UNIVERSITY OF LIVERPOOL | Management School

Working Paper in Economics

 # 202303

The Bias of the Modified Limited Information Maximum Likelihood Estimator (MLIML) in Static Simultaneous Equation Models

Gareth Liu-Evans & Garry DA Phillips

The Bias of the Modified Limited Information Maximum Likelihood Estimator (MLIML) in Static Simultaneous Equation Models

Gareth Liu-Evans & Garry DA Phillips

Keywords: LIML, Modified LIML, 2SLS, bias approximation, bias correction

Abstract

A higher-order approximation is made to the bias of the modified LIML (MLIML) estimator due to Fuller. It is demonstrated via simulation that the asymptotic approximation can be used to reduce estimation bias, including in cases where instrument strength is relatively weak, and that the approximation also mirrors the behaviour of the true bias. It is possible to see via the asymptotic approximation why MLIML estimation bias is often found to be very small in two equation models where the order of overidentification is small, and to predict, in simple models where the approximation is specialised, how the order of overidentification will relate nonlinearly to the bias. An asymptotic approximation is also obtained for the pseudo-bias of the LIML estimator. Finally, the bias-corrected MLIML estimator is used to re-examine the effect on the US college graduate wage premium of shifts in the relative supply of young college workers, following Fortin (2006).

1 Introduction

In simultaneous equation models it is well known that Ordinary Least Squares (OLS) is a biased and inconsistent estimator, and a good deal of research has been conducted to explore the nature of the bias and to develop less biased and consistent estimators. In particular the Two Stage Least Squares ($2SLS$) estimator, which is consistent but still biased in small samples emerged as the natural successor and has been in widespread use over many years. Its small sample properties were explored in the seminal paper by Nagar (1959), which has served to generate a great deal of research. The original Limited Information Maximum Likelihood ($LIML$) estimator, whose unconditional distribution has been studied recently by Giovanni and Jiang (2019), is consistent but does not have finite moments of any order. Anderson et al. (2011) show that it performs relatively well in terms of median bias and approaches normality faster than 2SLS when the number of instruments is large, but as a consequence of the moments issue LIML itself has not been in general use. However a modification of the estimator proposed by Fuller (1977) which we shall refer to as the Modified Limited Information Maximum Likelihood ($MLIML$) estimator is also consistent and has all necessary moments, while it was shown by Fuller to be unbiased to order T^{-1}.[1] Although

[1] We note that there have been other LIML estimators since with finite moments as well besides the MLIML estimator considered here, for example LIML using an alternative normalisation by Anderson (2010) and regularised LIML by Carrasco and Tchuente (2015).

the conventional $LIML$ estimator does not have a finite expectation, it is still possible to find an approximation to its central value, which has the interpretation of being an approximation to the mean of a distribution very close to that of the $LIML$ estimator and so can be interpreted as a pseudo bias; this bias is not zero to order T^{-1}. Hence the $MLIML$ estimator can be said to be less biased than $LIML$ (and, of course, $2SLS$).

In Mikhail (1972), the $2SLS$ bias approximation was extended to include higher order terms and an approximation to order T^{-2} was presented. It was shown that, in the case that the order of overidentification is $L = 1$, the bias is zero to order T^{-2}. However, when $L > 1$ the bias is of order T^{-1} and the higher order terms may be important, as was shown in Liu-Evans and Phillips (2019) so that in such cases bias correction should take account of this. Given that the $2SLS$ bias to order T^{-2} is known there is the opportunity to do this. But there is no possibility to do so for the $MLIML$ or $LIML$ estimators since there is, as yet, no bias approximation to order T^{-2} although there is, in each case, a known bias to order T^{-1}. Of course, having such an approximation is also important more generally; the $MLIML$ estimator, in particular, is seen as an especially important alternative to $2SLS$, so the more we know of its properties the better. The main purpose of this paper is to find approximations to the $MLIML$ and $LIML$ biases to order T^{-2}, which will be compared to the counterpart for $2SLS$.

2 The Simultaneous Equation Model

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

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

in which y_t is a $G \times 1$ vector of endogenous variables, z_t is a $K \times 1$ vector of strongly exogenous variables and u_t is a $G \times 1$ vector of 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

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

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

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

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 will be written as

$$y_1 = Y_2\beta + Z_1\gamma + u_1 \tag{4}$$

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. Z_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 $E(u_1u_1') = \sigma_{11}I_T$.

The reduced form of the system includes

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

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 $K \times (g_1 + 1)$ 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 Z is strongly exogenous and of rank K with limit matrix $\lim_{T\to\infty} T^{-1}Z'Z = \Sigma_{zz}$, which is $K \times K$ positive definite, and (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).

3 Large T-approximations for the bias of k-class Estimators

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

$$\begin{pmatrix} \hat{\beta}_k \\ \hat{\gamma}_k \end{pmatrix} = \begin{pmatrix} Y_2^{'}Y_2 - k\hat{V}_2\hat{V}_2 & Y_2^{'}Z_1 \\ Z_1^{'}Y_2 & Z_1^{'}Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2^{'}y_1 - k\hat{V}_2^{'}y_1 \\ Z_1^{'}y_1 \end{pmatrix} \tag{6}$$

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

$$\left| Y_1^{'}(I - P_{Z_1})Y_1 - \lambda Y_1^{'}(I - P_Z)Y_1 \right| = 0. \tag{7}$$

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), where $k_2 = K - k_1$ is the number of exogenous variables excluded from (4).

We shall find it convenient to rewrite (4) as

$$y_1 = R_1\alpha + u_1 \tag{8}$$

where $R_1 = (Y_2 : Z_1)$ 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 result for estimator bias is given as follows.

If we denote $\hat{\alpha}_k$ as the $k-$class estimator for α in (8) then, defining L as the degree of overidentification, the approximate bias is given by

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

where the degree of overidentification may be defined as

$$L = k_2 - g_1, \tag{10}$$

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

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

$$Q = \left[\begin{array}{cc} \bar{Y}_2'\bar{Y}_2 & \bar{Y}_2'Z_1 \\ Z_1'\bar{Y}_2 & Z_1'Z_1 \end{array} \right]^{-1}. \tag{11}$$

Further, we may write that $V_2 = W^* + u_1\pi'$ where u_1 and $W^* = (W : 0)$ are independent and

$$\frac{1}{T}\left(\begin{array}{c} E(V_2'u_1) \\ 0 \end{array} \right) = \sigma^2 \left(\begin{array}{c} \pi \\ 0 \end{array} \right) = q \tag{12}$$

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

$$C = E[\frac{1}{T}V_Z'V_Z] = \left[\begin{array}{cc} (1/T)E(V_2'V_2) & 0 \\ 0 & 0 \end{array} \right] = C_1 + C_2, \tag{13}$$

where $C_1 = \left[\begin{array}{cc} \sigma^2\pi\pi' & 0 \\ 0 & 0 \end{array} \right] = \sigma^2 qq'$ and $C_2 = \frac{1}{T}E(W^{*\prime}W^*) = \left[\begin{array}{cc} 1/TE(W'W) & 0 \\ 0 & 0 \end{array} \right]$.

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

$$E(\hat{\alpha} - \alpha) = (L-1)Qq + o(T^{-1}). \tag{14}$$

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

$$E(\hat{\alpha} - \alpha) = (L-1)[I + tr(QC)I - (L-2)QC]Qq \ + o(T^{-2}). \tag{15}$$

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 $K - (g_1 + k_1) = k_2 - g_1 = 1$. Finally, in a two-equation model $tr(QC)Qq = QCQq$, so that in this special case the higher order bias term becomes $-(L-1)(L-3)QCQq$; hence the higher order term also vanishes for $L = 3$ while the corresponding term of $O(T^{-1})$ remains.

The higher order bias approximation for the consistent fixed $k-$class estimator was given by Iglesias and Phillips (2008) as

$$\begin{aligned} E(\hat{\alpha}_k - \alpha) =& \left(L - 1 - \theta + \theta\frac{K}{T} \right) Qq + (L - 1 - 2\theta)tr(QC)Qq \\ & - [(L-1)(L-2) - \theta(2(L-2) - \theta)]QCQq + o(T^{-2}) \end{aligned}$$

It is seen that if θ is chosen equal to $L-1$ in the $k-$class bias approximation in (16), the bias disappears to order T^{-1} (though not to order T^{-2}). Hence when $k = 1 + \frac{L-1}{T}$ we have Nagar's unbiased estimator.

4 The Modified Limited Information Maximum Likelihood Estimator

A modification of the LIML estimator, which we call the Modified Limited Information Maximum Likelihood (MLIML) Estimator, was introduced by Fuller (1977). First note that from (6) the LIML estimator may be written in the form of a k-class estimator where k is stochastic as follows:

$$\begin{pmatrix} \hat{\beta}_{LIML} \\ \hat{\gamma}_{LIML} \end{pmatrix} = \begin{pmatrix} Y_2^{'}Y_2 - \lambda\hat{V}_2^{\prime}\hat{V}_2 & Y_2^{'}Z_1 \\ Z_1^{\prime}Y_2 & Z_1^{'}Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2^{'}y_1 - \lambda\hat{V}_2^{'}y_1 \\ Z_1^{'}y_1 \end{pmatrix} \tag{16}$$

but the 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. The estimator has (at least) finite first and second moments. Hence the $MLIML$ estimator is

$$\begin{pmatrix} \hat{\beta}_F \\ \hat{\gamma}_F \end{pmatrix} = \begin{pmatrix} Y_2^{'}Y_2 - (\lambda - \frac{\alpha}{T-K})\hat{V}_2^{\prime}\hat{V}_2 & Y_2^{'}Z_1 \\ Z_1^{'}Y_2 & Z_1^{'}Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2^{'}y_1 - (\lambda - \frac{\alpha}{T-K})\hat{V}_2^{'}y_1 \\ Z_1^{'}y_1 \end{pmatrix} \tag{17}$$

When $\alpha = 1$ is chosen, 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 studies have found that $MLIML$ may have good finite sample properties. Hahn, J., Hausman, J. & Kuersteiner, G. (2004) in particular suggest "that the Fuller estimator receive more attention and use than it seems to have received to date", and favourable Monte Carlo results are presented in Flores-Lagunes (2007).

It has been shown that the estimator has a relatively small bias when $\alpha = 1$, see Fuller (1977), where the bias is $o(T^{-1})$, however in a number of Monte Carlo experiments the $MLIML$ bias has been so small absolutely as to suggest that it may be of even smaller order, see, for example, Phillips G.D.A. and Yongdeng Xu (2017). In this paper an expression is found for the second order bias of the $MLIML$ estimator and, while the bias is not zero to $O(T^{-2})$, conditions are found under which the bias is very small, thus explaining the fact that very small biases are sometimes found in simulation studies.

5 Fuller Expansion to order T^{-2}

Applying the $Nagar$ expansion approach to the above (19) yields an asymptotic expansion of the estimation error

$$\begin{aligned} e_F &= \left[Q^{-1} + (X'V_Z + V_Z'X) + (1 - \lambda + \frac{1}{T-K})V_Z'(I - M^*)V_Z + V_Z'M^*V_Z\right]^{-1} \\ &\quad \times \left[X'u + (1 - \lambda + \frac{1}{T-K})V_Z'(I - M^*)u + V_Z'M^*u\right] \\ &= \left[I + Q\{(X'V_Z + V_Z'X + (1 - \lambda + \frac{1}{T-K})V_Z'(I - M^*)V_Z + V_Z'M^*V)\}\right]^{-1} Q \\ &\quad \times \left[(X'u + (1 - \lambda + \frac{1}{T-K})V_Z'(I - M^*)u + V_Z'M^*u)\right] \end{aligned}$$

where $M^* = Z(Z'Z)^{-1}Z'$. To order T^{-2} this yields

$$\begin{aligned}
e_F = {} & QX'u + QV_Z'M^*u + Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\
& - Q(X'V_Z+V_Z'X)QX'u - Q(X'V_Z+V_Z'X)QV_Z'M^*u \\
& - Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\
& - QV_Z'M^*V_ZQX'u - Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQX'u \\
& + Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)QX'u - QV_Z'M^*V_ZQV_Z'M^*u \\
& - QV_Z'M^*V_ZQ(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\
& - Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQV_Z'M^*u \\
& - Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQ(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\
& + Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)QV_Z'M^*u \\
& + Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\
& + QV_Z'M^*V_ZQ(X'V_Z+V_Z'X)QX'u + Q(X'V_Z+V_Z'X)QV_Z'M^*V_ZQX'u \\
& + Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)QX'u \\
& + Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQX'u \\
& + Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQ(X'V_Z+V_Z'X)QX'u + o_p(T^{-2}) \qquad (18)
\end{aligned}$$

This incorporates the expansion for the $2SLS$ estimator plus additional terms which involve (1-$\lambda + \frac{1}{T-K}$) and which represent the difference between the $2SLS$ and the $MLIML$ expansions to $O_p(T^{-2})$. Thus e_F may be written as

$$\begin{aligned} e_F &= e_1 + Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\ &-Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\ &-Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQX'u \\ &-QV_Z'M^*V_ZQ(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\ &-Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQV_Z'M^*u \\ &-Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQ(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\ &+Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\ &+Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQX'u \\ &+Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQ(X'V_Z+V_Z'X)QX'u+o_p(T^{-2}). \end{aligned} \quad (19)$$

where e_1 is the corresponding expansion for $2SLS$.

The bias approximation to order T^{-2} for $2SLS$ has already been found so to find the corresponding result for the Fuller estimator we shall need to evaluate the expectations of the nine additional terms to order T^{-2}.

Since the assumption is that the disturbances are normally distributed, terms involving a product of an odd number of normally distributed disturbances will have a zero expectation.

Examining the above we see that the orders of the nine additional terms are as follows:

1. $Q(1\text{-}\lambda+\frac{1}{T-K})V_Z'(I-M^*)u$ is $O_p(T^{-1})$ and has expectation $-(L-1)Qq+o(T^{-1})$. It does not have terms of higher order T^{-2}.
2. $-Q(1\text{-}\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQX'u$ is $O_p(T^{-\frac{3}{2}})$. It has expectation zero and does not have terms of higher order T^{-2}.
3. $-Q(X'V_Z+V_Z'X)Q(1\text{-}\lambda+\frac{1}{T-K})V_Z'(I-M^*)u$ is $O_p(T^{-\frac{3}{2}})$. Again it has expectation zero with no terms of higher order T^{-2}.

Thus terms 2. and 3. will not play a part in the approximation while the remaining terms 4. to 9. are all $O_p(T^{-2})$ and all have a role.

In the Appendix we have also evaluated the expectations of these terms as follows.

4. $E[-QV_Z'M^*V_ZQ(1\text{-}\lambda+\frac{1}{T-K})V_Z'(I-M^*)u]$

$$= K(L-1)QCQq + 2LQC_1Qq + o(T^{-2})$$

5. $E[-Q(1\text{-}\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQV_Z'M^*u]$

$$= K(L-1)QCQq + 2LQCQq + o(T^{-2})$$

6. $E[-Q(1\text{-}\lambda + \frac{1}{T-K})V_Z'(I - M^*)V_Z Q(1\text{-}\lambda + \frac{1}{T-K})V_Z'(I - M^*)u]$

$$= -(L^2 + 1)QCQq + o(T^{-2})$$

7. $E[+Q(X'V_Z + V_Z'X)Q(X'V_Z + V_Z'X)Q(1\text{-}\lambda + \frac{1}{T-K})V_Z'(I - M^*)u]$

$$= -(L-1)(k+g+2)QCQq - (L-1)trQC.Qq + o(T^{-2})$$

8. $E[+Q(X'V_Z + V_Z'X)Q(1\text{-}\lambda + \frac{1}{T-K})V_Z'(I - M^*)V_Z QX'u]$

$$= -(L-1)[QCQq + trQC.Qq] + o(T^{-2})$$

9. $E[+Q(1 - \lambda + \frac{1}{T-K})V_Z'(I - M^*)V_Z Q(X'V_Z + V_Z'X)QX'u]$

$$= -(k+g+1)(L-1)QCQq + o(T^{-2})$$

Gathering terms we find that the sum of 4. to 9 to order T^{-2} is

$$\begin{aligned}
&(2(L-1)(K-k-g-1) - (L-1)^2 - 2(L-1))QCQq - 2(L-1)trQC.Qq + 2LQC_1Qq \\
&\quad = ((L-1)^2 - 2(L-1))QCQq - 2(L-1)trQC.Qq + 2LQC_1Qq \\
&\quad = ((L-1)(L-3))QCQq - 2(L-1)trQC.Qq + 2LQC_1Qq \\
&\quad = (L-1)(L-3)QCQq - 2(L-1)trQC.Qq + 2LQC_1Qq.
\end{aligned}$$

To find the bias approximation of the Fuller estimator to $O(T^{-2})$ we need to add the higher order terms from 2SLS,

$$(L-1)trQC.Qq - (L-1)(L-2)QCQq,$$

which then gives the result in Theorem 1.

Theorem 1. *The bias of the Fuller MLIML estimator is*

$$E(e_F) = -(L-1)QCQq - (L-1)trQC.Qq + 2LQC_1Qq + o(T^{-2}) \tag{20}$$

In a two equation model $QCQq = trQC.Qq$, see Hadri and Phillips (1999), so that the bias approximation then reduces to

$$E(e_F) = -2(L-1)QCQq + 2LQC_1Qq = 2QCQq - 2LQC_2Qq + o(T^{-2}) \tag{21}$$

which does not vanish when $L = 1$. It then becomes $2QC_1Qq + o(T^{-2})$. In fact this result holds when $L = 1$ for any number of equations in the model.

5.1 The bias of the LIML estimator

It is of interest to compare this result to the corresponding approximation for $LIML$ which is obtained when $k = 1 - \lambda$. The required analysis is straightforward and proceeds from noting that $E(1-\lambda) = -\frac{L}{T-K} + o(T^{-1})$. The bias approximation is then found as follows.

For the $LIML$ case, the corresponding first three terms are:

1(a) $Q(1\text{-}\lambda)V_Z'(I-M^*)u$ is $O_p(T^{-1})$ whereas

$$E(Q(1-\lambda)V_Z'(I-M^*)u) = -LQq + o(T^{-2}).$$

[*Note:* the corresponding term for the $MLIML$ estimator replaces L with $L-1$ and we shall find this happens with all the terms 4(a)-9(a) below.]

2(a) $-Q(1\text{-}\lambda)V_Z'(I-M^*)V_ZQX'u$ is $O_p(T^{-\frac{3}{2}})$

3(a) $-Q(X'V_Z+V_Z'X)Q(1\text{-}\lambda)V_Z'(I-M^*)u$ is $O_p(T^{-\frac{3}{2}})$.

Both terms in 2(a) and 3(a) are shown to have expectation zero to order T^{-2} and so they play no part in the bias approximation. We shall, however, evaluate the remaining terms from the results in Appendix 1 for $MLIML$ given the close relationship between $LIML$ and $MLIML$. Thus we have:

4(a) $E[-QV_Z'M^*V_ZQ(1\text{-}\lambda)V_Z'(I-M^*)u]$

$$= KLQCQq + 2LQC_1Qq + o(T^{-2})$$

5(a) $-Q(1\text{-}\lambda)V_Z'(I-M^*)V_ZQV_Z'M^*u$

$$= KLQCQq + 2LQCQq + o(T^{-2})$$

6(a) $-Q(1\text{-}\lambda)V_Z'(I-M^*)V_ZQ(1\text{-}\lambda)V_Z'(I-M^*)u$

$$= -L^2QCQq + o(T^{-2})$$

7(a) $+Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)Q(1\text{-}\lambda)V_Z'(I-M^*)u$

$$= -L(k+g+2)QCQq - LtrQC.Qq + o(T^{-2})$$

8(a) $+Q(X'V_Z+V_Z'X)Q(1\text{-}\lambda)V_Z'(I-M^*)V_ZQX'u$

$$= -L[QCQq + trQC.Qq] + o(T^{-2})$$

9(a) $+Q(1\text{-}\lambda)V_Z'(I-M^*)V_ZQ(X'V_Z+V_Z'X)QX'u$

$$= -(k+g+1)LQCQq$$

Adding the terms 4(a) to 9(a) we have

$$\begin{aligned} &2L^2QCQq - 3LQCQq + 2LQCQq + 2LQC_1Qq - L^2QCQq - LQCQq - 2LtrQC.Qq \\ &= L^2QCQq - 2LQCQq - 2LtrQC.Qq + 2LQC_1Qq \\ &= L(L-2)QCQq - 2LtrQC.Qq + 2LQC_1Qq \end{aligned} \tag{22}$$

Adding to this the 2SLS approximation higher order terms, $(L-1)trQC.Qq-(L-1)(L-2)QCQq$, as well as the difference in the $O(T^{-1})$ terms, yields the $LIML$ higher order bias as follows:

Theorem 2. *The bias of the LIML estimator is*

$$E(\alpha_{LIML}) = -Qq - (L+1)trQC.Qq + (L-2)QCQq + 2LQC_1Qq + o(T^{-2}) \tag{23}$$

In the case of the two equation model, $trQC.Qq = QCQq$ and then the higher order part of the bias approximation becomes

$$-3QCQq + 2LQC_1Qq$$

So finally the bias of the $LIML$ estimator in a two equation model, to $O(T^{-2})$, is

$$E(e_{LIML}) = -Qq - 3QCQq + 2LQC_1Qq + o(T^{-2}) \tag{24}$$

which is likely to exceed the Fuller bias of $2QCQq - 2LQC_1Qq$ in absolute terms because of the presence of the term of order $O(T^{-1})$.

Interestingly the higher order part of the bias of $LIML$ is seen to be close to the negative of that of $MLIML$.

Lemma 1. *Suppose that A_T, B_T are each of order one, i.e. $O_p(1)$, while $A_T = E(A_T) + (A_T - E(A_T))$ where $(A_T - E(A_T)$ is $O_p(T^{-\frac{1}{2}})$ and, similarly, $B_T = E(B_T) + (B_T - E(B_T))$, where the expectations exist, then it follows that*

$$E(A_T B_T) = E(A_T)E(B_T) + o(T^{-\frac{1}{2}})$$

Clearly this is easily generalised.

Good use will be made of Lemma 1 in the subsequent analysis of the bias approximation that appears in the Appendix.

6 Numerical and Simulation Results

The bias of the MLIML estimator with $\alpha = 1$ is illustrated here and compared with the bias of an analytically corrected MLIML, which uses the $O(T^2)$ approximation in (21) and an initial MLIML estimation. We then consider the special case of $L = 1$, where the bias approximation is able to predict further aspects of the bias. All numerical results are for $T = 100$, and 100000 replications are used for the Monte Carlo.

The bias of the MLIML estimator is explored thoughout in a simple simultaneous equation model:

$$y_{1,t} = \beta_1 y_{2,t} + u_{1,t}, \tag{25}$$

$$y_{2,t} = \beta_2 y_{1,t} + \gamma' z_t + u_{2,t} \tag{26}$$

for $t = 1, 2, \ldots T$, where z_t is a $p \times 1$ vector of exogenous variables, and where the interest is in estimation of β_1. A similar model has been considered in Hahn, Hausman and Kuersteiner (2004) and others, most recently in Liu-Evans and Phillips (2018) for the 2SLS estimator. As in the latter, fixed exogenous data for each element z_{jt} of z_t, $j = 2, 3, \ldots,$ was drawn from an AR(1) model $z_{jt} = 0.9 z_{j,t-1} + \nu_t$ with $\nu_t \overset{i.i.d.}{\sim} N(0, 1)$, while z_{1t} was a constant for all t. The structural disturbances $(u_{1,t}, u_{2,t})'$ were jointly Normally distributed with mean 0 and covariance matrix Σ. As noted in the former, the MLIML estimator for this two-equation model may be written as

$$\hat{\beta}_{1,\text{MLIML}} = \frac{y_2' P y_1 - (\lambda - \frac{\alpha}{T-p}) y_2' M y_1}{y_2' P y_2 - (\lambda - \frac{\alpha}{T-p}) y_2' M y_2}.$$

The analytical bias correction using (22) requires estimates of Q, q, C and C_1. From (10), Q reduces to the scalar $(\bar{Y}_2' \bar{Y}_2)^{-1}$ where $\bar{Y}_2 = Z\Pi_2$, and we estimate the reduced form parameters Π_2 by ordinary least squares, yielding

$$\hat{Q} = (\hat{\Pi}_2' Z' Z \hat{\Pi}_2)^{-1}. \tag{27}$$

The following were used for the other terms using similar reductions:

$$\hat{q} = \frac{1}{T} \hat{V}_{2,OLS}' \hat{u}_{1,MLIML} \tag{28}$$

$$\hat{C} = \frac{1}{T} \hat{V}_{2,OLS}' \hat{V}_{2,OLS} \tag{29}$$

$$\hat{C}_1 = \frac{T}{\hat{u}_{1,MLIML}' \hat{u}_{1,MLIML}} \hat{q} \hat{q}' \tag{30}$$

Table 1 presents Monte Carlo results for the performance of the analytically corrected MLIML estimator vs the uncorrected MLIML estimator. This is for the case $\alpha = 1$, where the approximation has been developed in Section 5. The results in the table correspond to the following collection of models, which includes models with varying degrees of overidentification L and varying instrument strength:

Model Collection 1 ($L = 3, 4, 5, 6$ and various instrument strength)

$$\beta_1 = 2.73,\ \beta_2 = -16.39, \gamma = (12.00, 12.00, c\iota')), \dim(\iota) \in \{2, 3, 4, 5\},$$

$$\Sigma = \begin{pmatrix} 38.11 & -11.78 \\ -11.78 & 92.11 \end{pmatrix}, c \in \{0.1, 4, 5\}$$

The number of instruments considered ranges from four to seven, so that the degree of overidentification L ranges from 3 to 6. The structural coefficient on each additional instrument, c in each case, ranges from 0.1 to 5, with lower values of c corresponding to lower instrument strength. As a measure of overall instrument strength, the table reports the expected R^2 from a regression of the endogenous variable y_{2t} on the instruments in z_t, which is denoted by $E[\hat{\rho}]$. Some experimentation was required to find parameterisations where the overall instrument strength would not be too weak, but where including additional instruments would not increase the overall instrument strength too quickly.

It can be seen from the table that the bias correction works quite well in all cases, with a small or moderate increase in root mean squared error. The results in the final four rows, corresponding to cases with weaker instruments where $c = 0.1$ for $L = 3, 4, 5, 6$, illustrate that there is still a bias correction but that this is accompanied by a larger increase in RMSE. The extent of the bias correction appears to decline as L rises, and the root mean squared error of the corrected MLIML estimator becomes more similar to the original estimator. It was found in other simulation experiments not reported here that this pattern continues for larger L, with the bias and RMSE of the corrected estimator becoming increasingly similar to the original estimator.

Table 1: $T = 100$ MLIML and corrected, Model Collection 1

			MLIML		Bias-corrected MLIML		$O(T^{-2})$ approx.
	L	$E[\hat{\rho}]$	% Bias	RMSE	% Bias	RMSE	% Bias
$c = 4$	3	0.125	-3.09	0.68	-0.73	0.86	-1.32
	4	0.127	-4.57	0.69	-2.45	0.80	-1.79
	5	0.130	-5.83	0.69	-4.20	0.75	-2.18
	6	0.160	-3.21	0.68	-2.41	0.71	-1.34
$c = 5$	3	0.135	-2.42	0.67	-0.50	0.80	-1.20
	4	0.135	-3.71	0.68	-1.87	0.79	-1.53
	5	0.137	-4.82	0.69	-3.35	0.74	-1.86
	6	0.173	-2.17	0.67	-1.54	0.69	-1.00
$c = 0.1$	3	0.095	-6.96	0.69	-2.25	1.30	-2.55
	4	0.104	-6.98	0.70	-4.11	0.96	-2.56
	5	0.114	-6.99	0.70	-5.16	0.77	-2.56
	6	0.123	-6.98	0.70	-5.77	0.73	-2.52

The table presents the bias and RMSE for the MLIML estimator and an analytically corrected version. The corrected estimator uses estimates of the parameters via an initial MLIML estimation, as detailed in this section. The final column presents the $O(T^{-2})$ approximate bias, which is computed using true parameter values. The scalar c is given in Model Collection 1. $E[\hat{\rho}]$ is the expected R^2 from a regression of the endogenous variable y_{2t} on the instruments in z_t.

The MLIML bias appears to decrease in magnitude across c for each value of L, and this is expected as the overall instrument strength increases with c. More curiously, for each c the bias increases over $L = 3, 4, 5$ then decreases at $L = 6$. The approximate bias is conservative but mirrors this pattern. To understand the pattern further, note that for the simple model in this section the approximate MLIML bias in (21) can be reduced to the following:

$$-2LQC_2Qq + 2QCQq = 2(C - LC_2)Q^2q$$

with

$$Q = \frac{(1-\beta_1\beta_2)^2}{\gamma' Z' Z' \gamma}$$

If we assume that the coefficients in γ are all the same and equal to c, then

$$\gamma' Z' Z' \gamma = c^2 \iota_{k_2}' Z' Z \iota_{k_2}$$

where ι_{k_2} is a $k_2 \times 1$ vector of ones. We note that $Z\iota_{k_2} = k_2\tilde{c}_z$ with the $T \times 1$ vector $\tilde{c}_z = \sum_{j=1}^{k_2} z_j/k_2$ and write $\gamma' Z' Z' \gamma = c^2 k_2^2 \tilde{c}_z' \tilde{c}_z$. Further, with $\tilde{c}_{z^2} = \tilde{c}_z'\tilde{c}_z/T$ this leads to

$$\gamma' Z' Z' \gamma = c^2 k_2^2 \tilde{c}_{z^2} T$$

and the bias can be written

$$2(C - LC_2)Q^2q = 2(C - LC_2)q \times \frac{(1-\beta_1\beta_2)^4}{c^4(L+1)^4\tilde{c}_{z^2}^2 T^2} \tag{31}$$

using $k_2 = L + 1$. It can be seen from this that the effect on the bias of increasing L will eventually be dominated by the denominator, but that for relatively low L the magnitude of the bias can increase with L, for example if $C \approx C_2$ with C and C_2 sufficiently large. It can also be seen that the bias falls quickly with c for each order of overidentification L.

Table 2 presents simulation and numerical results for $L = 1$ cases. From Section 5 we know that MLIML still has a bias to order $O(T^{-2})$ in the case where $L = 1$, unlike 2SLS, though from the above we would still anticipate the numerical values for the bias being relatively small. In order to investigate further the ability of the bias approximation to predict the behaviour of the bias in special cases, we consider models in the following form for a range of values of c with Σ as earlier:

Model Collection 2 ($L = 1$, equal coefficients, various instrument strength)

$$\beta_1 = 2.73,\ \beta_2 = -16.39, \gamma = (c_1, c_2)), c_1 = c_2 \text{ for } c \in \{10, 12, 14, 16, 18, 20\}$$

The bias approximation in the $L = 1$ case reduces to

$$\begin{aligned} 2(C - LC_2)q \times \frac{(1-\beta_1\beta_2)^4}{c^4(L+1)^4\tilde{c}_{z^2}^2 T^2} &= C_1 q \times \frac{(1-\beta_1\beta_2)^4}{8c^4\tilde{c}_{z^2}^2 T^2} \\ &= \sigma^2 q q' q \times \frac{(1-\beta_1\beta_2)^4}{8c^4\tilde{c}_{z^2}^2 T^2} \\ &= (E[v_{2t}u_t])^3 \times \frac{(1-\beta_1\beta_2)^4}{8c^4\sigma^4\tilde{c}_{z^2}^2 T^2} \end{aligned} \tag{32}$$

and therefore the sign of the bias should be equal to the sign of $E[v_{2t}u_t]$. It can be seen that the biases in Table 2 are negative, coinciding with $E[v_{2t}u_t]$ being negative: $E[v_{2t}u_t] = E[e_2'B^{-1}u_t u_{1t}] = e_2'B^{-1}\begin{pmatrix}\sigma^2 \\ \sigma_{12}\end{pmatrix} = -13.90$, where $B = \begin{pmatrix}1 & -2.73 \\ 16.39 & 1\end{pmatrix}$, $\sigma^2 = 38.11$ and $\sigma_{12} = -11.78$. Moreover, the bias values should be approximately the same for positive and negative c, which is seen to be the case in the table where we compare columns 1 and 2. If z_t is scaled to $2^{-\frac{1}{4}}z_t$, the bias should also approximately double via the term $\tilde{c}^2_{z^2}$, and this can also be seen in the table.

Table 2: $T = 100$ MLIML bias, further investigation, Model Collection 2 ($L = 1$ cases)

	(Original z_t) % Bias			$2^{-\frac{1}{4}} \times z_t$ % Bias		
$\lvert c\rvert$	$c > 0$	$c < 0$	$E[\hat{\rho}]$	$c > 0$	$c < 0$	$E[\hat{\rho}]$
10	-14.48	-14.49	0.057	-24.58	-24.52	0.043
12	-6.97	-7.06	0.075	-14.03	-14.04	0.058
14	-3.29	-3.40	0.097	-7.60	-7.68	0.074
16	-1.62	-1.72	0.121	-4.03	-4.14	0.091
18	-0.84	-0.95	0.147	-2.18	-2.29	0.110
20	-0.47	-0.58	0.173	-1.23	-1.33	0.131

The scalar c here is, as described in Model Collection 2, the coefficient on each element of the 2×1 vector z_t. For each $|c|$ the values in columns $c > 0$ and $c < 0$ are for the bias where c is positive and negative, respectively. $E[\hat{\rho}]$ was computed using $c > 0$. For the final three columns, the exercise is repeated where z is replaced by $2^{-\frac{1}{4}} \times z_t$ in the data generation and estimation, where it is predicted by (27) that the bias doubles.

7 Application: college wage premia, Fortin (2006)

We use the higher-order MLIML and LIML bias approximations obtained in Section 5 to re-examine the effect on the US college graduate wage premium, originally estimated in Fortin (2006), of shifting the relative supply of young college workers. A similar exercise was carried out in Liu-Evans and Phillips (2018) in the context of 2SLS and allowing for the effect of asymmetric disturbances on the bias, though the estimated skewnesses were small in the case we consider here. Moreover, the estimated 2SLS biases were large in one case, making it interesting to revisit using the low bias ($\alpha = 1$) MLIML estimator.

Our 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:

$$r_{st} = \alpha_0 + \alpha_1 q_{st} + \alpha_2 q_s t^O + \alpha_3' Y_{st} + S_s + P_t + \varepsilon_{st}$$

where $r_{st} = ln(w^Y_{cst}/w^Y_{hst})$ is the college-high school wage gap for young workers, $q_{st} = ln(C^Y_{st}/H^Y_{st})$ is the relative supply of young workers with college education to those without, q^O_{st} 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 coefficient α_1 reflects the effect on the wage premium of shifting the relative supply of young college workers.

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, and 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 first column of Table 1 presents the estimates obtained by feasible weighted 2SLS (Fortin, 2006), MLIML, LIML and 2SLS, while the second column presents the corrected estimates. This is done for two different samples, one corresponding to US states with relatively low enrollment in private colleges, and one with high enrollment where the state educational policies under consideration do not apply. The estimated biases for MLIML are lower in both cases than for 2SLS, but at around 6% the estimated bias is still substantial for the High enrollment sample.

Table 3: Estimation of α_1, and bias estimates

Private Enrollment		Estimate	Corrected estimate
Low ($N = 217$)	W2SLS (Fortin, 2006)	-0.22	n/a
	MLIML	-0.318	-0.317
	2SLS (LEP, 2018)	-0.134	-0.143
	LIML	-0.365	-0.348
High ($N = 126$)	W2SLS (Fortin, 2006)	0.11	n/a
	MLIML	-0.0215	-0.0229
	2SLS (LEP, 2018)	-0.0259	-0.0172
	LIML	-0.0141	-0.0230

Feasible Weighted 2SLS results are due to Fortin (2006), see in particular Panel C of Table 8. 2SLS and bias-corrected 2SLS results are due to Liu-Evans and Phillips (2018).

8 Conclusions

There has been a resurgence of interest in LIML related estimation approaches, and it is now fairly common to see applications of the MLIML or "Fuller" estimator in empirical work, particularly when the model may be weakly identified or where there are many instruments. While it is well known that the MLIML bias is zero to order $O(T^{-1})$, there has not been an expression for the higher-order bias until now. A Nagar expansion of the MLIML estimation error can obtained in a similar way to other k-class estimators, but the calculation of the higher-order bias for MLIML is complicated by k being stochastic and driven by the smallest root of the LIML derminantal equation. With the higher-order analytical bias approximation that is obtained, though, it is possible to predict the behaviour of the bias well, and to suggest why very small biases are sometimes found in Monte Carlo studies. Finally, it is shown that the higher-order bias approximation can be used to achieve practical bias corrections.

Appendix 1

Theorem 1

In this appendix we use the result in Lemma 1 in evaluating the terms 4-9 in section which form the higher order bias of the $MLIML$ estimator.

To proceed we consider the expectation of the term in 4 given by

$$\begin{aligned} &E(-QV_Z'M^*V_ZQ(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u) \\ &\quad = E[-QV_Z'M^*V_ZQ(1-\lambda+\frac{1}{T-K})E(V_Z'(I-M^*)u)] + o_p(T^{-2}), \end{aligned}$$

which, using Lemma 1, is equal to $-(T-K)E(QV_Z'M^*V_ZQ(1\text{-}\lambda+\frac{1}{T-K}))q$ where $E(V_Z'(I-M^*)u)$ is replaced by $(T-K)q$.

Hence we now need to evaluate $E(QV_Z'M^*V_ZQ(1\text{-}\lambda+\frac{1}{T-K}))$ to order T^{-3} or effectively, $E(V_Z'M^*V_Z(1\text{-}\lambda+\frac{1}{T-K}))$ to order T^{-1} where

$$(1-\lambda+\frac{1}{T-K}) = \frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + \frac{1}{T-K} + o_p(T^{-1}).$$

Noting that $V_Z'M^*V_Z = V_Z'P_ZV_Z = W^{*\prime}P_ZW^* + qu'P_Zuq' + o_p(1)$, we need to find

$$E\left(W^{*\prime}P_ZW^*\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + \frac{1}{T-K}\right) + E\left(qu'P_Zuq'\left(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + \frac{1}{T-K}\right)\right)$$

where W is independent of u.

We find

$$\begin{aligned} E(W^{*\prime}P_ZW^*(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + \frac{1}{T-K})) &= E(W^{*\prime}P_ZW^*)E(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + \frac{1}{T-K}) \\ &= \frac{-(L-1)}{T-K}KC_2 \end{aligned}$$

$$\begin{aligned} E(qu'P_Zuq'(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + \frac{1}{T-K})) &= qq'E[(u'P_Zu)(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + \frac{1}{T-K})) \\ &= \frac{-(L-1)}{T-K}KC_1 - \frac{2LC_1}{T-K}. \end{aligned}$$

Hence

$$\begin{aligned} &(T-K)E(-QV_Z'M^*V_ZQ(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u) \\ &\quad = (L-1)KQ(C_1+C_2)Qq + 2LQC_1Qq + o(T^{-2}). \end{aligned}$$

Thus 4. is equal to

$$(L-1)KQCQq + 2LQC_1Qq + o(T^{-2})$$

The term in 5. is

$$-Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQV_Z'M^*u = -Q(1-\lambda+\frac{1}{T-K})(T-K)CQV_Z'M^*u + o_p(T^{-2})$$

where, with $E(V_Z'(I-M^*)V_Z) = (T-K)C$,

$$E(-Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQV_Z'M^*u) = E(-(T-K)QC(1-\lambda+\frac{1}{T-K})QV_Z'M^*u) + o(T^{-2})$$

so that the above, on introducing ( ), finally becomes

$$-(T-K)KQC\frac{-(L-1)}{T-K}Qq + 2LQCQq + o(T^{-2}).$$

Hence for 5. we have

$$(L-1)KQCQq + 2LQCQq + o(T^{-2})$$

Next we shall evaluate 6., in particular

$$\begin{aligned} &E[-Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQ(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u] \\ &\quad = E[-Q(1-\lambda+\frac{1}{T-K})(T-K)CQ(1-\lambda+\frac{1}{T-K})(T-K)q + o(T^{-2}) \\ &\quad = -(T-K)^2QCQqE(1-\lambda+\frac{1}{T-K})^2] + o(T^{-2}) \end{aligned}$$

where $$\begin{aligned} E((1-\lambda+\frac{1}{T-K})^2) &= E[(1-\lambda)^2 + 2\frac{(1-\lambda)}{T-K} + (\frac{1}{T-K})^2] \\ &= E[(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u})^2 + \frac{2}{T-K}\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u} + (\frac{1}{T-K})^2]. \end{aligned}$$

So we consider $E(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u})^2$ where $u'(\bar{P}_Z-\bar{P}_X)u$ is independent of $u'(\bar{P}_Z)u$ and $u'(\bar{P}_Z-\bar{P}_X)u = -u\prime(P_Z-P_X)u = u\prime(P_X-P_Z)u$.

By direct evaluation $E(u'(P_X-P_Z)u)^2 = L^2+2L$ and $E(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u})^2 = \frac{L^2+2L}{(T-K)^2} + o(T^2)$ while $E(\ \frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}) = \frac{-L}{T-K} + o(T^{-1})$. Hence

$$E((1-\lambda+\frac{1}{T-K})^2) = \frac{L^2+2L-2L+1}{(T-K)^2} = \frac{L^2+1}{(T-K)^2} + o(T^{-2})$$

so that for 6 we may write

$$\begin{aligned} &E\left[-Q\left(1-\lambda+\frac{1}{T-K}\right)V_Z'(I-M^*)V_ZQ\left(1-\lambda+\frac{1}{T-K}\right)V_Z'(I-M^*)u\right] \\ &\qquad = -(L^2+1)QCQq + o(T^{-2}) \end{aligned}$$

Now consider the sum of the two terms 7. and 8. which will be taken together as follows:

$$
\begin{aligned}
&+Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQX'u \\
&+Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQ(X'V_Z+V_Z'X)QX'u \\
&\quad=+Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})(T-K)CQX'u \\
&\qquad+Q(1-\lambda+\frac{1}{T-K})(T-K)CQ(X'V_Z+V_Z'X)QX'u+o_p(T^{-2}) \\
&\quad=(T-K)Q(X'V_Z+V_Z'X)QCQ(1-\lambda+\frac{1}{T-K})X'u \\
&\qquad+(T-K)QCQ(1-\lambda+\frac{1}{T-K})(X'V_Z+V_Z'X)QX'u+o_p(T^{-2}) \\
&\quad=\Big[(T-K)Q(X'(W^*+uq')+(W'^*+qu')X)QCQ((\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})X'u+o_p(T^{-2})\Big] \\
&\qquad+\Big[(T-K)QCQ(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})(X'(W^*+uq')+(W'^*+qu')X)QX'u+o_p(T^{-2})\Big] \\
&\quad=\Big[(T-K)QX'uq'QCQ((\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})X'u \\
&\qquad\qquad+(T-K)Qqu'XQCQ((\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})X'u+o_p(T^{-2})\Big] \\
&\qquad+\Big[(T-K)QCQ(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})X'uq'QX'u \\
&\qquad\qquad+(T-K)QCQ(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})qu'XQX'u+o_p(T^{-2})\Big]
\end{aligned}
$$

The terms may be written as

$$
\begin{aligned}
&\Big[(T-K)QX'uu'X\ \ ((\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})QCQq \\
&\qquad +(T-K)Qqu'XQCQ((\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})X'u+o_p(T^{-2})\Big] \\
&+\Big[(T-K)QCQ(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})X'uu'XQq \\
&\qquad +(T-K)QCQ(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})qu'XQX'u+o_p(T^{-2})\Big] \\
=&\Big[(T-K)QX'u(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})u'XQCQq \\
&\qquad +(T-K)Qqu'XQCQX'u\ (\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})+o_p(T^{-2})\Big] \\
&+\Big[(T-K)QCQX'u(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})u'XQq \\
&\qquad +(T-K)QCQq(\frac{u'(\bar{P}_Z-\bar{P}_X)u}{u'(\bar{P}_Z)u}+\frac{1}{T-K})u'XQX'u+o_p(T^{-2})\Big]
\end{aligned}
$$

where now the terms are in a form where evaluation is relatively straightforward.

We have by direct evaluation that the above reduces to

$$
\begin{aligned}
&-(L-1)QCQq-(L-1)trQC.Qq-(L-1)QCQq-(k+g)(L-1)QCQq \\
&\qquad =-(k+g+2)(L-1)QCQq-(L-1)trQC.Qq
\end{aligned}
$$

Hence it has been shown that

$$
\begin{aligned}
(7)+(8)=&+Q(Z'V_Z+V_Z'Z)Q(Z'V_Z+V_Z'Z)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\
&+Q(Z'V_Z+V_Z'Z)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)V_ZQZ'u
\end{aligned}
$$

has expectation

$$-(L-1)(k+g+2)QCQq-(L-1)trQC.Qq+o(T^{-2})$$

Finally we shall evaluate the last term.

$$
\begin{aligned}
(9)=&+Q(X'V_Z+V_Z'X)Q(X'V_Z+V_Z'X)Q(1-\lambda+\frac{1}{T-K})V_Z'(I-M^*)u \\
=&\,Q(X'(W^*+uq')+(W'^*+qu')X)Q(X'(W^*+uq') \\
&\qquad +(W'^*+qu')X)Q(1-\lambda+\frac{1}{T-K})(T-K)q
\end{aligned}
$$

which has 8 components as follows.

(i) $QX'W^*QX'W^*Q$(1-$\lambda+\frac{1}{T-K})(T-K)q$

$(ii)\ QW'^{*}XQX'W^{*}Q(1\text{-}\lambda+\frac{1}{T-K})(T-K)q$
$(iii)\ QX'W^{*}QW'^{*}XQ(1\text{-}\lambda+\frac{1}{T-K})(T-K)q$
$(iv)\ QW'^{*}XQW'^{*}XQ(1\text{-}\lambda+\frac{1}{T-K})(T-K)q$
$(v)\ QX'uq'QX'uq'Q(1\text{-}\lambda+\frac{1}{T-K})(T-K)q$
$(vi)\ QX'uq'Qqu'XQ\ (1\text{-}\lambda+\frac{1}{T-K})(T-K)q$
$(vii)\ Qqu'XQX'uq'Q\ (1\text{-}\lambda+\frac{1}{T-K})(T-K)q$
$(viii)\ Qqu'XQqu'XQ(1\text{-}\lambda+\frac{1}{T-K})(T-K)q$

Noting that W^* is distributed independently of u, the expected values of the above terms are shown to be

(i) $QX'E(W^*QX'W^*)QE((1\text{-}\lambda+\frac{1}{T-K}))(T-K)q = QX'XQC_2Q\frac{-(L-1)}{T-K}(T-K)q$
$= -(L-1)QC_2Qq + o(T^{-2})$

(ii) $QE(W'^*XQX'W^*)QE((1\text{-}\lambda+\frac{1}{T-K}))(T-K)q = -(L-1)(g+k)QC_2Qq + o(T^{-2})$

(iii) $QX'E(W^*QW'^*)XQE((1\text{-}\lambda+\frac{1}{T-K}))(T-K)q = -(L-1)trQC_2.Qq + o(T^{-2})$

(iv) $QE(W'^*XQW'^*)XQE((1\text{-}\lambda+\frac{1}{T-K}))(T-K)q = -(L-1)QC_2Qq + o(T^{-2})$

For the corresponding terms in u we have

(v) $QE(X'(uq'QX'uq'Q(1\text{-}\lambda+\frac{1}{T-K})(T-K)q = QE(X'uu'XQqq'Q(1\text{-}\lambda+\frac{1}{T-K}))(T-K)q$
$= -(L-1)Qqq'Qq = -(L-1)QC_1.Qq + o(T^{-2})$

(vi) $QE(X'uq'Qqu'XQ(1-\lambda+\frac{1}{T-K}))(T-K)q = -(L-1)Qqq'Qq = -(L-1)trQC_1.Qq + o(T^{-2})$

(vii) $QqE(u'XQX'uq'Q(1-\lambda+\frac{1}{T-K}))(T-K)q = -(L-1)(g+k)Qqq'Qq$
$= -(L-1)(g+k)QC_1.Qq + o(T^{-2})$

(viii) $+QqE(u'XQqu'XQ(1\text{-}\lambda+\frac{1}{T-K}))(T-K)q = QqE(q'QX'uu'XQ(1\text{-}\lambda+\frac{1}{T-K}))(T-K)q$
$= -(L-1)QqqQq = -(L-1)QC_1.Qq + o(T^{-2})$

Adding these 8 terms we have for the required expectation:

$$(9) = -(L-1)QCQq - (L-1)(g+k)QCQq - (L-1)trQC.Qq - (L-1)QCQq + o(T^{-2})$$
$$= -(L-1)(g+k+2)QCQq - (L-1)trQC.Qq + o(T^{-2})$$

Adding the terms (4) to (9) yields

$$(L-1)KQCQq + 2LQC_1Qq + (L-1)KQCQq + 2LQCQq - (L^2+1)QCQq - (L-1)(k+g+2)QCQq$$
$$-(L-1)trQC.Qq - (L-1)(g+k+2)QCQq - (L-1)trQC.Qq$$
$$= (L-1)(L-3)QCQq + 2LQC_1Qq - 2(L-1)trQC.Qq$$

Adding to this the 2SLS approximation higher order terms $(L-1)\ trQC.Qq - (L-1)(L-2)QCQq$ we find the $MLIML$ bias as

$$E(e_{MLIML}) = -(L-1)QCQq - (L-1)trQC.Qq + 2LQC_1Qq + o(T^{-2}).$$

In the two equation model $QCQq = trQC.Qq$ and the above becomes

$$-2(L-1)QCQq + 2LQC_1Qq = 2QCQq - 2LQC_2Qq$$

Hence it is found that whereas the bias is non-zero to order T^{-2} in general, it is likely to be close to zero in the special case of a two equation model when L is small, which explains why, in simulations which typically use two-equation models, the bias is often found to be very small.

References

Anderson, T.W. (2010). The LIML estimator has finite moments! *Journal of Econometrics* 157 (2), 359-361.

Anderson, T.W., N. Kunitomo and Y. Matsushita (2011). On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments*Journal of Econometrics* 165, 58-69.

Carrasco, M. and G. Tchuente (2015). Regularized LIML for many instruments. *Journal of Econometrics* 186, 427-442

Davison, R. and J. MacKinnon (1993). Estimation and Inference in Econometrics. Oxford University Press.

Flores-Lagunes, A. (2007). Finite sample evidence of IV estimators under weak instruments. *Journal of Applied Econometrics* 22, 677-694.

Fuller, W.A. (1977). Some Properties of a Modification of the Limited Information Estimator. *Econometrica* 45 (4), 939-953.

Giovanni, F. and B. Jiang (2019). The unconditional distributions of the OLS, 2SLS and LIML estimators in a simple structural equations model. *Econometric Reviews* 38(2), 208-247

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.

Hahn, J., Hausman, J. and G. Kuersteiner (2004). Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *Econometrics Journal* 7, 272-306.

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

G.D.A. Phillips and Y. Xu (2017). Almost Unbiased Variance Estimation in Simultaneous Equation Models. *Working Paper.*

Rothenberg (1983), Asymptotic properties of some estimators in structural models. In "Studies in Econometrics, Time Series, and Multivariate Statistics", Editors S. Karlin, T. Amemiya and L. Goodman. Academic Press, 1983, pages 153-168.