**Cardiff Economics
Working Papers**

Emma M. Iglesias and Garry D.A. Phillips

*Almost Unbiased Estimation in Simultaneous Equations
Models with Strong and / or Weak Instruments*

E2011/19

Almost Unbiased Estimation in Simultaneous Equations Models with Strong and / or Weak Instruments[*]

EMMA M. IGLESIAS and GARRY D.A. PHILLIPS

University of A Coruña Cardiff University

This version: August 2011

Abstract

We propose two simple bias reduction procedures that apply to estimators in a general static simultaneous equation model and which are valid under reatively weak distributional assumptions for the erros.. Standard jackknife estimators, as applied to $2SLS$, may not reduce the bias of the exogenous variable coefficient estimators since the estimator biases are not monotonically non-increasing with sample size (a necessary condition for successful bias reduction) and they have moments only up to the order of overidentification. Our proposed approaches do not have either of these drawbacks. (1) In the first procedure, both endogenous and exogenous variable parameter estimators are unbiased to order T^{-2} and when implemented for $k-$class estimators for which $k < 1$, the higher order moments will exist. (2) An alternative second approach is based on taking linear combinations of $k-$class estimators for $k < 1$. In general, this yields estimators which are unbiased to order T^{-1} and which possess higher moments. We also prove theoretically how the combined $k-$class estimator produces a smaller mean squared error than $2SLS$ when the degree of overidentification of the system is larger than 8. Moreover, the combined $k-$class estimators remain unbiased to order T^{-1} even if there are redundant variables (including weak instruments) in any part of the simultaneous equation system, and we can allow for any number of endogenous variables. The performance of the two procedures is compared

1

with $2SLS$ in a number of Monte Carlo experiments using a simple two equation model. Finally, an application shows the usefulness of our new estimator in practice versus competitor estimators.

Keywords: Combined $k-$class estimators; Bias correction; Weak instruments; Endogenous and exogenous parameter estimators; Permanent Income Hypothesis.

JEL classification: C12; C13; C30; C51, D12, D31, D91, E21, E40.

1 Introduction

Instrumental variable (IV) estimators (see e.g. Hausman (1983), Staiger and Stock (1997), and Hahn and Hausman (2002a, 2002b, 2003)) are known to have very poor finite sample properties, both when the size of the sample is very small, and / or when the instruments are weak (see eg. Nelson and Startz (1990a, 1990b) and Staiger and Stock (1997)). Andrews and Stock (2006) recently provided an extensive and updated review of the literature of simultaneous equations systems and weak instruments.

If the problem is that we simply have a small sample then traditional higher order *"strong IV asymptotics"* may provide an appropriate means of analysis; however, this is not enough in the weak instruments context. Therefore, in the weak instruments literature, in order to analyze the performance of instrumental variable estimators, several different types of asymptotics have been proposed: for example, *"weak IV asymptotics"*, *"many IV asymptotics"* and *"many weak IV asymptotics"*, see Andrews and Stock (2006).

In relation to bias, Richardson and Wu (1971) found the exact bias of the Two Stage Least Squares ($2SLS$) estimator assuming normal errors while Sawa (1972) gave the exact bias for the general $k-$class estimator. More recently, Chao and Swanson (2007) have shown the asymptotic bias of the $2SLS$ estimator using the *"weak IV asymptotics"* of Staiger and Stock (1997), allowing for possibly non-normal errors and stochastic instruments. All these results were derived in the context of an equation containing two endogenous variables. The exact density for instrumental variable estimators in the general case is given in Phillips (1980). However, the resulting moment expressions are complex and we shall not consider them here.

In this paper we are interested in constructing two very simple bias correction procedures that can work in a range of cases, including when we have weak instruments. Recently Davidson and Mackinnon (2006), in a comprehensive Monte Carlo study of a two endogenous variable equation, found little evidence of support for the use of the Jackknife IV estimator introduced by Phillips and Hale (1977), and later used by Angrist, Imbens and Krueger (1999) and Blomquist and Dahlberg (1999), especially in the presence of weak instruments. In addition, they concluded that there is no

clear advantage in using Limited Information Maximum Likelihood ($LIML$) over $2SLS$ in practice, particularly in the context of weak instruments, while estimators which possess moments tend to perform better than those which do not. So there remain questions about the finite sample properties of the standard simultaneous equation estimators especially in the weak instrument context.

For our analysis, we deal with a broad setting where we allow for a general number of endogenous and exogenous variables in the system. We first develop a bias reduction approach that will apply to both endogenous and exogenous variable parameter estimators. We use both large$-T$ asymptotic analysis and exact finite sample results, to propose a procedure to reduce the bias of $k-$class estimators that works particularly well when the system includes weak instruments since, in this case, any efficiency loss is minimized; however, we do not assume that all instruments are weak. We examine the bias of $2SLS$ and also other $k-$class estimators, and we show that the bias to order T^{-2} is eliminated when the number of exogenous variables excluded from the equation of interest is chosen so that the parameters are "notionally" overidentified of order one. This is achieved through the introduction of redundant regressors. While the resulting bias corrected estimators are more dispersed, the increased dispersion will be less the weaker the instruments we have available in the system for use as redundant regressors; hence, the availability of weak instruments is actually a blessing in context. We show in our simulations for sample sizes $T = 100$, that the bias correction is so impressive, that there are some indications that the bias may be of even smaller order.

The above procedure, as applied to $2SLS$, has the main disadvantage that whereas the estimator has a finite bias, the variance does not exist. We therefore consider other $k-$class estimators where we can continue to get a reduction of the bias by using our approach, and where higher order moments will exist. In particular, we consider the estimators where $k = 1 - T^{-3}$ and $k = 1 - T^{-1}$. The former is especially interesting since its behavior is very close to that of $2SLS$ yet higher moments are defined. Note that $LIML$ is known to be median unbiased up order T^{-1} (see Rothenberg (1983)), while with our bias correction mechanism we can get estimators that are unbiased to order T^{-2}.

A second procedure that we analyze is based on a linear combination of $k-$class estimators with $k < 1$ that is unbiased to order T^{-1}, has higher moments, and generally improves on Mean Squared Error (MSE) compared to $2SLS$ when there are a moderate or larger number of (weak or strong) instruments in the system. Hence whereas the first of our bias correction procedures is likely to be attractive mainly when there are weak instruments in the system, the second procedure does not depend on this.

The bias correction procedures proposed in this paper have the particular advantage over other bias reduction techniques e.g, Fuller's corrected $LIML$ estimator, of not depending on assuming normality for the errors. Indeed they are valid under relatively weak assumptions and in the case of the second procedure the errors need not even be independently distributed.. In addition they

may have several other advantages.. *First,* the standard delete one jackknife estimator referred to as $JN2SLS$ recently proposed by Hahn, Hausman and Kuersteiner (2004) does generally reduce the bias of the endogenous variable coefficient estimator but it does not necessarily reduce the bias of non-redundant exogenous variable coefficient estimators since, as Ip and Phillips (1998) show, the exogenous coefficient bias is not monotonically decreasing with the sample size. Our bias corrections will work for both endogenous and exogenous coefficient estimators.

Second, the Chao and Swanson (2007) bias correction does not work well for orders of overidentification smaller than 8 whereas our procedures will successfully reduce bias in such cases. Hansen, Hausman and Newey (2008) point out that in checking five years of the microeconometric papers published in the *Journal of Political Economy*, in the *Quarterly Journal of Economics* and in the *American Economic Review*, 50 percent had at least 1 overidentifying restriction, 25 percent had at least 3 and 10 per cent had 7 or more. This justifies the need to have procedures available that can work well in practice, at different orders of overidentification and with a range of instruments. When there is only one overidentifying restriction, the frequently used $2SLS$ does not have a finite variance and this applies in 50% of the papers analyzed above. However, our procedures based on $k-$class estimators for which $k < 1$ lead to estimators that have higher order moments and so will dominate $2SLS$ in this crucial case. Indeed, we show in our simulations that both our procedures work well not only for the case of one overidentifying restriction but, more generally, for both small and large numbers of instruments.

Third, following Hansen, Hausman and Newey (2008), we have checked the microeconometric papers published in 2004 and 2005 for the same journals and we found that 40 per cent of them used more than 1 included endogenous variable in their simultaneous equation models. Moreover, $2SLS$ is the most commonly chosen estimation procedure. This supports the need to develop attractive estimation methods that can deal with more than 1 included endogenous variable in the system. Note that, in the actual econometrics literature, some of the procedures that have been developed only allow for 1 included endogenous regressor variable (see e.g. Donald and Newey (2001)) and they have used $2SLS$ as the main estimation procedure. We propose in this paper an alternative estimation procedure that allows for any number of included endogenous variables in the equation.

Fourth, our procedures are very easy to implement and do not use a traditional "bias correction" approach whereby the estimated bias is subtracted from the uncorrected estimate. At the same time, the bias correction works even if we are introducing irrelevant variables in any equation of the system. *Finally*, we are able to prove theoretically that both of our approaches produce unbiased estimators up to, at least, order T^{-1}, and also, that the approach that combines $k-$class estimators produces a smaller MSE than $2SLS$ when the equation of interest is just identified, overidentified of order one and overidentified at least of order 8.

A final remark is that there is a growing econometric literature concerned with nearly-weak-identification (see e.g. Caner (2007), and Antoine and Renault (2009)), where standard strong asymptotics is valid. Our bias correction mechanisms will work for this case also.

The structure of the paper is as follows. Section 2 presents the model and it gives a summary of large$-T$ approximations which provide a theoretical underpinning for the bias correction. We also discuss some of the exact finite sample theory results for $2SLS$ which provide further theoretical support, and we examine the weak instrument model of Staiger and Stock (1997) and Chao and Swanson (2007). In Staiger and Stock only weak instruments are available, while in the strong asymptotic setting, we have a mixture of weak and strong instruments. In Section 3 we discuss in detail one possible procedure for bias correction based on introducing redundant exogenous variables and we also consider how to choose the redundant variables by examining how the asymptotic covariance matrix changes with the choice made. This provides criteria for selecting such variables optimally. In Section 4 we develop a new estimator based a linear combination of $k-$class estimators with $k < 1$ which is unbiased to order T^{-1} and has higher moments and we examine how to estimate the variance of the new combined estimator. In Section 5 we present the results of simulation experiments which indicate the usefulness of our proposed procedures. Section 6 provides an empirical application where we show the performance of our new estimator and the results when compared with alternative estimators such as $2SLS$. 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 presented in Nagar (1959) and which appears in many textbooks. Hence we consider a simultaneous equation model containing G equations given by

$$(1) \qquad\qquad By_t + \Gamma x_t = u_t, \quad t = 1, 2, ..., T,$$

in which y_t is a $G \times 1$ vector of endogenous variables, x_t is a $R \times 1$ 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 $G \times R$. It is assumed that B is non-singular so that the reduced form equations corresponding to (1) are

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

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

the system as

(2) $$YB' + X\Gamma' = U.$$

Here, Y is a $T \times G$ matrix of observations on endogenous variables, X is a $T \times R$ 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

(3) $$y_1 = Y_2\beta + X_1\gamma + u_1,$$

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 $E(u_1 u_1') = \sigma_{11} I_T$.

We shall find it convenient to rewrite (3) as

(4) $$y_1 = Y_2\beta + X_1\gamma + u_1 = Z_1\alpha + u_1,$$

where $\alpha' = (\beta, \gamma)'$ and $Z_1 = (Y_2 : X_1)$.

The reduced form of the system includes

$$Y_1 = X\Pi_1 + V_1,$$

in which $Y_1 = (y_1 : Y_2)$, $X = (X_1 : X_2)$ is a $T \times R$ matrix of observations on R exogenous variables with an associated $R \times (g_1 + 1)$ 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 A. (i): The $T \times R$ matrix X is strongly exogenous and of rank K with limit matrix $\lim_{T \to \infty} T^{-1} X'X = \Sigma_{XX}$, which is $R \times R$ positive definite, and that (ii): Equation (3) is over-identified so that $K > g_1 + r_1$, 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 R exceeds $g_1 + r_1$ 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 estimator moments exist: see Sargan (1974).

Nagar, in deriving the moment approximations to be discussed in the next section, assumed that the structural disturbances were normally, independently and identically distributed as in the above model. Subsequently it has proved possible to obtain his bias approximation with much weaker

6

conditions. In fact, Phillips (2000, 2007) showed that neither normality nor independence is required and that a sufficient condition for the Nagar $2SLS$ bias approximation to hold is that the disturbances in the system be Gauss Markov (i.e., with zero expectation and a scalar covariance matrix), which includes, for example, disturbances generated by $ARCH/GARCH$ processes. To derive the higher order bias, which is required for the first of our bias reduction procedures, stronger assumptions are required. Again normality is not needed for the Mikhail approximation to be valid, but it is necessary to assume that the errors are independently and symmetrically distributed, see Phillips and Liu-Evans (2011).

2.1   Large T-approximations in the simultaneous equation model

In his seminal paper, Nagar (1959) presented approximations for the first and second moments of the $k-$class of estimators where $k = 1 + \theta/T$, 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 (4), then the bias of $\hat{\alpha}_k$ where we define L as the degree of overidentification, is given by

$$(5) \qquad\qquad E(\hat{\alpha}_k - \alpha) = [L - \theta - 1]Qq + o(T^{-1}).$$

2. The second moment matrix of the $k-$class estimator for α in (4) is given by

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

where

$$A^* = [(2\theta - (2L - 3))tr(C_1 Q) + tr(C_2 Q)]I + \{(\theta - L + 2)^2 + 2(\theta + 1)\}C_1 Q + (2\theta - L + 2)C_2 Q.$$

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

$$L = r_2 - g_1,$$

where $r_2 = R - r_1$ 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

$$Q = \begin{bmatrix} \bar{Y}_2' \bar{Y}_2 & \bar{Y}_2' X_1 \\ X_1' \bar{Y}_2 & X_1' X_1 \end{bmatrix}^{-1}$$

Further, we may write that

$$V_2 = u_1 \pi' + W,$$

7

where u_1 and W are independent and $\frac{1}{T}\begin{bmatrix} E(V_2'u_1) \\ 0 \end{bmatrix} = \sigma^2 \begin{bmatrix} \pi \\ 0 \end{bmatrix} = q.$

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

$$C = E[\frac{1}{T}V_z'V_z] = \begin{bmatrix} (1/T)E(V_2'V_2) & 0 \\ 0 & 0 \end{bmatrix} = C_1 + C_2,$$

where $C_1 = \begin{bmatrix} \sigma^2\pi\pi' & 0 \\ 0 & 0 \end{bmatrix}$ and $C_2 = \begin{bmatrix} 1/TE(W'W) & 0 \\ 0 & 0 \end{bmatrix}.$

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

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

while the second moment approximation is

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

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

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

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

Notice that this bias approximation contains the term $(L-1)Qq$ which, as we have seen, is the approximation of 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 $R - (g_1 + r_1) = r_2 - g_1 = 1$.

It is seen that if θ is chosen equal to $L-1$ in the $k-$class bias approximation in (5), the bias disappears to order T^{-1}. Hence when $k = 1 + \frac{L-1}{T}$ we have Nagar's unbiased estimator and its second moment approximation can be found by setting $\theta = L - 1$ in (6). Thus writing the estimator as $\hat{\alpha}_N$ we have

$$E(\hat{\alpha}_N - \alpha)(\hat{\alpha}_N - \alpha)' = \sigma^2 Q[I + A^*_{L-1}] + o(T^{-2})$$

$$A^*_{L-1} = [tr(C_1Q) + tr(C_2Q)]I + (1 + 2L)C_1Q + LC_2Q.$$

Later in the paper we shall wish to compare the second moment approximations for Nagar's unbiased estimator and the $2SLS$ estimator.

It is well known that in simultaneous equation estimation there is an existence of moments problem; the $LIML$ estimator has no moments of any order in the classical framework and neither does any $k-$class estimator when $k > 1$. Hence Nagar's unbiased estimator has no moments and this has made the estimator less attractive than it might otherwise be. In particular, there is the usual question of the interpretation of the approximations while the estimator itself may be subject to outliers. On the other hand, the $2SLS$ estimator ($k = 1$) has moments up to the order of overidentification, L, while for $0 \leq k < 1$, moments may be shown to exist up to order $(T - g_1 - r_1 + 2)$, see, in particular, Kinal (1980). Little use has been made of $k-$class estimators for which $0 < k < 1$ despite the fact that there is no existence of moments problem. However, we shall see in the next section that there are situations, especially when bias correction is required, where such estimators can be most helpful.

2.2   Exact finite sample theory

Exact finite sample theory in the classical simultaneous equation model has a long history. Many of the developments are reviewed in the recent book by Ullah (2004). Much of this work was focused on a structural equation in which there were just two endogenous variables. Consider the equation

$$y_1 = \beta y_2 + X_1 \gamma + u_1,$$

which includes as regressors, one endogenous variable and a set of exogenous variables and which is a special case of (3). Assuming that the equation parameters are overidentified at least of order two, Richardson and Wu (1971) provided expressions for the exact bias and MSE of both the $2SLS$ and Ordinary Least Squares (OLS) estimators. In particular, the bias of the $2SLS$ estimator of β was shown to be

(6) $$E(\hat{\beta} - \beta) = -\frac{\sigma_{22}\beta - \sigma_{12}}{\sigma_{22}} e^{-\frac{\mu'\mu}{2}} {}_1F_1\left(\frac{r_2}{2} - 1; \frac{r_2}{2}; \frac{\mu'\mu}{2}\right),$$

where ${}_1F_1$ is the standard confluent hypergeometric function, see Slater (1960), σ_{22} is the variance of the disturbance in the second reduced form equation, σ_{12} is the covariance between the two reduced form disturbances. Also r_2 is the number of exogenous variables excluded from the equation which appear elsewhere in the system and $\mu'\mu$ is the concentration parameter. In this framework Owen (1976) noted that the bias and mean squared error of $2SLS$ are monotonically non-increasing functions of the sample size, a result that has implications for successful jackknife estimation. Later Ip and Phillips (1998) extended Owen's analysis to show that the result does not carry over to the $2SLS$ estimator of the exogenous coefficients.

Hale, Mariano and Ramage (1980) considered different types of misspecification in the context of (3) where the equation with two included endogenous variables is part of a general system. Specifically

9

they were concerned with the cases where exogenous variables are incorrectly included/omitted in parts of the system. Suppose that we include redundant variables in the first equation where we can also allow for other equations of the system to be overspecified by adding redundant variables as well. Then under this type of overspecification, the exact bias of the general $k-$class estimator of the coefficient of the endogenous regressor is given in Hale, Mariano and Ramage (1980). For $k = 1$, the $2SLS$ case, the bias *has the same structure* as when there is no overspecification.

Hale, Mariano and Ramage (1980) have two main conclusions

1. The fact of adding redundant exogenous variables in the first equation will decrease both $\mu'\mu$, the concentration parameter, and $m = (T - r_1)/2$, in relation to the correctly specified case. The direction of this effect on bias involves a trade-off: the bias is decreased because m decreases and is increased by decreasing $\mu'\mu$.

2. The MSE seems to increase with this type of overspecification mainly because of the decrease of $\mu'\mu$.

In relation to the above we should note that when the redundant variable is a weak instrument, the effect on the concentration parameter is likely to be particularly small. As a result the MSE may not increase and could well decrease. As we shall see below, this observation has important implications for the development of bias corrected estimators.

Finally we may deduce from the exact bias expression that, conditional on the concentration parameter, the bias is minimized for $r_2 = 2$, which is the case where L, the order of overidentification, is equal to one for the first equation of a two equation model.

To see why this is so we note that, following Ullah (2004, page 196), we may expand the confluent hypergeometric function in (10) as:

$$
{}_1F_1\left(\frac{r_2}{2} - 1; \frac{r_2}{2}; \frac{\acute{\mu}\mu}{2}\right) = 1 + \frac{\left(\frac{r_2}{2} - 1\right)\acute{\mu}\mu}{\frac{r_2}{2}}\frac{\acute{\mu}\mu}{2} + \frac{\left(\frac{r_2}{2} - 1\right)\left(\frac{r_2}{2}\right)}{2\left(\frac{r_2}{2}\right)\left(\frac{r_2}{2} + 1\right)}\left(\frac{\acute{\mu}\mu}{2}\right)^2 + ...,
$$

from which we may deduce that the bias is minimized for a given value of $\acute{\mu}\mu$ when $r_2 = 2$.

Hansen, Hausman and Newey (2008) recently proposed the use of the Fuller (1977) estimator with Bekker (1994) standard errors in order to improve the estimation and testing results in the case of many instruments. However, their procedure can still produce very large biases, and our procedure can complement theirs, since in our case, we can get a nearly unbiased estimator. Moreover, since we deal with $k-$class estimators where $k < 1$, our estimator has higher moments, and also we can allow for the existence of redundant variables in any part of the simultaneous equation system.

2.3 The weak instrument case

In the previous sections, we were dealing with a system where enough strong instruments were available to identify the structural parameters, together with some or many redundant variables that may exist in the structural equation of interest (weak instruments). So, in the reduced form, we have a mixture of weak and strong instruments. In this case, standard Nagar expansions are valid, and in this paper we find a way to reduce the bias, even if we have many redundant variables in the system.

Another different setting is the one proposed in Staiger and Stock (1997), where *only weak instruments are available* (locally unidentified case) in the reduced form equation to be used as instruments (excluding the exogenous variables which appear in the first equation). No strong instrument is available.

Chao and Swanson (2007) provide the asymptotic bias and the MSE results in the same framework of weak asymptotics. Suppose a system of the type

$$
\begin{aligned}
y_1 &= \beta y_2 + X\gamma + u \\
y_2 &= Z\Pi + X\Phi + v,
\end{aligned}
$$

where y_1 and y_2 are $T \times 1$ vectors of observations on the two endogenous variables, X is an $T \times r_1$ matrix of observations on r_1 exogenous variables included in the structural equation and Z is a $T \times r_2$ matrix of instruments which contains observations on r_2 exogenous variables. The asymptotic bias of the IV estimator takes the same form as in (6); hence we find that the asymptotic bias in this weak instrument model, takes the same form as the exact bias in the strong instrument case. Furthermore the above result goes through without an assumption of normality and in the presence of stochastic instruments. Note also that the framework of Staiger and Stock (1997) and Chao and Swanson (2007) allow for weak instruments to be in the equation of interest, but not in any other equation of the system. Chao and Swanson also assume that $r_2 \geq 4$ to ensure that the result encompasses the Gaussian case.

We can also provide an expansion up to $O\left(T^{-2}\right)$ with the approach of the "weak IV asymptotics" by using the bias expressions of Chao and Swanson (2007). The results of the previous theorems stay the same and also the bias is minimized for $r_2 = 2$, i.e, the equation is overidentified of order one.

3 Bias reduction

When the matrix X_1 in (4) is augmented by a set of r_1^* redundant exogenous variables X^*, which appear elsewhere in the system, so that $X_1^* = (X_1 : X^*)$, the equation to be estimated by $2SLS$ is

now

(7)
$$y_1 = Y_2\beta + X_1\gamma + X^*\delta + u_1,$$

where δ is zero but is estimated as part of the coefficient vector $\alpha^* = (\beta', \gamma', \delta')$.

The bias in estimating α^* takes the same form as (5) except that in the definition of Q, X_1 is replaced by X_1^*. If r_1^* is chosen so that the notional order of overidentification, defined as $R - (g_1 + r_1 + r_1^*)$ is equal to unity, then the bias of the resulting $2SLS$ estimator $\hat{\alpha}^*$, is zero to order T^{-2}. It follows that $\hat{\alpha}^*$ is unbiased to order T^{-2} and the result holds whether or not there are weak instruments in the system as will be discussed below. Of course, the introduction of the redundant variables changes the variance too. In fact the asymptotic covariance matrix of $\hat{\alpha}$ is given by $\lim_{T\to\infty} \sigma^2 TQ$, whereas the asymptotic covariance matrix for $\hat{\alpha}^*$ is given by $\lim_{T\to\infty} \sigma^2 TQ^*$; where Q^* is obtained by replacing X_1 with X_1^* in Q. Hence, the estimator $\hat{\alpha}^*$ will not be asymptotically efficient. However if the redundant variables are weak instruments the increase in the small sample variance will be small and the MSE may be reduced. We can find a second order approximation to the variance of $\hat{\alpha}^*$ from a simple extension of the earlier results of Nagar (1959) given above.

Although the proposed procedure is based on large sample asymptotic analysis, the case is strengthened by the exact finite sample results in Section 4 since we have observed there that setting $r_2 = 2$ minimizes the exact bias for a given value of the concentration parameter although it does not completely eliminate it. Of course, as we have noted, the introduction of the redundant exogenous variables alters the value of the concentration parameter so we cannot be certain that the bias is reduced. However, all our results indicate that bias reduction will be successful.

Effectively, we are using in our analysis an "optimal" number of redundant variables for inclusion in the first equation to improve on the finite sample properties of the $2SLS/IV$ instrumental variable estimator of α. By doing so we know that, in the two-equation model, this type of overspecification has the drawback that it decreases the concentration parameter $\mu'\mu$, so we want to choose the redundant variables from the weaker instruments. In the definition of weak instruments, it is usual to refer to such instruments as those which make only a small contribution to the concentration parameter (see eg. Stock, Wright and Yogo (2002) and Davidson and Mackinnon (2006)). So this adds support for what we do and guides our selection of variables in choosing the redundant set. This is further discussed in the next section.

A related approach is to commence from the correctly specified equation and achieve the bias reduction by suitably reducing the number of instruments included at the first stage. If the number of retained instruments at the first stage includes the exogenous variables that appear in the correctly specified equation to be estimated while the total number is such as to make the equation notionally overidentified of order one, then the resulting IV estimator will also be unbiased to order T^{-2}. An

12

obvious way to choose the instruments to be removed from the first stage would be to select the weakest. We shall consider this approach again in Section 3.1.

We have seen in Section 2 that the limiting bias in the weak instrument model of Staiger and Stock (1997) and of Chao and Swanson (2007), takes the same form as the exact bias in finite samples. Insofar as we can minimize the exact bias by setting $r_2 = 2$, it follows that we also minimize the limiting bias in the weak instrument case. As a result we can reasonably expect our bias reduction procedure will work in models even where the weak instrument problem is severe.

We have earlier noted that the existence of moments depends on the "notional" order of overidentification. To see why this is so we consider a simple system of 2 endogenous variables as given in (7), and a reduced form equation

$$(8) \qquad\qquad\qquad y_2 = X\pi_2 + v_2,$$

then the $2SLS$ estimator of β has the form $(y_2'Py_2)^{-1}(y_2'Py_1)$, where $P = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1'$ and $\operatorname{rank}(P) = p = \operatorname{rank}(X) - \operatorname{rank}(X_1)$. To examine the existence of moments, we can take a look at the canonical case where the covariance matrix of (u_1, v_2) is just the identity, and $\beta = 0 = \gamma = \delta = \pi_2$. Then, conditional on y_2, the distribution of the $2SLS$ estimator is $N\left(0, (y_2'Py_2)^{-1}\right)$, so the conditional moments all exist, the odd vanish, and the unconditional even moment of order $2r$ exists only if $E\left((y_2'Py_2)^{-r}\right)$ exists. But $y_2'Py_2 \sim \chi_p^2$, so this it exists only for $p < 2r$, or $2r \le p - 1$. In the non-canonical case the odd-order moments do not vanish, but the argument is the same. This provides the intuition of why, by adding redundant exogenous in the system, in this case by altering p, we alter the notional order of identification and it is this which determines the existence of moments[1].

Note that in our bias results, α corresponds to the full vector of parameters so that the bias is zero to order $1/T^2$ for the exogenous coefficient estimators too. This is important because not all bias corrected estimators proposed in the literature have this quality. For example, the standard delete-one jackknife in Hahn, Hausman and Kuersteiner (2004) is shown to perform quite well and it has moments whenever $2SLS$ has moments. However, the case analyzed does not include any exogenous variables in the equation that is estimated so no results are given for the jackknife applied to exogenous coefficient estimators. In fact, a necessary condition for the jackknife to reduce bias is that the estimator has a bias that is monotonically non-increasing with the sample size. As noted previously, the bias of the $2SLS$ estimator of the endogenous variable coefficient is monotonically non-increasing and so in this case the jackknife satisfies the necessary condition for bias reduction to be successful. However, Ip and Phillips (1998) show that the same does not apply to the exogenous coefficient bias. In fact, even the MSE of the $2SLS$ estimator in this case is not monotonically non-increasing in the sample size which is a disquieting limitation of the $2SLS$ method.

---

[1]We thank Grant Hillier for providing this simple proof.

As noted in Section 2, the Nagar bias approximation is valid under conditional heteroskedasticity so that we can prove that the bias disappears up to $O\left(T^{-1}\right)$ even when conditional heteroskedasticity is present. The essential requirement is that the disturbances be Gauss Markov which covers a wide range of generalized-autoregressive conditional heteroskedastic ($ARCH/GARCH$) processes (see, e.g. Bollerslev (1986)).

The bias corrected estimator that is discussed above is essentially a $2SLS$ estimator of an over-specified equation. It will have moments up to the notional order of overidentification as can be seen from Mariano (1972) so that whereas the first moment exists, the second and all other higher moments do not. However we prefer to use a bias corrected estimator which has higher moments and this can be achieved with an alternative member of the $k-$class.

Careful examination of the analysis employed by Nagar and Mikhail reveals that if k is chosen to be $1 - T^{-3}$, then the bias approximations to order T^{-1} and T^{-2} are the same as those for $2SLS$, i.e. when $k = 1$. Consequently, when redundant variables are added to reduce the notional order of overidentification L to unity, the estimator with $k = 1 - T^{-3}$ will be unbiased to order T^{-2} and its moments will exist. The estimator is, essentially, $2SLS$ with moments and is therefore to be preferred to $2SLS$.

One of the key issues about bias correcting with the introduction of redundant variables concerns the choice of the redundant variable set. We now consider this.

3.1   Choosing the Redundant Variables

Asymptotically efficient $k-$class estimators have an asymptotic covariance matrix given by

$$\sigma^2 \text{plim} T \left[\begin{array}{cc} \Pi_2'X'X\Pi_2 & \Pi_2'X'X_1 \\ X_1'X\Pi_2 & X_1'X_1 \end{array} \right]^{-1},$$

and the $2SLS$ small sample variances are typically estimated from the matrix

$$\hat{\sigma}^2 \left[\begin{array}{cc} \hat{\Pi}_2'X'X\hat{\Pi}_2 & \hat{\Pi}_2'X'X_1 \\ X_1'X\hat{\Pi}_2 & X_1'X_1 \end{array} \right]^{-1}.$$

Suppose that the relevant equation has a redundant variable set X^*, contained in the matrix X, where $X_1^* = (X_1 : X^*)$; then it is easily shown that the part of the asymptotic covariance matrix that relates to the coefficients of the non-redundant regressors is proportional to the limiting value of

$$V = \left[\begin{array}{cc} \Pi_2'X'M_{X^*}X\Pi_2 & \Pi_2'X'M_{X^*}X_1 \\ X_1'M_{X^*}X\Pi_2 & X_1'M_{X^*}X_1 \end{array} \right]^{-1},$$

where $M_{X^*} = I - X^*\left[(X^*)'X^*\right]^{-1}(X^*)'$.

It follows that we may use this result in developing criteria for choosing the redundant variable set. We should choose the redundant regressors X^* to minimize the difference, in some sense, between the two matrices. One approach is to minimize the difference between the generalized variances. In effect this means choosing X^* to minimize the determinant of V. A simpler approach is to minimize the difference between the traces of the two matrices. This amounts to choosing X^* so that the sum of the variances increases the least. This is clearly simple to implement since it implies that we merely compare the main diagonal terms of the estimated covariance matrices.

In the simple model in which there are no exogenous variables in the equation of interest, the above reduces to

$$V = (\pi_2' X' M_{X^*} X \pi_2)^{-1}.$$

If we partition the X matrix as $X = \begin{bmatrix} X^{**} & X^* \end{bmatrix}$ and the π_2 vector as $\pi_2 = (\pi_{21.}' \pi_{22}')'$ where $X \pi_2 = X^{**} \pi_{21} + X^* \pi_{22}$, then the above further reduces to

$$V = (\pi_{21}' (X^{**})' M_{X^*} X^{**} \pi_{21})^{-1}.$$

We recognize this as being proportional to the inverse of the concentration parameter. Hence, we see that in the simple model our criterion involves minimizing the reduction in the concentration parameter. Again, the practical implementation requires that the unknown reduced form parameter vector be replaced with an estimate. Hence, X^* is chosen to minimize the asymptotic variance of the endogenous coefficient estimator. Recall that in the previous section it was noted that choosing the redundant variables to be the weakest instruments will also minimize the reduction in the concentration parameter: hence we see that the two criteria are equivalent in this case.

It was noted in Section 3 that a biased corrected estimator can also be found by manipulating the number of instruments at the first stage. The resulting estimator then takes the form

$$\hat{\alpha} = \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} Y_2' P_{x^{**}} Y_2 & Y'_2 X_1 \\ X_1' Y_2 & X_1' X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2' P_{x^{**}} \\ X_1' \end{pmatrix} y_1,$$

where X^{**} is the matrix of instruments (which includes X_1) and $P_{X^{**}}$ is the projection matrix $X^{**}(X^{**\prime} X^{**})^{-1} X^{**\prime}$.

The asymptotic covariance matrix for this estimator is

$$\sigma^2 plim T \begin{bmatrix} \Pi_2' X' P_{X^{**}} X \Pi_2 & \Pi_2' X' X_1 \\ X_1' X \Pi_2 & X_1' X_1 \end{bmatrix}^{-1}.$$

Notice that only the top left hand quadrant is affected by the choice of instruments. The estimator will be fully efficient when $X^{**} = X$. In the case considered above where the equation to be estimated has no included exogenous variables, the asymptotic variance is then $\sigma^2 p \lim T (\pi_2' X' P_{X^{**}} X \pi_2)^{-1}$.

It does not appear possible to show that either estimator dominates the other on a MSE criterion. However, it can easily be shown that when X^* is orthogonal to X^{**}, $i.e$, the redundant variables in the estimated equation are orthogonal to all the other instruments, and whenever the alternative method is used it is the X^*variables which are excluded from the first stage, the two approaches lead to exactly the same estimators. Some simulations have shown that the two estimators give quite similar results. One reason for preferring the redundant variable estimator is that it is simply a special case of $2SLS$ (or $k-$class) about which a good deal is known, including its exact properties in simple cases. In what follows we shall be primarily concerned with the redundant variable method.

Later in the paper we shall examine the properties of the redundant variable estimator in a series of Monte Carlo experiments but first we consider an alternative approach to bias correction by combining $k-$class estimators.

4   A new estimator as a linear combination of k-class estimators with $k < 1$

In this section we develop a new estimator as a linear combination of $k-$class estimators with $k < 1$. As we shall see the new estimator has all higher moments and is unbiased to order T^{-1}.

To proceed we shall denote the estimator when $k = 1 - T^{-3}$ as $\widehat{\alpha}_{k_1}$ and the estimator for which $k = 1 - T^{-1}$ as $\widehat{\alpha}_{k_2}$. According to (5), the bias of $\hat{\alpha}_{k_1}$ up to T^{-1}takes the form

$$Bias\left(\widehat{\alpha}_{k_1}\right) = (L - 1) Qq,$$

and $\widehat{\alpha}_{k_2}$,the $k-$class estimator for which $k = 1 - \frac{1}{T}$, has bias given by

$$Bias\left(\widehat{\alpha}_{k_2}\right) = LQq.$$

It follows that the linear combination of $\widehat{\alpha}_{k_1}$ and $\widehat{\alpha}_{k_2}$ given by

$$(9) \qquad \widehat{\alpha}_{k_3} = L\widehat{\alpha}_{k_1} - (L - 1)\widehat{\alpha}_{k_2}$$
$$(10) \qquad = \widehat{\alpha}_{k_1} + (L - 1)(\widehat{\alpha}_{k_1} - \widehat{\alpha}_{k_2})$$

is unbiased to order T^{-1}.This is the combined estimator.

It is of interest that the combined estimator is approximately unbiased under much less restricted conditions than those imposed by Nagar (1959) as are the estimators considered in the previous section. The $2SLS$ case is considered in Phillips (2000, 2007) where it was shown that the bias approximation holds even in the presence of $ARCH/GARCH$ disturbances while Phillips (1997) had extended the result to cover the general $k-$class estimator. An earlier attempt to reduce bias by

combining $k-$class estimators is given in Sawa (1972) who showed that a linear combination of OLS and $2SLS$ can be found to yield an estimator which is unbiased to order σ^2. However the estimator is not unbiased to order T^{-1} and has been found to perform relatively poorly in some cases. In fact, the $O(T^{-1})$ bias approximation for OLS differs considerably from the $O(\sigma^2)$ approximation as can be seen in Kiviet and Phillips (1996). A further disadvantage is that a variance approximation is not available.

The next Theorem shows how the combined estimator may have smaller MSE than $2SLS$ for systems that have, at least, a moderate degree of overidentification. That means that when we have a reasonable number of instruments (upwards of 7), our combined estimator will not only be unbiased to order T^{-1}, but also it will improve on $\widehat{\alpha}_{k_1}$ in terms of MSE. Moreover, note that the combined estimator is always unbiased to order T^{-1} even if we have added irrelevant variables in any part of the system (as Hale, Mariano and Ramage (1980) point out). This also includes the case where we have many weak instruments in the system.

Theorem 1 *In the model of (1) and under Assumption A in Section 2, the difference to order T^{-2} between the mean squared error of the combined estimator $\widehat{\alpha}_{k_3}$ and that of the $k-$class estimator for which $k = 1 - T^{-3}$, $\widehat{\alpha}_{k_1}$, is given by*

$$E(\widehat{\alpha}_{k_1} - \alpha)(\widehat{\alpha}_{k_1} - \alpha)' - E(\widehat{\alpha}_{k_3} - \alpha)(\widehat{\alpha}_{k_3} - \alpha)'$$
$$= \sigma^2(L-1)[(L-7)QC_1Q - 2QC_2Q - 2(tr(C_1Q)Q - QC_1Q)] + o(T^{-2}).$$

The proof of this Theorem is given in the Appendix where it is shown that the Nagar expansion of the estimation error of the combined estimator $\widehat{\alpha}_{k_3}$ to $O_p(T^{\frac{-3}{2}})$ is exactly the same as that for the Nagar unbiased estimator for which $k = 1 + \frac{L-1}{T}$. Second moment approximations are based only on terms up to $O_p(T^{\frac{-3}{2}})$ so that the second moment approximation for the combined estimator will be the same as that of the Nagar unbiased estimator to order T^{-2}. Hence the combined estimator is essentially the same as Nagar's unbiased estimator with all necessary moments.

It is of interest to compare the combined estimator with a bias corrected estimator proposed by Donald and Newey (2001, page 1164). Their $B2SLS$ estimator is given as

$$\hat{\delta} = \left(Z'P_X Z - \bar{\Lambda} \ Z'Z)\right)^{-1} \left(Z'P_X \ \ y_1 - \bar{\Lambda} \ Z'y_1\right),$$

which is clearly a $k-$class estimator for which $\frac{(1-k)}{k} = -\bar{\Lambda}$. Solving for k yields $(1-k) + k\bar{\Lambda} = 0$ or $1 = k(1 - \bar{\Lambda})$ which gives $k = \frac{1}{(1-\bar{\Lambda})}$. Now Donald and Newey (2001) state that $\bar{\Lambda} = (R - r_1 - 2)/T$ (in out notation) so that $k = \frac{1}{1-(R-r_1-2)/T} = 1 + \frac{(R-r_1-2)/T}{1-(R-r_1-2)/T} > 1$.

Nagar's unbiased estimator for which $k = 1 + (L-1)/T$ is not the same as the Donald-Newey estimator. However to order T^{-1} the Donald and Newey (2001) estimator has $k = 1 + (R - r_1 -$

2)/T, which coincides with Nagar's estimator when there is only one endogenous regressor but not otherwise. Therefore, the unbiasedness of the Donald and Newey (2001) estimator in the context of more than 1 endogenous variables in the system is not a trivial extension. Moreover, the Donald and Newey estimator is momentless for the case of one endogenous regressor . One of the main novelties of our paper, is that we present an unbiased estimator to order T^{-1} in the setting of any number of endogenous variables and that allows for the existence of any number of irrelevant variables in the system. Moreover, our estimator has moments when dealing with the $k-$class where $k < 1$.

The precise value for L at which the result in Theorem 1 above is positive semi-definite so that the MSE of the combined estimator is less than that of $2SLS$, depends upon the relationship between QC_1Q and QC_2Q. This is especially so in the two equation model where $tr(C_1Q)Q = QC_1Q$. However it is to be expected that $QC_1Q \geq QC_2Q$, especially when the simultaneity is strong , so that the combined estimator is preferred on a MSE criterion for $L > 7$. Of course, when $L = 1$ the MSE of $2SLS$ does not exist and the combined estimator is preferred. In this case $\widehat{\alpha}_{k_3} = \widehat{\alpha}_{k_1}$; thus the combined estimator reduces to the $k-$class estimator for which $k = 1 - T^{-3}$. Although we show that the combined estimator only dominates in other cases for $L = 0$ and for values of $L > 7$, it should be noted that the combined estimator is always approximately unbiased, has all necessary moments and may be preferred in cases where biases are large even if it it is not optimal on a MSE criterion.

Note that the Chao and Swanson (2007) bias correction procedure does not work very well when the number of excluded regressors is less than 11 (see Chao and Swanson (2007), Table 2) while we have a very simple procedure that is unbiased up to T^{-1} for any number of excluded regressors. Our combined estimator, which does not assume that all instruments are weak, will work particularly well in the case of many (weak) instruments and moderate number of (weak) instruments. The result is the same regardless of whether our instruments are weak or strong and the unbiasedness result holds under possible overspecification in terms of including irrelevant variables in other equations. Note also that Chao and Swanson (2007) need a correctly specified system, while our combined estimator will be unbiased up to T^{-1} even if we have irrelevant exogenous variables in any part of the system. As Hansen, Hausman and Newey (2008) point out, there are many papers that need in practice to use a moderate/large number of instruments, and Angrist and Krueger (1991) is an example where a large number of instruments is used.

As a final remark, note that in the case $L = 0$, the combined estimator reduces to the $k-$class estimator with $k = 1 - T^{-1}$. This has moments and clearly dominates $2SLS$ and the $k = 1 - T^{-3}$ estimator on a MSE criterion in this situation. Actually this is quite obvious since the variance of the $k-$class will decline as k is reduced and so the estimator with $k = 1 - T^{-1}$ will have a smaller variance while at the same time it has a zero bias to order T^{-1} when $L = 0$. This estimator is of considerable importance since it has higher moments when $2SLS$ does not yet, as noted in the

introduction, and it is very commonly found in practice.

4.1 The combined estimator covariance matrix

While the combined estimator, $\hat{\alpha}_{k_3}$, is asymptotically efficient and has the same asymptotic variance as that of $\hat{\alpha}_{k_1}$, the estimator that is preferred to $2SLS$, their respective small sample variances may differ considerably. To note that, from Theorem 1 we have shown that, if we denote

$$E(\hat{\alpha}_{k_1} - \alpha)(\hat{\alpha}_{k_1} - \alpha)' = MSE\left(\hat{\alpha}_{k_1}\right)$$

and

$$E(\hat{\alpha}_{k_3} - \alpha)(\hat{\alpha}_{k_3} - \alpha)' = MSE\left(\hat{\alpha}_{k_3}\right),$$

then

$$(11) \qquad\qquad MSE\left(\hat{\alpha}_{k_3}\right) + \Delta = MSE\left(\hat{\alpha}_{k_1}\right).$$

where

$$\Delta = \sigma^2(L-1)[(L-7)QC_1Q - 2QC_2Q - 2(tr(C_1Q)Q - QC_1Q)].$$

Also, $\hat{\alpha}_{k_3}$ is unbiased to order T^{-1} so that

$$(12) \qquad\qquad MSE\left(\hat{\alpha}_{k_3}\right) = var\left(\hat{\alpha}_{k_3}\right) + o\left(T^{-2}\right).$$

Now, $\hat{\alpha}_{k_1}$ is not unbiased to order T^{-1} and

$$MSE\left(\hat{\alpha}_{k_1}\right) = var\left(\hat{\alpha}_{k_1}\right) + (bias\left(\hat{\alpha}_{k_1}\right))^2 = var\left(\hat{\alpha}_{k_1}\right) + (L-1)^2 Qqq'Q + o(T^{-2})$$
$$(13) \qquad\qquad = var\left(\hat{\alpha}_{k_1}\right) + (L-1)^2 \sigma^2 QC_1Q + o(T^{-2}).$$

From (11), (12) and (13) it follows that

$$var\left(\hat{\alpha}_{k_3}\right) = var\left(\hat{\alpha}_{k_1}\right) + \sigma^2\left(L-1\right)^2 QC_1Q - \sigma^2(L-1)[(L-7)QC_1Q - 2QC_2Q$$
$$-2(tr(C_1Q)Q - QC_1Q)] + o(T^{-2})$$
$$= var\left(\hat{\alpha}_{k_1}\right) + \sigma^2\left(L-1\right)\left[5QC_1Q + 2\left(tr(QC_1)Q\right) + 2QC_2Q\right] + o\left(T^{-2}\right).$$

It is clear that for large L there will be a large difference in the variances and we should sensibly adjust for this in any inference situation.

In general it is not easy to find a satisfactory estimator of the covariance matrix of a combined estimator when the component estimators are correlated. We could base an estimate for the variance of the combined estimator on the above by replacing the unknown terms with estimates but this

is not likely to be successful. It turns out that we can overcome this problem by noting that the combined estimator has the same approximate variance as a member of the $k-$class.We first note that the approximate second moment matrix of the general $k-$class estimator for α is given in section 2 as

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

where

(14)
$$\begin{aligned} A^* &= [2\theta - (2L - 3)tr(C_1Q) + tr(C_2Q)]I + \{(\theta - L + 2)^2 \\ &\quad + 2(\theta + 1)\}C_1Q + (2\theta - L + 2)C_2Q. \end{aligned}$$

and all terms are as defined there.

1. For 2SLS we have, on putting $\theta = 0$,

$$A_0^* = [(L^2 - 4L + 6)C_1Q + (-L + 2)C_2Q. - (2L - 3)tr(C_1Q)I + tr(C_2Q)]I.$$

2. For Nagar's unbiased estimator we put $\theta = L - 1$ in the above A^* which yields

$$A_{L-1}^* = (2L + 1)C_1Q + LC_2Q + tr(C_1Q)I + tr(C_2Q)I.$$

3. For the combined estimator, we shall write the associated A^* as A_c^* and using the result in Theorem 1 and (14), it may be shown that it is given by

$$A_c^* = [(2L + 1)C_1Q + LC_2Q + tr(C_1Q)I + tr(C_2Q)I].$$

   which is the same as the approximation for Nagar's unbiased estimator.

Since the latter two estimators are unbiased to order T^{-1}, it follows that the approximate variances for the combined estimator and Nagar's unbiased estimator are the same to order T^{-2}. Hence the combined estimator behaves like Nagar's unbiased estimator but it has moments whereas Nagar's estimator does not. To estimate the variance of the general $k-$class we use

(15)
$$\sigma^2 \begin{pmatrix} Y_2'Y_2 - k\hat{V}_2'\hat{V}_2 & Y_2'X_1 \\ X_1'Y_2 & X_1'X_1 \end{pmatrix}^{-1},$$

so that for Nagar's unbiased estimator we simply set $k = 1 + \frac{L-1}{T}$. Clearly this can also provide a consistent variance estimate for the combined estimator notwithstanding that the Nagar estimator does not have moments of any order.

5 Simulations

5.1 Results for the redundant variable estimator

We show in this section how our procedure works in practice. Table 1 provides simulations for a sample of size of 100 observations based on 5000 replications, and the structure we consider is of the form

$$(16) \qquad \begin{aligned} y_{1t} &= \beta_1 y_{2t} + x'_{1t}\gamma + u_{1t} \\ y_{2t} &= \beta_2 y_{1t} + x'_t \gamma_1 + u_{2t} \end{aligned}$$

$$(17) \qquad y_{2t} = x'_t \pi_2 + v_{2t}.$$

In many of our simulations, γ was chosen to be zero so that the first equation in (16) did not contain any non-redundant exogenous variables. However, in some of our experiments (EXP) we included one or more non-redundant exogenous variables whereupon γ had some non-zero components. 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

$$(18) \qquad YB' + X\Gamma' + U = 0.$$

In all our experiments the X matrix contains a first column of ones, while the other exogenous variables are generated as normal random variables with a mean zero and variance 10. The endogenous coefficient matrix was chosen as

$$B' = \begin{pmatrix} -1 & 0.267 \\ 0.222 & -1 \end{pmatrix},$$

in all experiments. Note that the coefficient of the endogenous regressor, β_1, which is the object of interest in our first five experiments (Table 1), has the value 0.222. The experiments differ through changing the coefficient matrix of the exogenous coefficients, Γ. In the first five experiments this is effected by changing the exogenous coefficients in the second equation only since no exogenous coefficients appear in the first equation. The disturbances are generated as the product of a 100×2 matrix of $N(0,1)$ times a Cholesky decomposition matrix (C). This was chosen as $C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix}$ in all experiments except experiment 2 where it was changed to $C = \begin{pmatrix} 10 & -1 \\ -1 & 4 \end{pmatrix}$.

The model has been estimated first by $2SLS$ and then by $2SLS$ with redundant exogenous regressors. We show the consequences of including exogenous variables that act as weak or strong

instruments, and we consider different structures in the disturbances. We give details of the individual experiments below. Note also that the standard error (se) of $2SLS$ when we add two redundants does not exist, and that is why we also report the interquartile range (IR).

Moreover, we also give the results for an alternative $k-$class estimators: $k = 1 - T^{-3}$ which yields $\widehat{\beta}_{k_1}$. This $k-$class estimators behaves very much like $2SLS$ $\left(\text{i.e. } \widehat{\beta}_1\right)$ but the variance will exist when the system is overidentified of order 1.

Experiment 1

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1.47 & 0.246 & 0.136 & 0.043 \end{pmatrix}, \quad C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix},$$

Here the equation specified contains two endogenous variables and no exogenous ones. There are four exogenous variables in the second equation so that the first equation is overidentified of order three.

Notice that the fourth coefficient in Γ is numerically small. This corresponds to a weak instrument. To create a situation where the first equation is overidentified of order one so that the $2SLS$ estimator is unbiased to order T^{-2}, we shall need to augment the first equation with two redundant variables. There are four variables to choose from. In the first experiment the the augmentation proceeds in two stages. First we introduce variables three and four which are the weaker instruments. Then an alternative ordering was tried with variables two and three. Thus now one of the redundant variables is a relatively strong instrument. A comparison of these two cases will show the different effects of redundant weak and strong instruments.

As our theory predicts, the bias is drastically reduced when either pair of exogenous variables are added as redundant, but it is the weaker instruments that increase the standard error the least. Moreover, if we add three exogenous variables then the redundant $2SLS$ estimator does not have moments of any order and it creates outliers in our simulations.

Experiment 2

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.5 & 0.100 & 0.036 & 0.043 \end{pmatrix}, \quad C = \begin{pmatrix} 10 & -1 \\ -1 & 4 \end{pmatrix}.$$

Here the coefficients of the Γ matrix are different to those in experiment 1 and the structural disturbance in the first equation is much smaller. The third and fourth coefficients are small too so now there are two instruments that are relatively weak. In this experiment we choose variables three and four for the redundant pair and so in this case the augmenting variables are both relatively weak. Again the bias is drastically reduced.

Experiment 3

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1.47 & 0.246 & 0.136 & 0.043 \end{pmatrix}, \ C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix}.$$

In this experiment the coefficients are the same as in experiment 1 but we introduce conditionally heteroskedastic structural disturbances. For reasons of operational simplicity we select the model of Wong and Li (1997) given by

$$
\begin{aligned}
E\left(u_{1t}^2|I_{t-1}\right) &= \alpha_0 + \alpha_1 u_{1t-1}^2 + \alpha_2 u_{2t-1}^2 \\
E\left(u_{2t}^2|I_{t-1}\right) &= \gamma_0 + \gamma_1 u_{1t-1}^2 + \gamma_2 u_{2t-1}^2,
\end{aligned}
$$

where $\alpha_0 = 81, \gamma_0 = 0.64, \alpha_1 = 0.25, \alpha_2 = \gamma_1 = \gamma_2 = 0.16$ and I_{t-1} denotes the sigma field generated by the past values of u_t. Here the redundant variables are variable three which is relatively strong and variable four which is weak. Again, in Table 1 it is seen that the bias is drastically reduced when two exogenous variables are added.

Experiment 4

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1.47 & 1.5 & 0.136 & 0.043 \end{pmatrix}, \ C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix}.$$

Note that experiment 4 differs from experiment 1 only in the fact that the exogenous variable in position 2 is now a much stronger instrument. Now, as seen in Table 1, the bias and standard error of the $2SLS$ estimator is drastically reduced.

In the simulations the weaker instruments are chosen as redundant. When these are introduced, the standard error increases only from 0.200 to 0.207. However when the second and third exogenous variables are chosen, where the second exogenous variable is clearly a very strong instrument, the more modest bias reduction is accompanied by very large increase in the variance.

Experiment 5

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1.47 & 0.5 & 0.0006 & 0.0003 \end{pmatrix}, \ C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix}.$$

This experiment is designed to show the consequences when we have very weak instruments in the system so that the coefficient values for variables three and four are very small. In the simulations again variables three and four were first chosen as the redundant ones. Thus only very weak instruments are used as redundant variables. This was followed by choosing variable two with variable four thus introducing a strong instrument with one weak one.

The bias of the $2SLS$ is obviously very large, and when we introduce the two weakest instruments in equation 1 as the redundant, we observe how the bias is drastically reduced, and also, how the standard error increases very little, as our theory predicts. However, when a strong instrument is introduced, the standard error increases considerably.

In Table 1 below, exogenous variable 2 is a strong instrument whereas exogenous variables 3 and 4 are both weak instruments.

Table 1: Simulation results for $2SLS$, with two exogenous redundant variables in the first equation

EXP	$2SLS$			$2SLS$, 2−redundants		
	bias$(\widehat{\beta}_1)$	$se(\widehat{\beta}_1)$	$IR(\widehat{\beta}_1)$	bias$(\widehat{\beta}_1^*)$	$se(\widehat{\beta}_1^*)$	$IR(\widehat{\beta}_1^*)$
1	-0.182	0.566	0.667	0.009	$0.657^{(1)}$	0.766
				0.004	$0.757^{(2)}$	0.859
2	-0.088	0.483	0.578	-0.022	$0.660^{(1)}$	0.632
3	-0.197	0.553	0.682	-0.024	$0.784^{(1)}$	0.789
4	-0.019	0.200	0.264	0.001	$0.207^{(1)}$	0.274
				-0.014	$1.648^{(2)}$	0.874
5	-0.123	0.442	0.547	0.001	$0.484^{(1)}$	0.597
				0.015	$0.808^{(2)}$	0.874

$^{(1)}$(exogenous variables 3 and 4 are the redundant set);

$^{(2)}$(exogenous variables 2 and 3 are the redundant set)

Table 1 (cont.): $IV-$estimator with $k = 1 - T^{-3}$

EXP	IV			IV, 2−redundants		
	bias$(\widehat{\beta}_{k_1})$	$se(\widehat{\beta}_{k_1})$	$IR(\widehat{\beta}_{k_1})$	bias$(\widehat{\beta}_{k_1}^*)$	$se(\widehat{\beta}_{k_1}^*)$	$IR(\widehat{\beta}_{k_1}^*)$
1	-0.178	0.554	0.693	$0.009^{(1)}$	$0.660^{(1)}$	$0.765^{(1)}$
2	-0.084	0.450	0.541	$0.002^{(1)}$	$0.487^{(1)}$	$0.545^{(1)}$
3	-0.178	0.543	0.623	$-0.020^{(1)}$	$0.942^{(1)}$	$0.760^{(1)}$
4	-0.021	0.201	0.264	$0.002^{(1)}$	$0.207^{(1)}$	$0.270^{(1)}$
5	-0.123	0.444	0.544	$-0.008^{(1)}$	$0.477^{(1)}$	$0.600^{(1)}$

$^{(1)}$(exogenous variables 3 and 4 are the redundant ones).

We present now another set of simulation experiments. Here we are interested in the bias corresponding to an estimate of an exogenous variable coefficient in the first equation. This is a situation where the Jackknife estimator of Hahn, Hausman and Kuersteiner (2004) will not work . Moreover,

we again show the results for the redundant variable estimators. Table 2 shows the results when in (16) we use the different estimators.

Experiment 6

$$\Gamma = \begin{pmatrix} 0 & 0.60 & 0 & 0 \\ 0.70 & 0.08 & 0.0006 & 0.0003 \end{pmatrix}, \; C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix},$$

Here γ is the parameter corresponding to the exogenous variable in the first equation. This is the object of our interest in this experiment and 0.60 is its true value. Again, the bias is greatly reduced (from 0.189 to 0.072) with the introduction of the redundant exogenous variable. $\widehat{\gamma}$ corresponds to $2SLS$ and $\widehat{\gamma}^*$ the bias corrected estimator by adding redundant variables.

Experiment 7

$$\Gamma = \begin{pmatrix} 0 & 0.60 & 0 & 0 \\ 0.70 & 0.08 & 0.3 & 0.00003 \end{pmatrix}, \; C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix}.$$

Here γ is again the parameter corresponding to the exogenous variable in the first equation and 0.60 is its true value. It is seen that the bias is greatly reduced, from 0.061 to 0.014, with the introduction of the redundant exogenous variable.

Table 2: Simulation results for $2SLS$, with one exogenous redundant variable in the first equation

EXP	$2SLS$			$2SLS$, 1−redundant		
	bias($\widehat{\gamma}$)	$se(\widehat{\gamma})$	$IR(\widehat{\gamma})$	bias($\widehat{\gamma}^*$)	$se(\widehat{\gamma}^*)$	$IR(\widehat{\gamma}^*)$
6	0.189	0.523	0.526	$0.072^{(1)}$	$0.632^{(1)}$	$0.581^{(1)}$
7	0.061	0.407	0.482	$0.014^{(2)}$	$0.505^{(2)}$	$0.514^{(2)}$

Table 2 (cont.): IV-estimator with $k = 1 - T^{-3}$

	IV			IV, 1−redundant		
	bias($\widehat{\gamma}_{k_1}$)	$se(\widehat{\gamma}_{k_1})$	$IR(\widehat{\gamma}_{k_1})$	bias($\widehat{\gamma}^*_{k_1}$)	$se(\widehat{\gamma}^*_{k_1})$	$IR(\widehat{\gamma}^*_{k_1})$
6	0.186	0.466	0.497	$0.067^{(2)}$	$0.792^{(2)}$	$0.566^{(2)}$
7	0.052	0.402	0.475	$0.005^{(2)}$	$0.464^{(2)}$	$0.513^{(2)}$

[1](exogenous variable 3 is the redundant one). [2](exogenous variable 4 is the redundant one)

Finally, we are also interested in knowing how our method works in the context of having many weak instruments in equation 2. Thus we perform two final experiments.

Experiment 8

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1.47 & 0.5 & 0.00001 & 0.00002 & ... & 0.00033 \end{pmatrix}, \; C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix}.$$

The Γ matrix is now (2×35) and includes the 33 values $0.00001, 0.00002, 0.00003, ..., 0.00033$ which are associated with weak instruments. Note that we have one strong and one reasonably strong instrument , and the rest are weak instruments.

We see in Table 3 the results for this experiment. The biases of the $2SLS$ estimator $\widehat{\beta}_1$ and that of $\widehat{\beta}_1^*$, our redundant-variables procedure estimator, are compared when we include the exogenous variables 3 to 35 (all weak instruments) in Γ. The reduction in the bias of $\widehat{\beta}_1$ is most impressive and the mean squared error is substantially reduced.

Experiment 9

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1.47 & 0.5 & 0.00001 & 0.00002 & ... & 0.00058 \end{pmatrix}, \; C = \begin{pmatrix} 112 & -1 \\ -1 & 4 \end{pmatrix}.$$

Γ is now a 2×60 matrix where we have 58 weak instruments in equation 2.

In Table 3 it is again seen that when we introduce the 58 weak instruments in equation 1 (the exogenous variables are in position 3 up to 60 in Γ) the bias reduction is impressive and there will be a relatively huge reduction in the MSE.

Table 3: Simulation results for $2SLS$, and with exogenous redundant variables in the first equation

EXP	$2SLS$			$2SLS-$redundants		
	bias$(\widehat{\beta}_1)$	$se(\widehat{\beta}_1)$	$IR(\widehat{\beta}_1)$	bias$(\widehat{\beta}_1^*)$	$se(\widehat{\beta}_1^*)$	$IR(\widehat{\beta}_1^*)$
8	-1.088	0.263	0.344	0.005	0.531	0.668
9	-1.409	0.216	0.290	-0.009	0.581	0.720

Table 3 (cont.): $IV-$estimator with $k = 1 - T^{-3}$

EXP	IV			$IV-$redundants		
	bias$(\widehat{\beta}_{k_1})$	$se(\widehat{\beta}_{k_1})$	$IR(\widehat{\beta}_{k_1})$	bias$(\widehat{\beta}_{k_1}^*)$	$se(\widehat{\beta}_{k_1}^*)$	$IR(\widehat{\beta}_{k_1}^*)$
8	-1.082	0.264	0.347	0.005	0.542	0.664
9	-1.413	0.213	0.283	-0.005	0.587	0.703

5.2 Results for the combined estimator

Finally, we simulate the combination of $k-$class estimators $\widehat{\beta}_{k_3}$. Since Table 1 refers to an equation that is overidentified of order 3, we use the estimator $\left(3\widehat{\beta}_{k_1} - 2\widehat{\beta}_{k_2}\right) = \widehat{\beta}_{k_3}$. We simulate again the five experiments of Table 1 with this mixture of estimators and the results are given in Table 4 below.

Table 4: $IV-$estimator $\widehat{\beta}_{k_3}$

EXP	IV		
	bias($\widehat{\beta}_{k_3}$)	$se(\widehat{\beta}_{k_3})$	$IR(\widehat{\beta}_{k_3})$
1	-0.028	0.658	0.760
2	-0.024	0.620	0.656
3	-0.039	0.708	0.748
4	-0.002	0.205	0.271
5	-0.016	0.497	0.617
8	-0.703	0.390	0.513
9	-0.971	0.292	0.381

By comparing the results of Table 1 for the first five experiments with those in Table 4, we see the clear advantage this combined estimator has in relation to the traditional $2SLS$ estimator mainly in terms of the biases. In the experiments where the weakest instruments were used for the redundant set, the redundant variable estimator compares favorably with the combined estimator in terms of the standard deviation and is, as expected, generally less biased. However when the redundant variables include at least one strong instrument the redundant variable estimator, while continuing to have very small bias, also suffered an increase in the standard deviation, see especially Experiment 4, indicating the need for caution in using the estimator. Of course, in practice, it is to be expected that the estimated variance would provide a warning against using the estimator.

We also show the behavior of the combined estimator for experiments 8 and 9. In these last two cases, we set $\widehat{\beta}_{k_3} = \left(34\widehat{\beta}_{k_1} - 33\widehat{\beta}_{k_2}\right)$ and $\widehat{\beta}_{k_3} = \left(59\widehat{\beta}_{k_1} - 58\widehat{\beta}_{k_2}\right)$ respectively.

In these experiments we have $L >> 8$, and as our theory predicts, the combined estimator is clearly superior in terms of MSE to $2SLS$. However the estimator is still very badly biased even though the bias of order T^{-1} has been removed. We also note that in this case, the redundant-variable method yielded a very small bias.

At first sight it is surprising that the bias of the combined estimator is so large. However, examining the higher order bias approximation in Section 2.1 for $2SLS$, we see that the bias term of order T^{-2} will be large for large L and this will apply to the combined estimator also. Whereas in many cases the bias term of order T^{-2} is relatively small, it is not when there are many instruments. Even when $L \leq 8$ (see experiments $1 - 5$), there are some cases where it is advisable in practice to calculate both $2SLS$ and our combined estimator and to compare the point estimates for the econometrician to uncover large biases.

. In the simulations of Hansen, Hausman and Newey (2008) and Chao and Swanson (2007), it is possible to see that $LIML$, Fuller or other estimators, may not be median unbiased or unbiased

in some situations and it is here where our proposal is useful. As Davidson and Mackinnon (2006) concluded, there is no clear advantage in using $LIML$ or $2SLS$ in practice (mainly in situations such as experiments 8 and 9 where we have many weak instruments). Here we find a case where our combined estimator generally improves on $2SLS$ in terms of MSE when $L \succeq 8$, and where the bias correction works when there are redundant variables in any equation of the simultaneous equation system whatever the order of identification. Therefore, if $L \succeq 8$ the researcher should use our combined estimator instead of $2SLS$; and even if $L < 8$, in some cases such as experiments 1 and 5, we can see by comparing Tables 1 and 4, that the bias correction is so large than the researcher may still consider it useful to use the combined estimator and compare it with $2SLS$.

Moreover, in the case of many instruments, the bias correction through the introduction of redundant variables produces a much larger bias reduction than our combined estimator. So in practice, it is advisable in the case of many instruments (more than 7) to compare also our combined estimator (clearly superior to the estimator with $k = 1 - T^{-3}$) with the redundant variable estimator.

Finally, in order to check our Theorem 1, we repeat Experiment 9 but instead of including 58 instruments we only include just 10 weak instruments with coefficients $0.00001, 0.00002, 0.00003, ...,$ 0.00010. This is what we call Experiment 10. Our Theorem 1 predicts that from the number of instruments 8 onwards, the combined estimator should improve on MSE in relation to $2SLS$. We show in Table 5 below that when we have 10 weak instruments, the combined estimator does indeed improve considerably on MSE in relation to $2SLS$. The same situation is seen in Experiments 8 and 9 when we compare Tables 3 and 4. Thus the combined estimator improves the MSE in relation to $2SLS$ in a very important way. The gains that we can get in terms of a reduced MSE from the combined estimator versus $2SLS$ are very impressive.

Table 5: Simulation results for $2SLS$, and with the combined estimator with 10 weak instruments. $\left(11\widehat{\beta}_{k_1} - 10\widehat{\beta}_{k_2}\right) = \widehat{\beta}_{k_3}$.

EXP	$2SLS$			combined-IV		
	bias$(\widehat{\beta}_1)$	$se(\widehat{\beta}_1)$	$MSE(\widehat{\beta}_1)$	bias$(\widehat{\beta}_{k_3})$	$se(\widehat{\beta}_{k_3})$	$MSE(\widehat{\beta}_{k_3})$
10	-0.485	0.358	0.363	-0.143	0.506	0.277

6 An empirical application

In this section we consider an application of bias correction in a simple case and we focus on the combined estimator which is likely to be of particular interest in practical cases.

One main advantage of our combined estimator is that it is very easy to compute. As is seen in (10), it simply requires the computation of two $k-$class estimators for $k = 1 - T^{-1}$ and $k = 1 - T^{-3}$.

The STATA package [2] has a module that implements straightforwardly k-class estimators[3] written by Baum, Schaffer and Stillman (2008)[4]. To compute the variance of our combined estimator, we can simply refer to (15), from where we find that our combined estimator has an asymptotic variance equivalent to the k-class estimator with $k = 1 + \frac{L-1}{T}$.

We are interested in analyzing the permanent income hypothesis, and for this purpose, we focus on equation (16.35) in Example 16.7 in Wooldridge (2008, page 563) where, following Campbell and Mankiw (1990), we can specify

(19) $$gc_t = \beta_0 + \beta_1 gy_t + \beta_2 r3_t + u_t,$$

where gc_t is the annual growth in real per capita consumption (excluding durables), gy_t is the growth in real disposable income and $r3_t$ is the (ex post) real interest rate as measured by the return on three-month T-bill rates.

The hypothesis of interest is that the pure form of the permanent income hypothesis states that $\beta_1 = \beta_2 = 0$. Campbell and Mankiw (1990) argue that β_1 is positive if some fraction of the population consumes current income instead of permanent income. And the permanent income hypothesis theory with a nonconstant real interest rate implies that $\beta_2 > 0$.

We use the same dataset as in Example 16.7 in Wooldridge (2008, page 563). We use the annual data from 1959 through 1995 given in CONSUMP.RAW to estimate the equation given by (19) and the results are given in Table 6. The two endogenous variables are gy_t and $r3_t$. Since a full three-equation model is not specified the choice of instrumental variables is ad hoc and in equation (16.35), Wooldridge uses as instruments gc_{t-1}, gy_{t-1} and $r3_{t-1}$ although the full model is not necessarily dynamic.We report the results when we apply the OLS and $2SLS$ estimators in rows (3) and (4) and the corresponding standard errors in parenthesis. These corresponds to the same reported outcomes in Wooldridge (2008, Example 16.7). From row (4), we can check how the pure form of the permanent income hypothesis is strongly rejected because the coefficient on gy_t is economically large and statistically significant. The coefficient on the real interest rate is very small and statistically insignificant. Campbell and Mankiw (1990) use different lags as instruments, so in order to confirm these results, we apply in row (5) $2SLS$ when the instruments are gc_{t-1}, gy_{t-1}, $r3_{t-1}$, gc_{t-2}, gy_{t-2}, $r3_{t-2}$, gc_{t-3}, gy_{t-3} and $r3_{t-3}$. The results when we compare rows (4) and (5) are very similar except that the estimated coefficient on gy_t increases slightly from 0.5904 to 0.6153, and as it is very usual (see for example Wooldridge (2008, Example 15.8)) the standard error of the $2SLS$

---

[2]see http://www.stata.com/

[3]IVREG2 is a STATA module available at http://ideas.repec.org/c/boc/bocode/s425401.html

[4]We are very grateful to Jeff Wooldridge for pointing out about the existence of this STATA module, and how our estimator can be easily computed in STATA. We are also in debt with Jeff Wooldridge for all his comments and suggestions that we have obtained about this empirical application.

is slightly reduced when more instruments are included. That is why we may prefer row (5) instead of row (4). That means that a 1% increase in disposable income increases consumption by 0.61%.

We also wanted to check the robustness of the results to the choice of the instruments, and we considered then another observable variable such as pop_t, that represents the population in thousands. We then apply again $2SLS$ with the new set of instruments gc_{t-1}, gc_{t-2}, gc_{t-3}, gc_{t-4}, pop_{t-1}, pop_{t-2}, pop_{t-3}, pop_{t-4} and pop_{t-5} accounting for lags of pop_t. Row (6) reports the results of $2SLS$, and the increased consumption is now reported to be 0.60%. Again, there is an important reduction in the standard error by increasing the number of instruments from 3 to 9.

We are particularly interested to apply our combined estimator which assumes that the instruments are exogenous. We found pop_t,which represents the population in thousands, to be a potentially suitable exogenous instrument. A final set of instruments uses only lags of population and the results are in row (7). In this case, the estimated elasticity is still statistically significant and it increases until 0.70%. So there is evidence that the estimated increase in consumption is significantly larger than the 0.58% reported by OLS and it varies depending on the choice of instruments. $2SLS$ reports estimated increases in the range of 0.59% to 0.70%.

Given the small sample sizes that usually available when dealing specially with macroeconomic data, we wish to check the previous results when using our combined estimator and we use as instruments gc_{t-1}, gy_{t-1}, $r3_{t-1}$, gc_{t-2}, gy_{t-2}, $r3_{t-2}$, gc_{t-3}, gy_{t-3} and $r3_{t-3}$. We compute then the estimator given in (10) and we also compute the $k-$class estimator where $k = 1 + \frac{L-1}{T}$, to obtain the asymptotic variance of our combined estimator. These are given in Table 6. We repeat the same procedure also when we use the set of instruments that include lags of gc_t and pop_t. And finally, we also show the results when lags of pop_t are the instruments. Clearly pop_t is exogenous and when we regressed the endogenous variables on lags of pop_t we found many of the coefficients to be significant. So a final set of instruments uses only lags of pop_t. When we use a sufficient number of lags, e.g, nine in this case, the combined estimator is expected to have better properties than the corresponding estimator that uses lags of the endogenous variables, on the basis of Theorem 1.

Table 6: Testing the Permanent Income Hypothesis (Wooldridge (2008), example 16.7)

Dependent variable: gc_t	Variables (standard errors are given in parenthesis)					
	gy_t		$r3_t$		$intercept$	
OLS	0.5781	(0.0715)	-0.0002	(0.0006)	0.0082	(0.0020)
2SLS*	0.5904	(0.1215)	-0.0002	(0.0007)	0.0079	(0.0027)
2SLS**	0.6153	(0.1204)	-0.0003	(0.0007)	0.0076	(0.0027)
2SLS***	0.6011	(0.0988)	-0.0003	(0.0007)	0.0079	(0.0024)
2SLS****	0.7046	(0.1627)	-0.0005	(0.0007)	0.0061	(0.0032)
Combined $k-$class**	0.6336	(0.1402)	-0.0003	(0.0007)	0.0076	(0.0030)
Combined $k-$class***	0.6088	(0.1013)	-0.0003	(0.0007)	0.0079	(0.0024)
Combined $k-$class****	0.8068	(0.2559)	-0.0005	(0.0010)	0.0047	(0.0047)

*Instruments are gc_{t-1}, gy_{t-1} and $r3_{t-1}$.

**Instruments are gc_{t-i}, gy_{t-i}, $r3_{t-i}$, $i = 1, 2, 3$.

***Instruments are gc_{t-i}, $i = 1, 2, 3, 4$, pop_{t-j}, $i = 1, ..., 5$.

****Instruments are pop_{t-i}, $i = 1, ..., 9$.

Table 6 shows the dangers of applying OLS, and the important bias that affects $2SLS$. As noted above our new combined estimator is very easy to compute and it provides significantly different results to $2SLS$. The combined estimator with the first set of 9 instruments confirms that the pure form of the permanent income hypothesis is rejected and that some fraction of the population consumes current income instead of permanent income. We find that a 1% increase in disposable income does not increase consumption only by 0.61% (as $2SLS$ would suggest with the second set of instruments), but by 0.63%, as our combined estimator suggests, and it is statistically significant with a t-statistic of 0.6336/0.1402=4.52. When lags of gc_t and pop_t are considered, again, our new combined $k-$class estimator provides a higher estimated increase (0.61%) in consumption than when $2SLS$ is used (0.60%). When the last set of instruments that contain lags of pop_t is used, $2SLS$ reports an estimated increase in consumption that is 0.70%, while with our combined $k-$class estimator that provides improved finite sample properties, the estimate is 0.81% and it is statistically significantly different from zero. In this latter case, the context is similar to that in which the combined estimator was shown to be unbiased to order T^{-1}.

In short, with $2SLS$ the estimated increase in consumption is sensitive to the choice of instruments and it varies in the range from 0.59% to 0.70%. While with our combined estimator, the estimated increase varies from 0.61% to 0.81% depending on the choice of instruments. In all cases, for any given choice of set of instruments that we take, there is evidence that the estimated increase in consumption is significantly larger than the 0.58% that OLS reports, and also than the estimate

that $2SLS$ reports since our combined $k-$class estimator consistently provides a higher estimate that is also statistically significant.

7 Conclusion

In this paper we propose two very simple bias reduction procedures which are valid under relatively weak disributional assumptions and which may be applied to $k-$class estimators in a general static simultaneous equation model, the first of which works particularly well when we have weak instruments in the system. In fact, for this approach, which employs redundant exogenous variables to achieve bias reduction, the presence of weak instruments is a definite advantage since their use will minimise the resulting increase in the variance of the estimator. Effectively we design the optimal structure of the simultaneous equations system to minimize the bias of our estimator through the introduction of redundant exogenous variables. When we choose an estimator from the $k-$class where $k < 1$, especially $k = 1 - T^{-1}$ and $k = 1 - T^{-3}$, then after applying our bias correction,the estimator will have a finite variance and may have a smaller MSE than $2SLS$. The bias corrected $k = 1 - T^{-3}$ estimator is of particular interest; it has a very similar behavior to $2SLS$ and it is unbiased to order T^{-2}. However the procedure must be used with care and the best results will be obtained when the redundant regressors are weak instruments. We show in simulations the advantages of our approach, including the settings of (many) weak instruments in the system.

The second approach constructs an estimator as a linear combination of $k-$class estimators with $k < 1$. It achieves a bias reduction without the need to introduce redundant variables in the system and it will work well whether or not there are weak instruments present. However it is likely to be of particular interest when there is a large number of instruments. We show that the combined estimator has the same MSE to order T^{-2}as the well known Nagar unbiased estimator and yet it has all necessary moments and generally has a smaller MSE than $2SLS$ when $L = 0$ or 1 and $L > 7$. This covers the majority of applications noted in the Hausman, Hansen and Newey (2008) paper and referred to in the Introduction. In addition the estimator remains unbiased up to order T^{-1} even when we have redundant variables (including weak instruments) in any other equation of the simultaneous equation system. Therefore we can allow for overspecification in any equation of the system in our bias correction procedure.

As mentioned earlier, there is evidence in the literature, for example see Davidson and Mackinnon (2006), that concludes that there is no clear advantage in using $LIML$ or $2SLS$ in practice, particularly in the context of weak instruments. In this paper we show that in some circumstances, $2SLS$ is clearly dominated in MSE by a combination of $k-$class estimators where $k < 1$. From our simulations, the gains in MSE that we can obtain from our new combined $k-$class estimators

versus traditional $k-$class estimators, can be very impressive when the equation of interest is heavily overidentified. Finally, an application shows the usefulness of our new estimator in practice, and how it is very easy to compute.

8   Appendix

Proof of Theorem 1.

Following the approach of Nagar (1959), we may write

$$
\begin{aligned}
\hat{\alpha}_{k_1} - \alpha &= Q[\bar{Z}_1' u_1 + V_z' M^* u_1] - Q[\bar{Z}_1' V_z + V_z' Z_1 + V_z' M^* V_z]Q \times [\bar{Z}_1' u_1 + V_z' M^* u_1] + o_p(T^{-\frac{3}{2}}) \\
&= Q[\bar{Z}_1' u_1 + V_z' M^* u_1] - Q[\bar{Z}_1' V_z + V_z' \bar{Z}_1]QV_z' M^* u_1 \\
&\quad - Q[\bar{Z}_1' V_z + V_z' \bar{Z}_1]Q\bar{Z}_1' u_1 - QV_z' M^* V_z Q Z_1' u_1 + o_p(T^{-\frac{3}{2}}).
\end{aligned}
$$

where all the terms used are defined in Section 2 except for $M^* = I - X(X'X)^{-1}X'$.and $\bar{Z}_1 = (X\Pi_2 : X_1)$.

Also

$$
\begin{aligned}
\hat{\alpha}_{k_2} - \alpha &= Q[\bar{Z}_1' u_1 + \frac{1}{T}V_z' u_1 + (1 - \frac{1}{T})\, V_z' M_1^* u_1] - Q[\bar{Z}_1' V_z + V_z' \bar{Z}_1 + V_z' M^* V_z \\
&\quad + \frac{1}{T}V_z' V_z]Q \times [\bar{Z}_1' u_1 + \frac{1}{T}V_2' u_1 + V_z' M^* u_1] + o(T^{-\frac{3}{2}}) \\
&= Q[\bar{Z}_1' u_1 + \frac{1}{T}V_z' u_1 + (1 - \frac{1}{T})\, V_z' M^* u_1] - .Q[\bar{Z}_1' V_z + V_z' \bar{Z}_1 + V_z' M^* V_z]Q\bar{Z}\prime u \\
&\quad - Q\frac{1}{T}V_z' V_z Q\bar{Z}_1' u_1 - Q(\bar{Z}_1' V_z + V_z' \bar{Z}_1)QV_z' M^* u_1 - Q(\bar{Z}_1' V_z + V_z' \bar{Z}_1)Q\frac{1}{T}V_z' u_1 + o_p(T^{\frac{-3}{2}})
\end{aligned}
$$

Then defining $\hat{\alpha}_{k_3} = L\hat{\alpha}_{k_2} - (L-1)\hat{\alpha}_{k_1}$we find from the above expansions that

$$
\begin{aligned}
\hat{\alpha}_{k_3} - \alpha &= Q\bar{Z}_1' u_1 + QV_z' M^* u_1 - Q[\bar{Z}_1' V_z + V_z' X]Q\bar{Z}_1' u_1 - (L-1)Q\frac{1}{T}V_z' u_1 \\
&\quad - QV_z' M^* V_z Q\bar{Z}_1' u_1 - Q[\bar{Z}_1' V_z + V_z' \bar{Z}_1]QV_z' M^* u_1 \\
&\quad + (L-1)Q\bar{Z}_1' V_z Q\frac{1}{T}V_z' u_1 + (L-1)QV_z' \bar{Z}_1 Q\frac{1}{T}V_z' u_1 + (L-1)Q\frac{1}{T}V_z' V_z Q\bar{Z}_1' u + o_p(T^{-\frac{3}{2}}),
\end{aligned}
$$

where the first term in the above is $O_p(T^{-\frac{1}{2}})$, the terms in the second line are $O_p(T^{-1})$ and the remaining terms are $O_p(T^{-\frac{3}{2}})$.This is the Nagar expansion to $O_p\,(T^{\frac{-3}{2}})$ for $k-$class estimator when $k = 1 + \frac{L-1}{T}$.

To complete the proof we need only compare the second moment approximations. The difference between then has the form

$$
\sigma^2 Q[I + A_0^*] - \sigma^2 Q[I + A_c^*] = \sigma^2 Q(A_0^* - A_c^*).
$$

where we have shown in Section 4 that

$$A_0^* = [(L^2 - 4L + 6)C_1Q + (-L + 2)C_2Q - (2L - 3)tr(C_1Q)I + tr(C_2Q)I],$$

and

$$A_c^* = [(2L + 1)C_1Q + LC_2Q + tr(C_1Q)I + tr(C_2Q)I].$$

By subtraction we find

$$A_0^* - A_c^* = (L - 1)[(L - 7)C_1Q - 2C_2Q - 2(tr(C_1Q)I - C_1Q)).$$

Multiplying through by Q completes the proof.■

References

[1] Andrews, D. W. K. and J. H. Stock (2006), Inference with Weak Instruments, in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Vol III., ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge, UK: Cambridge University Press.

[2] Angrist, J. D., G. W. Imbens and A. B. Krueger (1999), Jackknife Instrumental Variables Estimation, *Journal of Applied Econometrics* 14, 57-67.

[3] Angrist, J. D., and A. B. Krueger (1991), Does Compulsory School Attendance Affect Schooling and Earnings, *Quarterly Journal of Economics* 106, 979-1014.

[4] Antoine, B. and E. Renault (2009), Efficient GMM with Nearly-Weak Identification, *Econometrics Journal* 12, 1, S135-S171.

[5] Baum, C., M. Schaffer and S. Stillman (2008), IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation, Software component provided by Boston College Department of Economics in its series Statistical Software Components with number S425401.

[6] Bekker, P. A. (1994), Alternative Approximations to the Distributions of Instrumental Variable Estimators, *Econometrica* 63, 657-681.

[7] Blomquist, S. and M. Dahlberg (1999), Small Sample Properties of LIML and Jackknife IV estimators: Experiments with Weak Instruments, *Journal of Applied Econometrics* 14, 69-88.

[8] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* 31, 307-327.

[9] Campbell, J. Y. and N. G. Mankiw (1990), Permanent Income, Current Income and Comsumption, *Journal of Business and Economic Statistics* 8, 265-279.

[10] Caner, M. (2007), Testing, Estimation and Higher Order Expansions in GMM with Nearly-Weak Instruments, Working Paper.

[11] Chao, J. and N. R. Swanson (2007), Alternative Approximations of the Bias and MSE of the IV Estimator Under Weak Identification with Application to Bias Correction, *Journal of Econometrics* 137, 2, 515-555.

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

[13] Donald, S. G. and W. K. Newey (2001), Choosing the Number of Instruments, *Econometrica* 69, 5, 1161-1191.

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

[15] Hahn J. and J. A. Hausman (2002a), A New Specification Test for the Validity of Instrumental Variables, *Econometrica* 70, 163-189.

[16] Hahn J. and J. A. Hausman (2002b), Notes on Bias in Estimators for Simultaneous Equation Models, *Economics Letters* 75, 2, 237-41.

[17] Hahn J. and J. A. Hausman (2003), Weak Instruments: Diagnosis and Cures in Empirical Econometrics, *American Economic Review,* 93, 2, 118-125.

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

[19] Hale, C., R. S. Mariano and J. G. Ramage (1980), Finite Sample Analysis of Misspecification in Simultaneous Equation Models, *Journal of the American Statistical Association* 75, 370, 418-427.

[20] Hansen, C., J. Hausman and W. Newey (2008), Estimation with Many Instrumental Variables, *Journal of Business and Economic Statistics* 26, 4, 398-422.

[21] Hausman, J. A. (1983), Specification and Estimation of Simultaneous Equation Models" in Griliches, Zvi and Intriligator, Michael, eds., *Handbook of Econometrics*, Volume 1, Amsterdam: North Holland.

[22] Ip, C. W. and G. D. A. Phillips (1998), The Non-Monotonicity of the Bias and Mean Squared Error of the Two Stage Least Squares Estimators of Exogenous Variable Coefficients, *Economics Letters* 60, 3, 8303-310.

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

[24] Kiviet J.F. and G.D.A. Phillips (1996), The Bias of the Oridinary Least Squares Estimator in Simultaneous Equation Models, *Economics Letters* 53, 161-167.

[25] Mariano, R. (1972), The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators, *Econometrica* 40, 4, 643-652.

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

[27] 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.

[28] Nelson, C. R. and R. Startz, (1990a), The Distribution of the Instrumental Variables Estimator and its T-ratio when Instruments is a Poor One, *Journal of Business* 63, 125-140.

[29] Nelson, C. R. and R. Startz, (1990b), Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator, *Econometrica* 58, 967-976.

[30] Owen, A. D. (1976), A proof that both the Bias and the Mean Square Error of the Two Stages Least Squares Estimator are Monotonically Non-increasing Functions of Sample Size, *Econometrica* 44, 2, 409-411.

[31] Phillips, G. D. A. (1997), An Alternative Approach to Obtaining Nagar-Type Moment Approximations in Simultaneous Equation Models, University of Exeter Discussion Paper.

[32] 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.

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

[34] Phillips, G. D. A. and C. Hale (1977), The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems, *International Economic Review* 18, 219-228.

[35] Phillips, G. D. A. and G.Liu-Evans (2011), The Robustness of the Higher Order $2SLS$ and General $k-$Class Estimator Bias Approximations to Non-Normal Disturbances. Working paper.

[36] Phillips, P. C. B., (1980), The exact distribution of instrumental variable estimators in an equation containing n +1 endogenous variables. *Econometrica* 48, 861–878.

[37] Richardson, D. H. and D.-M. Wu (1971), A Note on the Comparison of Ordinary and Two-State Least Squares Estimators, *Econometrica* 39, 973-982.

[38] Rothenberg, T. (1983), Asymptotic Properties of Some Estimators in Structural Models, in S. Karlin, T. Amemiya and L. Goodman, eds., Studies in Econometrics, Time Series and Multivariate Statistics, Academic Press.

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

[40] Sawa, T. (1972), Finite Sample Properties of the k-class estimators, *Econometrica* 40, 653-680.

[41] Slater, L. J. (1960), *Confluent Hypergeometric Functions*, Cambridge University Press, London.

[42] Staiger, D. and J. H. Stock (1997), Instrumental Variables Regressions with Weak Instruments, *Econometrica* 65, 557-586.

[43] Stock, J. H., J. H. Wright and M. Yogo (2002), A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics* 20, 4, 518-529.

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

[45] Wong, H. and W. K. Li (1997), On a Multivariate Conditional Heteroscedastic Model, *Biometrika* 84, 1, 111-123.

[46] Wooldridge, Jeffrey M. (2008), *Introductory Econometrics: A Modern Approach*, 4th Edition, South-Western College.