

On the use of higher order bias approximations for 2SLS and k-class estimators with non-normal disturbances and many instruments

Abstract

The first and second moment approximations for the k-class of estimators were originally obtained in a general static simultaneous equation model under the assumption that the structural disturbances were i.i.d. and normally distributed. Later, higher-order bias approximations were obtained and were shown to be important especially in highly over-identified cases. It is shown that the higher order bias approximation continues to be valid under symmetric, but not necessarily normal, disturbances with an arbitrary degree of kurtosis, but not when the disturbances are asymmetric. A modified higher-order approximation for the bias is then obtained which includes the case of asymmetric disturbances. The effect of asymmetry in the disturbances is explored in the context of a two equation model where it is shown that the bias of 2SLS may be substantially changed when the skewness factor increases. The use of the bias approximation is illustrated using empirical examples from the literature on return to schooling, which employs a model with many instruments, and on higher education wage premia.

Keywords:

bias approximation, 2SLS, k-class, simultaneous equation model, many instruments, weak instruments

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 $O(T^{-1})$ bias approximation for the 2SLS case to a higher order, viz. $O(T^{-2})$, 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 Srinivasan (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 overidentification. Nagar type approximations have also been used in other contexts, see e.g. Bun and Windmeijer (2010).

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 Nagar 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.

While the Nagar bias approximation has attracted considerable attention, this has not been the case for the higher order approximation perhaps because of the relatively strict assumptions under which it has been presented, such as requiring normality, but also it may be seen as adding little to the Nagar result. In this paper it is shown, firstly, 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. An important and new contribution of this paper is to present the *2SLS* higher order bias approximation in the context of asymmetrically distributed disturbances. An extension of the results to the consistent members of the *k-class* is available in a Supplementary Appendix. Secondly, it is shown that in strongly overidentified cases the Nagar approximation may overstate the bias while the higher order bias approximation may be far more accurate. This arises when the additional terms are opposite in sign to the first order approximation. It is then argued that bias correction is better conducted based upon the higher order bias approximation especially in cases where the number of instruments is large.

The effect of asymmetry on the bias of *2SLS* is initially explored in a simple two-equation simultaneous equation model. It is found that the bias may be significantly affected as the degree of asymmetry increases. Both the approximation terms and the actual percentage bias figures show that asymmetry can play a greater role in *2SLS* bias than suggested in Knight (1984), where the allowable skewness and excess kurtosis values were relatively small (in asymmetric cases). After presenting the main results, the *2SLS* bias approximation is applied to estimates for the return to schooling in Staiger and Stock (1997) based on Angrist and Krueger (1991), and to an empirical model in Fortin (2006) for higher education wage premia.

2. A simple case

The effects of asymmetry on *2SLS* estimation bias are explored in a very simple simultaneous equation model in this section, in an attempt to isolate the key factors, and to demonstrate that the asymmetry effect can be substantial even in cases where the bias values are moderate. The approximate bias expressions for *2SLS* bias in Section 6 and *k-class* bias in Supplementary Appendix A provide a means for estimating the biases in the context of a general model.

The following simple model is considered:

$$y_{1,t} = \beta_1 y_{2,t} + u_{1,t}, \quad (1)$$

$$y_{2,t} = \beta_2 y_{1,t} + \gamma' z_t + u_{2,t} \quad (2)$$

$$t = 1, 2, \dots, T,$$

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

$$\begin{aligned} y_{2,t} &= \beta_2(\beta_1 y_{2,t} + u_{1,t}) + \gamma' z_t + u_{2,t} \\ &= \beta_1 \beta_2 y_{2,t} + \gamma' z_t + u_{2,t} + \beta_2 u_{1,t} \\ &= \pi_2' z_t + v_t \end{aligned} \quad (3)$$

where $\pi_2' = \frac{\gamma'}{1-\beta_1\beta_2}$ and $v_t = \frac{u_{2,t} + \beta_2 u_{1,t}}{1-\beta_1\beta_2}$.

Theorem 2 in Section 6 presents the approximate *2SLS* bias to order $O(T^{-2})$ for estimation of a general model without assuming normality or symmetry in the structural disturbances, and the part due to the asymmetry of the disturbances can be specialised to the following, see Appendix 2, for the model above in the special case of β_1 :

$$\begin{aligned} \Delta^* &= (1 - \beta_1 \beta_2)(\sigma_{111} \beta_2^2 + 2\sigma_{112} \beta_2 + \sigma_{122}) \\ &\times \left[\frac{4 \sum (\gamma' z_t)^3}{(\sum (\gamma' z_t)^2)^3} - \frac{3 \sum (\gamma' z_t) z_t' (Z' Z)^{-1} z_t}{(\sum (\gamma' z_t)^2)^2} \right] \end{aligned} \quad (4)$$

where Z is a $T \times p$ matrix with t -th row z_t' , $t = 1, \dots, T$, and $\sigma_{ijk} = E[u_{it} u_{jt} u_{kt}]$ for $i, j, k = 1, 2$.

Importantly, note that the second factor does not involve β_1 or β_2 or the disturbance third moments but depends upon the exogenous variable coefficient vector γ which can be varied independently so it is clear that, for non-zero third moments of the disturbances and appropriate choice of β_1 and β_2 , Δ^* may become large.

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 in Supplementary Appendix A.

2.1. Numerical and Simulation Results

It is possible to investigate the relationship between the third moment parameters σ_{111} , σ_{112} and σ_{122} and the bias numerically, via the approximation Δ^* and also by Monte Carlo, for given values of β_1 and β_2 . The two-equation model above can be used to investigate the performance of the new higher-order bias approximation (4) in asymmetric cases, and to illustrate how the bias can be shaped by the third moments of structural disturbances.

The estimation of β_1 is considered here. 50 million Monte Carlo replications are used for each moment computation, so that the MC simulated moments may reasonably be called the "true" moments. The sample size is $T = 50$, and fixed exogenous data for each element z_{jt} of z_t , $j = 1, 2, 3$, was drawn from an AR(1) model $z_{jt} = 0.9z_{j,t-1} + \nu_t$ with $\nu_t \stackrel{i.i.d.}{\sim} N(0, 1)$. Attention is drawn in particular to Figure 1, where the bias is depicted for a large number of different skewness cases in two separate models. It is found that the skewness of the structural disturbances in each equation can have a substantial effect on the bias of the *2SLS* estimator, and that the approximation does well in capturing the effect. Table 1 summarises the improvement compared with the approximations due to Nagar and Mikhail over a number of different skewness cases for the structural disturbances.

When generating the data the following are specified: the coefficients of the structural model, the structural covariance matrix Σ , and the u_1 and u_2 skewness coefficients γ_1^1 and γ_1^2 , respectively. Two different ways of generating the structural disturbance term u_t are considered, "Beta" and "Lognormal". The Lognormal cases are genuine multivariate lognormal, while the "Beta" cases are a linear combination of Beta random variables. The underlying parameters of the distributions in both cases are chosen numerically to yield structural disturbances with the desired covariance matrix and with specified skewness values γ_1^1 and γ_1^2 - this is then repeated for various choices of the skewness values. Full details about the data generation are given in Supplementary Appendix C, where, in particular, the values used for the underlying distributional parameters are provided. Summary results are presented in the present section for the Beta cases, where the skewness can be positive or negative, while Supplementary Appendix B presents additional results including the Lognormal cases.

The parameters of the main structural model considered are below, and were chosen numerically with a constraint that the *2SLS* bias should be in a mild or moderate range of 10-20% in absolute value. It is not clear from (4) that large differences in third moment values *necessarily* lead to large changes in bias, as the effect depends on other parameters and the exogenous data. This was found to be the case in simulations, and the parameterisations below were selected as cases where the size of the asymmetry effect was of practical relevance. If the constraint on the size of the *2SLS* bias is removed, there are parameterisations where the effect of varying the third moment values can be very large, but where the *2SLS* biases are relatively extreme. Models A and B below are therefore examples of what can happen due to asymmetry, while avoiding the relatively extreme cases. Values for

the first stage "population" F statistic $F_{pop} = \pi_2 Z' Z \pi_2$, see Cruz and Moreira (2005), are given as a measure of the strength of the instruments. Neither model is considered a weak instruments case ($F_{pop} \leq 1$), while Model A is a good instruments case ($F_{pop} \geq 10$). The reduced form covariance matrix Ω , implied by the choices of Σ , β_1 and β_2 , is also given below.

Model A

$$\beta_1 = 2.733, \beta_2 = -16.388, \gamma = (38.126, 6.205, 3.870)',$$

$$\Sigma = \begin{pmatrix} 38.106 & -11.780 \\ -11.780 & 92.107 \end{pmatrix}, \Omega = \begin{pmatrix} 0.316 & 0.068 \\ 0.068 & 5.111 \end{pmatrix}, F_{pop} = 35.66.$$

Model B

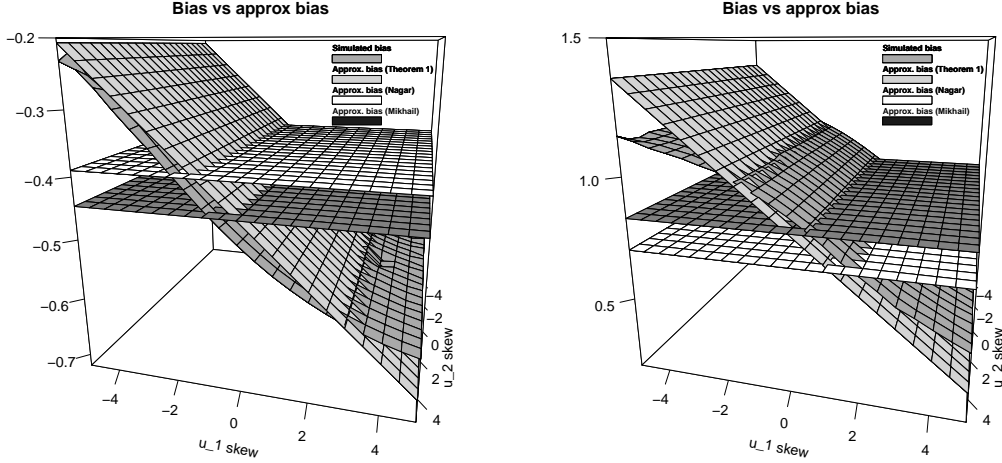
$$\beta_1 = -3.92, \beta_2 = 47.04, \gamma = (39.84, -12.94, -10.71)',$$

$$\Sigma = \begin{pmatrix} 13.06 & 7.48 \\ 7.48 & 60.98 \end{pmatrix}, \Omega = \begin{pmatrix} 0.03 & -0.03 \\ -0.03 & 0.86 \end{pmatrix}, F_{pop} = 4.74.$$

A grid of u_1 and u_2 skewness values is considered in Figure 1, and the true and approximate bias values are computed at each point for Models A and B. Sets of Beta distribution parameters (α_1, β_1) are chosen, see Supplementary Appendix C, to achieve u_1 skewness values in the set $s = \{-5, -4.5, \dots, -0.5, 0, +0.5, \dots, +4.5, +5\}$, and the same is done for the (α_2, β_2) corresponding to u_2 skewness values in this set. $S = s \times s$ then represents a grid of skewness values for u_1 and u_2 . $Bias$ denotes a vector of Monte Carlo simulated ("true") bias values corresponding to members of S , while $Bias_{Nagar}$, $Bias_{Mikhail}$, and $Bias_{New}$ are vectors of the corresponding approximate values using, respectively, the approximations by Nagar, Mikhail, and the result in Theorem 2.

Figure 1 plots the Monte Carlo simulated bias and the improved bias approximation for Models A and B over the skewness pairs, and it is seen that the bias can vary substantially with skewness. The results suggest that the skewness of the structural disturbances can have a substantial effect on the *2SLS* estimation bias, and that the $O(T^{-2})$ approximation taking third moments into account can capture this well. The bias approximations that do not take into account the asymmetry of model disturbances are given by the horizontal planes. The true bias is nonlinear in the skewnesses, but is approximated well by the expression in Theorem 2.

Figure 1: Simulated Bias vs the new $O(T^{-2})$ approximate Bias
 Model A Model B



The nearness of $Bias_{Nagar}$ and $Bias_{New}$ to $Bias$ can be compared in terms of Euclidean distance, and this is also done for the approximation due to Mikhail. The distances d_{Nagar}^{Bias} , $d_{Mikhail}^{Bias}$ and d_{New}^{Bias} 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. The results in Table 1 indicate that the new bias approximation does best in this overall sense for both Model A and Model B.

Table 1: Distance from true bias

	d_{Nagar}^{Bias}	$d_{Mikhail}^{Bias}$	d_{New}^{Bias}
Model A	3.02	2.69	0.63
Model B	5.65	4.70	2.21

The values are Euclidean distances between the vector of simulated biases, $Bias$, and the vectors of approximate biases. For example, $d_{New}^{Bias} = \|Bias - Bias_{New}\|$.

3. Model and Notation

A simultaneous equation model given by

$$By_t + \Gamma z_t = u_t \tag{5}$$

is considered, 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 (5) 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 the system may be written as

$$YB' + Z\Gamma' = U. \quad (6)$$

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

$$y_1 = Y_2\beta + Z_1\gamma + u_1, \quad (7)$$

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:

1. Equation (7) 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).

2. 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 $\Sigma_{ZZ} = \lim_{T \rightarrow \infty} T^{-1}Z'Z$. Following Anderson et. al. (1986, p7) it will also be assumed that $T^{-1}Z'Z = \Sigma_{ZZ} + o(T^{-1})$.

4. Nagar Approximations to the bias

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

$$\hat{\alpha} = \begin{pmatrix} Y_2'Y_2 - \hat{V}_2'\hat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2' - \hat{V}_2' \\ Z_1' \end{pmatrix} y_1. \quad (8)$$

The Nagar approximation for the bias of the *2SLS* estimator for α is given by

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

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 $Q = (X'X)^{-1}$ where $X = (Z\Pi_2 : Z_1)$.

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

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

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

$$C = \begin{bmatrix} (1/T)E(V_2'V_2) & 0 \\ 0 & 0 \end{bmatrix}.$$

It is apparent that when L is relatively large these added terms can be important. Also, in the two-equation case, $\text{tr}(QC)Qq = QCQq$ so that the higher order terms cancel for $L = 3$ while for $L > 3$ the higher order terms will be opposite in sign to the leading bias term, see Table 2 of Hadri and Phillips (1999) where this is noted. Hence in models with a large number of instruments the higher order approximation will be of particular value since reliance on the leading bias term may severely overstate the bias; in such a case bias correction may fail. Some evidence for this is given in Iglesias and Phillips (2012), and the issue will be considered again in an empirical application with large L in Section 7.1.

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. It is shown in Section 6 that the assumption of normality for disturbances can be relaxed, and that the approximation is modified when the disturbances are asymmetric.

5. An Alternative Approach to Approximating the **2SLS** Bias

In Phillips (2000) an alternative approach to finding **2SLS** moment approximations was introduced. To illustrate, consider the estimation of the equation given in (7) by the method of **2SLS**. It is well known that the **2SLS** estimator can be written in the form

$$\hat{\alpha} = \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \hat{\Pi}'_2 Z' Z \hat{\Pi}_2 & \hat{\Pi}'_2 Z' Z_1 \\ Z'_1 Z \hat{\Pi}_2 & Z'_1 Z_1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\Pi}'_2 Z' Z \hat{\pi}_1 \\ Z'_1 Z \hat{\pi}_1 \end{pmatrix} \quad (11)$$

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). As shown in Phillips (2000), it is possible to write both $\hat{\alpha} = f(\text{vec}\hat{\Pi}_1)$ and $\alpha = f(\text{vec}\Pi_1)$, which enables a Taylor expansion of the estimation error $\hat{\alpha}_i - \alpha_i$ about the point $\text{vec}\Pi_1$ as follows:

$$\begin{aligned} f_i(\text{vec}\hat{\Pi}_1) &= f_i(\text{vec}\Pi_1) + (\text{vec}(\hat{\Pi}_1 - \Pi_1))' f_i^{(1)} \\ &+ \frac{1}{2!} (\text{vec}(\hat{\Pi}_1 - \Pi_1))' f_i^{(2)} (\text{vec}(\hat{\Pi}_1 - \Pi_1)) \\ &+ \frac{1}{3!} \sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec}(\hat{\Pi}_1 - \Pi_1))' f_{i,rs}^{(3)} (\text{vec}(\hat{\Pi}_1 - \Pi_1)) \\ &+ \frac{1}{4!} F(\text{vec}(\hat{\Pi}_1 - \Pi_1)) + o_p(T^{-2}) \end{aligned} \quad (12)$$

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

The bias approximation to order T^{-1} is obtained by taking expectations of the first two terms of the stochastic expansion to yield:

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

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 $\Omega_1^{\text{vec}} = \Omega_1 \otimes I_T$, 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:

$$E(\hat{\alpha}_i - \alpha_i) = e_i' Q q + o(T^{-1}), \quad (13)$$

where e_i is (here and throughout) an appropriately sized row vector of zeros with a 1 in position i . However, to obtain his approximation Nagar assumed that the disturbances were normally distributed while here it need only be assumed that the row vectors of V_1 obey the *Gauss Markov* assumptions so that the row vectors are serially uncorrelated and homoscedastic.

To find the bias approximation to order T^{-2} it will also be necessary to evaluate the expected values for each of the terms in the expansion in (12). It has proved possible to find an explicit representation for the first three terms, see Phillips (2000), but it is quite difficult to do so for the fourth term. Notice that the third term

$$\frac{1}{3!} \sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec}(\hat{\Pi}_1 - \Pi_1))' f_{i,rs}^{(3)} (\text{vec}(\hat{\Pi}_1 - \Pi_1)) \quad (14)$$

is a linear function of products of three components of $\text{vec}(\hat{\Pi}_1 - \Pi_1)$ and the bounded third order derivatives which are evaluated at $\text{vec}(\Pi_1)$. It may be shown that the third moment of the least squares regression estimator is $O(T^{-2})$, see for example Phillips and Liu-Evans (2011), from which it may be deduced that the expectation of the third term in (12) is also $O(T^{-2})$ and this is evaluated in Appendix 1.

While an explicit representation cannot be found for the fourth term in the expansion, $F(\text{vec}(\hat{\Pi}_1 - \Pi_1))$, it turns out that we do not need to do so. It may readily be deduced that it is a linear function of fourth order products of the components of $\text{vec}(\hat{\Pi}_1 - \Pi_1)$ and the bounded fourth order derivatives evaluated at $\text{vec}(\Pi_1)$. 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 Phillips and Liu-Evans (2011) where we demonstrate that the fourth moment of the least squares regression estimator in the general linear regression model 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, plays no role in our approximation to $O(T^{-2})$. This latter result is also implicit in the work of Ullah, see Ullah (2004).

Because of this, the expectation of the fourth term in (12) to the order of the approximation will not depend upon the actual distribution of the errors provided the moment condition on $\text{vec}V_1$ is satisfied. Hence the expectation based upon the normal distribution, which has already been found by Mikhail, can also be employed for other distributions and in finding the higher order bias approximation to order T^{-2} the relevant part of the Mikhail result will simply be added.

6. The Higher Order Bias Approximations for **2SLS**

In this section the bias approximation is presented 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 (12) is trivially zero while the evaluation

of the fourth term has already been done by Mikhail for the normal distribution and, as noted at the end of Section 5, 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. The following theorem is deduced:

Theorem 1. *In the model of Section 3 where the errors are symmetrically but not necessarily normally distributed, the bias of the i^{th} component of the 2SLS estimator in (11) is given by*

$$E(\hat{\alpha}_i - \alpha_i) = (L - 1)[e_i'Qq + \text{tr}(QC)e_i'Qq - (L - 2)e_i'QCQq] + o(T^{-2})$$

for $i = 1, 2, \dots, g + k$.

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 (12) it is found that the revised approximation is given in the following.

Theorem 2. *In the model of Section 3 where the errors may be asymmetrically distributed, the bias of the i^{th} component of the 2SLS estimator is given by*

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

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 Ω_{ijs} , $s = 1, \dots, (g+1)$, which have ij -th element $\omega_{ijs} = E(v_{pi}v_{pj}v_{ps})$, $p = 1, \dots, T$. The $T \times 1$ vector Δ_{xz} has p^{th} component $x_p'(X'X)^{-1}x_p - z_p'(Z'Z)^{-1}z_p$, $\beta_0 = (-1, \beta)'$, and $H = \begin{pmatrix} 0 & I_g \\ 0 & 0 \end{pmatrix}$ is a $(g+k) \times (g+1)$ selection matrix. When Ω^* is zero the bias approximation reduces to that of Mikhail (1972).

The proof of the above is given in Appendix 1. If it is required to express the asymmetry effect in terms of the structural parameters one can replace the transpose of Ω^* with its

structural parameter representation, viz, $\Omega^{*'} = ((B')_{g+1}^{-1})' \Sigma^* ((B')_{g+1}^{-1} \otimes (B')_{g+1}^{-1})$ where $(B')_{g+1}^{-1}$ comprises the first $g + 1$ columns of $(B')^{-1}$, Σ^* is the $G \times G^2$ matrix formed as $\Sigma^* = (\Sigma_{ij1}, \Sigma_{ij2}, \dots, \Sigma_{ijG})$ and Σ_{ijk} is a $G \times G$ symmetric matrix with general element equal to the third moment $\sigma_{ijk} = E[u_{it}u_{jt}u_{kt}]$, $i, j, k = 1, \dots, G$.

Notice that the asymmetry effect does not depend explicitly on $L - 1$ and so it is present whatever the order of overidentification; in particular, the asymmetry effect does not go to zero in this case. The plots in Figure 2 are based on estimates of β_1 in a sequence of models where $L = 1$ and where different values for the coefficient on the excluded non constant exogenous variable are considered, see Model Group L below. While the other structural parameters are the same as Model A and were kept fixed, a choice of $\gamma^* = (24.49, -0.89)$ for the vector of reduced form coefficients yielded the largest asymmetry effect that could be found in the region $[-100, 100]^2$.

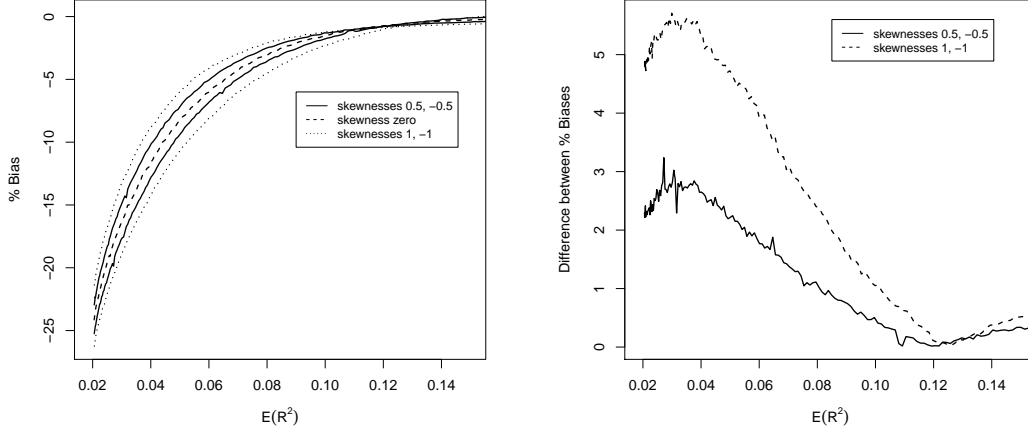
Model Group L

$$\beta_1 = 2.73, \beta_2 = -16.39, \gamma = (24.49, -0.89(0.5 + s))',$$

$$\Sigma = \begin{pmatrix} 38.11 & -11.78 \\ -11.78 & 92.11 \end{pmatrix}.$$

On the left in Figure 2 the percentage biases are plotted against $E[R^2]$, the Monte Carlo average of the sample R^2 values for the first stage reduced form estimate. Lower values for $E[R^2]$ indicate that the instruments are weaker, and correspond to smaller values of s . The simulations are run for three of the Beta cases used earlier. In particular, the two dotted lines correspond to structural disturbances with u_1 and u_2 skewnesses (γ_1^1, γ_1^2) of either $(-1, 1)$ or $(1, -1)$, while the two solid lines correspond to skewnesses of either $(-0.5, 0.5)$ or $(0.5, -0.5)$, and the dashed line is a case of zero skewness. On the right in Figure 2, the dotted and solid lines plot the vertical difference between the respective lines in the previous plot. It is evident here that the moderate levels of skewness are having a substantial effect on the bias when the instrument is weak but not too weak. When $E[R^2] = 0$, the *2SLS* estimator will be equivalent to *OLS*, which has a smaller asymmetry effect in this example.

Figure 2: Simulated Bias of $\hat{\beta}_{1,2SLS}$ when $L = 1$
 % Bias Differences in % bias



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, there have been no results for substantial departures from symmetry until now nor, indeed, for cases with a large number of instruments.

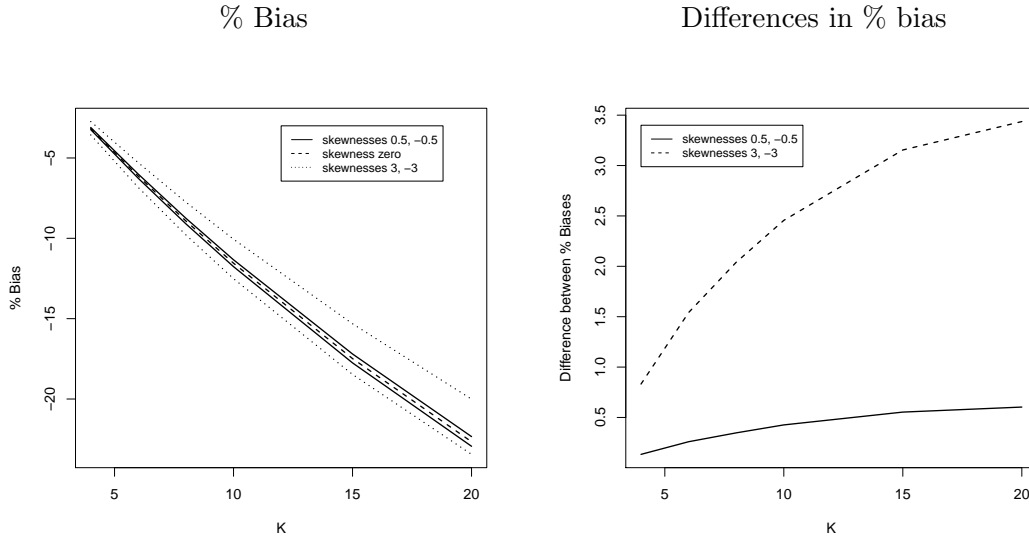
Theorem 2 suggests that the magnitude of the bias will increase with K , due to the three terms involving Δ_{xz} . Appendix 3 shows that the first of these is bounded below as follows

$$|c'\Delta_{xz}| \geq \frac{K - g - k}{\sqrt{T}} \|c\|_2 |\tilde{c}|$$

where $c = (e'_i QH(\beta'_0 \otimes I_{g+1})\Omega^* H Q X')'$ while \tilde{c} and $\|c\|_2$ are scalars that do not go to zero with K . For a given sample size T , this is increasing in K , and the other two terms in Theorem 2 involving Δ_{xz} are of the same form and therefore have similar lower bounds that increase with T . In practice there will be many cases where the effect of asymmetry does not increase with K . Theorem 2 applies to *2SLS*, but the estimator will be equal to *OLS* when $K = T$, and the *OLS* asymmetry effect may be relatively low. For sufficiently weak instruments, there may not be much change at all in the asymmetry effect from increasing K , and it could be decreasing over some or all values of K .

Figure 3 presents percentage biases and differences as in Figure 2, but now these are plotted against K as additional instruments are added to the model. To see a case where the asymmetry effect would increase with K an initial model with relatively strong instruments was chosen ($F_{pop} = 320.89$), then, to avoid $E[R^2]$ growing too quickly towards 1, the additional instruments were given a very small reduced form coefficient. The results in the figure are for a similar model to Model A initially, but with the three reduced form coefficients multiplied by three, so that $\gamma = (3 * 38.126, 3 * 6.21, 3 * 3.87)'$. Additional instruments are added with a reduced form coefficient of 0.000000001. It can be seen that the effect of asymmetry on bias does increase in this example, though the effect is more noticeable at relatively high skewness values of 3 and -3.

Figure 3: Simulated asymmetry effect on the bias of $\hat{\beta}_{1,2SLS}$ for different values of K



7. Empirical applications

7.1. Compulsory school attendance and earnings, Angrist and Krueger (1991)

Angrist and Krueger (1991) investigate the effect of compulsory school attendance on schooling and earnings levels using US Census data. The topic continues to be interest, see e.g. Devereux and Hart (2010). Others have investigated the wider benefits of compulsory education, see Stephens and Yang (2014). Angrist and Krueger (1991) used *2SLS* estimates to comment on the bias in *OLS* estimation of the return to education. Previous studies had focused on correcting for *OLS* estimation bias caused by omitted variables that would be positively correlated with education years, such as "innate ability", but the *2SLS* estimates

in Angrist and Krueger suggested a negative bias in the *OLS* estimation of the return to education. That is to say, the *2SLS* results suggested that the returns to education were higher than previously thought.

While the original analysis was concerned with the *OLS* bias, measured using *2SLS*, this section uses the approximation in Theorem 2 to account for the estimation bias that arises when using *2SLS*. The study by Angrist and Krueger is well known, making it appealing for the purpose of illustration. However, the possibility of weak instruments, see Bound, Jaeger and Baker (1995), may affect the accuracy of the bias approximations. Thus the specification in Case 2 of Staiger and Stock (1997) is used, which Cruz and Moreira (2005) find to be free from the weak instruments problem. The sample sizes are large, ranging from 247,199 to 486,926 across the three survey cohorts, yet *2SLS* biases of around 3 to 4% in absolute value persist. The interest is in estimation of α in the following structural equation for log wage, $\ln(W)$, where i denotes the i th individual while c and j denote the year and quarter in which the individual was born:

$$\begin{aligned}\ln W_i &= X_i' \beta + \sum_c Y_{ic} \zeta_c + \alpha E_i + u_i \\ E_i &= X_i' \pi + \sum_j Q_{ij} \delta_j + \sum_c \sum_j Y_{ic} Q_{ij} \theta_{jc} + \epsilon_i.\end{aligned}$$

Here E_i denotes the education years of the i th individual, X_i is a vector of covariates, Q_{ij} is a dummy variable taking value 1 if individual i was born in quarter j , while Y_{ic} is a dummy variable taking value 1 if individual i was born in year c . The quarter of birth Q_{jc} and interaction terms $Y_{ic} Q_{ij}$ are correlated with education years, but seem unlikely to be correlated with omitted variables from the wage equation.

Column (1) of Table 2 replicates the estimates for Case II in Tables 2 of Staiger and Stock (1997), and provides estimates of the skewness of the wage equation structural disturbances, while Column 2 provides the estimated percentage biases, obtained using Theorem 2, with $E[V_2' u_1]$ estimated by $\hat{V}_{2,OLS}' \hat{u}_{1,2SLS}$ and the elements ω_{ijs} of Ω^* estimated by $\hat{\omega}_{ijs} = \frac{1}{T} \sum_{t=1}^T \hat{v}_{ti} \hat{v}_{tj} \hat{v}_{ts}$, where \hat{v}_{tl} is the tl -th element of \hat{V}_{OLS} . *Matlab* code for this is available upon request. It can be seen that the estimated biases, as a percentage of the 2SLS estimate, are between -3.47% and 3.82%, depending on the sample. The $O(T^{-2})$ estimates for the 1930-39 cohort in Column (2) suggest that the returns to schooling were 3.47% higher, while the estimated returns for the 1920-29 and 1940-49 cohorts were biased upwards by 2.05% and 3.82%, respectively.

Table 2: Estimation of α , and the *2SLS* bias, Staiger and Stock (1997)

		(1)	(2)	(3)
		Staiger and Stock	$\widehat{\% \text{ Bias}}$	$\widehat{\% \text{ Bias}}_{Nagar}$
1920-29	α	0.0633	2.05	2.53
($n = 247199$)	$\hat{\gamma}(u_1)$	-2.23		
1930-39	α	0.0806	-3.47	-4.22
($n = 329509$)	$\hat{\gamma}(u_1)$	-2.40		
1940-49	α	0.0393	3.82	4.33
($n = 486926$)	$\hat{\gamma}(u_1)$	-2.31		

Column (1) replicates values from Case 2 of Staiger and Stock (1997). Column (2) estimates the bias using Theorem 2, while (3) estimates the bias using the original Nagar approximation.

Despite the estimated skewnesses being substantial, the amount of bias that can be attributed to the third moments of the structural disturbances, according to the analytical approximation, turns out to be negligible in the present example. It can be seen, though, that the higher-order bias terms taken as a whole are important. The $O(T^{-2})$ bias terms are non trivial and opposite in sign to the Nagar $O(T^{-1})$ bias, so that the total bias, $\widehat{\% \text{ Bias}}$, is relatively small. Thus if the Nagar approximation were used to bias-correct in this case it would make the bias worse than if the *2SLS* estimator were not corrected at all. When the bias approximation up to order $O(T^{-2})$ is used, though, this mistake is avoided. The importance of the higher-order terms may be expected, as there are a total of 30 instruments in this specification, making the order of overidentification relatively high at $L = 29$. The higher-order bias terms will tend to become more important at higher values of L , and it was noted in Section 4 that the contribution will be opposite in sign to the $O(T^{-1})$ bias.

7.2. The own-cohort supply effect on college wage premia, Fortin (2006)

Fortin (2006) investigates the relationship between state-specific supply of higher education college labour in the US and the wage premia obtained by college graduates between 1979 and 2002. The interest is in the estimation of α_1 in the following inverse relative demand equation for state s at time t , a 3-year pooled time period:

$$r_{st} = \alpha_0 + \alpha_1 q_{st} + \alpha_2 q_{st}^O + \alpha_3' Y_{st} + S_s + P_t + \varepsilon_{st}$$

where $r_{st} = \ln(w_{cst}^Y/w_{hst}^Y)$ is the college-high school wage gap for young workers, $q_{st} = \ln(C_{st}^Y/H_{st}^Y)$ is the relative supply of young workers with college education to those without, q_{st}^O is the same but for old workers, Y_{st} is a vector of observable demand variables, while S_s and P_t represent state and time effects, respectively. The coefficients can be collected in a vector α . The coefficient α_1 therefore reflects the effect on the wage premium of shifting

the relative supply of young college workers, and it has implications for higher education policies and wage inequality.

The relative supply of new college graduates, q_{st} , is likely to be influenced by ε_{st} , the inverse relative demand shocks to the college wage premium, though, and it is therefore an endogenous variable. One of the ways Fortin (2006) accounts for this endogeneity is by using a feasible weighted *2SLS* estimation, with a number of instruments for q_{st} . There are four instruments used in Panel C of Table 8 in the paper by Fortin: three supply-related determinants of lagged enrollment rates in public colleges, along with a variable representing the lagged level of enrollment in private colleges, making the order of overidentification $L = 3$.

The coefficient on q_{st} is re-estimated by *2SLS* here, and the bias is estimated using the approximation in Theorem 2. The first column of Table 3 presents the feasible weighted *2SLS* estimates of α_1 due to Fortin, while the second column is *2SLS*, with heteroskedasticity robust standard errors in parenthesis. Two estimates are obtained, corresponding to US states with relatively low and high enrollment in private colleges, where the state educational policies under consideration do not apply. The results are qualitatively similar to those of Fortin, in that the point estimates suggest a greater supply effect on the wage premium in states where enrollment in private colleges is lower; α_1 is still significant at the 10% level in this part of the sample, and not significant for the states with relatively high private enrollment, the point estimate also changes sign in this latter case, so that it is now the same as for the Low Enrollment sample. Moreover, the difference between the estimates in the two samples is not as pronounced.

Column 3 provides the estimated percentage biases, obtained using Theorem 2 as before. It can be seen that the estimated biases are quite substantial at 6.6% and 33.6%. In parenthesis are the proportions of the biases that can be attributed to the third moments of the structural disturbances according to the analytical approximation. The values are still low in the present example, though this may be expected given the estimated skewness of the structural equation disturbances, which is almost zero.

Table 3: Estimation of α_1 , and bias statistics

		(1)	(2)	(3)	(4)
Private Enrollment		Fortin (2006)	<i>2SLS</i>	$\widehat{\% \text{ Bias}}$	$\hat{r}_{10\%}$
Low ($N = 217$)	α_1	-0.22 (0.08)	-0.13 (0.08)	6.6% (0.39%)	6.75% (0.36%)
	$\hat{\gamma}(u_1)$		-0.06		-0.15
High ($N = 126$)	α_1	0.11 (0.14)	-0.026 (0.08)	33.6% (0.63%)	93.9% (0.95%)
	$\hat{\gamma}(u_1)$		0.20		0.30

Column (1) replicates the Feasible Weighted 2SLS results, from Panel C of Table 8 in Fortin (2006). Column (2) presents 2SLS estimates with robust standard errors, (3) presents estimates of the 2SLS bias using Theorem 2 with the proportion attributed to skewness in parenthesis, and (4) presents values for the measure of relative bias risk and the proportion attributed to skewness. The *2SLS* results are different to what would be obtained using the `Stata` code provided on the AER website after removing the weighting for heteroskedasticity, due to the way in which `Stata` removes colinear variables in *2SLS* regression. In particular, Z in the present setup contains all exogenous data.

The fourth column reports an approximate measure of bias-related risk using the new analytical bias approximation. This would be relatively challenging to compute for the Angrist and Krueger (1991) application due to the large sample sizes, but it is straightforward in the present setting. Recall that, given a model defined in terms of a parameter vector Θ belonging to a space C , a standard measure of global estimation risk for a single parameter θ , see for example E.L. Lehmann (1983), is the *maximum risk* given by $r = \sup_{\Theta \in C} R(\theta, \delta)$ where δ is the estimator of θ and $R(\theta, \delta) = E_{\Theta}[L(\theta, \delta)]$ is a risk function defined in terms of a loss function L . If the loss function is set to $L(\theta, \delta) = |\theta - \delta|$ then R is the absolute bias in estimation of θ at a particular point $\Theta \in C$ and r is the maximum bias over C . The term $\hat{r}_{10\%}$ is an empirical version of this, where the approximate relative bias using Theorem 2 is numerically maximised over values for the Equation 1 structural coefficients that lie within narrow 10% confidence intervals. A small confidence level of 10% is chosen here to compare the bias risk in the Low and High private enrollment states, because wider confidence intervals for α_1 in the High private enrollment states include zero, where the % bias is undefined; the confidence interval at 10% is (-0.04, -0.016).

If the bias were simulated for the Fortin model in order to calculate R for a particular parameterisation, an underlying parametric distribution for the structural disturbances would be chosen, along with the structural coefficients for both equations. The absolute bias could be calculated this way for many different choices of the structural coefficients, and then the maximum could be taken as an estimate of the global risk r . The approximate bias, however, just requires a specification for $E[V_2' u_1]$ in q , and Ω^* . By setting the Equation 1 structural coefficients to a new set of values $\tilde{\alpha}$, different to the *2SLS* estimates, and using the observed data y_1 , Y_2 , and Z_1 , the term $E[V_2' u_1]$ can be set to $\hat{V}_{2,OLS}' \tilde{u}_1$, where $\tilde{u}_1 = y_1 - Z_1 \tilde{\alpha}$. By doing this, the structural covariance matrix, and the endoge-

nous variable coefficient in the relative supply equation (see Fortin (2006)), are implicitly constrained to take values within the set where the expected value $E[V_2' u_1]$ is equal to $\hat{V}_{2,OLS}' \tilde{u}_1$. The approximate *relative* biases are computed for each $\tilde{\alpha}$, and $\hat{r}_{10\%}$ denotes the maximum obtained numerically.

The statistic suggests that there is more (relative) bias risk associated with *2SLS* estimation in the high private enrollment sample than in the low private enrollment sample, in addition to the higher point estimate of the bias given in Column 3.

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 more that is known about its properties the better. 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 particular importance when equations are heavily overidentified since then the higher order terms may be relatively large. The fact that the approximation holds under symmetric distributions and any degree of kurtosis obviously increases its applicability in practical cases. However, when the errors are asymmetrically distributed it is seen that the Mikhail approximation to order T^{-2} no longer holds and we have presented the correct approximation for such cases.

The earlier work of Knight indicated that the *2SLS* bias is not much affected by disturbance kurtosis and our analysis supports this since the $O(T^{-2})$ bias approximation is unchanged in its presence. In fact kurtosis is only relevant at the $O(T^{-3})$ level of approximation. Knight also concluded that the *2SLS* bias was relatively robust to skewness in the disturbances. Our analysis supports this too for small values of the skewness measure since the asymmetry effect is of order T^{-2} . However the effect of asymmetry on the bias can be significant at larger skewness values. More generally, the larger the degree of skewness the greater the effect on the *2SLS* bias although the effect of asymmetry on the bias is not unidirectional. It is found that an increase in disturbance skewness may have a different impact depending on the sign of the skewness measure. It is shown, using two simple examples, that increasing positive skewness can lead to a reduction in positive bias or an increase in negative bias, while an increase in negative skewness may reduce negative bias or increase positive bias. Thus an estimator may have more/less bias as a result of asymmetry.

Given the above, it is clear that asymmetry is an important concern when estimating by *2SLS*, as its contribution to bias can be substantial and in either direction. As long as there is significant estimator bias we shall wish to reduce it and, fortunately, we have the means of doing so since a bias approximation to order T^{-2} , which includes the asymmetry effect, is available along with code for implementing it. This can be used directly for bias reduction. It is also possible to estimate the incremental bias that is due to asymmetry and check the effect it has on the overall bias. Bai & Ng (2005) provide a test for skewness of

error terms in multiple regression models, and the analysis above suggests that an extension to models containing one or more endogenous variables would be worthwhile.

While we give particular attention to the effect of skewness on estimator bias, it has been shown that with or without skewness the higher order bias approximation is to be preferred, particularly in strongly overidentified cases, since bias correction can break down if the higher order bias terms are neglected. Indeed the foregoing discussion makes the case for bias correction to be generally based on the higher order bias approximation.

Finally, the *k-class* of estimators where $k < 1$ are also of interest partly because estimators in this class have all necessary moments while *2SLS* has moments up to the order of overidentification. As mentioned above, the higher order bias approximation is available for this class of estimators too.

References

- Anderson, T.W., N. Kunitomo and K. Morimune (1986). Comparing Single Equation Estimators In A Simultaneous Equation System, *Econometric Theory*, 2, 1-32.
- Angrist, J.D. and A.B. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*. 106(4), 979-1014.
- Bai, J. and S. Ng (2005). Tests for Skewness, Kurtosis, and Normality for Time Series Data. *Journal of Business and Economic Statistics*. 23 (1), 49-60.
- Bound, J., Jaeger, D., and R. Baker (1995). Problems With Instrumental Variables Estimation When the Correlation Between Instruments and the Endogenous Explanatory Variable Is Weak, *Journal of the American Statistical Association*. 90, 443-450.
- Bun, M.G. and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal*. 13(1), 95-126.
- Cruz, L.M. and M.J. Moreira (2005). On the Validity of Econometric Techniques With Weak Instruments: Inference on Returns to Education Using Compulsory Schooling Laws. *Journal of Human Resources*. 40, 393-410.
- Devereux, P.J. and R.A. Hart (2010). Forced to be rich? Returns to Compulsory Schooling in Britain. *Economic Journal*. 120, 1345-1364.
- Fortin, N.M. (2006). Higher-Education Policies and the College Wage Premium: Cross-State Evidence from the 1990s. *American Economic Review*. 96(4), 959-987.
- Hadri, K. and G.D.A. Phillips (1999). The accuracy of the higher order bias approximation for the 2SLS estimator, *Economics Letters*. 62(2), 167-174.
- Hansen, C., Hausman, J. and W. Newey (2008). Estimation with Many Instrumental Variables. *Journal of Business and Economic Statistics*. 26(4), 398-422.
- 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 (2010). The Bias to order T^{-2} for the general k -class estimator in a simultaneous model. *Economics Letters*. 109, 42-45.
- 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 Economics Statistics*. 30 (4), 505-520.
- 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. *Journal of Econometrics* 27, 39-60
- Lehmann, E.L. (1983). *Theory of Point Estimation*. London: Chapman and Hall.
- Magnus, J. and Neudecker, H. (1979). The commutation matrix: some properties and applications. *Annals of Statistics* 7, 381-394.
- 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*. Cambridge University 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.

Phillips, G.D.A. and Gareth Liu-Evans Cardiff Business (2011). The Robustness of Higher-Order 2SLS and General k-Class bias Approximations to Non-Normal disturbances. E2011/20, Cardiff Economics Working Papers.

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

Srinivasan, T.N (1970). Approximations to the Finite Sample Moments of Estimators Whose Finite sample Distributions are Unknown. *Econometrica* 38, p533-541.

Staiger, D., and J. Stock (1997). Instrumental Variables Regression With Weak Instruments. *Econometrica*. 65, 557-586.

Stephens, M. and D. Yang (2014). Compulsory Education and the Benefits of Schooling. *American Economic Review*. 104(6), 1777-1792.

Ullah, A. (2004). *Finite Sample Econometrics*. Oxford University Press.

Appendix

Appendix 1: Theorem 2

In this Appendix we evaluate the expectation of the skewness term

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

We commence with the following:

Lemma 1. $E\{(\hat{\pi}_{rs} - \pi_{rs})(\text{vec}V_1)(\text{vec}V_1)'\} = \Omega_{ijs} \otimes \text{Diag}(z_r)$ where

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

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

To see this we proceed from

$$\begin{aligned} E\{(\hat{\pi}_{rs} - \pi_{rs})(\text{vec}V_1)(\text{vec}V_1)'\} &= E\{e_r'(Z'Z)^{-1}Z'v_s(\text{vec}V_1)(\text{vec}V_1)'\} \\ &= E\left\{e_r'(Z'Z)^{-1}Z'v_s \begin{bmatrix} v_1v_1' & v_1v_2' & \dots & v_1v_{g+1}' \\ v_2v_1' & v_2v_2' & \dots & v_2v_{g+1}' \\ \cdot & \cdot & \dots & \dots \\ v_{g+1}v_1' & v_{g+1}v_2' & \dots & v_{g+1}v_{g+1}' \end{bmatrix}\right\} \end{aligned}$$

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 $E(\bar{z}_r'v_s v_i v_j')$ with general term $E(\bar{z}_r'v_s v_{pi} v_{qj})$ for $p, q = 1, 2, \dots, T$, which is non-zero only when the stochastic terms are of the same time period. When $p = q$ it is seen that $E(\bar{z}_r'v_s v_{pi} v_{pj}) = E(\bar{z}_{pr} v_{ps} v_{pi} v_{pj}) = \bar{z}_{pj} \omega_{ijs}$ where \bar{z}_{pr} is the p^{th} component of \bar{z}_r and $E(v_{pi} v_{pj} v_{ps}) = \omega_{ijs}$. More generally,

$$\begin{aligned} E(\bar{z}_r'v_s v_i v_j') &= \omega_{ijs} \begin{bmatrix} \bar{z}_{r1} & 0 & \dots & 0 \\ 0 & \bar{z}_{r2} & & 0 \\ & & \cdot & \cdot \\ 0 & 0 & \dots & \bar{z}_{rT} \end{bmatrix} \\ &= \omega_{ijs} \text{Diag}(\bar{z}_r) \text{ for } i, j, s = 1, 2, \dots, g+1. \end{aligned}$$

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

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

is equal to the sum of the following **three terms**:

$$\begin{aligned} & \sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec}V_1)' \{H' Q e_i \beta_0' E_{rs}' Z' X Q H \otimes (P_X - P_Z)\} \text{vec}V_1 \\ & + \sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec}V_1)' \{H' Q (X' Z E_{rs} H' + H E_{rs}' Z' X) Q e_i \beta_0' \otimes (P_X - P_Z)\} \text{vec}V_1 \\ & + \sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec}V_1)' \{\beta_0 e_i' Q X' Z E_{rs} H' Q X' \otimes X Q H\} I^* \text{vec}V_1 \end{aligned}$$

where E_{rs} is a $K \times (g+1)$ matrix of rank one with unity in the r, s^{th} position and zeroes elsewhere, and I^* is a $T(g+1) \times T(g+1)$ commutation matrix, see Magnus and Neudecker (1979).

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.

(a) First term. We examine

$$\begin{aligned} & (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec}V_1)' \{H' Q e_i \beta_0' E_{rs}' Z' X Q H \otimes (P_X - P_Z)\} \text{vec}V_1 \\ & = \text{tr}\{(\hat{\pi}_{rs} - \pi_{rs}) (\text{vec}V_1) (\text{vec}V_1)' \{H' Q e_i \beta_0' E_{rs}' Z' X Q H \otimes (P_X - P_Z)\}\} \end{aligned} \quad (16)$$

where $\hat{\pi}_{rs} - \pi_{rs}$, which is the $K(r-1) + s^{\text{th}}$ component of $\text{Vec}(\hat{\Pi}_1 - \Pi_1)$, 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 $T \times 1$ 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

$$\text{tr}\{e_r' (Z' Z)^{-1} Z' v_s (\text{vec}V_1) (\text{vec}V_1)' \{H' Q e_i \beta_0' E_{rs}' Z' X Q H \otimes (P_X - P_Z)\}\}. \quad (17)$$

We have shown above in Lemma 1 that

$$E\{e_r' (Z' Z)^{-1} Z' v_s (\text{vec}V_1) (\text{vec}V_1)'\} = \Omega_{ijs} \otimes \text{Diag}(\bar{z}_r)$$

so it follows that

$$\begin{aligned} & E\{\text{tr}\{e_r' (Z' Z)^{-1} Z' v_s (\text{vec}V_1) (\text{vec}V_1)' \{H' Q e_i \beta_0' E_{rs}' Z' X Q H \otimes (P_X - P_Z)\}\}\} \\ & = \text{tr}\{\Omega_{ijs} \otimes \text{Diag}(\bar{z}_r) \{H' Q e_i \beta_0' E_{rs}' Z' X Q H \otimes (P_X - P_Z)\}\} \\ & = \text{tr}\{\Omega_{ijs} H' Q e_i \beta_0' E_{rs}' Z' X Q H\} \text{tr}\{\text{Diag}(\bar{z}_r) (P_X - P_Z)\}. \end{aligned}$$

Some simplification is possible by writing

$$\text{tr}\{Diag(\bar{z}_r)(P_X - P_Z)\} = \bar{z}'_r \Delta_{xz}.$$

Next we shall write $E_{rs} = e_r e'_s$ where e_s is a $(g+1) \times 1$ 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

$$e'_i Q H \Omega_{ijs} H' Q X' Z e_r \beta_{s0} \bar{z}'_r \Delta_{xz}. \quad (18)$$

Finally we need to find the value of

$$\sum_{r=1}^K \sum_{s=1}^{g+1} e'_i Q H \Omega_{ijs} H' Q X' Z e_r \beta_{s0} \bar{z}'_r \Delta_{xz}. \quad (19)$$

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

$$\begin{aligned} & \sum_{r=1}^K e'_i Q H \Omega_{ijs} H' Q X' Z e_r \beta_{s0} \bar{z}'_r \Delta_{xz} \\ &= \beta_{s0} e'_i Q H \Omega_{ijs} H' Q X' Z \sum_{r=1}^K e_r e'_r (Z' Z)^{-1} Z' \Delta_{xz} \\ &= \beta_{s0} e'_i Q H \Omega_{ijs} H' Q X' \Delta_{xz} \end{aligned}$$

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

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

$$\begin{aligned} & E \sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec} V_1)' \{H' Q e_i \beta'_0 E'_{rs} Z' X Q H \otimes (P_X - P_Z)\} \text{vec} V_1 \\ &= e'_i Q H \left(\sum_{s=1}^{g+1} \beta_{s0} \Omega_{ijs} \right) H' Q X' \Delta_{xz} \\ &= e'_i Q H (\beta'_0 \otimes I_{g+1}) \Omega^* H' Q X' \Delta_{xz} \end{aligned} \quad (20)$$

Here we have used the result that $\sum_{s=1}^{g+1} \beta_{s0} \Omega_{ijs}$ 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 Ω_{ijs} , $s = 1, \dots, g+1$.

(b) Second term. Recall that this is

$$\sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs}) (\text{vec} V_1)' \{H' Q (X' Z E_{rs} H' + H E'_{rs} Z' X) Q e_i \beta'_0 \otimes (P_X - P_Z)\} \text{vec} V_1.$$

The expected value for this term can be found using the same approach as for the above term as is shown in Phillips and Liu-Evans (2011). In fact

$$\begin{aligned}
& E\left\{\sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs})(\text{vec}V_1)' \{H'Q(X'ZE_{rs}H' + HE'_{rs}Z'X)Qe_i\beta'_0 \otimes (P_X - P_Z)\} \text{vec}V_1\right\} \\
&= e'_i (QH\Omega^*(I_{g+1} \otimes \beta_0)H' + \text{tr}(QH\Omega^*(I_{g+1} \otimes \beta_0)H').I_{g+k}) QX'\Delta_{xz}. \\
&= e'_i (QH\Omega^*(I_{g+1} \otimes \beta_0)H' + \text{tr}(QH(I_{g+1} \otimes \beta'_0)\Omega^*H').I_{g+k}) QX'\Delta_{xz}.
\end{aligned}$$

(c) **Third and final term.** The expected value is

$$\begin{aligned}
& E\left\{\sum_{r=1}^K \sum_{s=1}^{g+1} (\hat{\pi}_{rs} - \pi_{rs})(\text{vec}V_1)' \{\beta_0 e'_i QX'ZE_{rs}H'QX' \otimes XQH\} I^* \text{vec}V_1\right\} \\
&= \text{tr}\{(I_{g+1} \otimes \beta'_0)\Omega^*H'QX' \text{Diag}(XQe_i)XQH\}
\end{aligned}$$

where $\text{Diag}(XQe_i)$ is a diagonal matrix with j, j^{th} component $e'_j XQe_i = x'_j Qe_i$.

Summing the three terms in (a), (b) and (c) above yields finally:

$$\begin{aligned}
& E\left\{\frac{1}{3!} \sum \sum (\hat{\pi}_{rs} - \pi_{rs})(\text{vec}(\hat{\Pi}_1 - \Pi_1))' f_{i,rs}^{(3)}(\text{vec}(\hat{\Pi}_1 - \Pi_1))\right\} \\
&= e'_i QH(\beta'_0 \otimes I_{g+1})\Omega^*H'QX'\Delta_{x,z} \\
&\quad + e'_i (QH\Omega^*(I_{g+1} \otimes \beta_0)H' + \text{tr}(QH(I_{g+1} \otimes \beta'_0)\Omega^*H').I_{g+k}) QX'\Delta_{xz} \\
&\quad + \text{tr}\{(I_{g+1} \otimes \beta'_0)\Omega^*H'QX' \text{Diag}(XQe_i)XQH\}.
\end{aligned}$$

The result is in terms of the $(g+1) \times (g+1)^2$ matrix Ω^* which itself is obtained by stacking the matrices Ω_{ijs} where the ij^{th} element of Ω_{ijs} is $\omega_{ijs} = E[v_{it}v_{jt}v_{st}]$. Ω^* may be expressed in terms of the structural parameters, see Phillips and Liu-Evans (2011) as follows:

$$\Omega^* = ((B')_{g+1}^{-1})' \Sigma'^* ((B')_{g+1}^{-1} \otimes (B')_{g+1}^{-1}) \quad (21)$$

where $(B')_{g+1}^{-1}$ is a $G \times (g+1)$ matrix containing the first $(g+1)$ columns of $(B')^{-1}$ and where Σ'^* is a $G \times G^2$ matrix given by $\Sigma'^* = [\Sigma_{ij1}, \dots, \Sigma_{ijG}]$.

Appendix 2

In this appendix we find expressions for the asymmetric terms in the context of the simple model in Section 2.

Note that $y_2 = Z\pi_2 + v_2$ from which we shall write $X = Z\pi_2$. We shall also require $\beta_0 = (-1, \beta'_1)'$, $e_i = 1$, $Q = (\pi'_2 Z' Z \pi_2)^{-1} = \frac{1}{\pi'_2 Z' Z \pi_2}$, $H = (0, 1)$, and

$$\Omega^{*'} = \begin{bmatrix} \omega_{111} & \omega_{211} & \omega_{112} & \omega_{212} \\ \omega_{121} & \omega_{221} & \omega_{122} & \omega_{222} \end{bmatrix}.$$

We now define the vector

$$\Delta_{x,z} = \begin{bmatrix} x'_1(X'X)^{-1}x_1 - z'_1(Z'Z)^{-1}z_1 \\ x'_2(X'X)^{-1}x_2 - z'_2(Z'Z)^{-1}z_2 \\ \dots \\ x'_T(X'X)^{-1}x_T - z'_T(Z'Z)^{-1}z_T \end{bmatrix} = \begin{bmatrix} \frac{z'_1(\pi_2\pi'_2)z_1}{\pi'_2Z'Z\pi_2} - z'_1(Z'Z)^{-1}z_1 \\ \frac{z'_2(\pi_2\pi'_2)z_2}{\pi'_2Z'Z\pi_2} - z'_2(Z'Z)^{-1}z_2 \\ \dots \\ \frac{z'_T(\pi_2\pi'_2)z_T}{\pi'_2Z'Z\pi_2} - z'_T(Z'Z)^{-1}z_T \end{bmatrix}$$

where $X' = \pi'_2(z_1, z_2, \dots, z_T)$, and $X'\Delta_{xz} = \frac{\sum(z'_j\pi_2)^3}{\pi'_2Z'Z\pi_2} - \sum \pi'_2z_jz'_j(Z'Z)^{-1}z_j$. All summations run from 1 to T .

The first of the asymmetric terms is

$$\begin{aligned} & e'_iQH(\beta'_0 \otimes I_{g+1})\Omega^*H'QX'\Delta_{x,z} \\ &= \frac{1}{\pi'_2Z'Z\pi_2}(0, 1) \begin{bmatrix} 1 & 0 & -\beta_1 & 0 \\ 0 & 1 & 0 & -\beta_1 \end{bmatrix} \begin{bmatrix} \omega_{111} & \omega_{121} \\ \omega_{211} & \omega_{221} \\ \omega_{112} & \omega_{122} \\ \omega_{212} & \omega_{222} \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ & \times \frac{1}{\pi'_2Z'Z\pi_2} \left[\frac{\sum(z'_j\pi_2)^3}{\pi'_2Z'Z\pi_2} - \sum \pi'_2z_jz'_j(Z'Z)^{-1}z_j \right] \\ &= (\omega_{221} - \beta_1\omega_{222}) \left[\frac{\sum(z'_j\pi_2)^3}{(\pi'_2Z'Z\pi_2)^3} - \frac{\sum \pi'_2z_jz'_j(Z'Z)^{-1}z_j}{(\pi'_2Z'Z\pi_2)^2} \right] \end{aligned} \quad (22)$$

The second asymmetric term is, using the fact that $\omega_{122} = \omega_{221}$,

$$\begin{aligned} & e'_i(QH\Omega^*(I_{g+1} \otimes \beta_0)H' + \text{tr}(QH(I_{g+1} \otimes \beta'_0)\Omega^*H').I_{g+k})QX'\Delta_{xz} \\ &= 2(\omega_{221} - \beta_1\omega_{222}) \left[\frac{\sum(z'_j\pi_2)^3}{\pi'_2Z'Z\pi_2} - \frac{\sum \pi'_2z_jz'_j(Z'Z)^{-1}z_j}{(\pi'_2Z'Z\pi_2)^2} \right] \end{aligned} \quad (23)$$

The third asymmetric term is

$$\begin{aligned}
& tr((I_{g+1} \otimes \beta'_0) \Omega^* H' Q X' Diag(X Q e_i) X Q H) \\
& = tr\{X Q H (I_{g+1} \otimes \beta'_0) \Omega^* H' Q X' Diag(X Q e_i)\} \\
& = tr\left\{ \frac{Z \pi_2}{\pi'_2 Z' Z \pi_2} (0, 1) \begin{bmatrix} 1 & -\beta_1 & 0 & 0 \\ 0 & 0 & 1 & -\beta_1 \end{bmatrix} \begin{bmatrix} \omega_{111} & \omega_{121} \\ \omega_{211} & \omega_{221} \\ \omega_{112} & \omega_{122} \\ \omega_{212} & \omega_{222} \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right. \\
& \quad \times \left. \frac{\pi'_2 Z'}{\pi'_2 Z' Z \pi_2} \begin{bmatrix} \frac{z'_1 \pi_2}{\pi'_2 Z' Z \pi_2} & 0 & 0 & 0 \\ 0 & \frac{z'_2 \pi_2}{\pi'_2 Z' Z \pi_2} & & 0 \\ \cdot & & \cdot & \\ 0 & & & \frac{z'_T \pi_2}{\pi'_2 Z' Z \pi_2} \end{bmatrix} \right\} \\
& = tr\left\{ \frac{1}{\pi'_2 Z' Z \pi_2} (\omega_{221} - \beta_1 \omega_{222}) \right. \\
& \quad \times \left. \frac{\pi'_2 Z'}{\pi'_2 Z' Z \pi_2} \begin{bmatrix} \frac{z'_1 \pi_2}{\pi'_2 Z' Z \pi_2} & 0 & 0 & 0 \\ 0 & \frac{z'_2 \pi_2}{\pi'_2 Z' Z \pi_2} & & 0 \\ \cdot & & \cdot & \\ 0 & & & \frac{z'_T \pi_2}{\pi'_2 Z' Z \pi_2} \end{bmatrix} \begin{bmatrix} z'_1 \pi_2 \\ z'_2 \pi_2 \\ \cdot \\ z'_T \pi_2 \end{bmatrix} \right\} \\
& = (\omega_{221} - \beta_1 \omega_{222}) \left[\frac{\sum (z'_j \pi_2)^3}{(\pi'_2 Z' Z \pi_2)^3} \right]. \tag{24}
\end{aligned}$$

Finally, summing (22)-(24), we find that the asymmetric effect

$$\begin{aligned}
& e'_i Q H (\beta'_0 \otimes I_{g+1}) \Omega^* H' Q X' \Delta_{x,z} + e'_i ((Q H \Omega^* (I_{g+1} \otimes \beta_0) H' \\
& + tr(Q H (I_{g+1} \otimes \beta'_0) \Omega^* H') \cdot I_{g+k}) Q X' \Delta_{xz}) \\
& + tr((I_{g+1} \otimes \beta'_0) \Omega^* H' Q X' Diag(X Q e_i) X Q H) + o(T^{-2})
\end{aligned}$$

is equal to

$$\begin{aligned}
& 3(\omega_{221} - \beta_1 \omega_{222}) \left[\frac{\sum (z'_j \pi_2)^3}{(\pi'_2 Z' Z \pi_2)^3} - \frac{\sum \pi'_2 z_j z'_j (Z' Z)^{-1} z_j}{(\pi'_2 Z' Z \pi_2)^2} \right] \\
& + (\omega_{221} - \beta_1 \omega_{222}) \left[\frac{\sum (z'_j \pi_2)^3}{(\pi'_2 Z' Z \pi_2)^3} \right] + o(T^{-2}) \tag{25}
\end{aligned}$$

for 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 $(\omega_{221} - \beta_1 \omega_{222})$ plays a key role.

It is helpful to interpret the disturbance skewness factors in terms of the structural parameters. Noting that

$$\begin{aligned}
\omega_{221} - \beta_1\omega_{222} &= E(v_{1,t}v_{2,t}^2) - \beta_1E(v_{2,t}^3) \\
&= E\left(\frac{\varepsilon_{1,t} + \beta_1\varepsilon_{2,t}}{1 - \beta_1\beta_2}\right)\left(\frac{\beta_2\varepsilon_{1,t} + \varepsilon_{2,t}}{1 - \beta_1\beta_2}\right)^2 \\
&\quad - \beta_1E\left(\frac{\beta_2\varepsilon_{1,t} + \varepsilon_{2,t}}{1 - \beta_1\beta_2}\right)^3
\end{aligned} \tag{26}$$

which with some manipulation simplifies to

$$\omega_{221} - \beta_1\omega_{222} = \frac{\sigma_{111}\beta_2^2 + 2\sigma_{112}\beta_2 + \sigma_{122}}{(1 - \beta_1\beta_2)^2} \tag{27}$$

it is clear that this term can be made large for suitable choice of the parameters, especially since β_1 and β_2 are unrestricted other than the requirement that $\beta_1\beta_2 \neq 1$.

Consider now the part not involving $\omega_{221} - \beta_1\omega_{222}$. If we put $\pi'_2 = \frac{\gamma'}{1 - \beta_1\beta_2}$ then

$$4 \frac{\sum (z'_j\pi_2)^3}{(\pi'_2 Z'Z\pi_2)^3} - 3 \frac{\sum \pi'_2 z_j z'_j (Z'Z)^{-1} z_j}{(\pi'_2 Z'Z\pi_2)^2} = (1 - \beta_1\beta_2)^3 \left[4 \frac{\sum (z'_j\gamma)^3}{(\gamma'Z'Z\gamma)^3} - 3 \frac{\sum \gamma' z_j z'_j (Z'Z)^{-1} z_j}{(\gamma'Z'Z\gamma)^2} \right] \tag{28}$$

and the expression in (25) is

$$\begin{aligned}
&\{(1 - \beta_1\beta_2)(\sigma_{111}\beta_2^2 + 2\sigma_{112}\beta_2 + \sigma_{122})\} \\
&\times \left[4 \frac{\sum (\gamma'z_j)^3}{(\sum (\gamma'z_j)^2)^3} - 3 \frac{\sum (\gamma'z_j)z'_j (Z'Z)^{-1} z_j}{(\sum (\gamma'z_j)^2)^2} \right]
\end{aligned} \tag{29}$$

The terms involving the $z_j, j = 1, \dots, T$, are $O(T^{-2})$ and so are likely to become small quite rapidly as T gets large. However the expression in the numerator $(1 - \beta_1\beta_2)(\sigma_{111}\beta_2^2 + 2\sigma_{112}\beta_2 + \sigma_{122})$ can be made large through suitable choice of coefficients β_1 and β_2 and the third moment parameters; hence there will be structures where the non-symmetry effect on the bias in estimating β_2 will be non-trivial despite the fact of being $O(T^{-2})$.

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

Appendix 3 (Asymmetry and K)

Approximate lower bound

The first asymmetry term is

$$e'_i QH(\beta'_0 \otimes I_{g+1})\Omega^* H Q X' \Delta_{xz} \tag{30}$$

where Δ_{xz} has p -th element $x'_p(X'X)^{-1}x_p - z'_p(Z'Z)^{-1}z_p$. This can be written as

$$c' \Delta_{xz} = c'(c_x - c_z) \quad (31)$$

where $c = \{e'_i QH(\beta'_0 \otimes I_{g+1})\Omega^* H Q X'\}'$, c_x has p -th element $x'_p(X'X)^{-1}x_p$ and c_z has p -th element $z'_p(Z'Z)^{-1}z_p$. The absolute value can be written as

$$|c'(c_x - c_z)| = \left| \|c\|_2 \|c_x - c_z\|_2 \tilde{c} \right| \quad (32)$$

where $\|\cdot\|_2$ is the Euclidean norm. The term \tilde{c} is the cosine of the angle between the vectors $\{e'_i QH(\beta'_0 \otimes I_{g+1})\Omega^* H Q X'\}'$ and Δ_{xz} , which does not go to zero with K . Similarly, $\|c\|_2$ does not generally decrease with K . The following lower bound on the term $\|c_x - c_z\|_2$ is increasing in K though:

$$\|c_x - c_z\|_2 \geq \frac{\|c_x - c_z\|_1}{\sqrt{T}} \quad (33)$$

$$\geq \frac{\|c_x\|_1 - \|c_z\|_1}{\sqrt{T}} \quad (34)$$

$$= \frac{\text{trace}(P_x) - \text{trace}(P_z)}{\sqrt{T}} \quad (35)$$

$$= \frac{K - g - k}{\sqrt{T}} \quad (36)$$

So that

$$|c'(c_x - c_z)| \geq \frac{K - g - k}{\sqrt{T}} \|c\|_2 |\tilde{c}| \quad (37)$$

with $\|c\|_2$ and $|\tilde{c}|$ not generally decreasing with K .