

Finite Sample Bias Corrected IV Estimation for Weak and Many Instruments*

Matthew Harding[†] Jerry Hausman[‡] Christopher Palmer[§]

November 6, 2015

Abstract

This paper considers the finite sample distribution of the 2SLS estimator and derives bounds on its exact bias in the presence of weak and/or many instruments. We then contrast the behavior of the exact bias expressions and the asymptotic expansions currently popular in the literature, including a consideration of the no-moment problem exhibited by many Nagar-type estimators. After deriving a finite sample unbiased k -class estimator, we introduce a double k -class estimator based on Nagar (1962) that dominates k -class estimators (including 2SLS), especially in the cases of weak and/or many instruments. We demonstrate these properties in Monte Carlo simulations showing that our preferred estimators outperforms Fuller (1977) estimators in terms of mean bias and MSE.

JEL: C31, C13, C15

Keywords: instrumental variables, weak and many instruments, finite sample, k -class estimators

1. Introduction

The failure of standard first-order asymptotic theory to provide a reliable guide to estimation in the widely used simultaneous equations model in the presence of weak and/or many instruments has led to renewed interest in the finite sample properties of estimators for these models. Weak instruments refers to the situation where in finite samples the endogenous regressor is only weakly correlated to the instrument set, while the many instruments case refers to situations where the dimensionality of the instrument set is large relative to the sample size.

*We are grateful to Tim Armstrong for providing excellent research assistance. Thanks to Plamen Koev for sharing his code for the computation of hypergeometric functions.

[†]Sanford School of Public Policy and Department of Economics, Duke University, Durham, NC 27708; Phone: (919) 613 9300; Email: matthew.harding@duke.edu

[‡]Department of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02142; Email: jhausman@mit.edu

[§]Haas School of Business, University of California at Berkeley, Berkeley, CA 94720; Email: cjpalmer@berkeley.edu

Faced with the failure of the traditional asymptotic framework in these models, econometricians have turned their attention to alternative asymptotic frameworks such as higher-order expansions (Hahn, Hausman and Kuersteiner, 2004), local-to-zero asymptotics (Staiger and Stock, 1997; Poskitt and Skeels, 2006) and many/weak instrument (Bekker) asymptotics (Bekker, 1994; Hansen, Hausman and Newey, 2008). While all these methods have provided valuable insights, we are nevertheless faced with uncertainty as to how successful these approximations can be in the difficult cases of weak and/or many instruments. As with all approximations, we face the choice of how many expansion terms to include before we can be confident in the accuracy of estimations.

Exact finite sample results on the distribution of instrumental variable estimators (IV) have been known for many years but have largely remained outside the grasp of practitioners due to the lack of computational tools for the evaluation of the complicated functions on which they depend in practical time (Anderson and Sawa, 1979; Nagar, 1959; Richardson, 1968; Richardson and Wu, 1971; Sawa, 1972; Phillips, 1980, 1983). Recent computational advances have removed this constraint, and we are now able to employ the exact finite sample methodology to better understand the statistical behavior of these models in the case of weak and/or many instruments. Still, Hirano and Porter (forthcoming) show that a finite sample unbiased estimate is unattainable without imposing further assumptions on the structure of the problem. As an example of this, Andrews and Armstrong (2015) develop an unbiased estimator when the sign of the first stage is known; although it appears the estimator does not do as well as Fuller (1977), there is no absolute ranking.

In this paper, we approach the problem of estimating simultaneous equations models from the perspective of the exact finite sample properties of these estimators. We consider broad classes of estimators such as the k -class estimators and evaluate their promises and limitations as methods to correctly provide finite sample inference on the structural parameters in simultaneous equations. To contextualize the performance of our estimator, Hahn, Hausman and Kuersteiner (2004) demonstrate that with weak instruments, Fuller outperforms LIML. While LIML does not have finite sample moments, which in practice means that it provides very dispersed estimates, Fuller has all moments and removes substantial finite sample bias. In this paper, we consider the “classic” situation of one endogenous variable and homoskedastic normal disturbances and find that our preferred double- k class estimators typically outperform Fuller estimates in mean bias and MSE and rarely do worse.

In Section 2, we start by considering the exact finite sample bias of the 2SLS estimator and analyze its dependence on the goodness of fit as measured by the concentration parameter in the first stage regression and the number of instruments employed. We show that the finite sample bias is strictly increasing in the number of instruments K and is always positive for $K > 1$. The bias decreases exponentially with the concentration parameter and is negligible for values of the concentration parameter greater than 60. Furthermore, we show that the maximum finite sample bias for weak

and/or many instruments cannot exceed the ratio of the covariance between the first stage and the structural equation and the variance of the first stage.

We also derive the exact finite sample unbiased estimator as a member of the k -class of IV estimators. Additionally, we show that by employing the k -class we face a fundamental trade-off. Estimators which aim to limit the finite sample bias are also estimators which lack finite sample moments and are thus infeasible in practice. We show using a Monte Carlo design that for severe cases of many and weak instruments there is no member of the k -class that also has moments which can adequately minimize the effect of finite sample bias.

In order to address this problem we consider a modification of the k -class, originally due to Nagar (1962), which allows for distinct parameters (k_1, k_2) to enter the estimating equations. We use a Monte Carlo design to show that this class of estimator is rich enough to allow for estimators with finite sample moments which are also unbiased and have very attractive MSE properties. We derive the optimal choice of parameters for a set of double k -class estimators based on 2SLS.

In Section 3, we resurrect the “double k -class” of estimators and extend Nagar (1962) to consider how finite sample bias and MSE are affected by choice of the double k -class parameters (k_1, k_2) in the presence of weak and/or many instruments. Section 4 derives double k -class estimators with certain optimality properties (e.g. in terms of their finite sample bias and MSE) in such circumstances. Section 5 presents Monte Carlo results contrasting the performance of our preferred double k -class estimators with 2SLS and Fuller. We find that our preferred bias-minimizing double k -class estimator outperforms 2SLS and Fuller’s minimum-bias estimator (“Fuller1”) even in the presence of weak and/or many instruments and that our preferred MSE-minimizing double k -class estimator generally outperforms Fuller’s minimum-MSE estimator (“Fuller4”) in MSE, especially in many-instrument settings. Section 6 concludes.

2. Finite Sample Behavior of k -class Estimators

We consider the following specification for a system of equations with two endogenous variables y_1 and y_2 :

$$(2.1) \quad \begin{aligned} y_1 &= \beta y_2 + \epsilon = \beta z\pi + w \\ y_2 &= z\pi + v, \end{aligned}$$

where $w = \beta v + \epsilon$. We let y_1 and y_2 be $N \times 1$ vectors of observations on the endogenous variables and denote by z the $N \times K$ matrix of observations on the predetermined (or exogenous) variables. The first equation is the (structural) equation of interest which measures the association between the two endogenous variables while the second equation is more commonly referred to as “the first stage regression”. The number of instruments is $\dim(\pi) = K$.

This system of equations admits the reduced form:

$$(2.2) \quad (y_1 \ y_2) = z(\beta\pi \ \pi) + (w \ v),$$

where the rows of $(w \ v)$ are independently and identically distributed as $N(0, \Omega)$ for

$$(2.3) \quad \Omega = \begin{pmatrix} \sigma_{ww} & \sigma_{wv} \\ \sigma_{wv} & \sigma_{vv} \end{pmatrix}.$$

Furthermore, we denote the variance of the error term in the structural equation by $\sigma_{\epsilon\epsilon}$ and the covariance between the structural equation for y_1 and the reduced form equation for y_2 by $\sigma_{\epsilon v}$.

An important statistic for the finite sample behavior of IV estimators is the concentration parameter μ^2 , where

$$(2.4) \quad \mu^2 = \frac{\pi' z' z \pi}{\sigma_{vv}}.$$

The concentration parameter is a measure of non-centrality and will play a quintessential role in the bias expressions for IV estimators we shall encounter throughout this paper. It is occasionally informative to impose the normalization $\sigma_{\epsilon\epsilon} = \sigma_{vv} = 1$. Using this normalization, a simple approximation of μ^2 is given by:

$$(2.5) \quad \mu^2 \approx \frac{E(\hat{\pi}' z' z \hat{\pi})}{E(\hat{v}' \hat{v}) / (N - K)} = \frac{(N - K) R^2}{1 - R^2},$$

where R^2 is the (theoretical) first stage coefficient of determination. Note that there is a close connection between the expression above and the first stage F -statistic implying that $\mu^2 \approx KF$ (Buse, 1992 and Hansen, Hausman and Newey, 2008).

Before proceeding we need to define the *confluent hypergeometric function* ${}_1F_1$ given by the expansion:

$$(2.6) \quad {}_1F_1(a; b; c) = \sum_{j=1}^{\infty} \frac{(a)_j c^j}{(b)_j j!}.$$

This function is one of the independent solutions to the differential equation $cd^2q/dc^2 + (b - c)dq/dc - ac = 0$ and is commonly used to characterize the non-central χ^2 distribution. As we shall see, it emerges in the expression for the bias of the 2SLS estimator below. Note that the confluent hypergeometric function is defined in terms of Pochhammer's symbol $(a)_j$ corresponding to the ascending factorial (Abadir, 1999):

$$(2.7) \quad (a)_j = \prod_{k=0}^{j-1} (a + k) = a(a + 1)(a + 2) \dots (a + j - 1) \quad \text{for } (a)_0 = 1.$$

The series expansion of the confluent hypergeometric function ${}_1F_1(a; b; c)$ converges for all finite c and $a > 0$ and $b > 0$ (Muller, 2001). While analytically attractive, its computation presents a number of numerical challenges for some regions of the parameter space (a, b, c) . In particular,

the efficient computation of hypergeometric functions has to be addressed with care (Koev and Edelman, 2006).

If we define the commonly used projection operators $P_z = z'(z'z)^{-1}z'$ and $Q_z = I - P_z$ we can derive the 2SLS estimator of β as:

$$(2.8) \quad \hat{\beta}_{2SLS} = \frac{y_2' P_z y_1}{y_2' P_z y_2}.$$

Recently, econometricians have devoted considerable attention to the behavior of this estimator in the presence of weak and/or many instruments (see Poskitt and Skeels, 2013 for a survey, as well as Nelson and Startz, 1990; Staiger and Stock, 1997; Hahn and Hausman, 2002a; Hahn, Hausman and Kuersteiner, 2004; Hansen, Hausman and Newey, 2008; Hirano and Porter, forthcoming).

To better understand the issues involved let us consider the exact finite sample bias of the 2SLS estimator, as derived by Richardson (1968) and Richardson and Wu (1971):

$$(2.9) \quad E(\hat{\beta}_{2SLS}) = \beta - \left(\beta - \frac{\sigma_{wv}}{\sigma_{vv}} \right) \exp\left(-\frac{\mu^2}{2}\right) {}_1F_1\left(\frac{K}{2} - 1; \frac{K}{2}; \frac{\mu^2}{2}\right).$$

Equivalent expressions are also found in Chao and Swanson (2003) and Ullah (2004). Note in particular that since $\sigma_{wv} = \beta\sigma_{vv} + \sigma_{ev}$, we have

$$(2.10) \quad E(\hat{\beta}_{2SLS} - \beta) = \frac{\sigma_{ev}}{\sigma_{vv}} \exp\left(-\frac{\mu^2}{2}\right) {}_1F_1\left(\frac{K}{2} - 1; \frac{K}{2}; \frac{\mu^2}{2}\right).$$

The exact finite sample bias of 2SLS is linear in the correlation between the structural equation and the first stage, σ_{ev} , and nonlinear in the concentration parameter μ^2 and the number of instruments K . Let us now direct our attention to the non-linear part of the bias expression which we shall denote by $B(\mu^2, K) = \exp(\mu^2/2) {}_1F_1(K/2 - 1; K/2; \mu^2/2)$.

In Figure 1, we plot the term $B(\mu^2, K)$ for a range of parameters $\mu^2 \in [0, 70]$ and $K \in \{1, 2, 3, 5, 15, 100\}$ in order to illustrate the effect of weak instruments (low μ^2) and many instruments (high K) on the bias of the 2SLS estimator.¹ Many of the simulations herein hold μ^2 fixed while varying K . Notice that the expression for B does not depend on additional sampling information beyond μ^2 and K . For all values of K the bias term B goes to 0 for values of $\mu^2 > 60$. For low values of μ^2 the weak instruments bias is compounded in the presence of many instruments. The bias resulting from the use of many instruments is eliminated by high values of the concentration parameter. As Equation 2.4 shows, for a given sample size we obtain a high value of the concentration parameter if the value of the first stage R^2 is large. Many instruments can thus be safely used if the first stage is strongly identified by those instruments.

¹Note that in general, the concentration parameter is weakly increasing in K and unless proper normalizations are made to the coefficients, increasing K for a fixed μ^2 shrinks the reduced-form variance. See Section 5 for more details on our simulations data-generating process.

Furthermore, notice that for any given value of the concentration parameter the bias is increasing with the number of instruments (consistent with Owen, 1976), but nevertheless, it remains the case that the bias term B is bounded and less than 1. For $K > 1$ the bias term B is always positive and decreases monotonically with μ^2 .² The just-identified case allows for both positive and negative bias. In fact, it appears that for most cases with a low first stage R^2 it is safer to choose two weak instruments than a single one. The bias does however increase rapidly as one adds further weak instruments. Recall however that the exact final sample bias for 2SLS is scaled by $\frac{\sigma_{ev}}{\sigma_{vv}}$ and thus the extent of the correlation between the first stage and the structural equation, σ_{ev} plays a central role.

It is possible to derive the exact bias expression for the 2SLS estimator for the over-identified case with $K = 2$. In this case we have

$$(2.11) \quad E(\hat{\beta}_{2SLS} - \beta) = \frac{\sigma_{ev}}{\sigma_{vv}} \exp\left(-\frac{\mu^2}{2}\right),$$

which shows that the bias decreases exponentially with the concentration parameter, thus emphasizing the importance of the first stage fit. While no equally simple expression can be derived for the just-identified case with $K = 1$, we can nevertheless obtain an approximation for small values of μ^2 :

$$(2.12) \quad E(\hat{\beta}_{2SLS} - \beta) = \frac{\sigma_{ev}}{\sigma_{vv}} \exp\left(-\frac{\mu^2}{2}\right) \left(\mu^2 - \frac{\mu^4}{6}\right) + O\left(\left(\sqrt{\frac{\mu^2}{2}}\right)^5\right)$$

The additional polynomial terms in the bias expression allow for both positive and negative bias at different levels of the concentration parameter.

We can ask how big the bias due to weak and/or many instruments can be by looking at the limits of the term $B(\mu^2, K)$. The bias is strictly increasing in the number of instruments K . Furthermore, we can derive the following limits:

$$(2.13) \quad \lim_{\mu^2 \rightarrow 0} B(\mu^2, K) = \lim_{K \rightarrow \infty} B(\mu^2, K) = 1.$$

This implies that the maximum finite sample bias for weak and/or many instruments cannot exceed σ_{ev}/σ_{vv} . Hahn and Hausman (2002b) show that under no identification $\pi = 0$ the bias of the OLS estimator is approximately equal to σ_{ev}/σ_{vv} . Thus, the weak instruments bias of the 2SLS estimator cannot exceed the corresponding OLS bias.

Nagar (1959) derived a large N expansion of the bias in a fairly complicated fashion, while a simpler derivation was given in Hahn and Hausman (2002b). In fact, the same second order approximation

²Note that these numerical results are consistent with earlier analytical work by Richardson and Wu (1971) and Chao and Swanson (2007). In particular, see Proposition 3.1 of Chao and Swanson (2007).

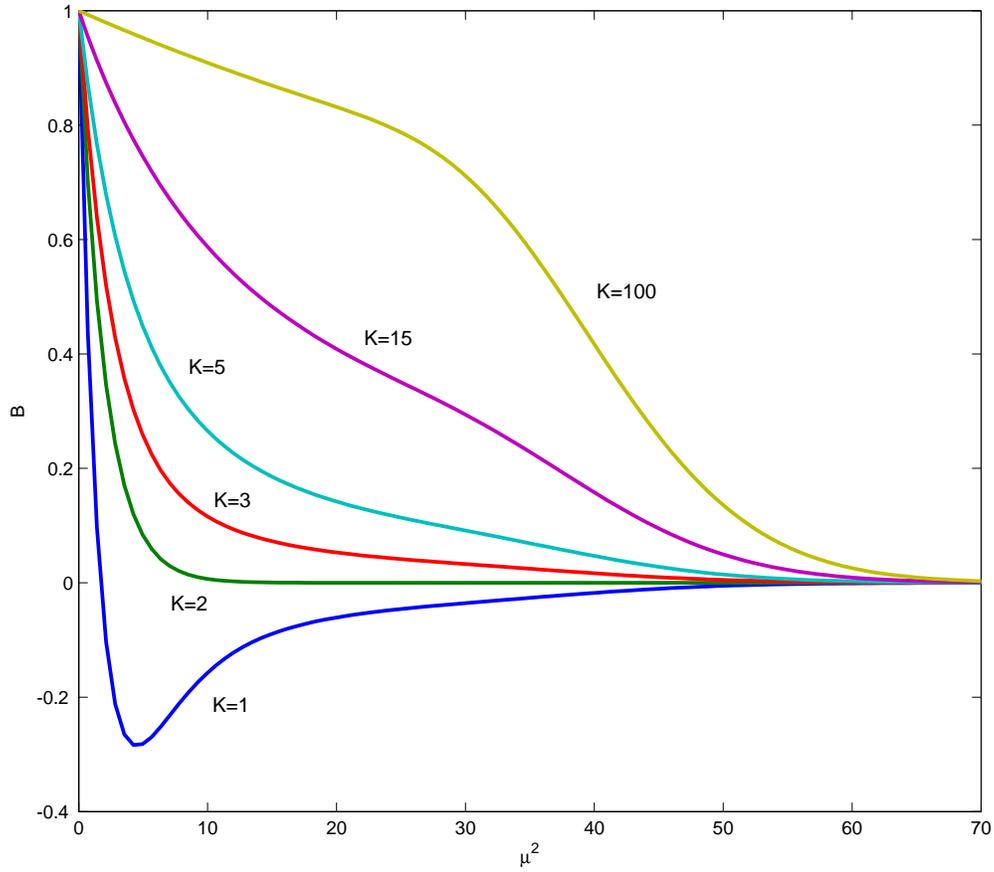


FIGURE 1. Finite Sample Bias Term B as a Function of the Concentration Parameter μ^2 for Different Numbers of Instruments K

to the bias can be directly obtained from the properties of the bias term $B(\mu^2, K)$ above. Let us consider a large μ^2 approximation of B . Using expression 13.1.4 in Slater (1964) we have:

$$(2.14) \quad B(\mu^2, K) = 2 \frac{\Gamma(K/2)}{\Gamma(K/2 - 1)} \left(\frac{1}{\mu^2} \right) + O \left(\left(\frac{\mu^2}{2} \right)^{-2} \right).$$

But since $\Gamma(K/2) = (K/2 - 1)\Gamma(K/2 - 1)$ we have

$$(2.15) \quad B(\mu^2, K) = \frac{K - 2}{\mu^2} + O \left(\left(\frac{\mu^2}{2} \right)^{-2} \right).$$

Returning to the 2SLS bias expression we obtain:

$$(2.16) \quad E(\hat{\beta}_{2SLS} - \beta) \approx \frac{\sigma_{ev}}{\sigma_{vv}} \frac{K-2}{\mu^2} = \frac{\sigma_{ev}}{\sigma_{vv}} \frac{(K-2)(1-R^2)}{(N-K)R^2},$$

where R^2 corresponds to the R^2 of the first stage regression. This expression is basically the same as the approximate bias expression given in Hahn and Hausman (2002 a,b) and which can be employed to derive a bias corrected estimator:

$$(2.17) \quad \beta_{BC} = \frac{\hat{\beta}_{2SLS} - \frac{\sigma_{wv}}{\sigma_{vv}} B(\mu^2, K)}{1 - B(\mu^2, K)}.$$

In order to operationalize this estimator one would additionally require an estimate of the concentration parameter μ^2 . Since in finite samples this estimator itself will be biased, Hansen, Hausman and Newey (2008) suggest the unbiased estimator $\mu_*^2 = \hat{\mu}^2 - K$.

Alternatively, we can proceed from equation 2.10 and employ the approach of Hahn and Hausman (2002b) by first noting that:

$$(2.18) \quad \sigma_{ev} = E\left[\frac{1}{N-K} (y_2' Q_z)(y_1 - y_2 \beta)\right].$$

Then we have,

$$(2.19) \quad E(\beta_{2SLS}) = \beta + E\left[\frac{N-K}{y_2' Q_z y_2} \left(\frac{1}{N-K} y_2' Q_z y_1 - \frac{1}{N-K} y_2' Q_z y_2 \beta\right) B(\mu^2, K)\right]$$

$$(2.20) \quad E(\beta_{2SLS}) = E\left[\beta + \frac{y_2' Q_z y_1}{y_2' Q_z y_2} B(\mu^2, K) - \beta B(\mu^2, K)\right].$$

Hence we can solve for β as:

$$(2.21) \quad \beta_{BC} = \frac{y_2' P_z y_1 - \kappa y_2' Q_z y_1}{y_2' P_z y_2 - \kappa y_2' Q_z y_2},$$

where

$$(2.22) \quad \kappa = B(\mu^2, K) \frac{y_2' P_z y_2}{y_2' Q_z y_2} = B(\mu^2, K) \frac{\mu^2}{N-K} = B(\mu^2, K) \frac{R^2}{1-R^2}.$$

This estimator has zero finite sample bias by construction. Its actual sample performance however will depend on the extent to which we can accurately determine the concentration parameter in a given sample. The performance of this particular estimator was found to be unreliable in simulations—it greatly outperformed all other estimators apart from infrequent, extreme outliers.

Still, it is worth noticing that the finite sample bias corrected estimator derived above is a member of the Nagar (1959) k -class family of estimators, which generalizes the 2SLS estimator.³

Recent work has devoted significant attention to the behavior of k -class estimators. This remarkable class of estimators includes many of the commonly used estimators for simultaneous equations. Thus, using the notation introduced above we obtain the k -class estimator β_{kIV} as:

$$(2.23) \quad \beta_{kIV} = \frac{y_2' P_z y_1 - k y_2' Q_z y_1}{y_2' P_z y_2 - k y_2' Q_z y_2},$$

for different values of the parameter k . By setting $k = -1$ we obtain the OLS estimator and $k = 0$ leads to the 2SLS estimator. But other leading estimators can be obtained as a member of the k -class. Thus, LIML corresponds to the estimator with $k = \phi$ where ϕ is set to the smallest eigenvalue of $W' P_z W (W' Q_z W)^{-1}$ for $W = (y_1, y_2)$. The estimator of Fuller (1977) corresponds to $k = \phi - 1/(N - K)$ while the Nagar (1959) estimator corresponds to $k = (K - 2)/(N - K)$.⁴

Sawa (1972) shows that for $k \leq 0$ the k class estimators have moments while for $k > 0$ they may not even have the first moment. Unfortunately, the estimators we are interested in such as LIML, Nagar or the exact finite sample unbiased estimator derived above all have values of k greater than 0.⁵ While these estimators have the potential to remove the 2SLS finite sample bias by adjusting k away from 0, their practical usefulness is stymied by their lack of moments. As Fuller (1977) realized, however, there is a small region of values of $k > 0$ for which the estimators do in fact have moments. An estimator using a value of $k > 0$ but less than the values of k prescribed by LIML or Nagar will remove some of the finite sample bias while retaining moments. This explains the popularity of Fuller as a solution to the weak and/or many instruments problem.

We do, however, face the question as to how useful k -class estimators ultimately are given that we face a trade-off between bias reduction and the lack of moments. In order to gain further insight into this problem we introduce a simple Monte Carlo simulation design for the model in equations 2.1. For a choice of values for the parameters N , K , ρ and R^2 we simulate the model 50,000 times and estimate the structural parameter β using all the k -class estimators for values of k between

³Simulation results for this estimator are not reported in this paper, but are available from the authors. A two-part estimator that uses the β_{BC} when the bias-corrected concentration parameter estimate is positive ($\hat{\mu}^2 - K > 0$) and Fuller otherwise significantly outperforms other estimators in both mean bias and MSE.

⁴See Hahn, Hausman and Kuersteiner (2004) for a detailed review of the performance of these estimators and their corresponding higher-order bias and MSE approximations. Note however that there is a minor but sometimes confusing difference in the way these estimators are defined in the recent literature (including this paper) relative to the notation employed in earlier research such as Nagar (1959) or Sawa (1972) who define the k parameter as $1 + k$ for the values of k discussed above. Furthermore, it is important to keep in mind that some of the values of k are fixed while others are stochastic, which leads to subtle differences in the way we think about the properties of these estimators.

⁵See Kinal (1980) for the definitive treatment of the existence of moments for k -class estimators.

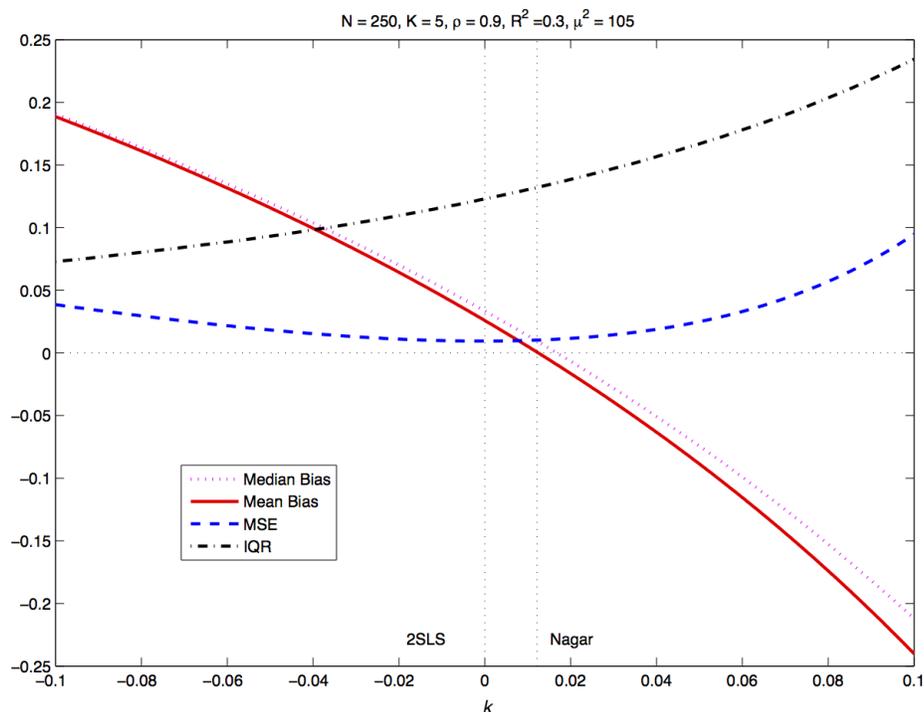


FIGURE 2. Monte Carlo simulation of k -class estimators for the “normal” case with a high value of the concentration parameter and $K = 5$ instruments

$k_l < 0 < k_u$. In practice we choose 300 values of k on the line $[k_l, k_u]$. Thus each value of k on this line defines a k class estimator which is then applied to all the simulated samples. Furthermore, for each value of k we compute the mean bias, MSE, median bias and interquartile range of the estimates for β .

In order to better understand the finite sample behavior of the k -class estimators, we plot the above summary statistics over a range of k -class estimators for three cases of interest to us: a benign case that does not exhibit weak instruments and employs only a small number of instruments, and two more pathological cases, one with weak instruments and the other with weak *and* many instruments. The figures discussed below were generated using a typical choice of parameter values, but it is easy to replicate the same qualitative properties for other parameter values falling within the same three cases of interest. Thus, in the simulations we use $N = 250$ and $\rho = 0.9$. The value of ρ is higher than what we would expect in typical applications. As we have seen from our discussion of the bias earlier on, ρ scales the bias upward. This allows us to better visualize the results in the figures below without changing their interpretation. Lower values of ρ lead to very similar results.

Let us first consider the “normal” case which does not exhibit either weak or many instruments. To simulate this case we use $K = 5$ and choose a value of $R^2 = 0.3$ for the theoretical first stage fit. This leads to a value of the concentration parameter of $\mu^2 = 105$ which we deem to be large enough to avoid the weak instruments problem. We plot our summary statistics for the k -class estimators for $k \in [-0.1, 0.1]$ in Figure 2. We notice that the mean bias and median bias are monotonically decreasing over the chosen range of k . In particular, note that at the value of $k = 0$ that corresponds to the standard 2SLS estimator, the estimator still has some small amount of finite sample bias of just under 5%. Notice however that if we choose a value of $k = 0.012$ which corresponds to the value recommended by the Nagar estimator, the bias is eliminated altogether. Furthermore, the MSE is small at values of k close to 0, but it increases as k deviates from 0 in either direction. The interquartile range increases with k over this range. Notice that for this range of values of k all the estimators have moments. However, as we increase k away from 0 both the mean bias and the MSE will diverge. This example shows that the range of values of $k > 0$ for which the estimators have moments is nevertheless large enough to allow for the unbiased estimation of the structural parameter β by employing a member of k -class such as the Nagar estimator or the exact finite sample unbiased estimator derived above. As we decrease k towards -1 which corresponds to the OLS estimator, both the mean bias and MSE will continue to increase but remain bounded.

Let us now turn our attention to the weak (but not many) instruments case. To simulate this case we use $K = 5$ and choose a value of $R^2 = 0.1$ for the theoretical first stage fit. This leads to a value of the concentration parameter of $\mu^2 = 27.2$ which we deem to be small enough to lead to the weak instruments problem. We plot our summary statistics for the k -class estimators for $k \in [-0.1, 0.1]$ in Figure 3. The 2SLS estimator at $k = 0$ now has substantial finite sample bias of about 15%. Furthermore, perhaps the most striking aspect of this figure is that the estimators lose moments for values of $k > 0$. As k increases the first moment becomes unstable (hence the oscillations visible in the numerical simulations) and eventually diverges. Similarly, the MSE explodes for the value of k used by the Nagar estimator. Notice however that if we were to choose a value of k less than that prescribed by the Nagar estimator it would still be possible to eliminate most of the finite sample bias of 2SLS. The Fuller estimator does in fact choose such a value of $k > 0$ allowing the estimator to have moments. This is the type of situation for which the Fuller estimator performs very well. Notice however that the interquartile range increases sharply with values of k , which is confirmed in simulations by the fact that the Fuller estimator exhibits a fairly wide distribution relative to the 2SLS estimator.

Finally, let us focus our attention on the weak and many instruments case. To simulate this case we use $K = 30$ and choose a value of $R^2 = 0.1$ for the theoretical first stage fit. This leads to a value of the concentration parameter of $\mu^2 = 27.2$ which we deem to be small enough to lead to the weak instruments problem. The resulting finite sample problem is amplified by the presence of a large number of instruments. We plot our summary statistics for the k -class estimators for

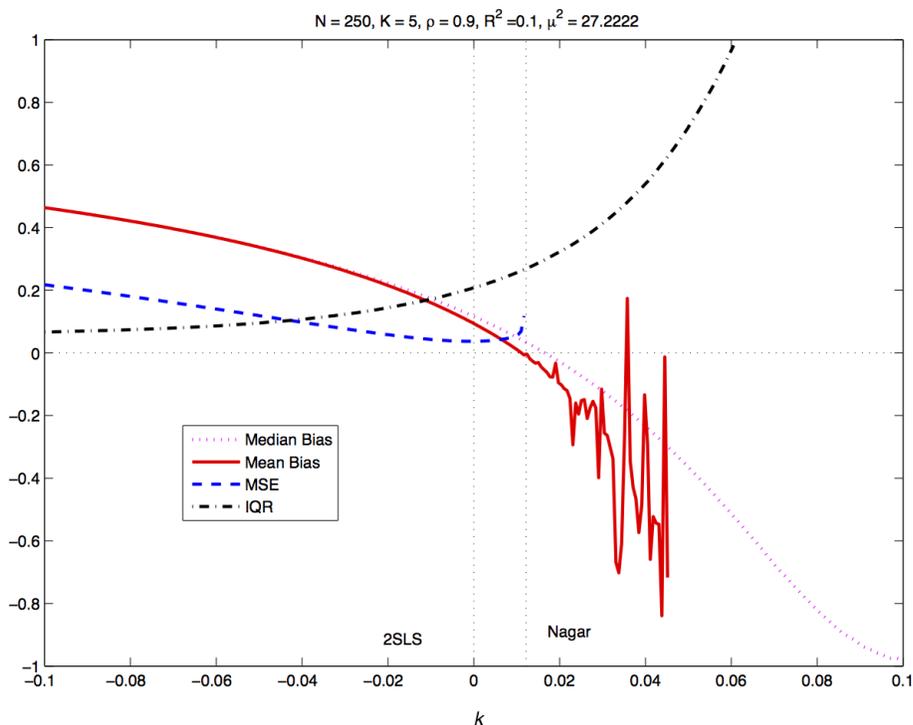


FIGURE 3. Monte Carlo simulation of k -class estimators for the weak-instruments case with a low value of the concentration parameter and $K = 5$ instruments

$k \in [-0.1, 0.15]$ in Figure 4. The 2SLS estimator at $k = 0$ now has very substantial finite sample bias of at least 45%. As we increase k above 0 the bias decreases but very soon, at value of k of about 0.06 the estimators lose moments. Both the first and second moments diverge rapidly and the interquartile range explodes too. The value of k recommended by the Nagar estimator which would be required to eliminate the bias altogether is well within the “no moments” range of values of k and thus this estimator is no help. Furthermore, notice that an estimator such as Fuller which does have moments cannot remove the bias since the range of values of k for which such a k -class estimator retains moments still corresponds to a bias of at least 30%. There is thus no range of values of k for which a k -class estimator has moments while also successfully removing the bias.

These results, while purely computational in nature, reveal a serious limitation of k -class estimators. Values of the parameters k which would be required to eliminate the finite sample bias due to many weak instruments lie outside the range of values for which this class of estimators has moments. Thus, unfortunately, no member of the k -class can achieve the dual goal of removing bias while maintaining moments.

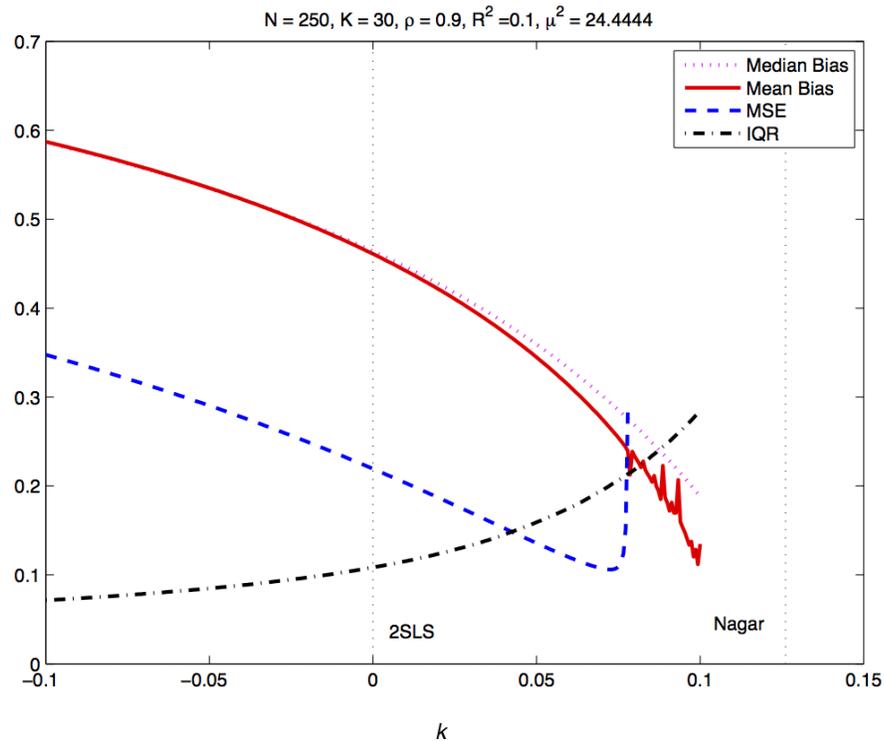


FIGURE 4. Monte Carlo simulation of k -class estimators for the many-instruments case with a low value of the concentration parameter and $K = 30$ instruments

This is a serious concern that prompts us to investigate alternatives to the k -class of estimators as a solution to the weak and many instruments problem. In the next section we explore a modification of the k -class due to Nagar (1972) as a solution to the finite sample unbiased estimation of the structural parameters in a system of simultaneous equations.

3. Finite Sample Bias Correction in the Double k -class

In the previous section we have uncovered a fundamental problem with the k -class of estimators for simultaneous equations. The use of a single parameter k to adjust for the finite sample bias in the weak and/or many instruments case is too restrictive and leads to estimators with no moments. To illustrate, consider the 2SLS case with $k = 0$, by construction the estimator has moments but also substantial finite sample bias. On the other hand, although the Nagar estimator is second-order unbiased (through an appropriate choice of $k = (K - 2)/(N - K)$), for the case where $\pi = 0$, it is easy to see that the denominator of the expression for the estimator will also equal 0 in expectation

since

$$(3.1) \quad E \left[y_2' P_z y_2 - \frac{K-2}{N-K} (y_2' Q_z y_2) \right] = (K-2)\sigma_{vv} - \frac{K-2}{N-K} (N-K)\sigma_{vv} = 0.$$

Intuitively, the k -class is too restrictive since the same parameter k enters both sides of the estimating equations. Manipulating this one parameter we cannot achieve the two goals of reducing bias and estimators with finite sample moments.

In a now largely forgotten contribution, Nagar (1962) suggested a minor change to the estimating equations of IV estimators by allowing for two distinct values of k to enter the expression for the k -class IV estimators:

$$(3.2) \quad \beta_{2kIV}(k_1, k_2) = \frac{y_2' P_z y_1 - k_2 y_2' Q_z y_1}{y_2' P_z y_2 - k_1 y_2' Q_z y_2}.$$

We denote the resulting class of estimators the double k -class indexed by (k_1, k_2) . Thus the case of $k = k_1 = k_2$ becomes a special case on which we hope to improve upon by choosing different values of k_1 and k_2 .

We can immediately rewrite Equation 3.2 as,

$$(3.3) \quad \beta_{2kIV}(k_1, k_2) = \beta_{kIV}(k_1) + (k_1 - k_2) \frac{y_2' Q_z y_1}{y_2' P_z y_2 - k_1 y_2' Q_z y_2},$$

where $\beta_{kIV}(k_1)$ is the k -class IV estimator with a value of $k = k_1$. For any given k -class estimator with $k = k_1$ we can construct a corresponding double k -class estimator by moving k_2 away from k_1 by some fixed amount. Intuitively, we can see from Equation 3.3 that for $k_1 \neq k_2$ the additional term on the right may potentially offset the finite sample bias of the k -class estimator on which it is based.

The double k -class of IV estimators has received very limited attention over the years (Nagar, 1962; Srivastava, Agnihotri and Dwivedi, 1980; Dwivedi and Srivastava, 1984; Gao and Lahiri, 2002) and to our knowledge has never been considered as an option for the estimation of systems of equations with weak and/or many instruments. The double k -class is also a generalization of a now-forgotten class of estimators proposed by Theil (1958) and called the h -class, and which appears to correspond to the case of $k_1 = h^2$ and $k_2 = h$ for a real valued parameter h .

We first wish to evaluate the finite sample performance of double k -class estimators using 2SLS as the benchmark, leaving the discussion of the fully general case for future research. Thus we will look at two particular subsets of the double k -class,

$$(3.4) \quad \beta_{2kIV}(\Delta, 0) = \frac{y_2' P_z y_1}{y_2' P_z y_2 - \Delta y_2' Q_z y_2} = \beta_{kIV}(\Delta) + \Delta \frac{y_2' Q_z y_1}{y_2' P_z y_2 - \Delta y_2' Q_z y_2}.$$

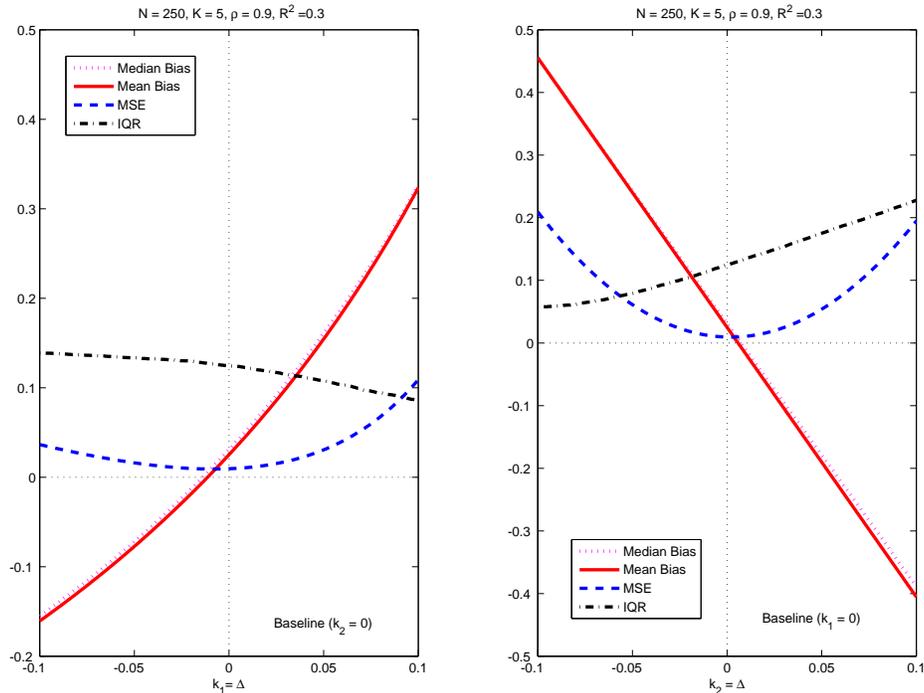


FIGURE 5. Monte Carlo simulation of double k -class estimators for the “normal” case with a high value of the concentration parameter and $K = 5$ instruments

and

$$(3.5) \quad \beta_{2kIV}(0, \Delta) = \frac{y_2' P_z y_1 - \Delta y_2' Q_z y_1}{y_2' P_z y_2} = \beta_{2SLS} - \Delta \frac{y_2' Q_z y_1}{y_2' P_z y_2}.$$

We will employ Monte Carlo simulations to show that both of these modifications of 2SLS can be successful in solving the weak and/or many instruments problem for well-chosen values of Δ . Note that these simulations are constructed under the assumption that the underlying parameters are known. In practice some of these parameters themselves need to be estimated and thus the finite sample performance of these estimators will differ. Therefore, in this section we document that a solution exists, while in the next section we discuss procedures for implementing these solutions in practice.

Let us first consider the baseline case which does not exhibit either weak or many instruments. To simulate this case we use $K = 5$ and choose a value of $R^2 = 0.3$ for the theoretical first stage fit. This leads to a value of the concentration parameter of $\mu^2 = 105$ which we deem to be large enough to avoid the weak instruments problem. We plot our summary statistics for the double k -class

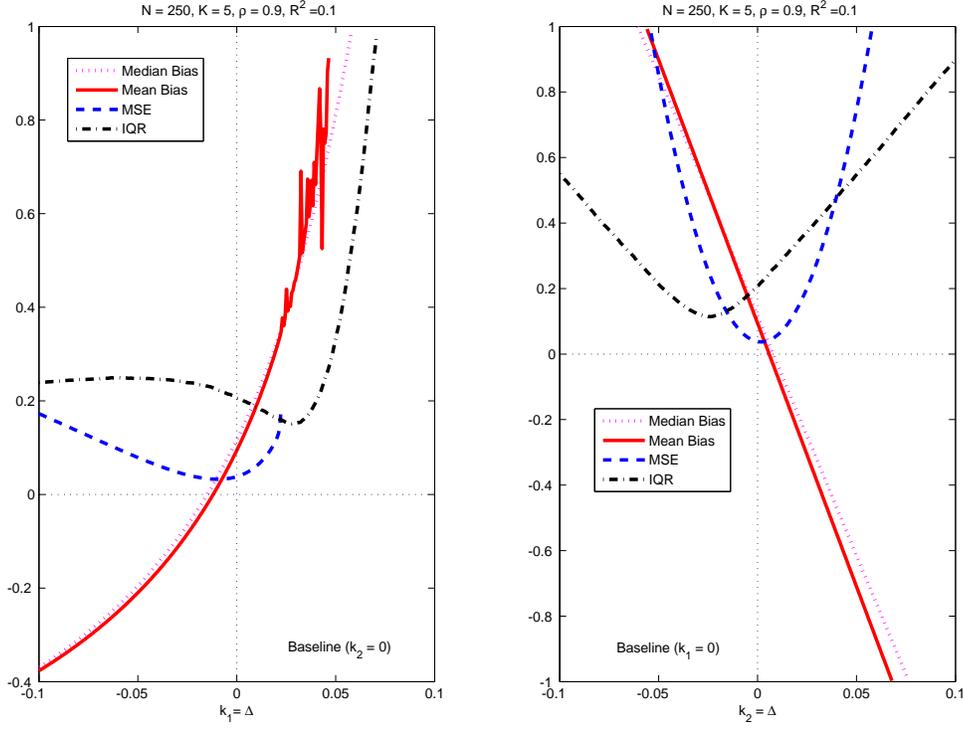


FIGURE 6. Monte Carlo simulation of double k -class estimators for the “weak instruments” case with a low value of the concentration parameter and $K = 5$ instruments

estimators in Figure 5. In the left panel we plot our summary statistics for the double k -class estimator with $k_1 = \Delta$ and $k_2 = 0$ (Equation 3.4), while in the second panel we plot the statistics for the double k -class estimators with $k_1 = 0$ and $k_2 = \Delta$ (Equation 3.5). We choose a range of $\Delta \in [-0.1, 0.1]$. We notice that for $\Delta = 0$ which corresponds to the 2SLS case we have a small amount of finite sample bias. This finite sample bias can be removed by choosing a small negative value of k_1 in the left panel or a small positive value of k_2 in the second panel. Furthermore, this can be achieved at a modest cost in MSE, which increases slowly as Δ increases or decreases away from 0. It appears that the resulting estimators maintain moments over the the chosen range on Δ . Notice that a particularly attractive feature of the estimator in Equation 3.5 is that the bias is linear in Δ while the MSE is quadratic.

Let us now turn our attention to the “weak instruments” case which does not employ too many instruments. To simulate this case we use $K = 5$ and choose a value of $R^2 = 0.1$ for the theoretical first stage fit. This leads to a value of the concentration parameter of $\mu^2 = 27.2$ which we deem

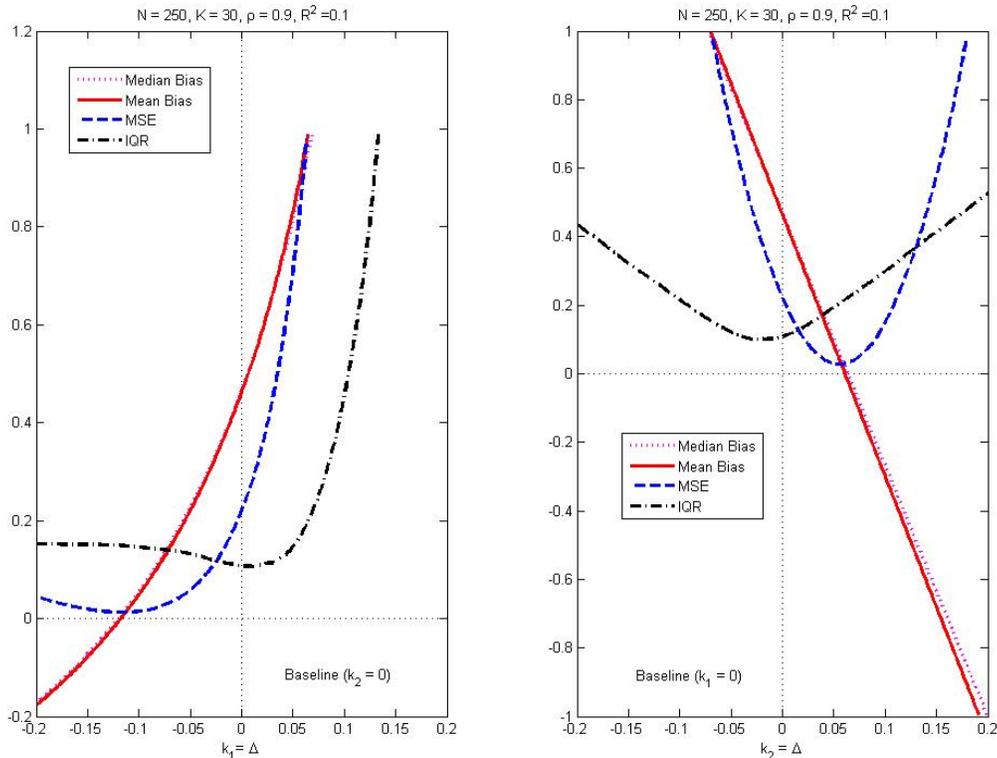


FIGURE 7. Monte Carlo simulation of double k -class estimators for the “many instruments” case with a low value of the concentration parameter and $K = 30$ instruments

to be small enough to lead to the weak instruments problem. We plot our summary statistics for the double k -class estimators in Figure 6. In the left panel we plot our summary statistics for the double k -class estimator with $k_1 = \Delta$ and $k_2 = 0$ (Equation 3.4), while in the second panel we plot the statistics for the double k -class estimators with $k_1 = 0$ and $k_2 = \Delta$ (Equation 3.5). We choose a range of $\Delta \in [-0.1, 0.1]$. We notice that for $\Delta = 0$ which corresponds to the 2SLS case we now have a substantial amount of finite sample bias. This finite sample bias can be removed by choosing a small negative value of k_1 in the left panel or a small positive value of k_2 in the second panel. Furthermore, this can be achieved at a modest cost due to increasing MSE. Notice that while the estimators in the right panel maintain moments for all values of Δ , those in the left panel do not for $\Delta > 0$. It is important to note, however, that the values of Δ required to address the finite sample bias of 2SLS are such that $\Delta < 0$ and thus we avoid the moment problem as well.

Finally, let us return to the “weak and many instruments” case. As before, we simulate this case using $K = 30$ and set the theoretical first stage $R^2 = 0.1$. We plot our summary statistics for the double k -class estimators in Figure 7. In the left panel we plot our summary statistics for the double k -class estimator with $k_1 = \Delta$ and $k_2 = 0$ (equation 3.4), while in the second panel we plot the statistics for the double k -class estimators with $k_1 = 0$ and $k_2 = \Delta$ (Equation 3.5). We choose a range of $\Delta \in [-0.1, 0.1]$. We note that for $\Delta = 0$, which corresponds to the 2SLS case, we now have a very substantial amount of finite sample bias. This finite sample bias can be removed by choosing a small negative value of k_1 in the left panel or a small positive value of k_2 in the second panel. Furthermore, this can be achieved at a modest cost due to increasing MSE.

Qualitatively, these three cases are very similar in the sense that for a fixed (finite) sample size, we can find values of Δ for each case such that the bias is completely eliminated while avoiding the no-moments problems of the estimators in the k -class. Having shown that the double k -class estimators can provide a satisfactory solution to the weak and/or many instruments problem of the baseline 2SLS estimator in theory, we now need to discuss the choice of Δ .

4. Optimal Parameter Choice for Double k -class Estimators

Dwivedi and Srivastava (1984) derive the exact finite sample bias expression for the double k -class estimators. Define $m = N/2$ and $n = (N - K)/2$. As before, denote by μ^2 the first stage concentration parameter. Then for non-negative integers a, b, c, d we can define,

$$(4.1) \quad \phi(a, b) = \exp\left(\frac{-\mu^2}{2}\right) \sum_{j=0}^{\infty} \frac{\Gamma(m - n + j - a) (\mu^2/2)^j}{\Gamma(m - n + j + b) j!}$$

and

$$(4.2) \quad \psi_d(a; b; c) = \exp\left(\frac{-\mu^2}{2}\right) \sum_{\alpha=0}^{\infty} \sum_{j=0}^{\infty} (d\alpha + 1)(k_1 + 1)^\alpha \frac{\Gamma(m + j - a - 1)\Gamma(n + \alpha + b) (\mu^2/2)^j}{\Gamma(m + j + \alpha + c)\Gamma(n) j!}.$$

Then, Dwivedi and Srivastava (1984) show that the bias of the double k -class is,

$$(4.3) \quad E(\hat{\beta}_{2kIV}) - \beta = \left(\beta - \frac{\sigma_{wv}}{\sigma_{vv}}\right) \left(\frac{\mu^2}{2} \psi_0(1; 0; 1) - 1\right) + (k_1 - k_2) \frac{\sigma_{wv}}{\sigma_{vv}} \psi_0(1; 0; 1).$$

Furthermore, if $K > 1$ and $k_1 = 0$ we also have,

$$(4.4) \quad E(\hat{\beta}_{2kIV}) - \beta = \left(\beta - \frac{\sigma_{wv}}{\sigma_{vv}}\right) \left(\frac{\mu^2}{2} \phi(0; 1) - 1\right) - k_2 \frac{\sigma_{wv}}{\sigma_{vv}} n \phi(1; 0).$$

4.1. Mean Unbiased Estimation

Using the properties above, for a given value of k_1 we can derive the value of k_2 such that the resulting double k -class estimator is unbiased. If $k_1 \neq 0$, then,

$$(4.5) \quad k_2^* = k_1 + \frac{\sigma_{vv}}{\sigma_{wv}} \left(\beta - \frac{\sigma_{wv}}{\sigma_{vv}} \right) \left[\frac{\frac{\mu^2}{2} \psi_0(1; 0; 1) - 1}{\psi_0(1; 0; 1)} \right],$$

while if $k_1 = 0$ this reduces to,

$$(4.6) \quad k_2^* = \frac{\sigma_{vv}}{\sigma_{wv}} \left(\beta - \frac{\sigma_{wv}}{\sigma_{vv}} \right) \left[\frac{\frac{\mu^2}{2} \phi(0; 1) - 1}{n\phi(1; 0)} \right].$$

Then, using the fact that

$$(4.7) \quad - \left[\frac{\frac{\mu^2}{2} \phi(0; 1) - 1}{n\phi(1; 0)} \right] = \frac{K - 2}{N - K},$$

we have a simplified version of k_2^* that does not require estimating μ^2 :

$$(4.8) \quad k_2^* = - \left(\frac{K - 2}{N - K} \right) \frac{\sigma_{vv}}{\sigma_{wv}} \left(\beta - \frac{\sigma_{wv}}{\sigma_{vv}} \right)$$

Notice two facts about the optimal choice of the k_2 . First, for a fixed K , k_2^* tends to 0 as $N \rightarrow \infty$, a requirement for consistency. Second, k_2^* depends on the unknown value of the structural parameter β . This suggests a two-step estimation approach in order to operationalize this estimator. In the tables, we denote this as the “2-step k_2^* ” estimator, which simply estimates σ_{vv} , σ_{wv} from the reduced form and β by Fuller:

$$(4.9) \quad \hat{k}_2^{2-step} = - \left(\frac{K - 2}{N - K} \right) \frac{\hat{\sigma}_{vv}}{\hat{\sigma}_{wv}} \left(\hat{\beta}_{Fuller} - \frac{\hat{\sigma}_{wv}}{\hat{\sigma}_{vv}} \right)$$

The resulting estimator is the double k -class estimator with $k_1 = 0$ and $k_2 = \hat{k}_2^{2-step}$.

4.2. Minimum MSE Estimation

Alternatively, we can choose the value of k_2 which minimizes the MSE . Using the expressions in Dwivedi and Srivastava (1984) we can define,

$$(4.10) \quad \bar{\sigma} = \sigma_{wv} - \frac{\sigma_{wv}^2}{\sigma_{vv}}.$$

Then the parameter choice which minimizes the MSE of the estimator is given by

$$(4.11) \quad k_2^* = \frac{\frac{\sigma_{wv}}{\sigma_{vv}} \left(\beta - \frac{\sigma_{wv}}{\sigma_{vv}} \right) \left(\frac{\mu^2}{2} \phi(1; 1) - \phi(1; 0) \right)}{\frac{\bar{\sigma}}{4\sigma_{vv}} \phi(2; 0) + \frac{n+1}{2} \left(\frac{\sigma_{wv}}{\sigma_{vv}} \right)^2 \phi(1; 0)}$$

Again, note that the optimal choice of k_2 depends on unknown parameters β , σ_{vv} , and σ_{wv} and additionally on μ^2 . Mirroring the solution strategy discussed before we can construct a feasible two-step estimator by plugging in the Fuller estimator of β (this time setting the Fuller parameter $\kappa = 4$ so that $k^{Fuller} = \phi - 4/(N - K)$ to minimize MSE) and the using reduced-form estimates of the remaining parameters (bias correcting μ^2 by K) to obtain $\hat{k}_2^{2-step}(\kappa = 4)$.

5. Monte Carlo Simulations

We now explore the finite sample performance of the two feasible double k -class estimators introduced above that either minimize the bias or MSE. Since these choices are functions of the unknown structural parameter β , the feasible estimators plug in an initial estimate obtained from Fuller's estimator.

Following the simulation design in Hahn and Hausman (2002a), we generate data using the following design with two endogenous variables y_1 and y_2 :

$$(5.1) \quad \begin{aligned} y_1 &= \beta y_2 + \epsilon = \beta z\pi + w \\ y_2 &= z\pi + v, \end{aligned}$$

where $w = \beta v + \epsilon$. We let y_1 and y_2 be $N \times 1$ vectors of observations on the endogenous variables and denote by z the $N \times K$ matrix of K instruments.

If we define:

$$(5.2) \quad \Omega = \begin{pmatrix} \sigma_{ww} & \sigma_{wv} \\ \sigma_{wv} & \sigma_{vv} \end{pmatrix},$$

then our design has $\sigma_{vv} = \sigma_{ww} = 1$ and we let $\sigma_{wv} = -0.3$. The parameter of interest β is given by $\beta = 2\sigma_{wv}$, which pins down $\sigma_{\epsilon v} = 0.3$. Each element of $\pi = \sqrt{\frac{\mu^2}{(N-K)K}}$, where μ^2 corresponds to the theoretical concentration parameter

$$(5.3) \quad \mu^2 = \frac{\pi' z' z \pi}{\sigma_{vv}}.$$

In the Monte Carlo experiments below we consider permutations of $\mu^2 \in \{8, 12, 24, 32\}$, $K \in \{8, 24\}$, and $N \in \{200, 800\}$. Both of the sample sizes we consider are quite small, highlighting the problems with traditional IV estimators in finite samples. The various values of μ^2 and K allow us to consider the case of weak and many instruments, respectively, potentially occurring simultaneously. We use a log normal distribution for the instrumental variables standardized with mean zero and variance one, and the stochastic disturbance is either normally distributed or has a t -distribution with 12 degrees of freedom, again standardized to have mean zero and variance one.

Table 1 reports Monte Carlo simulation results for the data-generating process with normally distributed disturbances. The top panel reports results for $N = 800$ and the bottom panel reports

results for $N = 200$, while the left-hand columns display mean bias results and the right-hand columns indicate the MSE of each estimator. Given that the true coefficient is $\beta = -0.6$, the 2SLS mean bias results are large. For example, the mean bias of 2SLS for weak but not many instruments ($\mu^2 = K = 8$) is 0.147, which represents a bias of nearly 25%. Fuller1 (which denotes Fuller’s k -class estimator with $k = \phi - 1/(N - K)$) consistently outperforms 2SLS in mean bias, sometimes by an order of magnitude. Our preferred estimator to minimize mean bias is labeled “2-step k_2^* min bias” and is defined in equation (4.9) above. This estimator uses the Fuller1 estimate of β to construct k_2^* and does not require estimating μ^2 . In mean bias, this estimator generally has almost 50% less mean bias than Fuller1, especially in the rows corresponding to weak instruments, and failing to have lower mean bias than Fuller1 only in a few specifications.

Turning to minimum MSE estimation, the Fuller estimator that has been shown to perform best in terms of MSE is $k = \phi - 4/(N - K)$, and we benchmark our results against this estimator, which we label Fuller4. As a general rule, 2SLS has low MSE because its variance is so small despite its location being badly biased. Fuller4, by contrast, has mean bias that is around half of the mean bias of 2SLS but MSE that is up to 20% higher than the MSE of 2SLS. Our preferred estimator for minimizing MSE is labeled “2-step k_2^* min MSE.” While this estimator dominates 2SLS in terms of mean bias, it generally has similar or higher mean bias as Fuller4. The MSE comparison between Fuller4 and this optimal double k -class estimator is more nuanced. When many instruments is an issue (i.e. $K = 24$), the feasible double k -class estimator has MSE nearly 50% less than Fuller4. When many instruments are not a problem (i.e. in the $K = 8$ specifications), the two-step double k -class estimator has mean similar or slightly higher MSE.

To test whether these results rely on the relatively thin tails of the normal distribution, in Table 2, we show Monte Carlo results when the stochastic disturbances have thicker tails, which we simulate by drawing from a t -distribution with 12 degrees of freedom. The results are broadly consistent with Table 1. 2SLS is badly biased, and Fuller1 shows tremendous finite sample bias reduction. Our preferred estimator for minimizing bias almost always dominates Fuller1 in mean bias. While Fuller4 offers a nice alternative to 2SLS in terms of minimizing MSE (mainly because Fuller4 is less badly biased than 2SLS), when many instruments are an issue, the feasible double k -class estimator that minimizes MSE outperforms Fuller4 in MSE.

6. Conclusion

In this paper, we present new tools for addressing the finite sample bias of instrumental variable estimators in the presence of weak and/or many instruments. Currently practitioners have a number of estimators at their disposal, but the most commonly used ones such as 2SLS suffer from substantial bias. Fuller (1977) performs reasonably well in the case of weak and many instruments but is not unbiased.

TABLE 1. Monte Carlo Results for Normally Distributed Errors

		<u>Mean Bias</u>					<u>MSE</u>				
		I. Results with $N = 800$					II. Results with $N = 200$				
μ^2	K	2-step k_2^*		2-step k_2^*		2-step k_2^*		2-step k_2^*		2-step k_2^*	
		2SLS	Fuller1	min bias	Fuller4	min MSE	2SLS	Fuller1	min bias	Fuller4	min MSE
8	8	0.147	0.048	0.027	0.122	0.111	0.088	0.246	0.313	0.106	0.141
8	24	0.224	0.076	0.051	0.141	0.205	0.080	0.438	0.564	0.174	0.093
12	8	0.115	0.015	0.008	0.083	0.088	0.066	0.163	0.178	0.080	0.089
12	24	0.199	0.034	0.016	0.099	0.180	0.066	0.301	0.351	0.132	0.072
24	8	0.068	-0.005	0.001	0.037	0.059	0.037	0.064	0.061	0.043	0.040
24	24	0.148	-0.002	-0.005	0.042	0.135	0.041	0.111	0.109	0.066	0.042
32	8	0.053	-0.005	0.001	0.026	0.048	0.029	0.043	0.041	0.033	0.030
32	24	0.126	-0.003	-0.002	0.029	0.117	0.033	0.066	0.063	0.046	0.033

Notes: Table shows mean bias (left-hand columns) and MSE results (right-hand columns) for estimation of β from 10,000 Monte Carlo simulations using the estimators described in the text. Panel I reports results for $N = 800$ and panel II reports results for $N = 200$. The data generating process is described in the text with the true $\beta = -0.6$ and the true $\sigma_{\varepsilon} = 0.3$.

We reintroduce Nagar's double k -class estimators to construct optimal estimators that minimize mean bias or MSE. To do so, we derive the exact finite sample bias of the estimators in terms of hypergeometric functions and employ recent computational advances in the speed with which hypergeometric functions can be evaluated to develop a computationally attractive estimator that weakly dominates k -class estimators, including 2SLS. Monte Carlo results demonstrate that in simulations our estimator has lower mean bias than Fuller and does comparably well for MSE, outperforming Fuller in terms of MSE in the case of many instruments. Additional research may suggest further solutions to the problem of weak and many instruments in the double k -class of estimators, including generalizing our estimator to accommodate conditional heteroskedasticity or multiple endogenous regressors, and determining whether the optimal double k -class estimators proposed here also remove asymptotic bias under many/weak instrument (Bekker) asymptotics.

TABLE 2. Monte Carlo Results for Thick-Tailed Errors

		Mean Bias					MSE				
		I. Results with $N = 800$					II. Results with $N = 200$				
μ^2	K	2SLS	Fuller1	2-step k_2^* min bias	Fuller4	2-step k_2^* min MSE	2SLS	Fuller1	2-step k_2^* min bias	Fuller4	2-step k_2^* min MSE
8	8	0.152	0.053	0.034	0.126	0.116	0.091	0.246	0.339	0.106	0.152
8	24	0.225	0.077	0.055	0.142	0.207	0.081	0.447	0.577	0.177	0.095
12	8	0.120	0.022	0.015	0.089	0.093	0.068	0.165	0.196	0.081	0.097
12	24	0.199	0.032	0.016	0.098	0.181	0.066	0.312	0.362	0.135	0.073
24	8	0.073	0.000	0.005	0.041	0.063	0.037	0.066	0.064	0.044	0.041
24	24	0.148	-0.007	-0.008	0.039	0.135	0.042	0.117	0.114	0.068	0.042
32	8	0.057	0.000	0.005	0.030	0.052	0.029	0.044	0.042	0.033	0.030
32	24	0.126	-0.007	-0.005	0.026	0.116	0.033	0.071	0.066	0.048	0.033

Notes: Table shows mean bias (left-hand columns) and MSE results (right-hand columns) for estimation of β from 10,000 Monte Carlo simulations using the estimators described in the text. Panel I reports results for $N = 800$ and panel II reports results for $N = 200$. The data generating process is described in the text with thick-tailed errors (a t -distribution with 12 degrees of freedom), the true $\beta = -0.6$ and true $\sigma_{ev} = 0.3$.

References

- ABADIR, K. (1999): "An Introduction to Hypergeometric Functions for Economists," *Econometric Reviews*, 18, 287–330.
- ANDREWS, I., AND T. B. ARMSTRONG (2015): "Unbiased Instrumental Variables Estimation Under Known First-Stage Sign," mimeo.
- BUSE, A. (1992): "The bias of instrumental variable estimators," *Econometrica: Journal of the Econometric Society*, pp. 173–180.
- CHAO, J., AND N. SWANSON (2007): "Alternative Specifications of the Bias and MSE of the IV Estimator under Weak Identification with an Application to Bias Correction," *Journal of Econometrics*, 137, 515–555.
- CHUANMING, G., AND K. LAHIRI (2001): "A Note on the Double k-class Estimator in Simultaneous Equations," *Journal of Econometrics*, 108, 101–111.
- DWIVEDI, T. D., AND V. K. SRIVASTAVA (1984): "Exact Finite Sample Properties of Double k-class Estimators in Simultaneous Equations," *Journal of Econometrics*, 25, 263–283.

- FULLER, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–953.
- HAHN, J., AND J. A. HAUSMAN (2002a): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189.
- (2002b): "Notes on Bias in Estimators for Simultaneous Equation Models," *Economics Letters*, 75, 237–241.
- HANSEN, C., J. HAUSMAN, AND W. NEWEY (2008): "Estimation with many instrumental variables," *Journal of Business & Economic Statistics*, 26(4).
- HILLIER, G., T. KINAL, AND V. K. SRIVASTAVA (1984): "On the Moments of Ordinary Least Squares and Instrumental Variables Estimators in a General Structural Equation," *Econometrica*, 52, 185–202.
- HIRANO, K., AND J. R. PORTER (forthcoming): "Location Properties of Point Estimators in Linear Instrumental Variables and Related Models," *Econometric Reviews*.
- KINAL, T. W. (1980): "The existence of moments of k-class estimators," *Econometrica: Journal of the Econometric Society*, pp. 241–249.
- KOEV, P., AND A. EDELMAN (2006): "The Efficient Evaluation of the Hypergeometric Function of a Matrix Argument," *Mathematics of Computation*, 75, 833–846.
- MULLER, K. E. (2001): "Computing the Confluent Hypergeometric Function $M(a,b,x)$," *Numerische Mathematik*, 90, 179–196.
- NAGAR, A. L. (1959): "The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575–595.
- (1962): "Double k-class Estimators of Parameters in Simultaneous Equations and their Small Sample Properties," *International Economic Review*, 3, 168–188.
- OWEN, A. D. (1976): "A Proof that Both the Bias and the Mean Square Error of the Two-Stage Least Squares Estimator are Monotonically Non-increasing Functions of the Sample Size," *Econometrica*, 44, 409–411.
- PHILLIPS, P. (1980): "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing $n+1$ Endogenous Variables," *Econometrica*, 48, 861–878.
- POSKITT, D., AND C. SKEELS (2013): "Inference in the presence of weak instruments: A selected survey," *Foundations and Trends in Econometrics*, 6(1), 1–102.
- POSKITT, D. S., AND C. L. SKEELS (2009): "Assessing the magnitude of the concentration parameter in a simultaneous equations model," *The Econometrics Journal*, 12(1), 26–44.
- RICHARDSON, D. H. (1968): "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, pp. 1214–1226.
- RICHARDSON, D. H., AND D. M. WU (1971): "A Note on the Comparison of Ordinary and Two-Stage Least Squares Estimators," *Econometrica*, 39, 973–981.
- SAWA, T. (1969): "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, pp. 923–937.

- (1972): “Finite Sample Properties of the k-class Estimators,” *Econometrica*, 40.
- SLATER, L. J. (1964): “Confluent Hypergeometric Functions,” in *Handbook of Mathematical Functions*, ed. by M. Abramowitz, and I. A. Stegun. National Bureau of Standards.
- SRINIVASAN, T. N. (1970): “Approximations to Finite Sample Moments of Estimators Whose Exact Sampling Distributions are Unknown,” *Econometrica*, 38, 533–541.
- SRIVASTAVA, V. K., B. S. AGNIHOTRI, AND T. D. DWIVEDI (1980): “Dominance of Double k-class Estimators in Simultaneous Equations,” *Annals of the Institute for Statistical Mathematics*, 32A, 387–392.
- ULLAH, A. (2004): *Finite Sample Econometrics*. Oxford.
- ZELLNER, A. (1963): “Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results,” *Journal of the American Statistical Association*, 58, 977–992.

APPENDIX A: Derivations of Expressions in Section 2

In a few of these derivations, we use the fact that

$$\begin{aligned} B(m, k) &= e^{-m/2} {}_1F_1(k/2 - 1; k/2; m/2) \\ &= e^{-m/2} e^{m/2} {}_1F_1(k/2 - (k/2 - 1); k/2; -m/2) = {}_1F_1(1; k/2; -m/2). \end{aligned}$$

The second equality is known as Kummer's Transformation (see equation (25) in Abadir, 1999).

part (a)

$$B(m, 2) = e^{-m/2} {}_1F_1(0; 1; m/2) = e^{-m/2} \sum_{j=0}^{\infty} \frac{(0)_j}{(1)_j} \frac{(m/2)^j}{j!} = e^{-m/2}$$

since the first term of the series is 1 and $(0)_j = 0$ for $j \geq 1$.

part (b)

$$\begin{aligned} \lim_{k \rightarrow \infty} B(m, k) &= \lim_{k \rightarrow \infty} {}_1F_1(1; k/2; -m/2) \\ &= \lim_{k \rightarrow \infty} 1 + \sum_{j=1}^{\infty} ((k/2)_j)^{-1} \frac{(-m/2)^j}{j!} \\ &= 1 + \sum_{j=1}^{\infty} \lim_{k \rightarrow \infty} ((k/2)_j)^{-1} \frac{(-m/2)^j}{j!} = 1 \end{aligned}$$

part (c)

$B(m, k)$ is continuous in m , so

$$(6.1) \quad \lim_{m \rightarrow 0} B(m, k) = B(0, k) = e^0 {}_1F_1(k/2 - 1; k/2; 0) = 1$$

since hypergeometric functions evaluated at zero are equal to 1, the first term in the infinite series.

part (d)

When we do the substitution we get $1 + (1/6)e^{-m/2}(-6 - m)m$ (the last minus sign in this expression is a plus sign in the corresponding equation). To derive this, expand $e^{-m/2}$ as $1 - m/2 + \mathcal{O}(m^2)$. Then

$$\begin{aligned} 1 + (1/6)e^{-m/2}(-6 - m)m &= 1 - (1/6)m(6 + m)(1 - m/2 + \mathcal{O}(m^2)) \\ &= 1 - (1/6)m(6 - 3m + m - m^2/2) + \mathcal{O}(m^3) = 1 - (1/3)m(3 - m) + \mathcal{O}(m^3). \end{aligned}$$

$B(m, 1)$ can be expanded as

$$\begin{aligned} B(m, 1) &= {}_1F_1(1; 1/2; -m/2) = 1 + \frac{1}{1/2}(-m/2) + \frac{1 \cdot 2}{(1/2)(3/2)} \frac{(-m/2)^2}{2} + \mathcal{O}(m^3) \\ &= 1 - m + \frac{2 \cdot 2}{3 \cdot 4} m^2 + \mathcal{O}(m^3) = 1 - (1/3)m(3 - m) + \mathcal{O}(m^3) \end{aligned}$$

so $B(m, 1) = 1 + (1/6)e^{-m/2}(-6 - m)m + \mathcal{O}(m^3)$.

part (e)

It suffices to show that ${}_1F_1(k/2 - 1; k/2; m/2)$ is strictly increasing in k , which will follow if we can show that each term in the infinite series is strictly increasing in k . The j th term is

$$(6.2) \quad \frac{(k/2 - 1)_j (m/2)^j}{(k/2)_j j!} = \frac{k/2 - 1}{k/2 + j - 1} \frac{(m/2)^j}{j!}$$

which is strictly increasing in k for $m \geq 0$ since

$$(6.3) \quad \frac{d}{dk} \left(\frac{k/2 - 1}{k/2 + j - 1} \right) = \frac{(1/2)(k/2 + j - 1) - (k/2 - 1)(1/2)}{(k/2 + j - 1)^2} = \frac{j}{2(k/2 + j - 1)^2} > 0.$$