A note on approximating moments of least squares estimators

Liu-Evans, Gareth

University of LIverpool

24 July 2014

A note on approximating moments of least squares estimators

Gareth Liu-Evans*

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.

Keywords: asymptotic approximation, bias, least squares, time series, simulteneity

*University of Liverpool
Email address: gareth.liu-evans@liverpool.ac.uk (Gareth Liu-Evans)

Introduction

We present results to facilitate the asymptotic approximation of the moments of least squares coefficient estimators under similar assumptions to Phillips (2000), but focussing on the OLS estimator. The expansion method in Phillips (2000) is valid for consistent k-class estimation of equations from static simultaneous equation systems, but does not apply to OLS or dynamic models. We instead use the validity framework in Kiviet and Phillips (2014), 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 Neudecker (1969), Neudecker and Wansbeek (1983), and Magnus and Neudecker (1979, 2002). The approach builds on Marriott and Pope (1954) and Kendall (1954), 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 Nagar (1959).

There are large-T asymptotic approximations of the OLS estimator moments using Nagar (1959)-type methodologies in Kiviet and Phillips (2005, 2012) and Bao (2007) in the context of stationary autoregressive models, and in Kiviet and Phillips (1996) for equations from static simultaneous equation models. Building on Rilestone et al. (1996), Bao and Ullah (2007) present a general method for estimators of time-series models. See also Sargan (1974, 1976).

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 Shao and Tu (1995) section 2.4 and Shao (1988), and for the rest see Kiviet and Phillips (2014), 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 $J = E[\hat{\delta}\hat{\delta}']$.

H_i is found below using the Second Identification Theorem in Magnus and Neudecker (2002). 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 $H_i = \frac{1}{2}(A_i + A_i')$, where A_i will be in terms of $Z'Z$ and $\hat{\zeta}$. In the following we write $A \otimes A \otimes ... \otimes A = A^{\otimes m}$, where $\otimes$ is the Kronecker product and A appears m times. Since $(A \otimes A \otimes ... \otimes A)^{-1} = A^{-1} \otimes A^{-1} \otimes ... \otimes A^{-1}$ when A is invertible, we denote this by $A^{\otimes(-m)}$.

The first differential of $\hat{\alpha}$ is

$$\begin{aligned} d\hat{\alpha} &= (d(Z'Z)^{-1})\hat{\zeta} + (Z'Z)^{-1}d\hat{\zeta} \\ &= -(\hat{\zeta}' \otimes I_N)(Z'Z)^{\otimes(-2)}vec(d(Z'Z)) + (Z'Z)^{-1}d\hat{\zeta}. \end{aligned} \tag{7}$$

Here we have made use of

Lemma 1. *(See Magnus and Neudecker (2002)) For matrices U, V and W,*

1. $d(UV) = (dU)V + UdV$
2. $dU^{-1} = -U^{-1}(dU)U^{-1}$
3. $vec(UVW) = (W' \otimes U)vec(V)$

Writing $vec(\hat{\zeta}) = \Gamma_1\hat{\delta}$ and $vec(Z'Z) = \Gamma_2\hat{\delta}$ where $\Gamma_1 = [0_{N\times N^2} : I_N]$ and $\Gamma_2 = [I_{N^2} : 0_{N^2\times N}]$ this becomes

$$d\hat{\alpha} = Nd\hat{\delta}, \tag{8}$$

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

$$d\hat{\alpha} = vec(I_N N(d\hat{\delta})) = ((d\hat{\delta})' \otimes I_N)vecN, \tag{9}$$

which is a convenient form for calculating the second differential of $\hat{\alpha}_i$:

$$d^2\hat{\alpha}_i = d(vec(N))'((d\hat{\delta}) \otimes I_N)e_i. \tag{10}$$

In the above we use $d(d\hat{\delta}') = 0$, since $d\hat{\delta}$ is a constant vector increment in $d\hat{\alpha}$.

We note that $\{(d\hat{\delta}) \otimes I_N\}e_i$ can be written in the form $B_i(d\hat{\delta})$. The term $(d\hat{\delta}) \otimes I_N$ 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

$$\{(d\hat{\delta}) \otimes I_N)\}e_i = (d\hat{r}_1 e_i, ..., d\hat{r}_{N(N+1)}e_i) = B_i d\hat{\delta} \tag{11}$$

for some constant $N^2(N+1) \times N(N+1)$ matrix B_i. The $nm-th$ element of B_1 is $(B_1)_{nm} = 1$ 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 $(B_2)_{nm} = 1$ for $n = qN+2$ and $m = q+1$, up to $q = N^2+N-1$, and zero otherwise. For general i we have $(B_i)_{nm} = 1$ for $n = qN+i$, $m = q+1$ and $q = 1, ..., N^2+N-1$.

Therefore

$$d^2\hat{\alpha}_i = (dvec(N))' B_i (d\hat{\delta}), \tag{12}$$

and the remaining task is to put this in the form

$$d^2\hat{\alpha}_i = (d\hat{\delta})' M B_i (d\hat{\delta}) \tag{13}$$

for some M, so that the Hessian can be identified. We do this by finding the matrix M such that $dvec(N) = M' d\hat{\delta}$.

From N in (14),

$$dvec(N) = dvec\{(Z'Z)^{-1}\Gamma_1\} - dvec\{(\hat{\zeta}' \otimes I_N)(Z'Z)^{\otimes(-2)}\Gamma_2\}. \tag{14}$$

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 using Lemmas 1 and 2:

$$\begin{aligned} &vec((d(\hat{\zeta}' \otimes I_N))(Z'Z)^{\otimes(-2)}\Gamma_2 + (\hat{\zeta}' \otimes I_N) d[(Z'Z)^{\otimes(-2)}\Gamma_2]) \\ &\quad = (((Z'Z)^{\otimes(-2)}\Gamma_2)' \otimes I_N)(I_N \otimes V_1) vec(d\hat{\zeta}') \\ &\qquad - [\Gamma_2' \otimes (\hat{\zeta}' \otimes I_N)] vec((Z'Z)^{\otimes(-2)} d((Z'Z)^{\otimes 2})(Z'Z)^{\otimes(-2)}). \end{aligned} \tag{15}$$

where $V_1 = (K_{21} \otimes I_2)(I_1 \otimes vec(I_N)) = (K_{N1} \otimes I_N) vec(I_N) = vec(I_N)$.

Lemma 2. *(See Magnus and Neudecker (2002)) For matrices P $(m \times n)$, Q $(p \times q)$, U and V and where K_{cd} is a $cd \times cd$ commutation matrix,*

1. $d(U \otimes V) = (dU) \otimes V + U \otimes dV$
2. $vec(P \otimes Q) = (I_n \otimes U) vec(P) = (V \otimes I_p) vecB$,
   where $U = (K_{qm} \otimes I_q)(I_m \otimes vecB)$ *and* $V = (I_n \otimes K_{qm})(vecA \otimes I_q)$

Using the second result in Lemma 2 again this becomes

$$\begin{aligned} &(((Z'Z)^{\otimes(-2)}\Gamma_2)' \otimes I_N)(I_N \otimes V_1)\Gamma_1 d\hat{\delta} \\ &\qquad - [\Gamma_2' \otimes (\hat{\zeta}' \otimes I_N)](Z'Z)^{\otimes(-4)}[(I_N \otimes V_2) + (V_3 \otimes I_N)]\Gamma_2 d\hat{\delta}, \end{aligned} \tag{16}$$

where $V_2 = (K_{NN} \otimes I_N)(I_N \otimes vec(Z'Z))$ and $V_3 = (I_N \otimes K_{NN})(vec(Z'Z) \otimes I_N)$.

The term $dvec(N)$ is now in the required form so that $d^2\hat{\alpha}_i = (d\hat{\delta})' M B_i (d\hat{\delta})$ with

$$M = \Big(([\Gamma_2' \otimes (\hat{\zeta}' \otimes I_N)](Z'Z)^{\otimes(-4)}[(I_N \otimes V_2) + (V_3 \otimes I_N)]\Gamma_2 - (\Gamma_1' \otimes I_N)(Z'Z)^{\otimes(-2)}\Gamma_2 - [((Z'Z)^{\otimes(-2)}\Gamma_2)' \otimes I_N](I_N \otimes V_1)\Gamma_1 \Big)'. \quad (17)$$

and the Hessian result is summarised in Theorem 2 below.

Theorem 2. *The unevaluated Hessian matrix of* $f_i(\hat{\delta})$ *is*

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

To evaluate the Hessian at δ, which is required in Theorem 1, we replace $\hat{\zeta}$ and $Z'Z$ with their expected values.

In the endogenous regressor case, where $E[Z'u] \neq 0$, note that

$$\alpha = (E[Z'Z])^{-1}E[Z'y] - (E[Z'Z])^{-1}E[Z'u] \Rightarrow \alpha_i = f_i(\delta) - e_i' E_1^{-1} E_2, \quad (18)$$

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

$$E[\hat{\alpha}_i] = f_i(\delta) + \frac{1}{2}\left(tr(H_i|_\delta J) + \delta' H_i|_\delta \delta\right) + o(T^{-1}), \quad (19)$$

the bias in OLS estimation is

$$E[\hat{\alpha}_i - \alpha_i] = \frac{1}{2}\left(tr(H_i|_\delta J) + \delta' H_i|_\delta \delta\right) + e_i' E_1^{-1} E_2 + o(T^{-1}). \quad (20)$$

The approximation is valid to order $O(T^{-1})$ under the same conditions as before, where the expected values E_1 and E_2 are required for $H_i|_\delta$ already.

AR(1) illustration

We do a simple illustration of the expansion here for an AR(1) with known mean. A similar illustration for the case with constant, along with a more general result for an AR(p) with k added exogenous variables, but without the $o(T^{-1})$ terms removed from the approximation, is available from the author in the form of a Technical Appendix[1].

The model with added exogenous data and general lag length, an ARX(p), may be written

$$y = \lambda_1 y_{-1} + ... + \lambda_p y_{-p} + X\beta + u, \tag{21}$$

where $u = \Gamma_3 v$ with $\Gamma_3 = [0_{T\times p} : I_T]$, X is a $T \times k$ matrix of fixed or strongly exogenous regressors, and where v is a $(T+p) \times 1$ random vector. It is assumed that 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. Moreover, it is assumed 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. These two assumptions make 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. It is relatively straightforward to write $vec(\hat{\zeta})$ and $vec(Z'Z)$ for the ARX(p) case in the following form:

$$vec(\hat{\zeta}) = P_1 + P_2 v + \sum_{i=1}^{p} P_{3i} v' P_{4i} v \tag{22}$$

$$vec(Z'Z) = A_1 + A_2 v + \sum_{i,j=1}^{p} A_{3ij} v' A_{4ij} v, \tag{23}$$

where P_1, P_{3i}, A_1, A_{3ij} are vectors and P_2 P_{4i}, A_2, A_{4ij} are matrices.

In the case where $p = 1$ and $\beta = 0$, the indices drop out and $A_1 = A_2 = P_1 = P_2 = 0$, $A_3 = P_3 = 1$, $A_4 = G'G$, and $P_4 = G'(\lambda G + [0_{T\times 1} : I_T])$, where the matrix G is defined in Kiviet and Phillips (2012). By expanding the product $\hat{\delta}\hat{\delta}'$ using this, and using results in Appendix 5 of Ullah (2004) for expected values of products of quadratic forms and vectors in non-normal random variables, one finds $J = J_1 + \gamma_2 J_2$ where J_1 and J_2 are given in the Appendix.

[1]This unfiltered (see Kiviet and Phillips (2012)) approximation for the more general model builds on the filtered approximations in Kiviet and Phillips (1994) for the normal case and Bao and Ullah (2007) for the model with $p = 1$.

To obtain $H_{i|\delta}$ we drop the index i and note that $H_i = H = 2BM'$ for the case at hand since H, M and B are scalars. Also $B = 1$, so $H = M'$. It follows that

$$H = \begin{pmatrix} 2\hat{\zeta}(Z'Z)^{-3} & -(Z'Z)^{-2} \\ -(Z'Z)^{-2} & 0 \end{pmatrix} \text{ and } H|_\delta = \begin{pmatrix} 2\zeta(E[Z'Z])^{-3} & -(E[Z'Z])^{-2} \\ -(E[Z'Z])^{-2} & 0 \end{pmatrix}. \quad (24)$$

The expected values in $H|_\delta$ are $E[Z'Z] = \sigma^2 tr(G'G)$ and $\zeta = \sigma^2\lambda tr(G'G)$, and the largest terms in each are $O(T)$, so $H|_\delta$ is at most $O(T^{-2})$. Since we may discard contributions of order $O(\lambda^{sT})$ for $s > 0$ in both $H|_\delta$ and J, we use an approximation[2] of $tr(G'G)$ up to order $O(1)$ and find

$$H|_\delta = \tilde{H}|_\delta + o(T^{-2}) \text{ with } \tilde{H}|_\delta = \begin{pmatrix} \frac{2\lambda(1-\lambda^2)^2}{\sigma^4 T^2} & -\frac{(1-\lambda^2)^2}{\sigma^4 T^2} \\ -\frac{(1-\lambda^2)^2}{\sigma^4 T^2} & 0 \end{pmatrix}. \quad (25)$$

For the product $H|_\delta J$ to contain no $o(T^{-1})$ terms, we discard any $o(T)$ terms from J. For J_1 we obtain

$$J_1 = \tilde{J}_1 + o(T) \text{ with } \tilde{J}_1 = \sigma^4 \begin{pmatrix} \frac{T\{2+T-(T-2)\lambda^2\}}{(1-\lambda^2)^3} & \frac{T\lambda\{4+T(1-\lambda^2)\}}{(1-\lambda^2)^3} \\ \frac{T\lambda\{4+T(1-\lambda^2)\}}{(1-\lambda^2)^3} & \frac{T\{(T+4)\lambda^2-(T+1)\lambda^4+1\}}{(1-\lambda^2)^3} \end{pmatrix}. \quad (26)$$

The term J_2 can be written as

$$J_2 = \gamma_2\sigma^4 \begin{pmatrix} \sum_{i=1}^{T+1}(G'G)_{ii}^2 & \lambda\sum_{i=1}^{T+1}(G'G)_{ii}^2 \\ \lambda\sum_{i=1}^{T+1}(G'G)_{ii}^2 & tr[P_4'(I_{T+1}\circ P_4)] \end{pmatrix}, \quad (27)$$

with $(G'G)_{ii}$ denoting the ith diagonal entry of $G'G$, and we see from above that $tr(\tilde{H}|_\delta J_2) = o(T^{-1})$, since $(\tilde{H}_i|_\delta J_2)_{11} + (\tilde{H}_i|_\delta J_2)_{22} = 0$. Using (see earlier for the components) $\delta = \sigma^2 tr(G'G)e_1 + \sigma^2\lambda tr(G'G)e_2$ it is straightforward to show that $\delta'\tilde{H}|_\delta\delta = 0$, and the $O(T^{-1})$ bias is therefore $tr(\tilde{H}|_\delta\tilde{J}_1)$ which simplifies to

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

agreeing with Kendall (1954) and Marriott and Pope (1954).

[2]See e.g. Kiviet and Phillips (2012).

Conclusion
The results here facilitate the direct application of the validity framework outlined in Kiviet and Phillips (2014). The approach is applicable in principle to models with both dynamics and endogeneity. It may be possible in the future to extend the methodology to 2SLS moment approximation.

Appendix

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

where $\circ$ is the Hadamard matrix product.

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

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

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., 2012. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. Computational Statistics and Data Analysis 56(11), 3706–3729.

Kiviet, J., Phillips, G., 2014. Improved variance estimation of maximum likelihood estimators in stable first-order dynamic regression models. Computational Statistics and Data Analysis 76, 424–448.

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

Magnus, J., Neudecker, H., 2002. Matrix differential calculus with applications in statistics and econometrics (Revised edition). Wiley.

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.

Neudecker, H., 1969. Some theorems on matrix differentiation with special reference to kronecker matrix products. Journal of the American Statistical Association 64 (327), pp. 953–963.

Neudecker, H., Wansbeek, T., 1983. Some results on commutation matrices, with statistical applications. The Canadian Journal of Statistics / La Revue Canadienne de Statistique 11 (3), pp. 221–231.

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

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.

Sargan, J., 1976. Econometric estimators and the Edgeworth expansion. Econometrica 44, 421–448.

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, 1st Edition. Springer-Verlag, 175th Avenue, New York.

Ullah, A., 2004. Finite Sample Econometrics. Oxford University Press.