4	22.4	25.1	25.3
400/200	22.9	27.9	23.4	23.6	26.7	23.9	19.2	22.5	22.6	19.3	22.2	24.2
1600/400	23.9	30.1	24.1	23.6	28.4	23.9	14.4	18.1	21.5	13.1	16.6	21.3
$H_a^2 : \|\beta\|^2 = 2 \times \frac{\sqrt{p}}{n}$												
6/25	30.0	35.9	33.6	29.2	33.1	32.2	28.4	32.5	32.2	29.1	33.0	33.3
25/50	37.3	44.8	39.3	38.0	42.8	39.4	36.6	42.5	39.4	36.5	41.9	39.8
100/100	42.6	53.7	43.8	43.4	51.3	44.2	39.7	46.4	42.9	40.4	47.9	44.2
400/200	45.4	57.3	45.9	47.1	56.4	47.7	39.1	48.4	44.6	38.1	46.9	44.8
1600/400	46.7	59.9	47.1	48.9	58.7	49.2	33.0	42.6	42.8	29.7	39.2	41.7
$H_a^3 : \|\beta\|^2 = 5 \times \frac{\sqrt{p}}{n}$												
6/25	56.0	68.4	59.7	50.2	59.4	53.7	49.0	57.8	53.3	50.8	60.2	55.4
25/50	70.3	83.5	72.0	67.1	76.1	68.8	64.8	74.7	67.6	65.4	74.5	68.4
100/100	79.8	91.7	80.6	78.7	88.1	79.5	75.3	85.2	77.9	74.9	84.6	77.7
400/200	85.0	94.5	85.3	86.6	94.5	86.8	79.9	90.6	83.5	78.2	89.1	82.8
1600/400	87.5	96.3	87.7	90.0	96.7	90.1	78.7	90.5	84.9	74.8	88.5	83.8

(iii) $\xi = \frac{U_n \Lambda_n^{s/2} \mathbf{1}_p}{\left\| U_n \Lambda_n^{s/2} \mathbf{1}_p \right\|}$ with $s = 0, 0.5, 1, 1.5$,

where $U_n = (u_1, \ldots, u_n)$ and $\Lambda_n = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ contain the eigenvectors and eigenvalues of the sample covariance matrix S_n respectively. We only consider the regular case with $s \geq 0$, where the large variance principal components contains more information in forecasting the response variables; see the aforementioned paper for more discussions. Following the setup therein we use $p = 52$ and $n = 294$, leading to a contraction ratio $p/n \approx 0.177$ and an order of alternatives $\sqrt{p}/n \approx 0.025$.

Table C.2: Size and power (%) of the tests against deterministic coefficient (ii) at level $\alpha = 5\%$ with $p/n = \frac{1}{4}, \frac{1}{2}, 1, 2, 4$ and $\sqrt{p}/n = 0.1$. The columns are for: (ii) the feasible test using $\widehat{\sigma}_n^2$ and assuming $\rho_n^2 = 0$, (ii^o) the oracle test using the true variance σ_n^2 and assuming $\rho_n^2 = 0$, (ii*) the robust test using $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n^2$.

	IID			CSD			MA1			AR1		
p/n	(ii)	(ii^o)	(ii*)	(ii)	(ii^o)	(ii*)	(ii)	(ii^o)	(ii*)	(ii)	(ii^o)	(ii*)
					$H_0 : \\|\beta\\|^2 = 0$							
6/25	4.6	4.3	5.5	4.6	4.8	5.8	5.7	5.6	7.1	5.2	5.6	6.3
25/50	4.9	5.0	5.4	5.8	6.1	6.5	5.4	5.8	6.4	5.9	5.8	7.3
100/100	5.5	5.8	5.9	5.9	5.8	6.2	4.3	5.1	5.7	5.6	5.3	6.8
400/200	6.0	6.1	6.2	5.8	5.7	5.9	4.7	4.6	6.1	4.8	4.7	6.7
1600/400	4.9	5.1	4.9	6.5	6.4	6.6	3.2	3.7	6.0	3.3	3.3	6.6
					$H_a^1 : \\|\beta\\|^2 = 1 \times \frac{\sqrt{p}}{n}$							
6/25	16.4	19.0	18.9	18.4	20.4	20.9	19.5	20.5	22.0	18.5	19.8	21.5
25/50	19.7	23.0	21.2	21.4	23.9	22.8	21.7	23.7	24.0	21.3	24.0	24.0
100/100	22.6	26.8	23.5	23.6	27.0	24.4	20.8	24.0	23.3	20.8	23.8	23.9
400/200	22.8	27.8	23.3	23.6	27.6	24.2	18.6	22.7	22.9	18.6	22.2	24.0
1600/400	22.9	29.2	23.2	25.0	29.5	25.3	14.6	18.4	21.8	13.8	17.4	21.8
					$H_a^2 : \\|\beta\\|^2 = 2 \times \frac{\sqrt{p}}{n}$							
6/25	29.3	34.7	33.1	31.4	35.1	34.8	31.9	35.1	35.7	30.6	35.1	34.9
25/50	35.4	43.8	37.2	38.2	44.4	39.7	37.8	43.1	40.6	37.1	43.2	40.7
100/100	41.9	52.2	42.9	43.8	51.0	44.6	39.4	46.5	42.3	39.0	45.8	42.6
400/200	45.4	57.0	46.2	47.1	55.7	47.6	39.0	47.8	44.8	37.9	46.6	45.0
1600/400	46.2	59.9	46.6	51.0	60.9	51.3	32.6	42.4	42.8	30.4	39.3	41.6
					$H_a^3 : \\|\beta\\|^2 = 5 \times \frac{\sqrt{p}}{n}$							
6/25	56.3	69.1	60.3	58.1	68.0	61.9	57.0	67.1	61.7	57.6	67.3	61.4
25/50	66.8	79.6	68.6	71.4	80.9	73.0	66.9	77.8	69.5	67.2	77.5	70.2
100/100	78.9	90.7	79.8	81.1	89.7	81.7	75.9	86.0	78.5	74.7	85.7	77.9
400/200	85.2	95.2	85.6	87.5	94.7	87.7	79.7	90.7	83.5	79.8	90.1	84.1
1600/400	88.2	96.9	88.4	91.5	97.3	91.6	78.4	90.5	84.7	74.9	88.3	83.4

Note that the regression coefficient vector is not free except the case with $s = 0$. We use the general asymptotic departure $\varpi_n = \varpi_n(s)$ given in Remark 1, rather the one for free alternatives, to generate the variance of regression errors $\sigma_n^2 = \varpi_n(s)/\sqrt{2}$. Hence, the asymptotic size and power only depends on the length of β under regular scenarios.

Table C.3 reports the results for the adaptive direction (iii) for different values of s. Again we report the size and power for three different tests: the feasible test using the estimated variance $\widehat{\sigma}_n^2$ for regular scenarios (i.e. assuming $\rho_n^2 = 0$), the oracle test using the true variance σ_n^2 for regular scenarios (i.e. assuming $\rho_n^2 = 0$), and the robust test using the estimated variance $\widehat{\sigma}_n^2$ and the

Table C.3: Size and power (%) of the tests against adaptive direction (iii) at level $\alpha = 5\%$ with $p = 52$ and $n = 294$. The columns are for: (iii) the feasible test using $\widehat{\sigma}_n^2$ and assuming $\rho_n^2 = 0$, (iii^o) the oracle test using the true variance σ_n^2 and assuming $\rho_n^2 = 0$, (iii*) the robust test using $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n^2$.

	IID			CSD			MA1			AR1		
s	(iii)	(iii^o)	(iii*)	(iii)	(iii^o)	(iii*)	(iii)	(iii^o)	(iii*)	(iii)	(iii^o)	(iii*)
$H_0 : \lVert\beta\rVert^2 = 0$												
0	5.7	5.7	5.9	6.1	6.0	6.3	5.9	5.9	6.1	6.3	6.1	6.6
0.5	5.4	5.5	5.4	6.2	6.5	6.3	5.6	5.6	5.7	5.6	5.8	5.8
1	5.9	6.3	6.1	5.5	5.9	5.7	6.7	6.3	6.9	6.5	6.4	6.8
1.5	6.2	6.2	6.2	6.0	5.9	6.1	6.0	6.0	6.2	6.7	6.8	7.0
$H_a^1 : \lVert\beta\rVert^2 = 1 \times \frac{\sqrt{p}}{n}$												
0	25.9	26.3	26.1	25.6	26.5	25.9	26.2	26.8	26.8	26.0	26.1	26.6
0.5	26.6	27.7	26.8	27.3	27.4	27.4	26.6	26.5	27.2	25.8	26.1	26.4
1	26.7	27.6	27.0	26.0	26.1	26.2	27.6	27.7	28.2	27.6	27.7	28.5
1.5	26.8	27.1	27.2	26.3	26.9	26.4	26.8	27.1	27.3	28.2	28.3	28.8
$H_a^2 : \lVert\beta\rVert^2 = 2 \times \frac{\sqrt{p}}{n}$												
0	53.0	55.6	53.4	52.3	53.9	52.6	51.5	53.0	52.1	51.7	53.3	52.6
0.5	53.9	55.8	54.1	52.6	53.3	52.7	51.5	52.9	52.2	51.1	52.8	51.9
1	53.8	55.8	54.2	51.8	52.4	52.2	52.4	53.5	53.2	51.4	52.5	52.5
1.5	55.0	56.4	55.3	52.5	53.2	52.7	52.4	52.8	53.0	53.0	53.6	53.7
$H_a^3 : \lVert\beta\rVert^2 = 5 \times \frac{\sqrt{p}}{n}$												
0	95.9	97.2	96.0	94.4	95.5	94.4	93.3	94.6	93.4	93.5	95.0	93.7
0.5	95.9	97.3	96.0	93.9	94.9	94.0	93.6	94.7	94.0	93.6	94.7	93.8
1	96.0	97.2	96.1	93.6	94.7	93.7	93.0	94.3	93.1	93.1	94.0	93.4
1.5	95.5	96.8	95.5	93.6	94.4	93.7	93.6	94.7	93.8	92.8	94.1	93.1

estimated irregularity coefficient $\widehat{\rho}_n^2$ for both regular and irregular scenarios. We observe that, for each departure value h, the size and power are stable over all scenarios. This clearly suggests the good performance of our general asymptotic approximations in Remark 1.

Appendix C.3. Simulation results under contemporary correlations

In this subsection we provide some additional simulations results to illustrate the power of our robust test when there are non-trivial contemporary (and lag) correlations between the nuisance variable z_t and the high-dimensional aggregate variable $x_t^T\beta$.

For simplicity we consider an univariate nuisance variable $z_t \in \mathbb{R}$, but the results are similar with multiple nuisance variables. We consider the regression model given by

$$y_t = \theta_0 + z_t\theta_1 + x_t^T\beta + \varepsilon_t$$

Table C.4: Empirical size and power (%) against uniform stochastic coefficient at level $\alpha = 5\%$ with $p/n = \frac{1}{4}, \frac{1}{2}, 1, 2, 4$ and $\sqrt{p}/n = 0.05$, using the least-squares variance estimator $\widehat{\sigma}_n^2$. The columns are for: (0) no contemporary dependence with $\rho = 0$ (+) positive contemporary dependence with $\rho = 0.2$ (-) negative contemporary dependence with $\rho = -0.2$.

	IID			CSD			MA1			AR1		
p/n	(0)	(+)	(−)	(0)	(+)	(−)	(0)	(+)	(−)	(0)	(+)	(−)
						$H_0 : \lVert\beta\rVert^2 = 0$						
25/100	5.8	5.8	5.8	6.0	6.0	6.0	6.4	6.4	6.4	6.3	6.3	6.3
100/200	5.8	5.8	5.8	5.7	5.7	5.7	5.0	5.0	5.0	5.7	5.7	5.7
400/400	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	4.9	4.9	4.9
1600/800	5.1	5.1	5.1	5.4	5.4	5.4	5.3	5.3	5.3	5.6	5.6	5.6
6400/1600	5.2	5.2	5.2	5.6	5.6	5.6	5.2	5.2	5.2	5.6	5.6	5.6
						$H_a^1 : \lVert\beta\rVert^2 = 1 \times \frac{\sqrt{p}}{n}$						
25/100	24.7	23.0	22.6	25.3	23.9	24.2	26.6	24.6	25.9	24.8	22.4	23.8
100/200	25.1	23.5	23.2	26.6	25.0	24.5	25.1	22.4	24.3	25.2	22.0	24.4
400/400	24.9	23.0	22.8	25.6	23.7	23.4	24.3	20.0	24.0	25.0	20.5	25.2
1600/800	24.0	21.6	21.8	25.8	23.4	23.4	24.6	18.0	26.9	24.7	18.5	27.1
6400/1600	23.7	21.6	21.6	25.6	23.4	23.4	24.4	15.6	30.7	24.8	14.7	32.3
						$H_a^2 : \lVert\beta\rVert^2 = 2 \times \frac{\sqrt{p}}{n}$						
25/100	46.1	43.4	42.9	46.8	43.6	44.2	45.3	42.5	43.4	45.2	42.1	43.0
100/200	51.6	47.9	47.0	51.1	47.6	47.3	49.8	44.6	48.0	50.1	45.1	48.1
400/400	51.7	47.3	47.8	52.7	48.9	49.0	52.0	44.6	50.8	52.6	44.5	51.6
1600/800	53.7	48.7	48.6	54.3	50.0	49.5	55.4	43.3	57.6	54.5	42.0	57.4
6400/1600	51.9	47.1	47.8	56.2	51.5	51.5	56.9	37.2	66.5	57.2	36.8	67.2
						$H_a^3 : \lVert\beta\rVert^2 = 5 \times \frac{\sqrt{p}}{n}$						
25/100	86.4	83.8	82.6	80.5	77.8	77.4	79.4	76.8	77.3	79.0	76.0	76.4
100/200	92.3	89.7	89.6	89.5	86.9	87.0	89.4	86.2	87.6	89.7	86.5	87.8
400/400	95.4	93.2	93.3	94.5	92.2	92.1	95.1	91.3	94.2	94.4	90.9	93.7
1600/800	97.1	95.2	95.3	97.2	95.5	95.9	96.9	93.0	97.2	97.8	93.7	97.9
6400/1600	97.7	95.9	95.7	97.8	96.5	96.7	98.5	93.0	99.2	98.7	92.4	99.5

where the intercept $\theta_0 = 0$ and $\theta_1 = 1$ without loss of generality; note that these values do not change the distribution of our test statistic. However, the true values are unknown to the statistician who always demeans the predictors in each sample and estimate the nuisance parameters. We generate the same regression errors $\varepsilon_t = \sigma_n \eta_t$ and the high-dimensional covariates $\{x_t\}$ from the DGPs 1–4 as in the main document. To save space, we only report the results for coefficient vector $\beta = \lVert\beta\rVert\, \xi_n$ with the direction ξ_n generated uniformly over the $\mathbb{R}^p$ unit sphere, that is, the case (i) in the main document; the results for the directions in case (ii) are similar and therefore omitted.

We choose the same pairs of $(p, n) \in \{(25, 100), (100, 200), (400, 400), (1600, 800), (6400, 1600)\}$ and same signal length $h_n^2 = \frac{n}{\sqrt{p}} \|\beta\| \in \{0, 1, 2, 5\}$ as in the main document.

We generate the nuisance variable z_t from the linear model given by

$$z_t = \psi_1 x_{t-1}^T \beta + \psi_1 \varepsilon_{t-1} + \psi_0 \cdot x_t^T \beta + v_t,$$

where the errors $v_t = \widetilde{\sigma}_n \widetilde{\eta}_t$ with $\widetilde{\eta}_t \overset{iid}{\sim} N(0, 1)$ independent of $\{x_t\}$ and $\{\varepsilon_t\}$. We set $\psi_1 = 0.3$ and vary the value of

$$\psi_0 = \begin{cases} \rho\sqrt{\frac{\text{var}(z_t)}{\text{var}(x_t^T\beta)}} - \psi_1 \text{corr}(x_{t-1}^T\beta, x_t^T\beta) & \beta \neq \mathbf{0}_p \\ 0 & \beta = \mathbf{0}_p, \end{cases}$$

for each DGP to ensure that the contemporary correlation

$$\rho := \text{corr}\left(z_t, x_t^T\beta\right) \in \{0, 0.2, -0.2\}.$$

Finally, we choose the standard deviation $\widetilde{\sigma}_n$ such that $\text{var}(z_t) = 1$ for each DGP.

Table C.4 reports the empirical size and power of our feasible robust test (14) in our main document. Our test maintains good sizes and non-trivial powers overall, whereas the power performance depends on the sign of the contemporary correlation ρ. In the regular cases with time independent covariates (IID and CSD), we observe a slight loss of finite-sample testing power regardless of the sign of the contemporary correlation ρ. In the irregular scenarios where the power bias term b_n become non-negligible (see Remark 6 in the main document), we observe that our test becomes less powerful when $\rho > 0$ but the power is (partially) recovered or even boosted when $\rho < 0$. The difference is particularly significant in higher dimensions with larger concentration ratio p/n according to our findings in Remark 6.

Appendix C.4. Robustness checks for our empirical analysis

We first report the rolling-window p-values, without robust corrections, for both the unadjusted and the time-variation adjusted data respectively in Figures C.1 and C.2. All the plots show very similar patterns to that in Section 4.

To illustrate how our test outcomes may explain the time-varying predictive gain by using the high-dimensional covariates, we compare the autoregressive forecasts with the ridge estimators which usually shows the best predictive performance among competitors in our empirical analysis; see, e.g., De Mol et al. (2008). Our conclusions remain qualitatively the same for lasso and principal component estimators. To keep the estimators comparable, we jointly estimate the nuisance parameter θ and the coefficient vector β, but penalize the L_2 norm of β only in our ridge estimation. We choose the optimal penalty coefficient by using the bias-corrected 10-fold cross-validation with the autoregressive residuals; see Liu and Dobribabn (2020). Following Stock and Watson (2002), in every month we use the observations over the last $n = 120$ months as our training data, and forecast the next-month industrial production growth ahead.

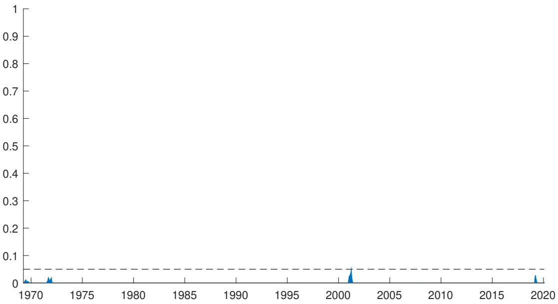
For every given $d \in \{0, 1, \ldots, 5\}$, in Figure C.3 we plot the time series of the relative out-of-sample R^2, beginning from December 1979, which is given by

$$\text{Relative Out-of-sample } R^2 = \frac{\sum_{t=m-119}^{m} (\widehat{y}_t - y_t)^2}{\sum_{t=m-119}^{m} (\widehat{y}_t^{\text{AR}} - y_t)^2},$$

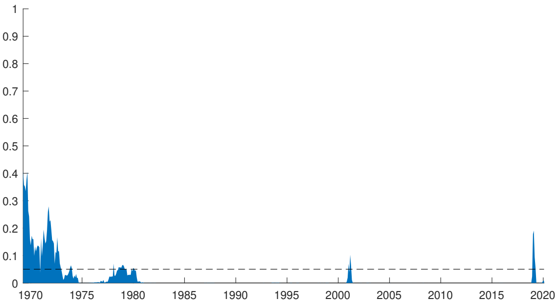
for the rolling window ending in month m, where $\widehat{y}_t$ denotes the ridge forecasts (or other forecasts of interest) and $\widehat{y}_t^{\text{AR}}$ denotes the autoregressive forecasts which are both available at time $t-1$. Overall, we observe a non-trivial reduction in forecast errors by using the covariates during the periods our tests are significant, whereas little gain or even a (large) loss in predictive accuracy when our tests are insignificant. Consistent with Figure 1 in the main document, our results suggest that the covariates become less useful when more lagged target values are included since early 2000.

We also report the rolling-window relative out-of-sample R^2 for LASSO and principal components estimators in Figure C.4. Like for the ridge estimator, we jointly estimate the nuisance parameter θ and the coefficient vector β, but penalize the L_1 norm of β only in our lasso estimation. We choose the optimal penalty coefficient by using the bias-corrected 10-fold cross-validation with the autoregressive residuals. For principal component regression, we report the results for using 3 principal components.

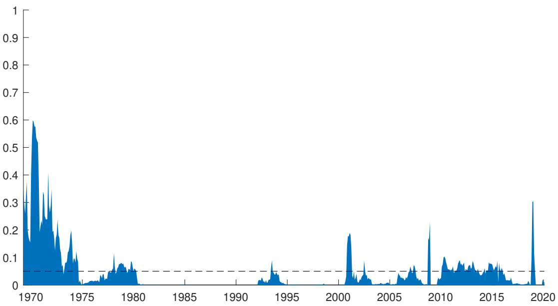
Figure C.1: Ten years ($n = 120$) rolling windows monthly unadjusted p values between March, 1969 and February, 2020 for different number of lags $d = 0, 1, 2, 3, 4, 5$.



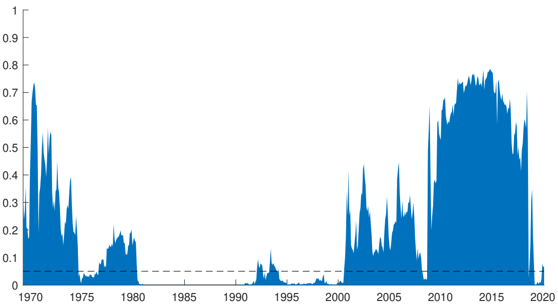
(a) No lag variable



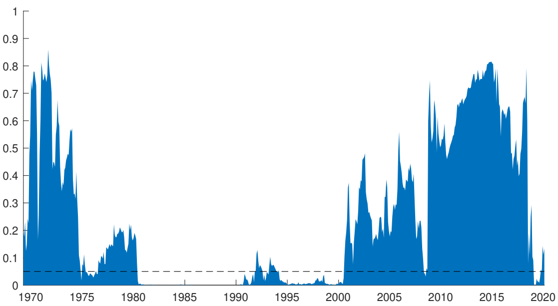
(b) lag $d = 1$



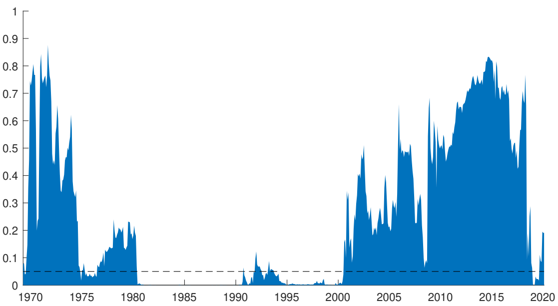
(c) lag $d = 2$



(d) lag $d = 3$

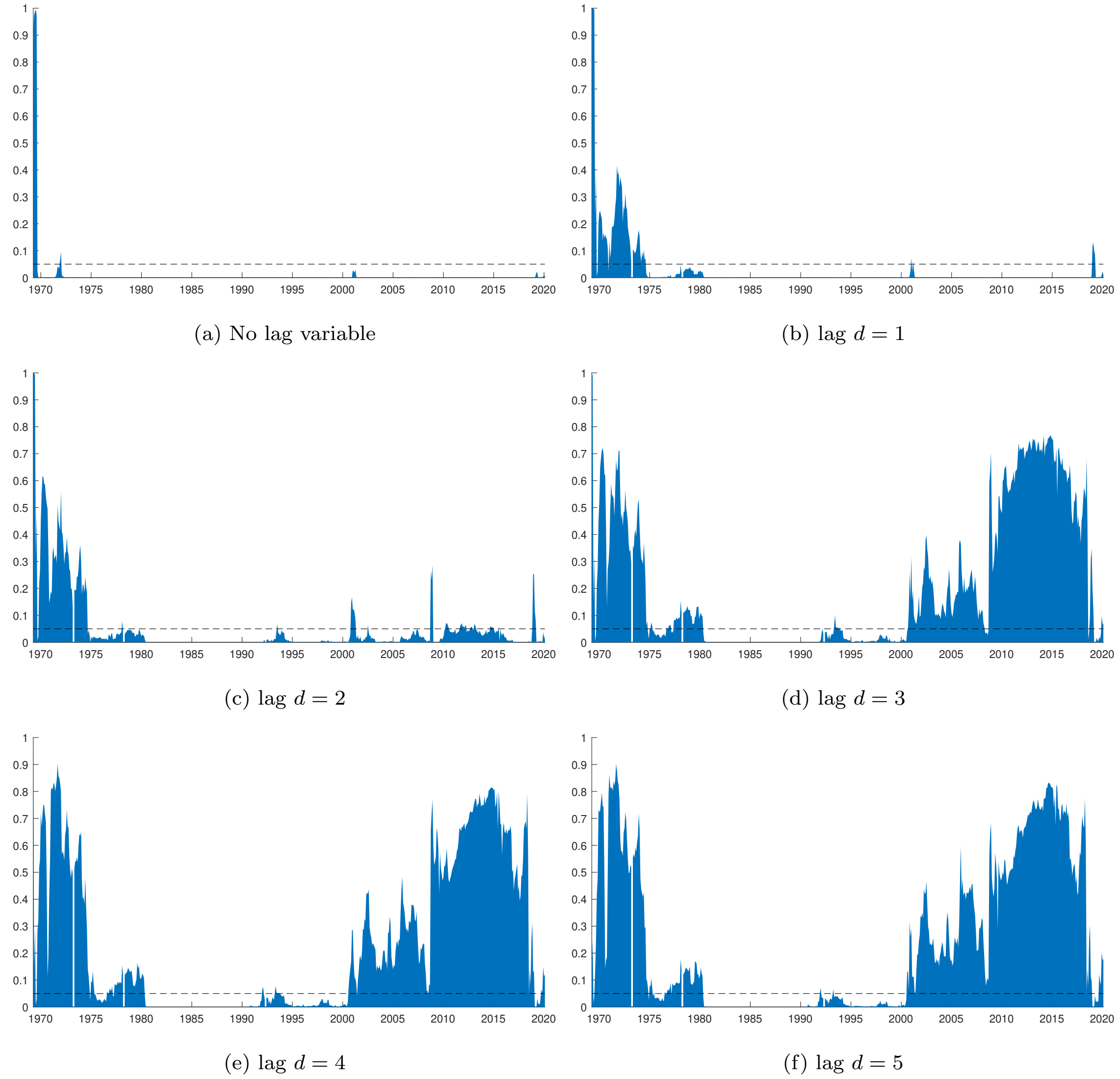


(e) lag $d = 4$



(f) lag $d = 5$

Figure C.2: Ten years ($n = 120$) rolling windows monthly time-variation adjusted p values between March, 1969 and February, 2020 for different number of lags $d = 0, 1, 2, 3, 4, 5$.



(a) No lag variable

(b) lag $d = 1$

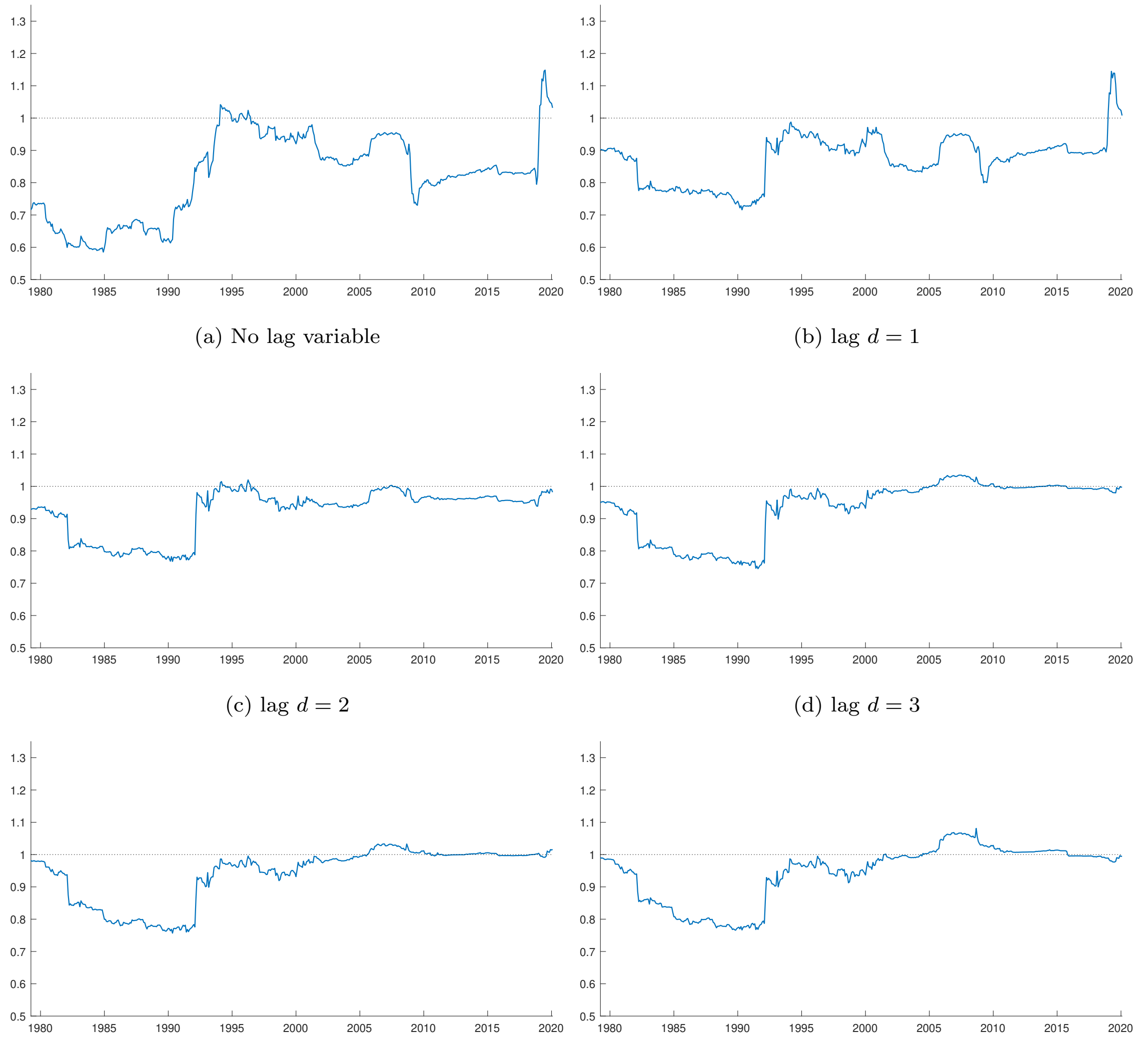
(c) lag $d = 2$

(d) lag $d = 3$

(e) lag $d = 4$

(f) lag $d = 5$

Figure C.3: Relative out-of-sample mean squared forecast errors for the ridge estimator (solid line) against the autoregressive forecasts (dotted reference line) in rolling windows of $n = 120$ months.



(a) No lag variable

(b) lag $d = 1$

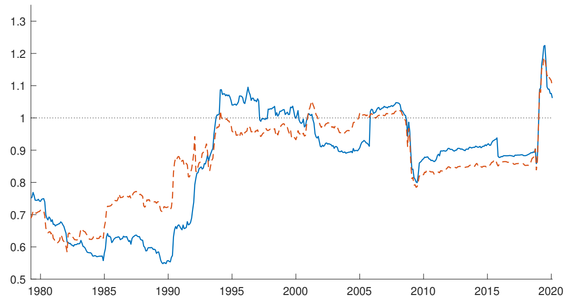
(c) lag $d = 2$

(d) lag $d = 3$

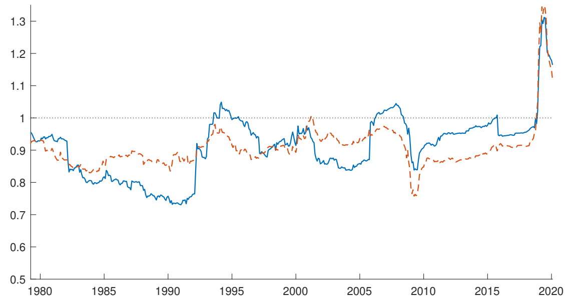
(e) lag $d = 4$

(f) lag $d = 5$

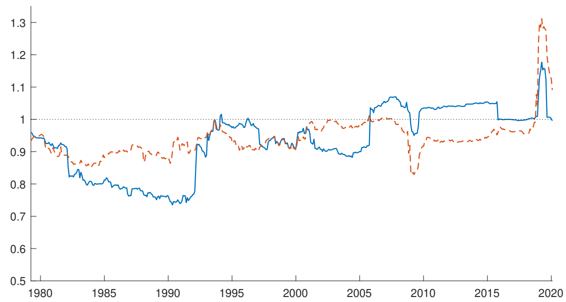
Figure C.4: Relative out-of-sample mean squared forecast errors for the LASSO estimator (solid line) and the principal component estimator (dashed line) against the autoregressive forecasts (dotted reference line) in rolling windows of $n = 120$ months.



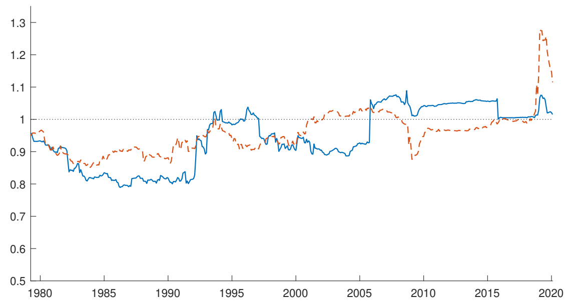
(a) No lag variable



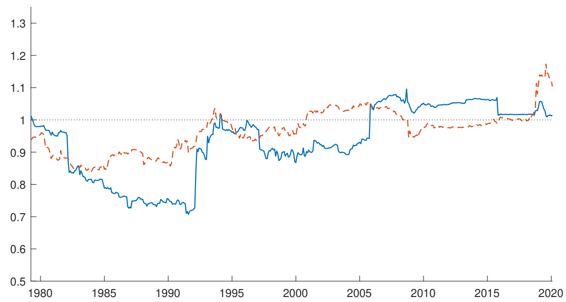
(b) lag $d = 1$



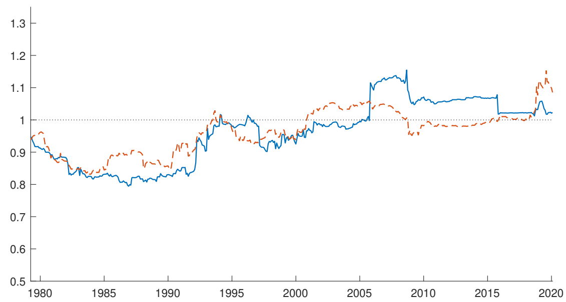
(c) lag $d = 2$



(d) lag $d = 3$



(e) lag $d = 4$



(f) lag $d = 5$

Appendix D. Checking technical conditions for example models

For all examples in this part we consider the standard asymptotic regime that $p/n \to c \in (0, \infty)$ in random matrix theory. Unless specified otherwise, all the inequalities hold with probability 1 and we do not repeat this argument for presentation convenience.

Appendix D.1. Time-independent model

Consider the time-independent model in Proposition 2, where $x_t = \Sigma^{1/2} v_t$ where $\{v_{t,i} : t = 1, \ldots, n, i = 1, \ldots, p\}$ is a double array of i.i.d. random variables with zero mean, unit variance and finite kurtosis bounded in n. Assume further than Σ has a bounded spectral norm in n.

First, we verify the condition (ii) and condition (iii) of Theorem 1. Let $\mathbb{S}_n = \frac{1}{n}\Sigma^{1/2} X^T X \Sigma^{1/2}$. By Bai and Silverstein (1998) we know that $\lambda_{\max}(\mathbb{S}_n) = O(1)$. Then

$$\lambda_{\max}(\underline{S}_n) = \lambda_{\max}(S_n) = \lambda_{\max}(\mathbb{S}_n - \bar{x}\bar{x}^T) \leq \lambda_{\max}(\mathbb{S}_n) = O(1).$$

It follows that

$$\|A_n\|_{sp} = \|\underline{S}_n - \text{diag}\,(\underline{S}_n)\|_{sp} \leq \|\underline{S}_n\|_{sp} + \|\text{diag}\,(\underline{S}_n)\|_{sp} \leq 2\,\|\underline{S}_n\|_{sp} = O(1).$$

On the other hand, $\|A_n\|^2 = \text{tr}\,(\underline{S}_n^2) - \text{tr}\left((\text{diag}\,(\underline{S}_n))^2\right) \leq \text{tr}\,(\underline{S}_n^2) - \frac{1}{n}\,\text{tr}^2\,(\underline{S}_n)$. Recall that $F^{\underline{S}_n}$ tends to a non-degenerate limit $\underline{F}$ with probability 1, and thus

$$\frac{1}{n}\,\text{tr}\,(\underline{S}_n^2) - \frac{1}{n^2}\,\text{tr}^2\,(\underline{S}_n) = \int x^2 dF^{\underline{S}_n} - \left(\int x dF^{\underline{S}_n}\right)^2 \xrightarrow{a.s.} \int x^2 d\underline{F} - \left(\int x d\underline{F}\right)^2 > 0.$$

Hence, $\|A_n\|_{sp}^2 / \|A_n\|^2 = O(n^{-1}) \to 0$. The condition (ii) then follows; see our arguments in Section 2. For condition (iii) it suffices to check condition (5). Let $\ell \in \{1, 2, \ldots, n\}$. Note that

$$A_n(t+\ell, t) = \frac{1}{n}\,(x_{t+\ell} - \bar{x})^T\,(x_t - \bar{x})\,.$$

By Cauchy–Schwarz inequality, it is easy to show that, for some absolute constant M

$$A_n^2(t+\ell, t) \leq M\left\{\frac{1}{n^2}(x_{t+\ell}^T x_t)^2 + \frac{1}{n^2} x_t^T x_t \cdot \bar{x}^T\bar{x} + \frac{1}{n^2} x_{t+\ell}^T x_{t+\ell} \cdot \bar{x}^T\bar{x} + \frac{1}{n^2}\bar{x}^T\bar{x}\right\}.$$

Recall from above that $1/\|A_n\|^2 = O(n^{-1})$. It remains to show that

$$\frac{1}{n^3}\sum_{t=1}^{n}\left(x_{t+\ell}^T x_t\right)^2 = o_{\mathbb{P}}(1), \quad \frac{1}{n^3}\sum_{t=1}^{n} x_t^T x_t \cdot \bar{x}^T\bar{x} = O_{\mathbb{P}}(1), \text{ and } \frac{1}{n^2}\bar{x}^T\bar{x} = o_{\mathbb{P}}(1).$$

By a direct calculation and the trace inequality (Lemma 3),

$$\mathbb{E}\left[\frac{1}{n^3}\sum_{t=1}^{n}\left(x_{t+\ell}^T x_t\right)^2\right] = \frac{\text{tr}(\Sigma^2)}{n^2} \leq \lambda_{\max}(\Sigma)\frac{\text{tr}(\Sigma)}{n^2} \to 0.$$

Moreover,

$$\mathbb{E}\left[\frac{1}{n^3}\sum_{t=1}^{n} x_t^T x_t\right] = \frac{\text{tr}(\Sigma)}{n^2} \to 0, \text{ and } \mathbb{E}\left[\bar{x}^T\bar{x}\right] = \frac{\text{tr}(\Sigma)}{n} = O(1).$$

The rest follows easily from the Markov inequality (Lemma 1).

Next, we verify conditions (i)–(iii) in Theorem 2. Condition (i) follows immediately from above and we omit the details. Condition (ii) follows as our model is a special case of that in Proposition 1. Let κ denote the kurtosis of $v_{t,i}$, and $a := (a_1, \ldots, a_p) := \Sigma^{1/2}\xi_n$. It is easy to check that, for some large M

$$\begin{aligned}\mathbb{E}\left(x_t^T\xi_n\right)^4 = \mathbb{E}\left(v_t^T\Sigma^{1/2}\xi_n\right)^4 =& \kappa\cdot\sum_{i=1}^{p} a_i^4 + 3\sum_{i\neq j}^{p} a_i^2 a_j^2 \\ \leq& M\left(\sum_{i=1}^{p} a_i^2\right)^2 = M\left(\xi_n^T\Sigma\xi_n\right)^2 = O(\lambda_{\max}^2(\Sigma)) = O(1).\end{aligned}$$

This is condition (iii).

Appendix D.2. High dimensional MA(1) model

Consider the first-order moving average model given by

$$x_t = \psi w_{t-1} + w_t$$

where $\psi \in (-1, 1)$ is a scalar lagged coefficient and $w_t = \Sigma^{1/2}v_t$ follows the time-independent model in the last section. With a slight abuse of notation, here Σ denotes the population covariance matrix of w_t rather that of x_t.

We first check the condition (ii) in Theorem 1. We skip the condition (iii) therein as it may not hold in general according to our simulations, but this is not an issue for our robust test. Using the same arguments (and the limiting spectral distribution in Jin et al., 2009) as that for the time-independent model, it suffices to show that

$$\lambda_{\max}(\mathbb{S}_n) = O(1).$$

Observe that

$$\begin{aligned}\mathbb{S}_n := \frac{1}{n}\sum_{t=1}^{n} x_t x_t^T &= \frac{1}{n}\sum_{t=1}^{n}\left(\psi w_{t-1} + w_t\right)\left(\psi w_{t-1} + w_t\right)^T \\ &= \psi^2\frac{1}{n}\sum_{t=1}^{n} w_{t-1}w_{t-1}^T + \psi\frac{1}{n}\sum_{t=1}^{n}\left(w_{t-1}w_t^T + w_t w_{t-1}^T\right) + \frac{1}{n}\sum_{i=1}^{n} w_t w_t^T \\ &=: \psi^2\mathbb{S}_{n,1} + \psi\mathbb{S}_{n,2} + \mathbb{S}_{n,3}.\end{aligned}$$

From the last section, we already know that $\lambda_{\max}(\mathbb{S}_{n,1}) = O(1)$, and $\lambda_{\max}(\mathbb{S}_{n,3}) = O(1)$. Using the triangle inequality for spectral norms, it remains to show that

$$\left\|\mathbb{S}_{n,2}\right\|_{sp} = O(1).$$

Let $\xi \in \mathbb{R}^p$ be an arbitrary unit vector.

$$\begin{aligned}\left|\xi^T \mathbb{S}_{n,2} \xi\right| \leq \frac{1}{n}\sum_{t=1}^{n} 2\left|\xi^T w_{t-1} w_t^T \xi\right| \leq &\frac{1}{n}\sum_{t=1}^{n} \xi^T w_{t-1} w_{t-1}^T \xi + \frac{1}{n}\sum_{t=1}^{n} \xi^T w_t w_t^T \xi \\ =&\xi^T \mathbb{S}_{n,1}\xi + \xi^T \mathbb{S}_{n,3}\xi \leq \lambda_{\max}\left(\mathbb{S}_{n,1}\right) + \lambda_{\max}\left(\mathbb{S}_{n,3}\right).\end{aligned}$$

Note that the last upper bound does not depend on ξ. Then using the fact that $\mathbb{S}_{n,2}$ is symmetric, $\|\mathbb{S}_{n,2}\|_{sp} = \sup_{\|\xi\|=1}\left|\xi^T \mathbb{S}_{n,2}\xi\right| \leq \lambda_{\max}\left(\mathbb{S}_{n,1}\right) + \lambda_{\max}\left(\mathbb{S}_{n,3}\right) = O(1)$.

Next, we verify conditions (i)–(iii) in Theorem 2. We can deduce from above that $\lambda_{\max}(S_n) \leq \lambda_{\max}\left(\mathbb{S}_n\right) = O(1)$ and $\lambda_{\max}\left(\mathbb{E}\left[x_t x_t^T\right]\right) = (\psi^2+1)\lambda_{\max}(\Sigma) = O(1)$. For condition (i), it remains to show that $\lambda_{\max}\left(\mathbb{E}[\bar{x}\bar{x}^T]\right) = O(1)$. By a direct calculation,

$$\begin{aligned}\mathbb{E}[\bar{x}\bar{x}^T] =&\frac{1}{n}\mathbb{E}\left(x_t x_t^T\right) + \frac{1}{n}\mathbb{E}\left(x_t x_{t-1}^T\right) + \frac{1}{n}\mathbb{E}\left(x_{t-1} x_t^T\right) \\ =&\frac{1}{n}\left(\psi^2+1\right)\Sigma + \psi\frac{1}{n}\Sigma + \psi\frac{1}{n}\Sigma = \frac{1}{n}\left(\psi+1\right)^2\Sigma.\end{aligned}$$

Hence,

$$\lambda_{\max}\left(\mathbb{E}[\bar{x}\bar{x}^T]\right) = \frac{1}{n}\left(\psi+1\right)^2\lambda_{\max}(\Sigma) \to 0.$$

The condition (ii) follows from Proposition 1 directly, by rewriting

$$x_t = \left[\psi I_p, I_p\right]\begin{bmatrix} w_{t-1} \\ w_t \end{bmatrix} = \left[\psi\Sigma^{1/2}, \Sigma^{1/2}\right]\begin{bmatrix} v_{t-1} \\ v_t \end{bmatrix}.$$

Regarding the condition (iii), for some absolute constant M,

$$\mathbb{E}(x_t^T\xi_n)^4 \leq M\left\{\psi^4\mathbb{E}(w_{t-1}^T\xi_n)^4 + \mathbb{E}(w_t^T\xi_n)^4\right\} = M(\psi^4+1)\mathbb{E}(w_t^T\xi_n)^4,$$

where the right-hand-side is bounded in n as w_t follows our time-independent model above.

Appendix D.3. High dimensional AR(1) model

Consider the autoregressive model given by

$$x_t = \phi x_{t-1} + w_t$$

where $\phi \in (-1,1)$ is a scalar autoregressive coefficient and $w_t = \Sigma^{1/2}v_t$ follows the time-independent model in the first section. With a slight abuse of notation, again here Σ denotes the population covariance matrix of w_t rather that of x_t. Inverting the autoregressive process, we can represent x_t as an infinite-order moving average process given by

$$x_t = \sum_{\ell=0}^{\infty}\phi^\ell w_{t-\ell}. \tag{D.1}$$

We first check the condition (ii) in Theorem 1. Like in the above section, it suffices to show that $\lambda_{\max}(\mathbb{S}_n) = O(1)$. We can expand that

$$\begin{aligned}\mathbb{S}_n = \frac{1}{n}\sum_{t=1}^{n} x_t x_t^T = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{\ell=0}^{\infty}\phi^{\ell} w_{t-\ell}\right)\left(\sum_{l=0}^{\infty}\phi^{\ell} w_{t-\ell}\right)^T \\ = \sum_{\ell=0}^{\infty}\phi^{2\ell}\frac{1}{n}\sum_{t=1}^{n} w_{t-\ell}w_{t-\ell}^T + \sum_{0\le \ell_1\neq \ell_2}\phi^{\ell_1+\ell_2}\frac{1}{n}\sum_{t=1}^{n}\left(w_{t-\ell_1}w_{t-\ell_2}^T + w_{t-\ell_1}w_{t-\ell_2}^T\right) \\ =: \mathbb{S}_{n,1} + \mathbb{S}_{n,2}.\end{aligned}$$

Now note that $\lambda_{\max}\left(\frac{1}{n}\sum_{t=1}^{n} w_{t-\ell}w_{t-\ell}^T\right) \le C$ with probability 1 where the constant C does not depend on ℓ, and the set of non-negative integers is countable. It follows that

$$\lambda_{\max}\left(\mathbb{S}_{n,1}\right) = \sum_{l=0}^{\infty}|\phi|^{2\ell}\cdot O(1) = O(1).$$

Moreover, using similar argument in the last section, we can show that

$$\begin{aligned}\lambda_{\max}\left(\mathbb{S}_{n,2}\right) =& \lambda_{\max}\left(\sum_{0\le \ell_1\neq \ell_2}\phi^{\ell_1+\ell_2}\frac{1}{n}\sum_{t=1}^{n}\left(w_{t-\ell_1}w_{t-\ell_1}^T + w_{t-\ell_2}w_{t-\ell_2}^T\right)\right) \\ \le& 2\sum_{\ell=0}^{\infty}|\phi|^{\ell}\left(\frac{1}{1-|\phi|} - |\phi|^{\ell}\right)\lambda_{\max}\left(\frac{1}{n}\sum_{t=1}^{n}\left(w_{t-\ell}w_{t-\ell}^T\right)\right) \\ =& \left\{\sum_{\ell=0}^{\infty}|\phi|^{\ell}\left(\frac{1}{1-|\phi|} - |\phi|^{\ell}\right)\right\}\cdot O(1) = O(1).\end{aligned}$$

The condition then follows.

Next, we verify conditions (i)–(iii) in Theorem 2. We can deduce from above that $\lambda_{\max}(S_n) \le \lambda_{\max}\left(\mathbb{S}_n\right) = O(1)$, and $\lambda_{\max}\left(\mathbb{E}\left[x_t x_t^T\right]\right) = \frac{1}{1-\phi^2}\lambda_{\max}(\Sigma) = O(1)$. For condition (i), it remains to show that $\lambda_{\max}\left(\mathbb{E}[\bar{x}\bar{x}^T]\right) = O(1)$. It is easy to verify that

$$\mathbb{E}\left(x_t x_{t-\ell}^T\right) = \mathbb{E}\left(x_{t-\ell}x_t^T\right) = \frac{\phi^{\ell}}{1-\phi^2}\Sigma,\ l = 0, 1, \ldots.$$

Then

$$\begin{aligned}\mathbb{E}[\bar{x}\bar{x}^T] =& \frac{1}{n}\mathbb{E}(x_t x_t^T) + \frac{2}{n^2}\sum_{\ell=1}^{n-1}(n-\ell)\mathbb{E}(x_t x_{t-\ell}^T) \\ =& \frac{1}{n}\left(\frac{1}{1-\phi^2} + \frac{2}{n}\sum_{\ell=1}^{n-1}(n-\ell)\frac{\phi^{\ell}}{1-\phi^2}\right)\Sigma.\end{aligned}$$

Hence,

$$\begin{aligned}\lambda_{\max}\left(\mathbb{E}[\bar{x}\bar{x}^T]\right) =& \frac{1}{n}\left(\frac{1}{1-\phi^2} + \frac{2}{n}\sum_{\ell=1}^{n-1}(n-\ell)\frac{\phi^{\ell}}{1-\phi^2}\right)\lambda_{\max}\left(\Sigma\right) \\ \le& \frac{1}{n}\left(\frac{1}{1-\phi^2} + \frac{2|\phi|}{(1-\phi^2)^2}\right)\lambda_{\max}\left(\Sigma\right) \to 0.\end{aligned}$$

Observe that it also implies that $\mathbb{E}\left(\bar{x}^T\bar{x}\right) = O_{\mathbb{P}}\left(\frac{\mathrm{tr}(\Sigma)}{n}\right)$. Then invoking the proof of Proposition 1 and note that $\frac{1}{n}x_t^T x_t \le \lambda_{\max}(\mathbb{S}_n) = O(1)$, for condition (ii) it suffices to prove that

$$\left|\frac{1}{n}x_t^T x_t - \frac{1}{n}\mathbb{E}\left[x_t^T x_t\right]\right| \xrightarrow{\mathbb{P}} 0, \text{ for each } t. \tag{D.2}$$

Now take a diverging sequence $K = K(n) \in \{1, 2, \ldots\} \to \infty$. Truncate the moving average form (D.1) at order K to get the approximation

$$\widehat{x}_t = \sum_{\ell=0}^{K} \phi^\ell w_{t-\ell} = \left[\phi^K \Sigma^{1/2}, \ldots, \phi\Sigma^{1/2}, \Sigma^{1/2}\right] \begin{bmatrix} v_{t-K} \\ \vdots \\ v_t \end{bmatrix}.$$

Now following the proof of Proposition 1,

$$\left|\frac{1}{n}\widehat{x}_t^T \widehat{x}_t - \frac{1}{n}\mathbb{E}\left[\widehat{x}_t^T \widehat{x}_t\right]\right| \xrightarrow{\mathbb{P}} 0.$$

Let $R_t = x_t - \widehat{x}_t = \sum_{\ell=K+1}^{\infty} \phi^\ell w_{t-\ell}$. Using Cauchy–Schwarz inequality, it is easy to show that

$$\left|\frac{1}{n}x_t^T x_t - \frac{1}{n}\widehat{x}_t^T \widehat{x}_t\right| \le \frac{2}{n}\sqrt{R_t^T R_t \cdot \widehat{x}_t^T \widehat{x}_t} + \frac{1}{n}\widehat{x}_t^T \widehat{x}_t.$$

Using Jensen's inequality and independence between R_t and $\widehat{x}_t$,

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{n}x_t^T x_t - \frac{1}{n}\widehat{x}_t^T \widehat{x}_t\right| \le& \frac{2}{n}\sqrt{\mathbb{E}\left[R_t^T R_t\right] \cdot \mathbb{E}\left[\widehat{x}_t^T \widehat{x}_t\right]} + \frac{1}{n}\mathbb{E}\left[R_t^T R_t\right] \\
=& \frac{\mathrm{tr}(\Sigma)}{n}\left\{\sqrt{\sum_{l=K+1}^{\infty}\phi^{2l} \cdot \sum_{l=0}^{K}\phi^{2l}} + \sum_{l=K+1}^{\infty}\phi^{2l}\right\} \to 0.
\end{aligned}$$

Then (D.2) follows by the triangle inequality. This completes the proof for condition (ii). Recall from the first subsection that $\mathbb{E}\left(v_t^T \Sigma^{1/2}\xi_n\right)^4 = O(\lambda_{\max}^2(\Sigma))$. Finally, recalling the moving average form (D.1) again, for all unit vector ξ_n

$$\begin{aligned}
\mathbb{E}\left(x_t^T \xi_n\right)^4 =& \mathbb{E}\left(\sum_{l=0}^{\infty}\phi^l v_t^T \Sigma^{1/2}\xi_n\right)^4 \\
=& \left(\sum_{l=0}^{\infty}\phi^{4l}\right)\mathbb{E}\left(v_t^T \Sigma^{1/2}\xi_n\right)^4 + 3\left(\sum_{0\le l_1 \ne l_2}\phi^{2l_1+2l_2}\right)\left(\xi_n^T S_n \xi_n\right)^2 \\
=& O_{\mathbb{P}}\left(\lambda_{\max}^2(\Sigma)\right) + O_{\mathbb{P}}\left(\lambda_{\max}^2(\Sigma)\right),
\end{aligned}$$

which is clearly bounded in n.

Appendix E. Additional asymptotic theory

Appendix E.1. Testing a non-zero null value

Our theory generalizes for testing any given direction of β, say, $\xi_n^{(0)}$ with the composite null hypothesis

$$H_0 : \beta \propto \xi_n^{(0)}, \text{ that is, } H_0 : \beta = \widetilde{\theta}\xi_n^{(0)}, \text{ for some } \widetilde{\theta} \in \mathbb{R}, \tag{E.1}$$

where $\propto$ means 'is proportional to' and the constant $\widetilde{\theta} \in \mathbb{R}$ can be unspecified. Now, if we decompose the aggregate variable $x_t^T\beta$ under the alternatives as given by

$$x_t^T\beta = \widetilde{\theta}x_t^T\xi_n^{(0)} + x_t^T\widetilde{\beta},$$

where $x_t^T\widetilde{\beta}$ is uncorrelated with the high-dimensional projection $x_t^T\xi_n^{(0)}$ for every n. The regression model (2) in the main document can be rewritten as

$$y_t = z_t^T\theta + x_t^T(\widetilde{\theta}\xi_n^{(0)} + \widetilde{\beta}) + \varepsilon_t = \underline{z}_t^T\underline{\theta} + x_t^T\widetilde{\beta} + \varepsilon_t, \tag{E.2}$$

where the extended nuisance input vector $\underline{z}_t = (z_t^T, x_t^T\xi_n^{(0)})^T \in \mathbb{R}^{d+2}$ satisfies the general decomposition (3), with the extended nuisance coefficient vector given by $\underline{\theta} = (\theta^T, \widetilde{\theta})^T$. Hence, testing a non-zero null hypothesis (E.1) is equivalent to testing the zero null $H_0 : \widetilde{\beta} = 0$ under the re-parameterized model (E.2). One may apply the results for the universal model in Section 2.3 when testing a non-zero null. The correlation conditions are trivial given the orthogonality between $x_t^T\widetilde{\beta}$ and $x_t^T\xi_n^{(0)}$.

Appendix E.2. Adjusting for time variations

In this section, we continue the discussions in Remark 2 for the time-variation adjusted data. Define the adjusted design matrix as

$$\widetilde{X}_{\text{adj}} = [\widetilde{x}_{1,\text{adj}}, \ldots, \widetilde{x}_{n,\text{adj}}]^T = D_n^{-1/2}\widetilde{X}, \text{ with } D_n = \text{diag}\left(\frac{\|\widetilde{x}_1\|^2}{\text{tr}(S_n)}, \ldots, \frac{\|\widetilde{x}_n\|^2}{\text{tr}(S_n)}\right),$$

and the adjusted preliminary weighting matrix as

$$\underline{S}_{n,\text{adj}} = \frac{1}{n}\widetilde{X}_{\text{adj}}\widetilde{X}_{\text{adj}}^T = \frac{1}{n}D_n^{-1/2}\widetilde{X}\widetilde{X}^T D_n^{-1/2}.$$

Observe that the diagonal element

$$\underline{S}_{n,\text{adj}}(t,t) = \frac{1}{n}\widetilde{x}_{t,\text{adj}}^T\widetilde{x}_{t,\text{adj}} = \frac{1}{n}\,\text{tr}(S_n), \ \forall t = 1, \ldots, n.$$

However, as the true coefficient vector β is associated with raw data x_t rather than $\widetilde{x}_{t,\text{adj}}$, the expression of the asymptotic power changes in general. More specifically, if β is also free against the cross-product matrix $\check{S}_n := \frac{1}{n}\widetilde{X}^T\widetilde{X}_{\text{adj}} = \frac{1}{n}\widetilde{X}^T D_n^{-1/2}\widetilde{X}$, it is not very hard to show that

$$\varpi_n = \frac{\int x^2 dF^{\check{S}_n}(x) - \frac{p}{n}\left(\int x dF^{S_n}(x)\right)^2}{\sqrt{\int x^2 dF^{S_{n,\text{adj}}}(x) - \frac{p}{n}(\int x dF^{S_{n,\text{adj}}}(x))^2}},$$

where $S_{n,\mathrm{adj}} = \frac{1}{n}\widetilde{X}_{\mathrm{adj}}^T\widetilde{X}_{\mathrm{adj}} = \frac{1}{n}\widetilde{X}^T D_n^{-1}\widetilde{X}$ is the adjusted sample covariance matrix. Now, when $\|D_n - I_n\|_{sp} = \max_{1\le t\le n}\left|\frac{\widetilde{x}_t^T\widetilde{x}_t}{\mathrm{tr}(S_n)} - 1\right| \xrightarrow{\mathbb{P}} 0$, by Lemma 1 in El Karoui (2009) we can show that ϖ_n reduces to that for the free models asymptotically. In the most general case, the asymptotic departures depends on the time variations such as for the elliptical model in the aforementioned paper; see also Zheng and Li (2011).

Appendix E.3. Relaxing condition (i) of Theorem 7

In this last section, we discuss how to relax condition (i) of Theorem 7 in two different senses. The first way is to generalize the asymptotic theory in the absence of the condition. The second way is to show that the condition is fulfilled for the design matrix with separable covariance structure under Assumption 3. Besides, we note that the condition is also satisfied for non-separable design matrix if the entries of $\beta = (\beta_1, \dots, \beta_p)^T$ are independently generated with zero mean, a (nonzero) common variance and a bounded kurtosis by a direct application of the concentration inequality for quadratic forms such as Lemma B.26 in Bai and Silverstein (2010).

First, we comment on the generalization of Theorem 7 in the absence of condition (17) by allowing the asymptotic power to be dependent on the unknown direction of the regression coefficients. By carefully checking the proof of the theorem, it is easy to substitute the $\rho_n(\delta, 1)$ in the numerator of the asymptotic departure by

$$\check{\rho}_n(\delta, 1) = \check{\mu}_n^T\mu_n(\delta),$$

where

$$\check{\mu}_n = \frac{1}{n^{1/2}\left\|\widetilde{A}_n\right\|}\Omega^{-1/2}\left[0, \frac{p}{n}\xi_n^T\widetilde{X}^T\Psi_1^T\widetilde{X}\xi_n, \dots, \frac{p}{n}\xi_n^T\widetilde{X}^T\Psi_d^T\widetilde{X}\xi_n\right]^T.$$

and ξ_n denotes the direction of the regression coefficients. Note that the above statistics may depend on δ, if we use $\widetilde{A}_n(\delta)$ rather than $\widetilde{A}_n$ everywhere. That is,

$$\frac{Q_n(\delta)}{\sigma_n^2\sqrt{1-\rho_n^2(\delta)}} - \frac{h^2}{\sqrt{2}\sigma_n^2}\frac{\varpi_n(\delta) - \check{\rho}_n(\delta, 1)\cdot\varpi_n}{\sqrt{1-\rho_n^2(\delta)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

We may replace $\varpi_n(\delta)$ by the general form in Remark 1, if we relax the freeness Assumption 3 as well. We omit the proofs. As we argued, the asymptotic limit becomes intractable to produce an interesting theory here, and thus we leave more detailed analysis for future study.

Next, we show that condition (i) of Theorem 7 is implied by Assumption 3 if the data matrix X has a separable covariance structure. In particular, we consider $X = \Upsilon^{1/2}F\Sigma^{1/2}$ where $\Upsilon \in \mathbb{R}^{n\times n}$ and $\Sigma \in \mathbb{R}^{p\times p}$ are the temporal and cross-sectional matrices with bounded spectral norms. The latent random matrix $F = [f_1, \dots, f_n]^T = \{f_{t,i}\}$ has independent entries with $\mathbb{E}f_{t,i} = 0$, $\mathbb{E}f_{t,i}^2 = 1$, and $\sup_{t,i}\mathbb{E}f_{t,i}^4 \le \nu_4$. Let $C \in \mathbb{R}^{n\times n}$ denote the centering matrix and $\Gamma_n := \frac{1}{n}\widetilde{X}^T\Psi_i^T\widetilde{X}$. It suffices to show that:

$$\Delta := \left\{\xi_n^T\Gamma_n\xi_n - \frac{1}{p}\,\mathrm{tr}\,\Gamma_n\right\} - \frac{\mathrm{tr}\left(\Upsilon^{1/2}C\Psi_i^TC\Upsilon^{1/2}\right)}{\mathrm{tr}\left(\Upsilon^{1/2}C\Upsilon^{1/2}\right)}\left\{\xi_n^TS_n\xi_n - \frac{1}{p}\mathrm{tr}(S_n)\right\} \xrightarrow{\mathbb{P}} 0,$$

because $\xi_n^T S_n \xi_n - \frac{1}{p}\text{tr}(S_n) \xrightarrow{\mathbb{P}} 0$ by Assumption 3 and $\left|\frac{\text{tr}\left(\Upsilon^{1/2}C\Psi_i^T C\Upsilon^{1/2}\right)}{\text{tr}\left(\Upsilon^{1/2}C\Upsilon^{1/2}\right)}\right| \le \left\|\Psi_i^T\right\|_{sp} = O(1)$ by Lemma 3 in the main document.

By a direct calculation,

$$\begin{aligned}\mathbb{E}\Delta =& \text{tr}\left(\Upsilon^{1/2}C\Psi_i^T C\Upsilon^{1/2}\right)\left(\xi_n^T\Sigma\xi_n - \frac{1}{p}\text{tr}\Sigma\right) \\ &- \frac{\text{tr}\left(\Upsilon^{1/2}C\Psi_i^T C\Upsilon^{1/2}\right)}{\text{tr}\left(\Upsilon^{1/2}C\Upsilon^{1/2}\right)} \cdot \text{tr}\left(\Upsilon^{1/2}C\Upsilon^{1/2}\right)\left(\xi_n^T\Sigma\xi_n - \frac{1}{p}\text{tr}\Sigma\right) = 0.\end{aligned}$$

It suffices to show that $\text{var}(\Delta) \to 0$ or to show that:

$$\text{var}(\xi_n^T\Gamma_n\xi_n),\ \text{var}\left(\frac{1}{p}\text{tr}\Gamma_n\right),\ \text{var}(\xi_n^T S_n\xi_n),\ \text{var}\left(\frac{1}{p}\text{tr}S_n\right) \to 0.$$

We only prove the first two parts, and the proofs for the last two are completely analogous (by replacing Ψ_i^T with identity matrix everywhere). Observe that

$$\xi_n^T\Gamma_n\xi_n = \frac{1}{n}\xi_n^T\Sigma^{1/2}F^T\Upsilon^{1/2}C\Psi_i^T C\Upsilon^{1/2}F\Sigma^{1/2}\xi_n = V^TAV, \tag{E.3}$$

where $A = \Upsilon^{1/2}C\Psi_i^T C\Upsilon^{1/2}$ and $V = [V_1, \ldots, V_n]^T$ with $V_t = \frac{1}{\sqrt{n}}f_t^T\Sigma^{1/2}\xi_n$. Note that V_i are independent, with $\mathbb{E}V_t = 0$ and $\mathbb{E}V_t^2 = \frac{1}{n}\xi_n^T\Sigma\xi_n$. Because V_t is a quadratic form of f_t, applying Lemma B.26 in Bai and Silverstein (2010), or Lemma 2 in our main document, yields that

$$\begin{aligned}\mathbb{E}V_i^4 =& (\mathbb{E}V_i^2)^2 + \mathbb{E}(V_i^2 - \mathbb{E}V_i^2)^2 \\ \le& \left(\frac{1}{n}\xi_n^T\Sigma\xi_n\right)^2 + M\nu_4\left(\frac{1}{n}\xi_n^T\Sigma\xi_n\right)^2 \le \frac{M}{n^2}(\xi_n^T\Sigma\xi_n)^2 \le Mn^{-2}\left\|\Sigma\right\|_{sp}^2,\end{aligned}$$

where M is some absolute constant not depending n nor ξ_n. Applying Lemma B.26 in Bai and Silverstein (2010), or Lemma 2 in our main document, again but to the quadratic form (E.3),

$$\text{var}\left(\xi_n^T\Gamma_n\xi_n\right) \le M\mathbb{E}V_i^4 \cdot \text{tr}A^2 \le Mn^{-2}\left\|\Sigma\right\|_{sp}^2 \cdot p\left\|A\right\|_{sp}^2 = O(p/n^2) \to 0,$$

where we recall that $\left\|\Sigma\right\|_{sp} = O(1)$ and $\left\|A\right\|_{sp} \le \left\|\Upsilon\right\|_{sp}\left\|C\right\|_{sp}^2\left\|\Psi_i^T\right\|_{sp} = O(1)$ in the last equality and M is a possibly different absolute constant. This completes the proof of $\text{var}(\xi_n^T\Gamma_n\xi_n) \to 0$. Note that the proofs above hold for any given unit vector ξ_n. It means that, for every standard basis vector $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)^T$ of $\mathbb{R}^p$, we also have that

$$\text{var}\left(\mathbf{e}_i^T\Gamma_n\mathbf{e}_i\right) \le Mn^{-2}\left\|\Sigma\right\|_{sp}^2 \cdot p\left\|A\right\|_{sp}^2, \quad i = 1, \ldots, p.$$

It follows that

$$\text{var}\left(\frac{1}{p}\text{tr}\Gamma_n\right) \le \frac{1}{p}\sum_{i=1}^p \text{var}(\mathbf{e}_i'\Gamma_n\mathbf{e}_i) \le Mn^{-2}\left\|\Sigma\right\|_{sp}^2 \cdot p\left\|A\right\|_{sp}^2 = O(p/n^2) \to 0.$$

Appendix F. Proof of Lemmas 1 – 14

Appendix F.1. Proof of Lemmas 1–3

Proof of Lemma 1. The lemma is straightforward by combining Markov inequality and the law of iterated expectations. We omit the details. □

Proof of Lemma 2. Let $A = \{A(s,t) : s,t = 1,\ldots,n\}$, where $A(s,t)$ denotes the entry of A in its s-th row and t-th column. Expanding the quadratic form,

$$\epsilon^T A\epsilon - \mathrm{tr}(A) = \sum_{t=1}^{n}(\varepsilon_t^2 - 1)A(t,t) + \sum_{1\le s<t\le n} \varepsilon_s\varepsilon_t\left(A(s,t) + A(t,s)\right) =: T_1 + T_2.$$

By Burkholder's inequality (e.g., Theorem 2.10 in Hall and Heyde, 1980), for some constant M

$$\mathbb{E}\left[|T_1|^{1+\iota} \mid \mathcal{F}_{n,0}\right] \le M\sum_{t=1}^{n}\mathbb{E}\left[|\varepsilon_t^2 - 1|^{1+\iota}|\mathcal{F}_{n,0}\right]|A(t,t)|^{1+\iota} \le M\kappa_n \cdot \sum_{t=1}^{n}|A(t,t)|^{1+\iota}.$$

Moreover, by a direct calculation and applying Cauchy–Schwarz inequality,

$$\begin{aligned}\mathbb{E}\left[T_2^2|\mathcal{F}_{n,0}\right] = \sum_{1\le s<t\le n}\left(A(s,t) + A(t,s)\right)^2 \le& 2\sum_{1\le s<t\le n}\left(A^2(s,t) + A^2(t,s)\right)\\ =& 2\left\|A - \mathrm{diag}(A)\right\|^2 \le 2\left\|A\right\|^2.\end{aligned}$$

Hence, using Jensen's inequality,

$$\begin{aligned}\mathbb{E}\left[|\epsilon^T A\epsilon - \mathrm{tr}(A)|^{1+\iota}|\mathcal{F}_{n,0}\right] \le& M\mathbb{E}\left[|T_1|^{1+\iota} + |T_2|^{1+\iota}|\mathcal{F}_{n,0}\right]\\ \le& M\mathbb{E}\left[|T_1|^{1+\iota}|\mathcal{F}_{n,0}\right] + M\left(\mathbb{E}\left[|T_2|^2|\mathcal{F}_{n,0}\right]\right)^{(1+\iota)/2}\\ \le& M\kappa_n \cdot \sum_{t=1}^{n}|A(t,t)|^{1+\iota} + M\left\|A\right\|^{1+\iota},\end{aligned}$$

where the constant M may be different in different inequalities. This is the first part of the lemma. For the rest we invoke Lemma 1 to get

$$\left|\epsilon^T A\epsilon - \mathrm{tr}(A)\right|^{1+\iota} = O_{\mathbb{P}}\left(\kappa_n\sum_{t=1}^{n}|A(t,t)|^{1+\iota} + \|A\|^{1+\iota}\right),$$

or equivalently

$$\left|\epsilon^T A\epsilon - \mathrm{tr}(A)\right| = O_{\mathbb{P}}\left(\kappa_n^{\frac{1}{1+\iota}}\left(\sum_{t=1}^{n}|A(t,t)|^{1+\iota}\right)^{\frac{1}{1+\iota}} + \|A\|\right).$$

The rest follows from the obvious inequality $|A(t,t)|^{1+\iota} \le |A(t,t)| \cdot \max_{1\le t\le n}|A(t,t)|^{\iota}$ and the triangle inequality. □

Proof of Lemma 3. Slightly abusing the notation, let $U\Lambda U^T$ be a spectral decomposition of A where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ is a diagonal matrix consists of the eigenvalues and $U = [u_1, \ldots, u_p]$ is an orthogonal matrix with columns being the corresponding eigenvectors. Hence, $A = \sum_{j=1}^p \lambda_j u_j u_j^T$, $\sum_{j=1}^p u_j u_j^T = UU^T = I$ and $\|A\|_{sp} = \max_j |\lambda_j|$. Then, noting that B is nonnegative definite,

$$
\begin{aligned}
|\text{tr}(AB)| &= \left| \sum_{j=1}^p \lambda_j u_j^T B u_j \right| \\
&\leq \sum_{j=1}^p |\lambda_j| u_j^T B u_j \leq \|A\|_{sp} \sum_{j=1}^p u_j^T B u_j = \|A\|_{sp} \,\text{tr}\left(B \sum_{j=1}^p u_j u_j^T \right) = \|A\|_{sp} \,\text{tr}\left(B\right).
\end{aligned}
$$

When $B = B'$ is symmetric, the lemma holds for arbitrary matrix A because

$$
\text{tr}(AB) = \frac{1}{2}\left(\text{tr}\left(AB\right) + \text{tr}\left(BA'\right)\right) = \text{tr}\left(\left(\frac{1}{2}(A + A')\right) B\right).
$$

The lemmas follows by replacing A with the symmetric matrix $\frac{1}{2}(A + A')$. □

Appendix F.2. Proof of Lemma 4

Let $b_t = (0, \ldots, 0, 1, 0, \ldots, 0)^T \in \mathbb{R}^n$ denote the unit vector with t-th entry equaling to 1 and all other entries equaling to 0. Rewrite the conditional variance into a quadratic form given by

$$
\mathbb{E}\left[\Delta_t^2 \mid \mathcal{F}_{n,t-1}\right] = \frac{2}{\|A_n\|^2} \left(\sum_{s=1}^{t-1} \varepsilon_s A(s,t) \right)^2 = \frac{1}{\left\|\widetilde{A}_n\right\|^2} \epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon.
$$

It suffices to show that $\max_{1 \leq t \leq n} \epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon = o_{\mathbb{P}}\left(\left\|\widetilde{A}_n\right\|^2\right)$. From Lemma 2,

$$
\begin{aligned}
&\mathbb{E}\left[\left|\epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon - b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t\right|^{1+\iota} \mid \mathcal{F}_{n,0}\right] \\
&\leq M \left(\kappa_n \sum_{s=1}^{t-1} |A_n(s,t)|^{2(1+\iota)} + (b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t)^{1+\iota} \right).
\end{aligned}
$$

Summing up over t and recalling the assumption that $\kappa_n = O_{\mathbb{P}}(1)$, it follows that

$$
\sum_{t=1}^n \mathbb{E}\left[\left|\epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon - b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t\right|^{1+\iota} \mid \mathcal{F}_{n,0}\right] = O_{\mathbb{P}}\left(\sum_{t=1}^n (b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t)^{1+\iota} \right),
$$

where we have also used the Jensen's inequality

$$
\sum_{t=1}^n \sum_{s=1}^{t-1} |A_n(s,t)|^{2(1+\iota)} \leq \sum_{t=1}^n \left(\sum_{s=1}^{t-1} A_n^2(s,t) \right)^{1+\iota} = \sum_{t=1}^n (b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t)^{1+\iota}.
$$

Note that $b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t$ is the t-th diagonal element of the matrix $\widetilde{A}_n \widetilde{A}_n^T$, $t = 1, \ldots, n$ and they are majorized by the eigenvalues (see, e.g., Theorem 4.3.45 in Horn and Johnson, 2012). Combining with the trace inequality (Lemma 3) and condition (ii) yields that

$$\sum_{t=1}^{n} (b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t)^{1+\iota} \leq \operatorname{tr}\left(\widetilde{A}_n \widetilde{A}_n^T\right)^{1+\iota}$$
$$= \operatorname{tr}\left(\widetilde{A}_n^T \widetilde{A}_n\right)^{1+\iota} \leq \lambda_{\max}^{\iota}\left(\widetilde{A}_n^T \widetilde{A}_n\right) \operatorname{tr}\left(\widetilde{A}_n^T \widetilde{A}_n\right) = o_{\mathbb{P}}\left(\left\|\widetilde{A}_n\right\|^{2\iota+2}\right).$$

Hence,

$$\mathbb{E}\left[\max_{1\leq t\leq n}\left|\epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon - b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t\right|^{1+\iota} \mid \mathcal{F}_{n,0}\right]$$
$$\leq \sum_{t=1}^{n} \mathbb{E}\left[\left|\epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon - b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t\right|^{1+\iota} \mid \mathcal{F}_{n,0}\right] = o_{\mathbb{P}}(\|A_n\|^{2\iota+2}).$$

It then follows from Lemma 1 that

$$\max_{1\leq t\leq n}\left|\epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon - \operatorname{tr}\left(\widetilde{A}_n^T b_t b_t^T \widetilde{A}_n\right)\right|^{1+\iota} = o_{\mathbb{P}}\left(\|A_n\|^{2\iota+2}\right),$$

or equivalently

$$\max_{1\leq t\leq n}\left|\epsilon^T \widetilde{A}_n^T b_t b_t^T \widetilde{A}_n \epsilon - b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t\right| = o_{\mathbb{P}}\left(\|A_n\|^2\right).$$

Using the definition of spectral norm,

$$\max_{1\leq t\leq n}\left|b_t^T \widetilde{A}_n \widetilde{A}_n^T b_t\right| \leq \lambda_{\max}\left(\widetilde{A}_n \widetilde{A}_n^T\right) = o_{\mathbb{P}}(\|A_n\|^2).$$

The rest follows by the triangular inequality.

Appendix F.3. Proof of Lemma 5

By Lemma 2 and recalling that $\kappa_n = O_{\mathbb{P}}(1)$,

$$\mathbb{E}\left[\left|\frac{2}{\|A_n\|^2}\sum_{t=1}^{n}\left(\sum_{s=1}^{t-1}\varepsilon_s \frac{1}{n}\widetilde{x}_s^T \widetilde{x}_t\right)^2 - 1\right|^{1+\iota} \mid \mathcal{F}_{n,0}\right]$$
$$= \mathbb{E}\left[\left|\frac{2}{\|A_n\|^2}\epsilon^T \widetilde{A}_n^T \widetilde{A}_n \epsilon - 1\right|^{1+\iota} \mid \mathcal{F}_{n,0}\right] \leq M\left(\frac{\sum_{t=1}^{n}\left(\widetilde{A}_n^T \widetilde{A}_n(t,t)\right)^{1+\iota}}{\|A_n\|^{2(1+\iota)}} + \frac{\left\|\widetilde{A}_n^T \widetilde{A}_n\right\|^{1+\iota}}{\|A_n\|^{2(1+\iota)}}\right),$$

where $\widetilde{A}_n^T \widetilde{A}_n(t,t)$ denotes the t-th diagonal element of $\widetilde{A}_n^T \widetilde{A}_n$. Using the majority property of eigenvalues against the diagonal elements and the trace inequality (Lemma 3),

$$\sum_{t=1}^{n}\left(\widetilde{A}_n^T \widetilde{A}_n(t,t)\right)^{1+\iota} \leq \operatorname{tr}\left(\widetilde{A}_n^T \widetilde{A}_n\right)^{1+\iota} \leq \lambda_{\max}^{\iota} \operatorname{tr}\left(\widetilde{A}_n^T \widetilde{A}_n\right) = o_{\mathbb{P}}\left(\left\|\widetilde{A}_n\right\|^{2(1+\iota)}\right).$$

On the other hand, using the the trace inequality (Lemma 3) again,

$$\left\|\widetilde{A}_n^T\widetilde{A}_n\right\| = \sqrt{\operatorname{tr}\left(\widetilde{A}_n^T\widetilde{A}_n\right)^2} \le \sqrt{\lambda_{\max}\left(\widetilde{A}_n^T\widetilde{A}_n\right)\operatorname{tr}\left(\widetilde{A}_n^T\widetilde{A}_n\right)} = o_{\mathbb{P}}\left(\left\|\widetilde{A}_n\right\|^2\right).$$

Using Lemma 1, $\left|\frac{2}{\|A_n\|^2}\sum_{t=1}^n\left(\sum_{s=1}^{t-1}\varepsilon_s\frac{1}{n}\widetilde{x}_s^T\widetilde{x}_t\right)^2 - 1\right|^{1+\iota} \xrightarrow{\mathbb{P}} 0$ and lemma follows.

Appendix F.4. Proof of Lemma 6

Without loss of generality, we may assume that $\mathbb{E}z_{t,i}^2 = 1$ by proper marginal scaling. Expanding the quadratic form,

$$\frac{1}{n}\epsilon^T ZZ^T\epsilon = \frac{1}{n}\sum_{t=1}^n \varepsilon_t^2 z_t^T z_t + \frac{1}{n}\sum_{t\neq s}\varepsilon_t\varepsilon_s z_t^T z_s.$$

Taking the expectation on both sides and using the law of iterated expectations,

$$\mathbb{E}\left[\frac{1}{n}\epsilon^T ZZ^T\epsilon\right] = \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[z_t^T z_t\right] = \sum_{i=1}^d \mathbb{E}\left[z_{t,i}^2\right] = O(d).$$

It follows from Lemma 1 that $\frac{1}{n}\epsilon^T ZZ^T\epsilon = O_{\mathbb{P}}(d)$. Finally,

$$\epsilon^T P_Z\epsilon = \frac{1}{n}\epsilon^T Z\widehat{\Omega}^{-1}Z^T\epsilon \le \lambda_{\min}^{-1}(\widehat{\Omega})\cdot\frac{1}{n}\epsilon^T ZZ^T\epsilon = O_{\mathbb{P}}(1/\lambda_{\min}(\widehat{\Omega}))\cdot O_{\mathbb{P}}(d),$$

using the definition of spectral norm.

Appendix F.5. Proof of Lemma 7

Let $\zeta := (\zeta_1,\ldots,\zeta_{d+1})^T := Z^T\widetilde{A}_n\epsilon$. It suffices to show that $\zeta_i = o_{\mathbb{P}}\left(\sqrt{n}\left\|\widetilde{A}_n\right\|\right)$ for each i. We invert the autoregressive process (under the null) into a moving average form given by

$$y_t = \alpha + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}, \tag{F.1}$$

where $\alpha = \theta_0\cdot\sum_{j=0}^{\infty}\psi_j$ and the sequence $\{\psi_j\}$ is absolutely summable, that is, $\sum_{j=0}^{\infty}|\psi_j| < \infty$ by Proposition 6.3 in Hayashi (2000). Now, for $i = 1,\ldots,d$, let the vector of lagged observations be

$$\mathbf{y}_{-i} = [y_{1-i},\ldots,y_{t-i},\ldots,y_{n-i}]^T = \alpha\mathbf{1}_n + \Psi_i\epsilon + \mathbf{v}_i, \tag{F.2}$$

where, like equation (13) in the main document,

$$\Psi_i = \sum_{j=0}^{\infty}\psi_j L_n^{i+j} = \sum_{j=0}^{n-i}\psi_j L_n^{i+j} \tag{F.3}$$

and L_n is the $n\times n$ lower shift matrix with ones on the subdiagonal and zeros elsewhere, and the entries of $\mathbf{v}_i = (v_{1,i},\ldots,v_{n,i})$ have bounded variances in n and, as a result, $\mathbf{v}_i^T\mathbf{v}_i = O_{\mathbb{P}}(\sum_{i=1}^n v_{n,i}^2) = O_{\mathbb{P}}(n)$.

Now, by a direct expansion, $\zeta_1 = \mathbf{1}_n^T \widetilde{A}_n \epsilon$ and

$$\zeta_{i+1} = \mathbf{y}_{-i}^T \widetilde{A}_n \epsilon = \alpha\zeta_1 + \epsilon^T \Psi_i^T \widetilde{A}_n \epsilon + \epsilon^T \widetilde{A}_n^T \mathbf{v}_i =: \alpha\zeta_1 + J_1 + J_2.$$

Using Lemma 1 and a direct calculation yields that

$$\zeta_1^2 = O_{\mathbb{P}}\left(\mathbb{E}\left[\zeta_1^2 \mid \mathcal{F}_{n,0}\right]\right) = O_{\mathbb{P}}\left(\mathbf{1}_n^T \widetilde{A}_n \widetilde{A}_n^T \mathbf{1}_n\right) = O_{\mathbb{P}}\left(n\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)\right). \tag{F.4}$$

Taking square root of both sides and using condition (ii) gives $\zeta_1 = o_{\mathbb{P}}\left(\sqrt{n}\left\|\widetilde{A}_n\right\|\right)$. To bound J_1, we first use the last part of Lemma 2 (taking $\iota = 0$ therein) to get that

$$J_1 = O_{\mathbb{P}}\left\{\sum_{t=1}^{n} |d_{i,t}| + \left\|\Psi_i^T \widetilde{A}_n\right\|\right\} =: O_{\mathbb{P}}\left(J_{1,1} + J_{1,2}\right),$$

where $d_{i,t} := \sum_{j=0}^{n-t-i} \psi_j A_n(t+i+j,t)$ is the t-th diagonal element of $\Psi_i^T \widetilde{A}_n = \sum_{j=0}^{n-i} \psi_j L_n^{i+j} \widetilde{A}_n$. By the triangle inequality and exchanging the order of summations,

$$\sum_{t=1}^{n} |d_{i,t}| \le \sum_{t=1}^{n} \sum_{j=0}^{n-t-i} |\psi_j| \, |A_n(t+i+j,t)| = \sum_{j=0}^{n-i} |\psi_j| \sum_{t=1}^{n-i-j} |A_n(t+i+j,t)| \,. \tag{F.5}$$

Let $\varsigma > 0$. By choosing a sufficiently large K and for a large constant M not depending on K nor n, we have

$$\begin{aligned}\sum_{t=1}^{n} |d_{i,t}| \le & \sum_{j=0}^{K} |\psi_j| \cdot \sum_{t=1}^{n-i-j} |A_n(t+i+j,t)| + \sum_{j=K+1}^{n-i} |\psi_j| \cdot \sum_{t=1}^{n-i-j} |A_n(t+i+j,t)| \\ \le & M \max_{1\le \ell \le K+i} \sum_{t=1}^{n-\ell} |A_n(t+\ell,t)| + \varsigma \cdot \max_{\ell > K} \sum_{t=1}^{n-\ell} |A_n(t+\ell,t)| \,.\end{aligned}$$

Furthermore, by Cauchy–Schwarz inequality,

$$\max_{\ell > K} \sum_{t=1}^{n-\ell} |A_n(t+\ell,t)| \le \sqrt{n-\ell}\sqrt{\max_{\ell > K} \sum_{t=1}^{n-\ell} A_n^2(t+\ell,t)} \le \sqrt{n}\left\|A_n\right\| .$$

Using condition (iii) and noting that ς can be arbitrarily small, we can show that

$$J_{1,1} = \sum_{t=1}^{n} |d_{i,t}| = o_{\mathbb{P}}(n^{1/2}\left\|A_n\right\|). \tag{F.6}$$

Next, applying the trace inequality (Lemma 3) and the triangle equality,

$$J_{1,2}^2 \le \left\|\Psi_i \Psi_i^T\right\|_{sp} \cdot \operatorname{tr}\left(\widetilde{A}_n \widetilde{A}_n^T\right) \le \left(\sum_{j=0}^{n-i} |\psi_j|\right)^2 \cdot \frac{1}{2}\left\|A_n\right\|^2 = O_{\mathbb{P}}\left(\left\|A_n\right\|^2\right). \tag{F.7}$$

Combining the bounds of ζ_1, $J_{1,1}$ and $J_{1,2}$, we can immediately conclude that

$$J_1 = o_{\mathbb{P}}(\sqrt{n}\,\|A_n\|) + o_{\mathbb{P}}(\sqrt{n}\,\|A_n\|) + o_{\mathbb{P}}(\|A_n\|) = o_{\mathbb{P}}(\sqrt{n}\,\|A_n\|).$$

Finally, using the square summability of $\mathbf{v}_i$ and condition (ii),

$$\mathbb{E}\left[J_2^2 \mid \mathcal{F}_{n,0}\right] = \mathbf{v}_i^T \widetilde{A}_n \widetilde{A}_n^T \mathbf{v}_i \le \mathbf{v}_i^T \mathbf{v}_i \cdot \lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right) = o_{\mathbb{P}}\left(n \left\|\widetilde{A}_n\right\|^2\right). \tag{F.8}$$

It follows from Lemma 1 that $J_2 = o_{\mathbb{P}}\left(\left\|\widetilde{A}_n\right\|\right)$. Our proof is now complete.

Appendix F.6. Proof of Lemma 8

The proof is very similar to that of Lemma 7, and hence we only sketch the differences. Under the alternatives, we replace (F.2) by

$$\mathbf{y}_{-i} = [y_{1-i}, \ldots, y_{t-i}, \ldots, y_{n-i}]^T = \alpha \mathbf{1}_n + \Psi_i \epsilon + \Psi_i X\beta + \mathbf{v}_i,\ i = 1, \ldots, d, \tag{F.9}$$

where Ψ_i is the same as in (F.3), and $\mathbf{v}_i = (v_{i,1}, \ldots, v_{i,n})$ depending on both $\{x_t^T\beta : t \le 0\}$ and $\{\varepsilon_t : t \le 0\}$. This introduces an additional term $J_3 := \frac{1}{\sqrt{n}\|\widetilde{A}_n\|}\beta^T X \Psi_i^T \widetilde{A}_n \epsilon$ in ζ_{i+1}, and it remains to show that, for any i,

$$J_3 = o_{\mathbb{P}}\left(\|\beta\|^2\right).$$

By a direct calculation,

$$\mathbb{E}\left[J_3^2 \mid X\right] = \frac{1}{n\left\|\widetilde{A}_n\right\|^2}\beta^T X^T \Psi_i^T \widetilde{A}_n \widetilde{A}_n^T \Psi_i X\beta \le \frac{\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)}{\left\|\widetilde{A}_n\right\|^2}\left\|\Psi_i^T \Psi_i\right\|_{sp} \beta^T \mathbb{S}_n \beta.$$

Note that

$$\left\|\Psi_i^T \Psi_i\right\|_{sp} \le \left(\sum_{j=0}^{\infty} |\psi_j|\right)^2 < \infty. \tag{F.10}$$

On the other hand, $\beta^T \mathbb{S}_n \beta = \beta^T S_n \beta - \beta^T \bar{x}\bar{x}^T \beta = O_{\mathbb{P}}\left(\|\beta\|^2\right)$ because

$$\beta^T S_n \beta = \|\beta\|^2 \int x dF^{S_n}(x;\beta) = \|\beta\|^2 \left(\int x dF^{S_n}(x) + o_{\mathbb{P}}(1)\right) = O_{\mathbb{P}}\left(\|\beta\|^2\right), \tag{F.11}$$

and, by Lemma 1,

$$\beta^T \bar{x}\bar{x}^T \beta = O_{\mathbb{P}}\left(\beta^T \mathbb{E}\left[\bar{x}\bar{x}^T\right]\beta\right) = O\left(\|\beta\|^2\right). \tag{F.12}$$

This completes the proof.

Appendix F.7. Proof of Lemma 9

Decompose that

$$\beta^T \widetilde{X}^T Z Z^T \widetilde{X} \beta = \sum_{i=1}^{d} \left(\mathbf{y}_{-i}^T \widetilde{X} \beta \right)^2 =: \sum_{i=1}^{d} \zeta_i^2,$$

with the vector $\mathbf{y}_{-i} = [y_{1-i}, \ldots, y_{n-i}]^T$. It suffices to prove that, for each i

$$\zeta_i^2 = O_p \left(n \left\| \beta \right\|^2 + n^2 \left\| \beta \right\|^4 \right).$$

Plugging in the expansion (F.9),

$$\zeta_i = \epsilon^T \Psi_i^T \widetilde{X} \beta + \beta^T \widetilde{X}^T \Psi_i^T \widetilde{X} \beta + \beta^T \bar{x} \mathbf{1}_n^T \Psi_i^T \widetilde{X} \beta + \mathbf{v}_i^T \widetilde{X} \beta =: \zeta_{i,1} + \zeta_{i,2} + \zeta_{i,3} + \zeta_{i,4}.$$

Now, invoking (F.10),

$$\mathbb{E}\left[\zeta_{i,1}^2 \mid X \right] = \beta^T \widetilde{X}^T \Psi_i^T \Psi_i \widetilde{X} \beta \le n \left\| \Psi_i^T \Psi_i \right\|_{sp} \beta^T S_n \beta = O_{\mathbb{P}} \left(n \left\| \beta \right\|^2 \right).$$

It follows from Lemma 1 that $\zeta_{i,1}^2 = O_{\mathbb{P}} \left(n \left\| \beta \right\|^2 \right)$. Similarly, by Cauchy–Schwarz inequality and the definition of spectral norm,

$$\zeta_{i,2}^2 \le n^2 \left\| \Psi_i^T \Psi_i \right\|_{sp} \left(\beta^T S_n \beta \right)^2 = O_{\mathbb{P}} \left(n^2 \left\| \beta \right\|^4 \right),$$

and

$$\zeta_{i,3}^2 \le n^2 \left(\bar{x}^T \beta \right)^2 \left\| \Psi_i^T \Psi_i \right\|_{sp} \beta^T S_n \beta = O_{\mathbb{P}} \left(n^2 \left\| \beta \right\|^4 \right),$$

where in the last step we invoke (F.12) as well. Finally, by Cauchy–Schwarz inequality,

$$\zeta_{i,4}^2 \le \beta^T \widetilde{X}^T \widetilde{X} \beta \cdot \mathbf{v}_i^T \mathbf{v}_i = n \beta^T S_n \beta \cdot O_{\mathbb{P}}(1) = O_{\mathbb{P}} \left(n \left\| \beta \right\|^2 \right), \tag{F.13}$$

by recalling that $\mathbf{v}_i^T \mathbf{v}_i = O_{\mathbb{P}}(1)$ because the entries $\{v_{t,i}\}$ in expansion (F.9) satisfy the linear difference equation $v_{t,i} = \sum_{\ell=1}^{d} \theta_\ell v_{t-\ell,i}$.

Appendix F.8. Proof of Lemma 10

By the definition of spectral norm,

$$\mu_n^T \mu_n \le \frac{\left(\lambda_{\min}(\Omega) \right)^{-1} \cdot \sum_{i=1}^{d} \left(\operatorname{tr}^2 (\Psi_i^T \widetilde{A}_n) \right)}{n \left\| \widetilde{A}_n \right\|^2}. \tag{F.14}$$

Similar to (F.5), exchanging order of summations and using triangle inequality,

$$\left| \operatorname{tr} \left(\Psi_i^T \widetilde{A}_n \right) \right| = \left| \sum_{t=1}^{n} \sum_{\ell=1}^{n-t} \psi_i(\ell) A_n(t+\ell, t) \right| \le \sum_{l=1}^{n-t} \left| \psi_i(\ell) \right| \left| \sum_{t=1}^{n-\ell} A_n(t+\ell, t) \right|$$

Following the proof of statement (F.6) and using condition (12), we can show that

$$\left| \operatorname{tr} \left(\Psi_i^T \widetilde{A}_n \right) \right| = o_{\mathbb{P}} \left(n^{1/2} \left\| \widetilde{A}_n \right\| \right),$$

that is, $\operatorname{tr}^2(\Psi_i^T \widetilde{A}_n) / \left(n \left\| \widetilde{A}_n \right\|^2 \right) \xrightarrow{\mathbb{P}} 0$. Summing over i and combining with (F.14) completes the proof.

Appendix F.9. Proof of Lemma 11

It suffices to show that every entry of $\zeta := \frac{1}{\sqrt{n}\|A_n\|} Z^T \widetilde{A}_n \epsilon - \Omega^{1/2}\mu_n \in \mathbb{R}^{d+1}$ converges to 0 in probability. Denote the observations for the i-th predictor, $i = 1, \dots, d$ by

$$\mathbf{z}_i := (z_{1,i}, \dots, z_{n,i})^T = \alpha_i \mathbf{1}_n + \Psi_i \epsilon + \mathbf{v}_i, \tag{F.15}$$

where Ψ_i is given in equation (13) in the main document. Denote the i-th entry of ζ by ζ_i, and then a direct calculation yields that $\zeta_1 = \frac{1}{\sqrt{n}\|\widetilde{A}_n\|} \mathbf{1}_n^T \widetilde{A}_n \epsilon$ and for $i = 1, \dots, d$

$$\begin{aligned}
\zeta_{i+1} =& \frac{1}{\sqrt{n}\left\|\widetilde{A}_n\right\|}\left(\mathbf{z}_i^T \widetilde{A}_n \epsilon - \operatorname{tr}\left(\Psi_i^T \widetilde{A}_n\right)\right) \\
=& \alpha_i \zeta_1 + \frac{1}{\sqrt{n}\left\|\widetilde{A}_n\right\|}\left(\epsilon^T \Psi_i^T \widetilde{A}_n \epsilon - \operatorname{tr}\left(\Psi_i^T \widetilde{A}_n\right)\right) + \frac{1}{\sqrt{n}\left\|\widetilde{A}_n\right\|} \mathbf{v}_i^T \widetilde{A}_n \epsilon =: \alpha_i \zeta_1 + \zeta_{i+1,1} + \zeta_{i+1,2}.
\end{aligned}$$

Recall from (F.4) that

$$\mathbb{E}\left[\zeta_1^2 \mid \mathcal{F}_{n,0}\right] \le \frac{n\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)}{n\left\|\widetilde{A}_n\right\|^2} = o_{\mathbb{P}}(1),$$

and here

$$\mathbb{E}\left[\zeta_{i+1,2}^2 \mid \mathcal{F}_{n,0}\right] = \frac{\mathbf{v}_i^T \widetilde{A}_n \widetilde{A}_n^T \mathbf{v}_i}{n\left\|\widetilde{A}_n\right\|^2} \le \frac{\mathbf{v}_i^T \mathbf{v}_i \lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)}{n\left\|\widetilde{A}_n\right\|^2} = \frac{O_{\mathbb{P}}\left(n\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)\right)}{n\left\|\widetilde{A}_n\right\|^2} \xrightarrow{\mathbb{P}} 0,$$

where we also use $\mathbf{v}_i^T \mathbf{v}_i = O_{\mathbb{P}}\left(\mathbb{E}\left[\mathbf{v}_i^T \mathbf{v}_i\right]\right) = O_{\mathbb{P}}(n)$. Therefore, $\zeta_1 = o_{\mathbb{P}}(1)$ and $\zeta_{i+1,2} = o_{\mathbb{P}}(1)$ by Lemma 1. It remains to prove that $\zeta_{i+1,1} = o_{\mathbb{P}}(1)$. Applying the last part of Lemma 2, we know

$$\begin{aligned}
\zeta_{i+1,1} =& O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}\left\|\widetilde{A}_n\right\|}\left(\sum_{t=1}^n |d_{i,t}|\right)^{\frac{1}{1+\iota}} \max_{1\le t\le n} |d_{i,t}|^{\frac{\iota}{1+\iota}} + \frac{\left\|\Psi_i^T \widetilde{A}_n\right\|}{\sqrt{n}\left\|\widetilde{A}_n\right\|}\right) \\
=:& O_{\mathbb{P}}\left(\zeta_{i+1,1,1} + \zeta_{i+1,1,2}\right),
\end{aligned}$$

where $d_{i,t} := \sum_{\ell=1}^{n-t} \psi_i(\ell) A_n(t+\ell, t)$ is the t-th diagonal element of $\Psi_i^T \widetilde{A}_n = \sum_{\ell=1}^n \psi_i(\ell) L_n^\ell \widetilde{A}_n$. Following the proof of statement (F.6) without using the condition (iii) in Theorem 1, we can show that

$$\sum_{t=1}^n |d_{i,t}| = o_{\mathbb{P}}(n^{(1+\iota)/2}\|A_n\|),$$

as by Cauchy–Schwarz inequality

$$\sum_{t=1}^{n-\ell} |A_n(t+\ell, t)| \le n^{1/2} \sqrt{\sum_{t=1}^{n-\ell} A_n^2(t+\ell, t)} \le n^{1/2}\|A_n\| = o_{\mathbb{P}}(n^{(1+\iota)/2}\|A_n\|).$$

On the other hand, for any constant $M > \sum_{\ell=1}^{\infty} |\psi_i(\ell)|$

$$\begin{aligned}\max_{1\le t\le n} |d_{i,t}| &\le \sum_{\ell=1}^{n} |\psi_i(\ell)| \max_{1\le t<s\le n} |A_n(s,t)| \\ &\le \sum_{\ell=1}^{n} |\psi_i(\ell)| \cdot \left\|\widetilde{A}_n\right\|_{sp} \le M \cdot \left(\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)\right)^{1/2} = o_{\mathbb{P}}\left(\|A_n\|\right).\end{aligned}$$

It follows that

$$\zeta_{i+1,1,1} = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}\,\|A_n\|}\left(n^{(1+\iota)/2}\|A_n\|\right)^{\frac{1}{1+\iota}} \cdot \|A_n\|^{\frac{\iota}{1+\iota}}\right) = o_{\mathbb{P}}(1).$$

Finally, we recall from (F.7) that

$$\zeta_{i+1,1,2}^2 \le \frac{1}{n\|A_n\|^2}\left\|\Psi_i\Psi_i^T\right\|_{sp} \cdot \mathrm{tr}\left(\widetilde{A}_n\widetilde{A}_n^T\right) = \frac{1}{n\|A_n\|^2} \cdot O_{\mathbb{P}}\left(\|A_n\|^2\right) \xrightarrow{\mathbb{P}} 0,$$

as $\left\|\Psi_i\Psi_i^T\right\|_{sp} \le \left(\sum_{\ell=1}^{n} |\psi_i(\ell)|\right)^2 < M^2$.

Appendix F.10. Proof of Lemma 12

Like (F.9), in general we can extend the expansion (F.15) and decompose that

$$\mathbf{z}_i = (z_{1,i}, \ldots, z_{n,i})^T = \alpha\mathbf{1}_n + \Psi_i\epsilon + \Psi_i X\beta + \mathbf{v}_i + \mathbf{r}_i,\ \ i = 1, \ldots, d, \tag{F.16}$$

where $\mathbf{v}_i = (v_{1,i}, \ldots, v_{n,i})^T$ with

$$v_{t,i} = \sum_{\ell=t}^{\infty} \psi_i(\ell) w_{t-\ell} = \sum_{\ell=t}^{\infty} \psi_i(\ell) x_{t-\ell}^T\beta + \sum_{\ell=t}^{\infty} \psi_i(\ell)\epsilon_{t-\ell},$$

and $\mathbf{r}_i = (r_{1,i}, \ldots, r_{n,i})^T$. We need an auxiliary lemma:

Lemma D.1. $\mathbf{v}_i^T\mathbf{v}_i = o_{\mathbb{P}}(n^{1/2})$ *for every* $i = 1, \ldots, d$.

Proof. Applying Cauchy-Schwarz inequality,

$$\mathbf{v}_i^T\mathbf{v}_i \le 2\sum_{t=1}^{n}\left(\sum_{\ell=t}^{\infty}\psi_i(\ell)x_{t-\ell}^T\beta\right)^2 + 2\sum_{t=1}^{n}\left(\sum_{\ell=t}^{\infty}\psi_i(\ell)\epsilon_{t-\ell}\right)^2 =: 2T_n$$

Taking expectation on the right-hand-side yields that

$$\begin{aligned}\mathbb{E}T_n =& O\left(\sum_{t=1}^{n}\sum_{\ell,\ell'=t}^{\infty}\psi_i(\ell)\psi_i(\ell')\mathrm{corr}(x_{t-\ell}^T\beta, x_{t-\ell'}^T\beta)\cdot\beta^T\Sigma\beta + \sum_{t=1}^{n}\sum_{\ell=t}^{\infty}\psi_i^2(\ell)\right) \\ =& O\left(\sum_{t=1}^{n}\sum_{\ell,\ell'=t}^{\infty}|\psi_i(\ell)||\psi_i(\ell')|\cdot\beta^T\Sigma\beta + \sum_{t=1}^{n}\sum_{\ell=t}^{\infty}\psi_i^2(\ell)\right).\end{aligned}$$

We claim that $\sum_{t=1}^{n}\sum_{\ell,\ell'=t}^{\infty}|\psi_i(\ell)||\psi_i(\ell')| = o(n)$. For any $\iota \in (0,1)$,

$$\begin{aligned}\sum_{t=1}^{n}\sum_{\ell,\ell'=t}^{\infty}|\psi_i(\ell)||\psi_i(\ell')| =& \sum_{t=1}^{n}\left(\sum_{\ell=t}^{\infty}|\psi_i(\ell)|\right)^2\\ =& \sum_{t=1}^{n^\iota}\left(\sum_{\ell=t}^{\infty}|\psi_i(\ell)|\right)^2 + \sum_{t=n^\iota+1}^{n}\left(\sum_{\ell=t}^{\infty}|\psi_i(\ell)|\right)^2\\ \leq& n^\iota\cdot\left(\sum_{\ell=1}^{\infty}|\psi_i(\ell)|\right)^2 + (n-n^\iota)\left(\sum_{\ell=n^\iota+1}^{\infty}|\psi_i(\ell)|\right)^2\\ =& O(n^\iota)+o(n)=o(n).\end{aligned}$$

Similarly, for any $\iota \in \left(\frac{1}{2+4q},\frac{1}{2}\right)$

$$\begin{aligned}\sum_{t=1}^{n}\sum_{\ell=t}^{\infty}\psi_i^2(\ell) =& \sum_{t=1}^{n^\iota}\sum_{\ell=t}^{\infty}\psi_i^2(\ell) + \sum_{t=n^\iota+1}^{n}\sum_{\ell=t}^{\infty}\psi_i^2(\ell)\\ \leq& n^\iota\sum_{\ell=1}^{\infty}\psi_i^2(\ell) + (n-n^\iota)\sum_{\ell=n^\iota+1}^{\infty}\psi_i^2(\ell)\\ =& O(n^\iota)+(n-n^\iota)\cdot O\left(n^{(-2(1+q)+1)\iota}\right) = o(n^\iota)+o(n^{1-(2q+1)\iota}) = o(n^{1/2}).\end{aligned}$$

Now we can conclude that $\mathbb{E}T_n = o(n\beta^T\Sigma\beta + n^{1/2}) = o(n^{1/2})$, and therefore $T_n = o_{\mathbb{P}}(n)$ by Markov inequality. It follows that $\mathbf{v}_i^T\mathbf{v}_i = O_{\mathbb{P}}(T_n) = o_{\mathbb{P}}(n^{1/2})$. □

Now we are ready to prove Lemma 12. Observe that

$$\left\|Z^T\widetilde{X}\beta\right\|^2 = \sum_{i=1}^{d}\left(\mathbf{z}_i^T\widetilde{X}\beta\right)^2.$$

It suffices to show that $\left(\mathbf{z}_i^T\widetilde{X}\beta\right)^2 = O_{\mathbb{P}}\left(n\|\beta\|^2 + n^2\|\beta\|^4\right) + O_{\mathbb{P}}\left(n^{3/2}\|\beta\|^2\right)$ for each i. Let $i \in \{1,\dots,d\}$. Invoking the decomposition (F.16),

$$\begin{aligned}\mathbf{z}_i^T\widetilde{X}\beta =& \epsilon^T\Psi_i^T\widetilde{X}\beta + \beta^T\widetilde{X}\Psi_i^T\widetilde{X}\beta + \beta^T\bar{x}\mathbf{1}_n^T\Psi_i^T\widetilde{X}\beta + \mathbf{v}_i^T\widetilde{X}\beta + \mathbf{r}_i^T\widetilde{X}\beta\\ =&: \zeta_{i,1}+\zeta_{i,2}+\zeta_{i,3}+\zeta_{i,4}+\zeta_{i,5}.\end{aligned}$$

From the proof of Lemma 9 we have that $\zeta_{i,1}^2 = O_{\mathbb{P}}(n\|\beta\|^2)$, $\zeta_{i,2}^2 = O_{\mathbb{P}}(n^2\|\beta\|^4)$, and $\zeta_{i,3}^2 = O_{\mathbb{P}}(n^2\|\beta\|^4)$. Furthermore, using the asymptotic bound $\mathbf{v}_i^T\mathbf{v}_i = o_{\mathbb{P}}(n^{1/2})$ from Lemma D.1 in equation (F.13) yields that $\zeta_{i,4}^2 = o_{\mathbb{P}}(n^{3/2}\|\beta\|^2)$. Furthermore, we can decompose that

$$\zeta_{i,5} = \mathbf{r}_i^TX\beta - \mathbf{r}_i^T\mathbf{1}_n\bar{x}^T\beta =: \zeta_{i,5,1}+\zeta_{i,5,2}.$$

It remains to check that $\zeta_{i,5,1}^2 = O_{\mathbb{P}}(n^{3/2}\|\beta\|^2)$ and $\zeta_{i,5,2}^2 = O_{\mathbb{P}}(n^{3/2}\|\beta\|^2)$.

First, we show that $\zeta_{i,5,1}^2 = O_{\mathbb{P}}(n^{3/2}\|\beta\|^2)$. Observe that

$$\zeta_{i,5,1}^2 = \left(\sum_{t=1}^n r_{t,i}\gamma_{n,t}\right)^2 \|\beta\|^2.$$

Taking expectation on both sides yields that

$$\begin{aligned}\mathbb{E}\zeta_{i,5,1}^2 =& (n\mathbb{E}[r_{t,i}\gamma_{n,t}])^2 \|\beta\|^2 + \mathrm{var}\left(\sum_{t=1}^n r_{t,i}\gamma_{n,t}\right)\|\beta\|^2 \\ =& O(n^{3/2}\|\beta\|^2) + \left(n\mathrm{var}(r_{t,i}\gamma_{n,t}) + \sum_{1\le s<t\le n}\mathrm{cov}(r_{t,i}\gamma_{n,t}, r_{s,i}\gamma_{n,s})\right)\|\beta\|^2 \\ =& O(n^{3/2}\|\beta\|^2) + O(n^{3/2}\|\beta\|^2) = O(n^{3/2}\|\beta\|^2).\end{aligned}$$

The rest follows by Markov inequality.

Finally, we show that $\zeta_{i,5,2}^2 = O_{\mathbb{P}}(n^{3/2}\|\beta\|^2)$. Observe that

$$\zeta_{i,5,2}^2 = \left(\mathbf{r}_i^T\mathbf{1}_n\right)^2 (\bar{x}^T\xi_n)^2 \|\beta\|^2.$$

By a direct calculation,

$$\mathbb{E}\left(\mathbf{r}_i^T\mathbf{1}_n\right)^2 = n\mathbb{E}r_{t,i}^2 + \sum_{1\le s<t\le n}\mathrm{cov}(r_{t,i}, r_{s,i}) = O(n) + \sum_{1\le s<t\le n}\mathrm{cov}(r_{t,i}, r_{s,i}).$$

On the other hand, using the definition of spectral norm we have that

$$\mathbb{E}\left(\xi_n^T\bar{x}\bar{x}^T\xi_n\right) = \xi_n^T\mathbb{E}\left[\bar{x}\bar{x}^T\right]\xi_n \le \lambda_{\max}\left(\mathbb{E}\left[\bar{x}\bar{x}^T\right]\right).$$

Now applying Markov inequality together with the last two equations yields that

$$\begin{aligned}\widetilde{\zeta}_{i,5,2}^2 =& O_{\mathbb{P}}\left(n\lambda_{\max}\left(\mathbb{E}\left[\bar{x}\bar{x}^T\right]\right) + \left(\sum_{1\le s<t\le n}\mathrm{cov}(r_{t,i}, r_{s,i})\right)\lambda_{\max}\left(\mathbb{E}\left[\bar{x}\bar{x}^T\right]\right)\right)\|\beta\|^2 \\ =& O_{\mathbb{P}}\left(n + n^{3/2}\right)\cdot\|\beta\|^2 = O_{\mathbb{P}}(n^{3/2}\|\beta\|^2)\end{aligned}$$

This completes the proof.

Appendix F.11. Proof of Lemma 13

Invoking the proof of Lemma 11, under the alternatives, we need to add an additional term into (F.15) to get:

$$\mathbf{z}_i := (z_{1,i},\ldots,z_{n,i})^T = \alpha_i\mathbf{1}_n + \Psi_i\epsilon + \Psi_i X\beta + \mathbf{v}_i,\ i = 1,\ldots,d. \tag{F.17}$$

This introduces an additional term in the entry ζ_{i+1} given by

$$\zeta_{i+1,3} := \frac{1}{\sqrt{n}\left\|\widetilde{A}_n\right\|}\beta^T X^T\Psi_i^T\widetilde{A}_n\epsilon.$$

By a direct calculation,

$$\mathbb{E}\left[\zeta_{i+1,3}^2|X\right] = \frac{1}{n\left\|\widetilde{A}_n\right\|^2}\beta^T X^T \Psi_i^T \widetilde{A}_n \widetilde{A}_n^T \Psi_i X\beta \leq \frac{\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)}{\left\|\widetilde{A}_n\right\|^2} \cdot \left\|\Psi_i^T \Psi_i\right\|_{sp} \beta^T \mathbb{S}_n \beta.$$

Note that $\left\|\Psi_i^T \Psi_i\right\|_{sp} \leq (\sum_{\ell=1}^{\infty} |\psi_i(\ell)|)^2 < \infty$, and recall that $\beta^T \mathbb{S}_n \beta = O_{\mathbb{P}}(\|\beta\|^2)$ by (F.11) and (F.12). Hence, using Lemma 1, $\zeta_{i+1,3}^2 = o_{\mathbb{P}}\left(\|\beta\|^2\right) \xrightarrow{\mathbb{P}} 0$. This completes the proof.

Appendix F.12. Proof of Lemma 14

Recall from the proof of Theorem 3 that, for $\delta(x) \equiv 1$, $\left\|\widetilde{A}_n\right\|/\sqrt{p} = \|A_n\|/\sqrt{2p} = \varpi_n/\sqrt{2} + o_{\mathbb{P}}(1)$. Then substituting $\varpi_n/\sqrt{2}$ by $\left\|\widetilde{A}_n\right\|/\sqrt{p}$, substituting h^2 by $\frac{n}{\sqrt{p}}\|\beta\|^2$, and using the additional freeness assumption in the theorem, we only need to show that

$$\frac{1}{\sqrt{n}} Z^T \widetilde{X}\beta - \frac{1}{\sqrt{n}}\left[0, \beta^T \widetilde{X}^T \Psi_1^T \widetilde{X}\beta, \ldots, \beta^T \widetilde{X}^T \Psi_d^T \widetilde{X}\beta\right] \xrightarrow{\mathbb{P}} 0,$$

that is,

$$\mathbf{z}_i^T \widetilde{X}\beta - \beta^T \widetilde{X}^T \Psi_i^T \widetilde{X}\beta = o_{\mathbb{P}}(n^{1/2}), \text{ for each } i = 1, \ldots, d.$$

Let $i \in \{1, \ldots, d\}$. Invoking the decomposition (F.16),

$$\begin{aligned}\mathbf{z}_i^T \widetilde{X}\beta - \beta^T \widetilde{X}\Psi_i^T \widetilde{X}\beta &= \epsilon^T \Psi_i^T \widetilde{X}\beta + \beta^T \bar{x}\mathbf{1}_n^T \Psi_i^T \widetilde{X}\beta + \mathbf{v}_i^T \widetilde{X}\beta + \mathbf{r}_i^T \widetilde{X}\beta \\ &=: \zeta_{i,1} + \zeta_{i,3} + \zeta_{i,4} + \zeta_{i,5}.\end{aligned}$$

Recall from the proof of Lemma 12 we already know that $\zeta_{i,1} = O_{\mathbb{P}}(n^{1/2}\|\beta\|) = o_{\mathbb{P}}(n^{1/2})$, $\zeta_{i,4} = o_{\mathbb{P}}(n^{3/4}\|\beta\|) = o_{\mathbb{P}}(n^{1/2})$. Furthermore, following the proof Lemma 12 therein but using the conditions $\lambda_{\max}(\mathbb{E}[\bar{x}\bar{x}^T]) = o_{\mathbb{P}}(1)$ and $\mathbb{E}[r_{t,i}\gamma_{n,t}] = o_{\mathbb{P}}(n^{-1/4})$, we can show that $\xi_{i,3} = o_{\mathbb{P}}(n\|\beta\|^2) = o_{\mathbb{P}}(n^{1/2})$ and $\zeta_{i,5} = o_{\mathbb{P}}(n^{3/4}\|\beta\|) = o_{\mathbb{P}}(n^{1/2})$.

Appendix G. Proof of Propositions 1 and 2

Appendix G.1. Proof of Proposition 1

We shall first show that the proposition holds for the oracle matrix $\underline{\mathbb{S}}_n$, and then we substitute it by the observed matrix $\underline{S}_n$. Note that

$$\underline{\mathbb{S}}_n(t,t) = \frac{1}{n} x_t^T x_t = \frac{1}{n} f_t^T \Phi^T \Phi f_t$$

Noting that $\mathbb{S}_n(t,t)$ are identically distributed (not necessarily independent) and following the proof of Lemma 2,

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{S}_n(t,t)-\frac{1}{n}\operatorname{tr}\left(\Phi^T\Phi\right)\right|^2\right] = \mathbb{E}\left|\mathbb{S}_n(t,t)-\frac{1}{n}\operatorname{tr}\left(\Phi^T\Phi\right)\right|^2$$
$$\leq M\left(\frac{1}{n^2}\sum_{i=1}^{k}\|\phi_i\|^4+\frac{1}{n^2}\left\|\Phi^T\Phi\right\|^2\right)$$
$$= M\left(\frac{1}{n^2}\sum_{i=1}^{k}\|\phi_i\|^4+\frac{1}{n^2}\|\Sigma\|^2\right).$$

Then, as the mean minimizes the mean squared error,

$$\frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{S}_n(t,t)-\frac{1}{n}\sum_{t=1}^{n}\mathbb{S}_n(t,t)\right|^2 \leq \frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{S}_n(t,t)-\frac{1}{n}\operatorname{tr}\left(\Phi^T\Phi\right)\right|^2$$
$$= O_{\mathbb{P}}\left(\frac{1}{n^2}\sum_{i=1}^{k}\|\phi_i\|^4+\frac{1}{n^2}\|\Sigma\|^2\right),$$

where in the last step we use Lemma 1. Observe that the last term is $O_{\mathbb{P}}(n^{-1})$ as $\|\Sigma\|^2 \leq n\|\Sigma\|_{sp}^2 = O(n)$ and $\sum_{i=1}^{k}\|\phi_i\|^4 = \left\|\operatorname{diag}(\Phi^T\Phi)\right\|^2 \leq \left\|\Phi^T\Phi\right\|^2 = \|\Sigma\|^2 = O(n)$.

Using the identity that $\widetilde{x}_t = x_t - \bar{x}$, we can calculate that

$$\underline{S}_n(t,t)-\mathbb{S}_n(t,t) =: -\frac{2}{n}\bar{x}^T x_t+\frac{1}{n}\bar{x}^T\bar{x},$$

and remove the last perturbation term in the demeaned diagonals to get that

$$\underline{S}_n(t,t)-\frac{1}{n}\sum_{t=1}^{n}\underline{S}_n(t,t) = \left\{\mathbb{S}_n(t,t)-\frac{1}{n}\sum_{t=1}^{n}\mathbb{S}_n(t,t)\right\}-\frac{2}{n}\bar{x}^T x_t.$$

Then, by Cauchy–Schwarz inequality, we can show that

$$\left|\underline{S}_n(t,t)-\frac{1}{n}\sum_{t=1}^{n}\underline{S}_n(t,t)\right|^2 \leq 2\left(\left|\mathbb{S}_n(t,t)-\frac{1}{n}\sum_{t=1}^{n}\mathbb{S}_n(t,t)\right|^2+\left|\frac{2}{n}\bar{x}^T x_t\right|^2\right)$$

Averaging over t yields that

$$\frac{1}{n}\sum_{t=1}^{n}\left|\underline{S}_n(t,t)-\frac{1}{n}\sum_{t=1}^{n}\underline{S}_n(t,t)\right|^2$$
$$\leq 2\left\{\frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{S}_n(t,t)-\frac{1}{n}\sum_{t=1}^{n}\mathbb{S}_n(t,t)\right|^2+\frac{1}{n}\sum_{t=1}^{n}\left|\frac{2}{n}\bar{x}^T x_t\right|^2\right\}.$$

The proposition then follows as

$$\frac{1}{n}\sum_{t=1}^{n}\left|\frac{2}{n}\bar{x}^T x_t\right|^2 = \frac{4}{n^2}\bar{x}^T S_n\bar{x} = \frac{1}{n^2}\|S_n\|_{sp}\|\bar{x}\|^2 = O_{\mathbb{P}}(n^{-1}).$$

Appendix G.2. Proof of Proposition 2

Our first lemma follows from the same arguments for equation (3.2) in Bai and Silverstein (1998) by combining Lemma 2.7 and Lemma 2.9 therein. Note that we have also used Jensen's inequality $(E|f_1|^4)^{\alpha/2} \leq E|f_1|^{2\alpha}$ for any $\alpha \geq 2$. We omit the details of the proof.

Lemma E.1 (Concentration inequality for quadratic forms)**.** *For A being a $n \times n$ matrix (complex), we have, for any $\alpha \geq 2$*

$$\mathbb{E}\left|f^T A f - \mathrm{tr}(A)\right|^\alpha \leq M \mathbb{E}|f_{1,1}|^{2\alpha} \|A\|^\alpha$$

where M is some absolute constant depending only on α.

Lemma E.2. $x_t^T x_t / p \xrightarrow{a.s.} \int x dH(x)$.

Proof. Applying Lemma E.1 with $\alpha = 2 + \iota/2$ and noting that $E|f_{1,1}|^{4+\iota}$ is bounded, for some large constant M

$$\begin{aligned}\mathbb{E}\left|\frac{1}{p}x_t^T x_t - \frac{\mathrm{tr}(\Sigma)}{p}\right|^{2+\iota/2} =& \mathbb{E}\left|\frac{1}{p}f_t^T \Sigma f_t - \frac{\mathrm{tr}(\Sigma)}{p}\right|^{2+\iota/2} \\ \leq& M p^{-(1+\iota/4)}\left(\frac{\mathrm{tr}(\Sigma^2)}{p}\right)^{1+\iota/4} = O(n^{-(1+\iota/4)}).\end{aligned}$$

By Markov inequality and Borel–Cantelli lemma, we can show that

$$\frac{1}{p}x_t^T x_t - \frac{\mathrm{tr}(\Sigma)}{p} \xrightarrow{a.s.} 0.$$

We complete the proof by checking that $\frac{\mathrm{tr}(\Sigma)}{p} = \int x dH_n(x) \to \int x dH(x)$ using the dominated convergence theorem. □

Lemma E.3. *Let $\mathbb{S}_n(t) = \mathbb{S}_n - \frac{1}{n}x_t x_t^T$ be the sample covariance matrix dropping x_t.*

$$\frac{1}{n}x_t^T(\mathbb{S}_n(t) - zI)^{-1}x_t \xrightarrow{a.s.} -1 - \frac{1}{\underline{m}(z)z},$$

where $\underline{m}(z) = \int \frac{1}{\lambda - z} d\underline{F}(\lambda)$ and $\underline{F} = (1-c)\, I_{[0,\infty)} + cF$ is the limiting spectral distribution of $\underline{S}_n$.

Proof. Note that x_t is independent of $\mathbb{S}_n(t)$. Let $z = a + b\mathrm{i}$, where i denotes the imaginary unit and $b > 0$. From the proof of Theorem 1 in Bai et al. (2007), e.g equation (2.9) therein, we know that

$$\frac{1}{x_t^T x_t}x_t^T(\mathbb{S}_n(t) - zI)^{-1}x_t - \frac{1}{x_t^T x_t}x_t^T\left(-z\underline{m}(z)\Sigma - zI\right)^{-1}x_t \xrightarrow{a.s.} 0$$

Combining with Lemma E.2 yield that

$$\frac{1}{n}x_t^T(\mathbb{S}_n(t) - zI)^{-1}x_t - \frac{1}{n}x_t^T\left(-z\underline{m}(z)\Sigma - zI\right)^{-1}x_t \xrightarrow{a.s.} 0.$$

Applying Lemma E.1 with $\alpha = 2 + \iota/2$ and noting that $\mathbb{E}|f_{1,1}|^{4+\iota}$ is bounded, for some large constant M

$$
\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}x_t^T\left(-z\underline{m}(z)\Sigma_n - zI\right)^{-1}x_t - \frac{1}{n}\operatorname{tr}\left(\Sigma\left(-z\underline{m}(z)\Sigma_n - zI\right)^{-1}\right)\right|^{2+\iota/2}\\
\leq& Mn^{-(2+\iota/2)}\left\|\Sigma\left(-z\underline{m}(z)\Sigma_n - zI\right)^{-1}\right\|^{2+\iota/2}\\
=& Mn^{-(2+\iota/2)}|z|^{-2}\left\|\Sigma\left(\underline{m}(z)\Sigma_n + I\right)^{-1}\right\|^{2+\iota/2}\\
\leq& Mn^{-(2+\iota/2)}|z|^{-2}\|\Sigma\|^{2+\iota/2}\left\|\left(\underline{m}(z)\Sigma + I\right)^{-1}\right\|_{sp}^{2+\iota/2}
\end{aligned}
$$

Recall from Silverstein (1995), the last paragraph in page 338, that $\left\|\left(\underline{m}(z)\Sigma + I\right)^{-1}\right\|_{sp}$ is bounded, and $|z| \geq b^2 > 0$ by construction. Hence, for some possibly different constant M, the last upper bound is further bounded by

$$
Mn^{-(2+\iota/2)}\|\Sigma\|^{2+\iota/2} = Mn^{-(1+\iota/4)}\left(\frac{p}{n}\right)^{1+\iota/4}\left(\frac{\operatorname{tr}(\Sigma^2)}{p}\right)^{1+\iota/4} = O(n^{-(1+\iota/4)}).
$$

Then, using Markov inequality and Borel–Cantelli lemma,

$$
\frac{1}{n}x_t^T(\mathbb{S}_n(t) - zI)^{-1}x_t - \frac{1}{n}\operatorname{tr}\left(\Sigma\left(-z\underline{m}(z)\Sigma - zI\right)^{-1}\right) \xrightarrow{a.s.} 0.
$$

Finally,

$$
\begin{aligned}
\frac{1}{n}\operatorname{tr}\left(\Sigma\left(-z\underline{m}(z)\Sigma - zI\right)^{-1}\right) =& -\frac{p}{n}\frac{1}{z}\int\frac{\lambda}{1+\underline{m}\lambda}dH_n(\lambda)\\
&\xrightarrow{a.s.} -\frac{1}{cz}\int\frac{\lambda}{1+\underline{m}\lambda}dH(\lambda) = -1 - \frac{1}{\underline{m}(z)z}.
\end{aligned}
$$

□

Proof of Proposition 2. Let $\delta_1(x) = \delta(x)\cdot x$, $x \in [0,\infty)$. It suffices to show that

$$
\frac{1}{n}\left\|\operatorname{diag}(W_n(\delta)) - \frac{1}{n}\operatorname{tr}(W_n(\delta))I_n\right\|^2 = \frac{1}{n}\sum_{t=1}^n\left(\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t - \frac{1}{n}\operatorname{tr}\delta_1(S_n)\right)^2 \xrightarrow{\mathbb{P}} 0.
$$

Using the fact that the sample mean minimizes the sample mean squared error, it suffices to show that there exists some constant $\mu = \mu(\delta) \in \mathbb{R}$ depending only on the function δ,

$$
\frac{1}{n}\sum_{t=1}^n\left(\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t - \mu\right)^2 \xrightarrow{\mathbb{P}} 0.
$$

We shall show that we only need to prove that, for each t,

$$
\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t \xrightarrow{\mathbb{P}} \mu(\delta). \tag{G.1}
$$

Note that $\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t$ is the t-diagonal element of $W_n(\delta) = \delta_1(\underline{S}_n)$. Applying Weierstrass theorem with the continuity of δ_1 yields that

$$\max_{1\leq t\leq n} \frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t \leq \|\delta_1(\underline{S}_n)\|_{sp} = \|\delta_1(S_n)\|_{sp} = O_{\mathbb{P}}(1). \tag{G.2}$$

Take an arbitrary constant $\varepsilon > 0$, with a slight abuse of notation. Using Markov inequality and the exchangeablity between $\{\widetilde{x}_t\}$ in terms of distribution,

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t - \mu\right)^2 > \varepsilon\right) \leq & \frac{\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t - \mu\right)^2 \mid \mathcal{E}_n\right]}{\varepsilon} + \mathbb{P}(\mathcal{E}_n^c) \\ = & \frac{\mathbb{E}\left[\left(\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t - \mu\right)^2 \mid \mathcal{E}_n\right]}{\varepsilon} + \mathbb{P}(\mathcal{E}_n^c).\end{aligned}$$

for the event $\mathcal{E}_n = \left\{\max_{1\leq t\leq n}\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t \leq M\right\}$. Taking M and n large enough, it follows from equation (G.2) that $\mathbb{P}(\mathcal{E}_n^c)$ can be arbitrarily small. Moreover, using the dominated convergence theorem and equation (G.1), the first term $\mathbb{E}\left[\left(\frac{1}{n}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t - \mu\right)^2 \mid \mathcal{E}_n\right]$ can also be arbitrarily small.

Hence, it remains to prove equation (G.1) for each t. Using the identity $\widetilde{x}_t = x_t - \bar{x}$, we can decompose the left-hand-side therein as

$$\frac{1}{p}\widetilde{x}_t^T\delta(S_n)\widetilde{x}_t = \frac{1}{p}x_t^T\delta(S_n)x_t + \frac{1}{p}\bar{x}^T\delta(S_n)\bar{x} - \frac{2}{p}x_t^T\delta(S_n)\bar{x}.$$

Note that

$$\frac{1}{p}\bar{x}^T\delta(S_n)\bar{x} \leq \|\delta(S_n)\|_{sp}\frac{1}{p}\bar{x}^T\bar{x} \xrightarrow{\mathbb{P}} 0,$$

as $\bar{x}^T\bar{x} = O_{\mathbb{P}}\left(\mathbb{E}\left[\bar{x}^T\bar{x}\right]\right) = \left(\frac{1}{n}\operatorname{tr}(\Sigma)\right) = (p/n)$. By Cauchy–Schwarz inequality,

$$\left|\frac{1}{p}x_t^T\delta(S_n)\bar{x}\right| \leq \sqrt{\frac{1}{p}x_t^T\delta(S_n)x_t\cdot\frac{1}{p}\bar{x}^T\delta(S_n)\bar{x}}.$$

It suffices to prove that

$$\frac{1}{p}x_t^T\delta(S_n)x_t \xrightarrow{\mathbb{P}} \mu(\delta). \tag{G.3}$$

Now, similar to $F^{S_n}(x;\beta)$, define an unproper weighted empirical spectral distribution

$$G_t^n(x) := \frac{1}{p}\sum_{i=1}^{p}(u_i^Tx_t)^2\mathbb{1}(\lambda_i \leq x).$$

Like in Bai et al. (2007), one can verify that the Stieltjes transform of $G_t^n(x)$ is given by

$$m_{G_t^n}(z) = \int\frac{1}{x-z}dG_t^n(x) = \frac{1}{p}x_t^T\left(S_n - zI\right)^{-1}x_t$$

for every $z \in \mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$. Using the expansions that $S_n = \mathbb{S}_n - \bar{x}\bar{x}^T$ and the Sherman–Morrison formula,

$$\begin{aligned} m_{G_t^n}(z) =& \frac{1}{p} x_t^T (\mathbb{S}_n - zI - \bar{x}\bar{x}^T)^{-1} x_t \\ =& \frac{1}{p} x_t^T (\mathbb{S}_n - zI)^{-1} x_t + \frac{1}{p} \frac{(\bar{x}^T (\mathbb{S}_n - zI)^{-1} x_t)^2}{1 - \bar{x}^T (\mathbb{S}_n - zI)^{-1} \bar{x}} =: T_1 + T_2. \end{aligned}$$

Recall that $\mathbb{S}_n = \mathbb{S}_n(t) + \frac{1}{n} x_t x_t^T$ and applying Sherman–Morrison formula,

$$T_1 = \frac{n}{p} \left(1 - \frac{1}{1 + \frac{1}{n} x_t^T (\mathbb{S}_n(t) - zI)^{-1} x_t} \right).$$

Applying Lemma E.3 yields that, for each t,

$$T_1 \xrightarrow{a.s.} \frac{1}{c} (1 + z\underline{m}(z)) = 1 + zm(z).$$

We shall show later that $T_2 \xrightarrow{a.s.} 0$, and therefore $m_{G_t^n}(z) \xrightarrow{a.s.} 1 + zm(z)$, where the limit does not depend on t. By the equivalence between Stieltjes transform and the associated measure, e.g., Theorem B.9 in Bai and Silverstein (2010), and noting that G_t^n has a bounded support with arbitrarily large probability, it follows that

$$\frac{1}{p} x_t^T \delta(S_n) x_t = \int \delta(x) dG_t^n(x) \xrightarrow{\mathbb{P}} \int \delta dG,$$

where the measure G has the Stieltjes transform $m_G(z) = 1 + zm(z)$. This is equation (G.3).

Finally, it remains to show that $T_2 \xrightarrow{a.s.} 0$. Let $\bar{x}_t = \bar{x} - \frac{1}{n} x_t$ be the sample average dropping x_t and recall that $\mathbb{S}_n(t) = \mathbb{S}_n - \frac{1}{n} x_t x_t^T$. Using Sherman–Morrison formula again and by a direct calculation yields that

$$\begin{aligned} \bar{x}^T (\mathbb{S}_n - zI)^{-1} x_t =& \left(\bar{x}_t + \frac{x_t}{n} \right)^T \left(\mathbb{S}_n(t) + \frac{1}{n} x_t x_t^T - zI \right)^{-1} x_t \\ =& \frac{1}{1 + \frac{1}{n} x_t (\mathbb{S}_n(t) - zI)^{-1} x_t} \left\{ \bar{x}_t^T (\mathbb{S}_n(t) - zI)^{-1} x_t + \frac{1}{n} x_t^T (\mathbb{S}_n(t) - zI)^{-1} x_t \right\}. \end{aligned}$$

Note that x_t is independent of $\bar{x}_t$ and $\mathbb{S}_n(t)$. From the proof of Theorem 2 in Pan (2014), by substituting $\frac{x_t}{\|x_t\|}$ for the unit vector $\mathbf{x}_n$ therein, we know that

$$\bar{x}_t^T (\mathbb{S}_n(t) - zI)^{-1} x_t = o(\|x_t\|) \ a.s.,$$

Recall from Lemma E.3 that the reminder term $\frac{1}{n} x_t^T (\mathbb{S}_n(t) - zI)^{-1} x_t = O(1)$ almost surely in the numerator, and furthermore the denominator

$$1 + \frac{1}{n} x_t (\mathbb{S}_n(t) - zI)^{-1} x_t \xrightarrow{a.s.} -\frac{1}{z\underline{m}(z)},$$

where the limit is bounded away from 0. Therefore, almost surely

$$\bar{x}^T(\mathbb{S}_n - zI)^{-1}x_t = o(\|x_t\|) + O(1).$$

Let $z = a + b\mathrm{i}$, where i denotes the imaginary unit and $b > 0$. Recall in the Theorem 2 in Pan (2014), using equation (2.27) therein,

$$\left|\frac{1}{1-\bar{x}^T(\mathbb{S}_n - zI)^{-1}\bar{x}}\right| = \left|1 + \bar{x}^T\left(S_n - zI\right)^{-1}\bar{x}\right| \le 1 + \bar{x}^T\bar{x}\left\|(S_n - zI)^{-1}\right\|_{sp} \le 1 + \bar{x}^T\bar{x}\frac{1}{b}.$$

On the other hand, for some large constant M, $\bar{x}^T\bar{x} = \bar{f}^T\Sigma\bar{f} \le M\bar{f}^T\bar{f} \xrightarrow{a.s.} Mc$. It follows that, almost surely

$$\left|\frac{1}{1-\bar{x}^T(\mathbb{S}_n - zI)^{-1}\bar{x}}\right| = O(1).$$

Hence, almost surely

$$T_2 = \frac{1}{p}\cdot\Big(o(\|x_t\|^2) + o(\|x_t\|) + O(1)\Big)\cdot O(1) = o\left(\frac{x_t^T x_t}{p}\right) \to 0,$$

by using Lemma E.2. Now the proof is complete. □

Appendix H. Proof of Corollaries 1–6

Proof of Corollary 1. It suffices to prove the consistency of the variance estimator (7). A direct calculation yields the matrix expression given by

$$\widehat{\sigma}^2 = \frac{1}{n-(d+1)}\epsilon^T\left(I - P_Z\right)\epsilon = \frac{n}{n-(d+1)}\frac{1}{n}\epsilon^T\epsilon - \frac{1}{n-(d+1)}\epsilon^T P_Z\epsilon =: T_1 + T_2.$$

Using the martingale law of large number and noting that $d/n \to 0$, we can show that $T_1 \xrightarrow{\mathbb{P}} 1$. It remains to show that $T_2 \xrightarrow{\mathbb{P}} 0$, which follows from Lemma 6. □

Proof of Corollary 2. Note that the support of F^{S_n} is bounded with probability tending to 1. It follows from Portmanteau Theorem (e.g. Theorem 2.1 in Billingsley) that $\varpi_n \xrightarrow{\mathbb{P}} \varpi$, where ϖ has the same expression but uses F instead of F^{S_n}. □

Proof of Corollaries 3 and 4. We only need to prove under the local alternatives (8) that $\widehat{\sigma}_n^2/\sigma_n^2 \xrightarrow{\mathbb{P}} 1$. Expanding

$$\begin{aligned}
&\frac{1}{n-(d+1)}y^T\left(I - P_Z\right)y\\
=&\frac{1}{n-(d+1)}\left(X\beta + \epsilon\right)^T\left(I - P_Z\right)\left(X\beta + \epsilon\right)\\
=&\frac{1}{n-(d+1)}\beta^T\widetilde{X}^T\left(I - P_Z\right)\widetilde{X}\beta + \frac{2}{n-(d+1)}\beta^T\widetilde{X}^T\left(I - P_Z\right)e + \frac{1}{n-(d+1)}e^T e\\
=:&T_1 + T_2 + T_3.
\end{aligned}$$

Note that Lemma 6 holds under the alternatives as well, and then by carefully checking the proof of Corollary 1, we already have $T_3 \xrightarrow{\mathbb{P}} 1$. Using the spectral norm inequality,

$$\begin{aligned} T_1 \leq & \|I - P_Z\|_{sp} \cdot \frac{1}{n-(d+1)} \beta^T \widetilde{X}^T \widetilde{X} \beta \\ \leq & 1 \cdot \frac{n}{n-(d+1)} \beta^T S_n \beta = O_{\mathbb{P}}\left(\|\beta\|^2\right) \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Finally, by Cauchy–Schwarz inequality $T_2^2 \leq T_1 \cdot T_3 \xrightarrow{\mathbb{P}} 0$. □

Proof of Corollary 5. It suffices to show that $\widehat{\rho}_n^2 - \rho_n^2 = \widehat{\rho}_n^2 - \|\mu_n\|^2 \xrightarrow{\mathbb{P}} 0$. Let

$$\widehat{\rho}_n^2 = \frac{e^T \widetilde{A}_n^T P_Z \widetilde{A}_n e / \left\|\widetilde{A}_n\right\|^2}{e^T \left(\widetilde{A}_n^T \widetilde{A}_n\right) e / \left\|\widetilde{A}_n\right\|^2} =: \frac{\Delta_1}{\Delta_2}.$$

It suffices to show that: (*) $\Delta_1 - \|\mu_n\|^2 \xrightarrow{\mathbb{P}} 0$; and (**) $\Delta_2 - 1 \xrightarrow{\mathbb{P}} 0$.

We expand that

$$\Delta_1 = \frac{\epsilon^T \widetilde{A}_n^T P_Z \widetilde{A}_n \epsilon}{\left\|\widetilde{A}_n\right\|^2} - \frac{2\epsilon^T P_Z \widetilde{A}_n^T P_Z \widetilde{A}_n \epsilon}{\left\|\widetilde{A}_n\right\|^2} + \frac{\epsilon^T P_Z \widetilde{A}_n^T P_Z \widetilde{A}_n P_Z \epsilon}{\left\|\widetilde{A}_n\right\|^2} =: \Delta_{1,1} - 2\Delta_{1,2} + \Delta_{1,3}.$$

By Lemma 11 and the assumption that $\widehat{\Omega} \xrightarrow{\mathbb{P}} \Omega$,

$$\begin{aligned} \Delta_{1,1} &= \left(\frac{1}{\sqrt{n}\,\|A_n\|} Z^T \widetilde{A}_n \epsilon\right)^T \widehat{\Omega}^{-1} \left(\frac{1}{\sqrt{n}\,\|A_n\|} Z^T \widetilde{A}_n \epsilon\right) \\ &= \left(\Omega^{1/2}\mu_n + o_{\mathbb{P}}(1)\right)^T \left(\Omega^{-1} + o_{\mathbb{P}}(1)\right) \left(\Omega^{1/2}\mu_n + o_{\mathbb{P}}(1)\right) = \mu_n^T \mu_n + o_{\mathbb{P}}(1). \end{aligned}$$

On the other hand, using the definition of spectral norms and Lemma 6,

$$0 \leq \Delta_{1,3} \leq \|P_Z\|_{sp} \frac{\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right)}{\left\|\widetilde{A}_n\right\|^2} \cdot \epsilon^T P_Z \epsilon = 1 \cdot o_{\mathbb{P}}(1) \cdot O_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0.$$

Now applying Cauchy–Schwarz inequality we also have that

$$|\Delta_{1,2}|^2 \leq \Delta_{1,1}\Delta_{1,3} \xrightarrow{\mathbb{P}} 0.$$

This completes the proof of statement (*). The proof of statement (**) is similar. We expand that

$$\Delta_2 = \frac{\epsilon^T \widetilde{A}_n^T \widetilde{A}_n \epsilon}{\left\|\widetilde{A}_n\right\|^2} - 2\frac{\epsilon^T P_Z \widetilde{A}_n^T \widetilde{A}_n \epsilon}{\left\|\widetilde{A}_n\right\|^2} + \frac{\epsilon^T P_Z \widetilde{A}_n^T \widetilde{A}_n P_Z \epsilon}{\left\|\widetilde{A}_n\right\|^2} =: \Delta_{2,1} - 2\Delta_{2,2} + \Delta_{2,3}.$$

Note that the diagonal elements of $\widetilde{A}_n^T\widetilde{A}_n$ are nonnegative and bounded by $\lambda_{\max}\left(\widetilde{A}_n^T\widetilde{A}_n\right)$, and their sum $\operatorname{tr}\left(\widetilde{A}_n^T\widetilde{A}_n\right) = \left\|\widetilde{A}_n\right\|^2$. Using Lemma 2,

$$\begin{aligned}\Delta_{2,1} - 1 =& O_{\mathbb{P}}\left(\left(\frac{\lambda_{\max}\left(\widetilde{A}_n^T\widetilde{A}_n\right)}{\left\|\widetilde{A}_n\right\|^2}\right)^{\frac{\iota}{1+\iota}} + \frac{\left\|\widetilde{A}_n^T\widetilde{A}_n\right\|}{\left\|\widetilde{A}_n\right\|^2}\right)\\ =& o_{\mathbb{P}}(1) + O_{\mathbb{P}}\left(\frac{\lambda_{\max}^{1/2}\left(\widetilde{A}_n^T\widetilde{A}_n\right)\cdot\sqrt{\operatorname{tr}\left(\widetilde{A}_n^T\widetilde{A}_n\right)}}{\left\|\widetilde{A}_n\right\|^2}\right) \xrightarrow{\mathbb{P}} 0.\end{aligned}$$

Finally, by the definition of spectral norm and Lemma 6,

$$0 \le \Delta_{2,3} \le \frac{\lambda_{\max}\left(\widetilde{A}_n^T\widetilde{A}_n\right)}{\left\|\widetilde{A}_n\right\|^2}\cdot\epsilon^T P_Z\epsilon = o_{\mathbb{P}}(1)\cdot O_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0,$$

and by Cauchy–Schwarz inequality $\Delta_{2,2}^2 \le \Delta_{2,1}\Delta_{2,3} \xrightarrow{\mathbb{P}} 0$. This completes the proof. □

Proof of Corollary 6. We need to prove that $\widehat{\rho}_n^2 - \rho_n^2 \xrightarrow{\mathbb{P}} 0$ under the alternatives. Let

$$\widehat{\rho}_n^2 = \frac{e^T\widetilde{A}_n^T P_Z\widetilde{A}_n e/\left\|\widetilde{A}_n\right\|^2}{e^T\left(\widetilde{A}_n^T\widetilde{A}_n\right)e/\left\|\widetilde{A}_n\right\|^2} =: \frac{\widetilde{\Delta}_1}{\widetilde{\Delta}_2}.$$

It suffices to show that: (*) $\widetilde{\Delta}_1 - \rho_n^2 \xrightarrow{\mathbb{P}} 0$; and (**) $\widetilde{\Delta}_2 - 1 \xrightarrow{\mathbb{P}} 0$.

By a direct calculation,

$$\begin{aligned}\widetilde{\Delta}_1 =& \frac{\epsilon^T(I-P_Z)\widetilde{A}_n^T P_Z\widetilde{A}_n(I-P_Z)\epsilon}{\left\|\widetilde{A}_n\right\|^2} + \frac{\beta^T\widetilde{X}^T(I-P_Z)\widetilde{A}_n^T P_Z\widetilde{A}_n(I-P_Z)\widetilde{X}\beta}{\left\|\widetilde{A}_n\right\|^2}\\ &+ 2\frac{\beta^T\widetilde{X}^T(I-P_Z)\widetilde{A}_n^T P_Z\widetilde{A}_n(I-P_Z)\epsilon}{\left\|\widetilde{A}_n\right\|^2} =: \Delta_1 + R_1 + R_2.\end{aligned}$$

Recall from the proof of Corollary 5 that $\Delta_1 - \rho_n^2 \xrightarrow{\mathbb{P}} 0$. For statement (*) it remains to show that $R_1 \xrightarrow{\mathbb{P}} 0$, as then by Cauchy–Schwarz inequality we have $R_2^2 \le 4\Delta_1 R_1 \xrightarrow{\mathbb{P}} 0$. Observe that $\lambda_{\max}(P_Z) = \lambda_{\max}(I-P_Z) = 1$. Then, using the definition of spectral norm,

$$R_1 \le \frac{\lambda_{\max}\left(\widetilde{A}_n^T\widetilde{A}_n\right)}{\left\|\widetilde{A}_n\right\|^2/n}\beta^T S_n\beta = O_{\mathbb{P}}\left(\|\beta\|^2\right) \xrightarrow{\mathbb{P}} 0,$$

where we used the facts that $\lambda_{\max}\left(\widetilde{A}_n^T \widetilde{A}_n\right) \leq \lambda_{\max}^2(S_n) = O_{\mathbb{P}}(1)$ and $\left\|\widetilde{A}_n\right\|^2/n = \|A_n\|^2/(2n) = \frac{p}{2n}\left(\varpi_n + o_{\mathbb{P}}(1)\right)$ which is bounded away from 0 with probability tending to 1.

The proof of statement (**) is completely analogous, after replacing $\widetilde{A}_n^T P_Z \widetilde{A}_n$ by $\widetilde{A}_n^T \widetilde{A}_n$ everywhere above. We omit the details. □

References

Bai, Z.D., Miao, B.Q., Pan, G.M., 2007. On asymptotics of eigenvectors of large sample covariance matrix. The Annals of Probability 35, 1532–1572.

Bai, Z.D., Silverstein, J.W., 1998. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. The Annals of Probability 26, 316–345.

Bai, Z.D., Silverstein, J.W., 2010. Spectral Analysis of Large Dimensional Random Matrices. Second ed., Springer, New York.

De Mol, C., Giannone, D., Reichlin, L., 2008. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? Journal of Econometrics 146, 318 – 328.

El Karoui, N., 2009. Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. The Annals of Applied Probability 19, 2362–2405.

Goeman, J.J., van de Geer, S.A., van Houwelingen, H.C., 2006. Testing against a high dimensional alternative. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68, 477–493.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. Academic Press, New York.

Hayashi, F., 2000. Econometrics. Princeton University Press.

Horn, R.A., Johnson, C.R., 2012. Matrix Analysis. Cambridge University Press, Cambridge.

Jin, B., Wang, C., Miao, B., Huang, M.N.L., 2009. Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA. Journal of Multivariate Analysis 100, 2112–2125.

Liu, S., Dobribabn, E., 2020. Ridge regression: Structure, cross-validation, and sketching, in: International Conference on Learning Representations.

Pan, G.M., 2014. Comparison between two types of large sample covariance matrices. Annales de l'Institut Henri Poincaré Probabilités et Statistiques 50, 655–677.

Silverstein, J.W., 1995. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. Journal of Multivariate Analysis 55, 331–339.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Zheng, X., Li, Y., 2011. On the estimation of integrated covariance matrices of high dimensional diffusion processes. The Annals of Statistics 39, 3121–3151.