Submitted to *Econometrics*. Pages 1 - 21.

OPEN ACCESS

Article

An alternative approach to approximating moments of least squares estimators

Gareth Liu-Evans

Management School, University of Liverpool, Chatham Street, Liverpool, United Kingdom

Version April 19, 2015 submitted to Econometrics. *Typeset by LaTeX using class file mdpi.cls*

Abstract: Results are presented for approximating the moments of least squares estimators, particularly those of the OLS estimator, and the methodology is illustrated using a simple dynamic model. An $O(T^{-1})$ approximation is presented for the bias in estimation of a general ARX(p) model, and this is specialised to known cases for the AR(1) with and without constant under an assumption of non normal model disturbances.

Keywords: asymptotic approximation; bias; least squares ; time series ; autoregression ; simulteneity

1. Introduction

We present results to facilitate the asymptotic approximation of the moments of least squares coefficient estimators under similar assumptions to [1], but focussing on the OLS estimator. The expansion method in [1] is valid for consistent k-class estimation of equations from static simultaneous equation systems, but the methodology does not apply to OLS or dynamic models. We instead use the validity framework in [2], and results are presented here to aid the direct use of the approach with autoregressive models as an illustration.

This is done using matrix differential calculus results from [3], [4], and [5,6]. The approach builds on [7] and [8], who consider the moments of the k-th order serial correlation coefficient in AR(1) models. It is shown how the approach is applicable in principle to models where both dynamics and simulteneity are present, as an alternative to methodology based on [9].

There are large-T asymptotic approximations of the OLS estimator moments using [9]-type methodologies in [10,11] and [12] in the context of stationary autoregressive models, and in [13] for equations from static simultaneous equation models. Building on [14], [15] present a general method for estimators of time-series models. See also [16,17] and [18].

2. The expansion method

Given a model

$$y = Z\alpha + u \tag{1}$$

with $E[Z'u] = 0$ (this assumption is dropped later), and with Z being a $T \times N$ matrix with rank N almost surely, the true coefficient α and its OLS estimator $\hat{\alpha}$ can be expressed in the same functional form so long as the expected values in (2) exist:

$$\hat{\alpha} = (Z'Z)^{-1}Z'y \ \text{ and } \ \alpha = (E[Z'Z])^{-1}E[Z'y]. \tag{2}$$

Equation (3) is obtained by premultiplying both sides of (1) by Z'. By defining 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 be expressed as

$$\hat{\alpha}_i = f_i(\hat{\delta}) \ \text{ and } \ \alpha_i = f_i(\delta), \tag{3}$$

respectively, for $i = 1, \ldots, N$, where $\hat{\delta} = vec(\hat{R})$ and $\delta = vec(R)$. This allows a Taylor series expansion of the following form:

$$\hat{\alpha}_i - \alpha_i = (\hat{\delta} - \delta)' f_i'(\delta) + \frac{1}{2}(\hat{\delta} - \delta)' H_i|_\delta (\hat{\delta} - \delta) + .., \tag{4}$$

where $H_i|_\delta$ is the Hessian matrix of f evaluated at $\delta = vec(R)$.

Using the extended mean value theorem we may write

$$\begin{aligned} f_i(\hat{\delta}) = f_i(\delta) + (\hat{\delta} - \delta)' f_i'(\delta) + \frac{1}{2}(\hat{\delta} - \delta)' H_i|_\delta (\hat{\delta} - \delta) \\ + \frac{1}{3!}\sum_{j=1}^{r}(\hat{\delta}_j - \delta_j)(\hat{\delta} - \delta)' f_{ij}^{(3)}|_{\delta^\star}(\hat{\delta} - \delta) \end{aligned} \tag{5}$$

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 and that the derivatives are uniformly bounded in a neighbourhood of δ as $T \to \infty$, and with third-order derivatives that are continuous. This makes the fourth term $O_p(T^{-\frac{3}{2}})$. If the components of $\hat{\delta}$ are assumed to have finite moments up to third order we therefore have

$$E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2} tr(H_i|_\delta Var(\hat{\delta})) + o(T^{-1}), \tag{6}$$

where $Var(\hat{\delta})$ is the covariance matrix for $\hat{\delta}$. This final step is clear from [19] section 2.4 and [20], see also [2], Appendix A. Rearranging this 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 $J = E[\hat{\delta}\hat{\delta}']$.

Writing $\Gamma_1 = [0_{N\times N^2} : I_N]$, $\Gamma_2 = [I_{N^2} : 0_{N^2\times N}]$, $V_1 = vec(I_N)$, $V_2 = (K_{NN} \otimes I_N)(I_N \otimes vec(Z'Z))$ and $V_3 = (I_N \otimes K_{NN})(vec(Z'Z) \otimes I_N)$, where K_{nm} is an $nm \times nm$ commutation matrix and I_N is an $N \times N$ identity matrix, Theorem 2 below provides the Hessian matrix H_i in terms of $\hat{\zeta}$ and $Z'Z$, using the Second Identification Theorem in [6]. 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. In the following we write $A \otimes A \otimes \cdots \otimes A = A^{\otimes m}$, where $\otimes$ is the Kronecker product and A appears m times. Since $(A \otimes A \otimes \cdots \otimes A)^{-1} = A^{-1} \otimes A^{-1} \otimes \cdots \otimes A^{-1}$ when A is invertible, we denote this by $A^{\otimes(-m)}$.

Theorem 2. *The Hessian matrix is*

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

where

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

and where B_i is defined as

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

Proof. *See Appendix A.*

It may be possible to extend the methodology here to cases where $E[Z'u] \neq 0$, i.e. where endogeneity is present. We have the following by premultiplying $y = Z\alpha + u$ by Z' and taking the expected value as above:

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

where $E_1 = E[Z'Z]$ and $E_2 = E[Z'u]$, and where e_i is an $N \times 1$ unit vector with unity in position i. 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

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

and the approximation is valid to order $O(T^{-1})$ under the same conditions as before.

The following sections apply Theorem 1 to the ARX(p) with non-normal disturbances. As in [15] and [12] 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.

3. ARX(p) illustration

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 [11][1]:

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

[1] See also [21].

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 $y_{-i} = M_i Y_{-i}$ 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

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

where $G_i = M_i G$. When $p = 1$, M_i is the identity matrix and G is the same G that appears in [11].

Following [11] we write $Z = \bar{Z} + \tilde{Z}$, where $\bar{Z} = (\bar{y}_{-1}, ..., \bar{y}_{-p}, X)$ and $\tilde{Z} = (G_1 v, ..., G_p v, 0_{T\times k})$. We have $\tilde{Z} = (G_1 v, ..., G_p v)\Gamma_4$ where $\Gamma_4 = [I_p : 0_{p\times k}]$, which gives

$$Z = \bar{Z} + \sum_{i=1}^{p} G_i v e'_i \Gamma_4,$$

where e_i here is a $p \times 1$ vector with 1 in position i and 0 elsewhere. When $p = 1$, the result reduces to $Z = \bar{Z} + Gve'_1$, which appears in [11].

We may write $\hat{\delta}$ as a linear combination of v and quadratic forms in v, so that the expected value of $\hat{\delta}\hat{\delta}'$ can be calculated using existing results on the expected values of products of quadratic forms.[2] We note that $\hat{\delta} = \Gamma'_2 vec(Z'Z) + \Gamma'_1 vec(\hat{\zeta})$, where Γ_1 and Γ_2 are the same as in the preceding section with $N = p + k$, and then express $vec(Z'Z)$ and $vec(\hat{\zeta})$ in terms of v.

We may write y in terms of v:

---

[2] See e.g. the Appendix in [22].

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

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*

$$
\begin{aligned}
vec(\hat{\zeta}) &= P_1 + P_2 v + \sum_{i=1}^{p} P_{3i} v' P_{4i} v \\
vec(Z'Z) &= A_1 + A_2 v + \sum_{i,j=1}^{p} A_{3ij} v' A_{4ij} v,
\end{aligned}
$$

where

$$
\begin{aligned}
P_1 &= vec[\bar{Z}'\{(\bar{Z} \sum_{i=1}^{p} \lambda_i \bar{y}_{-i}) + X\beta\}] \\
P_2 &= \bar{Z}'\{(\sum_{i=1}^{p} \lambda_i G_i) + \Gamma_3\} + \sum_{i=1}^{p} ([\{(\sum_{j=1}^{p} \lambda_j \bar{y}_{-j}) + X\beta\}' G_i] \otimes \Gamma_4' e_i) \\
P_{3i} &= vec(\Gamma_4' e_i), \quad P_{4i} = G_i'\{(\sum_{j=1}^{p} \lambda_j G_j) + \Gamma_3\}, \quad A_1 = vec(\bar{Z}'\bar{Z}) \\
A_2 &= \sum_{i=1}^{p} \{(\Gamma_4' e_i \otimes \bar{Z}' G_i) + (\bar{Z}' G_i \otimes \Gamma_4' e_i)\}, \quad A_{3ij} = \Gamma_4' e_j \otimes \Gamma_4' e_i, \quad A_{4ij} = G_i' G_j
\end{aligned}
$$

Proof. *See Appendix B.1.*

Using Lemma 1, it is possible to calculate the expected value J, and this is provided in Lemma 2.

Lemma 2. *The expected value J is*

$$J = Q_1Q_1' + \sigma^2 \sum_{i,j=1}^{p} tr(Q_{4ij}')Q_1Q_{3ij}' + \sigma^2 \sum_{i=1}^{p} tr(Q_{6i}')Q_1Q_{5i}' + \sigma^2 Q_2Q_2'$$
$$+ \sigma^2 \sum_{i,j=1}^{p} tr(Q_{4ij})Q_{3ij}Q_1' + \sigma^4 \sum_{i,j,k,l=1}^{p} Q_{3ij} tr[Q_{4lm}'\{tr(Q_{4ij})I_{T+1} + Q_{4ij} + Q_{4ij}'\}]Q_{3lm}'$$
$$+ \sigma^4 \sum_{i,j,l=1}^{p} Q_{3ij} tr[Q_{6l}'\{tr(Q_{4ij})I_{T+1} + Q_{ij4} + Q_{4ij}'\}]Q_{5l}'$$
$$+ \sigma^2 \sum_{i=1}^{p} tr(Qi_6)Q_{5i}Q_1' + \sigma^4 \sum_{i,j,l=1}^{p} Q_{5i} tr[Q_{4jl}'\{tr(Q_{6i}I_{T+1} + Q_{6i} + Q_{6i}')\}]Q_{3jl}'$$
$$+ \sigma^4 \sum_{i,j=1}^{p} Q_{5i} tr[Q_{6j}'\{tr(Q_{6j})I_{T+1} + Q_{6j} + Q_{6j}'\}]Q_{5j}'$$
$$+ \gamma_1\sigma^3\{Q_2 \sum_{i,j=1}^{p} (I_{T+1} \circ (Q_{4ij}'))iQ_{3ij}' + Q_2 \sum_{i=1}^{p} (I_{T+1} \circ (Q_{6i}'))iQ_{5i}'$$
$$+ \sum_{i,j=1}^{p} Q_{3ij}\{(I_{T+1} \circ Q_{4ij}')i\}'Q_2' + \sum_{i=1}^{p} Q_{5i}\{(I_{T+1} \circ Q_{6i})i\}'Q_2'\}$$
$$+ \gamma_2\sigma^4\{ \sum_{i,j,l,m=1}^{p} Q_{3ij}Q_{3lm}' tr[Q_{4lm}'(I_{T+1} \circ Q_{4ij})] + \sum_{i,j,l=1}^{p} Q_{3ij}Q_{5l}' tr[Q_6'(I_{T+1} \circ Q_{4ij})]$$
$$+ \sum_{i,j,l=1}^{p} Q_{5i}Q_{3jl}' tr[Q_{4jl}'(I_{T+1} \circ Q_{6i})] + \sum_{i,j=1}^{p} Q_{5i}Q_{5j}' tr[Q_{6j}'(I_{T+1} \circ Q_{6i})]\}.$$

where "$\circ$" is the Hadamard matrix product, and where $Q_1 = \Gamma_2' A_1 + \Gamma_1' P_1$, $Q_2 = \Gamma_2' A_2 + \Gamma_1' P_2$, $Q_{3ij} = \Gamma_2' A_{3ij}$, $Q_{4ij} = A_{4ij}$, $Q_{5i} = \Gamma_1' P_{3i}$ and $Q_{6i} = P_{4i}$.

Proof. See Appendix B.2.

The OLS bias is then given by

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, $\delta = Q_1 + \sigma^2 \sum_{i,j=1}^{p} Q_{3ij} tr(Q_{4ij}) + \sigma^2 \sum_{i=1}^{p} Q_{5i} tr(Q_{6i})$, and where the Hessian is evaluated at δ, or, equivalently, at $E[Z'Z] = \bar{Z}'\bar{Z} + \sigma^2 \sum_{i,j=1}^{p} \Gamma_4' e_i e_j' \Gamma_4 tr(G_i'G_j)$ and at $E[\zeta] = P_1 + \sigma^2 \sum_{i=1}^{p} P_{3i} tr(P_{4i})$.

Proof. See Appendix B.3.

The approximation in Theorem 3 is 'unfiltered', to use the terminology of [11], by which we mean there are likely to be unnecessary $o(T^{-1})$ terms in the approximation that could in principle be removed. Doing this was found to be difficult, and would for example require a generalisation of the results in Lemma 2 below for traces of products of the matrices G_i, $i = 1, \ldots, p$. It may well be, as found in [15] for the ARX(1), that no kurtosis coefficient γ_2 need enter the approximation to order $O(T^{-1})$. In what follows we specialise Theorem 3 to AR(1) cases with and without constant, remove the $o(T^{-1})$ terms, and agree with the results in [15] for these cases.

3.1. AR(1) models

We drop the subscript in G_1 for notational convenience in the following: since $p = 1$, the matrices $G_2, \ldots, G_p$ are no longer used. Moreover, the form of the expected value J is the same as in Lemma 2, but we may remove the redundant summation over indices i, j, l and m. Some of the matrices used above for the ARX(p) adjust straightforwardly when $p = 1$, for example

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

and the matrices B_i are as in Theorem 2 with $N = k+1$. Some specialisations for others are collected in Corollary 1 below.

Corollary 1. *The matrices $P_1, ..., P_4$ and $A_1, ..., A_4$ in Lemma 1 specialise to the following when $p = 1$:*

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

3.1.1. AR(1) with known mean

[8] and [7] 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. [15] 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 [11], the matix G 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. *(See [15] and [11])*

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

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 B.4.

This agrees with the [15] generalisation of the original [8] and [7] 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:

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

3.1.2. AR(1) with unknown mean

A similar reduction can be done for the case where $p = 1$ and the model includes a constant, see [8] and [7] for the original approximation under normality. Unlike in the case with known mean, it cannot be seen immediately that the four γ_1 terms in Lemma 2 sum to zero. In order to find this, it is necessary to use the following results

Lemma 5. *(See [11] and [15], respectively)*

$$i_t' F F' F = (\frac{1}{1-\lambda})(\frac{1}{1-\lambda^2}) + o(T), \quad i_T' C(I_T \circ C'C) i_T = \frac{T}{(1-\lambda^2)(1-\lambda)} + o(T)$$

Corollary 3. *The bias in OLS estimation of the autoregressive coefficient in the AR(1)+constant model, with γ_1 and γ_2 taking any finite value, is*

$$E[\hat{\lambda} - \lambda] = -\frac{1+3\lambda}{T} + o(T^{-1}).$$

Proof. See Appendix B.5.

Here we confirm that having skewness or kurtosis in the error terms does not affect the bias in these cases to order $O(T^{-1})$, see [15]. Note that the results here assume a random covariance-stationary startup, while [15] and [12] assume a fixed startup. [11] 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 there is still no effect to order $O(T^{-1})$ when the 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 [23] and ending with [24] also shows this, and shows the same for γ_2, though the validity conditions for these approximations include finite error moments up to the 12th and 16th orders for the known and unknown mean cases, respectively.

4. Conclusion

The results here facilitate the direct application of the validity framework outlined in [2]. The approach is applicable in principle to models with both dynamics and endogeneity. The methodology has been used to obtain an order $O(T^{-1})$ bias expression for the OLS coefficient estimator of a general ARX(p) model with non-normal disturbances. Moreover, we have confirmed the non normality results in [15] for the AR(1) with and without constant. It may be possible in the future to extend the methodology to other estimators such as 2SLS.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Phillips, G. An Alternative Approach to Obtaining Nagar-Type moment Approximations in Simultaneous Equation models. *Journal of Econometrics* **2000**, *97(2)*, 345–364.
2. Kiviet, J.; Phillips, G. Improved variance estimation of maximum likelihood estimators in stable first-order dynamic regression models. *Computational Statistics and Data Analysis* **2014**, *76*, 424–448.
3. Neudecker, H. Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products. *Journal of the American Statistical Association* **1969**, *64*, pp. 953–963.
4. Neudecker, H.; Wansbeek, T. Some Results on Commutation Matrices, with Statistical Applications. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **1983**, *11*, pp. 221–231.
5. Magnus, J.; Neudecker, H. The commutation matrix: some properties and applications. *The Annals of Statistics* **1979**, *7(2)*, 318–394.
6. Magnus, J.; Neudecker, H. *Matrix differential calculus with applications in statistics and econometrics (Revised edition)*; Wiley, 2002.
7. Marriott, F.; Pope, J. Bias in the estimation of autocorrelations. *Biometrika* **1954**, *61*, 393–403.
8. Kendall, M. Note on bias in the estimation of autocorrelation. *Biometrika* **1954**, *61*, 403–404.
9. Nagar, A. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* **1959**, *27*, 575–595.
10. Kiviet, J.; Phillips, G. Moment approximation for least-squares estimators in dynamic regression models with a unit root. *Econometrics Journal* **2005**, *8*, 1–28.
11. Kiviet, J.; Phillips, G. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. *Computational Statistics and Data Analysis* **2012**, *56(11)*, 3706–3729.
12. Bao, Y. The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. *Econometric Theory* **2007**, *23(05)*, 1013–1021.
13. Kiviet, J.; Phillips, G. The bias of the ordinary least squares estimator in simultaneous equation models. *Economics Letters* **1996**, *53*, 161–167.
14. Rilestone, P.; Srivastava, S.; Ullah, A. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* **1996**, *75*, 369–395.
15. Bao, Y.; Ullah, A. The second-order bias and mean-squared error of estimators in time-series models. *Journal of Econometrics* **2007**, *140(2)*, 650–669.
16. Sargan, J. The validity of Nagar's expansion for the moments of econometric estimators. *Econometrica* **1974**, *42*, 169–176.
17. Sargan, J. Econometric estimators and the Edgeworth expansion. *Econometrica* **1976**, *44*, 421–448.
18. Phillips, P. Approximations to some finite sample distributions associated with a first-order stochastic difference equation. *Econometrica* **1977**, *45*, 463–485.
19. Shao, J.; Tu, D. *The Jackknife and Bootstrap*, first ed.; Springer-Verlag: 175th Avenue, New York, 1995.

20. Shao, J. On resampling methods for variance and bias estimation in linear models. *Annals of Statistics* **1988**, *16*, 986–1008.
21. Kiviet, J.; Phillips, G. Bias assessment and reduction in linear error-correction models. *Journal of Econometrics* **1994**, *63*, 215–243.
22. Ullah, A. *Finite Sample Econometrics*; Oxford University Press, 2004.
23. Bhansali, R. Effects of not knowing the order of an autoregressive process on the mean squared error of prediction-I. *Journal of the Americal Statistical Association* **1981**, *76*, 588–597.
24. Shaman, P.; Stine, R. The bias of autoregressive coefficient estimators. *Journal of The American Statistical Association* **1988**, *83*, 842–848.

A. Theorem 2

A.1. Derivation of H_i

H_i can be found using the Second Identification Theorem in [6][3]. 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 $vec(Z'Z) = \Gamma_2\hat{\delta}$. 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

---

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

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

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

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

for some M, then the Hessian can be identified as $H_i = \frac{1}{2}\left(MB_i + B_i' M'\right)$. 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

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

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:

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

where, following the result in the exercise on p48 of [6], $V_1 = (K_{21} \otimes I_2)(I_1 \otimes vec(I_N)) = (K_{N1} \otimes I_N) vec(I_N) = vec(I_N)$, $V_2 = (K_{NN} \otimes I_N)(I_N \otimes vec(Z'Z))$ and $V_3 = (I_N \otimes K_{NN})(vec(Z'Z) \otimes I_N)$.

From the above we have

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

therefore $d^2\hat{\alpha}_i = (d\hat{\delta})' M B_i (d\hat{\delta})$ and the Hessian is

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

A.2. 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 $q = 1, ..., N^2 + N - 1$.

B. ARX models

B.1. Lemma 1

We have

$$\begin{aligned}
vec(\hat{\zeta}) &= vec(Z'y) \\
&= vec[(\bar{Z} + \sum_{i=1}^{p} G_i v e_i \Gamma_4)'\{(\sum_{j=1}^{p} \lambda_j \bar{y}_{-j}) + X\beta + \{(\sum_{j=1}^{p} \lambda_j G_j) + \Gamma_3\}v\}] \\
&= vec[\bar{Z}'\{(\sum_{j=1}^{p} \lambda_j \bar{y}_{-j}) + X\beta\}] + \{\bar{Z}'\{(\sum_{j=1}^{p} \lambda_j G_j) + \Gamma_3\} \\
&\quad + \sum_{i=1}^{p}([\{(\sum_{j=1}^{p} \lambda_j \bar{y}_{-j}) + X\beta\}' G_i] \otimes \Gamma_4' e_i)\}v + \sum_{i=1}^{p} \Gamma_4' e_i v' G_i' \{(\sum_{j=1}^{p} \lambda_j G_j) + \Gamma_3\}v \\
&= P_1 + P_2 v + P_{3i} v' P_{4i} v.
\end{aligned}$$

Similarly,

$$\begin{aligned}
vec(Z'Z) &= vec[(\bar{Z} + \sum_{i=1}^{p} G_i v e_i' \Gamma_4)'(\bar{Z} + \sum_{i=1}^{p} G_i v e_i \Gamma_4)] \\
&= vec(Z'Z) + \sum_{i=1}^{p}\{(\Gamma_4' e_i \otimes \bar{Z}' G_i) + (\bar{Z}' G_i \otimes \Gamma_4' e_i)\}v + \sum_{i,j=1}^{p}(\Gamma_4' e_j \otimes \Gamma_4' e_i) v' G_i' G_j v \\
&= A_1 + A_2 v + \sum_{i,j=1}^{p} A_{3ij} v' A_{4ij} v.
\end{aligned}$$

B.2. Lemma 2

Since $\hat{\delta} = \Gamma_2' vec(Z'Z) + \Gamma_1' vec\hat{\zeta}$ we have $\hat{\delta} = Q_1 + Q_2 v + \sum_{i,j=1}^{p} Q_{3ij} v' Q_{4ij} v + Q_{5i} v' Q_{6i} v$ from Lemma 1 where $Q_1 = \Gamma_2' A_1 + \Gamma_1' P_1$, $Q_2 = \Gamma_2' A_2 + \Gamma_1' P_2$, $Q_{3ij} = \Gamma_2' A_{3ij}$, $Q_{4ij} = A_{4ij}$ and $Q_{5i} = \Gamma_1' P_{3i}$, $Q_{6i} = P_{4i}$. To calculate J we can write

$$\begin{aligned}
\hat{\delta}\hat{\delta}' = (Q_1 + Q_2 v + \sum_{i,j=1}^{p} Q_{3ij} v' Q_{4ij} v + \sum_{i=1}^{p} Q_{5i} v' Q_{6i} v) \\
\times (Q_1 + Q_2 v + \sum_{i,j=1}^{p} Q_{3ij} v' Q_{4ij} v + \sum_{i=1}^{p} Q_{i5} v' Q_{6i} v)',
\end{aligned}$$

then after eliminating terms that have zero expectation we have

$$E[\hat{\delta}\hat{\delta}'] = Q_1Q_1' + Q_1\sum_{i,j=1}^{p} E[v'Q_4'v]Q_{3ij}' + Q_1\sum_{i=1}^{p} E[v'Q_{6i}'v]Q_{5i}' + Q_2E[vv']Q_2'$$
$$+ Q_2\sum_{i,j=1}^{p} E[vv'Q_{4ij}'v]Q_{3ij}' + Q_2\sum_{i=1}^{p} E[vv'Q_{6i}'v]Q_{5i}' + \sum_{i,j=1}^{p} Q_{3ij}E[v'Q_{4ij}v]Q_1'$$
$$+ \sum_{i,j=1}^{p} Q_{3ij}E[v'Q_{4ij}vv']Q_2' + \sum_{i,j,k,l=1}^{p} Q_{3ij}E[v'Q_{4ij}vv'Q_{4lm}'v]Q_{3lm}' + \sum_{i,j,l=1}^{p} Q_{3ij}E[v'Q_{4ij}vv'Q_{6l}'v]Q_{5l}'$$
$$+ \sum_{i=1}^{p} Q_{5i}E[v'Q_{6i}v]Q_1' + \sum_{i=1}^{p} Q_{5i}E[v'Q_{6i}vv']Q_2' + \sum_{i,j,l=1}^{p} Q_{5i}E[v'Q_{6i}vv'Q_{4jl}'v]Q_{3jl}'$$
$$+ \sum_{i,j=1}^{p} Q_{5i}E[v'Q_{6i}vv'Q_{6j}'v]Q_{5j}'.$$

In order to calculate these expected values, the results in Appendix A.5 of [22] are used. Doing this gives

$$J = Q_1Q_1' + \sigma^2\sum_{i,j=1}^{p} tr(Q_{4ij}')Q_1Q_{3ij}' + \sigma^2\sum_{i=1}^{p} tr(Q_{6i}')Q_1Q_{5i}' + \sigma^2Q_2Q_2'$$
$$+ \sigma^3\gamma_1Q_2\sum_{i,j=1}^{p}(I_{T+1}\circ(Q_{4ij}'))iQ_{3ij}' + \sigma^3\gamma_1Q_2\sum_{i=1}^{p}(I_{T+1}\circ(Q_{6i}'))iQ_{5i}'$$
$$+ \sigma^2\sum_{i,j=1}^{p} tr(Q_{4ij})Q_{3ij}Q_1' + \sigma^3\gamma_1\sum_{i,j=1}^{p} Q_{3ij}\{(I_{T+1}\circ Q_{4ij}')i\}'Q_2'$$
$$+ \sigma^4\sum_{i,j,k,l=1}^{p} Q_{3ij}tr[Q_{4lm}'\{\gamma_2(I_{T+1}\circ Q_{4ij}) + tr(Q_{4ij})I_{T+1} + Q_{4ij} + Q_{4ij}'\}]Q_{3lm}'$$
$$+ \sigma^4\sum_{i,j,l=1}^{p} Q_{3ij}tr[Q_{6l}'\{\gamma_2(I_{T+1}\circ Q_{4ij}) + tr(Q_{4ij})I_{T+1} + Q_{4ij} + Q_{4ij}'\}]Q_{5l}'$$
$$+ \sigma^2\sum_{i=1}^{p} tr(Q_{6i})Q_{5i}Q_1' + \sigma^3\gamma_1\sum_{i=1}^{p} Q_{5i}\{(I_{T+1}\circ Q_{6i})i\}'Q_2'$$
$$+ \sigma^4\sum_{i,j,l=1}^{p} Q_{5i}tr[Q_{4jl}'\{\gamma_2(I_{T+1}\circ Q_{6i}) + tr(Q_{6i})I_{T+1} + Q_{6i} + Q_{6i}')\}]Q_{3jl}'$$
$$+ \sigma^4\sum_{i,j=1}^{p} Q_{5i}tr[Q_{6j}'\{\gamma_2(I_{T+1}\circ Q_{6i}) + tr(Q_{6i})I_{T+1} + Q_{6i} + Q_{6i}'\}]Q_{5j}'.$$

B.3. Theorem 3

From Theorem 1 we have $E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2}\left(tr(H_i|_\delta J) - \delta' H_i|_\delta \delta\right) + o(T^{-1})$. From Lemma 2 we have $\delta = E[Q_1 + Q_2v + Q_3v'Q_4v + Q_5v'Q_6v] = Q_1 + Q_3tr(Q_4) + Q_5tr(Q_6)$. The Hessian H_i was found in Theorem 2, and to evaluate it note that $E[Z'Z] = E[(\bar{Z} + G_bv\Gamma_4)'(\bar{Z} + G_bv\Gamma_4)]$ and $E[\zeta] = E[(\bar{Z} + G_bv\Gamma_4)'\{(\sum_{i=1}^{p}\lambda_i\bar{y}_{-i}) + X\beta + [(\sum_{i=1}^{p}\lambda_iG_i)\Gamma_3]v\}]$. These two expected values are then calculated in the same way as Lemma 2.

B.4. 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

$$H = e_1\hat{\zeta}'(Z'Z)^{\otimes(-4)}(V_2 + V_3)e_1' - e_2(Z'Z)^{\otimes(-2)}e_1' - e_1(Z'Z)^{\otimes(-2)}V_1e_2'.$$

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

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

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

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

(i) Specialising the matrix J

When $p = 1$ we have

$$
\begin{aligned}
J = {} & Q_1Q_1' + \sigma^2 tr(Q_4')Q_1Q_3' + \sigma^2 tr(Q_6')Q_1Q_5' + \sigma^2 Q_2Q_2' \\
& + \sigma^3\gamma_1 Q_2(I_{T+1} \circ (Q_4'))iQ_3' + \sigma^3\gamma_1 Q_2(I_{T+1} \circ (Q_6'))iQ_5' \\
& + \sigma^2 tr(Q_4)Q_3Q_1' + \sigma^3\gamma_1 Q_3\{(I_{T+1} \circ Q_4')i\}'Q_2' \\
& + \sigma^4 Q_3 tr[Q_4'\{\gamma_2(I_{T+1} \circ Q_4) + tr(Q_4)I_{T+1} + Q_4 + Q_4'\}]Q_3' \\
& + \sigma^4 Q_3 tr[Q_6'\{\gamma_2(I_{T+1} \circ Q_4) + tr(Q_4)I_{T+1} + Q_4 + Q_4'\}]Q_5' \\
& + \sigma^2 tr(Q_6)Q_5Q_1' + \sigma^3\gamma_1 Q_5\{(I_{T+1} \circ Q_6)i\}'Q_2' \\
& + \sigma^4 Q_5 tr[Q_4'\{\gamma_2(I_{T+1} \circ Q_6) + tr(Q_6)I_{T+1} + Q_6 + Q_6')\}]Q_3' \\
& + \sigma^4 Q_5 tr[Q_6'\{\gamma_2(I_{T+1} \circ Q_6) + tr(Q_6)I_{T+1} + Q_6 + Q_6'\}]Q_5'
\end{aligned}
$$

where the redundant summation over i, j, l and m has been removed. 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:

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

$$\begin{aligned}tr(P_4'P_4') &= tr\{(\lambda G + [0_{T\times1} : I_T])'G(\lambda G + [0_{T\times1} : I_T])'G\}\\ &= \lambda^2 tr(G'GG'G) + 2\lambda tr\{[0_{T\times1} : I_T]'GG'G)\\ &\quad + tr([0_{T\times1} : I_T]'(\omega F : C)[0_{T\times1} : I_T]'(\omega F : C)\}\\ &= \lambda^2 T(\lambda^2+1)(\frac{1}{1-\lambda^2})^3 + 2\lambda\{T\lambda(\frac{1}{1-\lambda^2})^2\} + o(T).\end{aligned}$$

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:

$$\begin{aligned}J_1 = \sigma^4&\left\{\frac{T(2+T-(T-2)\lambda^2}{(1-\lambda^2)^3}\right\}e_1e_1'\\ &+\sigma^4\left\{\frac{T\lambda(4+T(1-\lambda^2))}{(1-\lambda^2)^3}\right\}e_1e_2'\\ &+\sigma^4\left\{\frac{T\lambda(4+T(1-\lambda^2))}{(1-\lambda^2)^3}\right\}e_2e_1'\\ &+\sigma^4\left\{\frac{T\{(T+4)\lambda^2-(T+1)\lambda^4+1\})}{(1-\lambda^2)^3}\right\}e_2e_2' + o(T).\end{aligned}$$

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 is unnecessary to find $\tilde{J}_2$, because it can already be shown (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

$$\begin{aligned}(\tilde{H}_i|_\delta J_2)_{11} &= \frac{2\lambda(1-\lambda^2)\sum_{i=1}^{T+1}(G'G)_{ii}}{T^2\sigma^4} - \frac{\lambda(1-\lambda^2)\sum_{i=1}^{T+1}(G'G)_{ii}}{T^2\sigma^4}\\ &\quad + o(T^{-1})\\ \text{and}\quad (\tilde{H}_i|_\delta J_2)_{22} &= -\frac{\lambda(1-\lambda^2)\sum_{i=1}^{T+1}(G'G)_{ii}}{T^2\sigma^4} + o(T^{-1}),\end{aligned}$$

showing that $tr(\tilde{H}|_\delta J_2)$ is $o(T^{-1})$.

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.

B.5. Corollary 3

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

For the Corollary we are only interested in the $i = 1$ case. We have $B_1 = (e_1, 0, e_2, 0, e_3, 0, e_4, 0, e_5, 0, e_6, 0)'$ where the e_i here are 12×1 vectors with 1 in position i and 0 elsewhere. The matrix M is written in terms of Γ_1, Γ_2, $\hat{\zeta}$, $E[Z'Z]$, V_2 and V_3. These can be specialised as follows: $E[Z'Z] = \bar{Z}'\bar{Z} + \Gamma_4'\Gamma_4 tr(G'G) = \begin{pmatrix} T\left(\frac{\beta}{1-\lambda}\right)^2 + tr(G'G) & T\left(\frac{\beta}{1-\lambda}\right) \\ T\left(\frac{\beta}{1-\lambda}\right) & T \end{pmatrix}$

$$V_2 = (K_{22} \otimes I_2)(I_2 \otimes vec(E[Z'Z])) = \begin{pmatrix} T\left(\frac{\beta}{1-\lambda}\right)^2 + tr(G'G) & 0 \\ T\left(\frac{\beta}{1-\lambda}\right) & 0 \\ 0 & T\left(\frac{\beta}{1-\lambda}\right)^2 + tr(G'G) \\ 0 & T\left(\frac{\beta}{1-\lambda}\right) \\ T\left(\frac{\beta}{1-\lambda}\right) & 0 \\ T & 0 \\ 0 & T\left(\frac{\beta}{1-\lambda}\right) \\ 0 & T \end{pmatrix}$$

$V_3 = V_2$

$$\hat{\zeta} = P_1 + P_3 tr(P_4) = \lambda \begin{pmatrix} T\left(\frac{\beta}{1-\lambda}\right)^2 \\ \frac{T\beta}{1-\lambda} \end{pmatrix} + \beta \begin{pmatrix} \frac{T\beta}{1-\lambda} \\ T \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 0 \end{pmatrix} (\lambda tr(G'G) + tr(G'\Gamma_3)).$$

(i) Specialising the matrix J

The matrix J is in terms of Q_1 to Q_6. These are

$$Q_1 = \begin{pmatrix} T(\beta/(1-\lambda))^2 \\ T\beta/(1-\lambda) \\ T\beta/(1-\lambda) \\ T \\ T\{\lambda(\beta/(1-\lambda))^2 + \beta^2/(1-\lambda)\} \\ T\{\lambda\beta/(1-\lambda) + \beta\} \end{pmatrix}, Q_2 = \begin{pmatrix} (2\beta/(1-\lambda))i'G \\ i'G \\ i'G \\ (2\lambda\beta/(1-\lambda) + \beta)i'G + (\beta/(1-\lambda))i'\Gamma_3 \\ \lambda i'G + i'\Gamma_3 \end{pmatrix}$$

$Q_3 = (1, 0, 0, 0, 0, 0)'$, $Q_4 = G'G$, $Q_5 = (0, 0, 0, 0, 1, 0)'$,
$Q_6 = \lambda G'G + G'[0_{T\times 1} : I_T]$.

These were found using

$$A_1 = \begin{pmatrix} T(\beta/(1-\lambda))^2 \\ T\beta/(1-\lambda) \\ T\beta/(1-\lambda)T \end{pmatrix}, A_2 = \begin{pmatrix} 2\bar{y}'_{-1}G \\ i'G \\ i'G \\ 0 \end{pmatrix}, A_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, A_4 = G'G,$$

$$P_1 = \lambda \begin{pmatrix} T(\beta/(1-\lambda))^2 \\ T\beta/(1-\lambda) \end{pmatrix} + \beta \begin{pmatrix} T\beta/(1-\lambda) \\ T \end{pmatrix}$$

$$P_2 = \begin{pmatrix} (\lambda\bar{y}_{-1} + i\beta)'G \\ 0 \end{pmatrix} + \lambda Z'G + \bar{Z}'\Gamma_3, P_3 = (1,0)', P_4 = \lambda G'G + G'\Gamma_3.$$

(iii) Finding the bias result

To find the result in Corollary 3 it is necessary to evaluate $H|_\delta$ using Γ_1, Γ_2, $\hat{\zeta}$, $E[Z'Z]$, V_2 and V_3, and to calcualte J for the specific versions of Q_1 to Q_6 in (ii). Due to the term $E[Z'Z]^{\otimes(-4)}$ in particular, it seemed necessary to use a symbolic programming tool: a Mathematica programme is available from the author's website, where the final reduction is made. A program for Corollary 2 is also available. As before, we must use the $O(T)$ approximations in Lemma 3 in order to filter out any smaller $o(T^{-1})$ terms in the approximation. To deal with the terms in γ_1 we have

$Q_2(I_{T+1} \circ (Q_4'))iQ_3' = A(1,0,0,0,0,0)$, $Q_2(I_{T+1} \circ (Q_6'))iQ_5' = \lambda A(0,0,0,0,1,0)$,

$Q_3\{(I_{T+1} \circ Q_4')i\}'Q_2' = \{A(1,0,0,0,0,0)\}'$, $Q_5\{(I_{T+1} \circ Q_6)i\}'Q_2' = \lambda\{A(0,0,0,0,1,0)\}'$ where

$$A = \begin{pmatrix} (2\beta/(1-\lambda))A_1 \\ A_1 \\ A_1 \\ 0 \\ ((2\lambda\beta/(1-\lambda)) + \beta)A_1 + (\beta/(1-\lambda))tr(C'C) \\ \lambda A_1 + tr(C'C) \end{pmatrix}$$

and $A_1 = i_T'\omega^3 FF'F + i_T'C(I_T \circ C'C)i_T$. We then use the results in Lemma 4 as before, along with those in Lemma 5.