Weighted Least Squares Approximate Restricted Likelihood Estimation for Vector Autoregressive Processes

BY WILLA W. CHEN

Department of Statistics, Texas A&M University College Station, Texas 77843, U.S.A.

wchen@stat.tamu.edu

AND ROHIT S. DEO

New York University, 44 West 4th Street, New York, New York 10012, U.S.A.

rdeo@stern.nyu.edu

SUMMARY

We derive a weighted least squares approximate restricted likelihood estimator for a k-dimensional pth order autoregressive model with intercept, for which exact likelihood optimization is generally infeasible due to the parameter space which is complicated and high-dimensional, involving pk^2 parameters. The weighted least squares estimator has significantly reduced bias and mean squared error than the ordinary least squares estimator for both stationary and non-stationary processes. Furthermore, at the unit root, the limiting distribution of the weighted least squares approximate restricted likelihood estimator is shown to be the zero-intercept Dickey–Fuller distribution, unlike the ordinary least squares with intercept estimator which has a different distribution with significantly higher bias.

Some key words: Autoregressive; Bias; Restricted maximum likelihood; Unbiased estimating equations.

1. INTRODUCTION

The pth order autoregressive models, AR(p), are commonly used in multiple time series as they are easy to estimate via least squares and to use in forecasting. Including an intercept in univariate autoregressive models increases the finite sample bias of the ordinary least squares and maximum likelihood estimators (Shaman & Stine, 1988, Cheang & Reinsel, 2000). For example, in the AR(1) with coefficient α and no intercept, the bias of both the maximum likelihood and ordinary least squares estimators based on n observations is $-2\alpha n^{-1}$, but becomes $-(1+3\alpha)\,n^{-1}$ when an intercept is included. Thus, the bias is almost doubled when the autoregressive coefficient is close to unity. At the unit root the intercept has an impact that persists even asymptotically, with different limiting distributions for the ordinary least squares estimator with and without intercept.

Cheang & Reinsel (2000) found that restricted maximum likelihood estimation of autoregressive models with intercept yields estimates with reduced bias which is identical up to $O\left(n^{-1}\right)$ to that of the maximum likelihood and ordinary least squares estimates for zero-intercept autoregressive models. See also Kang et al (2003). Hence, including an intercept has no impact on the restricted maximum likelihood estimator and its bias does not increase, unlike that of the maximum likelihood and ordinary least squares estimators. Other advantages of the restricted maximum likelihood estimator in time series applications include better model selection properties (Tunnicliffe-Wilson, 1989) and well-behaved likelihood ratio tests (Chen & Deo, 2009).

However, the restricted likelihood optimisation is infeasible in vector autoregressive models, where the parameter space is very complicated and defined implicitly through the roots of a characteristic equation. We obtain a weighted least squares approximate restricted maximum likelihood estimator which is easy to compute and has the same superior bias properties of the ordinary least squares estimator without intercept. At the unit root the limiting distribution of the proposed estimator for the AR model with intercept is identical to the Dickey–Fuller distribution for the zero-intercept model, unlike that of the ordinary least squares with intercept estimator. Our results have consequences for forecasting, unit root tests and cointegration tests.

2. THE APPROXIMATE RESTRICTED LIKELIHOOD ESTIMATOR

2·1. *Motivation*

Assume that $Y = (Y_1^{\mathrm{T}}, ..., Y_n^{\mathrm{T}})^{\mathrm{T}}$ follows the k dimensional AR(p) process, given by

$$Y_t = \mu + \widetilde{Y}_t, \quad \widetilde{Y}_t = \sum_{i=1}^{p} A_i \widetilde{Y}_{t-i} + e_t \tag{1}$$

where e_t is an independent $N(0, \Sigma_e)$ series and the roots of $\left|I_k z^p - \sum_{i=1}^{p} A_i z^{p-i}\right| = 0$ are at most one in absolute value. The initial values of $\widetilde{Y}_t$ are assumed to be $\widetilde{Y}_t = 0$ for $t = 0, -1, ..., -p+1$. Letting $\Sigma = \mathrm{var}(Y)$, the log restricted likelihood for Y is

$$\mathrm{RL} = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}\log\left|X^{\mathrm{T}}\Sigma^{-1}X\right| - \frac{1}{2}Q, \tag{2}$$

where $Q = Y^{\mathrm{T}}\Sigma^{-1}Y - Y^{\mathrm{T}}\Sigma^{-1}X\left(X^{\mathrm{T}}\Sigma^{-1}X\right)^{-1}X^{\mathrm{T}}\Sigma^{-1}Y$, and X is the $nk \times k$ matrix $(I_k, ..., I_k)^{\mathrm{T}}$. For simplicity, we illustrate the weighted least squares approximate restricted maximum likelihood estimator through the univariate AR(1) model, where $\widetilde{Y}_t = \alpha\widetilde{Y}_{t-1} + e_t$ in (1). From (2), the restricted maximum likelihood estimator, $\hat{\alpha}_{\mathrm{REML}}$ is the minimiser of

$$\mathrm{RL}(\alpha, \sigma_e^2) = -\frac{(n-1)}{2}\log\sigma_e^2 + \frac{1}{2}\log w(\alpha) - \frac{1}{2}Q, \tag{3}$$

where $w(\alpha) = \{1 + (n-1)(1-\alpha)^2\}^{-1}$. The ordinary least squares with intercept estimator, $\hat{\alpha}_{\mathrm{OLS}}$, minimises the objective function

$$Q_{\mathrm{OLS}}(\alpha) = \frac{1}{\sigma_e^2}\sum_{t=2}^{n}\left\{Y_t - \bar{Y}_{(1)} - \alpha\left(Y_{t-1} - \bar{Y}_{(0)}\right)\right\}^2$$

where $\bar{Y}_{(1)} = (n-1)^{-1}\sum_{t=2}^{n} Y_t$ and $\bar{Y}_{(0)} = (n-1)^{-1}\sum_{t=2}^{n} Y_{t-1}$. The log restricted likelihood in (3) has an unbiased estimating equation, i.e., $E(\partial\mathrm{RL}/\partial\alpha) = 0$ (Cheang & Reinsel, 2000). However, $Q_{\mathrm{OLS}}(\alpha)$ does not satisfy this condition due to the mean correction, which increases the bias of $\hat{\alpha}_{\text{OLS}}$. Lemma 1 below shows that the exponent term Q in (3) can be written as

$$Q(\alpha) = Q_{\text{OLS}}(\alpha) + w(\alpha)\, Q_{\text{C}}(\alpha), \tag{4}$$

where $Q_{\text{C}}(\alpha) = (n-1)^{-1}\sigma_e^2\{\bar{Y}_{(1)} - Y_1 - \alpha(\bar{Y}_{(0)} - Y_1)\}^2$. Since $E_\alpha(Q_{\text{C}}) = w^{-1}(\alpha)$, we get

$$0 = E_\alpha\left\{\frac{\partial \text{RL}(\alpha)}{\partial\alpha}\right\} = -E_\alpha\left\{\frac{\partial Q_{\text{OLS}}(\alpha)}{\partial\alpha} + w(\alpha)\,\frac{\partial Q_{\text{C}}(\alpha)}{\partial\alpha}\right\}.$$

Thus, the objective function, $Q_{\text{I}}(\alpha) = Q_{\text{OLS}}(\alpha) + w(\alpha_0)\, Q_{\text{C}}(\alpha)$, where α_0 is the true value of α, also has an unbiased estimating equation. Furthermore, $\partial Q_{\text{I}}(\alpha) \simeq \partial \text{RL}(\alpha)$ in the sense that both first derivatives are $O_p(n)$ and their difference, $w^{-1}(\alpha)\,\partial w(\alpha) - Q_{\text{C}}\partial w(\alpha) = O_p(1)$. Therefore, the objective functions RL and Q_{I} have similar estimating equations and the minimiser, $\hat{\alpha}_{\text{IWLS}} = \arg\min_\alpha Q_{\text{I}}(\alpha)$ can be thought of as an infeasible weighted least squares approximation to $\hat{\alpha}_{\text{REML}}$ and should have less bias than $\hat{\alpha}_{\text{OLS}}$, just like $\hat{\alpha}_{\text{REML}}$.

The weighted least squares approximate restricted maximum likelihood estimator that we consider uses a consistent initial estimate of α, say $\hat{\alpha}$, in the weight function $w(\cdot)$ in Q_{I} to yield a feasible estimator, $\hat{\alpha}_{\text{WLS}} = \arg\min_\alpha\{Q_{\text{OLS}}(\alpha) + w(\hat{\alpha})\, Q_{\text{C}}(\alpha)\}$. Since generally $(\hat{\alpha} - \alpha_0) = O_p(n^{-1/2})$ in the stationary case and $(\hat{\alpha} - \alpha_0) = O_p(n^{-1})$ in the unit root case, replacing $w(\alpha_0)$ by $w(\hat{\alpha})$ results in an error of $O_p(n^{-3/2})$ in the stationary case and $O_p(n^{-1})$ in the unit root case, suggesting that the difference between $\hat{\alpha}_{\text{IWLS}}$ and $\hat{\alpha}_{\text{WLS}}$ is small. Indeed, both our theoretical results in Theorems 1 and 2 as well as our simulations show that the feasible weighted least squares estimator approximates the restricted maximum likelihood estimator.

2·2. *The vector autoregressive case*

Let $H = (H_1, ..., H_p)$ where $H_1 = \sum_{i=1}^p A_i$ and $H_i = -\sum_{j=i}^p A_j$ for $i = 2, ..., p$ are the coefficient matrices of the AR(p) model expressed in Dickey–Fuller form. We first give the weighted least squares objective function for the vector autoregressive process. Let $n_p = n - p$, $\bar{Y}_{(1)} = n_p^{-1}\sum_{t=p+1}^n Y_t$, $\bar{Y}_{(0)} = n_p^{-1}\sum_{t=p+1}^n Y_{t-1}$ and $\bar{Z}_s = n_p^{-1}\sum_{t=p+1}^n \Delta Y_{t-s}$

for $s = 2, ..., p$. Set $L_t^{\mathrm{T}} = \{(Y_{t-1} - \bar{Y}_{(0)})^{\mathrm{T}}, (\Delta Y_{t-1} - \bar{Z}_1)^{\mathrm{T}}, ..., (\Delta Y_{t-p+1} - \bar{Z}_{p-1})^{\mathrm{T}}\}$. Define $R = (R_1, ..., R_p)$, where $R_i = Y_{i+1} - Y_1$, $i = 1, \ldots, p-1$, $R_p = n_p^{-1/2} \sum_{t=p+1}^{n} (Y_t - Y_1)$ and $U = (U_1, ..., U_p)$, where $U_s^{\mathrm{T}} = \{(Y_s - Y_1)^{\mathrm{T}}, \Delta Y_s^{\mathrm{T}}, ..., \Delta Y_2^{\mathrm{T}}, 0_{(p-s)\times 1}^{\mathrm{T}}\}$ for $s = 1, ..., p-1$,

$$U_p^{\mathrm{T}} = n_p^{-1/2} \left\{ \sum_{t=p+1}^{n} (Y_{t-1} - Y_1)^{\mathrm{T}}, \sum_{t=p+1}^{n} \Delta Y_{t-1}^{\mathrm{T}}, ..., \sum_{t=p+1}^{n} \Delta Y_{t-p+1}^{\mathrm{T}} \right\}.$$

Also, let $D = \{(I_k - H_1 - H_2)^{\mathrm{T}}, ..., (I_k - H_1 - H_p)^{\mathrm{T}}, n_p^{1/2} (I_k - H_1)^{\mathrm{T}}\} \Sigma_e^{-1/2}$.

LEMMA 1. *The log restricted likelihood in (2) can be written as* $\mathrm{RL} = -1/2\{(n-1)\log|\Sigma_e| - \log|W| + Q\}$, *where* $W = (I_{pk} + D^{\mathrm{T}}\Sigma_e D)^{-1}$ *is positive definite and*

$$Q = \sum_{t=p+1}^{n} \left(Y_t - \bar{Y}_{(1)} - HL_t\right)^{\mathrm{T}} \Sigma_e^{-1} \left(Y_t - \bar{Y}_{(1)} - HL_t\right) + \mathrm{vec}\,(R - HU)^{\mathrm{T}} (I_p \otimes \Sigma_e^{-1/2}) W (I_p \otimes \Sigma_e^{-1/2} \mathrm{vec}\,(R - HU)\,.$$

The expression for the log restricted likelihood in the general case mirrors that for the simple univariate $\mathrm{AR}(1)$ stated in (3) and (4). The first term in Q is the sum of squares for the ordinary least squares estimator, while the second term can be interpreted as a correction term. Though Σ_e and W are unknown, they can be estimated using any consistent estimator of H, such as the ordinary least squares estimator. Let $\hat{\Sigma}_e$ and $\hat{W} = (\hat{W}_{ij})_{i,j=1}^{p}$ be any such consistent estimates of Σ_e and W. The weighted least squares approximate restricted maximum likelihood estimator is the minimiser of Q evaluated at $\hat{\Sigma}_e$ and $\hat{W}$, given by

$$\mathrm{vec}(\hat{H}_{\mathrm{WLS}}) = \left\{ \sum_{t=p+1}^{n} L_t L_t^{\mathrm{T}} \otimes 2\hat{\Sigma}_e^{-1} + \sum_{i=1}^{p}\sum_{j=1}^{p} \left(U_i U_j^{\mathrm{T}} + U_j U_i^{\mathrm{T}}\right) \otimes \hat{\Sigma}_e^{-1/2} \hat{W}_{ij} \hat{\Sigma}_e^{-1/2} \right\}^{-1} \times \mathrm{vec} \left\{ 2\hat{\Sigma}_e^{-1} \sum_{t=p+1}^{n} \left(Y_t - \bar{Y}_{(1)}\right) L_t^{\mathrm{T}} + \sum_{i=1}^{p}\sum_{j=1}^{p} \hat{\Sigma}_e^{-1/2} \hat{W}_{ij} \hat{\Sigma}_e^{-1/2} \left(R_i U_j^{\mathrm{T}} + R_j U_i^{\mathrm{T}}\right) \right\}. \quad (5)$$

The finite sample bias of the ordinary least squares estimator of vector AR models both with and without intercept has been obtained by Yamamoto & Kunitomo (1984). The next theorem shows that the bias of $\hat{H}_{\mathrm{WLS}}$ is identical up to $o\left(n^{-1}\right)$ to that of the ordinary least squares estimator in the model without intercept, $\hat{H}_{\mathrm{OLS},0}$.

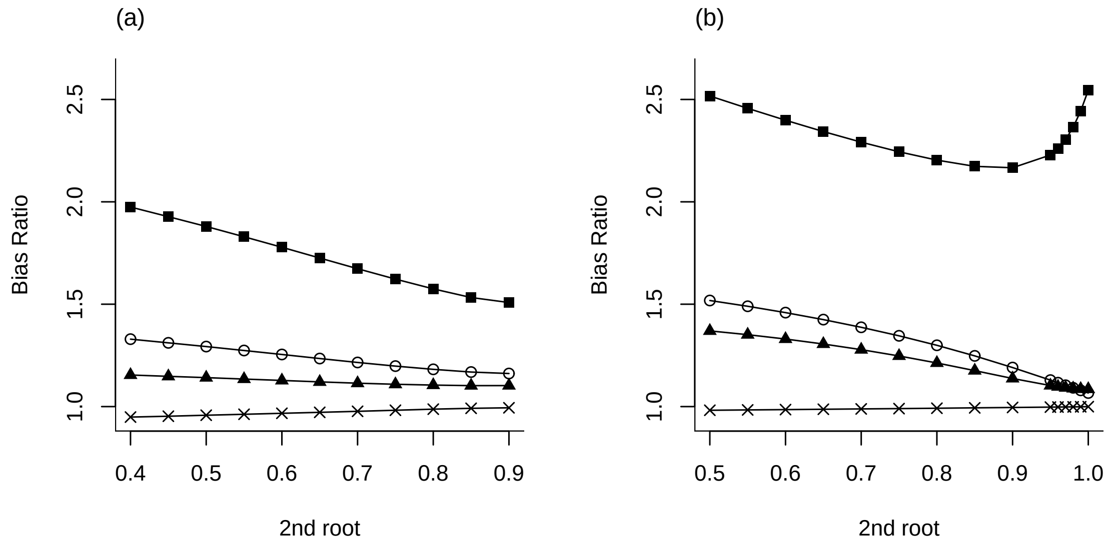

Fig. 1. Bias ratios of $\hat{H}_{\mathrm{OLS}}$ (square), $\hat{H}_{\mathrm{WLS}}$, (circle), $\hat{H}_{\mathrm{WLS}_2}$ (triangle) and $\hat{H}_{\mathrm{OLS},0}$ (cross) against $\hat{H}_{\mathrm{REML}}$ based on 10,000 replications of bivariate AR(1) with $n = 100$. (a) The first root is 0.8. (b) the first root is 1.

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

The next theorem shows that the bias of the exact restricted likelihood estimator from equation (2) obtained under the assumption that Σ_e is known, $\hat{H}_{\mathrm{REML},\Sigma_e}$, is also identical up to $o\left(n^{-1}\right)$ to that of $\hat{H}_{\mathrm{OLS},0}$.

THEOREM 2. *Let Y_t follow (1) and assume that Σ_e is known. Then $E(\hat{H}_{\mathrm{REML},\Sigma_e} - H) = E(\hat{H}_{\mathrm{OLS},0} - H) + o(n^{-1})$, where the bias of $\hat{H}_{\mathrm{OLS},0}$, which is $O(n^{-1})$, is given in Yamamoto & Kunitomo (1984).*

The bias of the restricted maximum likelihood estimator when Σ_e is unknown, $\hat{H}_{\mathrm{REML}}$, is expected to be no better than that of $\hat{H}_{\mathrm{REML},\Sigma_e}$, which is the same as that of $\hat{H}_{\mathrm{OLS},0}$. Indeed, Figure 1 shows that the bias of $\hat{H}_{\mathrm{REML}}$ is almost identical to that of $\hat{H}_{\mathrm{OLS},0}$. Thus, Theorems 1 and 2 together show that $\hat{H}_{\mathrm{WLS}}$ attains the bias of both $\hat{H}_{\mathrm{REML}}$ and $\hat{H}_{\mathrm{OLS},0}$.

For a bivariate AR(1) model with $n = 100$ observations, Figure 1 compares the bias of the ordinary least squares with and without intercept estimators, denoted by $\hat{H}_{\rm OLS}$ and $\hat{H}_{\rm OLS,0}$ respectively, the weighted least squares estimator, $\hat{H}_{\rm WLS}$, and the iterated weighted least squares estimator, $\hat{H}_{\rm WLS_2}$, computed using $\hat{H}_{\rm WLS}$ in $\hat{W}$, to that of the restricted maximum likelihood estimator, $\hat{H}_{\rm REML}$. The coefficient matrix H was configured as (6) in §3 with one root fixed at 0.8 in plot (a) and at 1 in plot (b), while the other root varied from 0.4 up. Thus, the two plots encompass full-stationarity, one unit root and two unit roots. Each curve compares the square root of the sum of squared bias of an estimator over the four elements of $\hat{H}$ to that of $\hat{H}_{\rm REML}$. As expected, $\hat{H}_{\rm OLS,0}$ has bias that is almost identical to that of $\hat{H}_{\rm REML}$, while $\hat{H}_{\rm OLS}$ has uniformly significantly larger bias. The bias of $\hat{H}_{\rm WLS}$ is uniformly significantly smaller than that of $\hat{H}_{\rm OLS}$ and approaches that of $\hat{H}_{\rm REML}$ and $\hat{H}_{\rm OLS,0}$, while the iterated weighted least squares $\hat{H}_{\rm WLS_2}$ has uniformly smaller bias than $\hat{H}_{\rm WLS}$. The next theorem shows that the limiting distribution of $\hat{H}_{\rm WLS}$ at the unit root is unaffected by the intercept, unlike that of the ordinary least squares estimator.

THEOREM 3. *Let $H_1 = I$ and $\hat{H}$ be any initial estimator such that $n(\hat{H}_1 - I_k) = O_p(1)$ and $n^{1/2}(\hat{H}_i - H_i) = O_p(1)$ for $i = 2, ..., p$. Let $M = \mathrm{diag}\{nI_k, n^{1/2}I_{(p-1)k}\}$, then $(\hat{H}_{\rm WLS} - H)M \longrightarrow \Psi$ in distribution, where Ψ is the asymptotic distribution of the ordinary least squares estimator assuming $\mu = 0$ described in Theorem 10.3.3 of Fuller (1996).*

Both $\hat{H}_{\rm OLS}$ and thus $\hat{H}_{\rm WLS}$ satisfy the assumptions of Theorem 3.

3. SIMULATION STUDY

We compare through simulations the finite sample performance of the iterated weighted least squares estimator, $\hat{H}_{\rm WLS_2}$, computed using $\hat{H}_{\rm WLS}$ in $\hat{W}$, to that of the ordinary least squares estimator. Only results for the bivariate AR(1) are reported here, while results for higher order models can be found at http://www.stat.tamu.edu/∼wchen/WLSRLSupp.pdf. However, our con-

Table 1. *Simulation bias and root mean square error for sample size* $n = 100$

Parameters			$n\times$Bias				$\sqrt{n}\times$RMSE			
z_{11}	z_{22}	Est.	$\hat{a}_{11}$	$\hat{a}_{21}$	$\hat{a}_{12}$	$\hat{a}_{22}$	$\hat{a}_{11}$	$\hat{a}_{21}$	$\hat{a}_{12}$	$\hat{a}_{22}$
1	1	WLS$_2$	-0.83	-1.11	-1.65	-2.23	0.18	0.20	0.37	0.41
		REML	-0.76	-1.03	-1.51	-2.06	0.17	0.19	0.35	0.39
		OLS	-2.08	-2.63	-3.98	-5.07	0.28	0.33	0.55	0.65
1	0.95	WLS$_2$	-0.84	-1.12	-1.50	-2.03	0.21	0.23	0.41	0.45
		REML	-0.76	-1.03	-1.33	-1.85	0.19	0.21	0.38	0.41
		OLS	-1.90	-2.38	-3.04	-3.92	0.28	0.33	0.52	0.58
0.8	0.8	WLS$_2$	-0.81	-0.96	-0.52	-1.20	0.29	0.30	0.75	0.77
		REML	-0.70	-0.86	-0.37	-1.15	0.27	0.28	0.68	0.70
		OLS	-1.43	-1.71	-0.11	-1.30	0.32	0.33	0.68	0.71

RMSE, root mean square error; WLS$_2$, the iterated weighted least squares estimator; REML, the restricted maximum likelihood estimator; OLS,the ordinary least squares estimator.

clusions for the AR(1) model reported here are representative of the other models we studied. Here, we also report results for the restricted maximum likelihood estimator, since it is feasible to compute it for the bivariate AR(1) model.

Tables 1 and 2 report results from 10,000 replications of a bivariate AR(1) model with sample sizes $n = 100$ and $n = 200$ respectively. Let z_{11} and z_{22} be characteristic roots of H, and set

$$Z = \begin{pmatrix} z_{11} & 0 \\ 0.4 & z_{22} \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0.5 \\ -1 & 1 \end{pmatrix}, \quad \Sigma_e = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}. \tag{6}$$

The matrix H is defined to be BZB^{-1}. The iterated weighted least squares estimator almost always has smaller bias and root mean square error than the ordinary least squares estimator and the bias reduction can be substantial, even far from the unit root. Near the unit root, this bias reduction is attained without an excessive increase in variance as seen by the root mean square

Table 2. *Simulation bias and root mean square error for sample size* $n = 200$

Parameters			$n\times$Bias				$\sqrt{n}\times$RMSE			
z_{11}	z_{22}	Est.	$\hat{a}_{11}$	$\hat{a}_{21}$	$\hat{a}_{12}$	$\hat{a}_{22}$	$\hat{a}_{11}$	$\hat{a}_{21}$	$\hat{a}_{12}$	$\hat{a}_{22}$
1	1	WLS$_2$	-0.86	-1.09	-1.71	-2.19	0.13	0.14	0.26	0.29
		REML	-0.83	-1.06	-1.66	-2.12	0.13	0.14	0.26	0.28
		OLS	-2.11	-2.64	-4.13	-5.19	0.20	0.24	0.40	0.48
1	0.95	WLS$_2$	-0.86	-1.09	-1.55	-1.98	0.17	0.18	0.34	0.36
		REML	-0.81	-1.04	-1.46	-1.88	0.16	0.17	0.32	0.34
		OLS	-1.80	-2.23	-2.96	-3.73	0.21	0.24	0.40	0.44
0.8	0.8	WLS$_2$	-0.82	-0.99	-0.40	-1.15	0.27	0.27	0.74	0.74
		REML	-0.73	-0.91	-0.32	-1.15	0.26	0.26	0.68	0.69
		OLS	-1.34	-1.65	-0.03	-1.26	0.28	0.29	0.68	0.69

RMSE, root mean square error; WLS$_2$, the iterated weighted least squares estimator; REML, the restricted maximum likelihood estimator; OLS,the ordinary least squares estimator.

error. There are some instances of an increase in variance for the iterated weighted least squares estimator when far from the unit root. The restricted maximum likelihood estimator reduces bias at all parameter values without inflating the variance. As predicted by Theorems 1 and 2, the bias of the weighted least squares estimator approaches that of the restricted likelihood estimator as the sample size increases. This convergence is more apparent in the unit root case, since then the estimators are known to converge to their limit distribution at a faster rate than in the stationary case. This phenomenon is also observed in Figure 1.

Acknowledgement

Chen's research was supported by a grant from the U.S. National Science foundation.

APPENDIX

Proofs

Let $\lambda_{\max}(B)$ and $\lambda_{\min}(B)$ be the largest and smallest eigenvalue of a matrix B respectively and let $\|B\| = \lambda_{\max}^{1/2}(B^{\mathrm{T}}B)$ be its matrix norm. As $\hat{H}_{\mathrm{WLS}}$ is mean invariant, we assume that $\mu = 0$. Proofs of Lemma 1, Theorems 2 and 3 are at www.stat.tamu.edu/∼wchen/WLSRLSupp.pdf.

LEMMA 2. *Let Z_t be a stationary Gaussian vector time series and $S = \sum_{t=1}^{n}\left(Z_t - \bar{Z}\right)\left(Z_t - \bar{Z}\right)^{\mathrm{T}}$. Then, for any positive integer h, $E\left\|S^{-1}\right\|^h = O\left(n^{-h}\right)$ and $E\left\|nS^{-1} - E^{-1}(n^{-1}S)\right\|^h = O(n^{-h/2})$.*

Proof of Lemma 2. For a positive integer m to be defined later, $n^{-1}S = n^{-1}\sum_{s=1}^{n/m} B_s$, where $B_s = \sum_{t=1+sm}^{(s+1)m}\left(Z_t - \bar{Z}\right)\left(Z_t - \bar{Z}\right)^{\mathrm{T}}$. We have $\lambda_{\min}\left(n^{-1}S\right) \geq n^{-1}\sum_{s=1}^{n/m}\lambda_{\min}\left(B_s\right)$. Hence, by inequality of harmonic and arithmetic means of positive numbers and Jensen's inequality

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

For any $1 \leq s \leq n/m$, the vectors $\left\{Z_t - \bar{Z}\right\}_{t=1+sm}^{(s+1)m}$ have a Gaussian distribution with a nonsingular variance matrix uniformly in s. The proof of Lemma 8 of Chen & Hurvich (2006) yields $\sup_n \sup_{1\leq s\leq n/m} E\left\{\lambda_{\min}^{-h}\left(B_s\right)\right\} < C_0$, where C_0 is the hth moment of $\lambda_{\max}$ of an inverse Wishart matrix with m degrees of freedom. Choosing $m > 2h + k - 1$ ensures that this hth moment is finite (von Rosen, 1988), giving $C_0 < \infty$ and proving the first part. The second assertion follows by the fact that $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ and Hölder's inequality. □

LEMMA 3. *Let $h > 0$, then*

(i) for $\hat{H} = \hat{H}_{\mathrm{OLS},0}$ or $\hat{H}_{\mathrm{OLS}}$, $E\left\|\hat{\Sigma}_e - \Sigma_e\right\|^h = O\left(n^{-h/2}\right)$, $E\left\|\hat{H} - H_0\right\|^h = O\left(n^{-h/2}\right)$;
(ii) for the hth moments of $\hat{\Sigma}_e^{-1}$, we have $E\left\|\hat{\Sigma}_e^{-1}\right\|^h = O(1)$, $E\left\|\hat{\Sigma}_e^{-1} - \Sigma_e^{-1}\right\|^h = O\left(n^{-h/2}\right)$; and
(iii) for the hth moments of $\hat{\Sigma}_e^{-1/2}\hat{W}\hat{\Sigma}_e^{-1/2}$, $E\left\|\hat{\Sigma}_e^{-1/2}\hat{W}\hat{\Sigma}_e^{-1/2} - \Sigma_e^{-1/2}W\Sigma_e^{-1/2}\right\|^h = O\left(n^{-3h/2}\right)$.

Proof of Lemma 3. Hölder's inequality and Lemma 2 yield (*i*). For (*ii*), let $\Psi = (n-1)^{-1}\sum_{t=2}^{n}\eta_t\eta_t^{\mathrm{T}}$, where $\eta_t = \{\left(Y_{t-1} - \bar{Y}_{(0)}\right)^{\mathrm{T}}, \left(e_t - \bar{e}_{(1)}\right)^{\mathrm{T}}\}^{\mathrm{T}}$, then $\hat{\Sigma}_e$ is the Schur complement of Ψ relative to $(n-1)S_Y^{-1}$. Hence, $E|\hat{\Sigma}_e|^{-h} = E\{|\Psi|^{-h}\left|n^{-1}S_Y\right|^h\} \leq E^{1/2}\left|\Psi\right|^{-2h}E^{1/2}\left|n^{-1}S_Y\right|^{2h} = O\left(1\right)$ by Lemma 2.

Lemma 2 and the fact that $\|\hat{\Sigma}_e^{-1}\| \le |\hat{\Sigma}_e^{-1}|\ \|\hat{\Sigma}\|^{k-1}$ yields (*ii*). By (i), (ii) and Lemma 2, $E\left\|\hat{W}\right\|^h = O(n^{-h})$ and $E\left\|\hat{W}-W\right\|^h = O\left(n^{-3h/2}\right)$, thus (*iii*) follows by Hölder's inequality. □

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

$$-\partial Q(\hat{H}_{\text{OLS},0}) = 2\hat{\Sigma}_e^{-1}(\hat{H}_{WLS} - \hat{H}_{\text{OLS},0})\sum_{t=2}^{n}\left(Y_{t-1}-\bar{Y}_{(0)}\right)\left(Y_{t-1}-\bar{Y}_{(0)}\right)^{\mathrm{T}}$$
$$+2(n-1)\hat{\Sigma}_e^{-1/2}\hat{W}\hat{\Sigma}_e^{-1/2}(\hat{H}_{\text{WLS}}-\hat{H}_{\text{OLS},0})\left(\bar{Y}_{(0)}-Y_1\right)\left(\bar{Y}_{(0)}-Y_1\right)^{\mathrm{T}}.$$

Thus $\operatorname{vec}(\hat{H}_{\text{WLS}}-\hat{H}_{\text{OLS},0}) = (G+J)^{-1}\operatorname{vec}\{\partial Q(\hat{H}_{\text{OLS},0})\}$, where $J=(n-1)\Gamma\otimes\hat{\Sigma}_e^{-1/2}\hat{W}\hat{\Sigma}_e^{-1/2}$, $\Gamma = \left(\bar{Y}_{(0)}-Y_1\right)\left(\bar{Y}_{(0)}-Y_1\right)'$ and $G = S_Y\otimes\hat{\Sigma}_e^{-1}$, $S_Y = \sum_{t=2}^n\left(Y_{t-1}-\bar{Y}_{(0)}\right)\left(Y_{t-1}-\bar{Y}_{(0)}\right)'$.

Let Q_{I} be the objective function of $\hat{H}_{\text{IWLS}}$, it is sufficient to show (*i*) $E\,(G+J)^{-1} = (n-1)^{-1}\{\Sigma_Y^{-1}\otimes\Sigma_e\}+o(n^{-1})$, (*ii*) $E\left\|(G+J)^{-1}\right\|^2 = O(n^{-2})$, (*iii*) $E\{\partial Q_I(\hat{H}_{\text{OLS},0})\} = o(1)$, (*iv*) $E\left\|\partial Q_I(\hat{H}_{\text{OLS},0})\right\|^2 = O(1)$, and (*v*) $E\left\|\partial Q_I(\hat{H}_{\text{OLS},0}) - \partial Q(\hat{H}_{\text{OLS},0})\right\|^2 = o(n^{-2})$.

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

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

By Lemmas 2, 3, both terms on the right hand side are $O(n^{-3})$. Since $\{1+\operatorname{trace}\left(JG^{-1}\right)\}^{-1}\le 1$ and $E\left\|G^{-1}JG^{-1}\right\|^2 = O(n^{-4})$ by Lemma 3, (*i*) and (*ii*) follow. Noting that $\sum_{t=2}^n Y_tY_{t-1}^{\mathrm{T}} = \hat{H}_{\text{OLS},0}\sum_{t=2}^n Y_{t-1}Y_{t-1}$, we get $\partial Q_I(\hat{H}_{\text{OLS},0}) = 2(n-1)\{\Omega(W_0,\Sigma_e)+R(W_0,\Sigma_e)\}$, where

$$\Omega(W_0,\Sigma_e) = \Sigma_e^{-1/2}\left(I-W_0\right)\Sigma_e^{-1/2}\left(I-H_0\right)\bar{Y}_{(0)}\bar{Y}_{(0)}^{\mathrm{T}} - \Sigma_e^{-1/2}W_0\Sigma_e^{-1/2}\left(I-H_0\right)Y_1Y_1^{\mathrm{T}}, \qquad \text{(A1)}$$

$$R(W_0,\Sigma_e) = \Sigma_e^{-1/2}\left(I-W_0\right)\Sigma_e^{-1/2}\left\{\left(H_0-\hat{H}_{\text{OLS},0}\right)\bar{Y}_{(0)}+(n-1)^{-1}\left(Y_n-Y_1\right)\right\}\bar{Y}_{(0)}^{\mathrm{T}}$$
$$-\Sigma_e^{-1/2}W_0\Sigma_e^{-1/2}\left\{\left(H_0-\hat{H}_{\text{OLS},0}\right)Y_1-\left(I-\hat{H}_{\text{OLS},0}\right)\bar{Y}_{(0)}-(n-1)^{-1}\left(Y_n-Y_1\right)\right\}Y_1^{\mathrm{T}}.$$

Since $E(Y_1Y_1^{\mathrm{T}}) = \Sigma_e$, $E((n-1)\bar{Y}_{(0)}\bar{Y}_{(0)}^{\mathrm{T}}) = \left(I-H_0\right)^{-1}\Sigma_e\left(I-H_0^{\mathrm{T}}\right)^{-1}+o(1)$, $W_0 = O(n^{-1})$ and $I-W_0 = (n-1)W_0\Sigma_e^{-1/2}(I-H_0)\Sigma_e(I-H_0)^{\mathrm{T}}\Sigma_e^{-1/2}$, we have $E(\Omega) = o(n^{-1})$ from (A1). Noting that $I - W_0 = O(1)$, it is easy to verify that $E\,\|R(W_0, \Sigma_e)\|^2 = O(n^{-3})$ by Lemma 3. Finally, (*v*) also follows from Lemma 3. □

REFERENCES

CHEANG, W. K. & REINSEL, G. (2000). Bias Reduction of Autoregressive Estimates in Time Series Regression Model through Restricted Maximum Likelihood, *J. Am. Statist. Assoc.* **95**, 1173-1184.

CHEN, W. W. & HURVICH, C. M. (2006). Semiparametric estimation of fractional cointegrating subspaces, *Ann. Statist.* **34**, 2939-2979.

CHEN, W. W. & DEO, R. S. (2009). Bias Reduction and Likelihood Based Almost-Exactly Sized Hypothesis Testing in Predictive Regressions using the Restricted Likelihood, forthcoming *Econometric Theory*

FULLER, W. (1996). *Introduction to Statistical Time Series,* Second Edition, John Wiley & Sons, Inc.

KANG, W., SHIN, D. & LEE, Y. (2003). Biases of the Restricted Maximum Likelihood Estimators for ARMA Processes with Polynomial Time Trend, *J. Statist. Plan. Infer.* **116**, 163-176.

NICHOLLS, D. F. & POPE, A. I.(1988). Bias in the Dstimation of Multivariate Autoregressions. *J. Mult. Anal.* **30A**, 296-309.

SHAMAN, P. & STINE, R. (1988). The Bias of Autoregressive Coefficient Estimators, *J. Am. Statist. Assoc.* **83**, 842-848.

TUNNICLIFFE-WILSON, G. (1989). On the Use of Marginal Likelihood in Time Series Model Estimation. *J. R. Statist. Soc. B* **51**,15-27

VON ROSEN, D. (1988). Moments for the inverted Wishart distribution. *Scand. J. Statist.* **15**,97-109

YAMAMOTO, T.& KUNITOMO, N(1984). Asymptotic bias of the least squares estimator for multivariate autoregressive models. *Ann, Inst. Statist. Math.* **36**, 419-430.