An alternative approach to approximating the moments of least squares estimators

Liu-Evans, Gareth

Xi'an Jiaotong Liverpool University

9 November 2010

An alternative approach to approximating moments of least squares estimators

Gareth D. Liu-Evans

November 9, 2010

Abstract

A new methodology is presented for approximating the moments of least squares coefficient estimators in situations where endogeneity and dynamics are present. The OLS estimator is the focus here, but the method, which is valid under a simple set of smoothness and moment conditions, can be applied to related estimators. An $O(T^{-1})$ approximation is presented for the bias in OLS estimation of a general ARX(p) model.

Introduction

A recent summary of the work on asympotic approximation of moments in econometrics can be found in Ullah (2005). Two papers of interest include Phillips (2000), which presents new approximations for the bias and mean squared error in 2SLS estimation of a static simultaneous equation model[1], and Bao & Ullah (2007), where a method is presented for approximating the moments of time-series estimators, building on Rilestone et al. (1996) for a class of nonlinear estimators. An earlier version of Bao & Ullah (2007), namely Bao & Ullah (2002), applies their methodology to a structural model with autoregressive error terms. These papers all develop asymptotic approximation methods that are valid under straightforward smoothness conditions and moment bounds. This is arguably an improvement on earlier papers, where the validity conditions for the expansions were more complicated. A key contribution of Phillips (2000), for example, was to obtain the 2SLS moment approximations in Nagar (1959) using a more understandable set of assumptions[2].

The current paper develops a new method for asymptotically approximating the moments of least squares coefficient estimators under similar

[1]See also Phillips (2007) and Phillips & Liu-Evans (2010).

[2]A discussion of the Nagar expansion method can be found in Sargan (1974).

assumptions to those in the recent papers mentioned above, particularly Phillips (2000). The focus is on the OLS estimator, but it is shown that the method can be applied to other estimators as well. It can be applied when there is both simulteneity and dynamics in the model being estimated. For illustration the approach is used to give an $O(T^{-1})$ approximation to the bias in OLS estimation of a stationary autoregressive model with arbirary lag order, added exogenous regressors, and non-normal disturbances.

There has been some recent interest in the moments of the OLS coefficient estimator. In the context of stationary autoregressive models, Kiviet & Phillips (2010) provide an order $O(T^{-2})$ Nagar approximation to the bias in OLS estimation of the coefficient vector in an ARX(1) model with normal errors. Bao (2007), using a Nagar-type methodology developed in Kiviet & Phillips (1993), finds the $O(T^{-2})$ bias and mean squared error in estimation of the AR(1) model with and without constant, and with model errors that can be skewed and lepto/platykurtic. Bao & Ullah (2007) present an expression for the $O(T^{-1})$ bias in OLS estimation of an ARX(1), again with non-normal errors, but this time using their alternative expansion method. They do the same for a stationary VAR(1) model with non-normal errors, building on Kiviet & Phillips (1994) for the VAR(p) with normal errors. Kiviet & Phillips (2005) use a Nagar-type expansion to find the $O(T^{-1})$ bias in estimation of the unit-root ARX(1) with normal errors. In a non-dynamic context, Kiviet & Phillips (1996) use a Nagar-type methodology to find the $O(T^{-1})$ bias in OLS estimation of a simultaneous equation model, building on Kadane (1971) who developed a small-σ expansion for k-class estimation.

Kiviet & Phillips (2010) includes a review of the earlier moment expansion work for the OLS estimator in the context of first-order stationary autoregressive models. See in particular Kendall (1954) and Marriott & Pope (1954), who were first to present approximations for the bias in OLS estimation of an autoregressive model. Some examples of work on higher-order autoregressive models are Bhansali (1981), Maekawa (1983), Tjostheim & Paulsen (1983), Tanaka (1984), Yamamoto & Kunitomo (1984), Kunitomo & Yamamoto (1985) and Shaman & Stine (1988), who consider AR(p) models, and Kiviet & Phillips (1994), who consider the coefficient vector in an ARX(p) model under a normality assumption. The current paper extends this by considering the ARX(p) model without a normality assumption, it also builds on the ARX(1) illustration in Bao & Ullah (2007), where the skewness and kurtosis of the errors is included explicitly, by allowing lags up to order p.

The underlying approach in the current paper is similar to the early work by Marriott & Pope (1954) and Kendall (1954), where the main focus was the kth-order autocorrelation coefficient[3]. An observation, along with the matrix results in Magnus & Neudecker (1979) and Magnus & Neudecker (1988), allows a more general class of models and estimators to be considered.

The expansion method

Given an AR(1) model $y_t = \lambda y_{t-1} + u_t$, $t = 1, .., T$, with $u_t \overset{iid}{\sim} N(0, \sigma_u^2)$, Marriott & Pope (1954)[4] write the kth-order autocorrelation coefficient of the AR(1) as a ratio $\hat{\lambda}_k = A/B$ and take a second-order Taylor-series expansion of A/B around a and b, the means of A and B. After calculating the expected value of this series and excluding terms smaller $O(T^{-1})$, they obtain an asymptotic approximation to $E[\hat{\lambda}_1 - \lambda]$ and state that the remainder is $o(T^{-1})$.

We note here that, for a model $y = Z\alpha + u$ with $E[Z'u] = 0$ (this assumption is dropped later), the true coefficient α and its OLS estimator $\hat{\alpha}$ can always be expressed in the same functional form:

$$\begin{aligned} \hat{\alpha} &= (Z'Z)^{-1}Z'y \\ \alpha &= (E[Z'Z])^{-1}E[Z'y]. \end{aligned}$$

Defining the matrices $\hat{R} = [Z'Z : \hat{\zeta}]$ and $R = [E[Z'Z] : \zeta]$, where $\hat{\zeta} = Z'y$ and $\zeta = E[Z'y]$, the estimated and true coefficients can then be expressed as $\hat{\alpha}_i = f_i(\hat{\delta})$ and $\alpha_i = f_i(\delta)$, respectively, where $\hat{\delta} = vec(\hat{R})$ and $\delta = vec(R)$. This allows a Taylor series expansion of the following form:

$$\begin{aligned} f_i(\hat{\delta}) &= f_i(\delta) + (\hat{\delta} - \delta)' f_i'(\delta) + \frac{1}{2}(\hat{\delta} - \delta)' H_i|_\delta (\hat{\delta} - \delta) + .., \\ \text{i.e.} \quad \hat{\alpha}_i - \alpha_i &= (\hat{\delta} - \delta)' f_i'(\delta) + \frac{1}{2}(\hat{\delta} - \delta)' H_i|_\delta (\hat{\delta} - \delta) + .., \end{aligned}$$

where $H_i|_\delta$ is the Hessian matrix of f evaluated at $\delta = vec(R)$. Phillips (2000) uses a similar approach by forming an expansion around the vectorised reduced form parameters in a static simultaneous equation model.

Using the extended mean value theorem we can write

$$\begin{aligned} f_i(\hat{\delta}) = f_i(\delta) &+ (\hat{\delta} - \delta)' f_i'(\delta) + \frac{1}{2}(\hat{\delta} - \delta)' H_i|_\delta (\hat{\delta} - \delta) \\ &+ \frac{1}{3!} \sum_{j=1}^{r} (\hat{\delta}_j - \delta_j)(\hat{\delta} - \delta)' f_{ij}^{(3)}|_{\delta^\star} (\hat{\delta} - \delta) \end{aligned}$$

[3]As Marriott & Pope (1954) note, the OLS bias in estimation of an AR(1) is the same to order $O(T^{-1})$ as the bias in the $1st$-order autocorrelation coefficient of an AR(1).

[4]See also Kendall (1954).

for some $\delta^\star$, where r denotes the row dimension of δ, and where $f_{ij}^{(3)}$ is an $r \times r$ matrix of derivatives defined as $f_{ij}^{(3)} = \frac{\partial H_i}{\partial \delta_j}$. Here it is assumed that f_i is differentiable up to third order with derivatives that are uniformly bounded in a neighbourhood of δ as $T \to \infty$, and with third-order derivatives that are continuous. Given these assumptions, the fourth term above is $O_p(T^{-\frac{3}{2}})$. If the components of $\hat{\delta}$ are assumed to have finite moments up to third order we therefore have the following:

$$E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2}E[(\hat{\delta} - \delta)' H_i|_\delta (\hat{\delta} - \delta)] + o(T^{-1})$$

$$\text{or} \quad E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2} tr(H_i|_\delta Var(\hat{\delta})) + o(T^{-1}),$$

where $Var(\hat{\delta})$ is the covariance matrix for $\hat{\delta}$. This step is clear from Shao & Tu (1995), see in particular section 2.4 (see also Shao (1988)). Phillips (2000) uses a similar argument. Rearranging these slightly we have

Theorem 1. *Under the assumptions made,*

$$E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2}\left(tr(H_i|_\delta J) + \delta' H_i|_\delta \delta\right) + o(T^{-1}),$$

where $J = E[\hat{\delta}\hat{\delta}']$.

The following theorem provides the unevaluated Hessian matrix, H_i. To evaluate the Hessian at δ, we replace $\hat{\zeta}$ and $Z'Z$ with their expected values, and this is done later for autoregressive models. A 'Kronecker power' notation is introduced: $A \otimes A \otimes ... \otimes A = A^{\otimes m}$, where A appears m times. Since we have the well-known result that $(A \otimes A \otimes ... \otimes A)^{-1} = A^{-1} \otimes A^{-1} \otimes ... \otimes A^{-1}$ when A is invertible, we can write these as $A^{\otimes(-m)}$.

Theorem 2. *Let Z have dimensions $T \times N$, and let $\Gamma_1 = [0_{N \times N^2} : I_N]$ and $\Gamma_2 = [I_{N^2} : 0_{N^2 \times N}]$. Let $V_1 = vec(I_N)$, $V_2 = (K_{NN} \otimes I_N)(I_N \otimes vec(Z'Z))$ and $V_3 = (I_N \otimes K_{NN})(vec(Z'Z) \otimes I_N)$, where K_{nm} is an $nm \times nm$ commutation matrix. Let e_i be an $N \times 1$ unit vector with unity in position i. Then, using the identification theorems in Magnus & Neudecker (1988), the unevaluated Hessian matrix is*

$$H_i = \frac{1}{2}\left(MB_i + B_i' M'\right),$$

where

$$M = \Big(([\Gamma_2' \otimes (\hat{\zeta}' \otimes I_N)](Z'Z)^{\otimes(-4)}[(I_N \otimes V_2) + (V_3 \otimes I_N)]\Gamma_2 - (\Gamma_1' \otimes I_N)(Z'Z)^{\otimes(-2)}\Gamma_2 - [((Z'Z)^{\otimes(-2)}\Gamma_2)' \otimes I_N](I_N \otimes V_1)\Gamma_1 \Big)',$$

and where B_i is defined as

$$
\begin{aligned}
(B_i)_{n,m} &= 1 \ \ \textit{for} \ \ n = qN + i, \quad m = q+1, \quad q = 0, 1, ...N^2 + N - 1 \\
&= 0 \ \ \textit{otherwise.}
\end{aligned}
$$

Proof. See Appendix A.1.

Note that the expansion approach developed here will yield moment approximations where $E[Z'u] \neq 0$, i.e. where endogeneity is present. Premultiplying $y = Z\alpha + u$ by Z' and taking the expected value as above, we have the following:

$$
\begin{aligned}
\alpha &= (E[Z'Z])^{-1}E[Z'y] - (E[Z'Z])^{-1}E[Z'u] \\
\Rightarrow \alpha_i &= f_i(\delta) - e_i' E_1^{-1} E_2,
\end{aligned}
$$

where $E_1 = E[Z'Z]$ and $E_2 = E[Z'u]$. Since it is still true that

$$E[\hat{\alpha}_i] = f_i(\delta) + \frac{1}{2}\left(tr(H_i|_\delta J) + \delta' H_i|_\delta \delta\right) + o(T^{-1}),$$

the bias in OLS estimation is

$$E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2}\left(tr(H_i|_\delta J) + \delta' H_i|_\delta \delta\right) + e_i' E_1^{-1} E_2 + o(T^{-1}).$$

The approximation is valid to order $O(T^{-1})$ under the same conditions as before, because the expected values E_1 and E_2 are exact. Despite the additional calculation here in the case where $E[Z'u] \neq 0$, the burden of expectation calculation is lower than in the Nagar approach: the terms $H_i|_\delta$ and δ just require us to know the expected values of $Z'Z$ and $Z'y$, and J can be found using existing results on the moments of products of quadratic forms.

In the case where $E[Z'u] \neq 0$, the moments of the IV estimator may be of more interest, since in practice there are often instruments available that allow us to do better than OLS. The estimator is

$$\hat{\alpha}_{IV} = (Z'PZ)^{-1}Z'Py,$$

where P is a projection matrix of instruments, and we can write

$$\alpha = (E[Z'PZ])^{-1}E[Z'Py] - (E[Z'PZ])^{-1}E[Z'Pu].$$

We can therefore use essentially the same expansion methodology. Here it is more difficult to calculate the necessary expected values: we have $P = W(W'W)^{-1}W$ where the $N \times g$ matrix W is a matrix of g instruments, and it is relatively difficult to calculate expected values that involve a matrix inverse. A solution is to approximate $(W'W)^{-1}$ by taking a Taylor series around $vec(E[W'W])$ up to an appropriate order, then for all the expected value calculations we are back to familiar territory with no inverted matrices. In practice, the increased algebraic complexity is likely to demand a different calculation approach to the one used in this paper for OLS. Another application of interest is the Within Group (WG) estimator for dynamic panel data modeling. Here we replace P in the above by $D'D$, where D is a $1-0$ difference matrix.

The following sections apply the above to the ARX(p) with non-normal disturbances. As in Bao & Ullah (2007) and Bao (2007) the third and fourth moments of the model errors are expressed in terms of skewness and excess kurtosis parameters, so that the effects of departures from normality can be seen more easily.

ARX(p)

Consider an autoregressive model with p lags and k added exogenous variables:

$$y = \lambda_1 y_{-1} + ... + \lambda_p y_{-p} + X\beta + u,$$

where $u = \Gamma_3 v$ with $\Gamma_3 = [0_{T\times p} : I_T]$, and where v is a $(T+p)\times 1$ random vector with the following moment properties.

Assumption 1. *The $i-th$ elements of v have finite moments up to 6th order with:*

$$E[v_i] = 0, \quad E[v_i^2] = \sigma^2 \quad E[v_i^3] = \sigma^3\gamma_1 \quad E[v_i^4] = \sigma^4(\gamma_2 + 3),$$

where γ_1 and γ_2 are Pearson's measures of skewness and excess kurtosis.

It is also assumed that the process is stationary in the sense that all roots of $1 - \lambda_1 r - \lambda_2 r - ... - \lambda_p r = 0$ lie outside the unit circle. This assumption, combined with Assumption 1, makes the process covariance stationary. The assumption of finite moments to 6th order for v ensures that $\hat{\delta}$ has finite moments up to 3rd order, which is a condition for Theorem 1.

From this we can write the following for periods $1-p$ through to $T-1$, building on the approach in Kiviet & Phillips (2010)[5]:

$$\Lambda Y_{-1} = \bar{Y}^{\star}_{-1} + [I_{T+p-1} : 0_{(T+p-1)\times 1}]\Omega v,$$

[5] See also Kiviet & Phillips (1994).

where $Y_{-1} = (y_{1-p}, ..., y_{T-1})'$, $\bar{Y}^\star_{-1} = (\bar{y}_{1-p}, ..., \bar{y}_0, x'_1\beta, ..., x'_T\beta)'$ and $\bar{Y}_{-1} = \Lambda^{-1}\bar{Y}^\star_{-1}$. The matrices Λ and Ω are defined, respectively, as

$$\Lambda = \begin{pmatrix} 1 & 0 & . & & . & . & . & 0 \\ 0 & 1 & 0 & & & & & . \\ . & . & . & & . & & & . \\ 0 & . & 0 & & . & . & & . \\ -\lambda_p & . & . & & -\lambda_1 & . & 0 & \\ 0 & & . & & . & . & 1 & 0 \\ 0 & 0 & -\lambda_p & & . & & -\lambda_1 & 1 \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \omega & 0 & . & . & . & . & 0 \\ 0 & . & 0 & . & & . & . \\ . & . & . & . & & . & . \\ . & & & \omega & & & . \\ . & & & & 1 & . & . \\ . & & & & & . & 0 \\ 0 & . & . & . & . & 0 & 1 \end{pmatrix},$$

where Λ has λ_{-p} as the $(p+1)th$ element of the first column, and where the ω term, which is the standard deviation of y_t, appears p times in Ω. This can alternatively be written as

$$Y_{-1} = \bar{Y}_{-1} + Gv,$$

where $G = \Lambda^{-1}[I_{T+p-1} : 0_{(T+p-1)\times 1}]$, and we can note that $y_{-i} = M_iY_{-i}$ and $\bar{y}_{-i} = M_i\bar{Y}_{-i}$ for $i = 1, ..., p$, where $M_i = [0_{T\times(p-i)} : I_T : 0_{T\times(i-1)}]$. This implies

$$y_{-i} = \bar{y}_{-i} + G_iv,$$

where $G_i = M_iG$. When $p = 1$, M_i is the identity matrix and G is the same G that appears in Kiviet & Phillips (2010).

A decomposition of $Z = (y_{-1}, ..., y_{-p}, X)$ is also required, and the following proceeds in a similar way to Kiviet & Phillips (2010). First we can write $Z = \bar{Z} + \tilde{Z}$, where $\bar{Z} = (\bar{y}_{-1}, ..., \bar{y}_{-p}, X)$ and $\tilde{Z} = (G_1v, ..., G_pv, 0_{T\times k})$. We have $\tilde{Z} = (G_1v, ..., G_pv)[I_p : 0_{p\times k}]$, which gives

$$Z = \bar{Z} + (G_1, ..., G_p)I_bv[I_p : 0_{p\times k}],$$

where I_b is a $p(T+p)\times p(T+p)$ block-diagonal matrix made up of I_{T+p} identity matrices. When $p = 1$, the result reduces to $Z = \bar{Z} + Gve'_1$, which appears in Kiviet & Phillips (2010).

In order to calculate $E[\hat{\delta}\hat{\delta}']$, it is useful to write $\hat{\delta}$ in the form $\hat{\delta} = Q_1 + Q_2v + Q_3v'Q_4v + Q_5v'Q_6v$, so that the expected value of $\hat{\delta}\hat{\delta}'$ can be calculated using existing results on expectations of products of quadratic forms.[6] We can do this by noting $\hat{\delta} = \Gamma'_2vec(Z'Z) + \Gamma'_1vec(\hat{\zeta})$, where Γ_1 and Γ_2 are the same as in the preceding section with $N = p + k$, and then by expressing $vec(Z'Z)$ and $vec(\hat{\zeta})$ in terms of v. We already have Z in terms of v:

$$Z = \bar{Z} + G_bv\Gamma_4,$$

[6]See e.g. the Appendix in Ullah (2005).

where $G_b = (G_1, ..., G_p)I_b$ and $\Gamma_4 = [I_p : 0_{p\times k}]$. Recalling that $vec(\hat{\zeta}) = vec(Z'y)$, it will be useful to write y in terms of v as well:

$$\begin{aligned} y &= \sum_{i=1}^{p} \lambda_i(\bar{y}_{-i} + G_i v) + X\beta + u \\ &= (\sum_{i=1}^{p} \lambda_i \bar{y}_{-i}) + X\beta + \{(\sum_{i=1}^{p} \lambda_i G_i) + \Gamma_3\}v, \end{aligned}$$

where $u = \Gamma_3 v$ as before. Using these decompositions of y and Z, it is straightforward to express $vec(\hat{\zeta})$ and $vec(Z'Z)$ in the desired form, and this is done in Lemma 1 below.

Lemma 1. *In the ARX(p) model the terms $vec(\hat{\zeta})$ and $vec(Z'Z)$ are*

$$\begin{aligned} vec(\hat{\zeta}) &= P_1 + P_2 v + P_3 v' P_4 v \\ vec(Z'Z) &= A_1 + A_2 v + A_3 v' A_4 v, \end{aligned}$$

where

$$\begin{aligned} &P_1 = vec[\bar{Z}'\{(\bar{Z}\sum_{i=1}^{p} \lambda_i \bar{y}_{-i}) + X\beta\}] \\ &P_2 = ([\{(\sum_{i=1}^{p} \lambda_i \bar{y}_{-i}) + X\beta\}' G_b] \otimes \Gamma_4) + \bar{Z}'\{(\sum_{i=1}^{p} \lambda_i G_i) + \Gamma_3\} \\ &P_3 = vec(\Gamma_4'), \quad P_4 = G_b'\{(\sum_{i=1}^{p} \lambda_i G_i) + \Gamma_3\}, \quad A_1 = vec(\bar{Z}'\bar{Z}) \\ &A_2 = \{(\Gamma_4' \otimes \bar{Z}' G_b) + (\bar{Z}' G_b \otimes \Gamma_4')\}, \quad A_3 = \Gamma_4' \otimes \Gamma_4', \quad A_4 = G_b' G_b \end{aligned}$$

Proof. See Appendix A.2.

Using Lemma 1, it is possible to calculate the expectated value J, and this is done below.

Lemma 2. *The expected value J is*

$$\begin{aligned}
J = {} & Q_1 Q_1' + \sigma^2 tr(Q_4') Q_1 Q_3' + \sigma^2 tr(Q_6') Q_1 Q_5' \\
& + \sigma^2 tr(Q_4) Q_3 Q_1' + \sigma^4 Q_3 tr[Q_4' \{tr(Q_4) I_{T+1} + Q_4 + Q_4'\}] Q_3' \\
& + \sigma^4 Q_3 tr[Q_6' \{tr(Q_4) I_{T+1} + Q_4 + Q_4'\}] Q_5' \\
& + \sigma^2 tr(Q_6) Q_5 Q_1' + \sigma^4 Q_5 tr[Q_4' \{tr(Q_6) I_{T+1} + Q_6 + Q_6')\}] Q_3' \\
& + \sigma^4 Q_5 tr[Q_6' \{tr(Q_6) I_{T+1} + Q_6 + Q_6'\}] Q_5' \\
& + \sigma^3 \gamma_1 \{Q_2 (I_{T+1} \circ (Q_4')) i Q_3' + Q_2 (I_{T+1} \circ (Q_6')) i Q_5' \\
& + Q_3 \{(I_{T+1} \circ Q_4') i\}' Q_2' + Q_5 \{(I_{T+1} \circ Q_6) i\}' Q_2'\} \\
& + \sigma^4 \gamma_2 \{Q_3 Q_3' tr[Q_4' (I_{T+1} \circ Q_4)] + Q_3 Q_5' tr[Q_6' (I_{T+1} \circ Q_4)] \\
& + Q_5 Q_3' tr[Q_4' (I_{T+1} \circ Q_6)] + Q_5 Q_5' tr[Q_6' (I_{T+1} \circ Q_6)]\}.
\end{aligned}$$

where "$\circ$" is the Hadamard matrix product, and where $Q_1 = \Gamma_2' A_1 + \Gamma_1' P_1$, $Q_2 = \Gamma_2' A_2 + \Gamma_1' P_2$, $Q_3 = \Gamma_2' A_3$, $Q_4 = A_4$, $Q_5 = \Gamma_1' P_3$ and $Q_6 = P_4$.
Proof. See Appendix A.3.

Finally, we have the following key theorem:

Theorem 3. *The bias in OLS regression of the ARX(p) model is*

$$E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2}\{tr(H_i|_\delta J) - \delta' H_i|_\delta \delta\} + o(T^{-1}),$$

where J is given in Lemma 2, $\delta = Q_1 + Q_3 tr(Q_4) + Q_5 tr(Q_6)$, and where the Hessian is evaluated at δ, or, equivalently, at $E[Z'Z] = \bar{Z}'\bar{Z} + \Gamma_4'\Gamma_4 tr(G_b' G_b)$ and at $E[\zeta] = P_1 + P_3 tr(P_4)$.
Proof. See Appendix A.4.

ARX(1)
The bias in OLS estimation of ARX(1) models is a corollary of Theorem 3.

Corollary 1. *The bias expression for estimation of the ARX(1) is the same as in Theorem 3, but with the matrices $P_1, ..., P_4$ and $A_1, ..., A_4$ in Lemma 1 specialised to the following when $p = 1$:*

$$\begin{aligned}
& A_1 = vec(\bar{Z}'\bar{Z}), \quad A_2 = (e_1 \otimes \bar{Z}'G) + (\bar{Z}'G \otimes e_1), \quad A_3 = e_1 \otimes e_1 \quad A_4 = G'G \\
& P_1 = vec[\bar{Z}'(\lambda \bar{y}_{-1} + \beta)], \quad P_2 = \{(\lambda \bar{y}_{-1} + \beta)'G\} \otimes e_1 + \bar{Z}'(\lambda G + \Gamma_3), \\
& P_3 = e_1, \quad P_4 = G'(\lambda G + \Gamma_3),
\end{aligned}$$

and with the other vectors and matrices now reflecting the particular value of p chosen. The matrices B_i are as in Theorem 2 but with $N = k + 1$. For $p = 1$ the matrix Γ_4 reduces to e_1', which is defined in Theorem 2 and now has $N = k + 1$.
Proof. Follows from Theorem 3 and Lemma 1 by setting $p = 1$.

For illustration of the matrices being specialised in Corollary 1, we can see that Λ and Ω become

$$\Lambda = \begin{pmatrix} 1 & 0 & . & . & . & 0 \\ -\lambda & 1 & & & & . \\ 0 & -\lambda & 1 & & & . \\ . & . & . & . & & \\ . & & . & . & . & . \\ 0 & . & . & 0 & -\lambda & 1 \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \omega & 0 & . & . & . & 0 \\ 0 & 1 & 0 & & & . \\ 0 & 0 & 1 & & & . \\ & & & & . & \\ . & & & & . & 0 \\ 0 & . & . & . & 0 & 1 \end{pmatrix}.$$

AR(1)

Kendall (1954) and Marriott & Pope (1954) found the bias in estimation of the model with $p = 1$ and $k = 0$ to be $-\frac{2\lambda}{T}$ to order $O(T^{-1})$ under a normality assumption. Bao & Ullah (2007) show that the $O(T^{-1})$ bias for this model is the same when the disturbance terms are skewed with non-zero excess kurtosis. Here we confirm that skewness and kurtosis in the error terms does not affect the bias to order $O(T^{-1})$. To do this we need to specialise Theorem 3 to the pure AR(1) case, and we need to filter out any unnecessary $o(T^{-1})$ terms,only keeping the $O(T^{-1})$ part of the bias.

As in Kiviet & Phillips (2010), the matix G (our G_1 currently) can be written as $G = [\omega F : C]$, where $F' = (1, \lambda, \lambda^2, ..., \lambda^{T-1})$ and

$$C = \begin{pmatrix} 0 & 0 & . & . & . & 0 \\ 1 & 0 & & & & . \\ \lambda & 1 & 0 & 0 & & . \\ . & & . & . & & \\ . & & . & . & . & . \\ \lambda^{T-2} & . & . & \lambda & 1 & 0 \end{pmatrix}.$$

In order to obtain filtered results for the AR(1) with known mean, certain products of G, G', C, C' and F must be replaced with their $O(T)$ or $O(1)$ approximations, otherwise the resulting OLS bias approximation is accurate to $O(T^{-1})$ but includes some unnecessary $o(T^{-1})$ terms. The required approximations are summarised in

Lemma 3. *Kiviet & Phillips (2010) find the following:*

$$tr(C'C) = T\left(\frac{1}{1-\lambda^2}\right) + o(T), \quad tr(CC'C) = T\lambda\left(\frac{1}{1-\lambda^2}\right)^2 + o(T)$$
$$tr(C'CC'C) = T(\lambda^2+1)\left(\frac{1}{1-\lambda^2}\right)^3 + o(T),$$
$$tr(G'G) = T\left(\frac{1}{1-\lambda^2}\right) + \frac{\omega}{1-\lambda^2} - \left(\frac{1}{1-\lambda^2}\right)^2 + o(1)$$
$$tr(GG'C) = T\lambda\left(\frac{1}{1-\lambda^2}\right)^2 + o(T), \quad tr(GG'GG') = tr(C'CC'C) + o(T)$$

A further Lemma is also required:

Lemma 4. *In the AR(1) model with known mean, i.e.* $y = \lambda y_{-1} + u$*, the matrices* $P_1, ..., P_4$ *and* $A_1, ..., A_4$ *specialise to the following:*

$$A_1 = 0, \; A_2 = 0, \; A_3 = 1, \; A_4 = G'G$$
$$P_1 = 0, \; P_2 = 0, \; P_3 = 1, \; P_4 = G'(\lambda G + [0_{T\times 1} : I_T]),$$

giving

$$Q_1 = 0, \quad Q_2 = 0, \quad Q_3 = e_1, \quad Q_4 = A_4, \quad Q_5 = e_2, \quad Q_6 = P_4.$$

Here e_1 *and* e_2 *are* 2×1 *unit matrices with unity in rows* 1 *and* 2 *respectively.*
Proof. Follows from Corollary 1 by setting $k = 0$.

The following can now be stated:

Corollary 2. *The bias in OLS estimation of the autoregressive coefficient in the pure AR(1) model, with* γ_1 *and* γ_2 *taking any finite value, is*

$$E[\hat{\lambda} - \lambda] = -\frac{2\lambda}{T} + o(T^{-1}).$$

Proof. See Appendix A.5.

This agrees with the Bao & Ullah (2007) generalisation of the original Kendall (1954) and Marriott & Pope (1954) result. As a by-product, the following approximation for the evaluated Hessian matrix was found, where the subscript i is dropped given that there is only one element in δ:

$$H_i|_\delta = \begin{pmatrix} \frac{2\lambda(1-\lambda^2)^2}{T^2\sigma^4} & -\frac{(1-\lambda^2)^2}{T^2\sigma^4} \\ -\frac{(1-\lambda^2)^2}{T^2\sigma^4} & 0 \end{pmatrix} + o(T^{-2}).$$

This is explained in the proof to Corollary 1.

Note that the bias result here assumes a random covariance-stationary startup, while Bao & Ullah (2007) and Bao (2007) assume a fixed startup. Kiviet & Phillips (2010) show that this distinction has an effect on the bias in the case of normal errors to order $O(T^{-2})$ but not to order $O(T^{-1})$. Here we can see that the latter is still true when the model errors are non-normal. It is also clear that the skewness parameter γ_1 does not enter the $O(T^{-1})$ bias expression for the more general AR(p) with no intercept: in this case $\bar{Z} = \bar{y}_{-1} = 0$ so $Q_2 = 0$ from Lemmas 1 and 2. An earlier sequence of papers starting with Bhansali (1981) and ending with Shaman & Stine (1988) also finds this, though the validity conditions for these approximations are quite different, e.g. the bias approximations in Bhansali (1981) and Shaman & Stine (1988) assume finite error moments up to the 12th and 16th orders, respectively.

Conclusion

An alternative asymptotic expansion method is developed here for approximating the moments of least squares estimators, particularly those of the OLS estimator. The method is used to obtain the first $O(T^{-1})$ bias approximation for the OLS coefficient estimator of a general ARX(p) model with non-normal disturbances and arbitrary lag order. It is shown that the method can potentially be used to find the moments of other estimators.

Appendix

A.1. Proof of Theorem 2

(i) Derivation of H_i

H_i can be found using the Second Identification Theorem in Magnus & Neudecker (1988)[7]. This requres the second differential of $\hat{\alpha}_i = e_i'(Z'Z)^{-1}\hat{\zeta}$ to be expressed in the form $(d\hat{\delta})'A_i(d\hat{\delta})$ where A_i is a constant matrix. The Hessian is then $H_i = \frac{1}{2}(A + A')$. The first differential can be calculated as follows:

$$\begin{aligned} d\hat{\alpha} &= (d(Z'Z)^{-1})\hat{\zeta} + (Z'Z)^{-1}d\hat{\zeta} \\ &= -(\hat{\zeta}' \otimes I_N)(Z'Z)^{\otimes(-2)}vec(d(Z'Z)) + (Z'Z)^{-1}d\hat{\zeta}. \end{aligned}$$

We can write $vec(\hat{\zeta}) = \Gamma_1\hat{\delta}$ and $vec(Z'Z) = \Gamma_2\hat{\delta}$, where Γ_1 and Γ_2 are defined in Theorem 2. Using this gives

$$d\hat{\alpha} = Nd\hat{\delta},$$

where $N = (Z'Z)^{-1}\Gamma_1 - (\hat{\zeta}' \otimes I_N)(Z'Z)^{\otimes(-2)}\Gamma_2$. For the first differential we now have

$$\begin{aligned} d\hat{\alpha} &= vec(I_N N(d\hat{\delta})) \\ &= ((d\hat{\delta})' \otimes I_N)vecN, \end{aligned}$$

which is a convenient form for calculating the second:

$$d^2\hat{\alpha}_i = d(vec(N))'((d\hat{\delta}) \otimes I_N)e_i.$$

In the above we use $d(d\hat{\delta}') = 0$, since $d\hat{\delta}$ is the constant vector increment in the differential $d\hat{\alpha}$. Note that the term $((d\hat{\delta}) \otimes I_N)e_i$ can be written as $B_i\hat{\delta}$, where B_i is derived in part (ii) below, so that

$$d^2\hat{\alpha}_i = (dvec(N))'B_i(d\hat{\delta}).$$

The remaining task is to put the second differential in the form

$$d^2\hat{\alpha}_i = (d\hat{\delta})'MB_i(d\hat{\delta})$$

for some M, then the Hessian can be identified as $H_i = \frac{1}{2}(MB_i + B_i'M')$. We therefore need to put $dvec(N)$ in the form $dvec(N) = M'd\hat{\delta}$.

From $N = (Z'Z)^{-1}\Gamma_1 - (\hat{\zeta}' \otimes I_N)(Z'Z)^{\otimes(-2)}\Gamma_2$ we have

$$dvec(N) = dvec[(Z'Z)^{-1}\Gamma_1] - dvec[(\hat{\zeta}' \otimes I_N)(Z'Z)^{\otimes(-2)}\Gamma_2].$$

[7] In particular, see the second line of Table 1 in Chapter 10.

The first term of this can be written as $-(\Gamma_1' \otimes I_N)(Z'Z)^{\otimes(-2)}\Gamma_2 d\hat{\delta}$, and the second term can be written as follows:

$$\begin{aligned}
&vec((d(\hat{\zeta}' \otimes I_N))(Z'Z)^{\otimes(-2)}\Gamma_2 + (\hat{\zeta}' \otimes I_N)d[(Z'Z)^{\otimes(-2)}\Gamma_2]) \\
&= (((Z'Z)^{\otimes(-2)}\Gamma_2)' \otimes I_N)(I_N \otimes V_1)vec(d\hat{\zeta}') \\
&\quad - [\Gamma_2' \otimes (\hat{\zeta}' \otimes I_N)]vec((Z'Z)^{\otimes(-2)}d((Z'Z)^{\otimes 2})(Z'Z)^{\otimes(-2)}) \\
&= (((Z'Z)^{\otimes(-2)}\Gamma_2)' \otimes I_N)(I_N \otimes V_1)\Gamma_1\hat{\delta} \\
&\quad - [\Gamma_2' \otimes (\hat{\zeta}' \otimes I_N)](Z'Z)^{\otimes(-4)}[(I_N \otimes V_2) + (V_3 \otimes I_N)]\Gamma_2 d\hat{\delta},
\end{aligned}$$

where, following the result in the exercise on p48 of Magnus & Neudecker (1988), $V_1 = (K_{21} \otimes I_2)(I_1 \otimes vec(I_N)) = (K_{N1} \otimes I_N)vec(I_N) = vec(I_N)$, $V_2 = (K_{NN} \otimes I_N)(I_N \otimes vec(Z'Z))$ and $V_3 = (I_N \otimes K_{NN})(vec(Z'Z) \otimes I_N)$. The term K_{nm} is an $nm \times nm$ commutation matrix, as defined in the Theorem.

From the above we have

$$dvec(N) = M'd\hat{\delta},$$

where M is given in the Theorem. Therefore $d^2\hat{\alpha}_i = (d\hat{\delta})'MB_i(d\hat{\delta})$ and the Hessian is

$$H_i = \frac{1}{2}\left(MB_i + B_i'M'\right).$$

(ii) Derivation of B_i

$\{(\hat{\delta} \otimes I_N)\}$ is an $N^2(N+1) \times N$ matrix and e_i is an $N \times 1$ vector with unity in element i. Let $d\hat{\delta} = (d\hat{r}_1, ..., d\hat{r}_{N(N+1)})'$. Then we have the $N^2(N+1) \times 1$ vector $\{\hat{\delta} \otimes I_N)\}e_i = (d\hat{r}_1 e_i, ..., d\hat{r}_{N(N+1)}e_i) = B_i d\hat{\delta}$ for some constant $N^2(N+1) \times N(N+1)$matrix B_i. Consider the case $i = 1$, and let the $nm-th$ element of a matrix A be denoted as $(A)_{nm}$. Here we see that $(B_1)_{nm} = 1$ for $n = 1$ and $m = 1$, and for $n = N+1$ and $m = 2$, and more generally for $n = qN + 1$ and $m = q + 1$ up to $q = N^2 + N - 1$, with all other elements being zero. Similarly, we can see that $(B_2)_{nm} = 1$ for $n = qN + 2$ and $m = q + 1$, up to $q = N^2 + N - 1$, and zero otherwise. For general i we therefore have $(B_i)_{nm} = 1$ for $n = qN + i$, $m = q + 1$ and $q = 1, ..., N^2 + N - 1$.

A.2. Proof of Lemma 1

We have

$$
\begin{aligned}
vec(\hat{\zeta}) &= vec(Z'y) \\
&= vec[(\bar{Z} + G_b v\Gamma_4)'\{(\sum_{i=1}^{p} \lambda_i \bar{y}_{-i}) + X\beta + \{(\sum_{i=1}^{p} \lambda_i G_i) + \Gamma_3\}v\}] \\
&= vec[\bar{Z}'\{(\sum_{i=1}^{p} \lambda_i \bar{y}_{-i}) + X\beta\}] + \{([\{(\sum_{i=1}^{p} \lambda_i \bar{y}_{-i}) + X\beta\}'G_b] \otimes \Gamma_4) \\
&\quad + \bar{Z}'\{(\sum_{i=1}^{p} \lambda_i G_i) + \Gamma_3\}\}v + vec(\Gamma_4')v'G_b'\{(\sum_{i=1}^{p} \lambda_i G_i) + \Gamma_3\}v \\
&= P_1 + P_2 v + P_3 v' P_4 v.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
vec(Z'Z) &= vec[(\bar{Z} + G_b v\Gamma_4)'(\bar{Z} + G_b v\Gamma_4)] \\
&= vec(Z'Z) + \{(\Gamma_4' \otimes \bar{Z}'G_b) + (\bar{Z}'G_b \otimes \Gamma_4')\}v + (\Gamma_4' \otimes \Gamma_4')v'G_b'G_b v \\
&= A_1 + A_2 v + A_3 v' A_4 v.
\end{aligned}
$$

A.3. Proof of Lemma 2

Since $\hat{\delta} = \Gamma_2' vec(Z'Z) + \Gamma_1' vec\hat{\zeta}$ we have $\hat{\delta} = Q_1 + Q_2 v + Q_3 v' Q_4 v + Q_5 v' Q_6 v$ from Lemma 1 where $Q_1 = \Gamma_2' A_1 + \Gamma_1' P_1$, $Q_2 = \Gamma_2' A_2 + \Gamma_1' P_2$, $Q_3 = \Gamma_2' A_3$, $Q_4 = A_4$ and $Q_5 = \Gamma_1' P_3$, $Q_6 = P_4$. To calculate J we can write

$$
\hat{\delta}\hat{\delta}' = (Q_1 + Q_2 v + Q_3 v' Q_4 v + Q_5 v' Q_6 v)(Q_1 + Q_2 v + Q_3 v' Q_4 v + Q_5 v' Q_6 v)',
$$

then after eliminating terms that have zero expectation we have

$$
\begin{aligned}
E[\hat{\delta}\hat{\delta}'] &= Q_1 Q_1' + Q_1 E[v'Q_4' v]Q_3' + Q_1 E[v'Q_6' v]Q_5' + Q_2 E[vv']Q_2' \\
&\quad + Q_2 E[vv'Q_4' v]Q_3' + Q_2 E[vv'Q_6' v]Q_5' + Q_3 E[v'Q_4 v]Q_1' \\
&\quad + Q_3 E[v'Q_4 vv']Q_2' + Q_3 E[v'Q_4 vv'Q_4' v]Q_3' + Q_3 E[v'Q_4 vv'Q_6' v]Q_5' \\
&\quad + Q_5 E[v'Q_6 v]Q_1' + Q_5 E[v'Q_6 vv']Q_2' + Q_5 E[v'Q_6 vv'Q_4' v]Q_3' \\
&\quad + Q_5 E[v'Q_6 vv'Q_6' v]Q_5'.
\end{aligned}
$$

In order to calculate these expected values, the results in Appendix A.5 of

Ullah (2005) are used. Doing this gives

$$
\begin{aligned}
J = {} & Q_1 Q_1' + \sigma^2 tr(Q_4') Q_1 Q_3' + \sigma^2 tr(Q_6') Q_1 Q_5' + \sigma^2 Q_2 Q_2' \\
& + \sigma^3 \gamma_1 Q_2 (I_{T+1} \circ (Q_4')) i Q_3' + \sigma^3 \gamma_1 Q_2 (I_{T+1} \circ (Q_6')) i Q_5' \\
& + \sigma^2 tr(Q_4) Q_3 Q_1' + \sigma^3 \gamma_1 Q_3 \{(I_{T+1} \circ Q_4') i\}' Q_2' \\
& + \sigma^4 Q_3 tr[Q_4' \{\gamma_2 (I_{T+1} \circ Q_4) + tr(Q_4) I_{T+1} + Q_4 + Q_4'\}] Q_3' \\
& + \sigma^4 Q_3 tr[Q_6' \{\gamma_2 (I_{T+1} \circ Q_4) + tr(Q_4) I_{T+1} + Q_4 + Q_4'\}] Q_5' \\
& + \sigma^2 tr(Q_6) Q_5 Q_1' + \sigma^3 \gamma_1 Q_5 \{(I_{T+1} \circ Q_6) i\}' Q_2' \\
& + \sigma^4 Q_5 tr[Q_4' \{\gamma_2 (I_{T+1} \circ Q_6) + tr(Q_6) I_{T+1} + Q_6 + Q_6')\}] Q_3' \\
& + \sigma^4 Q_5 tr[Q_6' \{\gamma_2 (I_{T+1} \circ Q_6) + tr(Q_6) I_{T+1} + Q_6 + Q_6'\}] Q_5'.
\end{aligned}
$$

A.4. Proof of Theorem 3

From Theorem 1 we have $E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2}(tr(H_i|_\delta J) - \delta' H_i|_\delta \delta) + o(T^{-1})$. From Lemma 2 we have $\delta = E[Q_1 + Q_2 v + Q_3 v' Q_4 v + Q_5 v' Q_6 v] = Q_1 + Q_3 tr(Q_4) + Q_5 tr(Q_6)$. The Hessian H_i was found in Theorem 2, and to evaluate it note that $E[Z'Z] = E[(\bar{Z} + G_b v \Gamma_4)'(\bar{Z} + G_b v \Gamma_4)]$ and $E[\zeta] = E[(\bar{Z} + G_b v \Gamma_4)'\{(\sum_{i=1}^p \lambda_i \bar{y}_{-i}) + X\beta + [(\sum_{i=1}^p \lambda_i G_i)\Gamma_3]v\}]$. These two expected values are then calculated in the same way as Lemma 2.

A.5. Proof of Corollary 2

(i) Specialising the matrices H_i and $H_i|_\delta$

Recall that $H_i = MB_i + B_i' M'$, or $H = 2BM'$ for the case at hand since H, M and B are all scalars. We also have $B = 1$, so that $H = M'$. In the matrix representation of M' we have the following for the AR(1) with known mean:

$$
\begin{aligned}
M' = {} & [e_1 \otimes (\hat{\zeta}' \otimes I_1)](Z'Z)^{\otimes(-4)}[(I_1 \otimes V_2) + (V_3 \otimes I_1)]e_1' \\
& - (e_2 \otimes I_1)(Z'Z)^{\otimes(-2)} e_1' - [((Z'Z)^{\otimes(-2)} e_1')' \otimes I_1](I_1 \otimes V_1) e_2'.
\end{aligned}
$$

It follows that

$$
H = e_1 \hat{\zeta}' (Z'Z)^{\otimes(-4)} (V_2 + V_3) e_1' - e_2 (Z'Z)^{\otimes(-2)} e_1' - e_1 (Z'Z)^{\otimes(-2)} V_1 e_2'.
$$

Noting the specialisations $V_1 = I_1$ and $V_2 = V_3 = vec(Z'Z)$, the expression for H reduces to

$$
H = 2 e_1 \hat{\zeta}' (Z'Z)^{\otimes(-4)} vec(Z'Z) e_1' - e_2 (Z'Z)^{\otimes(-2)} e_1' - e_1 (Z'Z)^{\otimes(-2)} e_2'.
$$

Finally, recall that $Z'Z$ and $\hat{\zeta}$ are scalars, so that

$$
H = \begin{pmatrix} 2\hat{\zeta}(Z'Z)^{-3} & -(Z'Z)^{-2} \\ -(Z'Z)^{-2} & 0 \end{pmatrix} \text{ and } H|_\delta = \begin{pmatrix} 2\zeta(E[Z'Z])^{-3} & -(E[Z'Z])^{-2} \\ -(E[Z'Z])^{-2} & 0 \end{pmatrix}.
$$

The expected values in $H|_\delta$ are $E[Z'Z] = \sigma^2 tr(G'G)$ and $\hat{\zeta} = \sigma^2\lambda tr(G'G)$ from Theorem 2, and from Lemma 3 the largest terms in each are $O(T)$. More specifically, we have $tr(G'G) = (\frac{T}{1-\lambda^2}) + (\frac{\omega}{1-\lambda^2}) - (\frac{1}{1-\lambda^2})^2 + \frac{\lambda^{2T}}{1-\lambda^2}(\frac{1}{1-\lambda^2} - \omega)$. This means that the non-zero elements in $H|_\delta$ are at most $O(T^{-2})$, though there will also be some smaller $o(T^{-2})$ contributions due to the $o(T)$ components of $E[Z'Z]$ and $\hat{\zeta}$. In both $H|_\delta$ and J we can discard contributions of order $O(\lambda^{sT})$ for $s > 0$ since there no (explosive) $O(\lambda^{-sT})$ contributions in either. Any products in $H|_\delta J$ involving $O(\lambda^{sT})$ terms will be $o(T^{-1})$. Therefore we use the approximation of $tr(G'G)$ up to order $O(1)$ that appears in Lemma 3, omitting the term $\frac{\lambda^{2T}}{1-\lambda^2}(\frac{1}{1-\lambda^2} - \omega)$. After simplifying $H|_\delta$ we have

$$H|_\delta = \tilde{H}|_\delta + O(\lambda^{2T}),$$
$$\text{where} \quad \tilde{H}|_\delta = \begin{pmatrix} \frac{2\lambda(1-\lambda^2)^2}{\sigma^4 T^2} & -\frac{(1-\lambda^2)^2}{\sigma^4 T^2} \\ -\frac{(1-\lambda^2)^2}{\sigma^4 T^2} & 0 \end{pmatrix}.$$

(i) Specialising the matrix J

In Lemma 4 we see that $Q_2 = 0$ in the pure AR(1) case, therefore all the terms in γ_1 here are zero. The terms in γ_2 are not all zero, and to consider their combined influence on the bias we can use the decomposition $J = J_1 + J_2$ where the terms in γ_2 are collected in J_2:

$$\begin{aligned} J_1 = {} & Q_1Q_1' + \sigma^2 tr(Q_4')Q_1Q_3' + \sigma^2 tr(Q_6')Q_1Q_5' \\ & + \sigma^2 tr(Q_4)Q_3Q_1' + \sigma^4 Q_3 tr[Q_4'\{tr(Q_4)I_{T+1} + Q_4 + Q_4'\}]Q_3' \\ & + \sigma^4 Q_3 tr[Q_6'\{tr(Q_4)I_{T+1} + Q_4 + Q_4'\}]Q_5' \\ & + \sigma^2 tr(Q_6)Q_5Q_1' + \sigma^4 Q_5 tr[Q_4'\{tr(Q_6)I_{T+1} + Q_6 + Q_6')\}]Q_3' \\ & + \sigma^4 Q_5 tr[Q_6'\{tr(Q_6)I_{T+1} + Q_6 + Q_6'\}]Q_5' \\ J_2 = {} & \sigma^4\gamma_2\{Q_3Q_3' tr[Q_4'(I_{T+1} \circ Q_4)] + Q_3Q_5' tr[Q_6'(I_{T+1} \circ Q_4)] \\ & + Q_5Q_3' tr[Q_4'(I_{T+1} \circ Q_6)] + Q_5Q_5' tr[Q_6'(I_{T+1} \circ Q_6)]\}. \end{aligned}$$

We have $Q_2 = 0$, $Q_1 = 0$, $Q_3 = e_1$, $Q_4 = A_4$, $Q_5 = e_2$ and $Q_6 = P_4$, which enables further specialisation:

$$\begin{aligned} J_1 = {} & \sigma^4[\{tr(A_4)\}^2 + 2tr(A_4'A_4)]e_1e_1' + \sigma^4\{tr(A_4)tr(P_4') + 2tr(P_4'A_4)\}e_1e_2' \\ & + \sigma^4\{tr(P_4)tr(A_4) + 2tr(A_4'P_4)\}e_2e_1' + \sigma^4[\{tr(P_4)\}^2 + tr(P_4'P_4) \\ & + tr(P_4'P_4')]e_2e_2' \\ J_2 = {} & \sigma^4\gamma_2\{tr[A_4'(I_{T+1} \circ A_4)]e_1e_1' + tr[P_4'(I_{T+1} \circ A_4)]e_1e_2' \\ & + tr[A_4'(I_{T+1} \circ P_4)]e_2e_1' + tr[P_4'(I_{T+1} \circ P_4)]e_2e_2'\}. \end{aligned}$$

The next task is to make this more explicit in terms of λ. Moreover, since we wish to discard all $o(T^{-1})$ terms from the product $H|_\delta J$, and since the largest terms in $H|_\delta$ are $O(T^{-2})$, we must discard all $o(T)$ terms from J. To do this, recall from Lemma 4 that $A_4 = G'G$ and $P_4 = G'(\lambda G + [0_{T\times 1} : I_T])$ and that we have approximations for the traces of products of these in Lemma 3. Let $\tilde{J}, \tilde{J}_1$ and $\tilde{J}_2$ denote, respectively, the versions of J, J_1, J_2 where $o(T)$ terms are excluded.

We have the following for the first term in $\tilde{J}_1$:

$$\begin{aligned}
\{tr(A_4)\}^2 + 2tr(A_4'A_4) &= \{tr(G'G)\}^2 + 2tr(G'GG'G) \\
&= \{T(\frac{1}{1-\lambda^2})\}^2 + 2\{\frac{\omega}{1-\lambda^2} - (\frac{1}{1-\lambda^2})^2\}(\frac{T}{1-\lambda^2}) \\
&\quad + 2T(\lambda^2+1)(\frac{1}{1-\lambda^2})^3 + o(T) \\
&= \frac{T(2+T-(T-2)\lambda^2}{(1-\lambda^2)^3} + o(T)
\end{aligned}$$

We have the following for the second and third terms in $\tilde{J}_1$:

$$\begin{aligned}
tr(A_4)tr(P_4') + 2tr(P_4'A_4) &= tr(G'G)tr((\lambda G + [0_{T\times 1} : I_T])'G) \\
&\quad + 2tr((\lambda G + [0_{T\times 1} : I_T])'GG'G) \\
&\quad + tr\{(\omega F : C)(\omega F : C)'(\omega F : C)[0_{T\times 1} : I_T]')\}] \\
&= \lambda\{tr(G'G)\}^2 + 2[\lambda tr(G'GG'G) \\
&\quad + tr\{(\omega F : C)\begin{pmatrix} \omega^2 F'F & \omega F'C \\ \omega C'F & C'C \end{pmatrix}\begin{pmatrix} 0_{1\times T} \\ I_T \end{pmatrix}\}] \\
&= \lambda(\frac{T}{1-\lambda^2})^2 + 2\lambda\{\frac{\omega}{1-\lambda^2} - (\frac{1}{1-\lambda^2})^2\}(\frac{T}{1-\lambda^2}) \\
&\quad + 2T(\lambda^2+1)(\frac{1}{1-\lambda^2})^3 + T\lambda(\frac{1}{1-\lambda^2})^2\} + o(T) \\
&= \frac{T\lambda(4+T(1-\lambda^2))}{(1-\lambda^2)^3} + o(T)
\end{aligned}$$

For the fourth term in $\tilde{J}_1$ we have

$$\{tr(P_4)\}^2 + tr(P_4'P_4) + tr(P_4'P_4') = \lambda^2\{tr(G'G)\}^2 + tr(P_4'P_4) + tr(P_4'P_4').$$

We consider $tr(P_4'P_4)$ and $tr(P_4'P_4')$ individually now:

$$\begin{aligned}
tr(P_4'P_4) &= tr\{(\lambda G + [0_{T\times 1} : I_T])'GG'(\lambda G + [0_{T\times 1} : I_T])\} \\
&= \lambda^2 tr(G'GG'G) + \lambda tr([0_{T\times 1} : I_T]'GG'G) + \lambda tr([0_{T\times 1} : I_T]'GG'G) \\
&\quad + tr\{[0_{T\times 1} : I_T]'(\omega F : C)(\omega F : C)'[0_{T\times 1} : I_T]\} \\
&= \lambda^2 T(\lambda^2+1)(\frac{1}{1-\lambda^2})^3 + 2\lambda(T\lambda(\frac{1}{1-\lambda^2})^2) + T(\frac{1}{1-\lambda^2}) + o(T)
\end{aligned}$$

$$
\begin{aligned}
tr(P_4'P_4') &= tr\{(\lambda G + [0_{T\times 1} : I_T])'G(\lambda G + [0_{T\times 1} : I_T])'G\} \\
&= \lambda^2 tr(G'GG'G) + 2\lambda tr\{[0_{T\times 1} : I_T]'GG'G) \\
&\quad + tr([0_{T\times 1} : I_T]'(\omega F : C)[0_{T\times 1} : I_T]'(\omega F : C)\} \\
&= \lambda^2 T(\lambda^2+1)(\frac{1}{1-\lambda^2})^3 + 2\lambda\{T\lambda(\frac{1}{1-\lambda^2})^2\} + o(T).
\end{aligned}
$$

This gives

$$
\{tr(P_4)\}^2 + tr(P_4'P_4) + tr(P_4'P_4') = \frac{T\{(T+4)\lambda^2 - (T+1)\lambda^4 + 1\})}{(1-\lambda^2)^3},
$$

so that the final form of $\tilde{J}_1$ is as follows:

$$
\begin{aligned}
J_1 = \sigma^4 &\left\{\frac{T(2+T-(T-2)\lambda^2}{(1-\lambda^2)^3}\right\} e_1 e_1' \\
+ \sigma^4 &\left\{\frac{T\lambda(4+T(1-\lambda^2))}{(1-\lambda^2)^3}\right\} e_1 e_2' \\
+ \sigma^4 &\left\{\frac{T\lambda(4+T(1-\lambda^2))}{(1-\lambda^2)^3}\right\} e_2 e_1' \\
+ \sigma^4 &\left\{\frac{T\{(T+4)\lambda^2 - (T+1)\lambda^4 + 1\})}{(1-\lambda^2)^3}\right\} e_2 e_2' + o(T).
\end{aligned}
$$

We can do a similar specialisation for J_2, using $tr[Q_6(I_{T+1} \circ Q_4)] = \lambda \sum_{i=1}^{T+1} (G'G)_{ii}^2$ and $tr[Q_4'(I_{T+1} \circ Q_4)] = \sum_{i=1}^{T+1} (G'G)_{ii}^2$. We have

$$
J_2 = \gamma_2 \sigma^4 \begin{pmatrix} \sum_{i=1}^{T+1} (G'G)_{ii}^2 & \lambda \sum_{i=1}^{T+1} (G'G)_{ii}^2 \\ \lambda \sum_{i=1}^{T+1} (G'G)_{ii}^2 & tr[P_4'(I_{T+1} \circ P_4)] \end{pmatrix}.
$$

Here it seems unnecessary to find $\tilde{J}_2$, since it can already be seen (below) that $tr(\tilde{H}|_\delta J_2) = 0$ to order $O(T^{-1})$.

(iii) Finding the bias result

The relevant elements of $\tilde{H}|_\delta J_2$ are

$$
\begin{aligned}
(\tilde{H}_i|_\delta J_2)_{11} &= \frac{2\lambda(1-\lambda^2)\sum_{i=1}^{T+1}(G'G)_{ii}}{T^2\sigma^4} - \frac{\lambda(1-\lambda^2)\sum_{i=1}^{T+1}(G'G)_{ii}}{T^2\sigma^4} \\
&\quad + o(T^{-1})
\end{aligned}
$$

$$
\text{and} \quad (\tilde{H}_i|_\delta J_2)_{22} = -\frac{\lambda(1-\lambda^2)\sum_{i=1}^{T+1}(G'G)_{ii}}{T^2\sigma^4} + o(T^{-1}),
$$

showing that $tr(\tilde{H}|_\delta J_2)$ is $o(T^{-1})$. The $o(T^{-1})$ parts in these are $O(\lambda^{2T})$.

To complete the proof, we see that $tr(\tilde{H}|_\delta \tilde{J}_1)$ simplifies to $\frac{-2\lambda}{T}$ and, using $\delta = \sigma^2 tr(G'G)e_1 + \sigma^2 \lambda tr(G'G)e_2$ from Theorem 3, it is straightforward to show that $\delta' \tilde{H}|_\delta \delta$ is zero.

References

Bao, Y. (2007). The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. *Econometric Theory* **23(05)**, 1013–1021.

Bao, Y. & Ullah, A. (2002). The second-order bias and mean-squared error of nonlinear estimators in time-series. Manuscript, University of California, Riverside.

Bao, Y. & Ullah, A. (2007). The second-order bias and mean-squared error of estimators in time-series models. *Journal of Econometrics* **140(2)**, 650–669.

Bhansali, R. (1981). Effects of not knowing the order of an autoregressive process on the mean squared error of prediction-i. *Journal of the Americal Statistical Association* **76**, 588–597.

Kadane, J. (1971). Comparison of k-class estimators when the disturbances are small. *Econometrica* **39**, 723–737.

Kendall, M. (1954). Note on bias in the estimation of autocorrelation. *Biometrika* **61**, 403–404.

Kiviet, J. & Phillips, G. (1993). Alternative bias approximations in regressions with a lagged dependent variable. *Econometric Theory* **9**, 62–80.

Kiviet, J. & Phillips, G. (1994). Bias assessment and reduction in linear error-correction models. *Journal of Econometrics* **63**, 215–243.

Kiviet, J. & Phillips, G. (1996). The bias of the ordinary least squares estimator in simultaneous equation models. *Economics Letters* **53**, 161–167.

Kiviet, J. & Phillips, G. (2005). Moment approximation for least-squares estimators in dynamic regression models with a unit root. *Econometrics Journal* **8**, 1–28.

Kiviet, J. & Phillips, G. (2010). Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. Forthcoming at Computational Statistics and Data Analysis, 6th Special Issue on Computational Econometrics.

Kunitomo, N. & Yamamoto, T. (1985). Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association* **80**, 941–950.

Maekawa, K. (1983). An approximation to the distribution of the least squares estimator in an autoregressive model with exogenous variables. *Econometrica* **51**, 229–238.

Magnus, J. & Neudecker, H. (1979). The commutation matrix: some properties and applications. *The Annals of Statistics* **7(2)**, 318–394.

Magnus, J. & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics (Revised 2002 edition)*. Wiley, 2nd ed.

Marriott, F. & Pope, J. (1954). Bias in the estimation of autocorrelations. *Biometrika* **61**, 393–403.

Nagar, A. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* **27**, 575–595.

Phillips, G. (2000). An alternative approach to obtaining nagar-type moment approximations in simultaneous equation models. *Journal of Econometrics* **97(2)**, 345–364.

Phillips, G. (2007). Nagar-type moment approximations in simultaneous equation models: some further results. In *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, G. Phillips, ed. Cambridge University Press, pp. 60–99.

Phillips, G. & Liu-Evans, G. (2010). The robustness of the 2SLS moment approximations to non-normal disturbances. An earlier version was presented at the Econometric Society European Meeting 2008.

Rilestone, P., Srivastava, S. & Ullah, A. (1996). The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* **75**, 369–395.

Sargan, J. (1974). The validity of nagar's expansion for the moments of econometric estimators. *Econometrica* **42**, 169–176.

Shaman, P. & Stine, R. (1988). The bias of autoregressive coefficient estimators. *Journal of The American Statistical Association* **83**, 842–848.

Shao, J. (1988). On resampling methods for variance and bias estimation in linear models. *Annals of Statistics* **16**, 986–1008.

Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap.* 175th Avenue, New York: Springer-Verlag, 1st ed.

Tanaka, K. (1984). Asymptotic expansions associated with the AR(1) model with unknown mean. *Econometrica* **51**, 1221–1231.

Tjostheim, D. & Paulsen, J. (1983). Bias of some commonly-used time series estimates. *Biometrika* **70**, 389–399.

Ullah, A. (2005). *Finite Sample Econometrics.* Oxford University Press.

Yamamoto, T. & Kunitomo, N. (1984). Asymptotic bias of the least squares estimator for multivariate autoregressive models. *Annals of the Institute of Statistical Mathematics* **36**, 419–430.