Bias of ML Estimator for Multivariate Regression Model with Vector AR(1) Noise

Cheang Wai Kwong
waikwong.cheang@nie.edu.sg
National Institute of Education
Nanyang Technological University
Singapore

Abstract

This paper considers the use of technology to assess the adequacy of a theoretical bias result in the maximum likelihood (ML) estimation of multivariate regression model with vector autoregressive AR(1) noise. We develop a relatively explicit and conveniently computable approximation for the bias of the ML estimator of the AR parameters. This bias estimate can be used to obtain a bias-corrected ML estimate. To assess the adequacy of our bias approximation, R/S-PLUS programs are written to calculate the theoretical biases and simulate the empirical biases for polynomial regression. Simulation results suggest that the theoretical ML bias approximations are in reasonable agreement with the empirical biases when the mean is unknown. In the presence of a linear or quadratic trend, a longer series length is needed for the bias approximations to be adequate.

1 Introduction

Consider a multivariate linear regression model of the form

$$\boldsymbol{Y}_t' \;=\; \boldsymbol{x}_t'\mathcal{B} + \boldsymbol{N}_t', \qquad t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{Y}_t = (Y_{t1}, \ldots, Y_{tk})'$ is a k-dimensional time series vector of random variables, $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tr})'$ is a r-dimensional vector of deterministic or stochastic regressors, and $\mathcal{B} = (\boldsymbol{\beta}_{1\cdot}, \ldots, \boldsymbol{\beta}_{r\cdot})'$ is a $r \times k$ matrix of regression coefficients. The noise series $\{\boldsymbol{N}_t\}$ is assumed to be a stationary process following a k-dimensional vector AR(1) model,

$$\boldsymbol{N}_t \;=\; \Phi \boldsymbol{N}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{2}$$

where Φ is a $k \times k$ matrix with all eigenvalues less than one in absolute value, and $\{\boldsymbol{\varepsilon}_t\}$ is a vector white noise process with zero mean vector and covariance matrix Σ.

Let $\mathbf{Y} = \left[\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T\right]'$ and $\mathbf{N} = \left[\boldsymbol{N}_1, \ldots, \boldsymbol{N}_T\right]'$ be the $T \times k$ data and noise matrices, respectively. Also, let $\mathbf{y} = \mathrm{vec}(\mathbf{Y}') = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_T')'$, $\mathbf{n} = \mathrm{vec}(\mathbf{N}') = (\boldsymbol{N}_1', \ldots, \boldsymbol{N}_T')'$, $\boldsymbol{\beta} = \mathrm{vec}(\mathcal{B}')$, $\boldsymbol{\phi} = \mathrm{vec}(\Phi)$, and $\boldsymbol{\sigma} = \mathrm{vec}(\Sigma)$. Define the $T \times r$ matrix $\mathbf{X} = \left[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\right]'$, and assume that $\mathbf{X}$ is of full rank $r = \text{rank}(\mathbf{X})$. Then the regression model (1) may be expressed in matrix form as $\mathbf{Y} = \mathbf{X}\mathcal{B} + \mathbf{N}$, or in "vec" form as

$$\mathbf{y} \;=\; (\mathbf{X} \otimes \mathbf{I}_k)\boldsymbol{\beta} + \mathbf{n}, \tag{3}$$

where $\mathbf{I}_k$ is the identity matrix of order k.

Let $\Gamma(0) = \text{Cov}(\boldsymbol{N}_t) \equiv E(\boldsymbol{N}_t\boldsymbol{N}_t^{'})$ denote the covariance matrix of $\boldsymbol{N}_t$. From [8, p. 135–138], the $kT \times kT$ covariance matrix of $\mathbf{n}$ can be expressed as

$$\boldsymbol{\Gamma}_T \;=\; \text{Cov}(\mathbf{n}) \;=\; \boldsymbol{\Theta}^{-1}\text{Diag}\big\{\Gamma(0), (\mathbf{I}_{T-1} \otimes \Sigma)\big\}\boldsymbol{\Theta}^{'-1}, \tag{4}$$

where $\boldsymbol{\Theta} = \mathbf{I}_T \otimes \mathbf{I}_k - \mathbf{L} \otimes \Phi$, and $\mathbf{L}$ denotes the $T \times T$ lag matrix that has ones on the first sub-diagonal and zeros elsewhere. Note that $|\boldsymbol{\Theta}| = 1$, so that $|\boldsymbol{\Gamma}_T| = |\Gamma(0)||\mathbf{I}_{T-1} \otimes \Sigma| = |\Gamma(0)||\Sigma|^{T-1}$. Under the regression model (3) and normality assumption of the vector white noise process $\{\boldsymbol{\varepsilon}_t\}$, the exact log-likelihood function is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}) \;=\; -\tfrac{kT}{2}\log(2\pi) - \tfrac{T-1}{2}\log|\Sigma| - \tfrac{1}{2}\log|\Gamma(0)| - \tfrac{1}{2}S(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}), \tag{5}$$

where $S(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma})$ is the exact sum of squares function given by

$$\begin{aligned} S(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}) &= \big[\mathbf{y} - (\mathbf{X} \otimes \mathbf{I}_k)\boldsymbol{\beta}\big]^{'}\boldsymbol{\Gamma}_T^{-1}\big[\mathbf{y} - (\mathbf{X} \otimes \mathbf{I}_k)\boldsymbol{\beta}\big] \\ &= \mathbf{n}^{'}\boldsymbol{\Theta}^{'}\text{Diag}\big\{\Gamma(0)^{-1}, (\mathbf{I}_{T-1} \otimes \Sigma^{-1})\big\}\boldsymbol{\Theta}\mathbf{n} \\ &= \boldsymbol{N}_1^{'}\Gamma(0)^{-1}\boldsymbol{N}_1 + \textstyle\sum_{t=2}^{T}\boldsymbol{\varepsilon}_t^{'}\Sigma^{-1}\boldsymbol{\varepsilon}_t. \end{aligned}$$

We are interested in obtaining an approximation for the bias of the maximum likelihood (ML) estimator of Φ based on a sample of T vector observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$. This bias estimate can be used to obtain a bias-corrected ML estimate. To assess the adequacy of our bias approximation, we will perform a simulation study using the software R/S-PLUS.

2 ML bias in vector AR(1) noise with no regression component

[5] considered the bias of the ML estimator $\hat{\boldsymbol{\alpha}}_{\mathtt{M}}$ of an m-dimensional parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$ in a general setting. Under the assumptions that the expectations $E\big[\frac{\partial^2\ell}{\partial\alpha_i\partial\alpha_j}\big]$, $E\big[\frac{\partial^3\ell}{\partial\alpha_i\partial\alpha_j\partial\alpha_k}\big]$ and $E\big[\frac{\partial^2\ell}{\partial\alpha_i\partial\alpha_j}\frac{\partial\ell}{\partial\alpha_k}\big]$ are of order $O(T)$, where $\ell(\boldsymbol{\alpha})$ denotes the log-likelihood function and $I(\boldsymbol{\alpha}) = -E\big[\frac{\partial^2\ell}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'}\big]$ is the information matrix for $\boldsymbol{\alpha}$, [5] gave the following general representation for the bias of $\hat{\boldsymbol{\alpha}}_{\mathtt{M}}$,

$$E(\hat{\boldsymbol{\alpha}}_{\mathtt{M}} - \boldsymbol{\alpha}) \;=\; I(\boldsymbol{\alpha})^{-1}\;\mathbf{C}\;\text{vec}\big[I(\boldsymbol{\alpha})^{-1}\big] + O\left(\frac{1}{T^2}\right), \tag{6}$$

where $\mathbf{C} = -\frac{\partial}{\partial\boldsymbol{\alpha}'} \otimes I(\boldsymbol{\alpha}) - \frac{1}{2}E\big[\frac{\partial}{\partial\boldsymbol{\alpha}'} \otimes \frac{\partial^2\ell}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'}\big]$. Here, we introduce the notation $\frac{\partial}{\partial\boldsymbol{\alpha}'} \otimes \mathbf{A} = \big[\frac{\partial}{\partial\alpha_1}\mathbf{A}, \ldots, \frac{\partial}{\partial\alpha_m}\mathbf{A}\big]$. We will use the result (6) to obtain an explicit approximate expression for the bias of the ML estimator of $\boldsymbol{\phi} = \text{vec}(\Phi)$ in the vector AR(1) model.

[10] derived the asymptotic bias of the conditional least squares (LS) estimator for the vector autoregressive models, which should be asymptotically equivalent to the ML estimator.

In particular, for the vector AR(1) model with known zero mean (i.e., no regression component) they obtained

$$E(\hat{\Phi}_{\mathtt{CL}} - \Phi) \;=\; -\frac{1}{T}\,\Sigma \sum_{j=0}^{\infty} \left[\Phi'^{j}\mathrm{tr}\big(\Phi^{j+1}\big) + \Phi'^{2j+1}\right]\Gamma(0)^{-1} + O\left(\frac{1}{T^2}\right), \tag{7}$$

where the conditional LS estimator is $\hat{\Phi}_{\mathtt{CL}} = \hat{\Gamma}(1)'\hat{\Gamma}(0)^{-1}$, with $\hat{\Gamma}(0) = \frac{1}{T}\sum_{t=2}^{T} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}$ and $\hat{\Gamma}(1) = \frac{1}{T}\sum_{t=2}^{T} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t}$. An asymptotic bias expression for $\hat{\Phi}_{\mathtt{CL}}$ was also given by [6] in the form

$$E(\hat{\Phi}_{\mathtt{CL}} - \Phi) \;\approx\; -\frac{1}{T}E\Big[\big[\hat{\Gamma}(1)' - \Phi\hat{\Gamma}(0)\big]\Gamma(0)^{-1}\big[\hat{\Gamma}(0) - \Gamma(0)\big]\Big]\Gamma(0)^{-1}.$$

For an $m\times n$ matrix $\mathbf{A}$, let $\mathbf{I}_{m,n}$ be the $mn\times mn$ vec-permutation matrix that rearranges $\mathrm{vec}(\mathbf{A}')$ to $\mathrm{vec}(\mathbf{A})$, that is, such that $\mathrm{vec}(\mathbf{A}) = \mathbf{I}_{m,n}\mathrm{vec}(\mathbf{A}')$. For a vector AR(1) noise with no regression component, [2, p. 143–146] used (6) to show that the bias of the ML estimator $\hat{\boldsymbol{\phi}}_{\mathtt{M}} = \mathrm{vec}(\hat{\Phi}_{\mathtt{M}})$ of $\boldsymbol{\phi} = \mathrm{vec}(\Phi)$ is given approximately by

$$\begin{aligned} E(\hat{\boldsymbol{\phi}}_{\mathtt{M}} - \boldsymbol{\phi}) \;\approx\; & -\frac{1}{T}\,\left[\Gamma(0)^{-1}\otimes\Sigma\right]\left[\mathrm{vec}(\mathbf{I}_k\otimes\Sigma)'\otimes\mathbf{I}_{k^2}\right] \\ & \cdot\left[(\Phi'\otimes\mathbf{I}_k)(\mathbf{I}_{k^2} + \mathbf{I}_{k,k})\Delta'^{-1}\otimes\mathbf{K}\right]\,\mathrm{vec}\big(\mathbf{I}_{k^2}\otimes\boldsymbol{\sigma}^*\big), \end{aligned} \tag{8}$$

where $\boldsymbol{\sigma}^* = \mathrm{vec}(\Sigma^{-1})$, $\mathbf{K} = \mathbf{I}_k\otimes\mathbf{I}_{k,k}\otimes\mathbf{I}_k$, and $\Delta = \mathbf{I}_{k^2} - \Phi\otimes\Phi$. This expression provides a relatively explicit and convenient form for the bias of the ML estimator $\hat{\boldsymbol{\phi}}_{\mathtt{M}}$ which can be evaluated easily for any given values of the parameters Φ and Σ. The bias expression (7) can be evaluated by taking a large partial sum, and calculations using various Φ and Σ indicate that the two expressions (7) and (8) agree numerically. A program for performing calculations using (8), written in R/S-PLUS, is given in Appendix A1. This is consistent with the result in [3], where the bias of the ML estimator in the univariate AR(p) model with known mean, to order $1/T$ terms, coincides with the bias of the conditional LS estimator given by [9].

3 ML bias in vector AR(1) noise due to regression

3.1 Decomposition of ML bias

For the multivariate regression model (1) with vector AR(1) noise, the expression (6) can be used to give an approximation for the bias of the ML estimator of $\boldsymbol{\alpha} = (\boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\sigma}')'$. Using the same approach as in [4] for univariate regression model with fractional ARIMA noise, it can be shown that the bias approximation of the ML estimator $\hat{\boldsymbol{\phi}}_{\mathtt{M}}$ can be decomposed into two components,

$$\begin{aligned} E(\hat{\boldsymbol{\phi}}_{\mathtt{M}} - \boldsymbol{\phi}) \;\approx\; & -\frac{1}{T}\bar{I}(\boldsymbol{\phi})^{-1}\left[\mathrm{vec}(\mathbf{I}_k\otimes\Sigma)'\otimes\mathbf{I}_{k^2}\right]\left[(\Phi'\otimes\mathbf{I}_k)(\mathbf{I}_{k^2} + \mathbf{I}_{k,k})\Delta'^{-1}\otimes\mathbf{K}\right]\mathrm{vec}\big(\mathbf{I}_{k^2}\otimes\boldsymbol{\sigma}^*\big) \\ & + \frac{1}{T}\bar{I}(\boldsymbol{\phi})^{-1}\boldsymbol{\tau}_{\boldsymbol{\phi}}, \end{aligned} \tag{9}$$

where $\bar{I}(\boldsymbol{\phi}) = I(\boldsymbol{\phi})/T$ and $\boldsymbol{\tau}_{\boldsymbol{\phi}} = \frac{1}{2}\frac{\partial}{\partial\boldsymbol{\phi}}\log|I(\boldsymbol{\beta})|$, with the information matrices $I(\boldsymbol{\phi}) \approx T\Gamma(0)\otimes\Sigma^{-1}$ and $I(\boldsymbol{\beta}) = (\mathbf{X}'\otimes\mathbf{I}_k)\boldsymbol{\Gamma}_T^{-1}(\mathbf{X}\otimes\mathbf{I}_k)$. The first component in (9), given by the bias expression (8) for vector AR(1) with zero mean, is intrinsic to the vector AR(1) noise model. The second bias component, $\bar{I}(\boldsymbol{\phi})^{-1}\boldsymbol{\tau}_{\boldsymbol{\phi}}/T$, can be attributed to the estimation of regression parameters.

3.2 ML bias due to polynomial regression

We will obtain an approximate form of the bias term due to the special case of polynomial regression of degree $r-1$, with $\boldsymbol{x}_t' = (1, t, \ldots, t^{r-1})$. From (4), we have

$$\boldsymbol{\Gamma}_T^{-1} \;=\; \boldsymbol{\Theta}'\mathrm{Diag}\big\{\Gamma(0)^{-1}, (\mathbf{I}_{T-1}\otimes\Sigma^{-1})\big\}\boldsymbol{\Theta} \;\approx\; \boldsymbol{\Theta}'(\mathbf{I}_T\otimes\Sigma^{-1})\boldsymbol{\Theta},$$

where $\boldsymbol{\Theta} = \mathbf{I}_T\otimes\mathbf{I}_k - \mathbf{L}\otimes\Phi$. Then,

$$\begin{aligned}
I(\boldsymbol{\beta}) \;&\approx\; \big[\mathbf{X}'\otimes\mathbf{I}_k - (\mathbf{X}'\mathbf{L}')\otimes\Phi'\big]\big(\mathbf{I}_T\otimes\Sigma^{-1}\big)\big[\mathbf{X}\otimes\mathbf{I}_k - (\mathbf{L}\mathbf{X})\otimes\Phi\big]\\
&=\; \big[\mathbf{X}'\otimes(\mathbf{I}_k-\Phi') + (\mathbf{X}'(\mathbf{I}_T-\mathbf{L}'))\otimes\Phi'\big]\big(\mathbf{I}_T\otimes\Sigma^{-1}\big)\\
&\quad\;\cdot\big[\mathbf{X}\otimes(\mathbf{I}_k-\Phi) + ((\mathbf{I}_T-\mathbf{L})\mathbf{X})\otimes\Phi\big]\\
&=\; (\mathbf{X}'\mathbf{X})\otimes\big[(\mathbf{I}_k-\Phi')\Sigma^{-1}(\mathbf{I}_k-\Phi)\big] + \big[\mathbf{X}'(\mathbf{I}_T-\mathbf{L}')(\mathbf{I}_T-\mathbf{L})\mathbf{X}\big]\otimes(\Phi'\Sigma^{-1}\Phi)\\
&\quad\; + \big[\mathbf{X}'(\mathbf{I}_T-\mathbf{L}')\mathbf{X}\big]\otimes\big[\Phi'\Sigma^{-1}(\mathbf{I}_k-\Phi)\big] + \big[\mathbf{X}'(\mathbf{I}_T-\mathbf{L})\mathbf{X}\big]\otimes\big[(\mathbf{I}_k-\Phi')\Sigma^{-1}\Phi\big]\\
&\approx\; (\mathbf{X}'\mathbf{X})\otimes\big[(\mathbf{I}_k-\Phi')\Sigma^{-1}(\mathbf{I}_k-\Phi)\big],
\end{aligned}$$

since $\lim_{T\to\infty}\mathbf{D}_T^{-1}\mathbf{X}'\mathbf{X}\mathbf{D}_T^{-1} = \lim_{T\to\infty}\mathbf{D}_T^{-1}\mathbf{X}'\mathbf{L}\mathbf{X}\mathbf{D}_T^{-1}$ for polynomial regressors with normalizing matrix $\mathbf{D}_T = \mathrm{Diag}\big\{(\sum_{t=1}^T x_{t1}^2)^{\frac12}, \ldots, (\sum_{t=1}^T x_{tr}^2)^{\frac12}\big\}$ [e.g., [1, p. 582]]. Therefore,

$$\begin{aligned}
\log|I(\boldsymbol{\beta})| \;&\approx\; k\log|\mathbf{X}'\mathbf{X}| + r\log|(\mathbf{I}_k-\Phi')\Sigma^{-1}(\mathbf{I}_k-\Phi)|\\
&=\; k\log|\mathbf{X}'\mathbf{X}| - r\log|\Sigma| + 2r\log|\mathbf{I}_k-\Phi|. \qquad (10)
\end{aligned}$$

Note that $|\mathbf{I}_k-\Phi|\neq 0$ by the stationary condition of $\boldsymbol{N}_t$. Using the chain rule and the result $\frac{\partial}{\partial\mathrm{vec}(\mathbf{A})}\log|\mathbf{A}| = \mathrm{vec}(\mathbf{A}'^{-1})$ [e.g., [7, p. 473]], we obtain

$$\begin{aligned}
\boldsymbol{\tau}_{\boldsymbol{\phi}} \;\approx\; r\,\frac{\partial}{\partial\boldsymbol{\phi}}\log|\mathbf{I}_k-\Phi| \;&=\; r\,\frac{\partial}{\partial\boldsymbol{\phi}}\mathrm{vec}(\mathbf{I}_k-\Phi)'\,\frac{\partial}{\partial\mathrm{vec}(\mathbf{I}_k-\Phi)}\log|\mathbf{I}_k-\Phi|\\
&=\; -r\;\mathrm{vec}\big[(\mathbf{I}_k-\Phi')^{-1}\big].
\end{aligned}$$

Hence, the ML bias term of $\hat{\boldsymbol{\phi}}_{\mathtt{M}}$ due to polynomial regression of degree $r-1$ is

$$\frac{1}{T}\bar{I}(\boldsymbol{\phi})^{-1}\boldsymbol{\tau}_{\boldsymbol{\phi}} \;\approx\; -\frac{r}{T}\;\big[\Gamma(0)^{-1}\otimes\Sigma\big]\;\mathrm{vec}\big[(\mathbf{I}_k-\Phi')^{-1}\big]. \qquad (11)$$

For univariate AR(1) noise $N_t - \phi N_{t-1} = \varepsilon_t$, with $\Phi=\phi$, $\Sigma = \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ and $\Gamma(0) = \mathrm{Var}(N_t) = \sigma_\varepsilon^2/(1-\phi^2)$, this bias term is $-r(1+\phi)/T$, as given in [3].

4 A simulation study

In this section, we present theoretical calculations and empirical simulation results to assess adequacy of the ML bias approximations given by (8), (9) and (11) for polynomial regression.

Table 1 presents the theoretical biases for two pairs of (Φ, Σ) and for $r = 1, 2, 3$ when $T = 30$. Note that the eigenvalues of the first Φ are complex while those of the second Φ are real. For reference, Table 1 also presents the theoretical biases when the mean is known ($r = 0$). These values indicate that for $T = 30$, the biases of $\hat{\Phi}_{\mathtt{M}}$ are generally "negligible" when $r = 0$, but can be "appreciable" even when $r = 1$ as in the case of the second Φ. Interestingly, further theoretical calculations (not shown) suggest that a Φ matrix that has eigenvalues smaller in absolute value does not necessarily lead to ML biases that are smaller in magnitude.

For each combination of (Φ, Σ) and r, a simulation was performed using 1000 replications of bivariate AR(1) noise with $T = 30$. See Appendix A2 for the R/S-PLUS program. Without loss of generality, we take the regression coefficients as $\boldsymbol{\beta} = \mathbf{0}$ in generating the simulated data. The empirical biases of the ML estimates are approximated using the empirical biases of the conditional LS estimates (i.e., the average of the estimates over the 1000 replications minus the true value). Under the regression model (3), the conditional LS estimator $\hat{\Phi}_{\mathtt{CL}}$ is given by

$$\hat{\Phi}_{\mathtt{CL}} \;=\; \left[\sum_{t=2}^{T} \hat{\boldsymbol{N}}_{t-1}\hat{\boldsymbol{N}}_t^{'}\right]^{'} \left[\sum_{t=2}^{T} \hat{\boldsymbol{N}}_{t-1}\hat{\boldsymbol{N}}_{t-1}^{'}\right]^{-1},$$

where $\hat{\boldsymbol{N}}_t$ are "residuals" from the regression, i.e., $(\hat{\boldsymbol{N}}_1^{'}, \ldots, \hat{\boldsymbol{N}}_T^{'})' = \mathbf{y} - (\mathbf{X} \otimes \mathbf{I}_k)\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\beta}}$ is an estimate of $\boldsymbol{\beta}$ of the form

$$\hat{\boldsymbol{\beta}} \;=\; \left[(\mathbf{X} \otimes \mathbf{I}_k)'\hat{\boldsymbol{\Gamma}}_T^{-1}(\mathbf{X} \otimes \mathbf{I}_k)\right]^{-1}(\mathbf{X} \otimes \mathbf{I}_k)'\hat{\boldsymbol{\Gamma}}_T^{-1}\mathbf{y}.$$

For each replication, $\hat{\Phi}_{\mathtt{CL}}$ was calculated using a 10-step iteration, beginning with the ordinary least squares residuals obtained using $\hat{\boldsymbol{\beta}} = \left[(\mathbf{X} \otimes \mathbf{I}_k)'(\mathbf{X} \otimes \mathbf{I}_k)\right]^{-1}(\mathbf{X} \otimes \mathbf{I}_k)'\mathbf{y}$. The estimate of Σ required in $\hat{\boldsymbol{\Gamma}}_T^{-1}$ in each iteration was calculated as $\hat{\Sigma} = \frac{1}{T-1-k-r}\sum_{t=2}^{T} \hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t^{'}$, where $\hat{\boldsymbol{\varepsilon}}_t = \hat{\boldsymbol{N}}_t - \hat{\Phi}\hat{\boldsymbol{N}}_{t-1}$. This form of $\hat{\Sigma}$ is consistent with the estimator of Σ given in [8, p. 91] for $r = 1$.

Table 1: Approximate theoretical values of $E(\hat{\Phi}_{\mathtt{M}} - \Phi)$ for polynomial regression given by (8), (9) and (11), and empirical biases of $\hat{\Phi}_{\mathtt{CL}}$. The series length is $T = 30$.

Regression with vector AR(1) noise		Theoretical bias	Empirical bias
$\Phi = \begin{bmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{bmatrix}$	Mean known ($r=0$)	$\begin{bmatrix} -0.023 & -0.023 \\ -0.005 & -0.027 \end{bmatrix}$	$\begin{bmatrix} -0.023 & -0.015 \\ -0.010 & -0.024 \end{bmatrix}$
$\Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$	Mean unknown ($r=1$)	$\begin{bmatrix} -0.033 & -0.010 \\ -0.014 & -0.029 \end{bmatrix}$	$\begin{bmatrix} -0.038 & 0.001 \\ -0.022 & -0.026 \end{bmatrix}$
Eigenvalues of Φ: $\lambda \;=\; 0.7 \pm 0.52i$	Linear ($r=2$)	$\begin{bmatrix} -0.044 & 0.003 \\ -0.023 & -0.030 \end{bmatrix}$	$\begin{bmatrix} -0.052 & 0.016 \\ -0.034 & -0.031 \end{bmatrix}$
$\lvert\lambda\rvert \;=\; 0.872$	Quadratic ($r=3$)	$\begin{bmatrix} -0.054 & 0.016 \\ -0.032 & -0.032 \end{bmatrix}$	$\begin{bmatrix} -0.072 & 0.034 \\ -0.046 & -0.036 \end{bmatrix}$
$\Phi = \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix}$	Mean known ($r=0$)	$\begin{bmatrix} -0.058 & 0.047 \\ -0.012 & -0.017 \end{bmatrix}$	$\begin{bmatrix} -0.058 & 0.041 \\ -0.016 & -0.016 \end{bmatrix}$
$\Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$	Mean unknown ($r=1$)	$\begin{bmatrix} -0.094 & 0.092 \\ -0.022 & -0.034 \end{bmatrix}$	$\begin{bmatrix} -0.101 & 0.096 \\ -0.026 & -0.040 \end{bmatrix}$
Eigenvalues of Φ: $\lambda \;=\; 0.8,\; 0.5$	Linear ($r=2$)	$\begin{bmatrix} -0.130 & 0.138 \\ -0.033 & -0.050 \end{bmatrix}$	$\begin{bmatrix} -0.154 & 0.160 \\ -0.028 & -0.078 \end{bmatrix}$
	Quadratic ($r=3$)	$\begin{bmatrix} -0.167 & 0.183 \\ -0.044 & -0.066 \end{bmatrix}$	$\begin{bmatrix} -0.227 & 0.239 \\ -0.026 & -0.124 \end{bmatrix}$

Table 1 indicates that when $r = 0$ and 1, the theoretical ML biases given by (8), (9) and (11) are in reasonable agreement with the empirical biases. However, when $r = 2$ and 3, the empirical biases are generally under-estimated in magnitude by the theoretical bias approximations. Further simulation results (not shown) suggest that in the presence of a linear or quadratic trend, a longer series length (say $T = 50$) is needed for the bias approximations to be adequate.

5 Concluding remarks

We derived a relatively explicit and convenient approximate form for the bias of the ML estimator of vector AR(1) which can be evaluated easily for any given values of the parameters Φ and Σ. [3] established that for restricted maximum likelihood (REML) estimation of a regression model with univariate AR(p) noise, the bias term $\bar{I}(\boldsymbol{\phi})^{-1}\boldsymbol{\tau}_{\boldsymbol{\phi}}/T$ of the ML estimator due to regression is essentially eliminated for the REML estimator. *If* this result could be extended to vector AR(1), then (8) would provide a convenient approximation to the bias of the REML estimator of vector AR(1).

Appendix

A1. R/S-PLUS program to calculate ML bias

```
# Multivariate regression with vector AR(1) noise
#      Yt' = xt'*B + Nt', Nt = Phi*N(t-1) + at
# Calculate the theoretical ML bias of phi = vec(Phi) in polynomial regression

vec <- function(A)
{ m <- nrow(A)
  n <- ncol(A)
  B <- as.matrix(A[,1])
  for (j in 2:n) B <- rbind(B,as.matrix(A[,j]))
  B
}

# Vec-permutation matrix: For m x n matrix A, vec(A) = I(m,n) vec(A')
vecp <- function(n)
{ Sn <- diag(n^2)  # Sn = I(n,n)
  per <- matrix(1:(n^2),n,n)
  per <- matrix(t(per),n*n,1)
  Sn <- Sn[per,]
  Sn
}

k <- 2
n <- 30
```

```
r <- 2
phi <- c(0.2,-0.6,0.3,1.1)   # vec(Phi)
sigma <- c(4,1,1,2)          # vec(Sigma)

phi <- matrix(phi,k^2,1)
sigma <- matrix(sigma,k^2,1)
Phi <- matrix(phi,k,k)
Sigma <- matrix(sigma,k,k)
lambda <- eigen(Phi)$values
print(round(lambda,6))
print(round(abs(lambda),6))

Delta <- diag(k^2) - kronecker(Phi,Phi)
Dinv <- solve(Delta)
gamma0 <- Dinv %*% sigma
Gamma0 <- matrix(gamma0,k,k)
G0inv <- solve(Gamma0)
Sinv <- solve(Sigma)
vecSinv <- vec(Sinv)
Iinv <- kronecker(G0inv,Sigma)

Kmat <- kronecker(diag(k),vecp(k))
Kmat <- kronecker(Kmat,diag(k))
B1 <- t(vec(kronecker(diag(k),Sigma)))
B2 <- kronecker(t(Phi),diag(k)) %*% (diag(k^2) + vecp(k)) %*% t(Dinv)
B3 <- kronecker(diag(k^2),vecSinv)
bias0 <- Iinv %*% kronecker(B1,diag(k^2)) %*% kronecker(B2,Kmat) %*% vec(B3)
bias0 <- -(1/n)*bias0

out <- matrix(NA,k^2,3)
dimnames(out) <- list(rep("",k^2),c("ML bias","Due to AR(1)","Due to reg"))
out[,2] <- bias0
out[,3] <- -(r/n)*kronecker(G0inv,Sigma) %*% vec(solve(diag(k) - t(Phi)))
out[,1] <- out[,2] + out[,3]
print(round(out,6))
```

A2. R/S-PLUS program to perform simulation

```
# Simulate the empirical biases of conditional LSE of Phi.

repl <- 1000  # No. of replications
iter <- 10    # No. of iterations for approx ML
k <- 2
n <- 30
r <- 2
```

```
mu <- c(0,0)
Phi <- c(0.2,-0.6,0.3,1.1)  # vec(Phi)
Sigma <- c(4,1,1,2)         # vec(Sigma)

mu <- matrix(mu,k,1)
Phi <- matrix(Phi,k,k)
Sigma <- matrix(Sigma,k,k)
Phihat <- array(NA,c(k,k,repl))

if (r==1) X <- matrix(1,n,1)
if (r==2) X <- cbind(rep(1,n),1:n)
if (r==3) X <- cbind(rep(1,n),1:n,c(1:n)^2)
if (r > 0)
{ X <- kronecker(X,diag(k))
  Xt <- t(X)
  XtX1 <- solve(Xt %*% X)
  A <- XtX1 %*% Xt
  H <- diag(k*n) - X %*% A
}

var1sim <- function(Phi,Sigma,n,k)  # Simulate vector AR(1)
{ n0 <- n + 50
  Y <- matrix(NA,n0,k)
  e <- matrix(NA,n0,k)
  A <- chol(Sigma)  # Choleski decomposition of Sigma = A'A
  for (j in 1:k) e[,j] <- rnorm(n0,mean=0,sd=1)
  e <- e %*% A
  Y[1,] <- e[1,]
  for (i in 2:n0) Y[i,] <- c(Phi %*% Y[i-1,]) + e[i,]
  Y <- Y[(n0-n+1):n0,]
  Y
}

covinvVAR1 <- function(Phi,Sigma,n,k)
{ L <- matrix(0,n,n)
  L[row(L)-col(L)==1] <- 1
  Theta <- kronecker(diag(n),diag(k)) - kronecker(L,Phi)
  vecGamma0 <- solve(diag(k^2) - kronecker(Phi,Phi)) %*% matrix(Sigma,k^2,1)
  Gamma0 <- matrix(vecGamma0,k,k)
  V1 <- matrix(0,k*n,k*n)
  V1[1:k,1:k] <- solve(Gamma0)
  V1[(k+1):(k*n),(k+1):(k*n)] <- kronecker(diag(n-1),solve(Sigma))
  V1 <- t(Theta) %*% V1 %*% Theta
  V1
}
```

```
set.seed(238) # Put the random no. generator in a reproducible state

for (i in 1:repl)
{ N <- var1sim(Phi,Sigma,n,k)
  Y <- matrix(t(N),k*n,1)  # Y = Xbeta + N = N for beta = 0
  if (r > 0)
  { N <- H %*% Y  # Residuals from OLS regression
    N <- t(matrix(N,k,n))
  }
  Nt <- array(t(N),c(k,1,n))
  Gamma0 <- matrix(0,k,k)
  Gamma1 <- matrix(0,k,k)
  for (j in 2:n)
  { Gamma0 <- Gamma0 + Nt[,,j-1] %*% t(Nt[,,j-1])
    Gamma1 <- Gamma1 + Nt[,,j-1] %*% t(Nt[,,j])
  }
  Phat <- t(Gamma1) %*% solve(Gamma0)
  e <- N[2:n,] - N[1:(n-1),] %*% t(Phat)
  Shat <- (t(e) %*% e)/(n-1-k-r)

  if (r > 0)
  { for (h in 1:iter)
    { V1 <- covinvVAR1(Phat,Shat,n,k)
      B <- solve(Xt %*% V1 %*% X)
      glsbeta <- B %*% (Xt %*% V1 %*% Y)
      N <- Y - X %*% glsbeta
      N <- t(matrix(N,k,n))
      Nt <- array(t(N),c(k,1,n))
      Gamma0 <- matrix(0,k,k)
      Gamma1 <- matrix(0,k,k)
      for (j in 2:n)
      { Gamma0 <- Gamma0 + Nt[,,j-1] %*% t(Nt[,,j-1])
        Gamma1 <- Gamma1 + Nt[,,j-1] %*% t(Nt[,,j])
      }
      Phat <- t(Gamma1) %*% solve(Gamma0)
      e <- N[2:n,] - N[1:(n-1),] %*% t(Phat)
      Shat <- (t(e) %*% e)/(n-1-k-r)
    }
  }
  Phihat[,,i] <- Phat
}

Phihat <- apply(Phihat,c(1,2),mean)
print(round(Phihat-Phi,6))
```

References

[1] Anderson, T. W., *The Statistical Analysis of Time Series*, John Wiley, 1971.

[2] Cheang, W. K., *Issues on Estimation of Time Series Regression Model With Autocorrelated Noise*, PhD dissertation, University of Wisconsin – Madison, 2000.

[3] Cheang, W. K. and Reinsel, G. C., "Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood", *Journal of the American Statistical Association*, 2000, **95**, 1173–1184.

[4] Cheang, W. K. and Reinsel, G. C., "Finite sample properties of ML and REML estimators in time series regression models with long memory noise", *Journal of Statistical Computation and Simulation*, 2003, **73**, 233–259.

[5] Cordeiro, G. M. and Klein, R., "Bias correction in ARMA models", *Statistics and Probability Letters*, 1994, **19**, 169–176.

[6] Lewis, R. and Reinsel, G. C., "Prediction error of multivariate time series with mis-specified models", *Journal of Time Series Analysis*, 1988, **9**, 43–57.

[7] Lütkepohl, H., *Introduction to Multiple Time Series Analysis*, 2nd edition, Springer-Verlag, 1993.

[8] Reinsel, G. C., *Elements of Multivariate Time Series Analysis*, 2nd edition, Springer-Verlag, 1997.

[9] Shaman, P. and Stine, R. A., "The bias of autoregressive coefficient estimators", *Journal of the American Statistical Association*, 1988, **83**, 842–848.

[10] Yamamoto, T. and Kunitomo, N., "Asymptotic bias of the least squares estimator for multivariate autoregressive models", *Annals of the Institute of Statistical Mathematics*, 1984, **36**, 419–430.