Working Paper in Economics

202117

The Robustness of the 2SLS Mean Squared Error to Non-Normal Disturbances

Garry DA Phillips
Gareth Liu-Evans

https://www.liverpool.ac.uk/management/people/economics/

The Robustness of the 2SLS Mean Squared Error to Non-Normal disturbances

Garry DA Phillips & Gareth Liu-Evans

March 15, 2021

Abstract

In a seminal paper Nagar (1959) obtained first and second moment approximations for the k-class of estimators in a general static simultaneous equation model under the assumption that the structural disturbances were i.i.d. and Normally distributed. In this paper we show that the second moment approximation for 2SLS continues to be valid under symmetric, but not necessarily normal, disturbances with an arbtrary degree of kurtosis but not when the disturbances are asymmetric. A modified approximation for the second moment is then obtained which includes the case of asymmetric disturbances. A series of Monte Carlo simulations shows that the second moment can indeed vary considerably across different structural error skewness scenarios.

1 Introduction

There has recently been a resurgence of interest in the properties of intrumental varable estimators, particularly in the context of weak/many instruments, see Davidson and Mackinnon (2006). Given the importance of moment approximations in analysing the small sample properties of estimators, it is not surprising that such approximations continue to play an important part, see, for example, Donald and Newey (2001) and Hahn, Hausman and Kuersteiner (2004).

Moment approximations in simultaneous equation models have a long history. In a seminal paper, Nagar (1959) derived approximations to the first and second moments of the consistent k-class of estimators in a general simultaneous equation model with exogenous regressors and, in obtaining his results, it was assumed that the structural disturbances were independently and normally distributed. Later Mikhail (1972) extended Nagar's bias approximation under the same assumptions. Nagar's work led to a great deal of research concerned with the small sample properties of simultaneous equation estimators; in particular, various writers examined conditions under which Nagar's approximations were valid, see Srinavasan (1970). The main result was given by Sargan (1974) who showed that a necessary and sufficient condition was that the estimator moments should exist. Much work has been done to explore the existence of estimator moments especially in simplified models. A paper which is of particular relevance, given its generality, is Kinal (1980). His results show that in the general model chosen by Nagar, the $2SLS$ estimator has moments up to the order of overidentification. However, k-class estimators behave differently depending on the value taken by k. When $k > 1$, which includes the $LIML$ estimator, the estimators do not possess moments of any order while when $k < 1$ higher moments exist and this does not depend on the order of overidentication.

In Phillips (2000) it was shown that the Nagar bias approximation for the $2SLS$ estimator is correct under much less restricted conditions than assumed by Nagar. In particular, the result does not require the assumption of normality. In Phillips (2007) it was noted that for the bias approximation to hold a sufficient condition is that the disturbances obey the classical Gauss-Markov assumptions which includes, in particular, the class of conditionally heteroscedastic disturbances such as $ARCH/GARCH$. However, in both papers the normality assumption was retained when the second moment approximation was considered. Clearly the normality assumption is restrictive and it would be most helpful if the second moment approximation could also be obtained under weaker assumptions.

In this paper it is shown that the Nagar second moment approximation is valid without the normality assumption but assuming that disturbances are symmetrically distributed and are i.i.d. with a finite variance. When the disturbances are asymmetric, however, the Nagar second moment approximation needs to be extended and we give the corrected expression, mirroring the work in Liu-Evans and Phillips (2018) for the higher-order 2SLS bias under asymmetry. A modified approach is used when obtaining the asymptotic approximation, and a large number of Monte Carlo experiments are used to illustrate the effect of varying structural error asymmetry on the mean squared error of the 2SLS estimator.

2 Asymptotic Approximations

Suppose that a random variable z_T admits an asymptotic expansion of the form

$$z_T = a_{0T} + a_{1T} + a_{2T} + a_{3T} + ... \tag{1}$$

where a_{0T} is $O_p(1)$ while a_{jT} is $O_p(T^{-j/2})$ for $j = 1, 2, ...,$as $T \longrightarrow \infty$.

Suppose also that $E(a_{0T}) = 0$, for all T, while $E(a_{jT}) = \bar{a}_{jT}$ is $O(T^{\frac{-j}{2}})$, $j = 1, 2, ...$, then the expected value of z_T can be approximated as

$$E(z_T) = \sum_{j=1}^{r} \bar{a}_{jT} + o(T^{-\frac{r+1}{2}}), \tag{2}$$

which is typically the way in which Nagar approximations are presented.

If $E(T^{\frac{j+1}{2}} a_{jT}) = \alpha_j$, $j = 1, 2, ..r$, as $T \rightarrow \infty$, then we have an alternative large-T approximation:

$$E(z_T) = \frac{1}{T}\alpha_1 + \frac{1}{T^{\frac{3}{2}}}\alpha_2 + \frac{1}{T^2}\alpha_3 + ...+ ...$$

$$= \sum_{j=1}^{r} \frac{1}{T^{\frac{j+1}{2}}}\alpha_j + o(T^{-\frac{r+1}{2}}) \tag{3}$$

to r terms. Both forms of the approximation are valid to order $\frac{1}{T^{\frac{r}{2}}}$ under suitable moment conditions on the disturbances and for large T they are essentially equivalent. Nagar approximations usually take the form in (2); however, in cases where the $\alpha_{j,}$ $j = 2, 3, .., r$, take a simple form or when the a_j are more easily obtained than the $\bar{a}_{jT}$, the approximation in (3) may sometimes be preferred.

The restriction that $E(a_{jT}) = \bar{a}_{jT}$ is $O(T^{\frac{-j}{2}})$ is explicitly made because it is possible to find cases where it does not hold, in particular $E(a_{jT}\ )$ may be $o(T^{-\frac{j}{2}})$, so that the equivalence of (2) and (3) will not follow. A situation of this kind is discussed further later in the paper.

3 Model and Notation

We consider a simultaneous equation model containing G equations given by

$$By_t + \Gamma z_t = u_t, \quad t = 1, 2, ..., \tag{4}$$

in which y_t is a $G \times 1$ vector of endogenous variables, z_t is a $K \times 1$ vector of strongly exogenous variables and u_t is a $G \times 1$ vector of independently distributed structural disturbances with $G \times G$ positive definite covariance matrix Σ. The matrices of structural disturbances, B and Γ are, respectively, $G \times G$ and $G \times K$. It is assumed that B is non-singular so that the reduced form equations corresponding to (2) are:

$$y_t = -B^{-1}\Gamma z_t + B^{-1}u_t$$
$$= \Pi z_t + v_t,$$

where Π is a $G \times K$ matrix of reduced form coefficients and v_t is a $G \times 1$ vector of reduced form disturbances with a $G \times G$ positive definite covariance matrix Ω. With T observations we may write the system as

$$YB' + Z\Gamma' = U. \tag{5}$$

Here, Y is a $T \times G$ matrix of observations on endogenous variables, Z is a $T \times K$ matrix of observations on the strongly exogenous variables and U is a $T \times G$ matrix of structural disturbances all of which may be serially correlated. The first equation of the system will be written as

$$y_1 = Y_2\beta + Z_1\gamma + u_1, \tag{6}$$

where y_1 and Y_2 are, respectively, a $T \times 1$ vector and a $T \times g$ matrix of observations on $g + 1$ endogenous variables, Z_1 is a $T \times k$ matrix of observations on k

exogenous variables, β and γ are, respectively, $g \times 1$ and $k \times 1$ vectors of unknown parameters and u_1 is a $T \times 1$ vector of independently distributed disturbances with positive definite covariance matrix $E(u_1 u_1') = \Sigma_{11}$ and finite moments up to fourth order. The reduced form of the system includes $Y_1 = Z\Pi_1 + V_1$ in which $Y_1 = (y_1 : Y_2)$, $Z = (Z_1 : Z_2)$ is a $T \times K$ matrix of observations on K exogenous variables with an associated $K \times (g+1)$ matrix of reduced form parameters given by $\Pi_1 = (\pi_1 : \Pi_2)$, while $V_1 = (v_1 : V_2)$ is a $T \times (g+1)$ matrix of reduced form disturbances. The transpose of each row of V_1 is independently distributed with zero mean vector and (g+1)$\times$(g+1) positive definite matrix $\Omega_1 = (\omega_{ij})$ while the $T(g+1)$ vector $vecV_1$ has a positive definite covariance matrix of dimension $T(g+1) \times T(g+1)$ given by $Cov(vecV_1) = \Omega_1^{vec}$ and has finite moments up to fourth order. It is further assumed that:

(i) Equation (6) is over-identified so that $K > g + k$, i.e. the number of excluded variables exceeds the number required for the equation to be just identified. In cases where second moments are analysed we shall assume that K exceeds $g + k$ by at least two. These over-identifying restrictions are sufficient to ensure that the Nagar expansion is valid in the case considered by Nagar and that the estimator moments exist: see Sargan (1974).

(ii) The $T \times K$ matrix Z is strongly exogenous and of rank K and there exists a $K \times K$ positive definite matrix with limit matrix $\Sigma_{ZZ} = \lim_{T\to\infty} T^{-1}Z'Z$. Following Anderson et al (1986, p7) it will also be assumed that $T^{-1}Z'Z = \Sigma_{ZZ} + o(T^{-1})$.

3.1 Nagar Approximations to the first and second moments

In the Nagar approach to finding moment approximations for the 2SLS estimator, the estimation error is often written as

$$\hat{\alpha} - \alpha = \left(\begin{pmatrix} Y_2'Y_2 - \hat{V}_2'\hat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{pmatrix} \right)^{-1} \begin{pmatrix} Y_2 - \hat{V}_2 \\ Z_1' \end{pmatrix} u_1,$$

where $\hat{\alpha}$ is the 2SLS estimator of $\alpha = (\beta', \gamma')'$. Nagar commences by putting

$$\begin{aligned} \hat{\alpha} - \alpha &= \left[Q^{-1} + X'V_z + V_z'X + V_z'P_zV_z \right]^{-1} \left[X'u_1 + V_z'P_zu_1 \right] \\ &= \left[I + Q^{-1}\{X'V_z + V_z'X + V_z'P_zV_z\} \right]^{-1} Q \left[X'u_1 + V_z'P_zu_1 \right], \end{aligned}$$

where $X = (Z\Pi_2 : Z_1)$, $Q = (X'X)^{-1}$, $P_z = Z(Z'Z)^{-1}Z'$ and $V_z = (V_2 : 0)$. Putting $\Delta = X'V_z + V_z'X + V_z'P_zV_z$ and expanding the inverse $\left[I + Q^{-1}\Delta\} \right]^{-1}$ in a Taylor expansion yields

$$\begin{aligned} \hat{\alpha} - \alpha &= \left[I + Q\Delta\} \right]^{-1} Q \left[X'u_1 + V_z'P_zu_1 \right] \\ &= \left[I - Q\Delta + Q\Delta Q\Delta - +.... \right] Q \left[X'u_1 + V_z'P_zu_1 \right], \end{aligned}$$

where terms can be arranged in decreasing order of stochastic magnitude. In fact, if we write $u_1 = V_1\beta_0$ and $V_z = V_1 H'$, where $\beta_0 = (-1, \beta')'$ and $H = \begin{pmatrix} 0 & I_g \\ 0 & 0 \end{pmatrix}$ is a $(g+k) \times (g+1)$ selection matrix, then the Nagar expansion may be written in the form

$$\begin{aligned}\hat{\alpha} - \alpha = & \, QX'V_1\beta_0 + QHV_1'P_zV_1\beta_0 - QX'V_1H'QX'V_1\beta_0 - QHV_1'P_xV_1\beta_0 \\ & - QHV_1'P_zV_1H'QHV_1\beta_0 - QHV_1'P_zV_1H'QX'V_1\beta_0 \\ & - QX'V_1H'QHV_1'P_zV_1\beta_0 - QHV'XQHV'P_zV_1\beta_0 \\ & + QX'HV_1'QXV_1H'QX'V_1\beta_0 + QHV_1'P_xV_1H'QXV_1\beta_0 \\ & + QX'V_1HQHV_1'P_xV_1\beta_0 + QHV_1'XQHV_1'P_xV_1\beta_0 + o_p(T^{-\frac{3}{2}}). \quad (7)\end{aligned}$$

The Nagar bias approximation is found by summing the expectations of the terms up to order T^{-1} while the second moment approximation is found by squaring the above and summing the expectations of terms up to $O(T^{-2})$. We shall later compare this expansion with an alternative representation presented in Phillips (2000).

Notice that, if we require the Nagar expansion for a general element of the vector $\hat{\alpha}$, say $\hat{\alpha}_i$, i=1,..,$g+k$, then we may simply extract the required terms by premultiplying the expansion for $\hat{\alpha} - \alpha$ by e_i', where e_i is a $(g+k) \times 1$ unit vector.

The Nagar approximations for the $2SLS$ case are as follows:

(a) The bias of the $2SLS$ estimator for α in (6) is given by

$$E(\hat{\alpha} - \alpha) = [L-1]Qq + o(T^{-1}).$$

(b) The second moment matrix of the $2SLS$ estimator for α in (6) is given by

$$(E(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)) = \sigma^2 Q[I + A^*] + o(T^{-2}),$$

where $L = K - g_1 - k_1$ is the order of overidentification and

$$A^* = [-(2L-3)tr(C_1Q) + tr(C_2Q)]I + \{(-L+2)^2 + 2\}C_1Q + (-L+2)C_2Q.$$

Also $q = \frac{1}{T}\begin{bmatrix} E(V_2'u_1) \\ 0 \end{bmatrix} = \sigma^2 \begin{bmatrix} \pi \\ 0 \end{bmatrix}$, while $C = \begin{bmatrix} (1/T)E(V_2'V_2) & 0 \\ 0 & 0 \end{bmatrix} = C_1 + C_2$, $C_1 = \begin{bmatrix} \sigma^2\pi\pi' & 0 \\ 0 & 0 \end{bmatrix}$, $C_2 = \begin{bmatrix} 1/TE(W'W) & 0 \\ 0 & 0 \end{bmatrix}$, and $W = V_2 - u_1\pi'$.

Finally, a higher order bias approximation for $2SLS$ in the same framework as Nagar was developed by Mikhail (1972). It is given by

$$E(\hat{\alpha} - \alpha) = (L-1)[I + tr(QC)I - (L-2)QC]Qq \; + o(1/T^2).$$

The assumptions made by Mikhail in obtaining this result were the same as those used by Nagar so that normality was assumed for the disturbances. We shall examine this approximation later in the paper. It is of interest that the bias approximation is zero when $L = 1$, i.e. when the parameters of the equation are overidentified of order unity. The approximation may work well especially when L is not large, see Hadri and Phillips (1999) and Iglesias and Phillips (2008) for evidence of this.

4 Alternative Approximations to the First and Second Moments

We consider the estimation of the equation given in (4) by the method of $2SLS$. It is well known that the estimator can be written in the form

$$\hat{\alpha} = \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \hat{\Pi}_2' Z' Z \hat{\Pi}_2 & \hat{\Pi}_2' Z' Z_1 \\ Z_1' Z \hat{\Pi}_2 & Z_1' Z_1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\Pi}_2' Z' Z \hat{\pi}_1 \\ Z_1' Z \hat{\pi}_1 \end{pmatrix} \tag{8}$$

where $\hat{\Pi}_2 = (Z'Z)^{-1} Z' Y_2$ and $\hat{\pi}_1 = (Z'Z)^{-1} Z' y_1$. This representation of $2SLS$ was considered in Harvey and Phillips (1980) and in Phillips (2000, 2007). It is apparent that, conditional on the exogenous variables, the $2SLS$ estimators are functions of the matrix $\hat{\Pi}_1 = (\hat{\pi}_1 : \hat{\Pi}_2)$; hence we may write $\hat{\alpha} = f(vec\hat{\Pi}_1)$. As shown in Phillips (2000), the unknown parameter vector can be written as $\alpha = f(vec\Pi_1)$, so that the estimation error is $f(vec\hat{\Pi}_1) - f(vec\Pi_1)$. A Taylor expansion about the point $vec\Pi_1$ may then be employed directly to find a counterpart of the Nagar expansion. In fact, Phillips considered the general element of the estimation error $\hat{\alpha}_i - \alpha_i = e_i'(\hat{\alpha} - \alpha) = f_i(vec\hat{\Pi}_1) - f_i(vec\Pi_1)$, $i = 1, 2, ..., g + k$, where e_i' is a $1 \times (g + k)$ unit vector, and the bias approximation for the general case was found using the expansion:

$$\begin{aligned} f_i(vec\hat{\Pi}_1) \quad &= \quad f_i(vec\Pi_1) + (vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)} \\ &\quad + \frac{1}{2!}(vec(\hat{\Pi}_1 - \Pi_1))' f_i^{(2)} (vec(\hat{\Pi}_1 - \Pi_1)) \\ &\quad + \frac{1}{3!} \Sigma_{r=1}^{K} \Sigma_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1))' f_{i,rs}^{(3)} (vec(\hat{\Pi}_1 - \Pi_1)) \\ &\quad + o_p(T^{-\frac{3}{2}}), \end{aligned}$$

where $f_i^{(1)}$ is a $K(g+1)$ vector of first-order partial derivatives, $\frac{\partial f_i}{\partial vec\hat{\Pi}_1}$: $f_i^{(2)}$ is a $(K(g+1)) \times (K(g+1))$ matrix of second-order partial derivatives, $\frac{\partial^2 f_i}{\partial vec\hat{\Pi}_1 (\partial vec\hat{\Pi}_1)'}$, and $f_{i,rs}^{(3)}$ is a $(K(g + 1)) \times (K(g + 1))$ matrix of third-order partial derivatives defined as $f_{i,rs}^{(3)} = \frac{\partial f_i^{(2)}}{\partial \pi_{rs}}$, $r = 1, ..., K$, $s = 1, ..., g + 1$. All derivatives are evaluated at $vec\Pi_1$. The bias approximation to order T^{-1} is then obtained by taking expectations of the first two terms of the stochastic expansion to yield:

$$E(\hat{\alpha}_i - \alpha_i) = \frac{1}{2!} tr \left[(f_i^{(2)} (I \otimes (Z'Z)^{-1} Z') \Omega_1^{vec} (I \otimes Z(Z'Z)^{-1}) \right] + o(T^{-1})$$

When the partial derivatives $f_i^{(2)}$ are introduced and Ω_1^{vec} is interpreted in terms of the structural parameters, the bias approximation is readily found. It is of interest to examine this bias approximation further. Note that the approximation changes as the matrix Ω_1^{vec} changes. When $\Omega_1^{vec} = \Omega_1 \otimes I_T$, which is the case where the rows of the matrix V_1 are serially uncorrelated and homoscedastic, the approximation reduces to that given by Nagar. However, to obtain his approximation Nagar assumed that the disturbances were normally distributed while here we need only assume that the row vectors of V_1 obey the $Gauss\ Markov$ assumptions.

It is not immediately obvious that the above expansion is equivalent to that used by Nagar. Examining the Nagar expansion in (7), however, we note that the first term, which is $O_p(T^{-\frac{1}{2}})$, may be written as $e_i'QX'V_1\beta_0 = tr\{\beta_0 e_i'QX'V_1\} = tr\{\beta_0 e_i'QX'Z(Z'Z)Z'V_1\} = \{vec(Z'Z)^{-1}Z'V_1)\}'vec(\beta_0 e_i'QX'Z)$ $= (vec(\hat{\Pi}_1 - \Pi_1)'(\beta_0 \otimes Z'XQe_i) = (vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(1)}$, where $f_i^{(1)} = (\beta_0 \otimes Z'XQe_i)$ is derived in Phillips (2000). This is just the first term in the above expansion. By the same approach it may be shown that the $O_p(T^{-1})$ part of the Nagar expansion, which is given by the second, third and fourth terms, is equal to the second term in the above expansion, while the remaining terms of the Nagar expansion will form the $O_p(T^{\frac{-3}{2}})$ part, which is equal to the third term above. Expressions for $f_i^{(2)}$ and $f_{i,rs}^{(3)}$ are given in Phillips (2000).

To find the second moment approximation we shall need the following result:

Lemma 1. *If η is a random normal vector with mean zero and positive definite covariance matrix Ψ, i.e. $\eta \sim N(0,\Psi)$, and if A and B are any conformable matrices, then*

$$E(\eta' A\eta)(\eta' B\eta) = tr(A\Psi)tr(B\Psi) + tr(A\Psi B\Psi) + tr(A\Psi B'\Psi) \qquad (9)$$

A proof of this lemma appears, for example, in Magnus and Neudecker(1979) and it is used to find the second moment approximation by deriving the expectation of the square of the relevant terms in the following:

$$E\{(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(1)} + \tfrac{1}{2!}(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1) \\ + \tfrac{1}{3!}\sum_{r=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1)'f_{i,rs}^{(3)}(vec(\hat{\Pi}_1 - \Pi_1))\}^2$$

Collecting terms up to order T^{-2} we have:

$$\begin{aligned} E(\hat{\alpha}_i - \alpha_i)^2 = {} & E\{(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(1)})^2 \\ & + (vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(1)})(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1)) \\ & + \frac{1}{4}(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1))^2 \\ & + \frac{1}{3}(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(1)})\sum_{r=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1)'f_{i,rs}^{(3)}(vec(\hat{\Pi}_1 - \Pi_1))\} \\ & + o(T^{-2}). \end{aligned} \qquad (10)$$

The approximation is then found by taking expectations of the above terms. Under normality assumptions the Lemma can be applied directly to show that the approximation coincides with that of Nagar(1959). However, when normality is not assumed the required approximation has not been obtained. Here we shall show that Nagar′s result holds without the assumption of normality provided the disturbances are symmetric but when symmetry does not hold the approximation must be modified. First we propose a further lemma which is an extension of the earlier one.

Lemma 2. *Suppose ζ_T is a random vector of fixed dimension with mean zero and positive definite covariance matrix Ψ_T, and let A_T and B_T any conformable fixed matrices. Suppose also that, as $T \to \infty$, ζ_T converges in distribution to a vector ζ where $\zeta \sim N(0, \Psi)$ while A_T and B_T respectively converge to limit matrices A and B. Then, as $T \to \infty$,*

$$E(\zeta_T' A_T \zeta_T)(\zeta_T' B_T \zeta_T) = tr(A\Psi)tr(B\Psi) + tr(A\Psi B\Psi) + tr(A\Psi B'\Psi) \tag{11}$$

The result follows from Lemma 1 and the fact that $\zeta_T^{'} A_T \zeta_T$ and $\zeta_T^{'} B_T \zeta_T$ converge respectively to $\zeta^{'} A \zeta$ and $\zeta^{'} B \zeta$.

We shall use Lemma 2 in finding the approximation to the second moment as $T \to \infty$. We first note that under the assumptions made, $T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)$ is asymptotically normal and converges in distribution to $N(0, \Omega_1 \otimes \Sigma_{zz}^{-1})$ as $T \to \infty$ and for some of the analysis we shall use this result, We shall consider the following expectations:

1. $E(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})^2 = E(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)} (f_i^{(1)})' vec(\hat{\Pi}_1 - \Pi_1)$
where $f_i^{(1)} = \beta_0 \otimes Z'X(X'X)^{-1} e_i$ and $vec(\hat{\Pi}_1 - \Pi_1) = ((I \otimes (Z'Z)^{-1} Z') vec V_1$.

Noting that

$$vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)} (f_i^{(1)})' vec(\hat{\Pi}_1 - \Pi_1) = tr(f_i^{(1)} (f_i^{(1)})' vec(\hat{\Pi}_1 - \Pi_1) vec(\hat{\Pi}_1 - \Pi_1)')$$

we find

$$\begin{aligned} E(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)} (f_i^{(1)})' vec(\hat{\Pi}_1 - \Pi_1)) \\ &= tr[(\beta_0 \beta\prime_0 \otimes Z'X(X'X)^{-1} e_i e\prime_i (X'X)^{-1} X'Z)(\Omega_1 \otimes (Z'Z)^{-1})] \\ &= \beta_0' \Omega_1 \beta_0 e\prime_i (X'X)^{-1} e_i \end{aligned}$$

where $E(vec(\hat{\Pi}_1 - \Pi_1) vec(\hat{\Pi}_1 - \Pi_1)') = \Omega_1 \otimes (Z'Z)^{-1}$ is used.

Hence, we may write that, as $T \to \infty$,

$$E\{(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})^2 = \beta_0' \Omega_1 \beta_0 e_i' \Sigma_{xx}^{-1} e_i$$

so that

$$E\{(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})^2 = \frac{1}{T} \beta_0' \Omega_1 \beta_0 e_i' \Sigma_{xx}^{-1} e_i$$

is the first term in the approximation.

2. $E[(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1))]$

This can be evaluated by considering the following related expression:

$$E(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)}(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)).$$

As $T \to \infty$, $T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)$ converges in distribution to a normally distributed random vector ζ, $f_i^{(1)}$ converges to $\bar{f}_i^{(1)}$ and $f_i^{(2)}$ converges to $\bar{f}_i^{(2)}$. Hence, as $T \to \infty$, the above expectation reduces to

$$E(\zeta' \bar{f}_i^{(1)} \zeta' \bar{f}_i^{(2)} \zeta) = 0,$$

since each component of the vector is an expectation of a product of an odd number of normal random variables. Thus we have shown that, as $T \to \infty$,

$$E(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)}(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)) = 0 \qquad (12)$$

While this analysis is correct it can be shown that, when the disturbances are asymmetric, $E(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1))$ is $O(T^{-2})$ and not $O(T^{-\frac{3}{2}})$, so that the large-T approximation will be incorrect to order T^{-2} unless this is allowed for. This is an example of the phenomenon referred to in Section 2.

Putting $vec(\hat{\Pi}_1 - \Pi_1) = (I \otimes (Z'Z)^{-1} Z') vecV_1$ we find in Note 1 that

$$\begin{aligned} &E[(vecV_1)'(I \otimes Z(Z'Z)^{-1}) \\ &\quad \times f_i^{(1)} (vecV_1)'(I \otimes Z(Z'Z)^{-1}) f_i^{(2)} (I \otimes (Z'Z)^{-1} Z')(vecV_1)] \\ &\qquad = 2e_i'(X'X)^{-1} HD\beta_0 . tr\{(P_z - P_X)F\} \\ &\qquad\quad - 2e_i'(X'X)^{-1} X'FX(X'X)^{-1} HD\beta_0. \qquad (13) \end{aligned}$$

Here F is a $T \times T$ diagonal matrix with component $F_{j,j}$ given by $x_j' q_i$, where x_j' is the j^{th} row of X for $j = 1, 2, .., T$ and $q_i = Qe_i$ is the i^{th} column of $Q = (X'X)^{-1}$. The $(g+1) \times (g+1)$ matrix D has general element $D_{ij} = -\omega_{1,ij} + \beta_1 \omega_{2,ij} + \beta_2 \omega_{3,ij} + ... + \beta_g \omega_{g+1,ij}$, with $\omega_{s,ij} = E(v_{st} v_{it} v_{jt})$ for $s, i, j = 1, \ldots, g+1$ being a third moment of the reduced form disturbances. The expression in (14) is $O(T^{-2})$ since $tr\{(P_z - P_X)F\}$ is $O(T^{-1})$ while $X'FX$ and D are $O(1)$.

There is an alternative form for the matrix D which follows from noting that the structural disturbance u_{1t} may be written as: $u_{1t} = -v_{1t} + \beta_1 v_{2t} + \beta_2 v_{3t} + ... + \beta_{g+1} v_{gt}$ so that $D_{ij} = E(u_{1t} v_{it} v_{jt})$. This is given in Note 2. Finally we find that, as $T \to \infty$,

$$\begin{aligned} &E[T^2 (vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1))] \\ &\quad = 2e_i' \Sigma_{xx}^{-1} HD\beta_0 . \{tr\{(\Sigma_{zz}^{-1} \Sigma_{zFz} - \Sigma_{xx}^{-1} \Sigma_{xFx})\} - 2e_i' \Sigma_{xx}^{-1} \Sigma_{xFx} \Sigma_x^{-1} HD\beta_0\}, \end{aligned}$$

where $\lim X'FX = \Sigma_{xFx}$, $\lim Z'FZ = \Sigma_{zFz}$ and $\lim T tr\{(P_z - P_X)F\} = tr\{(\Sigma_{zz}^{-1} \Sigma_{zFz} - \Sigma_{xx}^{-1} \Sigma_{xFx})\}$. Hence, as $T \to \infty$,

$$\begin{aligned} &E[(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1))] \\ &\quad = \frac{1}{T^2} 2\{e_i' \Sigma_{xx}^{-1} HD\beta_0 . tr\{(\Sigma_{zz}^{-1} \Sigma_{zFz} - \Sigma_{xx}^{-1} \Sigma_{xFx})\} - e_i' \Sigma_{xx}^{-1} \Sigma_{xFx} \Sigma_{xx}^{-1} HD\beta_0\} \\ &\qquad (14) \end{aligned}$$

which is the second term in the approximation.

3. $E(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)} (vec(\hat{\Pi}_1 - \Pi_1))^2$

We commence from

$$E(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)} (T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1))^2$$

This is $E(\zeta' \bar{f}_i^{(2)} \zeta)^2 = (tr[\bar{f}_i^{(2)} (\Omega_1 \otimes \Sigma_{zz}^{-1})])^2 + 2tr[\bar{f}_i^{(2)} (\Omega_1 \otimes \Sigma_{zz}^{-1}) \bar{f}_i^{(2)} (\Omega_1 \otimes \Sigma_{zz}^{-1})]$ as $T \to \infty$ using Lemma 12. It therefore follows that, as $T \to \infty$,

$$\begin{aligned} &E(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)} (vec(\hat{\Pi}_1 - \Pi_1))^2 \\ &\quad = \frac{1}{T^2} (tr[\bar{f}_i^{(2)} (\Omega_1 \otimes \Sigma_{zz}^{-1})])^2 + 2tr[\bar{f}_i^{(2)} (\Omega_1 \otimes \Sigma_{zz}^{-1}) \bar{f}_i^{(2)} (\Omega_1 \otimes \Sigma_{zz}^{-1})] \end{aligned} \tag{15}$$

which is the third term in the approximation.

4. $E\{(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)}) \sum_{r=1}^{K} \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) \times (vec(\hat{\Pi}_1 - \Pi_1)' f_{i,rs}^{(3)} (vec(\hat{\Pi}_1 - \Pi_1)\}$

We commence from

$$\begin{aligned} &E\{(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)}) \\ &\quad \times \sum_{r=1}^{K} \sum_{s=1}^{g+1} T^{\frac{1}{2}} (\hat{\pi}_{rs} - \pi_{rs}) \times T^{\frac{1}{2}} (vec(\hat{\Pi}_1 - \Pi_1)' f_{i,rs}^{(3)} (T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1)\} \end{aligned}$$

This is $E(\zeta' \bar{f}_i^{(1)} \sum_{r=1}^{K} \sum_{s=1}^{g+1} E_{rs} \zeta \zeta' \bar{f}_{i,rs}^{(3)} \zeta) = \sum_{r=1}^{K} \sum_{s=1}^{g+1} E(\zeta' \bar{f}_i^{(1)} E_{rs} \zeta \zeta' \bar{f}_{i,rs}^{(3)} \zeta)$ as $T \to \infty$, where E_{rs} is a $K \times (g_1+1)$ vector with unity in the $r + (s-1)K^{th}$ position and zeros elsewhere so that $E_{rs}(T^{\frac{1}{2}} vec(\hat{\Pi}_1 - \Pi_1) = T^{\frac{1}{2}} (\hat{\pi}_{rs} - \pi_{rs})$. Lemma 2 can now be applied to yield

$$\begin{aligned} &E\{(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)}) \sum_{r=1}^{K} \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) \times (vec(\hat{\Pi}_1 - \Pi_1)' f_{i,rs}^{(3)} (vec(\hat{\Pi}_1 - \Pi_1))\} \\ &\qquad = \frac{1}{T^2} \{ \sum_{r=1}^{K} \sum_{s=1}^{g+1} (tr[\bar{f}_i^{(1)} E_{rs} (\Omega_1 \otimes \Sigma_{zz}^{-1}))] tr[\bar{f}_{i,rs}^{(3)} (\Omega_1 \otimes \Sigma_{zz}^{-1})] \\ &\qquad\quad + tr[\bar{f}_i^{(1)} E_{rs} (\Omega_1 \otimes \Sigma_{zz}^{-1}) \bar{f}_{i,rs}^{(3)} (\Omega_1 \otimes \Sigma_{zz}^{-1})] \\ &\qquad\quad + tr[\bar{f}_i^{(1)} E_{rs} (\Omega_1 \otimes \Sigma_{zz}^{-1}) (\bar{f}_{i,rs}^{(3)})' (\Omega_1 \otimes \Sigma_{zz}^{-1})]\} \end{aligned} \tag{16}$$

as $T \to \infty$, which is the final term in the approximation. Gathering terms from 1-4 we have the following theorem:

Theorem 3. *As $T \to \infty$,*

$$
\begin{aligned}
E(\hat{\alpha}_i - \alpha_i)^2 = & \frac{1}{T}\beta_0'\Omega_1\beta_0 e_i'\Sigma_{xx}^{-1}e_i \\
& + \frac{1}{T^2}2\{e_i'\Sigma_{xx}^{-1}HD\beta_0 .tr[(\Sigma_{zz}^{-1}\Sigma_{zFz} - \Sigma_{xx}^{-1}\Sigma_{xFx})] - e_i'\Sigma_{xx}^{-1}\Sigma_{xFx}\Sigma_{xx}^{-1}HD\beta_0\} \\
& + \ \frac{1}{T^2}\frac{1}{4}\{(tr[\bar{f}_i^{(2)}(\Omega_1 \otimes \Sigma_{zz}^{-1})])^2 + 2tr[\bar{f}_i^{(2)}(\Omega_1 \otimes \Sigma_{zz}^{-1})\bar{f}_i^{(2)}(\Omega_1 \otimes \Sigma_{zz}^{-1})]\} \\
& + \frac{1}{T^2}\frac{1}{3}\{\sum_{r=1}^{K}\sum_{s=1}^{g+1}(tr[\bar{f}_i^{(1)}E_{rs}(\Omega_1 \otimes \Sigma_{zz}^{-1})]tr[\bar{f}_{i,rs}^{(3)}(\Omega_1 \otimes \Sigma_{zz}^{-1})] \\
& + \ tr[\bar{f}_i^{(1)}E_{rs}(\Omega_1 \otimes \Sigma_{zz}^{-1})\bar{f}_{i,rs}^{(3)}(\Omega_1 \otimes \Sigma_{zz}^{-1})] \\
& + \ tr[\bar{f}_i^{(1)}E_{rs}(\Omega_1 \otimes \Sigma_{zz}^{-1})(\bar{f}_{i,rs}^{(3)})'(\Omega_1 \otimes \Sigma_{zz}^{-1})]\} + o(\frac{1}{T^2}). \qquad (17)
\end{aligned}
$$

This approximation is of the type given in (3) and so will differ from Nagar's approximation, which was obtained under normality assumptions for the structural disturbances. Nagar's approximation is obtained though if the matrix D is set to zero (thus the disturbances become symmetric) and if $\bar{f}_i^{(1)}$is replaced by $f_i^{(1)}$, $\bar{f}_i^{(2)}$ is replaced by $f_i^{(2)}$, $\bar{f}_{i,rs}^{(3)}$ is replaced by $f_{i,rs}^{(3)}$, while Σ_{xx}^{-1} and Σ_{zz}^{-1} are replaced by $(\frac{1}{T}X^{'}X)^{-1}$ and $(\frac{1}{T}Z'Z)^{-1}$, respectively. Both are large T approximations, valid to order $\frac{1}{T^2}$, and they are effectively equivalent. It follows that the second moment approximation of Nagar is valid under less restrictive conditions than those usually imposed, e.g. normality, though symmetry is required. What is needed, apart from symmetry, is a set of assumptions that ensure that, as $T \to \infty$, $T^{\frac{1}{2}}vec(\hat{\Pi}_1 - \Pi_1)$ converges in distribution to a normally distributed random vector, ζ. The standard condition for this, in terms of the specification of the structural disturbances, is that they be *i.i.d.* with a finite variance. In particular, no restrictions are placed on the higher moments of the disturbances, though for the moment approximation to have a remainder term of the appropriate order we require that the fifth order moments of the disturbances exist, see Phillips (2000). Thus while the structural disturbances are required to be symmetric for the Nagar second moment approximation to hold, their moments higher than the fifth may not even exist. This result is at variance with the second moment approximation presented in Peixe et al (2006) who considered the case of elliptically symmetric disturbances chosen because members of the class may have fat tailed distributions. We find that it is asymmetry and not kurtosis which changes the second moment approximation to order T^{-2}.

When the structural disturbances are specified to be *i.i.d.* with a finite variance, then $2SLS$ has the same well defined asymptotic normal distribution and, hence, the same asymptotic variance, regardless of the distribution of the errors. What the above results indicate is that under symmetry the finite sample variance does not vary greatly with the distribution of the errors either. Of course, the analysis is based upon asymptotic expansions which are less accurate the smaller the sample size but in reasonably large samples, say $\mathbf{T = 50}$, we can expect the results to be valid. At the same time, we should be aware that the accuracy of the approximations will be influenced by the characteristics of the underlying distributions to some degree. An interesting question concerns what happens when the assumption of symmetry is dropped. Now that we have the second moment approximation allowing for asymmetric disturbances we are in a position to address this. There is some evidence to support the proposition that the small sample variance of 2SLS is relatively robust to the distribution of the errors including the case of non-symmetry. For example Raj (1980, p226), using Monte Carlo simulations, examined the performance of several econometric estimators, including $2SLS$, in a simultaneous equation model with Uniform, Normal and Logmormal errors, standardised to have mean zero and variance unity. For all eight parameters estimated it was found that the variance of the $2SLS$ estimator was remarkably stable across the three distributions for a sample size of only 20. The results were obtained based on just 1000 replications as was common at the time, so the degree of accuracy might not satisfy modern standards but, nevertheless, this evidence is of some value. Knight (1985) examined the exact moments of 2SLS and OLS estimators of the endogenous regressor parameter in an equation with two included endogenous variables, in situations where the reduced form disturbances followed a non-normal distribution of the Edgeworth type with specified skewness (λ_3) and kurtosis (λ_4). Knight calculated the relative biases and mean squared errors over a range of values for the skewness and kurtosis, although the necessary conditions for the Edgeworth distribution to be a valid probability distribution imposed the restrictions $-0.5 \leq \lambda_3 \leq 0.5$ and $0.4 \leq \lambda_4 \leq 3.8$. Hence extreme cases of skewness were not covered by his study. Knight concluded that the effects of departures from normality of the error distribution on both bias and MSE for both estimators were very slight and became negligible as the non-centrality parameter increased. Thus the $2SLS$ and OLS estimators were shown to be robust to non-normal error distributions characterised by the Edgeworth family. These results suggest that the extra term in the second moment approximation, which acounts for the asymmetry, has a relatively small effect. This can be checked by direct calculation for a given structure.

In the next section, we consider a set of Monte Carlo experiments which will provide further evidence. We consider the $2SLS$ estimator in the context of symmetric model errors and model errors that are more asymmetric than the values allowed by an Edgeworth distribution, thus checking the MSE robustness conclusion of Knight (1985) for further departures from Normality. We employ a very large number of replications, 15 million, and compare the Monte Carlo simulated MSE against values from the new mean squared error approximation and the approximation obtained under a Normality assumption by Nagar (1959). When the second moment approximation is calculated under asymmetric disturbances, we simply add to the original Nagar approximation the extra term:

$$2\{e_i'(X'X)^{-1}HD\beta_0.tr\{(P_z - P_X)F\} - e_i'(X'X)^{-1}X'FX(X'X)^{-1}HD\beta_0\}$$

which is derived on page 24.

5 Numerical and simulation results

We use the same two-equation models as Liu-Evans and Phillips (2018), who investigate the performance of a higher-order bias approximation in asymmetric cases. We compare the improved MSE approximation with the original due to Nagar and illustrate how skewness in the structural errors in each equation, u_1 and u_2, can have a substantial effect on the MSE of the 2SLS estimator. Consider the simple model given by

$$\begin{aligned} y_{1t} &= \beta_1 y_{2t} + u_{1t} \\ y_{2t} &= \beta_2 y_{1t} + \gamma' z_t + u_{2t} \end{aligned} \tag{18}$$

for $t = 1, \ldots, T$, where $z_t = (1, z_{1t}, z_{2t})'$ is a 3×1 vector of exogenous variables so that the order of overidentification in the first equation is $L = 2$. The terms z_{1t} and z_{2t}, $t = 1, \ldots, T$, are generated once from an $AR(1)$ with zero mean and autoregressive coefficient 0.9, and fixed across Monte Carlo replications. To investigate the effects of skewness in u_1 and u_2, two parameterisations for the structural model coefficients and covariance matrices are fixed, while the skewness values are varied. We use the numerical approach in Phillips and Liu-Evans (2018) to obtain structural disturbances that have the desired covariance matrix and a relatively wide range of positive and negative skewness values. The sample size is set at $T = 50$.

The structural disturbances are generated from

$$U = \mathcal{E} Q', \qquad (u_{it} = \varepsilon_t' Q' e_i) \tag{19}$$

where $\mathcal{E}$ is a $T \times 2$ matrix with row $\varepsilon_t' = (\varepsilon_{it}, \varepsilon_{2t})$ where, for each i, ε_{it} is *i.i.d.* standardised Beta, $t = 1, \ldots, T$, with parameters (α^i, β^i), and where Q is a factor in the Choleski decomposition $\Sigma = QQ'$. It is possible to write the skewness of u_{it} in terms of the skewness of the underlying ε_{it} variables, $\gamma_1(B(\alpha^i, \beta^i)) := \gamma_1(\alpha^i, \beta^i)$. Let γ_1^i denote the skewness of u_{it}, $i = 1, 2$. Then, as shown in Phillips and Liu-Evans (2018),

$$\gamma_1^i = \frac{e_i' Q(tr(N_{ii}'\Theta_1), tr(N_{ii}'\Theta_2))'}{tr(N_{ii})^{\frac{3}{2}}}, \quad i = 1, 2 \tag{20}$$

where $N_{ij} = Q' e_i e_j' Q$, $\Theta_1 = \begin{pmatrix} \gamma_1(\alpha^1, \beta^1) & 0 \\ 0 & 0 \end{pmatrix}$ and $\Theta_2 = \begin{pmatrix} 0 & 0 \\ 0 & \gamma_1(\alpha^2, \beta^2) \end{pmatrix}$. The formulae for σ_{111}, σ_{112} and σ_{122} are similar. Structural disturbance vectors u_1 and u_2 with the desired skewness values γ_1^i are then obtained by choosing the Beta distribution parameters α^1, β^1, α^2 and β^2 appropriately by numerical computation; this was done for each different pair of u_1 and u_2 skewness values. Results for the two coefficient and covariance matrix parameterisations follow, based on 15m replications unless otherwise indicated.

Model A
Here $\beta_1 = 2.732654$, $\beta_2 = -16.388196$, $\gamma = (38.126172, 6.204823, 3.870217)'$,

and

$$\Sigma = \begin{pmatrix} 38.106464 & -11.779951 \\ -11.779951 & 92.106520 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0.31560 & 0.068204 \\ 0.068204 & 5.1107 \end{pmatrix}.$$

Table 1: (Beta) MSE vs approximation, Model A

Skewness values and third moments							
γ_1^1	γ_1^2	σ_{111}	σ_{112}	σ_{122}	σ_{222}	approx. MSE	MSE
-5	5	-1176.2	433.09	-49.951	4419.8	1.2078	1.1035
-3	3	-705.7	259.85	-29.970	2651.9	1.2135	1.1593
-1	1	-235.23	86.617	-9.9901	883.97	1.2191	1.2775
0	0	0	0	0	0	1.2228	1.3119
1	-1	235.23	-86.617	9.9901	-883.97	1.2248	1.3269
3	-3	705.7	-259.85	29.970	-2651.9	1.2305	1.3680
5	-5	1176.2	-433.09	49.951	-4419.8	1.2361	1.4085

Model B
Here $\beta_1 = -3.916203$, $\beta_2 = 47.041079$, $\gamma = (39.838039, -12.940695, -10.705920)'$, and

$$\Sigma = \begin{pmatrix} 13.056401 & 7.482774 \\ 7.482774 & 60.983007 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0.025934 & -0.029021 \\ -0.029021 & 0.86445 \end{pmatrix}.$$

Table 2: (Beta) MSE[1] vs approximation, Model B

Skewness values and third moments							
γ_1^1	γ_1^2	σ_{111}	σ_{112}	σ_{122}	σ_{222}	approx. MSE	MSE
-5	5	235.89	48.68	-152.08	-2381.1	3.1817	*2.3493*
-3	3	141.53	29.208	-91.246	-1428.7	3.2123	*2.6879*
-1	1	47.178	9.736	-30.415	-476.23	3.243	*3.2825*
0	0	0	0	0	0	3.2583	*3.6545*
1	-1	-47.178	-9.736	30.415	-476.23	3.2736	*3.5327*
3	-3	-141.53	-29.208	91.246	-1428.7	3.3012	*3.5705*
5	-5	-235.89	-48.68	152.08	-2381.1	3.3349	*3.4756*

The approximate MSE for Models A and B understates the true magnitude of the changes in MSE, but seems to capture the direction with respect to changes in u_1 and u_2 skewness. We confirm this below using a more extensive set of u_1 and u_2 skewness pairs.

A summary comparison of the MSE
The MSE values corresponding to a grid of u_1 and u_2 skewness values are compared below for Models A and B with the corresponding $O(T^{-2})$ approximate MSE values. We choose 21 sets of (α_1, β_1) to achieve u_1 skewness values in the

[1] These are based on 1m replications as the 15m run has not finished yet.

set $\gamma = \{-5, -4.5, \ldots, -0.5, 0, 0.5, \ldots, 4.5, 5\}$, and do the same for the (α_2, β_2) corresponding to u_2 skewness values. The grid of skewness values for u_1 and u_2 is then a set $S = \gamma \times \gamma$, and we denote by MSE (in italics) a vector of Monte Carlo simulated ("true") MSE values corresponding to members of S, and by $\widetilde{MSE}_{sym}$ and $\widetilde{MSE}_{asym}$ the vectors of corresponding approximate values under symmetry and asymmetry assumptions, respectively. We then compare how near $\widetilde{MSE}_{sym}$ and $\widetilde{MSE}_{asym}$ are to MSE in terms of Euclidean distance: $d^{MSE}_{sym} = ||MSE - \widetilde{MSE}_{sym}||$, $d^{MSE}_{asym} = ||MSE - \widetilde{MSE}_{asym}||$. The terms d^{MSE}_{sym} and d^{MSE}_{asym} are summary performance measures for the approximations over u_1 and u_2 skewness values in the interval $[-5, 5]$, based on a total of $21^2 = 441$ pairs of skewness values.

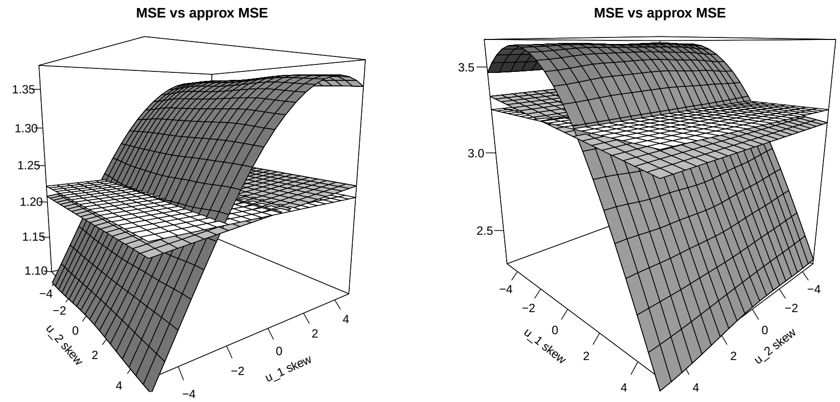
Each of the 441 pairs (α_1, β_1) and (α_2, β_2) was chosen numerically for Models A and B in the same way as earlier, so that the required structural disturbance skewness values were obtained but without changing the structural disturbance covariance matrices, it was also verified in each case that α_1, β_1, α_2 and β_2 were all positive, and that the implied values for the skewnesses were numerically correct to at least the 8th significant figure. The results are presented in Table 3 below.

Table 3: Distance between true and approximate MSE

	d_{sym}	d_{asym}
Model A	2.284503	2.139094
Model B	9.713572	8.695195

The results in Table 3 indicate that the new MSE approximation does better in this overall sense for both Model A and Model B. Figure 1 below plots the Monte Carlo simulated MSE and the MSE approximations for the 21^2 different skewness pairs, for Models A and B. It is clear from these figures that the MSE can change substantially when the skewnesses of the structural disturbances are varied, and that the approximations capture part of this nonlinear effect.

Figure 1: Monte Carlo simulated MSE vs $O(T^{-2})$ approximate MSE



5.1 Conclusions

This paper has demonstrated that the well known second moment approximation of Nagar, obtained to order $O(T^{-2})$ under a Normality assumption, continues to be valid under symmetric but not necessarily Normal disturbuances. Moreover, the result of Nagar has been extended to the case of asymmetric structural errors via an additional term that captures the effect of asymmetry.

Our results suggest that skewness in the error distribution has a greater effect on the 2SLS mean squared error than previously thought, e.g. by Knight (1985). This complements a similar finding in Liu-Evans and Phillips (2018) for the higher-order 2SLS bias, where it is shown that skewness can have a substantial effect despite being of order $O(T^{-2})$. More generally, it appears that asymmetry is a potentially important consideration when developing refined asymptotic procedures such as those discussed in Phillips and Tzavalis (2007).

References

Anderson, T.W., N. Kunimoto, and K. Morimune (1986). Comparing Single Equation Estimators In A Simultaneous Equation System, *Econometric Theory*, 2, 1-32.

Davidson,R and J.G. Mackinnon (2006). The Case Against JIVE, *Journal of Applied Econometrics* 21, 827-833.

Donald, S.G. and W.K. Newey (2001). Choosing the Number of Instruments, *Econometrica* 69, 1161-1191.

Hahn,J. J.A.Hausman and G. Kuersteiner (2004). Estimation with weak instruments. Accuracy of higher-order bias and MSE approximations, *Econometrics Journal* 71, 272-306.

Harvey,A.C. and G.D.A.Phillips(1980). Testing for Serial Correlation in Simultaneous Systems, *Econometrica* 48(3), 747-759.

Iglesias,E.M and G.D.A.Phillips (2012). Almost Unbiased Estimation in Simultaneous Equations with Strong and/or Weak Instruments, *Journal of Business and Economic Statistics* 30(4), 505-520.

Kinal, T. W. (1980). The Existence of Moments of k-Class Estimators, *Econometrica* 48, 1, 241-250.

Liu-Evans, G. and Phillips, G. D. A. (2018). On the use of higher order bias approximations for 2SLS and k-class estimators with non-normal disturbances and many instruments. *Econometrics and Statistics*, 6, 90-105.

Mikhail W. M. (1972). The Bias of the Two-Stage Least Squares Estimator, *Journal of the American Statistical Association* 67, 625-627.

Kiviet, J and G.D.A. Phillips (2002). Improved Coefficient and Variance Estimation in stable first-order dynamic regression models. Under revision.

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

Magnus, J and H. Neudecker (1979). The commutation matrix: some properties and applications. *The Annals of Statistics* 7, 381-394.

Nagar, A. L. (1959). The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations. *Econometrica* 27, 575-95.

Peixe, F.P.M., A.R. Hall, and K. Kyriakoulis (2006). The Mean Squared Error Of The Instrumental Variables estimator when the disturbance has an elliptical distribution, *Econometric Reviews* 25, 117-138.

Phillips, G. D. A. (2000). An Alternative Approach to Obtaining Nagar-Type moment Approximations in Simultaneous Equation models, *Journal of Econometrics* 97(2), 345-364.

Phillips, G. D. A. (2007). Nagar-type Moment Approximations in Simultaneous Equation Models: Some Further Results. In G.D.A. Phillips and E. Tzavalis (Eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Cambridge: Cambridge University Press.

Phillips, G.D.A. and E. Tzavalis(2007). *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Cambridge: Cambridge University Press.

Raj, B. (1980). A Monte Carlo Study of Small-Sample Properties of Simultaneous Equation Estimators With Normal and Nonnormal Disturbances. *Journal of the American Statistical Association* 75, 221-229.

Sargan, J. D. (1974). The Validity of Nagar's Expansion for the Moments of Econometric Estimators, *Econometrica* 42, 169-176.

Srinavasan, T.N. (1970). Approximations to the Finite Sample Moments of Estimators Whose Finite sample Distributions are Unknown, *Econometrica* 38, 533-541.

Appendix

Note 1

In this note, which was referred to above in the main part of the Appendix, we evaluate

$$E(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(1)})(vec(\hat{\Pi}_1 - \Pi_1)' f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1)), \tag{21}$$

which may be written as

$$E[(vecV_1)'(I \otimes Z(Z'Z)^{-1}) f_i^{(1)} (vecV_1)'(I \otimes Z(Z'Z)^{-1}) f_i^{(2)} (I \otimes (Z'Z)^{-1} Z')(vecV_1)]. \tag{22}$$

First we note that

$$\begin{aligned}(vecV_1)'(I \otimes Z(Z'Z)^{-1}) f_i^{(1)} &= (vecV_1)'(\beta_0 \otimes XQe_i) \\ &= -v_1' XQe_i + \beta_1 v_2' XQe_i + ... + \beta_g v\prime_{g+1} XQe_i \\ &\quad \text{since} \quad \beta_0 = (-1, \beta')\end{aligned}$$

where v_i, the i^{th} column of V_1, is $T \times 1$ for $i = 1, 2, ..., g+1$. Next we define

$$vecV_1(vec(V_1)' = \begin{pmatrix} v_1 v_1' & v_1 v_2' & v_1 v_3' & ... & v_1 v_{g+1}' \\ v_2 v_1' & v_2 v_2' & v_2 v_3' & ... & v_2 v_{g+1}' \\ v_3 v_1' & v_3 v_2' & v_3 v_3' & ... & v_3 v_{g+1}' \\ . & . & . & . & . \\ v_{g+1} v_1' & v_{g+1} v_2' & ..... & .... & v_{g+1} v_{g+1}' \end{pmatrix},$$

which is a $(T \times (g+1)) \times (T \times (g+1))$ matrix. It is our purpose to evaluate

$$\begin{aligned}&E\{vecV_1(vecV_1)'(vecV_1)'(\beta_0 \otimes XQe_i)\} \\ &\qquad = E\{-v_1' XQe_i + \beta_1 v_2' XQe_i + ... + \beta_g v\prime_{g+1} XQe_i) vecV_1 (vecV_1)'\}.\end{aligned}$$

The general "term" of $E[-vecV_1(vecV_1)'(v_1' XQe_i]$ is the $T \times T$ matrix

$$\begin{aligned}E(-v_i v_j' v_1' XQe_i) &= -E\{\begin{pmatrix} v_{i1} v_{j1} & v_{i1} v_{j2} & v_{i1} v_{j3} ... & v_{i1} v_{jT} \\ v_{i2} v_{j1} & v_{i2} v_{j2} & v_{i2} v_{j3} ... & v_{i2} v_{jT} \\ . & . & . & . \\ v_{iT} v_{j1} & v_{iT} v_{j2} & .... & v_{iT} v_{jT} \end{pmatrix} v_1' XQe_i\} \\ &= E\{\begin{pmatrix} -v_{i1} v_{j1} & 0 & .0.... & 0 \\ 0 & -v_{i2} v_{j2} & 0 & 0 \\ .. & 0 & .. & \\ 0 & . & . & -v_{iT} v_{jT} \end{pmatrix} v_1' XQe_i\}\end{aligned} \tag{23}$$

since the off diagonal terms will involve products of three reduced form disturbances that are not all of the same time period. On putting $XQe_i = (x_1, ..., x_T)' Qe_i$, where x_j' is the j^{th} row of X for $j = 1, 2, .., T$, it is seen that

$$-v_1' XQe_i = -v_{11} x_1' Qe_i \ - v_{12} x_2' Qe_i \ - .... - v_{1T} x_T' Qe_i. \tag{24}$$

We may now write the required expectation in (6.22) as:

$$E\{\begin{pmatrix} -v_{i1}v_{j1}v_{11}x_1'Qe_i & 0 & .0.. & 0 \\ 0 & -v_{i2}v_{j2}v_{12}x_2'Qe_i & & 0 \\ .0. & 0 & .. & .. \\ 0 & . & .0 & -v_{iT}v_{jT}v_{1T}x_T'Qe_i \end{pmatrix}$$

$$= -\omega_{1ij}\begin{pmatrix} x_1'Qe_i & & & \\ & x_2'Qe_i & & \\ & & .. & \\ & & & x_T'Qe_i \end{pmatrix},$$

since $E(v_{it}v_{jt}v_{1t}) = \omega_{1ij}$ for $t = 1, 2, ..., T$. In a similar way

$$E[\beta_1 v_2' XQe_i v_i v_j'] = \beta_1 \omega_{2_{ij}} \begin{pmatrix} x_1'Qe_i & & & \\ & x_2'Qe_i & & \\ & & .. & \\ & & & x_T'Qe_i \end{pmatrix},$$

and a similar result goes through for the remaining terms in $E\{-v_1'XQe_i + \beta_1 v_2' XQe_i + ... + \beta_g v\prime_{g+1} XQe_i) vecV_1 (vecV_1)'\}$. We have thus shown that the $T(g+1) \times T(g+1)$matrix of interest, $E\{vecV_1(vecV_1)'(vecV_1)'(\beta_0 \otimes XQe_i)\}$, has a general matrix term given by

$$(-\omega_{1ij} + \beta_1\omega_{2ij} + ... + \beta_g\omega_{g+1,ij}) \begin{pmatrix} x_1'Qe_i & & & \\ & x_2'Qe_i & & \\ & & .. & \\ & & & x_T'Qe_i \end{pmatrix},$$

and there are $(g+1)^2$ such matrices.

Let D be the $(g+1) \times (g+1)$ matrix where $\{D_{ij}\} = -\omega_{1ij} + \beta_1\omega_{2ij} + ... + \beta_g\omega_{g+1,ij}$. Then $E\{vecV_1(vecV_1)'(vecV_1)'(\beta_0 \otimes XQe_i)\}$ can be written as $D \otimes F$ where

$$F = \begin{pmatrix} x_1'Qe_i & & & \\ & x_2'Qe_i & & \\ & & .. & \\ & & & x_T'Qe_i \end{pmatrix}. \tag{25}$$

We now have all the terms we need to evaluate the required term in the second moment approximation which can be expressed as:

$$tr\{[I \otimes Z(Z'Z)^{-1}]f_i^2[I \otimes (Z'Z)^{-1}Z'][D \otimes F]\} \tag{26}$$

To proceed we note that, by direct multiplication,

$$[I \otimes Z(Z'Z)^{-1}]f_i^{(2)}[I \otimes (Z'Z)^{-1}Z'] =$$
$$H'(X'X)^{-1}e_i\beta_0' \otimes (P_z - P_x) - [H'X(X'X)^{-1} \otimes X(X'X)^{-1}e_i\beta_0']I^*$$
$$+ \beta_0 e_i'(X'X)^{-1}H \otimes (P_z - P_x) - [\beta_0 e_i'(X'X)^{-1}X' \otimes X(X'X)^{-1}H]I^{'*},$$

and from this $tr\{[I \otimes Z(Z'Z)^{-1}]f_i^{(2)}[I \otimes (Z'Z)^{-1}Z'][D \otimes F]\}$ has four components given by

$$
\begin{aligned}
& tr\{[H'(X'X)^{-1}e_i\beta_0' \otimes (P_z - P_x)][D \otimes F] \\
& - tr\{[H'X(X'X)^{-1} \otimes X(X'X)^{-1}e_i\beta_0']I^*[D \otimes F] \\
& + tr\{[\beta_0 e_i'(X'X)^{-1}H \otimes (P_z - P_x)][D \otimes F]\} \\
& - tr\{[\beta_0 e_i'(X'X)^{-1}X' \otimes X(X'X)^{-1}H]I^{'*}[D \otimes F]\} \\
= & \, tr\{H'(X'X)^{-1}e_i\beta_0'D\}tr\{(P_z - P_x)F\} \\
& - tr\{DH'(X'X)^{-1}X'FX(X'X)^{-1}e_i\beta_0'\} \\
& + tr\{\beta_0 e_i'(X'X)^{-1}HD\}tr\{(P_z - P_x)F\} \\
& - tr\{D\beta_0 e_i'(X'X)^{-1}X'FX(X'X)^{-1}H\},
\end{aligned}
$$

which simplifies to yield the result

$$
\begin{aligned}
& E(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(1)})(vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1)) \\
& \qquad = 2\{e_i'(X'X)^{-1}HD\beta_0\}tr\{(P_z - P_x)F\beta_0 - e_i'(X'X)^{-1}X'FX(X'X)^{-1}HD\beta_0\},
\end{aligned}
$$

which is $O(T^{-2})$.

Note 2

Here an alternative form is presented for $HD\beta_0$. This was mentioned in the main part of the Appendix. The expected values in Tables 3 and 4 refer to final form of $HD\beta_0$ presented in this note. We have seen that $u_{1t} = -v_{1t} + \beta_1 v_{2t} + \beta_2 v_{3t} + ... + \beta_{g+1} v_{gt}$, so that $D_{ij} = E(u_{1t} v_{it} v_{jt})$. The first row of D is given by $E(u_{1t}v_{1t}v_{1t}, u_{1t}v_{1t}v_{2t}, u_{1t}v_{1t}v_{3t}, ..., u_{1t}v_{1t}v_{g+1,t})$. Therefore the first component of $E\{D\beta_0\}$ is $E\{(u_{1t}v_{1t})(v_{1t}, v_{2t}, v_{3t}, ..., v_{g+1,t})\beta_0\}$, where $(v_{1t}, v_{2t}, v_{3t}, ..., v_{g+1,t})\beta_0 = u_{1t}$. It follows that the first component of the vector $E\{D\beta_0\}$ is $E\{u_{1t}^2 v_{1t}\}$. By a similar argument we can state that the r^{th} component is $E\{u_{1t}^2 v_{rt}\}$, $r = 1, 2, ..., g+1$. Hence we have shown that $E\{D\beta_0\} = E\{u_{1t}^2 v_1^*\}$ where $(v_1^*)'$ is the first row of V_1.

Noting that $(v_1^*)'$, the first row of V_1 is equal to $(u_1^*)' B_{g+1}^{-1}$, where $(u_1^*)'$ is the first row of the matrix U and B_{g+1}^{-1} is formed from the first $g+1$ columns of B^{-1}, it is clear that $v_1^* = (B_{g+1}^{-1})' u_1^*$.

Finally we have that $E\{D\beta_0\}$ is given by

$$\begin{aligned} E\{u_{1t}^2 v_1\} &= E\{(B_{g+1}^{-1})' u_{1.} u_{1t}^2\} \\ &= E\{(B_{g+1}^{-1})'(u_{1t}^3, u_{1t}^2 u_{2t}, u_{1t}^2 u_{3t}, .., u_{1t}^2 u_{Gt})'\} \end{aligned} \tag{27}$$

Hence it is required to find $(B_{g+1}^{-1})' E(u_{1t}^3, u_{1t}^2 u_{2t}, u_{1t}^2 u_{3t}, .., u_{1t}^2 u_{Gt})'$. In fact, we need H times this vector where $H = \begin{bmatrix} 0 & I_g \\ 0 & 0 \end{bmatrix}$, which has dimension $(g+k) \times (g+1)$.