A Online Supplements Part One: Additional Simulation Result: Joint Test of Significance

In order to illustrate the performance of the feasible GLS-AR procedure vis-à-vis other procedures for joint tests of significance we first augmented the simulated regression model (21) to include two additional exogenous variables $\{z_{1t}, z_{2t}\}$ where the generating process for z_{1t} followed an AR(1) process with the coefficient $\phi_{z_1} = \{0.0, 0.5, 0.9\}$ and z_{2t} was assumed to be i.i.d., i.e., $\phi_{z_2} = \{0.0\}$. Then using the regression model,

$$y_t = 2.0 + 0.5x_t + 0.0z_{1t} + 0.0z_{2t} + u_t, \tag{A.1}$$

we conducted an $F-$test to ascertain the validity of the null hypothesis $\beta_{z_1} = \beta_{z_2} = 0$. (In the above specification x_t was generated as an i.i.d. process.) Table 1 contains the percentages of the times the various approaches rejected the null hypothesis (with nominal level set at 5%) for several combinations of exogenous variables and a few selected residual serial correlation structures. We did not include results for joint tests based on HAC estimators because, as pointed out by Kiefer et al. (2000), 'in finite samples [they] can lead to tests that have substantial size distortions,' (p. 696). Moreover, from our review of the literature they are seldom used in practice, and to our knowledge, computer routines with such tests are not readily available. Kiefer, Vogelsang, and Bunzel, however, developed a "robust" joint test (denoted as KVB) that has a nonstandard asymptotic distribution that depends on the number of restrictions being tested and critical asymptotic values. In that study they reported results for their test statistic for several AR(1) and MA(1) serial correlation structures. Although neither their model specifications nor sample sizes were the same as ours, for completeness, we include their finite sample null rejection probabilities for the specifications that most closely matched ours, namely AR(1) ($\phi_1 = 0.5, 0.9$) HOMO $q = 2$ and MA(1) ($\theta_1 = \{0.5, 0.7, 0.99\}$) HOMO $q = 2$ for $n = 128$ and 256 observations.

From Table 1, we can see that for all but the AR(1) serial correlation structures, the feasible GLS-AR approach rejected the null hypothesis close to 5% of the time regardless of the structure of the exogenous variables. Improvement can be observed also when sample size is increased. For the AR(1) serial correlation structures the procedure tended to over-reject the null hypothesis. However, as anticipated, as sample size increased the probabilities of rejecting the null hypothesis began to approach their nominal levels. The performance of feasible GLS also improved as the sample increased. The $F-$tests based on feasible GLS were closer to their nominal levels for the AR(1) serial correlation structures than for the other forms of the autocorrelation reported here. For these latter structures, particularly for the smaller sample sizes, feasible GLS tended to over-reject (often by quite a substantial margin) the null hypothesis. Similar to the results of the t statistics, in finite samples, the F tests based on feasible GLS estimators are not well approximated by the F distributions unless the sample sizes are large (about 500 for some residual autocorrelation structures).

The null rejection probabilities using the KVB statistic were 9% and 27.3% (AR(1) ($\phi_1 = 0.5, 0.9$) and 7.8% (for the three MA(1) parametrizations) when $n = 128$ observations. When $n = 256$ only the probabilities for two AR(1) serial correlation structures were reported; they were 7% and 17%.

The performance of OLS, as expected, was very volatile and highly dependent of the structure of the serial correlation and the structure of the exogenous variables. For example, for the ARMA(1,1) serial correlation structure with $\phi = 0.8$ and $\theta = -0.7$, the OLS procedure tended to reject the null hypothesis approximately $5.18\%, 18.18\%$, and 40.94% when the AR(1) structure of z_1 was set at $0.0, 0.5$, and 0.9, respectively. For the MA(1) residual correlation structure, OLS had a tendency to under reject the null hypothesis when z_1 was generated by a nonzero autoregressive process. These results corroborate the general findings of Banerjee and Magnus (2000) that the $F-$statistic is not robust based on OLS residuals. However, they differ somewhat from the conclusion that "if the null hypothesis is accepted using the usual $F-$statistic, it [the null] will also be accepted if the disturbances are not white." This is because their study *only* considered AR(1) serial correlation structures (see also Rothengery (1988)).

B Online Supplements Part Two: Appendix

Notations and Preliminaries

For a vector x, $\|x\| = (x'x)^{1/2}$ denotes the Euclidean norm; for a symmetric matrix A, $\|A\| = \sup\{x'Ax : \ \|x\| = 1\}$; for two symmetric matrices A and B, $A \geqslant B$ means that $A - B$ is nonnegative definite; $A_n = o(1)B_n$ means that for any $\delta > 0$, $\delta B_n - A_n \geqslant 0$ and $A_n + \delta B_n \geqslant 0$ for sufficiently large n, where A_n and B_n are symmetric matrix.

A matrix $C = (c_{ij}, 1 \leqslant i, j \leqslant n)$, C is said to diagonally dominant if $\sum_{j \neq i} |c_{ij}| \leqslant |c_{ii}|$ for each $1 \leqslant i \leqslant n$. If C is symmetric and diagonally dominant with $c_{ii} > 0$ for each i, then C is nonnegative definite. As a consequence, for any symmetric matrix $C = (c_{ij})$,

$$C \leqslant \lambda I_n, \tag{A.2}$$

where $\lambda = \max_i(\sum_j |c_{ij}|)$ and I_n denotes the $n \times n$ identity matrix. Thus, by (7),

$$\Sigma_n \leqslant c_r I_n. \tag{A.3}$$

Let $A = (a_{ij})_{\infty\times\infty}$, where $a_{ij} = a_{j-i}$ with $a_0 = 1$ and $a_k = 0$ for $k < 0$; similarly set $B = (b_{ij})_{\infty\times\infty}$, where $b_{ij} = b_{j-i}$ with $b_0 = 1$ and $b_k = 0$ for $k < 0$. Clearly, (H1) and (8) imply that $A^{-1} = B$. By (H2) and (A.2), there exists $c_0 > 0$ such that $A'A \leqslant \frac{1}{c_0} I_\infty$ and hence $BB' = (A'A)^{-1} \geqslant c_0 I_\infty$. Since Σ_n is a principal submatrix of BB', we also have

$$\Sigma_n \geqslant c_0 I_n. \tag{A.4}$$

The following four lemmas show that $\hat{\Sigma}_n$ is close to Σ_n with high probability which makes it possible to replace $\hat{\Sigma}_n$ by Σ_n.

Lemma A1. *Let*

$$\eta_i = \sum_{-\infty<l<\infty} a_{i,l}\epsilon_l,$$

where $\{\epsilon_j, -\infty < j < \infty\}$ are i.i.d. with $E\epsilon_0 = 0$, $E\epsilon_0^2 = \sigma^2$, and $E\epsilon_0^4 < \infty$. Then

$$Var(\sum_{i=1}^m \eta_i\eta_{i+j}) \leqslant \sigma^{-4} E\epsilon_0^4 Tr(\Sigma_\eta^2), \tag{A.5}$$

where Σ_η is the covariance matrix of $(\eta_1, \cdots, \eta_{m+j})$.

Proof: We have

$$\eta_i\eta_{i+j} = \sum_l \sum_k \epsilon_l\epsilon_k a_{i,l}a_{i+j,k}$$

Table 1: Empirical Size of the F-Statistic Based on Different Estimators for a Variety of Serial Correlations (Nominal Size 5%)

	$\phi_{z_1} = -0.5^a$				$\phi_{z_1} = 0.0^a$				$\phi_{z_1} = 0.5^a$			
	Estimators[b]				Estimators[b]				Estimators[b]			
n	OLS	FGLS	FGLS-AR(1)	FGLS-AR($\tilde{p}$)	OLS	FGLS	FGLS-AR(1)	FGLS-AR($\tilde{p}$)	OLS	FGLS	FGLS-AR(1)	FGLS-AR($\tilde{p}$)
Residual Serial Correlation: AR(1) $\phi_1 = 0.5$												
50	4.2	6.3	6.3	7.2	10.8	5.4	5.4	7.0	17	6.8	6.8	11.2
100	6.8	5.9	5.9	8.5	10.7	5.9	5.9	8.9	17.8	6.2	6.2	8.9
200	4.7	5.5	5.5	6.6	11.4	5.1	5.1	5.8	19.6	5.7	5.7	8.2
500	5.1	4.6	5.2	5.9	12.2	5.4	5.2	5.6	19.6	5.3	5.1	6.1
1000	4.9	4.6	5.1	5.3	10.2	5.0	5.2	5.2	19.0	5.2	4.8	5.3
Residual Serial Correlation: AR(1) $\phi_1 = 0.9$												
50	4.7	4.4	4.4	7.1	16.8	4.6	4.6	6.3	39.2	4.8	4.8	7.6
100	4.5	5.5	5.5	7.6	16.3	5.1	5.1	7.9	44.2	5.6	5.6	9.1
200	4.7	5.3	5.3	7.4	18.2	4.8	4.8	6.2	46.7	5.2	5.2	6.6
500	5.3	5.4	5.3	6.2	21.2	5.6	5.1	5.8	51.8	5.4	5.4	6.2
1000	5.1	5.3	5.1	5.5	19.5	5.2	4.8	5.1	48.0	5.4	5.1	5.3
Residual Serial Correlation: MA(1) $\theta_1 = 0.5$												
50	5.4	7.3	3.5	6.2	3.3	6.6	2.7	6.2	1.9	6.5	2.1	6.5
100	5.8	6.2	3.8	5.2	2.8	5.1	3.5	5.8	1.6	5.0	2.4	6.3
200	5.0	4.8	4.1	5.5	3.0	4.8	3.3	4.9	1.9	4.7	3.0	5.2
500	5.3	4.5	4.9	5.4	2.7	4.6	3.2	5.3	1.9	5.1	3.2	5.5
1000	5.3	4.7	4.8	5.2	3.7	5.0	3.6	5.1	2.1	5.2	3.7	5.1
Residual Serial Correlation: MA(1) $\theta_1 = 0.9$												
50	6.2	21.6	3.3	4.3	3.1	24.1	1.4	4.4	1.5	26.0	1.4	3.3
100	5.2	13.3	3.4	4.6	2.6	14.0	2.3	4.6	1.9	15.6	1.6	3.4
200	4.7	6.8	4.0	4.7	2.5	6.1	2.6	4.7	1.6	6.7	2.2	3.6
500	4.8	4.7	3.9	4.8	2.0	4.5	2.7	4.5	1.5	5.3	2.5	4.3
1000	5.1	5.2	4.3	4.9	1.7	4.6	3.8	4.9	2.2	5.1	3.8	4.8
Residual Serial Correlation: ARMA(1,1) $\phi_1 = 0.8$; $\theta_1 = -0.7$												
50	5.0	8.5	0.9	5.9	15.6	9.4	1.8	6.7	37	9.7	2.9	6.5
100	5.1	5.3	1.5	4.8	17.0	5.6	2.0	4.9	40.0	5.7	2.7	4.7
200	4.7	4.7	2.2	4.8	17.0	4.9	2.6	5.3	39.9	4.4	2.8	5.7
500	5.6	4.8	1.9	5.1	19.2	4.7	2.5	5.5	42.1	5.5	3.1	5.7
1000	5.5	4.8	2.8	4.8	22.1	4.8	3.8	5.3	45.7	5.3	4.3	5.4
Residual Serial Correlation: ARMA(2,1) $\phi_1 = 1.4$, $\phi_2 = -0.6$; $\theta_1 = 0.8$												
50	5.1	15.8	0.4	3.8	16.8	16.0	1.5	4.3	36.1	15.5	2.3	4.3
100	5.4	8.5	0.8	4.7	16.7	9.3	2.4	4.8	36.2	8.3	2.7	4.8
200	5.2	8.0	1.2	4.7	19.0	5.7	2.5	4.6	39.5	5.4	3.4	4.9
500	5.5	4.4	1.1	4.9	19.3	4.4	2.6	5.3	36.2	5.2	3.8	5.5
1000	5.4	4.5	3.0	5.3	19.5	4.8	3.7	5.1	37.0	5.1	4.6	5.2

a. ϕ_{z_1}=AR coefficient associated the generating process for the exogenous variable z_{1t}.
b. Estimators: OLS= Ordinary Least Squares; AR(1)=Estimated Generalized Least Squares Estimator with AR(1) Correction for the Residual Serial Correlation; FGLS=Feasible Generalized Least Squares Estimator with the Corrected Residual Autocorrelation Structure; AR($\tilde{p}$)=Feasible Generalized Least Squares Estimator with AR($\tilde{p}$) Correction for the Residual Serial Correlation ($\tilde{p} = [2n^{1/4}]$).

and

$$\sum_{i=1}^{m} \eta_i \eta_{i+j} = \sum_l \sum_k \epsilon_l \epsilon_k \sum_{i=1}^{m} a_{i,l} a_{i+j,k} = \sum_l \epsilon_l^2 \sum_{i=1}^{m} a_{i,l} a_{i+j,l} + \sum_{l \neq k} \epsilon_l \epsilon_k \sum_{i=1}^{m} a_{i,l} a_{i+j,k}$$

Therefore

$$\begin{aligned} Var(\sum_{i=1}^{m} \eta_i \eta_{i+j}) &= \sum_l Var(\epsilon_l^2)(\sum_{i=1}^{m} a_{i,l} a_{i+j,l})^2 + \sum_{l \neq k} \sigma^4 (\sum_{i=1}^{m} a_{i,l} a_{i+j,k})^2 \\ &\leqslant E\epsilon_0^4 \sum_{l,k} (\sum_{i=1}^{m} a_{i,l} a_{i+j,k})^2 \qquad \text{(A.6)} \end{aligned}$$

Note that

$$E\eta_i \eta_{i'} = \sigma^2 \sum_l a_{i,l} a_{i',l}$$

and

$$\begin{aligned} \sum_{l,k} (\sum_{i=1}^{m} a_{i,l} a_{i+j,k})^2 &= \sum_{l,k} \sum_{i=1}^{m} \sum_{i'=1}^{m} a_{i,l} a_{i+j,k} a_{i',l} a_{i'+j,k} \\ &= \sum_{i=1}^{m} \sum_{i'=1}^{m} \sum_l a_{i,l} a_{i',l} \sum_k a_{i+j,k} a_{i'+j,k} \\ &= \sigma^{-4} \sum_{i=1}^{m} \sum_{i'=1}^{m} E(\eta_i \eta_{i'}) E(\eta_{i+j} \eta_{i'+j}) \\ &\leqslant \sigma^{-4} \sum_{i=1}^{m} \sum_{i'=1}^{m} (1/2)\{(E(\eta_i \eta_{i'}))^2 + (E(\eta_{i+j} \eta_{i'+j}))^2\} \\ &\leqslant \sigma^{-4} \mathrm{Tr}(\Sigma_\eta^2) \qquad \text{(A.7)} \end{aligned}$$

as desired.

Lemma A2. *We have*

$$Tr(\Sigma_{\hat{u}}^2) \leqslant c_r^2 n. \qquad \text{(A.8)}$$

Proof: Recall $\hat{u} = (I_n - X(X'X)^{-1}X')u$, we have

$$\Sigma_{\hat{u}} = (I_n - X(X'X)^{-1}X')\Sigma_n (I_n - X(X'X)^{-1}X') \qquad \text{(A.9)}$$

and by the fact that $(I_n - X(X'X)^{-1}X')^2 = I_n - X(X'X)^{-1}X'$,

$$\begin{aligned} \mathrm{Tr}(\Sigma_{\hat{u}}^2) &= \mathrm{Tr}\Big(I_n - X(X'X)^{-1}X')\Sigma_n (I_n - X(X'X)^{-1}X')^2 \Sigma_n (I_n - X(X'X)^{-1}X')\Big) \\ &= \mathrm{Tr}\Big(\Sigma_n (I_n - X(X'X)^{-1}X')^2 \Sigma_n (I_n - X(X'X)^{-1}X')^2\Big) \\ &= \mathrm{Tr}\Big(\Sigma_n (I_n - X(X'X)^{-1}X')\Sigma_n (I_n - X(X'X)^{-1}X')\Big) \\ &= \mathrm{Tr}\Big((\Sigma_n (I_n - X(X'X)^{-1}X'))^2\Big). \qquad \text{(A.10)} \end{aligned}$$

For $A = (a_{ij}, 1 \leqslant i, j \leqslant n)$, consider now $\mathrm{Tr}((\Sigma_n A)^2)$, the total sum of squared entries of $\Sigma_n A$. Since both Σ_n and A are symmetric, we have

$$\begin{aligned}
\mathrm{Tr}((\Sigma_n A)^2) &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\sum_{l=1}^{n}\sigma_{il}a_{lj})^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}(\sum_{l=1}^{n}|r_{i-l}|)(\sum_{l=1}^{n}|r_{i-l}|a_{lj}^2) \\
&\leqslant c_r\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}|r_{i-l}|a_{lj}^2 = c_r\sum_{l=1}^{n}\sum_{j=1}^{n}a_{lj}^2\sum_{i=1}^{n}|r_{i-l}| \\
&\leqslant c_r^2\sum_{l=1}^{n}\sum_{j=1}^{n}a_{lj}^2 = c_r^2\mathrm{Tr}(AA'). \qquad (A.11)
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{Tr}\Big((\Sigma_n(I_n - X(X'X)^{-1}X'))^2\Big) &\leqslant c_r^2\mathrm{Tr}((I_n - X(X'X)^{-1}X')^2) \\
&= c_r^2\mathrm{Tr}((I_n - X(X'X)^{-1}X')) = c_r^2(n-k) \qquad (A.12)
\end{aligned}$$

here we use the fact that $X(X'X)^{-1}X'$ is non-negative definitive and hence the diagonal is nonnegative.

Lemma A3. *Let $1 \leqslant p \leqslant n/2$. Then for $\delta > 0$,*

$$P(\sum_{j=0}^{p}|\hat{r}_j - E(\hat{r}_j)| > \delta) \leqslant P(\sum_{j=0}^{p}|\hat{r}_j - E(\hat{r}_j)|^2 > \delta^2/(p+1)) \leqslant \frac{8p^2E(\epsilon_0^4)c_r^2}{\sigma^4 n\delta^2}. \qquad (A.13)$$

Proof: The first inequality in (A.13) follows from the Cauchy inequality

$$(\sum_{j=0}^{p}|\hat{r}_j - E\hat{r}_j|)^2 \leqslant (p+1)\sum_{j=0}^{p}|\hat{r}_j - E\hat{r}_j|^2.$$

As to the second inequality, by Lemmas A1 and A2,

$$E(\hat{r}_j - E\hat{r}_j)^2 \leqslant \frac{E(\epsilon_0^4)c_r^2 n}{\sigma^4(n-p)^2} \leqslant \frac{4E(\epsilon_0^4){c_r}^2}{\sigma^4 n} \qquad (A.14)$$

and hence

$$P(\sum_{j=0}^{p}|\hat{r}_j - E\hat{r}_j|^2 > \delta^2/p) \leqslant \frac{p+1}{\delta^2}\sum_{j=0}^{p}E|\hat{r}_j - E\hat{r}_j|^2 \leqslant \frac{8p^2E(\epsilon_0^4){c_r}^2}{\sigma^4 n\delta^2}, \qquad (A.15)$$

as desired.

Lemma A4. *For $0 \leqslant j \leqslant p$, we have*

$$|E(\hat{r}_j) - r_j| \leqslant 3c_r h_n^* \qquad (A.16)$$

Proof: Let $H = X(X'X)^{-1}X'$. By (A.9), the covariance matrix of $\hat{u}$ is given by

$$\Sigma_{\hat{u}} = \Sigma_n - (\Sigma_n H + H\Sigma_n - H\Sigma_n H) \equiv \Sigma_n - K_n.$$

Recall $h_n^* = \max_i H_{ii}$. Since H is nonnegative definite, $|H_{ij}| \leqslant h_n^*$. Clearly,

$$|(\Sigma_n H)_{i,j}| = |\sum_{l=1}^{n} \sigma_{il} h_{lj}| \leqslant h_n^* \sigma_{il} |\sigma_{il}| \leqslant c_r h_n^*,$$

$$|(H\Sigma_n)_{i,j}| \leqslant c_r h_n^*,$$

and note $H\Sigma_n H$ is nonnegative definite, we have

$$\begin{aligned} |(H\Sigma_n H)_{ij}| &\leqslant (H\Sigma_n H)_{ii} = \sum_{l,l'} h_{il}\sigma_{l\,l'} h_{l'i} \leqslant (1/2) \sum_{l,l'} |\sigma_{l\,l'}|(h_{il}^2 + h_{l'i}^2) \\ &= \sum_l h_{il}^2 \sum_{l'} |\sigma_{l\,l'}| \leqslant c_r \sum_l h_{il}^2 = c_r (HH)_{ii} = c_r H_{ii} \leqslant c_r h_n^*. \end{aligned} \tag{A.17}$$

We are now ready to prove Lemmas 1 - 3.

Proof of Lemma 1. It is easy to see that

$$E(W_{n,1}W_{n,1}') = (X'\Sigma_n^{-1}X)^{-1}X'\Sigma_n^{-1}\Sigma_n\Sigma_n^{-1}X(X'\Sigma_n^{-1}X)^{-1} = (X'\Sigma_n^{-1}X)^{-1}$$

and hence by (H4),

$$n^\alpha \tau' E(W_{n,1}W_{n,1}')\tau \to \tau' A_0^{-1}\tau. \tag{A.18}$$

Note that $\tau' W_{n,1}$ is a linear combination of i.i.d. random variables, the central limit theorem (13) holds.

Proof of Lemma 2. Since $p = o(\sqrt{n})$ and $ph_n^* = o(1)$, by (A.13) and (A.16), there exists $\delta_n \to 0$ such that $P(D_n^c) \to 0$, where $D_n = \{\sum_{j=0}^{p} |\hat{r}_j - r_j| \leqslant \delta_n\}$. Thus, it suffices to show that

$$E|\tau'((X'\hat{\Sigma}_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1})X'\Sigma_n^{-1}u|1_{D_n} = o(n^{-\alpha/2}). \tag{A.19}$$

By the Cauchy inequality,

$$\begin{aligned} &E|\tau'((X'\hat{\Sigma}_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1})X'\Sigma_n^{-1}u|1_{D_n} \\ \leqslant\; &\big(E\|\tau'((X'\hat{\Sigma}_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1})\|^2 1_{D_n}\big)^{1/2}\big(E\|u'\Sigma_n^{-1}X\|^2\big)^{1/2} \\ \equiv\; &K_{n,1}^{1/2}\, K_{n,2}^{1/2}. \end{aligned} \tag{A.20}$$

Observe that

$$\begin{aligned} K_{n,2} &= E\big(\mathrm{Tr}(X'\Sigma_n^{-1}uu'\Sigma_n^{-1}X)\big) = \mathrm{Tr}\big(E((X'\Sigma_n^{-1}uu'\Sigma_n^{-1}X)\big) \\ &= \mathrm{Tr}\big(X'\Sigma_n^{-1}\Sigma_n\Sigma_n^{-1}X)\big) = \mathrm{Tr}\big(X'\Sigma_n^{-1}X\big) = O(n^\alpha). \end{aligned} \tag{A.21}$$

As to $K_{n,1}$, note that $\hat{\Sigma}_n$ is $(2p+1)$th diagonal. By (A.16), for $1 \leqslant i \leqslant n$ on the event D_n

$$\begin{aligned} \sum_{l=1}^{n} |(\hat{\Sigma}_n - \Sigma_n)_{il}| &\leqslant \sum_{l:|i-l|\leqslant p} |\hat{r}_{|i-l|} - r_{|i-l|}| + \sum_{l:|i-l|>p} |r_{|i-l|}| \\ &\leqslant \delta_n + \sum_{l:|i-l|>p} |r_{|i-l|}| = o(1). \end{aligned} \tag{A.22}$$

Therefore $\hat{\Sigma}_n - \Sigma_n = o(1)I_n$. By (A.4), $\hat{\Sigma}_n \geqslant c_0 I_n$ and hence $\hat{\Sigma}_n - \Sigma_n = o(1)\Sigma_n$, i.e., $\hat{\Sigma}_n = (1+o(1))\Sigma_n$. By the fact that $A \leqslant B$ implies $(X'A^{-1}X)^{-1} \leqslant (X'B^{-1}X)^{-1}$, we have for any $0 < \delta < 1/2$, on the event D_n

$$\begin{aligned} (X'\hat{\Sigma}_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1} &\leqslant (1+o(1))(X'\Sigma_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1} \\ &= o(1)(X'\Sigma_n^{-1}X)^{-1} = o(1)n^{-\alpha}I_k \end{aligned} \tag{A.23}$$

by (H4). Similarly,

$$(X'\Sigma_n^{-1}X)^{-1} - (X'\hat{\Sigma}_n^{-1}X)^{-1} = o(1)n^{-\alpha}I_k. \tag{A.24}$$

Now (A.23) and (A.24) yield

$$((X'\Sigma_n^{-1}X)^{-1} - (X'\hat{\Sigma}_n^{-1}X)^{-1})^2 = o(1)n^{-2\alpha}I_k \tag{A.25}$$

Therefore,

$$\begin{aligned} K_{n,1} &\leqslant E\tau'((X'\hat{\Sigma}_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1})(X'\hat{\Sigma}_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1})\tau 1_{D_n} \\ &\leqslant E(o(1)n^{-2\alpha}\tau'\tau) = o(n^{-2\alpha}). \end{aligned} \tag{A.26}$$

This proves (A.19) by combining results from (A.20), (A.21) and (A.26).
Proof of Lemma 3. Since $p = o(n^{1/4})$, by (A.13) again, there exists $\delta_n = o(n^{-1/4})$ such that $P(D_n^c) \to 0$, where $D_n = \{\sum_{j=0}^p |\hat{r}_j - E\hat{r}_j|^2 \leqslant \delta_n^2/(p+1)\}$. It suffices to show that

$$E(|\tau'(X'\hat{\Sigma}_n^{-1}X)^{-1}X'(\hat{\Sigma}_n^{-1} - G\Sigma_n^{-1})u|1_{D_n}) = o(n^{-\alpha/2}) \tag{A.27}$$

By the Cauchy inequality again, we have

$$E(|\tau'(X'\hat{\Sigma}_n^{-1}X)^{-1}X'(\hat{\Sigma}_n^{-1} - G\Sigma_n^{-1})u|1_{D_n}) \leqslant (K_{n,3}\,K_{n,4})^{1/2}, \tag{A.28}$$

where

$$K_{n,3} = E\|(X'\hat{\Sigma}_n^{-1}X)^{-1}\tau\,1_{D_n}\|^2, \quad K_{n,4} = E\|X'(\hat{\Sigma}_n^{-1} - \Sigma_n^{-1})u1_{D_n}\|^2$$

Similarly to (A.26),

$$K_{n,3} = O(n^{-2\alpha}). \tag{A.29}$$

To estimate $K_{n,4}$, let $G_n = E(\hat{\Sigma}_n)$ and write

$$\begin{aligned} X'(\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}) &= X'\hat{\Sigma}_n^{-1}(\Sigma_n - \hat{\Sigma}_n)\Sigma_n^{-1} \\ &= X'\hat{\Sigma}_n^{-1}(G_n - \hat{\Sigma}_n)\Sigma_n^{-1} + X'\hat{\Sigma}_n^{-1}(\Sigma_n - G_n)\Sigma_n^{-1} \\ &= X'G_n^{-1}(G_n - \hat{\Sigma}_n)\Sigma_n^{-1} + X'(\hat{\Sigma}_n^{-1} - G_n^{-1})(G_n - \hat{\Sigma}_n)\Sigma_n^{-1} \\ &\quad + X'\hat{\Sigma}_n^{-1}(\Sigma_n - G_n)\Sigma_n^{-1} \\ &= X'G_n^{-1}(G_n - \hat{\Sigma}_n)\Sigma_n^{-1} && \text{(A.30)} \\ &\quad + X'\hat{\Sigma}_n^{-1}(G_n - \hat{\Sigma}_n)G_n^{-1})(G_n - \hat{\Sigma}_n)\Sigma_n^{-1} \\ &\quad + X'\hat{\Sigma}_n^{-1}(\Sigma_n - G_n)\Sigma_n^{-1} && \text{(A.31)} \end{aligned}$$

Thus,

$$K_{n,4} \leqslant 3K_{n,5} + 3K_{n,6} + K_{n,7}, \tag{A.32}$$

where

$$\begin{aligned} K_{n,5} &= E\|X'G_n^{-1}(G_n - \hat{\Sigma}_n)\Sigma_n^{-1}u\,1_{D_n}\|^2, \\ K_{n,6} &= E\|X'\hat{\Sigma}_n^{-1}(G_n - \hat{\Sigma}_n)G_n^{-1}(G_n - \hat{\Sigma}_n)\Sigma_n^{-1}u\,1_{D_n}\|^2, \\ K_{n,7} &= E\|X'\hat{\Sigma}_n^{-1}(\Sigma_n - G_n)\Sigma_n^{-1}u\,1_{D_n}\|^2. \end{aligned}$$

Let $v = \Sigma_n^{-1} u$ and

$$A = X' \hat{\Sigma}_n^{-1} (G_n - \hat{\Sigma}_n) G_n^{-1} (G_n - \hat{\Sigma}_n).$$

Observe that $\|Av\|^2 \leqslant \mathrm{Tr}(AA')\|v\|^2 \leqslant k\, \lambda(AA')\|v\|^2$, where $\lambda(AA')$ denotes the maximum eigenvalue of AA'. To get an upper bound of the maximum eigenvalue of AA', we note that on the event D_n

$$\begin{aligned} AA' &\leqslant X' \hat{\Sigma}_n^{-1} (G_n - \hat{\Sigma}_n) G_n^{-1} \delta_n^2 G_n^{-1} (G_n - \hat{\Sigma}_n) \hat{\Sigma}_n^{-1} X \\ &\leqslant c\delta_n^2 X' \hat{\Sigma}_n^{-1} (G_n - \hat{\Sigma}_n)^2 \hat{\Sigma}_n^{-1} X \leqslant c\delta_n^4 X' \hat{\Sigma}_n^{-2} X \leqslant c\delta_n^4 n^\alpha I_k, \end{aligned} \tag{A.33}$$

where c denotes a finite constant whose value may be different at each appearance. Hence $\lambda(AA') = O(\delta_n^4 n^\alpha)$. It is easy to see that $E\|v\|^2 = O(n)$. Therefore,

$$K_{n,6} = o(n^\alpha) \tag{A.34}$$

by the choice of $\delta_n = o(n^{-1/4})$.

To deal with $K_{n,5}$, let $v = (v_1, v_2, \cdots, v_n) = \Sigma_n^{-1}$, $\bar{\sigma}_{ij} = \hat{\sigma}_{ij} - E\hat{\sigma}_{ij}$, $\bar{r}_j = \hat{r}_j - E\hat{r}_j$, and write $X' G_n^{-1} = (a_{ij}, 1 \leqslant i \leqslant k, 1 \leqslant j \leqslant n)$. Then

$$\|X' G_n^{-1} (G_n - \hat{\Sigma}_n) \Sigma_n^{-1} u\|^2 = \sum_{i=1}^k \Big(\sum_{j=1}^n a_{ij} \sum_{l=1}^n \bar{\sigma}_{jl} v_l\Big)^2 \tag{A.35}$$

and

$$\begin{aligned} \sum_{j=1}^n a_{ij} \sum_{l=1}^n \bar{\sigma}_{jl} v_l &= \sum_{j=1}^n a_{ij} \sum_{l=1}^n \bar{r}_{|j-l|} v_l \\ &= \sum_{j=1}^n a_{ij} \sum_l \bar{r}_{|l|} v_{j-l} \qquad\qquad [\text{here } v_k = 0 \text{ for } k < 0] \\ &= \sum_{|l| \leqslant p} \bar{r}_{|l|} \sum_{j=1}^n a_{ij} v_{j-l}. \end{aligned} \tag{A.36}$$

Thus,

$$\Big(\sum_{j=1}^n a_{ij} \sum_{l=1}^n \bar{\sigma}_{jl} v_l\Big)^2 1_{D_n} \leqslant \Big(\sum_{|l| \leqslant p} \bar{r}_l^2 1_{D_n}\Big)\Big(\sum_{|l| \leqslant p} \Big(\sum_{j=1}^n a_{ij} v_{j-l}\Big)^2\Big) \leqslant (\delta_n^2/p) \sum_{|l| \leqslant p} \Big(\sum_{j=1}^n a_{ij} v_{j-l}\Big)^2. \tag{A.37}$$

It is easy to see that the covariance matrix Σ_v of v satisfies $\Sigma_v \leqslant cI_n$ and hence

$$E\Big(\sum_{j=1}^n a_{ij} v_{j-l}\Big)^2 \leqslant c \sum_{j=1}^n a_{ij}^2.$$

Combining all inequalities above yields

$$E\Big(\sum_{j=1}^n a_{ij} \sum_{l=1}^n \bar{\sigma}_{jl} v_l\Big)^2 1_{D_n} \leqslant c_5 \delta_n^2 \sum_{j=1}^n a_{ij}^2$$

and

$$K_{n,5} \leqslant c_5 \delta_n^2 \sum_{i=1}^{k} \sum_{j=1}^{n} a_{ij}^2 = c_5 \delta_n^2 \mathrm{Tr}(X' G_n^{-1} G_n^{-1} X) = O(\delta_n^2 \, n^\alpha) = o(n^\alpha) \tag{A.38}$$

as desired. Similarly, we have

$$K_{n,7} = o(n^\alpha).$$

This completes the proof of Lemma 3 by putting the above inequalities together.