

Supplement to ” *Weighted Least Squares Approximate Restricted Maximum Likelihood Estimation of Vector Autoregressive Processes with Intercept*” by Willa W. Chen and Rohit S. Deo

Contents

Table of Contents	S-0
4 Appendix	S-1
4.1 Proof of Lemma 1	S-1
4.2 Proof of Theorem 1	S-4
4.3 Proof of Theorem 2	S-8
4.4 Proof of Theorem 3	S-11
5 Additional Simulations	S-12
5.1 Univariate AR processes	S-12
5.2 Bivariate AR processes	S-13

4. APPENDIX

4.1. Proof of Lemma 1

Proof. Since $CY = (e'_1, e'_2, \dots, e'_n)'$ where

$$C = \begin{bmatrix} I_k & 0 & & \dots & 0 \\ -A_1 & I_k & 0 & & \dots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & \dots & 0 & -A_p & \dots & -A_1 & I_k \end{bmatrix}, \quad (\text{A1})$$

we have $\text{Var}(Y) = \Sigma = C^{-1}(I_n \otimes \Sigma_e)C'^{-1}$ and so

$$\Sigma^{-1} = C'(I_n \otimes \Sigma_e^{-1})C. \quad (\text{A2})$$

We now simplify the various terms in $Q = Y'\Sigma_Y^{-1}Y - Y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$. We note that using (A2), $Y'\Sigma^{-1}Y = Y'_1\Sigma_e^{-1}Y_1 + \sum_{s=2}^p(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i})'\Sigma_e^{-1}(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i}) + \sum_{t=p+1}^n(Y_t - \sum_{i=1}^p A_i Y_{t-i})'\Sigma_e^{-1}(Y_t - \sum_{i=1}^p A_i Y_{t-i})$ and $X'\Sigma^{-1}X = \Sigma_e^{-1} + \sum_{s=1}^{p-1}(I_k - \sum_{i=1}^s A'_i)\Sigma_e^{-1}(I_k - \sum_{i=1}^s A_i) + (n-p)(I_k - H'_1)\Sigma_e^{-1}(I_k - H_1)$, where $H_1 = \sum_{i=1}^p A_i$. Next, we note that

$$\begin{aligned} X'\Sigma^{-1}Y &= \Sigma_e^{-1}Y_1 + \sum_{s=2}^p \left(I_k - \sum_{j=1}^{s-1} A'_j \right) \Sigma_e^{-1} \left(Y_s - \sum_{j=1}^{s-1} A_j Y_{s-j} \right) \\ &\quad + (I_k - H'_1) \Sigma_e^{-1} \sum_{t=p+1}^n (Y_t - A_1 Y_{t-1} - A_2 Y_{t-2} - \dots - A_p Y_{t-p}). \end{aligned} \quad (\text{A3})$$

Letting $Y_t^* = Y_t - Y_1$ for $t = 1, 2, \dots$, we now re-write each of the initial terms in (A3) above that are of the form $(I_k - \sum_{j=1}^{s-1} A'_j)\Sigma_e^{-1}(Y_s - \sum_{j=1}^{s-1} A_j Y_{s-j})$ for $s = 2, \dots, p$ as

$$\left(I_k - \sum_{j=1}^{s-1} A'_j \right) \Sigma_e^{-1} \left(Y_s^* - \sum_{j=1}^{s-1} A_j Y_{s-j}^* \right) + \left(I_k - \sum_{j=1}^{s-1} A'_j \right) \Sigma_e^{-1} \left(I_k - \sum_{j=1}^s A_j \right) Y_1 \quad (\text{A4})$$

and we re-write the last term in (A3) in the Dickey-Fuller form and then re-adjust Y_1 terms to get

$$\sum_{t=p+1}^n \left(Y_t - \sum_{s=1}^p A_s Y_{t-s} \right) = \sum_{t=p+1}^n \left(Y_t^* - H_1 Y_{t-1}^* - \sum_{s=2}^p H_s \Delta Y_{t-s+1} \right) + (n-p)(I_k - H_1)Y_1, \quad (\text{A5})$$

where $H_1 = \sum_{i=1}^p A_i$ and $H_i = -\sum_{j=i}^p A_j$ for $i = 2, 3, \dots, p$. Using the expressions in (A4) and (A5),

we get

$$\begin{aligned} X'\Sigma^{-1}Y &= (X'\Sigma^{-1}X)Y_1 + \sum_{s=2}^p \left(I_k - \sum_{j=1}^{s-1} A_j' \right) \Sigma_e^{-1} \left(Y_s^* - \sum_{j=1}^{s-1} A_j Y_{s-j}^* \right) \\ &\quad + (I_k - H_1') \Sigma_e^{-1} \sum_{t=p+1}^n \left(Y_t^* - H_1 Y_{t-1}^* - \sum_{s=2}^p H_s \Delta Y_{t-s+1} \right) \\ &= (X'\Sigma^{-1}X)Y_1 + DP_n, \end{aligned} \quad (\text{A6})$$

where $D = \{(I_k - A_1)' \Sigma_e^{-1/2}, \dots, (I_k - A_1 - A_2 - \dots - A_{p-1})' \Sigma_e^{-1/2}, (n-p)^{1/2} (I_k - H_1)' \Sigma_e^{-1/2}\}$

and

$$P_n = \begin{bmatrix} \Sigma_e^{-1/2} [Y_2^* - A_1 Y_1^*] \\ \vdots \\ \Sigma_e^{-1/2} [Y_p^* - \sum_{i=1}^{p-1} A_i Y_{p-i}^*] \\ (n-p)^{-1/2} \Sigma_e^{-1/2} \sum_{t=p+1}^n (Y_t^* - H_1 Y_{t-1}^* - \sum_{s=2}^p H_s \Delta Y_{t-s+1}) \end{bmatrix}.$$

From (A6) we get,

$$Y'\Sigma^{-1}X (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}Y = \{(X'\Sigma^{-1}X)Y_1 + DP_n\}' (X'\Sigma^{-1}X)^{-1} \{(X'\Sigma^{-1}X)Y_1 + DP_n\}$$

$$= Y_1' (X'\Sigma^{-1}X)Y_1 + 2Y_1' DP_n + P_n' D' (X'\Sigma^{-1}X)^{-1} DP_n = T_{p1} + T_{p2} + T_{p3}. \text{ To simplify } T_{p3}, \text{ we}$$

note that

$$X'\Sigma^{-1}X = \Sigma_e^{-1/2} \{I_k + \Sigma_e^{1/2} D D' \Sigma_e^{1/2}\} \Sigma_e^{-1/2} \quad (\text{A7})$$

and hence applying Corollary 18.2.9 of Harville (1997), we get

$$D' (X'\Sigma^{-1}X)^{-1} D = D' \Sigma_e^{1/2} \{I_k + \Sigma_e^{1/2} D D' \Sigma_e^{1/2}\}^{-1} \Sigma_e^{1/2} D = I_{pk} - (I_{pk} + D' \Sigma_e D)^{-1}.$$

Thus, T_{p3} simplifies as $T_{p3} = P_n' P_n - P_n' (I_{pk} + D' \Sigma_e D)^{-1} P_n$. Next, we partition P_n as $P_n =$

$(P_{n1}', P_{n2}')'$, where

$$\begin{aligned} P_{n1} &= \left\{ \Sigma_e^{-1/2} \left(Y_s - Y_1 - \sum_{i=1}^{s-1} A_i (Y_{s-i} - Y_1) \right) \right\}_{s=2}^p \\ &= \left\{ \Sigma_e^{-1/2} \left(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i} \right) \right\}_{s=2}^p - \left\{ \Sigma_e^{-1/2} \left(I_k - \sum_{i=1}^{s-1} A_i \right) Y_1 \right\}_{s=2}^p \end{aligned}$$

and

$$\begin{aligned} P_{n2} &= (n-p)^{1/2} \Sigma_e^{-1/2} \sum_{t=p+1}^n \left\{ (Y_t - Y_1) - H_1 (Y_{t-1} - Y_1) - \sum_{s=2}^p H_s \Delta Y_{t-s+1} \right\} \\ &= (n-p)^{1/2} \Sigma_e^{-1/2} \left\{ \bar{Y}_{(1)} - H_1 \bar{Y}_{(0)} - \sum_{s=2}^p H_s \bar{Z}_{s-1} \right\} - (n-p)^{1/2} \Sigma_e^{-1/2} (I_k - H_1) Y_1, \end{aligned}$$

where $\bar{Y}_{(1)} = (n-p)^{-1} \sum_{t=p+1}^n Y_t$, $\bar{Y}_{(0)} = (n-p)^{-1} \sum_{t=p+1}^n Y_{t-1}$ and $\bar{Z}_s = (n-p)^{-1} \sum_{t=p+1}^n \Delta Y_{t-s}$ for $s = 2, \dots, p$. Using this decomposition of P_{n1} and P_{n2} , we further

simplify the first term of T_{p3} above as $P'_n P_n = P'_{n1} P_{n1} + P'_{n2} P_{n2}$ where

$$\begin{aligned} P'_{n1} P_{n1} &= \sum_{s=2}^p \left(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i} \right)' \Sigma_e^{-1} \left(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i} \right) \\ &\quad + Y_1' \sum_{s=2}^p \left(I_k - \sum_{i=1}^{s-1} A_i' \right) \Sigma_e^{-1} \left(I_k - \sum_{i=1}^{s-1} A_i \right) Y_1 \\ &\quad - 2 \sum_{s=2}^p \left(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i} \right)' \Sigma_e^{-1} \left(I_k - \sum_{i=1}^{s-1} A_i \right) Y_1 \end{aligned}$$

and

$$\begin{aligned} P'_{n2} P_{n2} &= (n-p) \left\{ \bar{Y}_{(1)} - H_1 \bar{Y}_{(0)} - \sum_{s=2}^p H_s \bar{Z}_{s-1} \right\}' \Sigma_e^{-1} \left\{ \bar{Y}_{(1)} - H_1 \bar{Y}_{(0)} - \sum_{s=2}^p H_s \bar{Z}_{s-1} \right\} \\ &\quad + (n-p) Y_1' (I_k - H_1') \Sigma_e^{-1} (I_k - H_1) Y_1 \\ &\quad - 2(n-p) Y_1' (I_k - H_1') \Sigma_e^{-1} \left\{ \bar{Y}_{(1)} - H_1 \bar{Y}_{(0)} - \sum_{s=2}^p H_s \bar{Z}_{s-1} \right\}. \end{aligned}$$

Now we turn to T_{p2} and partitioning D as $D = (D_1, D_2)$ to conform with the rows of P_n we

get $T_{p2} = 2Y_1' D P_n = 2Y_1' D_1 P_{n1} + 2Y_1' D_2 P_{n2}$. The first term is $2Y_1' D_1 P_{n1} = 2Y_1' \sum_{s=2}^p (I_k - \sum_{i=1}^{s-1} A_i) \Sigma_e^{-1} (Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i}) - 2Y_1' \sum_{s=2}^p (I_k - \sum_{i=1}^{s-1} A_i') \Sigma_e^{-1} (I_k - \sum_{i=1}^{s-1} A_i) Y_1$ and the

second term is $2Y_1' D_2 P_{n2} = 2(n-p) Y_1' (I_k - H_1') \Sigma_e^{-1} \left\{ \bar{Y}_{(1)} - H_1 \bar{Y}_{(0)} - \sum_{s=2}^p H_s \bar{Z}_{s-1} \right\} - 2(n-p) Y_1' (I_k - H_1') \Sigma_e^{-1} (I_k - H_1) Y_1$. These calculations, along with the fact that

$$X' \Sigma^{-1} X = \Sigma_e^{-1} + \sum_{s=2}^p \left(I_k - \sum_{i=1}^{s-1} A_i' \right) \Sigma_e^{-1} \left(I_k - \sum_{i=1}^{s-1} A_i \right) + (n-p) (I_k - H_1') \Sigma_e^{-1} (I_k - H_1)$$

thus yield a simplified version of $T_{p1} + T_{p2} + T_{p3}$ given by

$$T_{p1} + T_{p2} + T_{p3} = Y_1' \Sigma_e^{-1} Y_1 + \sum_{s=2}^p \left(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i} \right)' \Sigma_e^{-1} \left(Y_s - \sum_{i=1}^{s-1} A_i Y_{s-i} \right)$$

$$\begin{aligned}
& + (n-p) \left(\bar{Y}_{(1)} - H_1 \bar{Y}_{(0)} - \sum_{s=2}^p H_s \bar{Z}_{s-1} \right)' \Sigma_e^{-1} \left(\bar{Y}_{(1)} - H_1 \bar{Y}_{(0)} - \sum_{s=2}^p H_s \bar{Z}_{s-1} \right) \\
& - P_n' (I_{pk} + D' \Sigma_e D)^{-1} P_n.
\end{aligned}$$

Using this expression, we get

$$\begin{aligned}
Q & = Y' \Sigma^{-1} Y - Y' \Sigma^{-1} X (X' \Sigma_Y^{-1} X)^{-1} X' \Sigma^{-1} Y = Y' \Sigma^{-1} Y - T_{p1} - T_{p2} - T_{p3} \\
& = \sum_{t=p+1}^n S_t' \Sigma_e^{-1} S_t + P_n' (I_{pk} + D' \Sigma_e D)^{-1} P_n,
\end{aligned}$$

where $S_t = (Y_t - \bar{Y}_{(1)}) - H_1 (Y_{t-1} - \bar{Y}_{(0)}) - \sum_{s=2}^p H_s (\Delta Y_{t-s+1} - \bar{Z}_{s-1})$. Writing the A_i in P_{n1} in terms of H_i , it follows that $P_n = (I_p \otimes \Sigma_e^{-1/2}) \text{vec}(R - HU)$ and the expression for Q in Theorem 1 is obtained. Finally, the proof is completed by using the expressions in (A2) and (A7) and the fact that for any matrix A , the non-zero eigenvalues of $A'A$ are identical to those of AA' , all of which yield $\log |\Sigma^{-1}| = -n \log |\Sigma_e|$ and

$$\log |X' \Sigma^{-1} X| = -\log |\Sigma_e| + \log |I_k + \Sigma_e^{1/2} D D' \Sigma_e^{1/2}| = -\log |\Sigma_e| + \log |I_{pk} + D' \Sigma_e D|. \quad \square$$

4.2. Proof of Theorem 1

Let $\lambda_{\max}(B)$ and $\lambda_{\min}(B)$ be the largest and smallest eigenvalue of a matrix B respectively and let $\|B\| = \lambda_{\max}^{1/2}(B'B)$ be its matrix norm. As \hat{A}_{WLS} is mean invariant, we assume that $\mu = 0$.

LEMMA 2. Let Z_t be a stationary Gaussian vector time series and $S = \sum_{t=1}^n (Z_t - \bar{Z})(Z_t - \bar{Z})'$. Then, for any positive integer h , $E \|S^{-1}\|^h = O(n^{-h})$ and $E \|nS^{-1} - E^{-1}(n^{-1}S)\|^h = O(n^{-h/2})$.

Proof of Lemma 2. We assume, without loss of generality that n/m is an integer. For a positive integer m to be defined later, $n^{-1}S = n^{-1} \sum_{s=1}^{n/m} B_s$, where $B_s = \sum_{t=1+sm}^{(s+1)m} (Z_t - \bar{Z})(Z_t - \bar{Z})'$.

$$\lambda_{\min}(n^{-1}S) = \lambda_{\min} \left(\frac{1}{n} \sum_{s=1}^{[n/m]} B_s \right) = \frac{1}{n} \lambda_{\min} \left(\sum_{s=1}^{[n/m]} B_s \right) \geq \frac{1}{n} \sum_{s=1}^{[n/m]} \lambda_{\min}(B_s).$$

Hence, by inequality of harmonic and arithmetic means of positive numbers and Jensen's inequality

$$\lambda_{\min}^{-h}(n^{-1}S) \leq n^h \left\{ \sum_{s=1}^{n/m} \lambda_{\min}(B_s) \right\}^{-h} \leq \left(\frac{1}{n} \sum_{s=1}^{n/m} \lambda_{\min}^{-1}(B_s) \right)^h \leq \frac{1}{n} \sum_{s=1}^{n/m} \lambda_{\min}^{-h}(B_s).$$

For any $1 \leq s \leq n/m$, the vectors $\{Z_t - \bar{Z}\}_{t=1+sm}^{(s+1)m}$ have a Gaussian distribution with a nonsingular variance matrix uniformly in s . The proof of Lemma 8 of Chen & Hurvich (2006) yields $\sup_n \sup_{1 \leq s \leq n/m} E \{\lambda_{\min}^{-h}(B_s)\} < C_0$, where C_0 is the h th moment of λ_{\max} of an inverse Wishart matrix with m degrees of freedom. Choosing $m > 2h + k - 1$ ensures that this h th moment is finite (von Rosen, 1988), giving $C_0 < \infty$ and proving the first part. For the second assertion, we note that $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$. Thus

$$E \|nS^{-1} - E^{-1}(n^{-1}S)\|^h \leq E^{1/2} \|nS^{-1}\|^{2h} E^{1/2} \|n^{-1}S - E(n^{-1}S)\|^{2h} E(n^{-1}S) = O(n^{-h/2}).$$

□

LEMMA 3. For any positive h ,

- (i) $E \|\hat{A}_{OLS,0} - A_0\|^h = O(n^{-h/2})$, $E \|\hat{A}_{OLS} - A_0\|^h = O(n^{-h/2})$, $E \|\hat{\Sigma}_e - \Sigma_e\|^h = O(n^{-h/2})$
- (ii) $E \|\hat{\Sigma}_e^{-1}\|^h = O(1)$, $E \|\hat{\Sigma}_e^{-1} - \Sigma_e^{-1}\|^h = O(n^{-h/2})$
- (iii) $E \|\hat{\Sigma}_e^{-1/2} \hat{W} \hat{\Sigma}_e^{-1/2} - \Sigma_e^{-1/2} W \Sigma_e^{-1/2}\|^h = O(n^{-3h/2})$

Proof of Lemma 3. For (i), Hölder's inequality and Lemma 2 yield

$$n^{h/2} E \|\hat{A}_{OLS,0} - A_0\|^h \leq n^{-h/2} E^{1/2} \left\| \sum_{t=2}^n e_t Y'_{t-1} \right\|^{2h} n^h E^{1/2} \left\| \sum_{t=2}^n Y_{t-1} Y'_{t-1} \right\|^{-2h} = O(1)$$

by Lemma 2 and Lemma 3.3 of Bhansali (1981). The results for \hat{A}_{ols} and $\hat{\Sigma}_e$ can be proved in a similar fashion. For (ii), we note that

$$\hat{\Sigma}_e = \frac{1}{n} \left\{ \sum_{t=2}^n (e_t - \bar{e}_{(1)}) (e_t - \bar{e}_{(1)})' - \sum_{t=2}^n (e_t - \bar{e}_{(1)}) (Y_{t-1} - \bar{Y}_{(0)})' S_Y^{-1} \sum_{t=2}^n (Y_{t-1} - \bar{Y}_{(0)}) (e_t - \bar{e}_{(1)})' \right\},$$

where $\bar{e}_{(1)} = (n-1)^{-1} \sum_{t=2}^n e_t$. Letting $\eta_t = \{(Y_{t-1} - \bar{Y}_{(0)})', (e_t - \bar{e}_{(1)})'\}'$ and $\Psi = (n-1)^{-1} \sum_{t=2}^n \eta_t \eta'_t$, then $\hat{\Sigma}_e$ is the Schur complement of Ψ relative to $(n-1)S_Y^{-1}$. Hence,

$$E|\hat{\Sigma}_e|^{-h} = E\{|\Psi|^{-h} |n^{-1}S_Y|^h\} \leq E^{1/2} |\Psi|^{-2h} E^{1/2} |n^{-1}S_Y|^{2h} = O(1)$$

by Lemma 2. The same lemma and the fact that $\|\hat{\Sigma}_e^{-1}\| \leq |\hat{\Sigma}_e^{-1}| \|\hat{\Sigma}\|^{k-1}$ yields first part of (ii). The second part of (i) can be shown by writing $\hat{\Sigma}_e^{-1} - \Sigma_e^{-1} = \hat{\Sigma}_e^{-1} (\Sigma_e - \hat{\Sigma}_e) \Sigma_e^{-1}$.

For (iii) we first show that $E \left\| \hat{W} \right\|^h = O(n^{-h})$. Defining the set $\Phi = \left\{ \left\| \hat{A}_{OLS} - A \right\| < \varepsilon \right\} \cap \left\{ \left\| \hat{\Sigma}_e - \Sigma_e \right\| < \varepsilon \right\} = \Phi_1 \cap \Phi_2$, we have

$$n^h E \left\| \hat{W} \right\|^h = n^h E \left\| \hat{W} \right\|^h I_\Phi + n^h E \left\| \hat{W} \right\|^h I_{\Phi^c}. \quad (\text{A8})$$

Since A has all eigenvalues less than unity, both $I - A$ and $I - \Sigma_e^{-1/2} A \Sigma_e^{1/2}$ are non-singular and hence $n \left(I + (n-1) \Sigma_e^{-1/2} (I - A) \Sigma_e (I - A)' \Sigma_e^{-1/2} \right)^{-1} = O(1)$ due to which $n^h E \left\| \hat{W} \right\|^h I_\Phi = O(1)$.

To show that the second term on the right in (A8) is bounded, we note that \hat{W} has all eigenvalues bounded by one and get $n^h E \left\| \hat{W} \right\|^h I_{\Phi^c} \leq n^h E (I_{\Phi^c})$. From (i) and Chebyshev's inequality,

$$n^k E I_{\Phi^c} = n^k P(\Phi^c) \leq n^k P(\Phi_1^c) + n^k P(\Phi_2^c) \leq n^k P\left(\left\| \hat{A}_{OLS} - A \right\| > \varepsilon\right) + n^k P\left(\left\| \hat{\Sigma}_e - \Sigma_e \right\| > \varepsilon\right) = O(1).$$

Thus $E \left\| \hat{W} \right\|^h = O(n^{-h})$. Furthermore, writing $\hat{W} - W_0 = \hat{W}^h (W_0^{-1} - \hat{W}^{-1}) W_0^h$, we get

$$\begin{aligned} E \left\| \hat{W} - W_0 \right\|^h &\leq C n^{-2h} E^{1/2} \left\| W_0^{-1} - \hat{W}^{-1} \right\|^{2h} \\ &\leq C n^{-2h} \left\{ E^{1/4} \left\| \hat{A}_{OLS} - A_0 \right\|^{4h} + E^{1/4} \left\| \hat{\Sigma}_e^{-1/2} - \Sigma_e^{-1/2} \right\|^{4h} + E^{1/4} \left\| \hat{\Sigma}_e - \Sigma_e \right\|^{4h} \right\} \\ &= O(n^{-2h} n^{1/2}) = O(n^{-3h/2}). \end{aligned}$$

Part (iii) follows from the fact that the left side is bounded by

$$\begin{aligned} &\Sigma_e^{-1/2} W_0 E \left\| \hat{\Sigma}_e^{-1} \right\|^h + \Sigma_e^{-1/2} E^{1/2} \left\| \hat{W} - W_0 \right\|^{2h} E^{1/2} \left\| \hat{\Sigma}_e^{-1} \right\|^{2h} \\ &+ E^{1/4} \left\| \hat{\Sigma}_e^{-1/2} - \Sigma_e^{-1/2} \right\|^{4h} E^{1/4} \left\| \hat{W} \right\|^{4h} E^{1/2} \left\| \hat{\Sigma}_e^{-1/2} \right\|^{2h}. \quad \square \end{aligned}$$

Proof of Theorem 1. As noted by Yamamoto & Kunitomo (1984) & Nicholls and Pope (1988) it suffices to obtain the bias for an AR(1) process since an AR(p) can always be re-expressed as a suitable AR(1).

Since $-\partial Q(\hat{A}_{OLS,0}) = \partial Q(\hat{A}_{WLS}) - \partial Q(\hat{A}_{OLS,0})$, we get

$$\begin{aligned} -\partial Q(\hat{A}_{OLS,0}) &= 2\hat{\Sigma}_e^{-1} (\hat{A}_{WLS} - \hat{A}_{OLS,0}) \sum_{t=2}^n (Y_{t-1} - \bar{Y}_{(0)}) (Y_{t-1} - \bar{Y}_{(0)})' \\ &+ 2(n-1) \hat{\Sigma}_e^{-1/2} \hat{W} \hat{\Sigma}_e^{-1/2} (\hat{A}_{WLS} - \hat{A}_{OLS,0}) (\bar{Y}_{(0)} - Y_1) (\bar{Y}_{(0)} - Y_1)'. \end{aligned}$$

Thus $vec(\hat{A}_{WLS} - \hat{A}_{OLS,0}) = (G + J)^{-1} vec\{\partial Q(\hat{A}_{OLS,0})\}$, where

$$J = (n-1) \Gamma \otimes \hat{\Sigma}_e^{-1/2} \hat{W} \hat{\Sigma}_e^{-1/2}, \quad \Gamma = (\bar{Y}_{(0)} - Y_1) (\bar{Y}_{(0)} - Y_1)' \quad (\text{A9})$$

$$G = S_Y \otimes \hat{\Sigma}_e^{-1}, \quad S_Y = \sum_{t=2}^n (Y_{t-1} - \bar{Y}_{(0)}) (Y_{t-1} - \bar{Y}_{(0)})'. \quad (\text{A10})$$

Let Q_0 be the objective function of \hat{A}_{IWLs} , it is sufficient to show (i) $E(G+J)^{-1} = (n-1)^{-1}\{\Sigma_Y^{-1} \otimes \Sigma_e\} + o(n^{-1})$, (ii) $E\|(G+J)^{-1}\|^2 = O(n^{-2})$, (iii) $E\{\partial Q_0(\hat{A}_{OLS,0})\} = o(1)$, (iv) $E\|\partial Q_0(\hat{A}_{OLS,0})\|^2 = O(1)$, and (v) $E\|\partial Q_0(\hat{A}_{OLS,0}) - \partial Q(\hat{A}_{OLS,0})\|^2 = o(n^{-2})$.

Note that J is ranked one, thus $(G+J)^{-1} = G^{-1} - \{1 + \text{trace}(JG^{-1})\}G^{-1}JG^{-1}$. Furthermore,

$$E\left\|G^{-1} - \frac{\Sigma_Y^{-1} \otimes \hat{\Sigma}_e}{(n-1)}\right\|^h \leq \frac{E\left\|\{(n-1)S_Y^{-1} - \Sigma_Y^{-1}\} \otimes \hat{\Sigma}_e\right\|^h}{(n-1)^h} + \frac{\Sigma_Y^{-h} \otimes E\|\hat{\Sigma}_e - \Sigma\|^h}{(n-1)^h}.$$

By Lemmas 2, 3 and Hölder's inequality, both terms on the right hand side are $O(n^{-3})$. Since $\{1 + \text{trace}(JG^{-1})\}^{-1}G^{-1}JG^{-1} \leq G^{-1}JG^{-1}$, we get $E\|G^{-1}JG^{-1}\|^2 = O(n^{-4})$ by Lemma 3. We have shown (i) and (ii). Noting that $\sum_{t=2}^n Y_t Y_{t-1}' = \hat{A}_{OLS,0} \sum_{t=2}^n Y_{t-1} Y_{t-1}'$, we get

$$\begin{aligned} \partial Q_0(\hat{A}_{OLS,0}) &= 2(n-1)\Sigma_e^{-1} \left(\bar{Y}_{(1)} - \hat{A}_{ols,0} \bar{Y}_{(0)} \right) \bar{Y}_{(0)}' \\ &\quad - 2(n-1)\Sigma_e^{-1/2} W \Sigma_e^{-1/2} \left(\{\bar{Y}_{(1)} - Y_1\} - \hat{A}_{ols,0} \{\bar{Y}_{(0)} - Y_1\} \right) (\bar{Y}_{(0)} - Y_1)' \\ &= 2(n-1)\{\Omega(W_0, \Sigma_e) + R(W_0, \Sigma_e)\}, \end{aligned}$$

$$\Omega(W_0, \Sigma_e) = \Sigma_e^{-1/2} (I - W_0) \Sigma_e^{-1/2} (I - A_0) \bar{Y}_{(0)} \bar{Y}_{(0)}' - \Sigma_e^{-1/2} W_0 \Sigma_e^{-1/2} (I - A_0) Y_1 Y_1', \quad (\text{A11})$$

$$\begin{aligned} R(W_0, \Sigma_e) &= \Sigma_e^{-1/2} (I - W_0) \Sigma_e^{-1/2} \left\{ (A_0 - \hat{A}_{ols,0}) \bar{Y}_{(0)} + (n-1)^{-1} (Y_n - Y_1) \right\} \bar{Y}_{(0)}' \\ &\quad - \Sigma_e^{-1/2} W_0 \Sigma_e^{-1/2} \left\{ (A_0 - \hat{A}_{ols,0}) Y_1 - 2(I - \hat{A}_{ols,0}) \bar{Y}_{(0)} - (n-1)^{-1} (Y_n - Y_1) \right\} Y_1'. \end{aligned}$$

From Brillinger (2001),

$$E\{(n-1)\bar{Y}_{(0)} \bar{Y}_{(0)}'\} = f(0) + o(1) = (I - A_0)^{-1} \Sigma_e (I - A_0')^{-1} + o(1),$$

where $f(\cdot)$ is the spectral density of Y_t . Moreover, $E(Y_1 Y_1') = \Sigma_e$, $W_0 = O(n^{-1})$ and $I - W_0 = (n-1)W_0 \Sigma_e^{-1/2} (I - A_0) \Sigma_e (I - A_0)' \Sigma_e^{-1/2}$, we have

$$\begin{aligned} E(\Omega) &= \Sigma_e^{-1/2} \left(W_0 \Sigma_e^{-1/2} (I - A_0) \Sigma_e (I - A_0)' \Sigma_e^{-1/2} \right) \Sigma_e^{-1/2} (I - A_0) (I - A_0)^{-1} \Sigma_e (I - A_0')^{-1} \\ &\quad - \Sigma_e^{-1/2} W_0 \Sigma_e^{-1/2} (I - A_0) \Sigma_e + o(1) \\ &= \Sigma_e^{-1/2} W_0 \Sigma_e^{-1/2} (I - A_0) \Sigma_e - \Sigma_e^{-1/2} W_0 \Sigma_e^{-1/2} (I - A_0) \Sigma_e + o(1) = o(1). \end{aligned}$$

from (A11). Noting that $I - W_0 = O(1)$, it is easy to verify that $E \|R(W_0, \Sigma_e)\|^2 = O(n^{-3})$ by (i) and (ii) of Lemma 3. We have shown (iii) and (iv). For (v), we write

$$\partial Q(\hat{A}_{OLS,0}) - \partial Q_0(\hat{A}_{OLS,0}) = 2(n-1)\{\Omega(\hat{W}, \hat{\Sigma}_e) - \Omega(W_0, \Sigma_e) + R(\hat{W}, \hat{\Sigma}_e) - R(W_0, \Sigma_e)\},$$

then (v) follows easily from Lemma 3 and Hölder's inequality. \square

4.3. Proof of Theorem 2

Assuming Σ is known, then

$$\partial_{\text{RL}} = -\frac{1}{2}\partial \log |W| + \frac{1}{2}\partial Q_R. \quad (\text{A12})$$

We further assume, without loss of generality, that $\Sigma = I$ since any VAR(1) can be pre-multiplied by $\Sigma^{-1/2}$,

$$\Sigma^{-1/2}Y_t = \Sigma^{-1/2}A\Sigma^{1/2}(\Sigma^{-1/2}Y_{t-1}) + \Sigma^{-1/2}e_t$$

to get $Z_t = \tilde{A}Z_{t-1} + \epsilon_t$, where $\tilde{A} = \Sigma^{-1/2}A\Sigma^{1/2}$ has the same characteristic roots of A and $\text{Var}(\epsilon_t) = I$. Let $W(A) = W$, the next lemma is similar to Lemma 3 for \hat{A}_{REML} . We thus omit the proof.

LEMMA 4. For any positive h ,

$$E \left\| \hat{A}_{REML} - A_0 \right\|^h = O(n^{-h/2}), \quad E \left\| W(\hat{A}_{REML}) - W(A_0) \right\|^h = O(n^{-3h/2}).$$

LEMMA 5. Let $\xi(A) = (\bar{Y}_{(1)} - Y_1) - A(\bar{Y}_{(0)} - Y_1)$, then for $\hat{A} = \hat{A}_{OLS}$, $\hat{A}_{OLS,0}$ and \hat{A}_{REML} ,

$$E \left\{ \xi(\hat{A}) \xi(\hat{A})' \right\} = (I - A_0)\Sigma(I - A_0') + o(1), \quad E \left\| \xi(\hat{A}) \xi(\hat{A})' \right\|^h = O(1)$$

Proof of Lemma 5. Writing $\xi(A) = -(I - A)Y_1 + (I - A)\bar{Y}_{(0)} + (n-1)^{-1}(Y_n - Y_1)$,

$$\begin{aligned} \xi(\hat{A}) \xi(\hat{A})' &= (I - \hat{A})Y_1Y_1'(I - \hat{A})' + (I - \hat{A})\bar{Y}_{(0)}\bar{Y}_{(0)}'(I - \hat{A})' + (n-1)^{-2}(Y_n - Y_1)(Y_n - Y_1)' \\ &\quad + (n-1)^{-1} \left\{ (Y_n - Y_1)(\bar{Y}_{(0)} - Y_1)'(I - \hat{A})' + (I - \hat{A})(\bar{Y}_{(0)} - Y_1)(Y_n - Y_1)' \right\} \\ &\quad - (I - \hat{A})\bar{Y}_{(0)}Y_1' - Y_1\bar{Y}_{(0)}'(I - \hat{A}'). \end{aligned}$$

Since $E \left\| \bar{Y}_{(0)} \bar{Y}'_{(0)} \right\|^h = O(n^{-h})$, $E \left\| \bar{Y}_{(0)} Y'_1 \right\| = O(n^{-h/2})$, $E \left\| I - \hat{A} \right\|^h = O(1)$, $E \left\| Y_1 Y'_1 \right\| = O(1)$, thus $E \left\| \xi(\hat{A}) \xi(\hat{A})' \right\|^h = O(1)$. Furthermore,

$$E \left\{ (I - \hat{A}) Y_1 Y_1 (I - \hat{A})' \right\} = (I - A_0) E (Y_1 Y_1) (I - A_0)' \\ + E \left\{ (I - \hat{A}) Y_1 Y_1 (I - \hat{A})' - (I - A_0) Y_1 Y_1 (I - A_0)' \right\}$$

The first term is $(I - A_0) \Sigma (I - A_0)'$ and the second term is bounded by

$$2E^{1/2} \left\| \hat{A} - A_0 \right\|^2 E \left\| Y_1 Y_1 (I - \hat{A})' \right\| = O(n^{-1/2})$$

by Lemma 4. □

LEMMA 6. Let Y_t follow (1), then $E(\hat{H}_{IWLs} - H) = E(\hat{H}_{OLS,0} - H) + o(n^{-1})$, where the bias of $\hat{H}_{OLS,0}$, which is $O(n^{-1})$, is given in Yamamoto & Kunitomo (1984).

Proof of Lemma 6. This lemma is a corollary of Theorem 1 □

Proof of Theorem 2. Following from Lemma 6, it is sufficient to show that $E\{\hat{A}_{REML} - \hat{A}_{IWLs}\} = o(n^{-1})$. Let S_Y and Γ be defined as (A10), then

$$\partial Q_R(A) = -2 \sum_{t=2}^n (Y_t - \bar{Y}_{(1)}) (Y_{t-1} - \bar{Y}_{(0)})' + 2AS_Y - 2(n-1)W(A)\xi(A)(\bar{Y}_{(0)} - Y_1)' \\ + (n-1) \left\{ \xi'(A) \frac{\partial W(A)}{\partial a_{ij}} \xi(A) \right\}_{i,j=1}^k$$

With tedious algebraic computation,

$$\left\{ \xi'(A) \frac{\partial W(A)}{\partial a_{ij}} \xi(A) \right\}_{i,j=1}^k = 2(n-1)W(A)\xi(A)\xi'(A)W(A)(I-A)$$

and

$$-\partial \log |W(A)| = - \left[\text{trace} \left\{ W^{-1}(A) \frac{\partial W(A)}{\partial a_{ij}} \right\} \right]_{i,j=1}^k = -2(n-1)W(A)(I_k - A).$$

In conjunction with (A12), we get

$$\partial RL(A) = - \sum_{t=2}^n (Y_t - \bar{Y}_{(1)}) (Y_{t-1} - \bar{Y}_{(0)})' + AS_Y - (n-1)W(A)\xi(A)(\bar{Y}_{(0)} - Y_1)' \\ + (n-1)^2 W(A)\xi(A)\xi'(A)W(A)(I-A) - (n-1)W(A)(I-A).$$

Since $\partial Q_0(\hat{A}_{IWLS}) = 0$

$$\sum_{t=2}^n (Y_t - \bar{Y}_{(1)}) (Y_{t-1} - \bar{Y}_{(0)})' = \hat{A}_{IWLS} S_Y - (n-1) W(A_0) \xi(\hat{A}_{IWLS})'$$

and since $\partial RL(\hat{A}_{REML}) = 0$, we have

$$0 = (\hat{A}_{IWLS} - \hat{A}_{REML}) S_Y + (n-1) W(A_0) \left\{ \hat{A}_{IWLS} - \hat{A}_{REML} \right\} \Gamma + R_1(\hat{A}_{REML}) + R_2(\hat{A}_{REML}),$$

where

$$R_1(A) = (n-1) \{W(A) - W(A_0)\} \{(\bar{Y}_{(1)} - Y_1) - A(\bar{Y}_{(0)} - Y_1)\} (\bar{Y}_{(0)} - Y_1)'$$

$$R_2(A) = -(n-1)^2 W(A) \xi(A) \xi'(A) W(A) (I - A) + (n-1) W(A) (I - A).$$

Thus

$$vec(\hat{A}_{REML} - \hat{A}_{IWLS}) = \{S_Y \otimes I + (n-1) \Gamma \otimes W(A_0)\}^{-1} vec\{R_1(\hat{A}_{REML}) + R_2(\hat{A}_{REML})\}.$$

Following from the proof of Theorem 1, $(n-1)E(S_Y \otimes I + (n-1) \Gamma \otimes W(A_0)) = \Sigma_Y + o(1)$ and $E \|S_Y \otimes I + (n-1) \Gamma \otimes W(A_0)\|^{-h} = O(n^{-h})$. We will complete the proof by showing

$$E \left\| R_1(\hat{A}_{REML}) \right\|^h = O(n^{-h/2}), \quad (A13)$$

$$E \left\{ R_2(\hat{A}_{REML}) \right\} = o(1), \quad E \left\| R_2(\hat{A}_{REML}) \right\|^h = O(1). \quad (A14)$$

Since $E(Y_1 Y_1) = \Sigma = I$, we get $E \left\| \xi'(\hat{A}) \xi(\hat{A}) - (I - A_0)(I - A_0') \right\|^h = O(n^{-h/2})$. By Lemmas 4 and 5, (A13) follows by Hölder's inequality. Let $R_{2a}(A) = -(n-1)^2 W(A) \xi(A) \xi'(A) W(A) (I - A)$ and $R_{2b}(A) = (n-1) W(A) (I - A)$ so that $R_2(A) = R_{2a}(A) + R_{2b}(A)$. Then

$$\begin{aligned} E \left\{ R_{2a}(\hat{A}_{REML}) \right\} &= E \left\{ R_{2a}(A_0) \right\} + E \left\{ R_{2a}(\hat{A}_{REML}) - R_{2a}(A_0) \right\} \\ &= -(n-1)^2 W(A_0) (I - A_0) (I - A_0') W(A_0) (I - A_0) + O(n^{-n/2}) \end{aligned}$$

by Lemma 4, 5 and Hölder's inequality, and similarly

$$E \left\{ R_{2b}(\hat{A}_{REML}) \right\} = (n-1) W(A_0) (I - A_0) + O(n^{-n/2}).$$

We have

$$\begin{aligned}
E \left\{ R_2 \left(\hat{A}_{REML} \right) \right\} &= -(n-1) W(A_0) (I - A_0) \{ I - (n-1) W(A_0) (I - A_0) (I - A'_0) \} + O(n^{-n/2}) \\
&= -(n-1) W(A_0) (I - A_0) W(A_0) + O(n^{-n/2}) \\
&= O(n^{-n/2}). \quad \square
\end{aligned}$$

4.4. Proof of Theorem 3

Proof. Letting $b_s = (H_1 + H_s - I_k) Y_1 + e_s$ for $s = 2, \dots, p$ we have $R_{s-1} = H U_{s-1} + b_s$ and letting $b_{p+1} = (n-p)^{1/2} (H_1 - I_k) Y_1 + (n-p)^{-1/2} \sum_{t=p+1}^n e_t$ gives $R_p = H U_p + b_p$. Hence the estimating equations for \hat{H} can be re-expressed as

$$\begin{aligned}
&2\hat{\Sigma}_e^{-1} \sum_{t=p+1}^n e_t L'_t + \sum_{i=1}^p \sum_{j=1}^p \hat{\Sigma}_e^{-1/2} \hat{W}_{ij} \hat{\Sigma}_e^{-1/2} (b_{i+1} U'_j + b_{j+1} U'_i) \quad (A15) \\
&= 2\hat{\Sigma}_e^{-1} (\hat{H} - H) \sum_{t=p+1}^n L_t L'_t + \sum_{i=1}^p \sum_{j=1}^p \hat{\Sigma}_e^{-1/2} \hat{W}_{ij} \hat{\Sigma}_e^{-1/2} (\hat{H} - H) (U_i U'_j + U_j U'_i).
\end{aligned}$$

From this stage onwards, we provide the proof for the $AR(1)$ case with $p = 1$ to avoid cumbersome notation, noting that the proof for the general $AR(p)$ case follows along similar lines. We will assume that the process Y_t has been expressed in canonical form (page 608, Fuller, 1996) so that $\Sigma_e = I$ (the result for the case of general Σ_e can then be obtained by suitable pre/post multiplication). Since $n(I_k - \hat{H}_1) = O_p(1)$ and $\hat{\Sigma}_e \rightarrow I_k$ in probability, we have

$$\hat{W} = \hat{W}_{11} = \left[I_k + (n-1) \hat{\Sigma}_e^{-1/2} (I_k - \hat{H}_1) \hat{\Sigma}_e (I_k - \hat{H}_1') \hat{\Sigma}_e^{-1/2} \right]^{-1} = I_k + o_p(1).$$

Furthermore, from Lemma 10.3.2 of Fuller (1996),

$$\left(\frac{1}{n} \sum_{t=2}^n e_t L'_t, \frac{1}{n} b_2 U'_1, \frac{1}{n^2} \sum_{t=p+1}^n L_t L'_t, \frac{1}{n^2} U_1 U'_1 \right) \rightarrow (\Upsilon' - W(1) \zeta', W(1) \zeta', G - \zeta \zeta', \zeta \zeta'),$$

in distribution, where $(\Upsilon, W(1), \zeta, G)$ are defined in Lemma 10.3.1 of Fuller (1996). Hence, dividing both sides of (A15) by n and applying the convergence results stated above yields

$$\begin{aligned}
&(I + o_p(1)) (\Upsilon' - W(1) \zeta' + o_p(1)) + (I + o_p(1)) (W(1) \zeta' + o_p(1)) \\
&= (I + o_p(1)) n (\hat{H} - I) (G - \zeta \zeta' + o_p(1)) + (I + o_p(1)) n (\hat{H} - I) (\zeta \zeta' + o_p(1)),
\end{aligned}$$

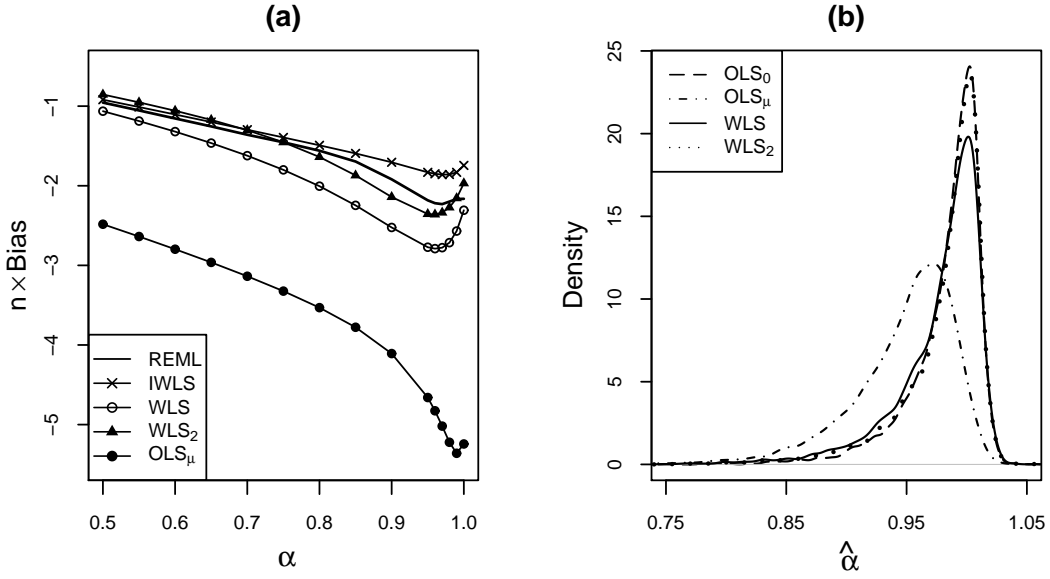


Fig. 2. (a) Bias of $\hat{\alpha}_{REML}$, $\hat{\alpha}_{IWLS}$, $\hat{\alpha}_{WLS}$, $\hat{\alpha}_{WLS_2}$ and $\hat{\alpha}_{OLS}$. (b) Empirical densities of $\hat{\alpha}_{OLS_0}$, $\hat{\alpha}_{OLS_\mu}$, $\hat{\alpha}_{WLS}$ and $\hat{\alpha}_{WLS_2}$ when $\alpha = 1$. All plots are based on simulations of 10,000 replications with $n = 100$.

and thus $n(\hat{H} - I) \rightarrow \Upsilon'G^{-1}$ in distribution. □

5. ADDITIONAL SIMULATIONS

5.1. Univariate AR processes

We generate 10,000 replications of size $n = 100$ AR(1) series for various values of α ranging from 0.5 to 1. We compare the bias properties of 5 estimators, the exact REML, $\hat{\alpha}_{REML}$, OLS with intercept, $\hat{\alpha}_{OLS}$, infeasible weighted, weighted and iterated weighted least squares, $\hat{\alpha}_{IWLS}$, $\hat{\alpha}_{WLS}$ and $\hat{\alpha}_{WLS_2}$ respectively, where $\hat{\alpha}_{REML}$ is the maximiser of (3) and

$$\hat{\alpha}_{IWLS} = \frac{\sum_{t=2}^n (Y_t - \bar{Y}_{(1)})(Y_{t-1} - \bar{Y}_{(0)}) + w(\alpha_0)(n-1)(\bar{Y}_{(1)} - Y_1)(\bar{Y}_{(0)} - Y_1)}{\sum_{t=2}^n (Y_{t-1} - \bar{Y}_{(0)})^2 + w(\alpha_0)(n-1)(\bar{Y}_{(0)} - Y_1)^2}. \quad (A16)$$

The two feasible weighted least squares $\hat{\alpha}_{WLS}$ and $\hat{\alpha}_{WLS_2}$ are computed analogously with $w(\alpha_0)$ being replaced by $w(\hat{\alpha}_{OLS})$ and $w(\hat{\alpha}_{WLS})$ respectively.

Table 2. Bias and root mean square error of univariate AR(1)

α	REML		WLS		WLS ₂		OLS	
	$n \times$ Bias	$\sqrt{n} \times$ RMSE	$n \times$ Bias	$\sqrt{n} \times$ RMSE	$n \times$ Bias	$\sqrt{n} \times$ RMSE	$n \times$ Bias	$\sqrt{n} \times$ RMSE
1	-2.164	2.002	-2.308	2.091	-1.969	2.016	-5.243	2.608
0.95	-2.183	2.345	-2.772	2.387	-2.355	2.349	-4.660	2.650
0.9	-1.919	2.494	-2.524	2.514	-2.141	2.504	-4.108	2.673
0.7	-1.359	2.823	-1.622	2.847	-1.302	2.876	-3.137	2.888
0.5	-0.956	3.005	-1.063	3.031	-0.854	3.053	-2.483	3.036

Figure 2(a) shows the bias of these 5 estimators. We see clearly that $\hat{\alpha}_{REML}$, $\hat{\alpha}_{IWLS}$, $\hat{\alpha}_{WLS}$ and $\hat{\alpha}_{WLS_2}$ have significantly less bias than $\hat{\alpha}_{OLS}$ for all values of α . Though the bias of $\hat{\alpha}_{WLS}$, which uses $\hat{\alpha}_{OLS}$ in the weight function $w(\cdot)$, increases slightly from that of $\hat{\alpha}_{REML}$, the bias of $\hat{\alpha}_{WLS_2}$, obtained by using $\hat{\alpha}_{WLS}$ in $w(\cdot)$, is as small as that of $\hat{\alpha}_{REML}$ and the infeasible estimator $\hat{\alpha}_{IWLS}$. Thus, further iterations are not likely to produce any significant improvement. The most dramatic improvement shown by $\hat{\alpha}_{WLS}$ and $\hat{\alpha}_{WLS_2}$ can be seen in Figure 2(b), where we plot their empirical densities, with those of $\hat{\alpha}_{OLS}$ and $\hat{\alpha}_{OLS,0}$ for $\alpha = 1$. It is well-known that at the unit root the distribution of $\hat{\alpha}_{OLS}$ is different from that of $\hat{\alpha}_{OLS,0}$ and centred further to the left, as seen in Figure 2(b), resulting in $\hat{\alpha}_{OLS}$ having a higher bias than $\hat{\alpha}_{OLS,0}$. The plot shows that $\hat{\alpha}_{WLS}$ has the same distribution as $\hat{\alpha}_{OLS,0}$ and is unaffected by the intercept in the model. Indeed, the density of $\hat{\alpha}_{WLS_2}$ sits almost exactly on that of $\hat{\alpha}_{OLS,0}$. This finding, shown theoretically in §2.2, is of importance for unit root tests and tests of cointegration in vector AR models, since it is known that the inclusion of an intercept significantly deteriorates the power of unit root tests based on estimators such as OLS, whose limiting distribution is affected by the intercept.

Table 2 reports $n \times$ bias as well as $\sqrt{n} \times$ RMSE of WLS, WLS₂, OLS and REML for a univariate AR(1). It is seen that WLS₂ has even less bias than WLS and both provide significant bias reduction over OLS even far from the unit root. The RMSE of both WLS and WLS₂ is almost uniformly less than that of OLS, and the reduction can be as high as 23% (40% in terms of MSE).

Table 3. *Bias and root mean square error of bivariate AR(1)– $z_{11} = 1$*

a_{22}	Est.	$n \times \text{Bias}$				$\sqrt{n} \times \text{RMSE}$			
		a_{11}	a_{21}	a_{12}	a_{22}	a_{11}	a_{21}	a_{12}	a_{22}
1	WLS	-4.452	0.044	0.032	-4.508	0.703	0.541	0.550	0.715
	WLS ₂	-4.015	0.051	0.027	-4.079	0.668	0.528	0.537	0.680
	OLS	-7.000	-0.003	0.037	-7.002	0.916	0.601	0.603	0.923
0.9	WLS	-2.993	-1.091	-0.186	-3.847	0.528	0.416	0.651	0.767
	WLS ₂	-2.515	-0.882	-0.247	-3.487	0.490	0.400	0.652	0.749
	OLS	-5.972	-2.129	0.298	-4.861	0.782	0.511	0.674	0.840
0.8	WLS	-2.806	-1.298	-0.183	-3.227	0.495	0.401	0.763	0.823
	WLS ₂	-2.343	-1.069	-0.245	-2.851	0.457	0.388	0.777	0.815
	OLS	-5.679	-2.413	0.357	-4.105	0.742	0.483	0.750	0.862
0.7	WLS	-2.744	-1.361	-0.174	-2.804	0.480	0.395	0.859	0.881
	WLS ₂	-2.272	-1.126	-0.247	-2.416	0.442	0.384	0.888	0.881
	OLS	-5.543	-2.509	0.403	-3.634	0.723	0.472	0.817	0.899

5.2. *Bivariate AR processes*

In Tables 3 and 4 we report the simulation results for a bivariate AR(1) model. Since the coefficient matrix can always be diagonalised and standard cointegration tests are invariant to pre/post multiplication by non-singular matrices, we considered a diagonal coefficient matrix A with different configurations for the diagonal entries (a_{11}, a_{22}) . The different parameter configurations for A include two unit roots, one unit root with the other either close to unity or far from unity, as well as both roots being less than unity. In all the cases, both WLS and WLS₂ offer dramatic bias reduction compared to OLS, and the RMSE of both weighted least squares estimators is uniformly better than that of OLS. The bias reduction is significant even for roots far from unity and is more pronounced for the iterated estimator WLS₂.

Finally, Tables 5, 6, 7 and 8, report results for a bivariate AR(2), in which both the coefficient matrices A_1 and A_2 were set to be diagonal. The parameter configurations were chosen in such a way that the

Table 4. Bias and root mean square error of bivariate AR(1)– $z_{11} = 0.95$

a_{22}	Est.	$n \times \text{Bias}$				$\sqrt{n} \times \text{MSE}$			
		a_{11}	a_{21}	a_{12}	a_{22}	a_{11}	a_{21}	a_{12}	a_{22}
1	WLS	-4.301	-0.101	-0.782	-3.354	0.756	0.603	0.444	0.587
	WLS ₂	-3.903	-0.147	-0.624	-2.859	0.731	0.596	0.428	0.547
	OLS	-5.633	0.266	-1.632	-6.323	0.869	0.655	0.542	0.836
0.9	WLS	-3.956	-0.512	0.309	-3.834	0.716	0.597	0.706	0.809
	WLS ₂	-3.445	-0.423	0.229	-3.454	0.690	0.590	0.705	0.796
	OLS	-5.788	-0.838	0.624	-5.048	0.857	0.644	0.741	0.887
0.8	WLS	-3.574	-0.994	0.487	-3.044	0.658	0.555	0.799	0.860
	WLS ₂	-3.061	-0.844	0.354	-2.712	0.633	0.550	0.810	0.859
	OLS	-5.430	-1.536	1.000	-4.061	0.802	0.594	0.813	0.902
0.7	WLS	-3.382	-1.192	0.535	-2.527	0.628	0.533	0.874	0.907
	WLS ₂	-2.866	-1.009	0.379	-2.232	0.602	0.529	0.896	0.913
	OLS	-5.202	-1.825	1.108	-3.467	0.770	0.570	0.870	0.931

two largest roots $(z_1, z_2)'$ of the two processes $(Y_{t,1}, Y_{t,2})'$ covered a range of different values, while the two smaller roots were fixed at 0.8 each. Tables 5 and 6 report the results for the entries of the matrix $H_1 = A_1 + A_2$, which is the coefficient matrix of Y_{t-1} in the Dickey-Fuller representation of Y_t , while Tables 7 and 8 have the corresponding results for the matrix $H_2 = -A_2$, which is the coefficient matrix of ΔY_{t-1} . The matrix H_1 is relevant in likelihood ratio type tests of cointegration, while H_2 would be relevant for forecasting purposes. The two weighted least squares REML estimators have uniformly smaller bias and RMSE than the OLS estimator for almost all the configurations of H_1 and the bias reduction can be substantial, even far from the unit root. From Tables 5–8 we see that our weighted least squares procedure has smaller bias and RMSE than the OLS for configurations that are closer to the unit root. For parameter values with all roots far from unity, we see that there are very few situations where OLS has a smaller bias than the weight least squares estimators but the corresponding RMSE is larger than, or equal to that of the

Table 5. *Bias and root mean square error of bivariate AR(2) – H₁*

z_1	z_2	Est.	$n \times \text{Bias}$				$\sqrt{n} \times \text{MSE}$			
			$h_{11}^{(1)}$	$h_{21}^{(1)}$	$h_{12}^{(1)}$	$h_{22}^{(1)}$	$h_{11}^{(1)}$	$h_{21}^{(1)}$	$h_{12}^{(1)}$	$h_{22}^{(1)}$
1	1	WLS	-0.953	0.009	0.021	-0.945	0.178	0.151	0.153	0.176
		WLS ₂	-0.942	0.006	0.018	-0.933	0.176	0.149	0.151	0.174
		OLS	-1.835	-0.001	0.015	-1.821	0.259	0.187	0.191	0.258
0.95		WLS	-0.555	-0.117	-0.033	-1.151	0.130	0.112	0.181	0.211
		WLS ₂	-0.543	-0.115	-0.034	-1.122	0.129	0.111	0.180	0.209
		OLS	-1.552	-0.412	0.063	-1.920	0.222	0.165	0.216	0.288
0.9		WLS	-0.466	-0.155	-0.041	0.113	0.113	0.099	0.214	0.251
		WLS ₂	-0.453	-0.151	-0.040	-1.305	0.111	0.097	0.212	0.247
		OLS	-1.440	-0.524	0.077	-2.087	0.204	0.151	0.245	0.320
0.8		WLS	-0.428	-0.192	-0.033	-1.812	0.103	0.090	0.292	0.340
		WLS ₂	-0.416	-0.187	-0.030	-1.703	0.102	0.088	0.289	0.332
		OLS	-1.346	-0.587	0.095	-2.583	0.189	0.138	0.318	0.405

z_1 and z_2 are the two largest roots of the determinant of AR polynomial. The other two roots are both .8.

weighted least squares estimates. The overall conclusion to be drawn from Tables 3-8 is that in general the weighted least squares REML estimates provide a significant advantage over OLS in terms of bias and RMSE reduction, not only near the unit root but also for stationary vector processes.

Table 6. Bias and root mean square error of bivariate AR(2) – H_1

z_1	z_2	Est.	$n \times \text{Bias}$				$\sqrt{n} \times \text{MSE}$				
			$h_{11}^{(1)}$	$h_{21}^{(1)}$	$h_{12}^{(1)}$	$h_{22}^{(1)}$	$h_{11}^{(1)}$	$h_{21}^{(1)}$	$h_{12}^{(1)}$	$h_{22}^{(1)}$	
0.95	0.95	WLS	-0.974	0.007	0.017	-0.967	0.200	0.178	0.181	0.200	
		WLS ₂	-0.934	0.006	0.016	-0.926	0.198	0.176	0.179	0.197	
		OLS	-1.918	-0.014	0.008	-1.901	0.292	0.231	0.230	0.290	
	0.90	WLS	-0.830	-0.054	0.036	-1.198	0.182	0.164	0.222	0.248	
		WLS ₂	-0.790	-0.053	0.035	-1.130	0.179	0.161	0.219	0.243	
		OLS	-1.811	-0.233	0.172	-2.008	0.274	0.215	0.266	0.327	
	0.80	WLS	-0.715	-0.119	0.043	-1.652	0.163	0.150	0.299	0.342	
		WLS ₂	-0.677	-0.117	0.045	-1.526	0.160	0.147	0.296	0.334	
		OLS	-1.648	-0.398	0.265	-2.415	0.249	0.195	0.334	0.408	
0.90	0.90	WLS	-1.081	0.011	0.005	-1.090	0.238	0.217	0.218	0.240	
		WLS ₂	-1.011	0.011	0.004	-1.020	0.233	0.214	0.216	0.236	
		OLS	-1.971	-0.010	-0.004	-1.973	0.320	0.263	0.261	0.321	
	0.80	WLS	-0.932	-0.057	0.024	-1.530	0.219	0.203	0.301	0.338	
		WLS ₂	-0.868	-0.058	0.026	-1.401	0.215	0.200	0.297	0.330	
		OLS	-1.830	-0.221	0.162	-2.345	0.298	0.243	0.337	0.405	
	0.80	0.80	WLS	-1.338	0.020	-0.016	-1.377	0.320	0.296	0.298	0.326
			WLS ₂	-1.218	0.019	-0.017	-1.255	0.314	0.292	0.294	0.320
			OLS	-2.204	0.008	-0.023	-2.245	0.389	0.332	0.330	0.394

z_1 and z_2 are the two largest roots of the determinant of AR polynomial. The other two roots are both .8.

Table 7. *Bias and root mean square error of bivariate AR(2)-H₂*

z_1	z_2	Est.	$n \times \text{Bias}$				$\sqrt{n} \times \text{MSE}$			
			$h_{11}^{(2)}$	$h_{21}^{(2)}$	$h_{12}^{(2)}$	$h_{22}^{(2)}$	$h_{11}^{(2)}$	$h_{21}^{(2)}$	$h_{12}^{(2)}$	$h_{22}^{(2)}$
1	1	WLS	-4.341	0.064	-0.031	-4.416	0.975	0.870	0.864	0.978
		WLS ₂	-4.336	0.054	-0.040	-4.411	0.975	0.870	0.864	0.978
		OLS	-6.556	0.053	-0.088	-6.666	1.133	0.917	0.906	1.144
0.95		WLS	-4.540	0.518	0.151	-3.187	0.990	0.889	0.865	0.925
		WLS ₂	-4.530	0.493	0.152	-3.168	0.990	0.889	0.865	0.925
		OLS	-6.540	0.616	-0.196	-4.344	1.137	0.931	0.897	1.009
0.9		WLS	-4.311	0.698	0.170	-2.766	0.974	0.887	0.873	0.922
		WLS ₂	-4.302	0.645	0.176	-2.737	0.973	0.886	0.873	0.921
		OLS	-6.292	0.800	-0.261	-3.659	1.116	0.929	0.898	0.984
0.8		WLS	-3.907	0.864	0.074	-2.214	0.948	0.880	0.899	0.931
		WLS ₂	-3.904	0.744	0.085	-2.170	0.946	0.877	0.899	0.929
		OLS	-5.858	0.915	-0.380	-2.835	1.082	0.918	0.915	0.969

z_1 and z_2 are the two largest roots of the determinant of AR polynomial. The other two roots are both .8.

Table 8. Bias and root mean square error of bivariate AR(2)-H₂

z_1	z_2	Est.	$n \times \text{Bias}$				$\sqrt{n} \times \text{MSE}$				
			$h_{11}^{(2)}$	$h_{21}^{(2)}$	$h_{12}^{(2)}$	$h_{22}^{(2)}$	$h_{11}^{(2)}$	$h_{21}^{(2)}$	$h_{12}^{(2)}$	$h_{22}^{(2)}$	
0.95	0.95	WLS	-3.227	0.070	-0.029	-3.307	0.937	0.883	0.877	0.941	
		WLS ₂	-3.215	0.064	-0.035	-3.296	0.936	0.884	0.877	0.941	
		OLS	-3.659	0.101	-0.032	-3.777	0.977	0.907	0.905	0.981	
	0.90	WLS	-3.241	0.338	-0.136	-2.887	0.933	0.888	0.899	0.949	
		WLS ₂	-3.233	0.303	-0.127	-2.874	0.933	0.888	0.899	0.949	
		OLS	-3.378	0.753	-0.587	-3.357	0.956	0.914	0.917	0.984	
	0.80	WLS	-3.043	0.691	-0.237	-2.403	0.917	0.887	0.932	0.969	
		WLS ₂	-3.037	0.595	-0.215	-2.385	0.916	0.885	0.932	0.969	
		OLS	-3.075	1.321	-1.007	-2.890	0.933	0.912	0.944	1.000	
0.90	0.90	WLS	-2.809	0.061	-0.004	-2.880	0.944	0.912	0.906	0.953	
		WLS ₂	-2.808	0.054	-0.012	-2.880	0.944	0.912	0.906	0.953	
		OLS	-2.853	0.089	0.005	-2.948	0.957	0.924	0.923	0.965	
	0.80	WLS	-2.707	0.398	-0.178	-2.400	0.931	0.912	0.954	0.988	
		WLS ₂	-2.716	0.335	-0.156	-2.403	0.932	0.912	0.955	0.988	
		OLS	-2.518	0.828	-0.632	-2.567	0.932	0.928	0.963	1.003	
	0.80	0.80	WLS	-2.279	0.032	0.047	-2.296	0.981	0.972	0.968	0.993
			WLS ₂	-2.305	0.022	0.037	-2.322	0.982	0.973	0.968	0.995
			OLS	-2.099	0.047	0.054	-2.128	0.981	0.977	0.976	0.993

z_1 and z_2 are the two largest roots of the determinant of AR polynomial. The other two roots are both .8.