Cardiff Economics Working Papers

Working Paper No. E2016/10

Almost Unbiased Variance Estimation in Simultaneous Equation Models

Garry DA Phillips and Yongdeng Xu

September 2016

ISSN 1749-6010

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

Almost Unbiased Variance Estimation in Simultaneous Equation Models

Garry DA Phillips (Cardiff University and University of Alicante)*
Yongdeng Xu (Cardiff University)†

September 2016

Abstract

While a good deal of research in simultaneous equation models has been conducted to examine the small sample properties of coefficient estimators there has not been a corresponding interest in the properties of estimators for the associated variances. In this paper we build on Kiviet and Phillips (2000) and explore the biases in variance estimators. This is done for the 2SLS and the MLIML estimators.The approximations to the bias are then used to develop less biased estimators whose properties are examined and compared in a number of simulation experiments. In addition, a bootstrap estimator is included which is found to perform especially well. The experiments also consider coverage probabilities/test sizes and test powers of the t-tests where it is shown that tests based on 2SLS are generally oversized while test sizes based on MLIML are closer to nominal levels. In both cases test statistics based on the corrected variance estimates generally have a higher power than standard procedures.

Keywords: Simultaneous equation models, 2SLS and Fuller's estimators, Bias corrected variance estimation, Inference and bias corrected variance

JEL Classification: C12; C13; C26; C30

*Email: phillipsgd1@cardiff.ac.uk; Cardiff Business School, Cardiff University, CF10 3EU, UK.

†Corresponding author. Email: xuy16@cf.ac.uk; Cardiff Business School, Cardiff University, CF10 3EU, UK

1 Introduction

In the static simultaneous equation model a considerable amount of research has been conducted to examine the small sample properties of coefficient estimators. Of particular relevance is the seminal paper by Nagar(1959) in which approximations were given for the bias of the 2SLS estimator to the order of $T^{-1}$and for its mean squared error to order T^{-2}, where T is the sample size. The analysis did not focus entirely on 2SLS however, but also covered the consistent $k-$class of estimators where k is fixed. When the square of the bias approximation is subtracted from the approximation for the MSE, we have an approximation for the variance. By comparing this approximate variance with an approximation to order T^{-2} for the expectation of the asymptotic variance estimator which is used in practice, we may find an approximation for the bias of the asymptotic variance estimator. This was explored by Kiviet and Phillips(2000) who were able to deduce that the bias, which is of order T^{-2}, is in general upwards in the 2SLS case. However no attempt was then made to explore the magnitude of the bias nor a method for bias correction although it is apparent that once a bias approximation is available a means for correcting for the bias is available too.

In this paper we examine bias correction for the variance estimator by first considering the traditional approach whereby the estimated approximate bias is subtracted from the estimator; then we examine its performance in a set on Monte Carlo experiments. We also include results for the bootstrap technique which provides a general method of estimating the variance.

In Kiviet and Phillips (2000) the focus of attention was the 2SLS estimator which, while in very common use, has certain drawbacks related to the non-existence of moments in some situations. In this paper we widen the investigation to include the Modified Limited Information Maximum Likelihood(MLIML) estimator proposed by Fuller (1977) which has attracted considerable attention in recent years and which does not suffer from the non-existence of moments problem. Our aim is to compare the properties of the variance estimators for 2SLS and the Fuller MLIML. It is found that these estimators may be badly biased and so bias corrected versions are developed. An alternative variance estimator based upon the bootstrap is also included in the comparison. Monte Carlo experiments are used explore the properties of these latter estimators in the context of standard inference procedures.

The structure of the paper is as follows. Section 2 presents the model and it gives a summary of large T approximations. It also discusses a problem with 2SLS variance estimation. In section 3 we derive approximations to the expectation of the asymptotic variance for 2SLS and LIML estimator. In section 4 we discuss the bias of these asymptotic variance estimators and develop the bias corrected variance estimators. Section 5, we include a bootstrap variance estimation method. In section 6, we present the results of simulation experiments that indicate the usefulness of the proposed procedures in a numerous cases. Finally, Section 7 concludes. The proofs are collected in the Appendix.

2 The Simultaneous Equation Model

The model we shall analyze is the classical static simultaneous equation model containing G equations given by

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

in which y_t is a $G \times 1$ vector of endogenous variables, x_t is a $K \times 1$ vector of strongly exogenous variables and u_t is a $G \times 1$ vector of 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 corresponding reduced form equations are

$$y_t = -B^{-1}\Gamma x_t + B^{-1}u_t = \Pi x_t + v_t, \tag{2}$$

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

$$YB' + X\Gamma' = U \tag{3}$$

Here, Y is a $T \times G$ matrix of observations on endogenous variables, X 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 will be written as

$$y_1 = Y_2\beta + X_1\gamma + u_1 \tag{4}$$

where y_1 and Y_2 are, respectively, a $T \times 1$ vector and a $T \times g_1$ matrix of observations on $g_1 + 1$ endogenous variables. X_1 is a $T \times r_1$ matrix of observations on r_1 exogenous variables, β and γ are, respectively, $g_1 \times 1$ and $r_1 \times 1$ vectors of unknown parameters and u_1 is a $T \times 1$ vector of normally distributed disturbances with covariance matrix $E(u_1 u_1') = \sigma_{11} I_T$.

The reduced form of the system includes

$$Y_1 = X\Pi_1 + V_1, \tag{5}$$

in which $Y_1 = (y_1 : Y_2)$, $X = (X_1 : X_2)$ is a $T \times K$ matrix of observations on K exogenous variables with an associated $K \times (g_1 + 1)$ matrix of reduced form parameters given by $\Pi_1 = (\pi_1 : \Pi_2)$, while $V_1 = (v_1 : V_2)$ is a $T \times (g_1 + 1)$ matrix of normally distributed reduced form disturbances. The transpose of each row of V_1 is independently and normally distributed with a zero mean vector and $(g_1 + 1) \times (g_1 + 1)$ positive definite matrix $\Omega_1 = (\omega_{ij})$. We also make the following assumption:

Assumption 1. (i): The $T \times K$ matrix X is strongly exogenous and of rank K with limit matrix $\lim_{T\to\infty} T^{-1}X'X = \Sigma_{XX}$, which is $K \times K$ positive definite, and that (ii): Equation (4) is over-identified so that $K > g_1 + k_1$, i.e. the number of excluded variables exceeds the number required for the equation to be just identified. In cases where second moments are analyzed we shall assume that K exceeds $g_1 + k_1$ by at least two. These over-identifying restrictions are sufficient to ensure that the Nagar expansion is valid in the case considered by Nagar and that the first two estimator moments for 2SLS exist: see Sargan (1974).

2.1 Large T-approximations for the Moments of k-class Estimators

The k-class estimator was introduced by Nagar (1959) and in the context of (4) it is given by

$$\begin{pmatrix} \hat{\beta}_k \\ \hat{\gamma}_k \end{pmatrix} = \begin{pmatrix} Y_2^{'}Y_2 - k\hat{V}_2\hat{V}_2 & Y_2^{'}X_1 \\ X_1^{'}Y_2 & X_1^{'}X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2^{'}y_1 - k\hat{V}_2^{'}y_1 \\ X_1^{'}y_1 \end{pmatrix}. \quad (6)$$

When $k = 1$ we have the 2SLS estimator while the Limited Information Maximum Likelihood (LIML) estimator is obtained when $k = \lambda \geq 1$, and λ is the smallest root of the determinantal equation

$$\left| Y_1^{'}(I - P_{X_1})Y_1 - \lambda Y_1^{'}(I - P_X)Y_1 \right| = 0.$$

Note that λ is stochastic and under the assumptions employed here, $T\lambda$ is asymptotically distributed as $\chi^2_{k_2 - g_1}$, see Fuller(1977). The LIML estimator has the drawback that it does not have finite moments of any order. To overcome this problem Fuller (1977) presented a Modified Limited Information Maximum Likelihood Estimator (MLIML) where λ is replaced by $\lambda - \frac{\alpha}{T-K}$ and α is a chosen positive integer, which has (at least) finite first and second moments . Hence the MLIML estimator is

$$\begin{pmatrix} \hat{\beta}_F \\ \hat{\gamma}_F \end{pmatrix} = \begin{pmatrix} Y_2^{'}Y_2 - (\lambda - \frac{\alpha}{T-K})\hat{V}_2\hat{V}_2 & Y_2^{'}X_1 \\ X_1^{'}Y_2 & X_1^{'}X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2^{'}y_1 - (\lambda - \frac{\alpha}{T-K})\hat{V}_2^{'}y_1 \\ X_1^{'}y_1 \end{pmatrix}. \quad (7)$$

When $\alpha = 1$ the estimator has small bias whereas when $\alpha = 4$ the estimator has smallest MSE but its bias is typically larger than when $\alpha = 1$. A number of recent studies have found that MLIML may have good finite sample properties, see for example, Hahn, Hausman and Kuersteiner (2004).

We shall later consider a set of Monte Carlo experiments in which the small sample properties of both of these estimators are explored.

We shall find it convenient to rewrite (4) as

$$y_1 = Z_1\alpha + u_1 \quad (8)$$

where $Z_1 = (Y_2 : X_1)$ and $\alpha = (\beta^{'}, \gamma^{'})^{'}$.In this context the $k-$class estimator will be written as $\hat{\alpha}_k$.

In his seminal paper, Nagar (1959) presented approximations for the first and second moments of the $k-$class of estimators where $k = 1 + \theta/T$ and θ is non-stochastic and may be any real number. Notice that $(1 - k)$ is of order T^{-1}. The main results are given by the following:

1. If we denote $\hat{\alpha}_k$ as the $k-$class estimator for α in (6) then, defining L as the degree of overidentification, the approximate bias is given by

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

2. The MSE of $\hat{\alpha}_k$ is given by

$$(E(\hat{\alpha}_k - \alpha)(\hat{\alpha}_k - \alpha)^{'}) = \sigma^2 Q[I + A^*] + o(T^{-2}), \tag{10}$$

where

$$\begin{aligned} A^* &= [(2\theta - (2L-3))tr(C_1Q) + tr(C_2Q)]I + \{(\theta - L + 2)^2 \\ &\quad +2(\theta + 1)\}C_1Q + (2\theta - L + 2)C_2Q. \end{aligned} \tag{11}$$

To interpret the above approximations we define the degree of overidentification as

$$L = k_2 - g_1, \tag{12}$$

where $k_2 = K - k_1$ is the number of exogenous variables excluded from the equation of interest.

Noting that $Y_2 = \bar{Y}_2 + V_2$ where $\bar{Y}_2 = X\Pi_2$,we define

$$Q = \left[\begin{array}{cc} \bar{Y}_2'\bar{Y}_2 & \bar{Y}_2'X_1 \\ X_1'\bar{Y}_2 & X_1'X_1 \end{array} \right]^{-1}. \tag{13}$$

Further, we may write that $V_2 = W + u_1\pi'$ where u_1 and W are independent and

$$\frac{1}{T}\left(\begin{array}{c} E(V_2'u_1) \\ 0 \end{array} \right) = \sigma^2 \left(\begin{array}{c} \pi \\ 0 \end{array} \right) = q. \tag{14}$$

Moreover, defining $V_z = [V_2 : 0]$ we have

$$C = E[\frac{1}{T}V_z'V_z] = \left[\begin{array}{cc} (1/T)E(V_2'V_2) & 0 \\ 0 & 0 \end{array} \right] = C_1 + C_2, \tag{15}$$

where $C_1 = \left[\begin{array}{cc} \sigma^2\pi\pi' & 0 \\ 0 & 0 \end{array} \right]$ and $C_2 = \left[\begin{array}{cc} 1/TE(W'W) & 0 \\ 0 & 0 \end{array} \right]$.

The approximations for the 2SLS estimator are found by setting $\theta = 0$ in the first expression above so that, for example, the 2SLS bias approximation is given by

$$E(\hat{\alpha} - \alpha) = (L - 1)Qq + o(1/T). \tag{16}$$

while the second moment approximation is

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

where

$$A^* = [-(2L-3)tr(C_1Q) + tr(C_2Q)]I + \{(-L+2)^2 + 2\}C_1Q + (-L+2)C_2Q. \tag{17}$$

The 2SLS bias approximation above was extended by Mikhail (1972) to

$$E(\hat{\alpha} - \alpha) = (L-1)[I + tr(QC)I - (L-2)QC]Qq \ + o(1/T^2). \tag{18}$$

Notice that this bias approximation contains the term $(L - 1)Qq$ which, as we have seen, is the approximation to order $1/T$ whereas the remaining term, $(L-1)[tr(QC)I-(L-2)QC]Qq$, is of order T^{-2}. This higher order approximation is of considerable importance for this paper. Note that the T^{-2} term includes a component $-(L-1)(L-2)QCQq$ which may be relatively large when L is large, a fact that will be commented on again later. It is also of particular interest that the approximate bias is zero to order T^{-2} when $L=1$, *i.e*, when $K-(g_1+k_1)=k_2-g_1=1$.

2.2 A Problem With 2SLS Variance Estimation.

There are, in effect, two stages to 2SLS variance estimation, the first of which requires the estimation of the disturbance variance σ^2. It turns out that the properties of the usual estimator of σ^2, which is obtained by dividing the sum of squared residuals by the sample size, is itself affected by non-existence of moments of the coefficient estimators. This is not a feature of regression models. For example, consider the classical linear regression model

$$Y = X\beta + \varepsilon \tag{19}$$

where $E(\varepsilon)=0, E(\varepsilon\varepsilon')=\sigma^2 I_T$ and X is $T\times k$ and is non-stochastic of rank k.

As is well known the least squares estimator, $\hat{\beta}=(X'X)^{-1}X'y$ is best linear unbiased (BLUE) with covariance matrix given by

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

The estimator for σ^2 is given by

$$\hat{\sigma}^2 = \sum_{t=1}^{T} \hat{e}_t^2/(T-k)$$

where $\hat{e}_t = y_t - x_t'\hat{\beta} = \varepsilon_t - x_t'(\hat{\beta}-\beta)$ and

$$\sum_{t=1}^{T} \hat{e}_t^2/(T-k) = \frac{1}{T-k}\{\sum_{t=1}^{T}\varepsilon_t^2 - 2\sum_{t=1}^{T}(\hat{\beta}-\beta)'x_t\varepsilon + (\hat{\beta}-\beta)'\sum_{t=1}^{T}x_t x'(\hat{\beta}-\beta)\}.$$

For $E(\hat{\sigma}^2)<\infty$ to hold, we require only that $E(\varepsilon_t^2)<\infty$ for all t, which is an assumption of the classical model. This also ensures that the second moment of $\hat{\beta}$ exists. However, if the fourth moment of ε_t does not exist, then the second moment of $\hat{\sigma}^2$ and the fourth moment of $\hat{\beta}$ will not exist either. Of course, if the ε_t are normally distributed then all moments exist for both $\hat{\sigma}^2$ and $\hat{\beta}$ but if this is not assumed it should be borne in mind that the existence of a second moment for $\hat{\sigma}^2$ is a prerequisite for valid finite sample inference since, otherwise, the variance estimator may be subject to extreme outliers.

A different situation occurs in the simultaneous equation model in section 2. Consider the equation to be estimated:

$$y_1 = Z_1\alpha + u_1$$

where $Z_1 = (Y_2 : X_1)$ and let the estimated version be

$$y_1 = Z_1\hat{\alpha} + \hat{u}_1 \tag{20}$$

so that the sum of squared residuals is

$$\hat{u}_1'\hat{u}_1 = u_1'u_1 - 2(\hat{\alpha} - \alpha)'Z_1'u_1 + (\hat{\alpha} - \alpha)'Z_1'Z_1(\hat{\alpha} - \alpha),$$

while the estimator for σ^2 may be represented as

$$\hat{\sigma}^2 = \frac{u_1'u_1 - 2(\hat{\alpha} - \alpha)'Z_1'u_1 + (\hat{\alpha} - \alpha)'Z_1'Z_1(\hat{\alpha} - \alpha)}{T - g_1 - k_1}. \tag{21}$$

If the estimator employed is 2SLS then it is well known that it has moments up to the order of overidentification of the associated parameters. Thus we see from (21) that even when components of u_t have a finite second moment this does not ensure that even the first moment of $\hat{\sigma}^2$ exists. It is necessary to also consider the existence of moments for $\hat{\alpha}$. In fact, for the existence of the first moment of $\hat{\sigma}^2$ it is required that the second moment of $\hat{\alpha}$ should exist and this, in turn, requires that the order of overidentification of the estimated parameters be at least two,i,e $L \geq 2$. Similarly, for $\hat{\sigma}^2$ to have a second moment we require that, in addition to the existence of the fourth moment of components of u_1,the fourth moment of $\hat{\alpha}$ should be finite also. This requires that $L \geq 4$.Thus even when normality for the disturbances is assumed, so that moments of all orders for the disturbances exist, finite sample inference based on an estimated variance may be problematic unless $L \geq 4$,see Murray D.Smith (1994) for a full analysis of structural disturbance estimators.This is obviously a serious restriction on the use of 2SLS in practice.

Of course the problem is avoided if we estimate the parameters using an estimator which has all necessary moments and, fortunately, such estimators are available. This applies in the case of the Fuller MLIML estimator in (7) and to any member of the consistent $k-$class for which $0 < k < 1$, see Kinal (1983).

3 Expectations of Asymptotic Variances

In this section we derive approximations to the expectation of the asymptotic variances for the MLIML estimator which is required for the variance bias approximation. In Kiviet and Phillips (2000) an approximation to order T^{-2} was given for the 2SLS case as was noted in section 1. We shall exploit these results to find the required expectations for the MLIML estimator but first we consider how results were obtained for 2SLS.

3.1 The 2SLS Asymptotic variance

We shall write the 2SLS asymptotic variance estimate as

$$\hat{V}ar(\hat{\alpha}) = \hat{\sigma}^2(\hat{Z}_1\hat{Z}_1)^{-1} \tag{22}$$

where

$$\hat{\sigma}^2 = \frac{\hat{u}_1'\hat{u}_1}{T-(g_1+k_1)} = \frac{u_1'u_1 - 2(\hat{\alpha}-\alpha)'Z_1'u_1 + (\hat{\alpha}-\alpha)'Z_1'Z_1(\hat{\alpha}-\alpha)'}{T-(g_1+k_1)} \qquad (23)$$

in which successive terms are $O_p(1), O_p(T^{-\frac{1}{2}})$ and $O_p(T^{-1})$.

The 2SLS estimation error may be expanded (see Nagar 1959) as

$$\hat{\alpha} - \alpha = Q\bar{Z}_1'u_1 + QV_z'M^*u_1 - Q[\bar{Z}_1'V_z + V_z'\bar{Z}]Q\bar{Z}_1'u_1 + o_p(T^{-1}). \qquad (24)$$

while the inverse matrix $(\hat{Z}_1\hat{Z}_1)^{-1}$ is similarly expanded to obtain:

$$\begin{aligned}(\hat{Z}_1'\hat{Z}_1)^{-1} &= [Q - Q(\bar{Z}'V_z + V_Z'\bar{Z})Q - QV_Z'MV_ZQ \\ &\quad +Q(\bar{Z}'V_Z + V_Z'\bar{Z})Q(\bar{Z}'V_z + V_Z'\bar{Z})Q] + o_p(T^{-2}) \end{aligned} \qquad (25)$$

where successive terms are $O(T^{-1}), O_pT^{\frac{-3}{2}})$ and $O_p(T^{-2})$.

To find an appropriate expansion for $\hat{\sigma}^2(\hat{Z}_1\hat{Z}_1)^{-1}$ we combine the expressions in (23) and (25) to yield

$$\begin{aligned}\hat{\sigma}^2(\hat{Z}_1\hat{Z}_1)^{-1} &= [[\frac{u_1'u_1}{T-(g+k)}[Q - Q(\bar{Z}'V_z + V_Z'\bar{Z})Q - QV_Z'MV_ZQ \\ &\quad +Q(\bar{Z}'V_Z + V_Z'\bar{Z})Q(\bar{Z}'V_z + V_Z'\bar{Z})Q] - \frac{2(\hat{\alpha}-\alpha)'Z_1'u_1}{T-(g+k)}] \\ &\quad \times[Q - Q(\bar{Z}'V_z + V_Z'\bar{Z})Q] + \frac{(\hat{\alpha}-\alpha)'Z_1'Z_1(\hat{\alpha}-\alpha)}{T-(g+k)}Q \\ &\quad +o_p(T^{-2}).\end{aligned} \qquad (26)$$

The expectation of this to order T^{-2} is given in Kiviet and Phillips (2000) as

$$\begin{aligned}E[\hat{\sigma}^2(\hat{Z}_1\hat{Z}_1)^{-1}] &= \sigma^2Q + \sigma^2[-(L-2)QCQ + 4QC_1Q \\ &\quad -2(L-1)tr(QC_1)Q + 2tr(QC)Q] + o(T^{-2}).\end{aligned} \qquad (27)$$

We shall make use of this in evaluating the counterpart results for MLIML.

3.2 The MLIML Asymptotic Variance

The estimate for the asymptotic variance of MLIML is given by

$$\hat{V}ar(\hat{\alpha}_F) = \hat{\sigma}_F^2 \begin{pmatrix} Y_2^{'}Y_2 - (\lambda - \frac{1}{T-K})\hat{V}_2\hat{V}_2 & Y_2^{'}X_1 \\ X_1^{'}Y_2 & X_1^{'}X_1 \end{pmatrix}^{-1}$$

which we shall write as

$$\hat{V}ar(\hat{\alpha}_F) = \hat{\sigma}_F^2(\hat{Z}_1'\hat{Z}_1)_F^{-1} \qquad (28)$$

where $\hat{\sigma}_F^2$ is obtained from the residuals following estimation by MLIML.

To find the counterpart expression to the 2SLS for MLIML, we first note that the relevant estimator for σ_F^2 is given by

$$\hat{\sigma}_F^2 = \frac{\hat{u}_{1F}'\hat{u}_{1F}}{T-(g_1+k_1)} = \frac{u_1'u_1 - 2(\hat{\alpha}_F-\alpha)'Z_1'u_1 + (\hat{\alpha}_F-\alpha)'Z_1'Z_1(\hat{\alpha}_F-\alpha)}{T-(g_1+k_1)} \tag{29}$$

where $\hat{\alpha}_F$ is the MLIML estimate of α.

Noting that $(\hat{\alpha}_F-\alpha)'Z_1'Z_1(\hat{\alpha}_F-\alpha) = (\hat{\alpha}-\alpha)'Z_1'Z_1(\hat{\alpha}-\alpha) + o_p(1)$, see Nagar(1961), we see that, using (23) and (29), we may write

$$\hat{\sigma}_F^2 - \hat{\sigma}^2 = -\frac{2(\hat{\alpha}_F-\hat{\alpha})'Z_1'u_1}{T-(g_1+k_1)} + o_p(T^{-1}). \tag{30}$$

Also, the asymptotic expansion for the estimation error of MLIML is given in Appendix A (A.3) by

$$\begin{aligned}\hat{\alpha}_F - \alpha &= Q\bar{Z}_1'u_1 + (1-\lambda+\frac{1}{T-K})QV_Z'u_1 + QV_Z'M^*u_1 \\ &\quad -QV_Z'\bar{Z}_1'QZu_1 - Q\bar{Z}_1'V_ZQ\bar{Z}_1'u_1 + o_p(T^{-1})\end{aligned} \tag{31}$$

where we have used the result that $\lambda QV_Z'M^*u_1 = QV_Z'M^*u_1 + o_p(T^{-1})$.

From (24) and (31) we find that

$$\hat{\alpha}_F - \hat{\alpha} = (1-\lambda+\frac{1}{T-K})QV_Z'u_1 + o_p(T^{-1}). \tag{32}$$

Substituting from (32) into (30) we have

$$\begin{aligned}\hat{\sigma}_F^2 - \hat{\sigma}^2 &= -\frac{2(\hat{\alpha}_F-\hat{\alpha})'Z_1'u_1}{T-(g_1+k_1)} + o_p(T^{-1}) \\ &= -\frac{2(1-\lambda+\frac{1}{T-K})u_1'V_ZQZ_1'u_1}{T-g_1-k_1} + o_p(T^{-1}).\end{aligned} \tag{33}$$

In what follows we shall need the following asymptotic expansions:

$$\begin{aligned}(\hat{Z}_1'\hat{Z}_1)_F^{-1} &= [Q - Q(\bar{Z}'V_z + V_Z'\bar{Z})Q - Q((1-\lambda)V_Z'V_Z \\ &\quad +\frac{1}{T-K}V_Z'V_Z)Q - Q\lambda V_Z'MV_ZQ \\ &\quad +Q(\bar{Z}'V_Z + V_Z'\bar{Z})Q(\bar{Z}'V_z + V_Z'\bar{Z})Q] \\ &\quad +o_p(T^{-2})\end{aligned} \tag{34}$$

where this expansion is given in Appendix A (A.2).

From (34) and (25), we have

$$(\hat{Z}_1'\hat{Z}_1)_F^{-1} - (\hat{Z}_1'\hat{Z}_1)^{-1} = -Q((1-\lambda)V_Z'V_Z + \frac{1}{T-K}V_Z'V_Z)Q + o_pT^{-2}. \tag{35}$$

We wish to find an approximation for the expected value of the asymptotic variance estimate

$$\begin{aligned}
\hat{\sigma}_F^2(\hat{Z}_1'\hat{Z}_1)_F^{-1} &= ((\hat{\sigma}_f^2 - \hat{\sigma}^2) + \hat{\sigma}^2)(\hat{Z}_1'\hat{Z}_1)^{-1} + ((\hat{Z}_1'\hat{Z}_1)_F^{-1} - (\hat{Z}_1'\hat{Z}_1)^{-1})) \\
&= \hat{\sigma}^2(\hat{Z}_1'\hat{Z}_1)^{-1} + (\hat{\sigma}_f^2 - \hat{\sigma}^2)(\hat{Z}_1'\hat{Z}_1)^{-1} \\
&\quad + \hat{\sigma}^2((\hat{Z}_1'\hat{Z}_1)_F^{-1} - (\hat{Z}_1'\hat{Z}_1)^{-1})) + o_p(T^{-2}) \\
&= \hat{\sigma}^2(\hat{Z}_1'\hat{Z}_1)^{-1} - \frac{2(1-\lambda+\frac{1}{T-K})u_1'V_Z Q Z_1' u_1}{T-g-k}(\hat{Z}_1'\hat{Z}_1)^{-1} \\
&\quad -\hat{\sigma}^2(Q((1-\lambda)V_Z'V_Z + \frac{1}{T-K}V_Z'V_Z)Q \\
&\quad + o_p(T^{-2}) \qquad (36)
\end{aligned}$$

from which it is seen that the asymptotic variance estimate for MLIML to $O_p(T^{-2})$ equals the asymptotic variance estimate for 2SLS minus two additional terms. Hence to get the required approximation to the expected value of $\hat{\sigma}_f^2(\hat{Z}_1'\hat{Z}_1)_F^{-1}$ we shall need to find expectations for these additional terms only. Thus we need

$$\begin{aligned}
&E[(-\frac{2(1-\lambda+\frac{1}{T-K})u_1'V_Z Q Z_1' u_1}{T-g-k})Q] \\
&-E[(\frac{u_1'u_1}{T-(g+k)})(Q((1-\lambda)V_Z'V_Z + \frac{1}{T-K}V_Z'V_Z)Q)] \qquad (37)
\end{aligned}$$

where terms of smaller order than T^{-2} have been ignored.

It is shown in the Appendix B (B.10), that the first part of the above is

$$E[(-\frac{2(1-\lambda+\frac{1}{T-K})u_1'V_Z Q Z_1' u_1}{T-g-k})Q] = 2\sigma^2(L-1)tr(QC_1)Q + o(T^{-2}). \qquad (38)$$

It is also shown in the Appendix B (B.17), that the second part of the above is:

$$\begin{aligned}
&-E[(\frac{u_1'u_1}{T-(g+k)})(Q((1-\lambda)V_Z'V_Z + \frac{1}{T-K}V_Z'V_Z)Q] \\
&= -E[\sigma^2(Q((1-\lambda)V_Z'V_Z + \frac{1}{T-K}V_Z'V_Z)Q] + o(T^{-2}) \\
&= \sigma^2(L-1)QCQ + o(T^{-2}) \qquad (39)
\end{aligned}$$

where we have used the fact that

$$\frac{u_1'u_1}{T-(g+k)} = \sigma^2 + o_p(1).$$

Finally, gathering terms, we have that

$$\begin{aligned}
E(\hat{\sigma}_F^2(\hat{Z}_1'\hat{Z}_1)_F^{-1}) &= E(\hat{\sigma}^2(\hat{Z}_1'\hat{Z}_1)^{-1}) + 2\sigma^2(L-1)tr(QC_1)Q \\
&\quad + \sigma^2(L-1)QCQ + o(T^{-2}) \\
&= \sigma^2 Q + \sigma^2[QCQ + 4QC_1Q + 2tr(QC)Q] \\
&\quad + o(T^{-2}). \qquad (40)
\end{aligned}$$

We shall use the results in (27) and (40) in the next section when we derive approximately unbiased estimates for the variances of 2SLS and MLIML .

4 The Bias of the Asymptotic Variance Estimators

From the results in Kiviet and Phillips (2000), we may deduce an approximation to the unknown variance of 2SLS as follows:

$$\begin{aligned} Var(\hat{\alpha}) &= \sigma^2 Q + \sigma^2[tr(QC)Q - 2(L-1)tr(QC_1)Q \\ &\quad -(L-3)QC_1Q - (L-2)QCQ] + o(T^{-2}). \end{aligned} \tag{41}$$

In Kiviet and Phillips (2000), it was also shown in (27) that, the expected value of the asymptotic variance estimator can be approximated to order T^{-2} by:

$$\begin{aligned} E[\hat{V}ar(\hat{\alpha})] &= \sigma^2 Q + \sigma^2[-(L-2)QCQ + 4QC_1Q \\ &\quad -2(L-1)tr(QC_1)Q + 2tr(QC)Q] + o(T^{-2}). \end{aligned} \tag{42}$$

It then follows that an approximation to order T^{-2} for the bias of the asymptotic variance estimator is given by

$$E[\hat{V}ar(\hat{\alpha})] - Var(\hat{\alpha}) = \sigma^2[tr(QC)Q + (L+1)QC_1Q] + o(T^{-2}). \tag{43}$$

Noting that the terms on the right hand side of the above are positive semi-definite, Kiviet and Phillips(2000) deduced that the bias to the order of the approximation is, in general, upwards.

An approach to finding a bias corrected estimate of the variance is immediate. We simply deduct an estimate of the bias in (43) from $\hat{V}ar(\hat{\alpha})$.Thus we have the result:

Theorem 1 *In the model of section 2, an unbiased estimate of the variance of the 2SLS estimator $\hat{\alpha}$ to order T^{-2},is given by*

$$\hat{V}ar(\hat{\alpha})_{BC} = \hat{V}ar(\hat{\alpha}) - \hat{\sigma}^2[tr(\hat{Q}\hat{C})\hat{Q} + (L+1)\hat{Q}\hat{C}_1\hat{Q}] \tag{44}$$

where $\hat{\sigma}^2, \hat{Q}$ and $\hat{C}$ are consistent estimates of σ^2, Q and C respectively.

A similar approach can be adopted in the case MLIML. For MLIML it has been shown by Anderson,Kunitomo and Morimune (1986) that an approximation for the MSE is

$$E(e_F e_F') = \sigma^2 Q + \sigma^2([trQC]\,Q - QC_1Q + LQC_2Q) + o(T^{-2}) \tag{45}$$

and since the estimator is unbiased to order T^{-1}, this is also the approximation to the variance. Hence we shall write:

$$\begin{aligned} Var(\hat{\alpha}_F) &= \sigma^2 Q + \sigma^2([trQC]\,Q - QC_1Q + LQC_2Q) \\ &\quad + o(T^{-2}). \end{aligned} \tag{46}$$

It is of interest to compare this approximation to the variance approximation for the 2SLS estimator given in (41). Subtracting (46) from (41) yields:

$$
\begin{aligned}
Var(\hat{\alpha}_F) - Var(\hat{\alpha}) &= 2(L-3)QC_1Q + 2(L-1)QC_2Q + 2(L-1)(trQC_1)Q \\
&\quad + o(T^{-2}).
\end{aligned} \tag{47}
$$

Upon noting that $(trQC_1)Q \geq QC_1Q$, it follows that the above is positive semidefinite for $L \geq 2$ so that the MLIML estimator has a larger variance than 2SLS to the order of the approximation.

Similarly the comparison of the mean squared errors to order T^{-2} yields:

$$
\begin{aligned}
& MSE(\hat{\alpha}_F) - MSE(\hat{\alpha}) \\
= \; & (-L^2 + 4L - 7)QC_1Q + 2(L-1)QC_2Q + 2(L-1)(trQC_1)Q \\
& + o(T^{-2}).
\end{aligned} \tag{48}
$$

It is not possible to sign this expression but the first term is always negative and will tend to dominate as L increases whereupon the MSE of MLIML will be less than that of 2SLS.

The estimate for the asymptotic variance of MLIML is given by

$$
\hat{V}ar(\hat{\alpha}_F) = \hat{\sigma}_F^2(\hat{Z}_1'\hat{Z}_1)_F^{-1} \tag{49}
$$

Its expectation is evaluated in (40) as:

$$
E(\hat{V}ar(\hat{\alpha}_F)) = \sigma^2 Q + \sigma^2[QCQ + 4QC_1Q + 2tr(QC)Q] + o(T^{-2}) \tag{50}
$$

from which it follows that an approximation to the bias of $\hat{V}ar(\hat{\alpha}_F)$ using (46) and (50) is:

$$
\begin{aligned}
& E(\hat{V}ar(\hat{\alpha}_F) - Var(\hat{\alpha}_F)) \\
= \; & \sigma^2[4QC_1Q + tr(QC)Q - (L-1)QC_2Q + o(T^{-2}).
\end{aligned} \tag{51}
$$

This leads immediately to the following bias corrected estimator:

Theorem 2 *In the model of section (2), an unbiased estimator to order T^{-2} for the variance of the MLIML estimator is given by*

$$
\hat{V}ar(\hat{\alpha}_F)_{BC} = \hat{V}ar(\hat{\alpha}_F) - \hat{\sigma}^2[4\hat{Q}\hat{C}_1\hat{Q} + tr(\hat{Q}\hat{C})\hat{Q} - (L-1)\hat{Q}\hat{C}_2\hat{Q} \tag{52}
$$

where $\hat{V}ar(\hat{\alpha}_F)$ is the asymptotic variance estimate and $\hat{\sigma}^2, \hat{Q}$, $\hat{C}$, $\hat{C}_1$ and $\hat{C}_2$ are consistent estimates of σ^2, Q, C, C_1 and C_2 respectively.

The properties of the approximately unbiased estimators in Theorems 1-2 will be explored in a set of Monte Carlo experiments in section 6. But first we note a problem with bias correction of variance estimates which, as here, involves substacting the estimated bias term from the original estimator in that the corrected version may become negative;clearly this is inappropriate for a variance estimate. When this happens a natural response is not to correct and this procedure was followed in the very small number of cases when it occurred in the simulations.

Finally we shall consider an alternative estimator for the variances based on the non-parametric bootstrap method.

5 Bootstrap Variance Estimation

In this section we consider a bootstrap approach to estimating the variances. Consider the model of section 2. Suppose that the model has been estimated; for simplicity we shall consider the case where estimation has been carried out using 2SLS. The estimated equation is then

$$y_1 = Z_1\hat{\alpha} + \hat{u}_1$$

where $\hat{u}_1$ is a $T \times 1$ vector of 2SLS residuals. As already noted a variance estimate is obtained as

$$\hat{V}ar(\hat{\alpha}) = \hat{\sigma}^2(\hat{Z}_1'\hat{Z}_1)^{-1} .$$

As part of the estimation procedure we also estimate the reduced form:

$$Y_2 = X\hat{\Pi}_2 + \hat{V}_2.$$

The bootstrap data are obtained by first sampling with replacement from the rows of $(\hat{V}_2, \hat{u}_1)$. This yields a corresponding matrix which we label $(V_2^* : u_1^*)$.From these data we can form new values for Y_2 and y_1 as follows:

$$\begin{aligned} Y_2^* &= X\hat{\Pi}_2 + V_2^*, \\ y_1^* &= Z_1\hat{\alpha} + u_1^*. \end{aligned}$$

Suppose that N bootstrap samples are drawn as above and for each sample an estimate for α is obtained. Denote these estimates as $\hat{\alpha}_j^*, , j = 1, 2, ..., N$. The variance estimate is formed as follows:

$$\hat{V}ar(\hat{\alpha})_B = \frac{1}{N}\sum_{j=1}^{N}(\hat{\alpha}_j^* - \hat{\alpha})(\hat{\alpha}_j^* - \hat{\alpha})' \tag{53}$$

which measures the variation of the bootstrap estimates about the original 2SLS estimate $\hat{\alpha}$. It would also be possible to replace $\hat{\alpha}$ with the sample mean $\frac{1}{N}\sum_{j=1}^{N}\hat{\alpha}_j^*$.It is, in principle, possible to develop a bias corrected version of this estimator but the procedure is quite complicated and is not likely to be used in practice; hence we shall not explore this. A question arises concerning the moments of this bootstrap variance estimator. There is no requirement to estimate σ^2 but the variance estimator involves the second moment of the bootstrap estimates with respect to the bootstrap distribution. In the case of a parametric bootstrap variance estimator a non-existence of moments problem will arise but it is not clear that this will carry over to the non-parameteric bootstrap.

For the MLIML estimator, a similar bootstrap variance estimator can be developed whereby there will not be a moments problem. We shall examine and compare the performance of the bootstrap estimator in the simulation experiments which follow.

6 The Simulations

In this section we present the results of a set of simulation experiments designed to examine and compare the approaches to variance estimation that have been developed in the preceding sections. The model specified contains two equations of the form

$$\begin{aligned} y_{1t} &= \alpha_{11} + \beta_{12} y_{2t} + \gamma_{11} x_{1t} + \gamma_{12} x_{2t} + u_{1t}, \\ y_{2t} &= \alpha_{21} + \beta_{21} y_{1t} + \sum_{j=3}^{k} \gamma_{2j} x_{jt} + u_{2t}\ ,\ t = 1, 2, ..., T.\ T = 50 \text{ or } 100. \end{aligned}$$

Interest centres on the estimation of the first equation whereby the order of overidentification of its parameters is determined by the number of exogenous variables in the second equation. Any redundant variables added to the first equation, were also included in the second equation. Only the first equation is estimated.

In matrix notation the system may be written as

$$YB' + X\Gamma' + U = 0.$$

The matrix of endogenous coefficients was chosen as

$$B = \begin{bmatrix} 1.00 & -0.20 \\ 0.00 & 1.00 \end{bmatrix}.$$

The matrix of the exogenous variable coefficients for the case $L = 2$ and where all instruments are strong, was

$$\Gamma = \begin{bmatrix} -1.00 & -0.60 & -1.20 & 0.00 & 0.00 & 0.00 \\ -1.00 & 0.00 & 0.00 & -0.30 & -0.30 & -0.30 \end{bmatrix},$$

and where all instruments are weak, was

$$\Gamma = \begin{bmatrix} -1.00 & -0.60 & -1.20 & 0.00 & 0.00 & 0.00 \\ -1.00 & 0.00 & 0.00 & -0.08 & -0.08 & -0.08 \end{bmatrix}.$$

And as L is increased , further additional columns $\begin{bmatrix} 0.00 \\ -0.30 \end{bmatrix}$ or $\begin{bmatrix} 0.00 \\ -0.08 \end{bmatrix}$ were added.

The corresponding reduced form matrices, Π, were, for $L = 2$,

$$\Pi = \begin{bmatrix} 1.200 & 0.600 & -1.200 & 0.060 & 0.060 & 0.060 \\ 1.000 & 0 & 0 & 0.300 & 0.300 & 0.300 \end{bmatrix}.$$

or

$$\Pi = \begin{bmatrix} 1.200 & 0.600 & -1.200 & 0.016 & 0.016 & 0.016 \\ 1.000 & 0 & 0 & 0.080 & 0.080 & 0.080 \end{bmatrix}$$

and as L increases further columns of $\begin{bmatrix} 0.060 \\ 0.300 \end{bmatrix}$ or $\begin{bmatrix} 0.016 \\ 0.080 \end{bmatrix}$ are added.

The covariance matrix of the structural disturbances was either

$$\Sigma = \begin{bmatrix} 2.00 & 1.80 \\ 1.80 & 2.00 \end{bmatrix} \quad \text{or} \quad \Sigma = \begin{bmatrix} 2.00 & 1.00 \\ 1.00 & 2.00 \end{bmatrix}$$

which gave rise to two different levels of simultaneity in the estimated equation.

The corresponding covariance matrix for the reduced form disturbances was either

$$\Omega = \begin{bmatrix} 2.80 & 2.20 \\ 2.20 & 2.00 \end{bmatrix} \quad \text{or} \quad \Omega = \begin{bmatrix} 2.48 & 1.40 \\ 1.40 & 2.00 \end{bmatrix}.$$

The matrix of reduced form parameters can be varied in the experiments so as to examine the performance of estimators under different levels of overidentification and instruments of different strenghts.

In all our experiments the X matrix contains a first column of ones. The other exogenous variables are generated independently as

$$x_{jt} = 0.95x_{j,t-1} + v_{jt}, \quad j = 1, 2, ..., k, \quad t = 1, 2, ..., T,$$

while the v_{jt} were generated from a $N(0,1)$ distribution. Once a sample of the required size has been generated, the sample was kept fixed over all replications.

The disturbances are generated as the product of a 100×2 matrix of $N(0,1)$ times a Cholesky decomposition matrix. In our experiments only the first structural equation is estimated and we focus on the following:

1. The simulated coefficient estimator bias,denoted $Bias$.
2. The simulated variance of the estimator,denoted Var.
3. The simulated expected value of the asymptotic variance, denoted $EstVar$.
4. The simulated expected value of the bias corrected asymptotic variance estimator,denoted $VarBC$.
5. The simulated expected value of the bootstrap variance estimator, denoted $VarB$.
6. The simulated expected value of the estimator for σ^2,which is denoted as either $\hat{\sigma}^2$ or $\hat{\sigma}^2_F$.

In each simulation experiment there were 20000 replications so that the simulated values for 1-6 above can be taken as essentially the true values. The order of overidentification was varied starting at L=2 so that the 2SLS estimator would have a second moment, and L was increased in steps of two so that typically results covered L=2,4 and 6. Generally estimator biases increase with L while the variance decreases.

It is seen that in each of the reported experiments, instruments may be weak or strong as measured by the R^2 in the reduced form regression of the endogenous variable regressor. Since the order of overidentification, L, is either 2,4 or 6, the R^2,and, hence, the instrument strength, will vary with L. However in the simulations when weak instruments are employed, the R^2 is typically in the range 0.09 to 0.15. In the strong instrument cases R^2 is in the range 0.26 to 0.48, deliberately chosen not to be so strong that there is little difference between the variance estimators. Also the degree of simultaneity may be moderate or strong depending on the correlation, ρ, between the endogenous regressor and the structural disturbance. The degree of simultaneity is moderate when ρ is 0.5 and strong when ρ is 0.9. This is not affected by the the order of overidentification. The sample size in the strong instrument case is chosen as T=50 but when the instruments are weak T=100. We found that in the weak instrument case when T=50 the results were too volatile to be of practical use. We conducted four sets of experiments: 1) strong instruments with strong simultaneity and sample size 50; 2) weak instruments with strong simultaneity and sample size 100; 3) weak instruments with moderate simultaneity and sample size 100, and 4) strong instruments with moderate simultaneity and sample size 50. We now proceed to examine the results which are presented below.

6.1 Simulation Results

In the Tables below it is seen that there are two sections. The first presents the main findings for variance estimation. Then the first two columns present the coefficient estimator bias and the sample variance as found in the 20000 simulations.These are taken to be the "true" values. The coefficient bias is included because of the effect the coefficient bias has on the coverage probabilities to be explored later. The third column gives the ratio of the value of the asymptotic variance, termed EstVar, to the variance, which indicates the proportionate error in practice in estimating the variance using EstVar. The fourth column provides a similar measure of the error in estimating the variance using the bias corrected variance estimate while the last column does the same in respect of the bootstrap variance estimate. Clearly a value close to unity for the corresponding ratio indicates a variance estimator with small bias. It is seen that the above results are given for 2SLS and MLIML and in respect of all three coefficients β, γ_1 and γ_2 in the estimated equation though the coefficient on the endogenous regressor, β, is of particular interest. The second part of the Table uses the first five columns to give more information on the sampling properties of the three variance estimators and the estimator of the disturbance variance σ^2 while the final two columns present the coverage probabilities and test sizes for the associated t-test of the null hypothesis $H_0 : \beta = 0.2$ based upon the three variance estimates.

Experiment 1: Strong Instruments and Strong Simultaneity

Table 1: L=2 T=50 ($R^2 = 0.26, \rho=$ **0.9**)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.067	0.074	1.52	1.09	1.27
	MLIML	0.015	0.065	1.77	1.19	1.01
$\gamma_1 = 0.6$	2SLS	-0.0140	0.014	1.21	1.02	1.08
	MLIML	-0.003	0.014	1.25	1.05	1.01
$\gamma_2 = -1.2$	2SLS	-0.001	0.007	1.10	0.98	0.99
	MLIML	0.001	0.007	1.09	0.99	1.00

2SLS Var($\widehat{\beta}$)=0.074

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.112	0.683	56.509	0.005	0.050	0.859	0.141
VarBC	0.081	0.679	56.509	0.000	0.039	0.830	0.170
VarBoot	0.093	0.203	11.677	0.005	0.049	0.813	0.187
$\hat{\sigma}^2$	1.914	1.327	37.881	0.409	1.602	Na	Na

MLIML Var($\widehat{\beta}_F$)=0.065

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.115	0.192	3.503	0.005	0.059	0.909	0.091
VarBC	0.077	0.165	3.503	0.000	0.041	0.880	0.121
VarBoot	0.065	0.053	0.683	0.005	0.052	0.860	0.140
$\hat{\sigma}^2$	2.064	1.097	10.488	0.439	1.783	Na	Na

In Table 1 above, the value for L is chosen as L=2 and, as noted in the text, the second moment of the 2SLS variance estimator will not exist. While we have shown that the 2SLS variance is smaller than that of MLIML, it is seen that this is not the case here and this may be attributed to the non-existence of the second moment and the outliers that result. The ratio of the asymptotic variance to the true variance is much greater than one for both estimators and, surprisingly, is worse for MLIML; thus both variance estimators are badly biased. The corresponding ratios for the corrected variance estimators are much closer to one and is almost one for the bootstrap variance estimator. Similar comments can be made for the variance estimators of the exogenous coefficients. The second part of the Table indicates that the 2SLS variance estimator is affected by large outliers and this applies also to the disturbance variance estimator although the associated bias of the disturbance variance estimator is relatively small. The coverage probabilities of the t-test are much below the notional 95% in the 2SLS case which is partly due to the fact that the coefficient estimator is very biased and so the distribution of the test t-statistic is shifted to the right in all three cases. Notice that the asymptotic variance estimator, EstVar, is strongly biased upwards which is in accordance with the findings of Kiviet and Phillips (1999) noted earlier; hence when the t-ratio is correctly centred at zero, the overstated variance/standard deviation will reduce the spread of the t-statistic and there will be a tendency for the t-test to under-reject. Thus the nominal size of the test is overstated.

However, when the t-statistic is not centred at zero due to coefficient estimator bias and /or the distribution of the test t-statistic is far from normal, test sizes greater than the notional will be expected and that would explain the large empirical sizes which are observed.

Table 2: L=4 T=50 $(R^2 = 0.44, \rho = 0.9)$

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.085	0.026	1.17	0.92	0.90
	MLIML	0.003	0.035	1.19	1.02	1.07
$\gamma_1 = 0.6$	2SLS	-0.014	0.010	1.06	1.00	0.92
	MLIML	-0.001	0.012	1.06	1.01	1.07
$\gamma_2 = -1.2$	2SLS	-0.017	0.007	1.05	0.98	0.90
	MLIML	-0.001	0.008	1.06	1.00	1.06

2SLS $\text{Var}(\widehat{\beta})$=0.026							
$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.031	0.033	2.591	0.004	0.024	0.846	0.154
VarBC	0.024	0.024	2.591	0.004	0.020	0.826	0.174
VarBoot	0.024	0.016	0.602	0.003	0.020	0.845	0.155
$\hat{\sigma}^2$	1.766	0.665	15.554	0.488	1.643	Na	Na
MLIML $\text{Var}(\widehat{\beta}_F)$=0.035							
$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.042	0.045	1.580	0.004	0.030	0.926	0.074
VarBC	0.035	0.034	1.580	0.004	0.027	0.921	0.079
VarBoot	0.037	0.029	0.748	0.004	0.029	0.910	0.090
$\hat{\sigma}^2$	2.055	0.857	11.246	0.505	1.875	Na	Na

In Table 2 where L=4, it is again seen that 2SLS is more biased while MLIML has virtually zero bias. The 2SLS variance estimator now has necessary moments and it is seen to be less that of MLIML in line with the earlier approximations. The ratios of the asymptotic variance to the actual variance are now closer to but greater than unity in all cases and the ratio which involves the bias corrected variance is very close to unity. The ratios for the bootstrap variance are noticeably less than unity for 2SLS and larger than unity for MLIML. In the second part of the Table, none of the variance estimates are affected by extreme outliers but it is noticeable that the bootstrap variance estimator is decidedly less volatile than the others. The disturbance variance estimator $\hat{\sigma}^2$ based on 2SLS is substantially biased whereas based on MLIML there is little bias. Examining the coverage probabilities, the 2SLS t-statistic has a coverage probability which leads to a test size about three times the notional 5% in all three cases while the MLIML statistics are much closer to 5%. However the t-statistics based on bias corrected variances have slightly larger size for 2SLS. Again we see that although the 2SLS asymptotic variance is biased upwards the test size is too large. This is at least partly explained by the fact that the coefficient estimator is considerably upwards biased so that the t-statistic is shifted to the right.

Table 3: L=6 T=50 ($R^2 = 0.48, \rho = 0.9$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.123	0.021	1.19	0.91	0.82
	MLIML	0.001	0.033	1.15	1.05	1.02
$\gamma_1 = 0.6$	2SLS	0.033	0.010	1.07	1.00	0.87
	MLIML	0.000	0.014	1.06	1.02	1.06
$\gamma_2 = -1.2$	2SLS	0.025	0.006	1.06	0.99	0.86
	MLIML	0.000	0.008	1.05	1.01	1.05

2SLS Var($\widehat{\beta}$)=0.021

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.024	0.018	0.529	0.004	0.020	0.785	0.215
VarBC	0.018	0.011	0.529	0.004	0.017	0.755	0.245
VarBoot	0.017	0.009	0.139	0.003	0.015	0.785	0.215
$\hat{\sigma}^2$	1.646	0.547	8.782	0.549	1.552	Na	Na

MLIML Var($\widehat{\beta}_F$)=0.033

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.037	0.039	1.450	0.004	0.027	0.928	0.073
VarBC	0.034	0.032	0.968	0.004	0.026	0.924	0.076
VarBoot	0.033	0.026	0.422	0.004	0.026	0.918	0.082
$\hat{\sigma}^2$	2.054	0.839	12.864	0.571	1.887	Na	Na

A feature of Table 3 where L is now increased to 6, is that the 2SLS coefficient estimator biases are increased while those for MLIML are essentially zero and the 2SLS variances are all less that those for MLIML .The ratios based on the bias corrected variances are close to unity and In terms of the ratios of the three variance estimates to the variance there is little change from Table 2 except that the 2SLS ratios based upon the bootstrap variance continue to be less than unity. In the second part of the Table it is seen that the 2SLS asymptotic variance is upwards biased while the other variance estimators are downwards biased. There is no sign of extreme outliers in either case but again the bootstrap variance estimator is less variable. The estimator for the disturbance variance is biased for both estimators and some relatively extreme values occur. The coverage probabilities for the 2SLS t-statistic are now smaller which is partly explained by the increased coefficient bias, and the associated test sizes are around four times the nominal 5%. The t-statistic based on the bias corrected variance has a larger size. In the MLIML case the disturbance variance estimator has very small bias while coverage probabilities and, correspondingly, the test sizes are much closer to the nominal levels. In particular the test sizes are now close to those in Table 2.

Experiment 2: Weak instruments, moderate simultaneity. T=100

Table 4: L=2 T=100 ($R^2 = 0.10, \rho = 0.5$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.054	0.106	1.55	1.24	1.28
	MLIML	0.027	0.104	1.41	1.10	1.01
$\gamma_1 = 0.6$	2SLS	-0.002	0.008	1.17	1.03	1.16
	MLIML	-0.000	0.008	1.14	1.02	1.12
$\gamma_2 = -1.2$	2SLS	-0.001	0.004	1.19	1.06	1.17
	MLIML	-0.000	0.004	1.16	1.03	1.13

2SLS Var($\widehat{\beta}$)=0.106

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.164	2.118	197.315	0.012	0.076	0.925	0.076
VarBC	0.132	2.117	197.315	0.002	0.066	0.906	0.094
VarBoot	0.136	0.538	53.235	0.011	0.081	0.877	0.123
$\hat{\sigma}^2$	2.108	1.090	60.893	0.924	1.873	Na	Na

MLIML Var($\widehat{\beta}_F$)=0.104

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.147	0.196	2.706	0.012	0.085	0.936	0.064
VarBC	0.115	0.162	2.706	0.000	0.073	0.920	0.080
VarBoot	0.106	0.068	0.733	0.012	0.088	0.874	0.126
$\hat{\sigma}^2$	2.156	0.825	12.625	0.924	1.929	Na	Na

In Table 4 , it is seen that the instruments are weak. Since L=2 there is a non-existence of moments problem for the 2SLS variance estimator and , possibly, for the bootstrap variance estimator. Both coefficient estimators are quite biased and the variance of 2SLS is shown to exceed that of MLIML which is explained by the occurrence of large outliers. The ratios of the 2SLS asymptotic variance to the variance is noticeably larger than unity but this is much improved when the bias corrected variance is considered. The ratio based on the bootstrap variance is too large for 2SLS but better when MLIML is employed. In the second part all three variance estimators based on 2SLS are seen to strongly biased upwards and there are extreme outliers in all three cases. Here there is an indication that there is a moments problem for the bootstrap estimator. In the MLIML case the asymptotic variance is again biased upwards but the biased corrected variance removes most of the bias, Also the bootstrap variance estimator is close to being unbiased. There is no indication of outliers when MLIML is used. The 2SLS estimator for the disturbance variance is much affected by extreme outliers as expected while there is little bias. The corresponding MLIML estimator is not affected by outliers and also has little bias.The coverage probabilities for 2SLS again lead to test sizes which are above the nominal 5% but not greatly so and now they are relatively close to those for MLIML. Here bias correcting the variances does not lead to improved test size.

Table 5: L=4 T=100 ($R^2 = 0.09, \rho = 0.5$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.141	0.082	1.25	0.95	0.83
	MLIML	0.043	0.141	1.31	1.13	0.84
$\gamma_1 = 0.6$	2SLS	-0.008	0.008	1.12	1.00	1.03
	MLIML	-0.002	0.009	1.19	1.10	1.15
$\gamma_2 = -1.2$	2SLS	-0.008	0.004	1.11	0.98	1.01
	MLIML	-0.002	0.005	1.17	1.08	1.10

2SLS Var($\widehat{\beta}$)=0.082

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.102	0.279	29.597	0.012	0.068	0.885	0.115
VarBC	0.078	0.274	29.597	0.000	0.058	0.859	0.141
VarBoot	0.068	0.042	0.573	0.010	0.057	0.880	0.120
$\hat{\sigma}^2$	1.919	0.701	34.988	0.839	1.772	Na	Na

MLIML Var ($\widehat{\beta}_F$)=0.141

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.185	0.265	4.234	0.013	0.092	0.922	0.078
VarBC	0.159	0.217	3.122	0.006	0.089	0.916	0.084
VarBoot	0.119	0.078	0.837	0.013	0.098	0.872	0.129
$\hat{\sigma}^2$	2.200	0.927	13.476	0.839	1.928	Na	Na

In Table 5 the 2SLS estimator is very biased and the MLIML estimator has a bias of 20%.The variance of 2SLS is much lower than that of MLIML. The variance ratios involving the asymptotic variance are well above unity for both estimators and for all parameters with those based on MLIML being larger. The bias corrected ratios are much closer to unity and the bootstrap variance ratios are clearly below unity for the endogenous parameter. Thus moving from L=2 to L=4 has a large effect on the bootstrap variance ratios which supports the non-existence of a second moment for the bootstrap variance when L=2. There is some evidence of extreme values for the 2SLS variance estimators excluding the bootstrap based estimator. The asymptotic variance estimator based on MLIML is heavily biased upwards while bias correction is effective. The bootstrap estimator is biased downwards. Examining the coverage probabilities and associated t-test sizes, it is again seen that 2SLS based tests are considerably oversized and more so than when L=2 while the MLIML based tests are less oversized. However the bootstrap based test with MLIML is the most oversized here. We again note that even though the MLIML variance estimator is heavily biased upwards whereas the coefficient bias is not especially large, the resulting tests still have an empirical size which is too large.

Table 6: L=6 T=100 $(R^2 = 0.16, \rho = 0.5)$

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.116	0.041	1.10	0.92	0.82
	MLIML	0.002	0.080	1.06	1.01	0.85
$\gamma_1 = 0.6$	2SLS	0.008	0.007	1.04	0.98	0.98
	MLIML	0.000	0.009	1.07	1.04	1.09
$\gamma_2 = -1.2$	2SLS	0.001	0.004	1.05	1.00	1.00
	MLIML	0.000	0.004	1.08	1.06	1.11

2SLS $\text{Var}(\widehat{\beta})$=0.041

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.045	0.033	1.072	0.009	0.037	0.874	0.126
VarBC	0.037	0.023	1.072	0.005	0.034	0.862	0.138
VarBoot	0.034	0.015	0.194	0.008	0.030	0.874	0.126
$\hat{\sigma}^2$	1.878	0.443	7.582	0.775	1.802	Na	Na

MLIML $\text{Var}(\widehat{\beta}_F)$=0.080

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.085	0.137	2.501	0.010	0.050	0.929	0.071
VarBC	0.081	0.123	2.289	0.010	0.050	0.927	0.073
VarBoot	0.068	0.049	0.801	0.010	0.054	0.907	0.093
$\hat{\sigma}^2$	2.161	0.798	11.662	0.776	1.965	Na	Na

In Table 6 where L=6, the instrument strength increases which makes a direction comparison with the previous Table 5 less possible. The coefficient biases and the associated variances are smaller now. As a result the variance ratios of interest are quite close to unity except for 2SLS and MLIML bootstrap ratios for the endogenous parameter which are well below unity as a result of the bootstrap variance estimators being considerably biased downwards. The other variance estimators have a small bias and bias correction works well. There is no evidence of extreme values. The estimators of the disturbance variance are little biased. The coverage probabilities for 2SLS and the associated test sizes are again too large and those based on the bias corrected variances a little higher than the others. The MLIML test sizes are quite close to the nominal level but, as in Table 5, the bootstrap based tests are the most oversized.

Experiment 3: Weak Instruments and Strong Simultaneity T=100

Table 7: L=2 T=100 ($R^2 = 0.10, \rho = 0.9$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.098	0.095	1.76	1.34	1.25
	MLIML	0.041	0.070	2.18	1.49	1.00
$\gamma_1 = 0.6$	2SLS	-0.005	0.009	1.22	1.07	1.02
	MLIML	-0.002	0.009	1.18	1.05	0.99
$\gamma_2 = -1.2$	2SLS	-0.003	0.004	1.20	1.06	1.00
	MLIML	-0.001	0.004	1.17	1.04	0.98

2SLS Var($\widehat{\beta}$)=0.095

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.167	0.865	50.696	0.005	0.059	0.824	0.176
VarBC	0.132	2.117	197.315	0.002	0.066	0.906	0.094
VarBoot	0.136	0.538	53.235	0.011	0.081	0.877	0.123
$\hat{\sigma}^2$	1.856	1.485	35.656	0.368	1.482	Na	Na

MLIML Var($\widehat{\beta}_F$)=0.070

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.153	0.229	2.909	0.005	0.075	0.894	0.106
VarBC	0.105	0.203	2.909	0.000	0.040	0.854	0.146
VarBoot	0.070	0.052	0.824	0.006	0.056	0.824	0.176
$\hat{\sigma}^2$	1.990	1.034	9.626	0.363	1.712	Na	Na

In Table 7 we consider the worst possible case for estimation; both weak instruments and strong simultaneity. Hence, unsurprisingly the 2SLS coefficient estimator for the endogenous variable regressor is strongly biased and the corresponding MLIML estimator has a substantial bias also. The ratios of the asymptotic variance to the variance are far from unity. This is not unexpected for 2SLS when L=2 but the very large value in the MLIML case is surprising. Both results reflect the large biases in the asymptotic variance.However the corrected variance ratio is much nearer unity for all coefficients and the bootstrap variance ratios even better especially for MLIML. As expected, there is evidence of large outliers in the 2SLS case. The estimated disturbance variance is clearly biased for 2SLS and extreme values are evident while MLIML shows little bias. Despite the large upward biases in the variance estimators the coverage probabilities are much higher than the notional level leading to size distortions in both cases especially for 2SLS. Tests based on the corrected variance have better size in the 2SLS case but are worse for MLIML. The bootstrap based test for MLIML is particularly oversized.

Table 8: L=4 T=100 ($R^2 = 0.09, \rho = 0.9$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.257	0.056	1.45	0.98	0.66
	MLIML	0.059	0.077	2.25	1.55	0.91
$\gamma_1 = 0.6$	2SLS	-0.015	0.005	1.17	1.04	0.80
	MLIML	-0.002	0.009	1.25	1.08	1.03
$\gamma_2 = -1.2$	2SLS	-0.015	0.003	1.15	1.01	0.77
	MLIML	-0.004	0.004	1.25	1.07	1.00

2SLS Var($\widehat{\beta}$)=0.056

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.081	0.187	8.587	0.005	0.043	0.646	0.354
VarBC	0.054	0.186	8.587	0.000	0.027	0.554	0.446
VarBoot	0.037	0.029	0.516	0.004	0.029	0.645	0.355
$\hat{\sigma}^2$	1.327	0.751	14.637	0.313	1.138	Na	Na

MLIML Var($\widehat{\beta}_F$)=0.077

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.173	0.250	3.299	0.006	0.082	0.876	0.124
VarBC	0.119	0.196	2.545	0.001	0.054	0.852	0.148
VarBoot	0.070	0.051	0.526	0.006	0.057	0.828	0.173
$\hat{\sigma}^2$	2.200	0.927	13.476	0.839	1.928	Na	Na

Table 8 shows that 2SLS is hugely biased when L is increased and the asymptotic variance ratio is much greater than unity. However the bias corrected variance ratio is very close to unity while the ratio based on the bootstrap variance estimator is far less than unity. For MLIML the asymptotic variance ratio is much greater than in the 2SLS case though the bias corrected variance ratio while much reduced is still well above unity for the endogenous regressor but close to unity for the other coefficients. The ratios based on the bootstrap give a different picture where they are well below unity for 2SLS and very close to unity for MLIML. The problems with the variance ratios are explained by the large upward bias in the variance estimators and this is especially so in the MLIML case.The associated estimator for the disturbance variance is greatly biased downwards for 2SLS and there is a moderate upward bias for MLIML. The coverage probabilities are far from the nominal level for 2SLS and the test sizes are over six times as large. The situation is much better in the MLIML case but still around two and a half times the nominal level. Bias correcting the variances makes matters worse for 2SLS and has little effect for MLIML.

Table 9: L=6 T=100 ($R^2 = 0.16, \rho = 0.9$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.206	0.029	1.27	0.82	0.68
	MLIML	0.008	0.056	1.47	1.17	0.94
$\gamma_1 = 0.6$	2SLS	0.014	0.005	1.07	0.99	0.79
	MLIML	0.000	0.009	1.10	1.02	1.07
$\gamma_2 = -1.2$	2SLS	0.002	0.003	1.06	0.99	0.79
	MLIML	0.001	0.004	1.08	1.01	1.07

2SLS Var($\widehat{\beta}$)=0.029

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.037	0.045	2.595	0.004	0.026	0.659	0.341
VarBC	0.024	0.037	2.595	0.000	0.020	0.594	0.406
VarBoot	0.020	0.011	0.197	0.003	0.017	0.659	0.341
$\hat{\sigma}^2$	1.414	0.550	10.657	0.457	1.297	Na	Na

MLIML Var($\widehat{\beta}_F$)=0.056

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.083	0.138	3.258	0.004	0.047	0.906	0.094
VarBC	0.066	0.105	2.899	0.004	0.041	0.898	0.102
VarBoot	0.053	0.042	0.526	0.004	0.041	0.892	0.108
$\hat{\sigma}^2$	2.080	1.026	12.848	0.470	1.827	Na	Na

In Table 9, it is again seen that the 2SLS coefficient estimator is badly biased but MLIML is virtually unbiased. Here the instruments are somewhat less weak than in Table 8. The strong upward biases in the asymptotic variance estimates are repeated with those based on MLIML again being the largest. Thus the asymptotic variance ratios are well above unity for the endogenous regressor though generally close to unity for the other coefficients. The bias correction is seen to work quite well in this case though the bootstrap variance ratios are well below unity. The estimator for the disturbance based on 2SLS is greatly biased downwards while there is a small upwards bias in the MLIML counterpart. The coverage probabilities and the associated test sizes and hugely far from the nominal levels. The situation is much better with the MLIML based tests however, and are better than those in Table 8. While bias correction works well when applied to the variance ratios it appears to have little effect on the tests.

Experiment 4: Strong Instruments and Moderate Simultaneity T=50

Table 10: L=2 T=50 ($R^2 = 0.26, \rho = 0.5$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.038	0.076	1.25	0.98	1.25
	MLIML	0.011	0.081	1.32	1.06	1.06
$\gamma_1 = 0.6$	2SLS	-0.008	0.015	1.12	0.99	1.14
	MLIML	-0.002	0.015	1.15	1.02	1.10
$\gamma_2 = -1.2$	2SLS	-0.001	0.007	1.08	0.98	1.10
	MLIML	-0.001	0.007	1.09	1.00	1.10

2SLS Var($\widehat{\beta}$)=0.076

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.094	0.231	16.175	0.008	0.059	0.934	0.066
VarBC	0.074	2.209	16.175	0.002	0.005	0.921	0.079
VarBoot	0.095	0.132	5.602	0.009	0.061	0.898	0.102
$\hat{\sigma}^2$	2.068	0.831	29.077	0.620	1.901	Na	Na

MLIML Var($\widehat{\beta}_F$)=0.081

$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.107	0.150	2.819	0.010	0.064	0.944	0.056
VarBC	0.086	0.120	2.640	0.002	0.059	0.934	0.066
VarBoot	0.086	0.060	0.744	0.009	0.069	0.890	0.110
$\hat{\sigma}^2$	2.132	0.815	10.267	0.622	1.950	Na	Na

The results in Table 10 are for the case most favourable to estimation. It is seen that coefficient biases are now much smaller. In contrast to the results in other Tables, when L=2 the 2SLS variance is found to be less than that of MLIML which suggests that extreme values had less influence in the results despite the non-existence of the second moment. The asymptotic variances are generally upward biased but the bias correction works very well to bring the variance ratios close to unity. The ratios based on the bootstrap are similarly close. The disturbance variance estimator has small bias though MLIML has more bias than 2SLS. The coverage probabilities and test sizes are now closer to nominal levels with the MLIML based tests being slightly better except for the bootstrap based tests. The standard t-tests are slightly superior here and the variance bias correction has little effect. However the bootstrap based tests are substantially over sized for both 2SLS and MLIML

Table 11: L=4 T=50 ($R^2 = 0.44, \rho = 0.5$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.049	0.030	1.08	0.97	0.96
	MLIML	0.000	0.040	1.03	0.99	1.01
$\gamma_1 = 0.6$	2SLS	-0.009	0.012	1.03	0.99	1.00
	MLIML	-0.001	0.013	1.04	1.02	1.06
$\gamma_2 = -1.2$	2SLS	0.025	0.006	1.04	0.99	1.00
	MLIML	-0.011	0.008	1.04	1.02	1.07

			2SLS Var($\widehat{\beta}$)=0.030				
$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.032	0.021	0.807	0.005	0.027	0.924	0.076
VarBC	0.029	0.015	0.807	0.005	0.025	0.917	0.083
VarBoot	0.029	0.015	0.339	0.005	0.025	0.923	0.077
$\hat{\sigma}^2$	1.963	0.533	7.937	0.692	1.891	Na	Na
			MLIML Var($\widehat{\beta}_F$)=0.040				
$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.041	0.048	2.845	0.005	0.031	0.943	0.057
VarBC	0.039	0.043	2.764	0.005	0.031	0.941	0.059
VarBoot	0.040	0.027	0.489	0.005	0.033	0.925	0.075
$\hat{\sigma}^2$	2.078	0.650	11.367	0.695	1.970	Na	Na

In Table 11 there is still a substantial 2SLS coefficient bias but none for MLIML. All the variance ratios are close to one reflecting the low variance estimation bias for all three variance estimators. The estimators of the disturbance variance have a very small bias. The coverage probabilities for the t-statistics based on 2SLS and the associated test sizes are close to the nominal values though the test sizes are about 50% too large. None the less they are the closest to the 5% level that have been found in the experiments. The coverage probabilities and test sizes for tests based on MLIML and the asymptotic variance or the bias corrected variance are very close to the 5% level while tests based on the bootstrap variance have a slightly larger size. Hence when the coefficient estimator bias and the variance estimator biases are small one can expect the tests to have approximately the correct size.

Table 12: L=6 T=50 ($R^2 = 0.48, \rho = 0.5$)

		Bias	Var	$\frac{\text{EstVar}}{\text{Var}}$	$\frac{\text{VarBC}}{\text{Var}}$	$\frac{\text{VarBoot}}{\text{Var}}$
$\beta = 0.2$	2SLS	0.070	0.025	1.08	0.97	0.90
	MLIML	0.000	0.037	0.99	0.98	0.97
$\gamma_1 = 0.6$	2SLS	0.018	0.012	1.03	0.99	0.97
	MLIML	0.000	0.014	1.02	1.01	1.04
$\gamma_2 = -1.2$	2SLS	0.015	0.008	1.04	1.00	0.98
	MLIML	0.001	0.009	1.03	1.02	1.06

			2SLS Var($\widehat{\beta}$)=0.025				
$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.027	0.014	0.351	0.005	0.023	0.905	0.095
VarBC	0.024	0.011	0.344	0.005	0.022	0.898	0.102
VarBoot	0.022	0.010	0.140	0.005	0.020	0.905	0.095
$\hat{\sigma}^2$	1.921	0.486	5.432	0.707	1.866	Na	Na
			MLIML Var($\widehat{\beta}_F$)=0.037				
$\beta = 0.2$	Mean	Sdv	Max	Min	Med	CovProb	size(t)
EstVar	0.037	0.043	2.168	0.006	0.028	0.941	0.059
VarBC	0.036	0.041	2.167	0.006	0.028	0.941	0.059
VarBoot	0.036	0.024	0.440	0.006	0.030	0.927	0.073
$\hat{\sigma}^2$	2.079	0.640	11.196	0.712	1.978	Na	Na

The results in Table 12 have much in common with Table 11. The coefficient estimator for the endogenous coefficient has a larger bias though of about 35% as L is increased to 6 while there is no bias in MLIML. All the variance ratios are close to unity reflecting the very small biases in the variance estimators. The estimators of the disturbance variance have a small bias. However the coverage probabilities for the 2SLS t-test statistics are below notional values and the associated test sizes are nearly twice the nominal value which stems from the bias in the coefficient estimator. The test sizes for the MLIML tests based on the asymptotic variance and the bias corrected variance are both very close to the 5% level while reflects the fact that there is no coefficient bias and the estimated variances are all virtually unbiased. Test sizes for tests based on the bootstrap variance are somewhat larger.

6.2 Power of the t-test

In the remaining part of the paper we examine the powers of the two types of t-test. We do this for the cases of $L = 4$, so that the 2SLS estimator has no moments problems. As we observed from the experiments the the coefficient estimator is biased upwards, so that the distribution of the t-statistic is shifted to the right. To get a better understanding, we examine the power of the two types of $t-$ test: left tail and right tail t-tests. In the left tail test, we test H_0^L : $\beta = 0$ against $H_1^L : \beta < 0$. In right tail test, we test $H_0^R : \beta = 0$ against H_1^R : $\beta > 0$. We do this for cases: i): Strong instruments and strong simultaneity, $T = 50$, $L = 4$. ii) Weak instruments and strong simultaneity $T = 100, L = 4$. The experiment setup and the choice of parameters in DGPs are same as in previous section, except for β. See the table below for the choice of β in the DGPs.

Table 13 : Values for β in t-test

Left tail test	Right tail test
0	0
−0.06	0.06
−0.12	0.12
−0.18	0.18
−0.24	0.24
−0.30	0.30
−0.36	0.36
−0.42	0.42
−0.48	0.48
−0.54	0.54

The results are reported in the figures below.The coefficient estimator bias of 2SLS is again quite large and positive so that the distribution of the t-statistic is shifted to the right. This has an effect on the standard t-statistic. It can be seen that the size of the left tail t-test is too small while the size of the right tail t-test is too large. This is true whenever we use an asymptotic or bias corrected variance estimator in the t-statistics. However, the power of the t-statistics when using the bias corrected variance estimator shows a clear difference. We can see that the power is higher when using the bias corrected variance than when using the asymptotic variance estimator. This is true for both the left tail and right tail test. This is also true when the instruments are strong or weak. The t-statistics when using the bootstrap variance estimator works in a way similar to the one using the bias corrected variance estimator. They have similar powers, which are both significantly higher than for the t-statistics which use the asymptotic variance. The bias corrected variance works slightly better than the bootstrap variance estimator in some cases. The results for MLIML are somewhat similar to the test powers in the 2SLS case. In all,what we can learn from this experiment is that the false hypothesis is more likely to be rejected when we correct the bias of the variance estimators in the t-statistics.

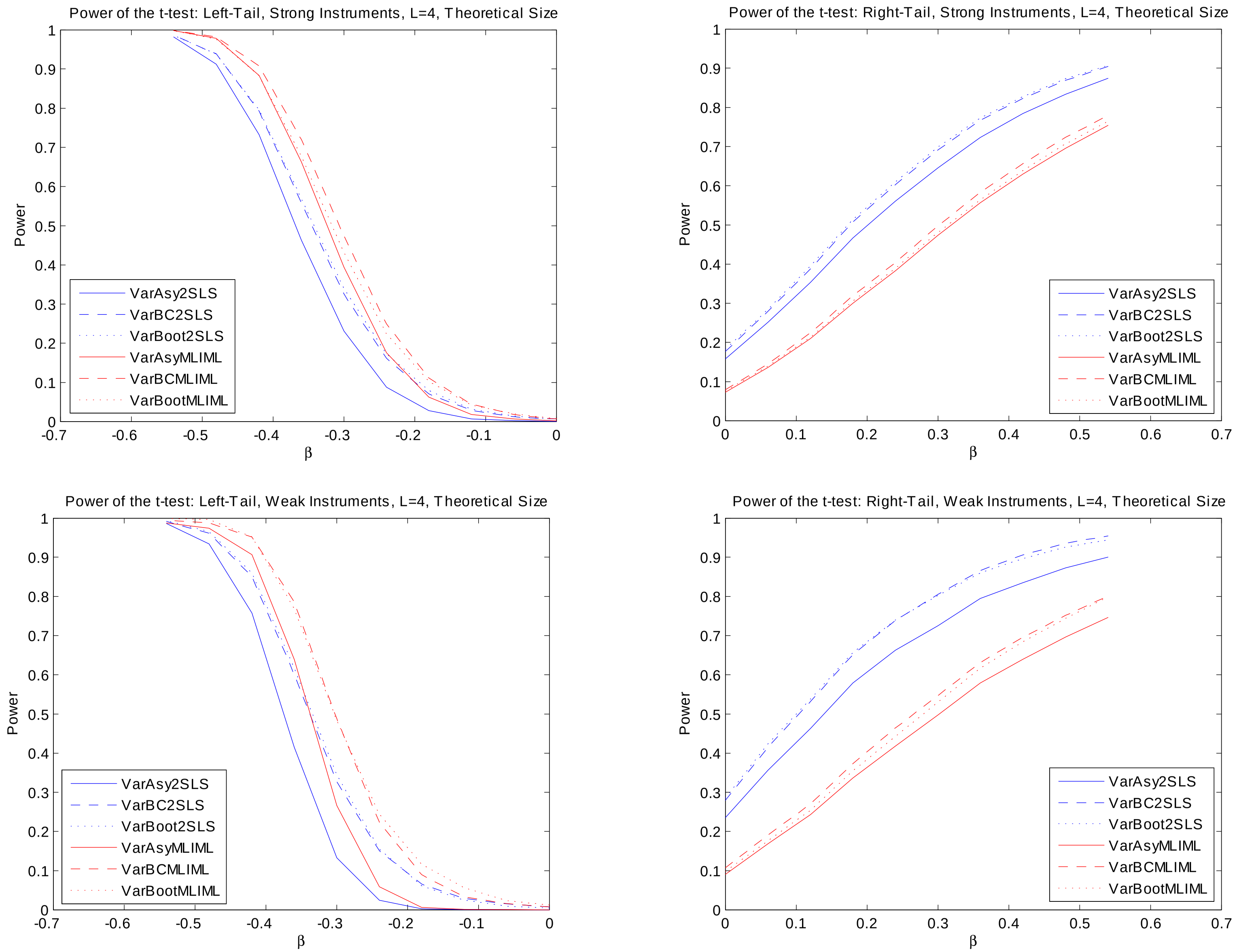

Figure 1: Power of t-test: one tail test.

7 Conclusion

This paper shows that the bias in the usual variance estimator for simultaneous equation coefficient estimators can be substantially reduced by using a bias correction. However, the bias corrected estimator may sometimes yield a negative estimate and when this occurs an obvious solution is to simply replace the bias corrected estimate by the usual estimate. This may not be a serious drawback to using bias correction since if negative estimates are relatively infrequent the bias corrected estimator will be little affected. In our experiments there was every sign that this approach is valid. An alternative is to use a bootstrap variance estimator. This bootstrap is shown to have good finite sample properties, including that of relatively small bias which in most cases rival those of the bias corrected variance estimator. Also, since the bootstrap estimator is inherently positive, there are no problems of negative estimates.

One particularly relevant result arising out of this paper is that the 2SLS estimator of the disturbance variance does not have a second moment unless L, the order of overidentification, is at least 4. Since the disturbance variance estimate is inherently part of the actual variance estimate, this imposes severe restrictions on the practical use of 2SLS for inference purposes unless some other variance estimate is used which does not depend on the estimated disturbance variance.This is not a limitation which applies to estimators which have all necessary moments such as the MLIML estimator explored in this paper although the variances associated with both estimators are affected by the volatility in the disturbance variance estimate. In situations where 2SLS is ruled out, MLIML is the best alternative since it has the smallest coefficient estimator bias and is more efficient.

The experiments which investigated coverage probabilities and the associated test sizes were very informative. It was seen that 2SLS based tests were generally seriously oversized whereas tests based on MLIML were clearly closer to nominal levels. Some results were also given for a comparison of t-test powers. These showed that the powers of tests based on 2SLS can be much affected by coefficient estimator bias which may enhance or reduce the test power. Hence in some cases the power may be low and substantially less than that of the MLIML based tests and in other cases corresponding higher.

Overall our results so far suggest that MLIML based inference is to be recommended, used in conjunction with either the bias corrected variance estimate or the associated bootstrap variance estimate although overall our simulations suggest that using the bias corrected variance estimate has some advantages in terms of test size and power.

Acknowledgement

The authors wish to acknowledge helpful comments from Grant Hillier and Peter C.B.Phillips which led to improvements in the paper.

References

[1] Anderson,T.W.,Kunitomo,K and Morimune K,(1986). Comparing Single-Equation estimators in a Simultaneous Equation System. *Econometric Theory*,2,1-32

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

[3] Fuller, W. A. (1977), Some Properties of a Modification of the Limited Information Estimator, *Econometrica* 45, 939-954.

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

[5] 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.

[6] Hahn, J., Hausman, J. & Kuersteiner, G. (2004). Estimation with weak instruments: accuracy of higher-order bias and MSE approximations. Econometrics Journal 7(1), 272-306.

[7] Kadane, J. (1971). Comparison of k-class estimators when the disturbances are small. Econometrica 39, 723-737

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

[9] Kiviet,J.F. and Phillips,G.D.A.(2000). The Bias of the Two Stage Least Squares Estimator. *Economics Letters 66,7-15.*

[10] Kiviet,J.F. and Phillips,G.D.A.(2014). Improved variance Estimation of maximum Likelihood Estimators in Stable First-Order Dynamic regression Models. *Computational Statistics and Data Analysis 76, 424-448.*

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

[12] 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.

[13] Nagar,A.L. (1961), A Note on the Residual Variance Estimator in Simultaneous Equations. Econometrica,29, 238-243.

[14] 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.

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

[16] Smith, Murray D. (1994), Exact densities for variance estimators of the structural disturbances in simultaneous equations models, *Journal of Econometrics 60*, 157-180.

Appendix

A. An approximation to the MLIML estimator

In the first part of this appendix we obtain a number of results which were used in section 3 in relation to the MLIML estimator. We commence with the asymptotic expansion for the estimation error. First we shall write the Fuller estimator error as

$$\alpha_F - \alpha = (\hat{Z}_1'\hat{Z}_1)_F^{-1}\hat{Z}_{1F}'u_1$$

where

$$(\hat{Z}_1\hat{Z}_1)_F = \bar{Z}_1'\bar{Z}_1 + V_Z'\bar{Z}_1 + \bar{Z}_1'V_Z + (1-\lambda)V_Z'V_Z + \frac{1}{T-K}V_Z'V_Z + \lambda V_Z'M^*V_Z - \frac{1}{T-K}V_Z'M^*V_Z.$$

Putting $\bar{Z}_1'\bar{Z}_1 = Q^{-1}$ and rearranging leads to

$$\begin{aligned}(\hat{Z}_1\hat{Z}_1)_F^{-1} &= [I + Q(V_Z'\bar{Z}_1 + \bar{Z}_1'V_Z + (1-\lambda)V_Z'V_Z \\ &\quad + \frac{1}{T-K}V_Z'V_Z + \lambda V_Z'M^*V_Z - \frac{1}{T-K}V_Z'M^*V_Z)]^{-1}Q \\ &= Q - Q(V_Z'\bar{Z}_1 + \bar{Z}_1'V_Z)Q - Q((1-\lambda)V_Z'V_Z + \frac{1}{T-K}V_Z'V_Z)Q \\ &\quad - Q\lambda V_Z'M^*V_ZQ + \frac{1}{T-K}QV_Z'M^*V_ZQ \\ &\quad + Q(V_Z'\bar{Z}_1 + \bar{Z}_1'V_Z)Q(V_Z'\bar{Z}_1 + \bar{Z}_1'V_Z)Q + o_p(T^{-2}). \qquad (A.1)\end{aligned}$$

The associated term $\hat{Z}_{1F}'u_1$is given by:

$$\bar{Z}_1'u_1 + V_Z'u_1 - (\lambda - \frac{1}{T-K})V_Z'(I - M^*)u_1. \qquad (A.2)$$

Multiplying the terms in (A.1) by (A.2) and retaining terms up to order $O_p(T^{-1})$ yields

$$\begin{aligned}\alpha_F - \alpha &= Q\bar{Z}_1'u_1 + Q(1 - (\lambda - \frac{1}{T-K}))V_Z'u_1 + \lambda QV_Z'M^*u_1 \\ &\quad - QV_Z'\bar{Z}Q\bar{Z}u_1 - Q\bar{Z}'V_ZQ\bar{Z}'u_1 + o_p(T^{-1}). \qquad (A.3)\end{aligned}$$

Taking the expected value of (A.3) shows that the bias to disappears to order T^{-1}.

B. An approximation to the two expectations in (37)

Next we consider two expectations that are required in (37), section 3.

$$(i) E[(-\frac{2(1-(\lambda-\frac{1}{T-K})u_1'V_ZQV_Z'u_1}{T-g-k}Q)]$$
$$(ii)\ E[Q((1-\lambda)V_Z'V_Z+\frac{1}{T-K}V_Z'V_ZQ].$$

(i) An approximation to $E[(-\frac{2(1-(\lambda-\frac{1}{T-K})u_1'V_ZQV_Z'u_1}{T-g-k}Q)]$

To proceed we note that from Kadane(1960) we have

$$1-\lambda=\frac{-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u_1'\bar{P}_Xu_1}+o_p(T^{-1}) \qquad \text{(B.1)}$$

so that for (i) we have

$$\begin{aligned} & E[(-\frac{(1-(\lambda-\frac{1}{T-K})u_1'V_ZQV_Z'u_1}{T-g-k}Q)] \\ = \quad & E[\frac{-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u'\bar{P}_Xu}\frac{u_1'V_ZQV_Z'u_1}{T-g-k}Q-\frac{u_1'V_ZQV_Z'u_1}{(T-K)(T-g-k)}Q]+o(T^{-2}) \end{aligned} \qquad \text{(B.2)}$$

To evaluate the first of these terms we shall put $V_Z=[W+u_1\pi':0]$ where W and u_1 are independent so that

$$\begin{aligned} & E[\frac{-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u_1'\bar{P}_Xu_1}\frac{u_1'V_ZQV_Z'u_1}{T-g-k}Q] \\ = \quad & E[\frac{-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u_1'\bar{P}_Xu_1}(u_1'WQ_{11}W'u_1+u_1'u_1\pi Q_{11}\pi'u_1'u_1)Q] \end{aligned} \qquad \text{(B.3)}$$

plus terms involving a product of W and an odd number of terms in u_1 which will have an expected value of zero. Noting that

$$\frac{-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u_1'\bar{P}_Xu_1}(u_1'WQ_{11}W'u_1)Q \qquad \text{(B.4)}$$

is of stochastic order T^{-3}, we may ignore this term so we focus on

$$E[\frac{-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u_1'\bar{P}_Xu_1}(u_1'u_1\pi Q_{11}\pi'u_1'u_1)Q]. \qquad \text{(B.5)}$$

Evaluating this is simplified by noting that

$$u_1'u_1=u_1'\bar{P}_Xu_1+u_1'P_Xu_1 \qquad \text{(B.6)}$$

where $u_1'P_Xu_1$ is $O_p(1)$.Hence we may replace $u_1'u_1$ with $u_1'\bar{P}_Xu_1$ in the above expectation which will be unchanged to order T^{-2}. So we consider

$$\begin{aligned}
& E[\frac{-u'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u}{u'\bar{P}_Xu}(u_1'u_1\pi Q_{11}\pi' u_1'u_1)Q] \\
= \;& E[-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1(\pi Q_{11}\pi' u_1'\bar{P}_Xu_1)Q]+o(T^{-2}) \\
= \;& -\pi Q_{11}\pi' E(u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1u_1'\bar{P}_Xu_1)Q+o(T^{-2}) \\
= \;& \sigma^2 Ltr(QC_1)Q+o(T^{-2}) \qquad \text{(B.7)}
\end{aligned}$$

This is found by straightforward evaluation of a product of quadratic forms in normal variables and by noting that $-tr(\bar{P}_{\bar{z}_1}-\bar{P}_X)=K-(g+k)=L$, the order of overidentification. Also we use the fact that $\pi Q_{11}\pi'=tr(QC_1)$,see (15).

Hence we have shown that

$$E[\frac{-2u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u_1'\bar{P}_Xu_1}\frac{u_1'V_ZQV_Z'u_1}{T-g-k}Q]=2\sigma^2Ltr(QC_1)Q+o(T^{-2}). \qquad \text{(B.8)}$$

Next we consider $E\left(-\frac{u_1'V_ZQV_Z'u_1}{(T-K)(T-g-k)}Q\right)$. Substituting $V_Z=[W+u_1\pi':0]$ and ignoring terms which involve products of W and u we find that:

$$\begin{aligned}
E\left(-\frac{u_1'V_ZQV_Z'u_1}{(T-K)(T-g-k)}Q\right) &= -E\left(\frac{u_1'u_1\pi' Q_{11}\pi u_1'u_1}{(T-K)(T-g-k)}\right)Q \\
&= -2\sigma^2 tr(QC_1)Q. \qquad \text{(B.9)}
\end{aligned}$$

Gathering terms from (B.8) and (B.9) we have shown that for (i)

$$E[(-\frac{2(1-(\lambda-\frac{1}{T-K}))u_1'V_ZQV_Z'u_1}{T-g-k}Q)]=2\sigma^2(L-1)tr(QC_1)Q+o(T^{-2}). \qquad \text{(B.10)}$$

(ii) An approximation to $E[Q((1-\lambda)V_Z'V_Z+\frac{1}{T-K}V_Z'V_ZQ]$

For (ii) we need to find

$$E[Q((1-\lambda)V_Z'V_Z+\frac{1}{T-K}V_Z'V_ZQ]=E[Q(1-\lambda)V_Z'V_ZQ]+E[Q\frac{1}{T-K}V_Z'V_ZQ]. \qquad \text{(B.11)}$$

Consider the first part $E[Q((1-\lambda)V_Z'V_ZQ]$.Using (B.1) and putting $V_Z=[W+u_1\pi':0]$,we may write

$$E[Q((1-\lambda)V_Z'V_ZQ]=E[\frac{-u_1'(\bar{P}_{\bar{z}_1}-\bar{P}_X)u_1}{u_1'\bar{P}_Xu_1}(QW'WQ+u_1'u_1Q\pi\pi'Q)]+o(T^{-2}) \qquad \text{(B.12)}$$

where the expectations products of W and u_1 have been ignored.

There are two terms to evaluate the first of which is

$$E[\frac{-u_1'(\bar{P}_{\bar{z}_1} - \bar{P}_X)u_1}{u_1'\bar{P}_X u_1}(QW'WQ)] = E(\frac{-u_1'(\bar{P}_{\bar{z}_1} - \bar{P}_X)u_1}{u_1'\bar{P}_X u_1})QE(W'W)Q$$
$$= -LQC_2Q \tag{B.13}$$

and the second is

$$E[\frac{-u_1'(\bar{P}_{\bar{z}_1} - \bar{P}_X)u_1}{u_1'\bar{P}_X u_1}u_1'u_1Q\pi\pi'Q]$$
$$= E[\frac{-u_1'(\bar{P}_{\bar{z}_1} - \bar{P}_X)u_1}{u_1'\bar{P}_X u_1}u_1'\bar{P}_X u_1Q\pi\pi'Q] + o(T^{-2})$$
$$= E[-u_1'(\bar{P}_{\bar{z}_1} - \bar{P}_X)u_1Q\pi\pi'Q] + o(T^{-2})$$
$$= -\sigma^2 LQC_1Q. \tag{B.14}$$

Gathering the two parts above we have shown that

$$E[Q(1-\lambda)V_z'V_zQ] = -LQCQ + o(T^{-2}) \tag{B.15}$$

where $C = \sigma^2 C_1 + C_2$,see (15).

Finally it may be shown directly that

$$E(Q\frac{1}{T-K}V_z'V_zQ) = QCQ + o(T^{-2}) \tag{B.16}$$

We thus have

$$E[Q((1-\lambda)V_z'V_z + \frac{1}{T-K}V_z'V_z)Q] = -(L-1)QCQ + o(T^{-2}). \tag{B.17}$$