CARDIFF BUSINESS SCHOOL
WORKING PAPER SERIES

Cardiff Economics Working Papers

Garry DA Phillips and Gareth Liu-Evans

The Robustness of the Higher-Order 2SLS and General k-Class Bias Approximations to Non-Normal Disturbances

E2011/20

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
www.cardiff.ac.uk/carbs

ISSN 1749-6101
August 2011

The Robustness of the Higher-Order 2SLS and General k-Class Bias Approximations to Non-Normal Disturbances

Garry DA Phillips
The Business School, Cardiff University

Gareth Liu-Evans
School of Management, Liverpool University

August 2011

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.Later Mikhail (1972) obtained a higher-order bias approximation for 2SLS under the same assumptions as Nagar while Iglesias and Phillips (2010) obtained the higher order approximation for the general k-class of estimators. These approximations show that the higher order biases can be important especially in highly overidentified cases. In this paper we show that Mikhail's higher order bias 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 2SLS bias is then obtained which includes the case of asymmetric disturbances. The results are then extended to the general $k-$class of estimators.

1 Introduction

Moment approximations of estimators in simultaneous equation models have a long history. The seminal paper was Nagar (1959) who derived approximations to the first and second moments of the consistent k-class of estimators in a general simultaneous equation model with exogenous regressors. In obtaining the results, it was assumed that the structural disturbances were independently and normally distributed. Later Mikhail (1972) extended Nagar's bias approximation for the $2SLS$ case to a higher order and under the same assumptions while Iglesias and Phillips (2010) give the higher order approximation for the consistent k-class estimator. 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. However, a paper which is of particular relevance, given its generality, is Kinal (1980). His results show that in the general simultaneous equation 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. In cases where $k > 1$, which includes the $LIML$ estimator, the k-class 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 nor, indeed, symmetry. 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 $ARCH/GARCH$. Neither paper considered the higher order approximation however,

In this paper it is shown that the Mikhail higher order bias approximation is valid without assuming normality for the disturbances. It does, however, require that the disturbances are distributed symmetrically. If disturbances have a skewed distribution then the approximation has to be modified and we give the corrected form.. The results are then extended to include the consistent nembers of the $k-$class.

2 Model and Notation

We consider a simultaneous equation model given by

$$By_t + \Gamma z_t = u_t \tag{1}$$

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 and identically distributed structural disturbances with $G \times G$ positive definite covariance matrix Σ. The matrices of structural parameters, 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 (1) are:

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

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

$$YB' + Z\Gamma' = U. \tag{2}$$

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.

The first equation of the system is given by

$$y_1 = Y_2\beta + Z_1\gamma + u_1, \tag{3}$$

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 and identically distributed disturbances with positive definite covariance matrix $E(u_1u_1') = \Sigma_{11}$. 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 and identically distributed with zero mean vector and (g+1)$\times$(g+1) positive definite covariance matrix $\Omega_1 = (\omega_{ij})$ while the $T(g + 1)$ vector $vecV_1$, obtained by stacking the columns of V_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 fifth order. This latter condition is required to ensure that the expansion used has a remainder term of appropriate order, see Phillips (2000). It is further assumed that:

(i) Equation (3) 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. This over-identifying restriction is sufficient to ensure that the Nagar expansion is valid in the case considered by Nagar and that, at least, the first estimator moment exists: 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 Nagar Approximations to the bias

The $2SLS$ estimator of $\alpha = (\beta', \gamma')'$is given by

$$\hat{\alpha} = \begin{pmatrix} Y_2'Y_2 - \hat{V}_2'\hat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2 - \hat{V}_2 \\ Z_1' \end{pmatrix}' y_1. \tag{4}$$

The Nagar approach to finding moment approximations for the 2SLS estimator, proceeds from the estimation error,

$$\hat{\alpha} - \alpha = \begin{pmatrix} Y_2'Y_{2-}\hat{V}_2'\hat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2 - \hat{V}_2 \\ Z_1' \end{pmatrix}' u_1, \tag{5}$$

To find the expansion we first write

$$\begin{aligned} \hat{\alpha} - \alpha &= \left[Q^{-1} + X'V_z + V_z'X + V_z'P_zV_z\right]^{-1}\left[X'u_1 + V_z'P_zu_1\right] \qquad (6) \\ &= \left[I + Q^{-1}\{X'V_z + V_z'X + V_z'P_zV_z\}\right]^{-1} Q\left[X'u_1 + V_z'P_zu_1\right]. \end{aligned}$$

$X = (Z\Pi_2 : Z_1)$, $Q = (X'X)^{-1}$, $P_z = Z(Z'Z)^{-1}Z'$ and $V_z = (V_2 : 0)$.

Setting $\Delta = X'V_z + V_z'X + V_z'P_zV_z$ and expanding the inverse $\left[I + Q^{-1}\Delta\}\right]^{-1}$ in a Taylor expansion yields

$$\begin{aligned} \hat{\alpha} - \alpha &= [I + Q\Delta\}]^{-1} Q\left[X'u_1 + V_z'P_zu_1\right] \\ &= [I - Q\Delta + Q\Delta Q\Delta - +....]\, Q\left[X'u_1 + V_z'P_zu_1\right] \end{aligned} \tag{7}$$

where terms can be arranged in decreasing order of stochastic magnitude. In fact, if we write $u_1 = V_1\beta_0$ and $V_z = V_1H'$, 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

$$\begin{aligned} \hat{\alpha} - \alpha = {} & QX'V_1\beta_0 + QHV_1'P_zV_1\beta_0 - QX'V_1H'QX'V_1\beta_0 - QHV_1'P_xV_1\beta_0 \\ & - QHV_1'P_zV_1H'QHV_1\beta_0 - QHV_1'P_zV_1H'QX'V_1\beta_0 \\ & - QX'V_1H'QHV_1'P_zV_1\beta_0 - QHV'XQHV'P_zV_1\beta_0 \\ & + QX'HV_1'QXV_1H'QX'V_1\beta_0 + QHV_1'P_xV_1H'QXV_1\beta_0 \\ & + QX'V_1HQHV_1'P_xV_1\beta_0 + QHV_1'XQHV_1'P_xV_1\beta_0 + o_p(T^{-\frac{3}{2}}). \end{aligned} \tag{8}$$

The Nagar bias approximation is found by summing the expectations of the terms up to order T^{-1}.We shall later compare this expansion with an alternative representation presented in Phillips (2000)., 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 approximation for the bias of the $2SLS$ estimator for α in (4) is given by

$$E(\hat{\alpha} - \alpha) = [L - 1]Qq + o(T^{-1}). \tag{9}$$

where $L = K - g - k$ is the order of overidentification, $q = \frac{1}{T}\begin{bmatrix} E(V_2'u_1) \\ 0 \end{bmatrix}$ and Q is as defined above.

The Mikhail higher order approximation for the $2SLS$ estimator for α in (4), in the same framework as Nagar but extending the expansion to include terms up to $O_p\left(T^{-2}\right)$, is given by

$$E(\hat{\alpha}-\alpha)=(L-1)[I+tr(QC)I-(L-2)QC]Qq\ +o(1/T^2). \tag{10}$$

which adds two terms to Nagar's result, namely, $(L-1)(tr(QC)Qq$ and -$(L-1)(L-2)QCQq$, both of which are $O(T^{-2})$. The $(g+k)\times(g+k)$ matrix C above is given by

$$C=\left[\begin{array}{cc}(1/T)E(V_2'V_2) & 0\\ 0 & 0\end{array}\right].$$

It is apparent that when L is relatively large these added terms can be important. Hence in models with a large number of intruments the higher order approximation will be of particular value. Some evidence for this is given in Iglesias and Phillips (2008).

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; in particular, we shall show that the assumption of normality for disturbances can be relaxed. We shall also consider how the approximation is modified when the disturbances are asymmetric. 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, see Hadri and Phillips (1999) and Iglesias and Phillips (2008) for evidence of this.

In Iglesias and Phillips (2010) Mikhail's approximation was extended, under the same assumptions as used by Mikhail, to include the general $k-$class of estimators as follows:

$$\begin{aligned}E(\hat{\alpha}_k-\alpha) \quad &= \quad (L-1-\theta)Qq+(L-1)-2\theta)trQC.Qq\\ &\quad -[(L-1)(L-2)-\theta 2(L-2)+\theta^2]QCQq\\ &\quad +\theta\frac{K}{T}Qq+o(T^{-2})\end{aligned} \tag{11}$$

Here $k=1+\frac{\theta}{T}$ and $\theta\leq 0$ is a real negative number. Notice that this approximation reduces to that of Mikhail when $\theta=0$ for then the $k-$class estimator is just $2SLS$.

4 An Alternative Approach to Approximating The 2SLS Bias

We consider the estimation of the equation given in (3) by the method of $2SLS$. It is well known that the estimator can be written in the form

$$\hat{\alpha}=\left(\begin{array}{c}\hat{\beta}\\ \hat{\gamma}\end{array}\right)=\left(\begin{array}{cc}\hat{\Pi}_2'Z'Z\hat{\Pi}_2 & \hat{\Pi}_2'Z'Z_1\\ Z_1'Z\hat{\Pi}_2 & Z_1'Z_1\end{array}\right)^{-1}\left(\begin{array}{c}\hat{\Pi}_2'Z'Z\hat{\pi}_1\\ Z_1'Z\hat{\pi}_1\end{array}\right) \tag{12}$$

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 to order $T^{-1}$was found using the first two terms of the expansion:

$$\begin{aligned} f_i(vec\hat{\Pi}_1) \quad = \quad & f_i(vec\Pi_1) + (vec(\hat{\Pi}_1 - \Pi_1)'f_i^{(1)} \\ & + \frac{1}{2!}(vec(\hat{\Pi}_1 - \Pi_1))'f_i^{(2)}(vec(\hat{\Pi}_1 - \Pi_1)) \\ & + \frac{1}{3!}\Sigma_{r=1}^{K}\Sigma_{s=1}^{g+1}(\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1))'f_{i,rs}^{(3)}(vec(\hat{\Pi}_1 - \Pi_1)) \\ & + \frac{1}{4!}F(vec((\hat{\Pi}_1 - \Pi_1)) + o_p(T^{-2}). \end{aligned} \tag{13}$$

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, $\frac{\partial^2 f_i}{\partial vec\hat{\Pi}_1(\partial vec\hat{\Pi}_1)'}$, $f_{i,rs}^{(3)}$ is a $(K(g+1)) \times (K(g+1))$ matrix of third-order partial derivatives defined as $f_{i,rs}^{(3)} = \frac{\partial f_i^{(2)}}{\partial \pi_{rs}}$, $r = 1, ..., K$, $s = 1, ..., g+1$. The derivatives, $f_i^{(1)}, f_i^{(2)}$ and $f_{i,rs}^{(3)}$ are given in Phillips(2000). The expression $F(vec((\hat{\Pi}_1 - \Pi_1))$ represents the unknown fourth term which will involve the fourth order partial derivatives and products of four components of $vec((\hat{\Pi}_1 - \Pi_1)$. All derivatives are evaluated at $vec\Pi_1$.

The bias approximation to order T^{-1} is 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;

$$E(\hat{\alpha}_i - \alpha_i) = e_i'Qq + o(T^{-1}). \tag{14}$$

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 so that the row vectors are serially uncorrelated and homoscedastic.

It is not immediately obvious that the above expansion in (13) is equivalent to that used by Nagar. Examining the Nagar expansion in (8), 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 (8). 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, equals the second term in (8), and so on.

To find the bias approximation to order T^{-2} we shall also need the next two terms in the expansion. It has proved possible to find an explicit representation for the third term but it is quite difficult to do so for the fourth term. Notice that the third term term

$$\frac{1}{3!}\Sigma_{r=1}^{K}\Sigma_{s=1}^{g+1}(\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1))'f_{i,rs}^{(3)}(vec(\hat{\Pi}_1 - \Pi_1) \qquad (15)$$

is a linear function of products of three components of $vec(\hat{\Pi}_1 - \Pi_1)$ and the bounded third order derivatives which are evaluated at $vec(\Pi_1)$. In Appendix 1 it is shown that the third moment of the least squares regression estimator is $O(T^{-2})$ from which we may deduce that the expectation of the third term in (13) is also $O(T^{-2})$ and we evaluate this in Appendix 2.

While we cannot easily find an explicit representation for the fourth term in the expansion, $F(vec((\hat{\Pi}_1 - \Pi_1))$, it turns out that we do not need to do so. We may readily deduce that it is a linear function of fourth order products of the components of $vec(\hat{\Pi}_1 - \Pi_1)$ and the bounded fourth order derivatives evaluated at $vec(\Pi_1)$.We find that not knowing its precise form is of no consequence in context because the fourth moment of the least squares regression estimator does not depend upon the kurtosis of the error distribution to the order of the approximation. This is shown in Appendix 2 where we demonstrate that the fourth moment has two components. The first of these is $O(T^{-2})$ while the second, which involves the kurtosis of the error distribution is $O(T^{-3})$ and, as such, it plays no role in our approximation to $O(T^{-2})$. Because of this the expectation of the fourth term to the order of the approximation will not depend upon the actual distribution of the errors provided the moment conditions are met. Hence the expectation based upon the normal distribution, which has already been found by Mikhail, will also be appropriate for other distributions and in finding the higher order bias approximation to order T^{-2} we shall simply add the relevant part of the Mikhail result. We shall see that the analysis can also be extended to find similar results for the $k-$class of estimators.

5 The Higher Order Bias Approximations For 2SLS

In this section we present the bias approximation under weaker conditions than those assumed by Mikhail. In case the disturbances are non-normal but symmetric, the evaluation of the expected value of the third term in (13) is trivally zero while the evaluation of the fourth term has already been done by Mkihail for the normal distribution and. as noted above, the same evaluation will apply here also. Hence the Mikhail approximation carries over directly for non-normal but symmetric distributions for which the moment conditions are met and does not depend upon kurtosis. We now state the following theorem.

.

Theorem 1

In the model of Section 2 where the errors are symmetrically but not necessarily normally distributed, the bias of the i^{th} component of the $2SLS$ estimator in (4), is given by

$$\begin{aligned} E(\hat{\alpha}_i - \alpha_i) = (L-1)[e_i'Qq + tr(QC)e_i'Qq \\ -(L-2)e_i'QCQq] \quad +o(T^{-2}),\ i = 1, 2, .., g+k. \end{aligned}$$

This is exactly the approximation found by Mikhail for the case of normally distributed errors and the proof of the theorem follows immediately from the preceding discussion. This result helps to explain the findings of Knight(1985) who, using exact finite sample theory in the context of a two equation model, found that a moderate level of kurtosis had little effect on the bias of the $2SLS$ estimator.

The second case of interest is where the errors are asymmetrically distributed. Now it is necessary to extend the Mikhail approximation to allow for asymmetry but, again, the approximation does not depend upon the kurtosis of the error distribution. Introducing the evaluation of the third term of (13) we find that the revised approximation is given in the following.

:

Theorem 2

In the model of Section 2 where the errors may be asymmetrically distributed, the bias of the i^{th} component of the $2SLS$ estimator is given by

$$\begin{aligned} E(\hat{\alpha}_i - \alpha_i) \quad = \quad & (L-1)[e_i'Qq + tr(QC).e_i'Qq - (L-2)e_i'QCQq] \\ & e_i'QH(\beta_0' \otimes I_{g+1})\Omega^* H'QX'\Delta_{x,z} + e_i'((QH(I_{g+1} \otimes \beta_0')\Omega^* H' + \\ & tr(QH(I_{g+1} \otimes \beta_0')\Omega^* H').I_{g+k})QX'\Delta_{xz} \\ & -tr((I_{g+1} \otimes \beta_0')\Omega^* \ H'QX'Diag(XQe_i)XQH) \ + o(T^{-2}) \end{aligned}$$

where the effects of the asymmetry of the disturbances are indicated by the presence of the $(g+1)^2 \times (g+1)$ matrix of third moments Ω^*which is obtained by stacking the $(g+1) \times (g+1)$ matrices $\Omega_{ijs}, s = 1, ., (g+1)$.The $T \times 1$ vector Δ_{xz} has p^{th} component $x_p'(X'X)^{-1}x_p - z_p'(Z'Z)^{-1}z_p$, $p = 1, 2, ..., T$ .When

Ω^* is zero the bias approximation reduces to that of Mikhail (1972) The proof of the above is given in Appendix 2.

Notice that the asymmetry effect does not depend explicity on L and so it is present whatever the order of overidentification; in particular, the asymmetry effect does not go to zero when L=1.

If it is required to express the asymmetry effect in terms of the structural parameters we can replace $\Omega'^*$with its structural parameter representation.viz, $\Omega'^* = ((B')^{-1}_{g+1})'\Sigma^\star((B')^{-1}_{g+1} \otimes (B')^{-1}_{g+1})$ where $(B')^{-1}_{g+1}$ comprises the first $g+1$ columns of $(B')^{-1}$, $\Sigma^*$is the $G \times G^2$ matrix formed as $\Sigma^* = (\Sigma_{ij1}, \Sigma_{ij2}, ..., \Sigma_{ijG})$ and Σ_{ijk} is a $G \times G$ symmetric matrix with general element equal to the third moment $\sigma_{ijk}, k = 1, .., G$.

It is apparent that the asymmetry effect is a complicated function of the endogenous variable parameters in the model and all the third moments of the structural disturbances. As such it is difficult to deduce its sign or magnitude in general though it is possible to calculate the value of the approximation for a given structure. The study by Knight (1985) referred to above also examined the effect of error skewness on the bias of $2SLS$ and found that a moderate degree of skewness appeared to have only a small effect; however, we have no results for substantial departures from symmetry nor, indeed, for cases with a large number of instruments.

6 An extension to the general k-class of estimators

In a recent paper Iglesias and Phillips (2010) have derived the higher order bias of the consistent $k-$class of estimators, thus extending the result for $2SLS$ in Mikhail (1972). This class of estimators for which $0 < k < 1$ is potentially interesting because the estimators have all necessary moments, see Kinal (1980), whereas, for example, $2SLS$ only has moments up to the order of overidentification. To obtain the higher order bias of the general $k-$class estimator under skewness and kurtosis, we shall need to modify the above approach. Consider the $k-$class estimator given by

$$\begin{aligned}
\hat{\alpha}_k &= \begin{pmatrix} \widehat{\beta}_k \\ \hat{\gamma}_k \end{pmatrix} = \begin{pmatrix} Y_2'Y_2 - k\hat{V}_2'\hat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2' - k\hat{V}_2' \\ X_1' \end{pmatrix} y_1 \\
&= \begin{pmatrix} \hat{\Pi}_2'Z'Z\hat{\Pi}_2 + (1-k)\hat{V}_2'\hat{V}_2 & \hat{\Pi}_2'Z'Z_1 \\ Z_1'Z\hat{\Pi}_2 & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\Pi}_2'Z'Z\hat{\pi}_1 + (1-k)\hat{V}_2'\hat{v}_1 \\ Z_1'Z\hat{\pi}_1 \end{pmatrix}.
\end{aligned}$$

Here it is clear that, conditional on the exogenous variables, the $k-$class estimators are functions not only of $\hat{\Pi}_1 = (\hat{\pi}_1 : \hat{\Pi}_2)$ but also of $\hat{V}_2'\hat{V}_2$ and $\hat{V}_2'\hat{v}_1$.

However by suitably manipulating the estimator it is possible to express **it** in the same form as (12) so that no new analysis, beyond that set out in Phillips (2000), is required to find the bias approximation.

To see this let $W = (Z : \bar{Z})$ be a $T \times T$ matrix of rank T obtained by augmenting the X matrix and adding $T - K$ linearly independent columns $\bar{Z}$. Then it is possible to write

$$Y_1 = (y_1 : Y_2) = W(\pi_1^* : \Pi_2^*),$$

where

$$(\pi_1^* : \Pi_2^*) = (W'W)^{-1}W'((y_1 : Y_2) = W^{-1}(y_1 : Y_2).$$

Let

$$W^* = [I - c(I - P_z)]\, W,$$

where $c = 1 + \sqrt{1-k}$. The corresponding $k-$class estimator may then be written as

$$\hat{\alpha}_k = \begin{pmatrix} \widehat{\beta}_k \\ \hat{\gamma}_k \end{pmatrix} = \begin{pmatrix} \Pi_2^{*\prime}(W^*)'W^*\Pi_2^* & \Pi_2^{*\prime}(W^*)'W_1^* \\ (W_1^*)'W^*\Pi_2^* & (W_1^*)'W_1^* \end{pmatrix}^{-1} \begin{pmatrix} \Pi_2^{*\prime}(W^{*\prime})'W^*\pi_1^* \\ (W_1^*)'W^*\pi_1^* \end{pmatrix}, \tag{16}$$

where W_1^* is a $T \times k_1$ matrix forming the first k_1 columns of W^*. Thus $W_1^* = Z_1$.

Now putting

$$\alpha_k = \begin{pmatrix} \beta_k \\ \gamma_k \end{pmatrix},$$

we may write that, conditional on W^*,

$$\alpha_k = f^*(vec\Pi_1^*),$$

where $E(\Pi_1^*) = \bar{\Pi}_1 = \begin{pmatrix} \Pi_1 \\ 0 \end{pmatrix}$ and the 0 matrix is $(T-K) \times (g+1)$.

Also,

$$\begin{aligned} W^*\bar{\Pi}_1 &= [I - c(I - P_Z)]\, W\bar{\Pi}_1 = [I - c(I - P_Z)]\,(Z : \bar{Z}) \begin{pmatrix} \Pi_1 \\ 0 \end{pmatrix} \\ &= [I - c(I - P_Z)]\, Z\Pi_1 \ = Z\Pi_1 = (Z\pi_1 : Z\Pi_2)\,. \end{aligned}$$

On putting $\left(\bar{\pi}_1 : \quad \bar{\Pi}_2 \right)$ in place of $\left(\pi_1^* : \quad \Pi_2^* \right)$ in (16), it is seen that

$$f^*(vec\bar{\Pi}_1) = f(vec\Pi_1) = \alpha.$$

Thus we have shown that $(\alpha_k)_i = f_i^*(vec\Pi_1^*)$ and $\alpha_i = f_1^*(vec\bar{\Pi}_1)$.

We may now write down a Taylor Series expansion analogous to (13) as follows

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

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

$$E((\hat{\alpha}_k)_i - \alpha_i) = \frac{1}{2!} tr\left[(f_i^{*(2)}(I \otimes (W^{-1})\Omega_1^{vec}(I \otimes (W)^{-1})\right] + o(T^{-1}),$$

where we have used the result that $vec(\Pi_1^* - \bar{\Pi}_1) = (I_{g+1} \otimes W^{-1})vecV_1$.

The introduction of the matrix W was made simply to write the $k-$class estimator in a form which enabled the bias approximation to be obtained directly using the same analysis as in the $2SLS$ case. When $f_i^{*(2)}$is interpreted in terms of the structural parameters, and Ω_1^{vec} is set equal to $\Omega_1 \otimes I_T$, W is cancelled out and then we have

$$\begin{aligned} E((\hat{\alpha}_k)_i - \alpha_i) &= tr\left[(HQe_i\beta_0' \otimes (I - P_Z) - k(I - P_X))\Omega_1^{vec}\right] \\ &\quad - tr[(I^*(XQe_i\beta_0' \otimes H'QX'))\Omega_1^{vec}] + o(T^{-1}). \qquad (17) \end{aligned}$$

which reduces to the result of Nagar(1959).

Notice that when $k = 1$ and $(I-P_Z)-k(I-P_X)$ is replaced by $P_X - P_Z$, the approximation reduces to that for $2SLS$ given above. Previously we noted that the $2SLS$ bias approximation was obtained under much weaker assumptions than were employed by Nagar. Here we see that we have obtained the bias approximation for the consistent $k-$class under the same weak assumptions; in particular, we *do not need normality nor independence* of the disturbances, merely that the row vectors of V_1 should satisfy the Gauss Markov conditions. Consequently, the Nagar bias approximation for $k-$class estimators is valid, for example, as it is for $2SLS$, under assumptions such as martingale differences and $ARCH/GARCH$ disturbances.

The higher order bias approximation can now be derived as in the $2SLS$ case and the bias to order T^{-2} does not depend upon the kurtosis of the error distribution Hence the higher order bias given in Iglesias and Phillips (2010) holds also under symmetric distributions whatever the degree of kurtosis. We may now state:

Theorem 3.

In the model of Section 2 where the errors may be asymmetrically distributed, the bias of the i^{th} component of the $k-$class estimator is given by

$$
\begin{aligned}
E((\hat{\alpha}_k)_i - \alpha_i) \quad = \quad & (L-1)[e_i' Qq + tr(QC)e'Qq - (L-2)e_i' QCQq] \\
& -[(L-1)(L-2) - \theta 2(L-2) + \theta^2]e'QCQq + \theta \frac{K}{T} e'Qq + \\
& e_i' QH(\beta_0' \otimes I_{g+1})\Omega^* H' QX' \Delta_{x,z}^{(k)} + e_i'((QH(I_{g+1} \otimes \beta_0')\Omega^* H' + \\
& tr(QH(I_{g+1} \otimes \beta_0')\Omega^* H').I_{g+k})QX' \Delta_{xz}^{(k)} \\
& -tr((I_{g+1} \otimes \beta_0')\Omega^* \; H' QX' Diag(XQe_i)XQH) \\
& +o(T^{-2}) \qquad (18)
\end{aligned}
$$

The part of the approximation which relates to the effects of the asymmetric errors differs from that of $2SLS$. In fact the effects of asymmetric errors involves the term Δ_{xz} in the $2SLS$ case which is a $T \times 1$ vector with p^{th} component $x_p'(X'X)^{-1}x_p - z_p'(Z'Z)^{-1}z_p$, $p = 1, 2, ..., T, .$ In the general $k-$class case examined here where $k = 1 + \frac{\theta}{T}$ and $\theta \leq 0$, this is replaced by $\Delta_{x,z}^k$ which is a $T \times 1$ vector with p^{th} component $1 - z_p'(Z'Z)^{-1}z_p - k(1 - x_p'(X'X)^{-1}x_p)$, $p = 1, 2, ..., T$. While the asymmetry effects depend directly on k through $\Delta_{xz}^{(k)}$, none of the higher order terms in the $k-$class bias approximation is explicit in $L-1$; hence, when $L = 1$ the higher order bias does not go to zero under asymmetric disturbances. It was seen earlier that this is the case for $2SLS$ also.

7 A Simple case

In the general model it is difficult to interpret the affects of asymmetry in the disturbances on the estimators and in this section we explore the effects in a very simple simultaneous equation model in an attempt to isolate the key factors.

We consider the simple model given by

$$
y_{2,t} = \beta_2 y_{2,t} + \gamma' z_t + u_{2,t}, t = 1, 2, ..., T, \qquad (19)
$$

where z_t is a $p \times 1$ vector of exogenous variables. The reduced form for $y_{2.t}$ is given by

$$
\begin{aligned}
y_{2,t} \quad &= \quad \beta_2(\beta_1 y_{2,t} + u_{1,t}) + \gamma' z_t + u_{2,t} \\
&= \quad \beta_1 \beta_2 y_{2,t} + \gamma' z_t + u_{2,t} + \beta_2 u_{1,t} \\
&= \quad \pi_2' z_t + v_t \qquad (20)
\end{aligned}
$$

where $\pi_2' = \frac{\gamma'}{1-\beta_1\beta_2}$.

Also $y_2 = Z\pi_2 + v_2$ from which we shall write $X = Z\pi_2$. We shall also require, $Q = (\pi_2' Z' Z\pi_2)^{-1} = \frac{1}{\pi_2' Z' Z \pi_2}, H = (0, 1),$

$(\Omega^*)' = \begin{bmatrix} \omega_{1,1,1} & \omega_{2,1,1} & \omega_{1,1,2} & \omega_{2,1,2} \\ \omega_{1,2,1} & \omega_{2,2,1} & \omega_{1,1,2} & \omega_{2,2,2} \end{bmatrix}, \beta_0 = \begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix}$ and $e_i = 1$.
We now define the vector

$$\Delta_{x,z} = \begin{bmatrix} x_1'(X'X)^{-1}x_1 - z_1'(Z'Z)^{-1}z_1 \\ x_2(X'X)^{-1}x_2 - z_2'(Z'Z)^{-1}z_2 \\ .... \\ x_T'(X'X)^{-1}x_T - z_T'(Z'Z)^{-1}z_T \end{bmatrix} = \begin{bmatrix} \frac{z_1'(\pi_2'\pi_2)z_1}{\pi_2 Z'Z\pi_2} - z_1'(Z'Z)^{-1}z_1 \\ \frac{z_2'(\pi_2'\pi_2)z_2}{\pi_2 Z'Z\pi_2} - z_2'(Z'Z)^{-1}z_2 \\ .... \\ \frac{z_T'(\pi_2'\pi_2)z_T}{\pi_2 Z'Z\pi_2} - z_T'(Z'Z)^{-1}z_T \end{bmatrix}$$

where $X'\Delta_{xz} = \pi_2'(z_1, z_2, ..., z_T)$, and $\Delta_{xz} = \frac{\sum(z_j'\pi_2)^3}{\pi_2 Z'Z\pi_2} - \sum \pi_2' z_j z_j'(Z'Z)^{-1}z_j$..
All summations run from 1 to T.

The first of the asymmetric terms is
$e_i'QH(\beta_0' \otimes I_{g+1})\Omega^* H'QX'\Delta_{x,z} =$

$\frac{1}{\pi_2'Z'Z\pi_2}(0,1) \begin{bmatrix} 1 & 0 & -\beta_1 & 0 \\ 0 & 1 & 0 & -\beta_1 \end{bmatrix} \begin{bmatrix} \omega_{1,1,1} & \omega_{1,2,1} \\ \omega_{2,1,1} & \omega_{2,2,1} \\ \omega_{1,1,2} & \omega_{1,2,2} \\ \omega_{2,1,2} & \omega_{2,2,2} \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
$\times \frac{1}{\pi_2'Z'Z\pi_2}[\frac{\sum(z_j'\pi_2)^3}{\pi_2 Z'Z\pi_2} - \sum \pi_2' z_j z_j'(Z'Z)^{-1}z_j]$

$$= (\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{\sum(z_j'\pi_2)^3}{(\pi_2 Z'Z\pi_2)^3} - \frac{\sum \pi_2' z_j z_j'(Z'Z)^{-1}z_j}{(\pi_2 Z'Z\pi_2)^2}] \tag{21}$$

The second asymmetric term is

$e_i'((QH(I_{g+1} \otimes \beta_0')\Omega^* H'QX'\Delta_{x,z} =$

$\frac{1}{\pi_2'Z'Z\pi_2}(0,1) \begin{bmatrix} \omega_{1,1,1} & \omega_{2,1,1} & \omega_{1,1,2} & \omega_{2,1,2} \\ \omega_{1,2,1} & \omega_{2,2,1} & \omega_{1,1,2} & \omega_{2,2,2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\beta_1 & 0 \\ 0 & 1 \\ 0 & -\beta_1 \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
$\times \frac{1}{\pi_2'Z'Z\pi_2}[\frac{\sum(z_j'\pi_2)^3}{\pi_2 Z'Z\pi_2} - \sum \pi_2' z_j z_j'(Z'Z)^{-1}z_j]$

$$= (\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{\sum(z_j'\pi_2)^3}{\pi_2 Z'Z\pi_2} - \frac{\sum \pi_2' z_j z_j'(Z'Z)^{-1}z_j}{(\pi_2 Z'Z\pi_2)^2}] \tag{22}$$

The third asymmetric term is
$tr\{HQH'((I_{g+1} \otimes \beta_0')\Omega^* e_i'QX'\Delta_{x,z}\} = tr\{QH'((I_{g+1} \otimes \beta_0')\Omega^* He_i'QX'\Delta_{x,z}\}$
$\frac{1}{\pi_2'Z'Z\pi_2}(0,1) \begin{bmatrix} 1 & -\beta_1 & 0 & 0 \\ 0 & 0 & 1 & -\beta_1 \end{bmatrix} \begin{bmatrix} \omega_{1,1,1} & \omega_{1,2,1} \\ \omega_{2,1,1} & \omega_{2,2,1} \\ \omega_{1,1,2} & \omega_{1,2,2} \\ \omega_{2,1,2} & \omega_{2,2,2} \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$\times\frac{1}{\pi_2' Z' Z\pi_2}[\frac{\sum(z_j'\pi_2)^3}{\pi_2 Z' Z\pi_2}-\sum\pi_2' z_j z_j'(Z'Z)^{-1}z_j]$

$$=(\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{\sum(z_j'\pi_2)^3}{(\pi_2 Z' Z\pi_2)^3}-\frac{\sum\pi_2' z_j z_j'(Z'Z)^{-1}z_j}{(\pi_2 Z' Z\pi_2)^2}] \tag{23}$$

on noting that $\omega_{1,2,2}=\omega_{2,2,1}$.

Finally the fourth asymmetric term is

$$-tr((I_{g+1}\otimes\beta_0')\Omega^*\ H'QX'Diag(XQe_i)XQH)$$
$$=-tr(XQH)((\beta_0'\otimes I_{g+1})\Omega^*\ H'QX'Diag(XQe_i))$$
$$=-tr\{\frac{Z\pi_2}{\pi_2' Z' Z\pi_2}(0,1)\begin{bmatrix}1 & -\beta_1 & 0 & 0\\ 0 & & 10 & -\beta_1\end{bmatrix}\begin{bmatrix}\omega_{1,1,1} & \omega_{1,2,1}\\ \omega_{2,1,1} & \omega_{2,2,1}\\ \omega_{1,1,2} & \omega_{1,2,2}\\ \omega_{2,1,2} & \omega_{2,2,2}\end{bmatrix}\begin{pmatrix}0\\1\end{pmatrix}$$
$$\times\frac{\pi_2' Z'}{\pi_2 Z' Z\pi_2}\begin{bmatrix}\frac{z_1'\pi_2}{\pi_2 Z' Z\pi_2} & 0 & 0 & 0\\ 0 & \frac{z_2'\pi_2}{\pi_2 Z' Z\pi_2} & & 0\\ . & & . & \\ 0 & & & \frac{z_T'\pi_2}{\pi_2 Z' Z\pi_2}\end{bmatrix}\}$$

$$=-tr\{\frac{1}{\pi_2' Z' Z\pi_2}(\omega_{2,2,1}-\beta_1\omega_{2,2,2})$$
$$\times\frac{\pi_2' Z'}{\pi_2 Z' Z\pi_2}\begin{bmatrix}\frac{z_1'\pi_2}{\pi_2 Z' Z\pi_2} & 0 & 0 & 0\\ 0 & \frac{z_{22}'\pi}{\pi_2 Z' Z\pi_2} & & 0\\ . & & . & \\ 0 & & & \frac{z_T'\pi_2}{\pi_2 Z' Z\pi_2}\end{bmatrix}\begin{bmatrix}z_1'\pi_2\\ z_2'\pi_2\\ .\\ z_T'\pi_2\end{bmatrix}\}$$

$$=(\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{\sum(z_j'\pi_2)^3}{(\pi_2 Z' Z\pi_2)^3}] \tag{24}$$

Finally, summing (21)-(24), we find that the asymmetric terms

$$e_i'QH(\beta_0'\otimes I_{g+1})\Omega^* H'QX'\Delta_{x,z}+e_i'((QH(I_{g+1}\otimes\beta_0')\Omega^* H'+$$
$$tr(QH(I_{g+1}\otimes\beta_0')\Omega^* H').I_{g+k})QX'\Delta_{xz}$$
$$-tr((I_{g+1}\otimes\beta_0')\Omega^*\ H'QX'Diag(XQe_i)XQH)\ +o(T^{-2})$$

are equal to

$$3(\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{\sum(z_j'\pi_2)^3}{(\pi_2 Z' Z\pi_2)^3}-\frac{\sum\pi_2' z_j z_j'(Z'Z)^{-1}z_j}{(\pi_2 Z' Z\pi_2)^2}] \tag{25}$$
$$-(\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{\sum(z_j'\pi_2)^3}{(\pi_2 Z' Z\pi_2)^3}]$$

in

this special case.It is seen that the above expression is of order T^{-2} as expected and the bracketed terms may go to zero quite quickly as T gets large. Clearly $(\omega_{2,2,1-}\beta_1\omega_{2,2,2})$ plays a key role. It is helpful to interpret the disturbance skewness factors in terms of the structural parameters. Noting that

$$\begin{aligned}\omega_{2,2,1-}\beta_1\omega_{2,2,2} &= E(v_{1,t}v_{2,t}^2) - \beta_1 E(v_{2,t}^3) \\ &= E(\frac{\varepsilon_{1,t}+\beta_1\varepsilon_{2,t}}{1-\beta_1\beta_2})(\frac{\beta_2\varepsilon_{1,t}+\varepsilon_{2,t}}{1-\beta_1\beta_2})^2 \\ &\quad -\beta_1 E(\frac{\beta_2\varepsilon_{1,t}+\varepsilon_{2,t}}{1-\beta_1\beta_2})^3 \end{aligned} \tag{26}$$

which

with some manipulation simplifies to

$$\omega_{2,2,1-}\beta_1\omega_{2,2,2} = \frac{\beta_1(\sigma_{1,1,1-}\sigma_{2,2,2})}{(1-\beta_1\beta_2)^3} + \frac{\sigma_{1,1,1}\beta_2^2+2\sigma_{1,1,2}+\sigma_{1,2,2}}{(1-\beta_1\beta_2)^2} \tag{27}$$

it

is clear that this term can be made large for suitable choice of the parameters, especially since $\beta_1$and β_2 are unrestricted other than the requirement that $\beta_1\beta_2 \neq 1$.

Consider now the part not involving $\omega_{2,2,1-}\beta_1\omega_{2,2,2}$. If we define $x_j = z_j'\pi_2$ and $\alpha_j = z_j'(Z'Z)^{-1}z_j$ then

$$\frac{\sum\left(z_j'\pi_2\right)^3}{(\pi_2'Z'Z\pi_2)^3} - \frac{\sum \pi_2' z_j z_j'(Z'Z^{-1})z_j}{(\pi_2'Z'Z\pi_2)^2} = \frac{\sum x_j^3}{(\sum x_j^2)^3} - \frac{\sum x_j\alpha_j}{(\sum x_j^2)^2} \tag{28}$$

Upon putting $x_j = rc_j$ where $r = (\sum x_j^2)^{\frac{1}{2}}$and $\sum c_j^2 = 1$, the above becomes $\frac{1}{r^3}(\sum c_j^3 - \sum c_j\alpha_j)$..If we allow all the z_j (and hence the x_j) to shrink at the same rate then c_j and α_j are unchanged while r becomes small. In this case in the limit, as $r \to 0$, of $\frac{i}{r^3}(\sum c_j^3 - \sum c_j\alpha_j)$.is unbounded. Similarly on noting that $\pi_2' = \frac{\gamma'}{1-\beta_1\beta_2}$ where the structural parameter vector γ can be varied independently of the all the other structural parameters and, in particular, it can be varied to make π_2 arbitrarily small (which might happen in the weak instrument case), it is clear that as π_2 becomes small (so that r becomes small), the overall expression could become relatively large despite the fact that it is of order T^{-2}. Thus there is another situation in which the asymmetry might lead to a significant bias.

This simple case provides evidence that skewness of disturbances seems likely to cause estimation biases to differ substantially in some situations compared to when disturbances are symmetric.

The analogous result for the general $k-$class of estimators in the simple case is given by

$$3(\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{1-k}{(\pi_2 Z'Z\pi_2)^2}+k\frac{\sum(z_j'\pi_2)^3}{(\pi_2 Z'Z\pi_2)^3}-\frac{\sum\pi_2'z_jz_j'(Z'Z)^{-1}z_j}{(\pi_2 Z'Z\pi_2)^2}]$$
$$-(\omega_{2,2,1-}\beta_1\omega_{2,2,2})[\frac{1-k}{(\pi_2 Z'Z\pi_2)^2}+k\frac{\sum(z_j'\pi_2)^3}{(\pi_2 Z'Z\pi_2)^3}] \tag{29}$$

Notice that when $k=1$ the above reduces to the result for $2SLS$.We shall not give a separate analysis for the general $k-$class estimator:since it is clear that the same observations can be made as in the case of $2SLS$.

8 Conclusion

The $2SLS$ estimator has an important place in the history of simultaneous equation estimation and continues to be frequently used in practice. Hence the results in this paper are of both theoretical and practical interest. As noted previously, the Mikhail $2SLS$ bias approximation is likely to be of importance when equations are heavily overidentified since then the higher order terms will not be neglible The fact that the approximation holds under symmetric distributions and any degree of kurtosis obviously increases its applicability in practical cases.When the errors are asymmetrically distributed we have seen that the Mikhail approximation no longer holds and we have presented the correct approximation for such cases.

The $k-$class of estimators where $k<1$ are also of interest partly because estimators in this class have all necessary moments. In Iglesias and Phillip(2008), a linear combination of $k-$class estimators for which $k<1$, was presented which was unbiased to order T^{-1} , had all necessary moments and dominated $2SLS$ on a MSE criterion in strongly overidentified cases. Here it has been shown that the terms in the higher order bias approximation which measured the asymmetry effects, depend directly on k while kurtosis did not play a role to the order of the approximation.

We cannot say, without further work, what the general effects of asymmetry are except that they are likely to be greater the larger the degree of skewness in the error distributions. We have examined a special case where it appears that the skewness effects can be significant. This can be explored numerically in more general cases by calculating the approximations for a variety of different structures; it can also be examined in Monte Carlo experiments.In fact some preliminary Monte Carlo results of a two-equation model indicate that when the skewness is considerable the effect on bias is far from trivial. A comprehensive study of the asymmetry effects would be a major exercise and a natural next step; however it lies outside the scope of this paper.

References

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

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 (2008), Almost Unbiased Estimation in Simultaneous Equations with Strong and /or Weak Instruments. Paper submitted to the *Journal of Econometrics*.

Iglesias, E.M and G.D.A.Phillips (2010). The Bias to order T^{-2} for the general $k-$class estimator in a simultaneous model. *Economics Letters* 109,42-45.

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

Knight, J.L. (1985) The Moments Of OLS And 2SLS When The Disturbances Are Non-Normal. J*ournal of Econometrics 27,39-60*

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

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.

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.*CambridgeUniversity Press.

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

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, p533-541.

Ullah, A.(2007) .*Finite Sample Econometrics.*Oxford University Press.

Appendix 1

In our examination of the Mikhail approximation we shall need the third and fourth moments of reduced form regression estimators where, for the general static simultaneous equation case, the reduced form model is equivalent to the classical linear regression model. In this appendix we derive the third and fourth moments of the OLS estimator in a linear regression model.

(i) The Third Moment of the OLS Estimator

In the regression model $y = X\beta + \varepsilon$ where the errors are $i.i.d, (0, \sigma^2)$ and X contains p exogenous regressors, the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ has an estimation error

$\hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$

The general element of this vector will be written as

$e_i'(\hat{\beta} - \beta) = \hat{\beta}_i - \beta_i = e_i'(X'X)^{-1}X'\varepsilon$.

Our interest centres on the third moment

$$E(\hat{\beta}_i - \beta_i)^3 = E(e_i'(X'X)^{-1}X'\varepsilon)^3 \tag{30}$$

To obtain this we first consider

$$(e_i'(X'X)^{-1}X'\varepsilon)^3 = e_i'(X'X)^{-1}X'\varepsilon e_i'(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}e_i \tag{31}$$

contains the stochastic component $(e_i'(X'X)^{-1}X'\varepsilon)\varepsilon\varepsilon'$

.

To find the expectation of this we first note the general term of the $T \times T$ matrix $\varepsilon\varepsilon'$ which is $\varepsilon_j\varepsilon_k$, $j, k = 1, 2, ..., T$. then find the expectation of $(e_i'(X'X)^{-1}X'\varepsilon)\varepsilon_j\varepsilon_k$ .Notice that $e_i'(X'X)^{-1}X'$ is the i^{th} row of $(X'X)^{-1}X'$ and we denote this row vector by α' so that $e_i'(X'X)^{-1}X'\varepsilon \equiv \alpha'\varepsilon = \sum_{r=1}^{T} \alpha_r\varepsilon$.

Now consider $E\{\sum_{r=1}^{T} \alpha_r\varepsilon_r\varepsilon_j\varepsilon_k\}$which will contain non-zero components only for $r = j = k$. Hence $E\{\sum_{r=1}^{T} \alpha_r\varepsilon_r\varepsilon_j\varepsilon_k\} = \alpha_r E(\varepsilon_r^3)$, $r = j = k$, $k = 1, 2, , T$, and is zero otherwise.

It follows that $E\{(e_i'(X'X)^{-1}X'\varepsilon)\varepsilon\varepsilon'\}$ is a diagonal matrix with r, r^{th} component given by $\alpha_r E(\varepsilon_r^3)$ where α_r is the r^{th} component of α. Note that $e_i'(X'X)^{-1}X'$ picks out the i^{th} row of $(X'X)^{-1}X'$ and so α_r is the r^{th}component of this row vector. We now write that

$(X'X)^{-1}X' = (X'X)^{-1}(x_1, x_2, ..., x_T)$, where x_j is the j^{th} column of X', from which it is seen that the r^{th} component of $e_i'(X'X)^{-1}X'$ is equal to $e_i'(X'X)^{-1}x_r$.

It follows that

$E(e_i'(X'X)^{-1}X'\varepsilon)^3 = E(\varepsilon_t^3)e_i'(X'X)^{-1}X'Diag(\alpha_r)X(X'X)^{-1}e_i$ where $Diag(\alpha_r)$ is a $T \times T$ diagonal matrix with the r, r^{th}component α_r.

Some simplification is possible by noting that $X'\ Diag(\alpha_r)X$ with α_r replaced by $e_i'X(X'X)^{-1}x_r, r = 1, 2, ..., T$, may be written as

$= \sum_{j=1}^{T} x_jx_j'e_i'(X'X)^{-1}x_j$.

Then $e_i'(X'X)^{-1}X'Diag(\alpha_r)X(X'X)^{-1}e_i$ is equal to

$e_i'(X'X)^{-1}\sum x_jx_j'e_i'(X'X)^{-1}x_j(X'X)^{-1}e_i = \sum(e_i'(X'X)^{-1}x_j)^3$

Hence the required third moment is

$$E(\hat{\beta}_i - \beta_i)^3 = E(\varepsilon_i^3)\sum_{j=1}^{T}(e_i'(X'X)^{-1}x_j)^3 \tag{32}$$

Finally, it is seen that $e_i'(X'X)^{-1}x_j$ is $O(T^{-1})$, $(e_i'(X'X)^{-1}x_j)^3$ is $O(T^{-3})$ and $\sum_{j=1}^{T}(e_i'(X'X)^{-1}x_j)^3$ is $O(T^{-2})$ and, hence, the third moment of the OLS estimator is $O(T^{-2})$.

(ii) The fourth moment of the OLS Estimator

The fourth moment is given by $E(e_i'(X'X)^{-1}X'\varepsilon)^4$ and we commence by writing

$$\begin{aligned}(e_i'(X'X)^{-1}X\varepsilon)^4 &= e_i'(X'X)^{-1}X\varepsilon\varepsilon'X(X'X)^{-1})e_i \\ &\quad \times e_i'(X'X)^{-1}X\varepsilon\varepsilon'X(X'X)^{-1})e_i\end{aligned} \tag{33}$$

We shall need to find the expected value of this but we can proceed by focusing on the stochastic part

$E(\varepsilon\varepsilon'X(X'X)^{-1})e_ie_i'(X'X)^{-1}X\varepsilon\varepsilon')$.

In fact we shall write

$$E(\varepsilon\varepsilon'X(X'X)^{-1}e_ie_i'(X'X)^{-1}X\varepsilon\varepsilon') = E(\varepsilon\varepsilon'A\varepsilon\varepsilon') \tag{34}$$

where the matrix A defined as

$$A = X(X'X)^{-1})e_ie_i'(X'X)^{-1}X \tag{35}$$

is $T \times T$ and symmetric.

Note that the $T \times T$ matrix $\varepsilon\varepsilon'$ has a general element $\varepsilon_i\varepsilon_j$while, in addition, we have

$\varepsilon'A\varepsilon = \sum_{i=1}^{T} a_{ii}^2\varepsilon_i^2 + \sum_{i=1}^{T}\sum_{j=1}^{T} a_{ij}\varepsilon_i\varepsilon_j, i \neq j,$

When $i = j = k$, we may write

$E(\varepsilon_i\varepsilon_j'\varepsilon A\varepsilon) = a_{kk}E(\varepsilon_k^4) + (trA - a_{kk})\sigma^4, k = 1, 2, ..., T$

whereas when $i \neq j$ but $i = k, j = l$,we have

$$\begin{aligned}E(\varepsilon_i\varepsilon_j'\varepsilon A\varepsilon) &= (a_{kl} + a_{lk})E(\varepsilon_k^2\varepsilon_l^2) \\ &= (a_{kl} + a_{lk})E(\varepsilon_k^2)E(\varepsilon_l^2) = (a_{kl} + a_{lk})\sigma^4, \\ &\quad k, l = 1, 2, ., T, k \neq l.\end{aligned}$$

.

It follows that

$$\begin{aligned}E(\varepsilon'A\varepsilon\varepsilon\varepsilon') &= E(\varepsilon_i^4)DiagA + (trA.I_T - DiagA)\sigma^4 \\ &\quad +2(A - DiagA)\sigma^4\end{aligned}$$

$$= (E(\varepsilon_i^4) - 3\sigma^4) Diag A + (2A + tr A.I_T)\sigma^4 \tag{36}$$

We now replace the matrix A with $X(X'X)^{-1})e_i e_i'(X'X)^{-1}X\ '$ and consider the earlier expression

$(e_i'(X'X)^{-1}X\varepsilon)^4 = e_i'(X'X)^{-1}X'\varepsilon\varepsilon' X(X'X)^{-1}e_i$
$\times e_i'(X'X)^{-1}X\varepsilon\varepsilon' X(X'X)^{-1}e_i.$

Taking expectations.

$$\begin{aligned} &E(e_i'(X'X)^{-1}X\varepsilon)^4 = (E(\varepsilon_i^4) - 3\sigma^4)e_i'(X'X)^{-1}X' \\ &\times Diag[[X(X'X)^{-1})e_i e_i'(X'X)^{-1}X'.]X(X'X)^{-1})e_i \\ &+3\sigma^4(e_i'(X'X)^{-1}e_{i.})^2 \end{aligned} \tag{37}$$

where this latter term above is $O(T^{-2})$.

Consider the matrix $Diag[[X(X'X)^{-1})e_i e_i'(X'X)^{-1}X'.]$. The first diagonal element of this matrix is

$e_1'X(X'X)^{-1})e_i e_i'(X'X)^{-1}X'e_1 = (e_i'(X'X)^{-1}X'e_1)^2.$

Here e_1 is a $T \times 1$ unit vector with unity in the first position and all other elements zero. Noting that $X'e_1 = x_1$,which is the transpose of the first row of X, we see $e_i'(X'X)^{-1}X'e_1 = e_i'(X'X)^{-1}x_1$and $Diag[[X(X'X)^{-1})e_i e_i'(X'X)^{-1}X'.]$ is a diagonal matrix with j, jth component $(e_i'(X'X)^{-1}x_j)^2,\ j = 1, 2, ..., T.$

The foregoing enables some simplication of the expression for the fourth moment since it is easy to see that

$$e_i'(X'X)^{-1}X' \times Diag[[X(X'X)^{-1})e_i e_i'(X'X)^{-1}X'.]]X(X'X)^{-1})e_i \tag{38}$$

reduces to $\sum_{j=1}^{T}(x_j'(X'X)^{-1}e_i)^4$.

Hence, finally it has been shown that

$$\begin{aligned} E(\hat{\beta}_i - \beta_i)^4 &= \sigma^4 3(e_i'(X'X)^{-1}e_i)^2 + \\ &\quad (E(u_i^4) - 3\sigma^4)\sum_{j=1}^{T}(x_j'(X'X)^{-1}e_i)^4 \end{aligned} \tag{39}$$

Notice that the first term is $O(T^{-2})$ and the second is $O(T^{-3})$. Thus the kurtosis of the error distribution does not affect the fourth moment of $\hat{\beta}_i$ to order T^{-2}.If the error distribution is normal then $E(u_i^4) - 3\sigma^4 = 0$ so that the second term disappears.In such a case $\hat{\beta}_i$ is normally distributed and again we see that the fourth moment is just three times the squared variance as required.

One further observation is that without assuming normality,

$$lim E(T^{\frac{1}{2}}(\hat{\beta}_i - \beta_i))^4 = 3\sigma^4((e_i'\Sigma_{xx}^{-1}e_i)^2$$

as $T \to \infty$, where $limT(X'X)^{-1} = \Sigma_{xx}^{-1}$.This does not involve kurtosis and. as expected, it is equal to three times the square of the limiting variance.

Appendix 2

In this Appendix we evaluate the expectation of the skewness term

$\frac{1}{3!}\sum\sum(\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1))' f_{i,rs}^{(3)}(vec(\hat{\Pi}_1 - \Pi_1))$.

We commence with the following:

Lemma 1

$E\{(\hat{\pi}_{rs} - \pi_{rs})(vecV_1)(vecV_1)'\} = \quad \Omega_{ijs} \otimes Diag(z_r)$

where

$$\Omega_{ijs=} \begin{bmatrix} \omega_{11s} & \omega_{12s} & \ldots\ldots & \omega_{1,g+1.s} \\ \omega_{21s} & \omega_{22s} & \ldots\ldots & \omega_{2,g+1,s} \\ . & . & . & . \\ . & . & . & . \\ \omega_{g+1,1,s} & \omega_{g+1,2,s} & \ldots. & \omega_{g+1,2,s} \end{bmatrix} \quad \text{and}$$

$$Diag(z_r) = \begin{bmatrix} z_{r1} & . & . & 0 \\ 0 & z_{r2} & & o \\ . & & . & \\ 0 & 0 & & z_{rT} \end{bmatrix}$$

Proof

To see this we proceed from

$$E\{(\hat{\pi}_{rs} - \pi_{rs})(vecV_1)(vecV_1)'\} = E\{e_r'(Z'Z)^{-1}Z'v_s(vecV_1)(vecV_1)'\}$$

$$= E\{e_r'(Z'Z)^{-1}Z'v_s \begin{bmatrix} v_1v_1' & v_1v_2' & \ldots & v_1v_{g+1}' \\ v_2v_1' & v_2v_2' & \ldots. & v_2v_{g+1}' \\ . & . & \ldots & \ldots. \\ v_{g+1}v_1' & v_{g+1}v_2' & \ldots\ldots & v_{g+1}v_{g+1}' \end{bmatrix}$$

where v_j is a $T \times 1$ vector forming the j^{th} column of V_1.

We shall write $e_r'(Z'Z)^{-1}Z' = \bar{z}_r'$ and $e_r'(Z'Z)^{-1}Z'v_s = \bar{z}_r'v_s$ and consider $E(\bar{z}_r'v_sv_iv_j')$ with general term $E(\bar{z}_r'v_sv_{pi}v_{qj})$ $p.q = 1.2....,T$ which is non-zero only when the stochastic terms are of the same time period. When $p = q$ it is seen that $E(\bar{z}_r'v_sv_{pi}v_{pj}) = E(\bar{z}_{pr}v_{ps}v_{pi}v_{pj}) = \bar{z}_{pj}\omega_{ijs}$ where $\bar{z}_{pj}$ is the p^{th} component of $\bar{z}_r$ and $E(v_{pi}v_{pj}v_{ps}) = \omega_{ijs}$. More generally,

$$E(\bar{z}_r'v_sv_iv_j') = \omega_{ijs} \begin{bmatrix} \bar{z}_{r1} & 0 & \ldots. & 0 \\ 0 & \bar{z}_{r2} & & 0 \\ & & . & . \\ 0 & 0 & \ldots. & \bar{z}_{rT} \end{bmatrix}$$

$$= \omega_{ijs}Diag(\bar{z}_r) \text{ for } i, j, s = 1, 2, ..., g + 1.$$

In Phillips (2000). it is shown that the term of interest

$$\frac{1}{3!}\sum_{r=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1))' f_{i,rs}^{(3)}(vec(\hat{\Pi}_1 - \Pi_1)) \tag{40}$$

is equal to the sum of the following three terms:

$$\sum_{r=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs}-\pi_{rs})(vecV_1)'\{H'Qe_i\beta_0'E_{rs}'Z'XQH\otimes(P_X-P_Z)\}vecV_1$$

$$+\sum_{r-=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs}-\pi_{rs})(vecV_1)'\{H'Q(X'ZE_{rs}H'+HE_{rs}'Z'X)Qe_i\beta_0'$$
$$\otimes(P_X-P_Z)\}vecV_1$$

$$-\sum_{r=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs}-\pi_{rs})(vecV_1)'\{\beta_0e_i'QX'ZE_{rs}H'QX'\otimes XQH\}I^*vecV_1$$

where E_{rs} is a $K\times(g+1)$ matrix of rank one with unity in the r,s^{th} position and zeroes elsewhere.

It is required to find the expected value of the above and we shall do so by evaluating each of the three components in turn.

First we examine

$$(a).\ (\hat{\pi}_{rs}-\pi_{rs})(vecV_1)'\{H'Qe_i\beta_0'E_{rs}'Z'XQH\otimes(P_X-P_Z)\}vecV_1 \tag{41}$$

$$=tr[(\hat{\pi}_{rs}-\pi_{rs})(vecV_1)(vecV_1)'\{H'Qe_i\beta_0'E_{rs}'Z'XQH\otimes(P_X-P_Z)\}$$

where $\hat{\pi}_{rs}-\pi_{rs}$,which is the $K(r-1)+s^{th}$ component of $Vec(\hat{\Pi}_1-\Pi_1)$,and which may be written as $\hat{\pi}_{rs}-\pi_{rs}=e_r'(Z'Z)^{-1}Z'v_s$. Here e_r' is a $1\times K$ unit vector with unity in the r^{th} position and zeroes elsewhere. Thus it picks out the r^{th}component of $(Z'Z)^{-1}Z'v_s$ where v_s is a $T\times1$ vector of reduced form disturbances appearing in the s^{th} reduced form equation, i.e, the s^{th} column of V_1.

The term of interest can then be written as

$$tr[e_r'(Z'Z)^{-1}Z'v_s(vecV_1)(vecV_1)'\{H'Qe_i\beta_0'E_{rs}'Z'XQH\otimes(P_X-P_Z)\}]. \tag{42}$$

We have shown above in Lemma 1 that
$E\{e_r'(Z'Z)^{-1}Z'v_s(vecV_1)(vecV_1)'\}=\Omega_{ijs}\otimes Diag(\bar{z}_r)$

so it follows that

$$Etr[e_r'(Z'Z)^{-1}Z'v_s(vecV_1)(vecV_1)'\{H'Qe_i\beta_0'E_{rs}'Z'XQH\otimes(P_X-P_Z)\}]$$
$$=tr[\Omega_{ijs}\otimes Diag(\bar{z}_r)\{H'Qe_i\beta_0'E_{rs}'Z'XQH\otimes(P_X-P_Z)\}]$$
$$=tr\{\Omega_{ijs}H'Qe_i\beta_0'E_{rs}'Z'XQH\}tr\{Diag(\bar{z}_r)(P_X-P_Z)\}$$

Some simplification is possible by writing
$tr\{Diag(\bar{z}_r)(P_X-P_Z)\}=\bar{z}_r'\ \Delta_{x,z}$
where $\Delta_{x,z}$ is a $T\times1$ vector with p^{th} component $x_p'(X'X)^{-1}x_p-z_p'(Z'Z)^{-1}z_p$, $p=1,2,......,T$.

Next we shall write $E_{rs}=e_re_s'$ where e_s is a $(g+1)\times1$ unit vector with unity in the s^{th}position. On putting $e_s'\beta_0=\beta_{s0}$, the s^{th} component of β_0, the above expression may be written as

$$e_i'QH\Omega_{ijs}H'QX'Ze_r\beta_{s0}\bar{z}_r'\ \Delta_{x,z} \tag{43}$$

Finally we need to find the value of

$$\sum_{r=1}^{K}\sum_{s=1}^{g+1} e_i' QH\Omega_{ijs}H'QX'Ze_r\beta_{s0}\bar{z}_r' \ \Delta_{x,z} \tag{44}$$

We shall proceed by first finding the summation for $r = 1, ..., K$ and so we consider

$\sum\limits_{r=1}^{K} e_i' QH\Omega_{ijs}H'QX'Ze_r\beta_{s0}\bar{z}_r' \ \Delta_{x,z}$

$= \beta_{s0}e_i' QH\Omega_{ijs}H'QX'Z \sum\limits_{r=1}^{K} e_r e_r'(Z'Z)^{-1}Z'\Delta_{x,z}$

$=\beta_{s0}e_i' QH\Omega_{ijs}H'QX'\Delta_{x,z}$

where we have used the fact that $\sum\limits_{r=1}^{K} e_r e_r' = I_K$ and $X'Z(Z'Z)^{-1}Z' = X'$.

To complete the evaluation we simply need to sum over s. Hence the final expression is

$E\sum\limits_{r=1}^{K}\sum\limits_{s=1}^{g+1}(\hat{\pi}_{rs} - \pi_{rs})(vecV_1)'\{H'Qe_i\beta_0' E_{rs}' Z'XQH \otimes (P_X - P_Z)\}vecV_1$

$= e_i' QH(\sum\limits_{s=1}^{g+1} \beta_{s0}\Omega_{ijs})H'QX'\Delta_{x,z}$

$$= e_i' QH(\beta_0' \otimes I_{g+1})\Omega^* H'QX'\Delta_{x,z} \tag{45}$$

Here we have used the result that $\sum\limits_{s=1}^{g+1} \beta_{s0}\Omega_{ijs}$can be written as $(\beta_0' \otimes I_{g+1})\Omega^*$ where Ω^* is a $(g + 1)^2 \times (g + 1)$ matrix obtained by stacking the matrices $\Omega_{ijs}, s = 1, ..., g + 1,$

(b).The second term of interest is

$\sum\limits_{r=1}^{K}\sum\limits_{s=1}^{g+1}(\hat{\pi}_{rs} - \pi_{rs})(vecV_1)'\{H'Q(X'ZE_{rs}H' + HE_{rs}'Z'X)Qe_i\beta_0'$

$\otimes(P_X - P_Z)\}vecV_1$

Again we shall initially disregard the summations and consider

$E[tr(e_r'(Z'Z)^{-1}Z'v_s vecV_1(vecV_1)')$

$\times\{H'Q(X'ZE_{rs}H' + HE_{rs}'Z'X)Qe_i\beta_0' \otimes (P_X - P_Z)\}]$

$= tr((\Omega_{ijs} \otimes Diag(\bar{z}_r))\{H'Q(X'ZE_{rs}H' + HE_{rs}'Z'X)Qe_i\beta_0' \otimes (P_X - P_Z)\})$

$=tr(\Omega_{ijs}(H'Q(X'ZE_{rs}H' + HE_{rs}'Z'X)Qe_i\beta_0'))tr(Diag(\bar{z}_r)(P_X - P_Z))$

$=e_i' QHE_{rs}'Z'XQH\Omega_{ijs}\beta_0\bar{z}_r'\Delta_{xz}+e_i' QX'ZE_{rs}H'QH\Omega_{ijs}\beta_0\bar{z}_r'\Delta_{xz}$

Putting $E_{rs} = e_r e_s'$ the above becomes

$e_i' QHe_s e_r' Z'XQH\Omega_{ijs}\beta_0\bar{z}_r'\Delta_{xz} + e_i' QX'Ze_r e_s' H'QH\Omega_{ijs}\beta_0\bar{z}_r'\Delta_{xz}$

$= e_i' QHe_s\Delta_{xz}'Z(Z'Z)^{-1}e_r e_r' Z'XQH\Omega_{ijs}\beta_0 + e_i' QX'Ze_r\bar{z}_r'\Delta_{xz\ r}e_s' H'QH\Omega_{ijs}\beta_0$

$= e_i' QHe_s\Delta_{xz}'Z(Z'Z)^{-1}e_r e_r' Z'XQH\Omega_{ijs}\beta_0 + e_i' QX'Ze_r e_r'(Z'Z)^{-1}Z'\Delta_{xz}e_s' H'QH\Omega_{ijs}\beta_0$

Summing over $r = 1, ..., K$ yields

$e_i'QHe_s\Delta_{xz}'XQH\Omega_{ijs}\beta_0 + e_i'QX'\Delta_{xz}e_s'H'QH\Omega_{ijs}\beta_0$

Finally, summing over $s = 1, ..., g+1$ gives

$$e_i'QH\sum_{s=1}^{g+1}(e_s\beta_0'\Omega_{ijs}')H'QX'\Delta_{xz} + e_i'QX'\Delta_{xz}tr(H'QH\sum_{s=1}^{g+1}\Omega_{ijs}\beta_0 e_s')$$

Alternatively, on noting that $\sum_{s=1}^{g+1}(e_s\beta_0'\Omega_{ijs}') = \Omega'^*(I_{g+1}\otimes\beta_0)$ where $\Omega^*$is a $(g+1)^2\times(g+1)$ matrix obtained by "stacking" the matrices Ω_{ijs} , $s = 1, .., g+1$, we may write

$$E\{\sum_{r=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs}-\pi_{rs})(vecV_1)'\{H'Q(X'ZE_{rs}H' + HE_{rs}'Z'X)$$
$$Qe_i\beta_0'\otimes(P_X - P_Z)\}vecV_1\}$$

$$= e_i'(QH\Omega^{*\prime}(I_{g+1}\otimes\beta_0)H' + tr(QH\Omega'^*(I_{g+1}\otimes\beta_0)H').I_{g+k})QX'\Delta_{xz} \qquad (46)$$

(c).The third and final term is

$$-\sum_{r=1}^{K}\sum_{s=1}^{g+1}(\hat{\pi}_{rs}-\pi_{rs})\ (vecV_1)'\ \{\beta_0 e_i'QX'ZE_{rs}H'QX'\otimes XQH\}I^*vecV_1$$

where intiially we focus on

$$-Etr[(\hat{\pi}_{rs}-\pi_{rs})vecV_1(vecV_1)'\{\beta_0 e_i'QX'ZE_{rs}H'QX'\otimes XQH\}I^*]$$
$$= -trE[(\hat{\pi}_{rs}-\pi_{rs})vecV_1(vecV_1)'\{\beta_0 e_i'QX'ZE_{rs}H'QX'\otimes XQH\}I^*]$$
$$= -tr[(\Omega_{ijs}\otimes Diag(\bar{z}_z))\{\beta_0 e_i'QX'ZE_{rs}H'QX'\otimes XQH\}I^*]$$
$$=-tr[(\Omega_{ijs}\beta_0 e_i'QX'ZE_{rs}H'QX'\otimes Diag(\bar{z}_r)XQH)I^*]$$
$$=-tr[(\Omega_{ijs}\beta_0 e_i'QX'ZE_{rs}H'QX'Diag(\bar{z}_r)XQH)]$$
$$=-tr[(\Omega_{ijs}\beta_0 e_i'QX'Ze_r e_s'H'QX'Diag(\bar{z}_r)XQH)].$$

Consider now $e_i'QX'Ze_r\bar{z}_{rj}$ in which $\bar{z}_{rj}$ is the jth component of $\bar{z}_r$ and $\bar{z}_{rj} = e_r'(Z'Z)^{-1}Z'e_j, j = 1, ..., T.$ where e_j is a $T\times1$ unit vector.Then $\sum_{r=1}^{K}e_i'QX'Ze_r\bar{z}_{rj} = \sum_{r=1}^{K}e_i'QX'Ze_r e_r'(Z'Z)^{-1}Z'e_j = e_i'QX'Z\sum_{r=1}^{K}e_r e_r'(Z'Z)^{-1}Z'e_j = e_i'QX'e_j.$on summing over $r = 1, ..., K.$

It follows that $\sum_{r=1}^{K}Diag(\bar{z}_r)e_i'QX'Ze_r$ is a diagonal matrix with j, j^{th} component $e_j'XQe_i = x_j'Qe_i$.We shall refer to this matrix as $Diag(XQe_i)$ whereupon we shall write

$$\sum_{r=1}^{K}tr[(\Omega_{ijs}\beta_0 e_i'QX'Ze_r e_s'H'QX'Diag(\bar{z}_r)XQH)]$$
$$= tr[\Omega_{ijs}\beta_0 e_s'H'QX'Diag(XQe_i)XQH].$$

Summing over $s = 1, .., g+1$ gives $tr[\sum_{s=1}^{g+1}(\Omega_{ijs}\beta_0 e_s')H'QX'Diag(XQe_i)XQH]$.

which when introducing $\sum_{s=1}^{g+1}(e_s\beta_0'\Omega_{ijs}') = \Omega'^*(I_{g+1}\otimes\beta_0)$ enables the final result as follows:

$$E\sum\sum(\hat{\pi}_{rs}-\pi_{rs})(vecV_1)'\{\beta_0 e_i'QX'ZE_{rs}H'QX'\otimes XQH\}I^*vecV_1$$

$$= tr\{\Omega'^{*}(I_{g+1} \otimes \beta_0\ )\ H'QX'Diag(XQe_i)XQH\}.$$

$$= tr\{(I_{g+1} \otimes \beta_0'\ )\Omega^{*}\ H'QX'Diag(XQe_i)XQH\} \tag{47}$$

Summing the three terms in (44),(45) and (46) above yields finally:

$$E\{\tfrac{1}{3!}\sum\sum(\hat{\pi}_{rs} - \pi_{rs})(vec(\hat{\Pi}_1 - \Pi_1))' f^{(3)}_{i,rs}(vec(\hat{\Pi}_1 - \Pi_1))\}$$

$$\begin{aligned} = \quad & e_i'QH(\beta_0' \otimes I_{g+1})\Omega^{*}H'QX'\Delta_{x,z} \\ & + e_i'(QH(I_{g+1} \otimes \beta_0')\Omega^{*}H' + tr(QH(I_{g+1} \otimes \beta_0')\Omega^{*}H').I_{g+k}))QX'\Delta_{xz} \\ & - tr\{(I_{g+1} \otimes \beta_0'\ )\ \Omega^{*}H'QX'Diag(XQe_i)XQH\}. \end{aligned} \tag{48}$$

The results are in terms of the $(g+1) \times (g+1)^2$ matrix $\Omega^{\star}$ which itself is obtained by stacking the matrices Ω_{ijs} where the ij^{th}element of Ω_{ijs} is $\omega_{ijs} = E[v_{it}v_{jt}v_{st}]$. We shall now express Ω^{*} in terms of the structural parameters.

First note that $v_{it} = e_t'U(B')^{-1}_{g+1}e_i$ where $(B')^{-1}_{g+1}$is a $G \times (g+1)$ matrix containing the first $(g+1)$ columns of $(B^{'})^{-1}$and e_t and e_i are $T \times 1$ and $(g+1) \times 1$ unit vectors, respectively. Therefore

$$v_{it} = u_t'b_i,$$

where u_t is a $G \times 1$ vector of structural disturbances at time t and $b_i = (B')^{-1}_{g+1}e_i, i = 1,..,(g+1)$, is a $G \times 1$ vector. Similarly we have

$$\begin{aligned} v_{jt} &= u_t'b_j \\ v_{st} &= u_t'b_s \end{aligned}$$

where $b_j = (B')^{-1}_{g+1}e_j$ and $b_s = (B')^{-1}_{g+1}e_s,\ j,s = 1,..,(g+1)$. With these definitions we can write

$$\begin{aligned} \omega_{ijs} &= E[v_{it}v_{jt}v_{st}] \\ &= E[u_t'b_iu_t'b_ju_t'b_s] \\ &= E[u_t'b_ib_j'u_tu_t'b_s], i,j,s = 1,..,g+1. \end{aligned}$$

Using result (A.27) of Ullah(2005) we then have

$$\omega_{ijs} = [tr(b_jb_i'\Sigma_{ij1}), \ldots, tr(b_jb_i'\Sigma_{ijG})]b_s,$$

where the ij^{th} element of the $G \times G$ matrix Σ_{ijs} is $E[u_{it}u_{jt}u_{st}] = \sigma_{ijs}$. We can rewrite this as

$$\begin{aligned} \omega_{ijs} &= [b_i'\Sigma_{ij1}b_j, \ldots, b_i'\Sigma_{ijG}b_j]b_s \\ &= (\sum_{p=1}^{G} b_i'\Sigma_{ijp}b_je_r')b_s. \end{aligned}$$

where e_p is a $G \times 1$ unit vector. Denoting the first column of Ω_{ijs} by Ω_{i1s} we have

$$\Omega_{i1s} = \begin{pmatrix} (\sum_{p=1}^{G} b_1' \Sigma_{ijp} b_1 e_p') b_s \\ . \\ . \\ . \\ (\sum_{p=1}^{G} b_{g+1}' \Sigma_{ijp} b_1 e_p') b_s \end{pmatrix} = \begin{pmatrix} b_1' (\sum_{p=1}^{G} \Sigma_{ijp} b_1 e_p') b_s \\ . \\ . \\ . \\ b_{g+1}' (\sum_{p=1}^{G} \Sigma_{ijp} b_1 e_p') b_s \end{pmatrix}.$$

Noting that $(B')^{-1}_{g+1} = (b_1, ..., b_{g+1})$,

$$\begin{aligned} \Omega_{i1s} &= ((B')^{-1}_{g+1})' (\sum_{p=1}^{G} \Sigma_{ijp} b_1 e_p') b_s \\ &= ((B')^{-1}_{g+1})' [\Sigma_{ij1}, \ldots, \Sigma_{ijG}] \begin{pmatrix} b_1 e_1' b_s \\ . \\ . \\ . \\ b_1 e_G' b_s \end{pmatrix} \\ &= ((B')^{-1}_{g+1})' \Sigma'^{\star} (b_s \otimes b_1), \end{aligned}$$

where $\Sigma^{'\star}$ is a $G \times G^2$ matrix given by
$\Sigma^{'\star} = [\Sigma_{ij1}, \ldots, \Sigma_{ijG}]$.
Generalising, the p^{th} column of Ω_{ijs} is

$$(B')^{-1}_{g+1})' \Sigma'^{\star} (b_s \otimes b_p),$$

so the matrix Ω_{ijs} is

$$\begin{aligned} \Omega_{ijs} &= ((B')^{-1}_{g+1})' \Sigma'^{\star} \sum_{p=1}^{G} (b_s \otimes b_p) e_p' \\ &= ((B')^{-1}_{g+1})' \Sigma'^{\star} \sum_{p=1}^{G} ((B')^{-1}_{g+1} e_s \otimes (B')^{-1}_{g+1} e_p) e_p' \\ &= ((B')^{-1}_{g+1})' \Sigma'^{\star} ((B')^{-1}_{g+1} \otimes (B')^{-1}_{g+1}) \sum_{p=1}^{G} (e_s \otimes e_p) e_p' \\ &= ((B')^{-1}_{g+1})' \Sigma^{'\star} ((B')^{-1}_{g+1} \otimes (B')^{-1}_{g+1}) (e_s \otimes I_{g+1}) \\ &= ((B')^{-1}_{g+1})' \Sigma^{'\star} (B')^{-1}_{g+1} e_s \otimes (B')^{-1}_{g+1}) \end{aligned}$$

The $(g+1) \times (g+1)^2$ matrix $\Omega'^{*}$is then given by
$\Omega'^{*} = \ (\Omega_{ij1,}\ \Omega_{ij2,...,} \Omega_{ij(g+1)})$
$= ((B')^{-1}_{g+1})' \Sigma'^{*} (B')^{-1}_{g+1} e_1 \otimes (B')^{-1}_{g+1}), ((B')^{-1}_{g+1})' \Sigma'^{\star} (B')^{-1}_{g+1} e_2 \otimes (B')^{-1}_{g+1}), ...$
$((B')^{-1}_{g+1})' \Sigma'^{\star} (B')^{-1}_{g+1} e_{g+1} \otimes B^{-1}_{g+1})$

$=((B')^{-1}_{g+1})'\Sigma^{'\star}((B')^{-1}_{g+1} \otimes (B')^{-1}_{g+1})$
on noting that $(e_1, e_2, ..., e_{g+1}) = I_{g+1}$.Hence finally

$$\Omega'^{*} = ((B')^{-1}_{g+1})'\Sigma^{'\star}((B')^{-1}_{g+1} \otimes (B')^{-1}_{g+1}) \tag{49}$$