Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations[1]

Jinyong Hahn, Brown University
Jerry Hausman, MIT
Guido Kuersteiner[2], MIT

January, 2004

[1]Please send editorial correspondence to: Guido Kuersteiner, MIT, Department of Economics, E-52-371A, 50 Memorial Drive, Cambridge, MA 02142

[2]Financial support from NSF grant SES-0095132 is gratefully acknowledged.

Abstract

In this paper we consider parameter estimation in a linear simultaneous equations model. It is well known that two stage least squares (2SLS) estimators may perform poorly when the instruments are weak. In this case 2SLS tends to suffer from substantial small sample biases. It is also known that LIML and Nagar-type estimators are less biased than 2SLS but suffer from large small sample variability. We construct a bias corrected version of 2SLS based on the Jackknife principle. Using higher order expansions we show that the MSE of our Jackknife 2SLS estimator is approximately the same as the MSE of the Nagar-type estimator. We also compare the Jackknife 2SLS with an estimator suggested by Fuller (1977) that significantly decreases the small sample variability of LIML. Monte Carlo simulations show that even in relatively large samples the MSE of LIML and Nagar can be substantially larger than for Jackknife 2SLS. The Jackknife 2SLS estimator and Fuller's estimator give the best overall performance. Based on our Monte Carlo experiments we conduct formal statistical tests of the accuracy of approximate bias and MSE formulas. We find that higher order expansions traditionally used to rank LIML, 2SLS and other IV estimators are unreliable when identification of the model is weak. Overall, our results show that only estimators with well defined finite sample moments should be used when identification of the model is weak.

Keywords: weak instruments, higher order expansions, bias reduction, Jackknife, 2SLS

JEL C13,C21,C31,C51

1 Introduction

Over the past few years there has been renewed interest in finite sample properties of econometric estimators. Most of the related research activities in this area are concentrated in the investigation of finite sample properties of instrumental variables (IV) estimators. It has been found that standard large sample inference based on 2SLS can be quite misleading in small samples when the endogenous regressor is only weakly correlated with the instrument. A partial list of such research activities is Nelson and Startz (1990), Maddala and Jeong (1992), Staiger and Stock (1997), and Hahn and Hausman (2002).

A general result is that controlling for bias can be quite important in small sample situations. Anderson and Sawa (1979), Morimune (1983), Bekker (1994), Angrist, Imbens, and Krueger (1995), and Donald and Newey (2001) found that IV estimators with smaller bias typically have better risk properties in finite sample. For example, it has been found that the LIML, the JIVE, or Nagar's (1959) estimator tend to have much better risk properties than 2SLS. Donald and Newey (2000), Newey and Smith (2001) and Kuersteiner (2000) may be understood as an endeavor to obtain a bias reduced version of the GMM estimator in order to improve the finite sample risk properties.

In this paper we consider higher order expansions of LIML, JIVE, Nagar and 2SLS estimators. In addition we contribute to the higher order literature by deriving the higher order risk properties of the Jackknife 2SLS. Such an exercise is of interest for several reasons. First, we believe that higher order MSE calculations for the Jackknife estimator have not been available in the literature. Most papers simply verify the consistency of the Jackknife bias estimator. See Shao and Tu (1995, Section 2.4) for a typical discussion of this kind. Akahira (1983), who showed that the Jackknife MLE is second order equivalent to MLE, is closest in spirit to our exercise here, although a *third* order expansion is necessary in order to calculate the higher order MSE.

Second, Jackknife 2SLS may prove to be a reasonable competitor to the LIML or Nagar's estimator despite the fact that higher order theory predicts it should be dominated by LIML. It is well-known that LIML and Nagar's estimator have the "moment" problem: With normally distributed error terms, it is known that LIML and Nagar do not possess any moments. See Mariano and Sawa (1972) or Sawa (1972). On the other hand, it can be shown that Jackknife 2SLS has moments up to the degree of overidentification. LIML and Nagar's estimator have better higher order risk properties than 2SLS, based on higher order expansions used by Rothenberg (1983) or Donald and Newey (2001). These results may however not be very reliable if the moment problem is not only a feature of the extreme end of the tails but rather affects dispersion of the estimators more generally. Large dispersions and lack of moments are technically distinct problems, but we identify these problems for historical reasons. On the other hand, these two problems seem to be identical problems for all practical purpose, as our Monte Carlo results demonstrate.

We conduct a series of Monte Carlo experiments to determine how well higher order approximations predict the actual small sample behavior of the different estimators. When identification of the model is weak the quality of the approximate MSE formulas based on higher order expansions turns out to be poor. This is particularly true for the LIML and Nagar estimators that have no moments in finite samples. Our calculations show that estimators that would be dismissed based on the analysis of higher order stochastic expansions turn out to perform much better than predicted by theory.

We point out that the literature on exact small sample distributions of 2SLS and LIML such as Anderson and Sawa (1979), Anderson, Morimune and Sawa (1983), Anderson, Naoto and Sawa (1982), Holly and Phillips (1979), Phillips (1980) and more recently Oberhelman and Kadiyala (2000) did not focus on the weak instrument case and therefore the small sample behavior of LIML under these circumstances is still an open issue. We show that the parametrizations used in these papers imply values for the first stage R^2 that are much larger than typically thought relevant for the weak instrument case.

These previous papers present finite sample results summarized around the "concentration parameter".[1] The concentration parameter δ^2 is approximately equal to $nR^2/\left(1-R^2\right)$ under our normalization, where n denotes the sample size.[2] In Anderson et. al. (1982, Table I, p. 1012) the minimum concentration parameter equals 30, which far exceeds the weak instrument cases we consider where the first stage R^2 is equal to 0.01. Rather, for the case of $n = 100$, one needs $R^2 = 0.3$ in order to get $\delta^2 = 30$, which is quite high. Subsequent tables in Anderson et.al. continue to have this problem with the minimum $\delta^2 = 10$ which for $R^2 = 0.1$ means that one needs $n = 1000$. Our Monte Carlo experiments show that the weak instrument problem pretty much disappears for n this large. Similarly, the graphs reported in Anderson et.al., p.1022-1023 do not apply to the weak instrument case because they are based on $\delta^2 = 100$ which would require $R^2 = 0.1$ for $n = 1000$. The parametrizations of Anderson et. al. were used by other researchers such as Oberhelman and Kadiyala (2000, p. 171). Anderson and Sawa (1973) use values of δ^2 ranging from 20 to 180 in their numerical work and Anderson and Sawa (1979) use values of δ^2 ranging from 10 to ∞. Holly and Phillips (1979) use $\delta^2 \geq 40$ and Phillips (1980) considers 2SLS estimators for the case of multiple endogenous variables such that his results are not directly comparable here. Nevertheless, the implied values of δ^2 are roughly 80 with a sample size of $n = 20$ which implies an R^2 of 0.8.

[1] The concentration parameter is formally defined later on page 11.

[2] See the discussion on page 11.

Precisely for the cases with very low first stage R^2 we find previously undocumented behavior of LIML estimators. In particular, LIML tends to be biased and has large inter-quantile ranges that can dominate those of 2SLS and certainly those of the Jackknife 2SLS advocated here. Anderson et. al. (1983, p. 233) claim: "The infinite moments are due to the behavior of the distributions outside of the range of practical interest to the econometrician." However, the approximations that they use to reach this conclusion do not work in the weak instrument case. In particular they claim on p.233 that, "The MSE of the asymptotic expansion to order $1/\delta^2$ of the LIML estimator is smaller than that of the TSLS estimator if $K_2 > 7$ and $|\alpha| > \sqrt{2/(K_2-7)}$...," where K_2 is the degree of overidentification. It can be checked easily that this inequality holds for all the parameter values that we consider in our simulations and we would thus expect to find smaller interquartile ranges for LIML compared to 2SLS. But quite to the contrary, our simulation results indicate that LIML does worse than 2SLS in the weak instrument cases. This is further evidence that the finite sample analysis did not address the weak instrument case. The reason is of course that in the weak instrument case of Staiger and Stock (1997) $\delta^2 = O(1)$ such that the approximations analyzed in Anderson et. al. (1983) do not converge and are thus not useful guides to assess the properties of estimators in the weak instrument case. Contrary to Anderson et.al. we find that the large dispersion of LIML is around parameter values very much of relevance to the econometrician. We also find that the actual median bias of LIML can be significant and is increasing in the number of instruments. Both of these findings are in contrast to the fact that the higher order median bias of LIML according to the traditional expansion is zero. We try to provide an explanation of this discrepancy by using an alternative asymptotic analysis put forth by Staiger and Stock (1997).

Based on our Monte Carlo experiments we conduct informal statistical tests of the accuracy of predictions about bias and MSE based on higher order stochastic expansions. We find that when identification of the model is weak such bias and MSE approximations perform poorly and selecting estimators based on them is unreliable. The issue of how a small concentration parameter may lead to a break down of the reliability of the traditional higher order expansion has been recognized in the literature, although the practical relevance of this problem does not seem to have been extensively investigated. See, e.g., Rothenberg (1984).

In this paper, we also compare the Jackknife 2SLS estimator with a modification of the LIML estimator proposed by Fuller (1977). Fuller's estimator does have finite sample moments so it solves the moment problems associated with the LIML and Nagar estimators. We find the optimum form of Fuller's estimator. Our conclusion is that both this form of Fuller's estimator and JN2SLS have improved finite sample properties and do not have the "moment" problem in comparison to the typically used estimators such as LIML. However, neither the Fuller estimator nor JN2SLS dominate each other in actual practice.

Our recommendation for the practitioner is thus to use only estimators with well defined finite sample moments when the model may only be weakly identified.

2 MSE of Jackknife 2SLS

The model we focus on is the simplest model specification with one right hand side (RHS) jointly endogenous variable so that the left hand side variable (LHS) depends only on the single jointly endogenous RHS variable. This model specification accounts for other RHS predetermined (or exogenous) variables, which have been "partialled out" of the specification. We will assume that

$$\begin{aligned} y_i &= x_i\beta + \varepsilon_i, \\ x_i &= f_i + u_i = z_i'\pi + u_i \qquad i = 1, \ldots, n \end{aligned} \tag{1}$$

Here, x_i is a scalar variable, and z_i is a K-dimensional nonstochastic column vector. The first equation is the equation of interest, and the right hand side variable x_i is possibly correlated with ε_i. The second equation represents the "first stage regression", i.e., the reduced form between the endogenous regressor x_i and the instruments z_i. By writing $f_i \equiv E\left[x_i \middle| z_i\right] = z_i'\pi$, we are ruling out a nonparametric specification of the first stage regression. Note that the first equation does not include any other exogenous variable. It will be assumed throughout the paper (except for the empirical results) that all the error terms are homoscedastic.

We focus on the 2SLS estimator b given by

$$b = \frac{x'Py}{x'Px} = \beta + \frac{x'P\varepsilon}{x'Px},$$

where $P \equiv Z\left(Z'Z\right)^{-1}Z'$. Here, y denotes $(y_1, \ldots, y_n)'$. We define x, ε, u, and Z similarly. 2SLS is a special case of the k-class estimator given by

$$\frac{x'Py - \kappa \cdot x'My}{x'Px - \kappa \cdot x'Mx},$$

where $M \equiv I - P$ and κ is a scalar. For $\kappa = 0$, we obtain 2SLS. For κ equal to the smallest eigenvalue of the matrix $W'PW\left(W'MW\right)^{-1}$, where $W \equiv [y, x]$, we obtain LIML. For $\kappa = \frac{K-2}{n} \Big/ \left(1 - \frac{K-2}{n}\right)$, we obtain B2SLS, which is Donald and Newey's (2001) modification of Nagar's (1959) estimator.

Donald and Newey (2001) compute the higher order mean squared error (MSE) of the k-class estimators. They show that n times the MSE of 2SLS, LIML, and B2SLS are approximately equal to

$$\frac{\sigma_\varepsilon^2}{H} + \frac{K^2}{n}\frac{\sigma_{u\varepsilon}^2}{H^2}$$

for 2SLS,

$$\frac{\sigma_\varepsilon^2}{H} + \frac{K}{n}\frac{\sigma_u^2\sigma_\varepsilon^2 - \sigma_{u\varepsilon}^2}{H^2}$$

for LIML and

$$\frac{\sigma_\varepsilon^2}{H} + \frac{K}{n}\frac{\sigma_u^2\sigma_\varepsilon^2 + \sigma_{u\varepsilon}^2}{H^2}$$

for B2SLS, where we define $H \equiv \frac{f'f}{n}$. The first term, which is common in all three expressions, is the usual asymptotic variance obtained under the first order asymptotics. Finite sample properties are captured by the second terms. For 2SLS, the second term is easy to understand. As discussed in, e.g., Hahn and Hausman (2001), 2SLS has an approximate bias equal to $\frac{K\sigma_{u\varepsilon}}{nH}$. Therefore, the approximate expectation for $\sqrt{n}\,(b-\beta)$ ignored in the usual first order asymptotics is equal to $\frac{K\sigma_{u\varepsilon}}{\sqrt{n}H}$, which contributes $\left(\frac{K\sigma_{u\varepsilon}}{\sqrt{n}H}\right)^2 = \frac{K^2}{n}\frac{\sigma_{u\varepsilon}^2}{H^2}$ to the higher order MSE. The second terms for LIML and B2SLS do not reflect higher order biases. Rather, they reflect higher order variance that can be understood from Rothenberg's (1983) or Bekker's (1994) asymptotics.

Higher order MSE comparison alone suggest that LIML and B2SLS should be preferred to 2SLS. Unfortunately, it is well-known that LIML and Nagar's estimator have the "moment" problem. If (ε_i, u_i) has a bivariate normal distribution, it is known that LIML and B2SLS do not possess any moments. On the other hand, it is known that 2SLS does not have a moment problem. See Mariano and Sawa (1972) or Sawa (1972). This theoretical property implies that LIML and B2SLS have thicker tails than 2SLS. It would be nice if the moment problem could be dismissed as a mere academic curiosity. Unfortunately, we find in Monte Carlo experiments that LIML and B2SLS tend to be more dispersed (measured in terms of interquartile range, etc) than 2SLS for some parameter combinations. This is especially true when identification of the model is weak. Under these circumstances higher order expansions tend to deliver unreliable rankings of estimators. In this sense, 2SLS can still be viewed as a reasonable contender to LIML and B2SLS.

Given that the poor higher order MSE property of 2SLS is based on its bias, we may hope to improve 2SLS by eliminating its finite sample bias through the jackknife. Jackknife 2SLS may turn out to be a reasonable contender given that it can be expressed as a linear combination of 2SLS, and hence, free of the moment problem. This is because the jackknife estimator of the bias is given by

$$\frac{n-1}{n}\sum_i\left(\frac{\widehat{\pi}_{(i)}'\sum_{j\neq i} z_j y_j}{\widehat{\pi}_{(i)}'\sum_{j\neq i} z_j x_j} - \frac{\widehat{\pi}'\sum_i z_i y_i}{\widehat{\pi}'\sum_i z_i x_i}\right) = \frac{n-1}{n}\sum_i\left(\frac{\widehat{\pi}_{(i)}'\sum_{j\neq i} z_j \varepsilon_j}{\widehat{\pi}_{(i)}'\sum_{j\neq i} z_j x_j} - \frac{x'P\varepsilon}{x'Px}\right) \tag{2}$$

and the corresponding jackknife estimator is given by

$$\begin{aligned} b_J &= \frac{\widehat{\pi}' \sum_i z_i y_i}{\widehat{\pi}' \sum_i z_i x_i} - \frac{n-1}{n} \sum_i \left(\frac{\widehat{\pi}'_{(i)} \sum_{j \neq i} z_j y_j}{\widehat{\pi}'_{(i)} \sum_{j \neq i} z_j x_j} - \frac{\widehat{\pi}' \sum_i z_i y_i}{\widehat{\pi}' \sum_i z_i x_i} \right) \\ &= n \frac{\widehat{\pi}' \sum_i z_i y_i}{\widehat{\pi}' \sum_i z_i x_i} - \frac{n-1}{n} \sum_i \frac{\widehat{\pi}'_{(i)} \sum_{j \neq i} z_j y_j}{\widehat{\pi}'_{(i)} \sum_{j \neq i} z_j x_j} \end{aligned}$$

Here, $\widehat{\pi}$ denotes the OLS estimator of the first stage coefficient π, and $\widehat{\pi}_{(i)}$ denotes an OLS estimator based on every observation except the ith. Observe that b_J is a linear combination of

$$\frac{\widehat{\pi}' \sum_i z_i y_i}{\widehat{\pi}' \sum_i z_i x_i}, \frac{\widehat{\pi}'_{(1)} \sum_{j \neq i} z_j y_j}{\widehat{\pi}'_{(1)} \sum_{j \neq i} z_j x_j}, \ldots, \frac{\widehat{\pi}'_{(n)} \sum_{j \neq i} z_j y_j}{\widehat{\pi}'_{(n)} \sum_{j \neq i} z_j x_j}$$

and all of them have finite moments if the degree of overidentification is sufficiently large ($K > 2$). See, e.g., Mariano (1972). Therefore, b_J has finite second moments if the degree of overidentification is large.

We show that, for large K, the approximate MSE for the jackknife 2SLS is the same as in Nagar's estimator or JIVE. As in Donald and Newey (2001), we let $h \equiv \frac{f'\varepsilon}{n}$. We impose the following assumptions. First, we assume normality[3]:

Condition 1 *(i) $(\varepsilon_i, u_i)'$ $i = 1, \ldots, n$ are i.i.d.; (ii) $(\varepsilon_i, u_i)'$ has a bivariate normal distribution with mean equal to zero.*

We also assume that z_i is a sequence of nonstochastic column vectors satisfying

Condition 2 $\max P_{ii} = O\left(\frac{1}{n}\right)$*, where P_{ii} denotes the (i,i)-element of $P \equiv Z\left(Z'Z\right)^{-1} Z'$.*

Condition 3 *(i) $\max |f_i| = \max |z_i'\pi| = O\left(n^{1/r}\right)$ for some r sufficiently large ($r > 3$); (ii) $\frac{1}{n}\sum_i f_i^6 = O(1)$.*[4]

After some algebra, it can be shown that

$$\widehat{\pi}'_{(i)} \sum_{j \neq i} z_j \varepsilon_j = x'P\varepsilon + \delta_{1i}, \quad \widehat{\pi}'_{(i)} \sum_{j \neq i} z_j x_j = x'Px + \delta_{2i},$$

where

$$\delta_{1i} \equiv -x_i \varepsilon_i + \left(1 - P_{ii}\right)^{-1} (Mx)_i \, (M\varepsilon)_i\,, \quad \delta_{2i} \equiv -x_i^2 + \left(1 - P_{ii}\right)^{-1} (Mx)_i^2\,.$$

[3]We expect that our result would remain valid under the symmetry assumption as in Donald and Newey (1998), although such generalization is expected to be substantially complicated.

[4]If $\{f_i\}$ is a realization of a sequence of *i.i.d.* random variables such that $E\left[|f_i|^r\right] < \infty$ for r sufficiently large, Condition 3 (i) may be justified in probabilistic sense. See Lemma 1 in Appendix.

Here, $(Mx)_i$ denotes the ith element of Mx, and $M \equiv I - P$. We may therefore write the jackknife estimator of the bias as

$$
\begin{aligned}
&\frac{n-1}{n}\sum_i\left(\frac{x'P\varepsilon+\delta_{1i}}{x'Px+\delta_{2i}}-\frac{x'P\varepsilon}{x'Px}\right)\\
&\qquad\qquad = \frac{n-1}{n}\sum_i\left(\frac{1}{x'Px}\delta_{1i}-\frac{x'P\varepsilon}{(x'Px)^2}\delta_{2i}-\frac{1}{(x'Px)^2}\delta_{1i}\delta_{2i}+\frac{x'P\varepsilon}{(x'Px)^3}\delta_{2i}^2\right)+R_n
\end{aligned}
$$

where

$$
R_n \equiv \frac{n-1}{n^4}\frac{1}{\left(\frac{1}{n}x'Px\right)^2}\sum_i\frac{\delta_{1i}\delta_{2i}^2}{\frac{1}{n}x'Px+\frac{1}{n}\delta_{2i}}-\frac{n-1}{n^4}\frac{\frac{1}{n}x'P\varepsilon}{\left(\frac{1}{n}x'Px\right)^3}\sum_i\frac{\delta_{2i}^3}{\frac{1}{n}x'Px+\frac{1}{n}\delta_{2i}}.
$$

By the Lemma 2 in the Appendix, we have

$$
n^{3/2}R_n = O_p\left(\frac{1}{n\sqrt{n}}\sum_i\left|\delta_{1i}\delta_{2i}^2\right|+\frac{1}{n\sqrt{n}}\sum_i\left|\delta_{2i}^3\right|\right)=o_p(1),
$$

and we can ignore it from our further computation.

We now examine the resultant bias corrected estimator (2) ignoring R_n:

$$
\begin{aligned}
&H\sqrt{n}\left(\frac{x'P\varepsilon}{x'Px}-\frac{n-1}{n}\sum_i\left(\frac{x'P\varepsilon+\delta_{1i}}{x'Px+\delta_{2i}}-\frac{x'P\varepsilon}{x'Px}\right)+R_n\right)\\
&= H\sqrt{n}\frac{x'P\varepsilon}{x'Px}\\
&\quad -\frac{n-1}{n}\frac{H}{\frac{1}{n}x'Px}\left(\frac{1}{\sqrt{n}}\sum_i\delta_{1i}\right)\\
&\quad +\frac{n-1}{n}\frac{H}{\frac{1}{n}x'Px}\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n}\sum_i\delta_{2i}\right)\\
&\quad +\frac{n-1}{n}\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{1}{n\sqrt{n}}\sum_i\delta_{1i}\delta_{2i}\right)\\
&\quad -\frac{n-1}{n}\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n^2}\sum_i\delta_{2i}^2\right)
\end{aligned}
\tag{3}
$$

Theorem 1 below is obtained by squaring and taking expectation of the RHS of (3):

Theorem 1 *Assume that Conditions 1, 2, and 3 are satisfied. Then, the approximate MSE of* $\sqrt{n}(b_J-\beta)$ *for the jackknife estimator up to* $O\left(\frac{K}{n}\right)$ *is given by*

$$
\frac{\sigma_\varepsilon^2}{H}+\frac{K}{n}\frac{\sigma_u^2\sigma_\varepsilon^2+\sigma_{u\varepsilon}^2}{H^2}.
$$

Proof. See Appendix. ■

Theorem 1 indicates that the higher order MSE of Jackknife 2SLS is equivalent to that of Nagar's (1959) estimator if the number of instruments is sufficiently large (see Donald and Newey (2001)). However, Jackknife 2SLS does have moments up to the degree of overidentification. Therefore, the Jackknife does not increase the variance too much. Although it has long been known that the Jackknife does reduce the bias, the literature has been hesitant in recommending its use primarily because of the concern that the variance may increase too much due to the Jackknife bias reduction. See Shao and Tu (1995, p. 65), for example.

Theorem 1 also indicates that the higher order MSE of Jackknife 2SLS is bigger than that of LIML. In some sense, this result is not surprising. Hahn and Hausman (2002) demonstrated that LIML is approximately equivalent to the optimal linear combination of the two Nagar estimators based on forward and reverse specifications. Jackknife 2SLS is solely based on forward 2SLS, and ignores the information contained in reverse 2SLS. Therefore, it is quite natural to have LIML dominating Jackknife 2SLS on a theoretical basis.

3 Fuller's (1977) Estimator

Fuller (1977) developed a modification of LIML of the form

$$\frac{x'Py - \left(\phi - \frac{a}{n-K}\right) \cdot x'My}{x'Px - \left(\phi - \frac{a}{n-K}\right) \cdot x'Mx}, \tag{4}$$

where ϕ is equal to the smallest eigenvalue of the matrix $W'PW\left(W'MW\right)^{-1}$ and $W \equiv [y, x]$. Here, $a > 0$ is a constant to be chosen by the researcher. Note that the estimator is identical to LIML if α is chosen to be equal to zero. We consider values of alpha equal to $a = 1$ and $a = 4$. The choice of $a = 1$ advocated, e.g. by Davidson and McKinnon (1993, p. 649), yields a higher order mean bias of zero while $a = 4$ has a nonzero higher mean bias, but a smaller MSE according to calculations based on Rothenberg's (1983) analysis.

Fuller (1977) showed that this estimator does not have the "moment problem" that plagues LIML. It can also be shown that this estimator has the same higher order MSE as LIML up to $O\left(\frac{K}{n^2}\right)$.[5] Therefore, it dominates Jackknife 2SLS on higher order theoretical grounds for MSE, although not necessarily for bias.

[5] See Appendix C for the higher order bias and MSE of the Fuller estimator.

4 Theory and Practice

In this section we report the results of an extensive Monte Carlo experiment and then do an econometric analysis to analyze how well the empirical results accord with the second order asymptotic theory that we explored previously. We have two major findings: (1) estimators that have good theoretical properties but lack finite sample moments should not be used. Thus, our recommendation is that LIML not be used in a "weak instruments" situation (2) approximately unbiased (to second order) estimators that have moments offer a great improvement. The Fuller adaptation of LIML and JN2SLS are superior to LIML, Nagar, and JIVE. However, depending on the criterion used, 2SLS does very well despite its second order bias properties. 2SLS's superiority in terms of asymptotic variance, as demonstrated in the higher order asymptotic expansions appears in the results. The second order bias calculation for 2SLS, e.g. Hahn and Hausman (2001), which demonstrates that bias grows with the number of instruments K so that the MSE grows as K^2, appears unduly pessimistic based on our empirical results. Thus, our suggestion is to use JN2SLS, a Fuller estimator, or 2SLS depending on the criterion preferred by the researcher.

4.1 Estimators Considered

We consider estimation of equation (1) with one RHS endogenous variable and all predetermined variables have been partialled out. We then assume (without loss of generality) that $\sigma_\varepsilon^2 = \sigma_u^2 = 1$ and $\sigma_{\varepsilon u} = \rho$. Thus, our higher order formula will depend on the number of instruments K, the number of observations n, ρ, and the (theoretical) R^2 of the first stage regression.[6] Using the normalization, the often used concentration parameter approach yields $\delta^2 \approx nR^2/\left(1 - R^2\right)$.

The estimators that we consider are:

- LIML - see e.g. Hausman (1983) for a derivation and analysis. LIML is known not to have finite sample moments of any order. LIML is also known to be median unbiased to second order and to be admissible for median unbiased estimators, see Rothenberg (1983). The higher order mean bias for LIML does not depend on K.

- 2SLS - the most widely used IV estimator. 2SLS has finite sample bias that depends on the number of instruments used K and inversely on the R^2 of the first stage regression, see e.g. Hahn and Hausman (2001). The higher order mean bias of 2SLS is proportional to K. However, 2SLS can have smaller higher order mean square error (MSE) than LIML using second order approximations when the number of instruments is not too large, see Bekker (1994) and Donald and Newey (2001).

[6] The theoretical R^2 is defined later in (5).

- Nagar - mean unbiased up to second order. For a simplified derivation see Hahn and Hausman (2001). The Nagar estimator does not have moments of any order.

- Fuller (1977) - this estimator is an adaptation of LIML designed to have finite sample moments. We consider three different estimators with the a parameter in (4) chosen to take on values 1 or 4 or the value that minimizes higher order MSE. The optimal estimator uses $a = 3 + 1/\rho^2$. This choice minimizes the higher order MSE regarded as a function of a. These three estimators will be abbreviated F(1), F(4), and F(opt) throughout the rest of the paper. For the optimal Fuller estimator, the higher order bias is greater, but the MSE is smaller. This last estimator is infeasible since ρ is unknown in an actual situation, but we explore it for completeness. The optimal estimator has the same higher order MSE as LIML up to $O\left(\frac{K}{n^2}\right)$ but unlike LIML also has existing finite sample moments.

- JN2SLS - the higher order mean bias does not depend on K, the number of instruments. JN2SLS has finite sample moments. However, as we discuss later, its MSE exceeds the other estimators in some situations.

- JIVE - the jackknifed IV estimator of Phillips and Hale (1977) and Angrist, Imbens, and Krueger (1999). This estimator is higher order mean unbiased similar to Nagar, but we conjecture that it does not have finite sample moments. The Monte Carlo results demonstrate a likely absence of finite sample moments.

- OLS - This estimator is to be considered as a benchmark.

Formal definitions of some of the k-class estimators for equation (1) are provided in Table 6:

4.2 Monte Carlo Design

We used the same design as in Hahn and Hausman (2002) with one RHS endogenous variable corresponding to equation (1). We let $\beta = 0$, and $z_i \sim \mathcal{N}(0, I_K)$. Let

$$\mathbb{R}_f^2 = \frac{E\left[(\pi' z_i)^2\right]}{E\left[(\pi' z_i)^2\right] + E\left[v_i^2\right]} = \frac{\pi'\pi}{\pi'\pi + 1} \tag{5}$$

denote the theoretical R^2 of the first stage regression. We specify $\pi = (\eta, \eta, \ldots, \eta)'$ so that

$$\mathbb{R}_f^2 = \frac{q \cdot \eta^2}{q \cdot \eta^2 + 1}$$

We use $n = (100, 500, 1000)$, $K = (5, 10, 30)$, $R^2 = (0.01, 0.1, 0.3)$, and $\rho = (0.5, 0.9)$, which are considered to be weak instrument situations. Our results, which are reported in Tables 1 - 5, are based on 5000 replications.

In order to highlight the weak instrument nature of our simulation design we compare it to parametrizations typically used in the exact finite sample distribution literature for LIML and 2SLS. In particular consider the parametrization in Anderson, Kunitomo and Sawa (1982). They define

$$\delta^2 = \frac{\pi' \left(\sum_{i=1}^n z_i z_i'\right) \pi}{\omega_{22}}$$

and

$$\alpha = \frac{\omega_{22}}{|\Omega|^{1/2}} \left(\beta - \frac{\omega_{12}}{\omega_{22}}\right)$$

where the reduced form covariance matrix is given by

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11} + 2\beta\sigma_{12} + \beta^2\sigma_{22} & \sigma_{12} + \beta\sigma_{22} \\ \sigma_{12} - \beta\sigma_{22} & \sigma_{22} \end{bmatrix}$$

for the structural errors $[\varepsilon_i, u_i] \tilde{} N(0, \Sigma)$ and Σ with elements σ_{ij}. Given our assumption about z_i it follows at once that $\delta^2 \approx n \cdot \pi'\pi = n \cdot q \cdot \eta^2$. This implies that $\delta^2 \approx nR^2 / \left(1 - R^2\right)$ and we also find that $\alpha = -\rho/\sqrt{1-\rho^2}$ or $\rho = -\alpha/\sqrt{1+\alpha^2}$. We compute values of δ for the parametrizations that we use in our simulations. The results are summarized in Table 7.

4.3 Monte Carlo Results

4.3.1 Median Bias

We first consider the median bias results. Especially for the situation of $R^2 = 0.01$ the absence of finite sample moments for LIML, Nagar, and JIVE is apparent. Among the three Fuller estimators, F(1) has the smallest bias, in accordance with the second order theory. Also, the median bias increases as we go to F(4) and F(opt), again as theory predicts. When R^2 increases to 0.1 the F(1) estimator often does better than JN2SLS, but not by large amounts. Lastly, when R^2 increases to 0.3, the finite sample problem ceases to be important, and LIML, Nagar and the other estimators do well. We conclude that for sample sizes above 100 that R^2=0.3 is high enough that finite sample problems cease to be a concern. Overall, the JN2SLS estimator does quite well in terms of bias - it is usually comparable and sometimes smaller than the "unbiased" F(1) estimator, although on average F(1) does better than JN2SLS. Overall, JN2SLS has smaller bias than either the F(4) estimator or the infeasible F(opt) estimator. JN2SLS also has smaller bias than the 2SLS estimator, as expected. In general, LIML seems to have the smallest median bias, at least under homoscedastic design. This is in agreement with the well-known fact that LIML has a zero higher order median bias. On the other hand, the actual median bias of LIML is not equal to zero, which can be explained by the alternative asymptotic analysis put forth by Staiger and Stock (1997). See Section 5.3.

4.3.2 MSE

For LIML, F(1), F(4), and F(opt), the approximate MSE is equal to

$$\frac{1-R^2}{nR^2} + K\frac{1-\rho^2}{n^2}\left(\frac{1-R^2}{R^2}\right)^2 + O\left(\frac{1}{n^2}\right) \tag{6}$$

To allow for a more refined expression for the MSE of the Fuller estimators, we also calculate the MSE of F(4) using the approach of Rothenberg (1983):

$$\frac{1-R^2}{nR^2} + \rho^2\frac{-1-K}{n^2}\left(\frac{1-R^2}{R^2}\right)^2 + \frac{K-6}{n^2}\left(\frac{1-R^2}{R^2}\right)^2 \tag{7}$$

For Nagar, JN2SLS, and JIVE, the approximate MSE is equal to

$$\frac{1-R^2}{nR^2} + K\frac{1+\rho^2}{n^2}\left(\frac{1-R^2}{R^2}\right)^2 + O\left(\frac{1}{n^2}\right) \tag{8}$$

Thus, note that JN2SLS has the same MSE as Nagar or JIVE, but JN2SLS has finite sample moments as we demonstrated above. For 2SLS, the MSE is equal to

$$\frac{1-R^2}{nR^2} + K^2\frac{\rho^2}{n^2}\left(\frac{1-R^2}{R^2}\right)^2 + O\left(\frac{1}{n^2}\right) \tag{9}$$

The first order term is identical for all the estimators as is well known. The second order terms depend on the number of instruments as K and for 2SLS as K^2. Note that for 2SLS the second order term is the square of the bias term of 2SLS from Hahn and Hausman (2001).

When we turn to the empirical results, we find that the theory does not give especially good guidance to the actual empirical results. Although Nagar is supposed to be equivalent to JN2SLS, it is not and performs considerably worse than JN2SLS when the model is weakly identified. Presumably the lack of moments invalidates the Nagar calculations. Indeed we strongly recommend that the "no-moment" estimators LIML, Nagar, and JIVE not be used in weak instrument situations. The ordering of the empirical MSE of the Fuller estimators is in accord with the higher order theory as discussed by Rothenberg (1983) and in the Appendix C. If we compare the best of the feasible Fuller estimators F(4) to JN2SLS, the F(4) estimator does better with a small number of instruments, but JN2SLS often does better when the number of instrument increases. However, we might give a "slight nod" to the F(4) estimator over JN2SLS here. Note that 2SLS turns in a respectable performance here, also.

2SLS tends to dominate OLS in terms of both bias and RMSE except most extreme cases very small R^2.

4.3.3 Interquartile Range (IQR)

We think that the IQR is a useful measure since extreme results do not matter. Thus, a reasonable conjecture is that the "no moment" estimators would be superior with respect to the IQR. This is not what we find however. Instead, LIML, Nagar, and JIVE are all found to have significantly larger IQR than the other estimators. Since the "no-moment" estimators also have inferior empirical mean bias and MSE performance, we suggest that they are not useful estimators in the weak instrument situation. For the IQR we find that the F(4) estimator does significantly better than the F(1) estimator. 2SLS does better than JN2SLS for the IQR, but often not by large amounts. The F(4) estimator has no ordering with respect to 2SLS and JN2SLS.

Based on the mean bias, the MSE, and the IQR we find no overall ordering among the 2SLS, F(4), and JN2SLS estimators. However, these estimators perform better than the "no moment" estimators. We suggest that the Fuller estimator receive more attention and use than it seems to have received to date. We also suggest that the 2SLS estimator and the JN2SLS be calculated in a weak instrument situation. These three estimators seem to have the best properties of the estimators we investigated. Overall, our finding is that 2SLS does better than would be expected based on the theoretical calculations.

4.3.4 A Heteroscedastic Design

We now consider a heteroscedastic design where $E\left[\varepsilon_i^2 \middle| z_i\right] = z_i' z_i / K$. We only consider F(4), 2SLS, and JN2SLS because the "no-moment" estimators continue to have similar problems as in the homoscedastic case. We find that in terms of mean bias that JN2SLS does better than either F(4) or 2SLS. For MSE, F(4) often does better than JN2SLS, but also often does considerably worse. 2SLS often does better than JN2SLS. Based on the MSE we thus again suggest considering all three estimators. Our suggested use of all three estimators remains the same based on IQR. Thus, the use of a heteroscedastic design continues to lead to the same suggestion as the homoscedastic design.

4.3.5 How Important are Outliers?

In Table 5 we analyze the importance of outliers by comparing the $5-95\%$ range of the empirical distribution of our estimators to the range implied by the asymptotic distribution. The table shows that the no-moment estimators, LIML, Nagar and Jive, have severely inflated ranges relative to their asymptotic distribution when $R^2 = 0.01$. This suggests that for the weak instrument case the nonexistence of moments affects the entire distribution and is not a feature of extreme tail behavior alone. On the other hand, the estimators with known finite moments, Fuller, 2SLS and JN2SLS do not show inter quantile ranges larger than predicted by the asymptotic distribution.

5 How Well Do the Higher Order Formulae Explain the Data?

All of our bias and MSE formula are higher order asymptotic expansions to $O\left(1/n^2\right)$. We have already ascertained that for the "no-moment" estimators the formulae are not useful in the weak instrument situation. More generally, we have determined above from the Monte Carlo results that the asymptotic expansions may not provide especially good guidance in the weak instrument situation. Thus, we now test the asymptotic expansions given the data obtained from the Monte Carlo experiments. We consider the formulae in two respects. We first take the MSE formulae given above and run a regression, using our Monte Carlo design results, of the empirical MSE on the theory predictions. We use a constant, which should be zero, and an intercept, which should be one, if the formulae hold true. We then alternatively run a regression using the first and second order terms separately from the MSE formulae. Each of the coefficients should be unity. This latter approach allows us to sort out the first and second order terms.

5.1 Basic Regression Results

We first run the "0-1" regression with a constant and a coefficient for the MSE formulae that we derived for the estimators. The results should be the constant=0 and the intercept coefficient =1 if the formulae are correct for our Monte Carlo weak instrument design. The results are given in Table 8.

Even for the estimators with finite sample moments, the higher formulae are all rejected since none of the intercepts equals anywhere near unity. The JN2SLS, Fuller (4) and 2SLS have some predictive power.

5.2 Further Regression Results

We now repeat the regressions, but we separate the RHS into the two terms corresponding to the first order and second order terms in the approximate MSE formulae. The first term with coefficient C1 is the first order term while the next term with coefficient C2 is the second order term: We present the results in Table 9.

All the coefficients should be unity if the formulae are correct. None of the estimates are unity. The first order terms are most important, as expected. The second order terms are typically small in magnitude, but often significant. However, the signs of the second order coefficients for JN2SLS and F(4) are incorrect while the second order coefficient for 2SLS is very small and not significant. The fit of the regression is improved by dividing up the terms. Thus, the second order terms do not do a good job in explaining the empirical results.

5.3 Numerical Calculation of the Median of the Weak Instrument Limit Distribution of LIML

Our Monte Carlo result can be related to the alternative asymptotic analysis put forth by Staiger and Stock (1997). Under their alternative approximation, our model is such that

$$y_i = x_i\beta + \varepsilon_i, \quad x_i = z_i'\pi + u_i$$

where $z_i \sim N(0, I_K)$, $v_i = (\varepsilon_i, u_i)'$ with $v_i \sim N(0, \Sigma)$ and

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

$\pi = (\eta, ..., \eta)' / \sqrt{n}$ for some constant η. Define $\lambda = \sqrt{n}(\eta, ..., \eta)'$, and let $\widetilde{\phi}$ be the smallest eigenvalue of the matrix $W'W(W'MW)^{-1}$. Denote by $\overline{\phi}$ the limit of $n\left(\widetilde{\phi} - 1\right)$. As in Staiger and Stock (1997), we may then assume that $\left(n\left(\widetilde{\phi} - 1\right), n^{-1/2}Z'\varepsilon, n^{-1/2}Z'u\right) \Rightarrow \left(\overline{\phi}, \Psi_{Z\varepsilon}, \Psi_{Zu}\right)$, where $\Psi_{Z\varepsilon}, \Psi_{Zu}$ have a marginal normal distribution $(\Psi_{Z\varepsilon}, \Psi_{Zu})' \sim N(0, \Sigma \otimes I_k)$. The limiting distribution of the LIML estimator under local to zero asymptotics then is defined by letting $v_1 = (\lambda + \Psi_{Zu})'(\lambda + \Psi_{Zu})$ and $v_2 = (\lambda + \Psi_{Zu})'\Psi_{Z\varepsilon}$, where $\lambda = (\eta, ..., \eta)'$. As Staiger and Stock (1997) show,

$$\beta_{LIML} - \beta \Rightarrow \left(v_2 - \overline{\phi}\rho\right) / \left(v_1 - \overline{\phi}\right).$$

We draw samples from the distribution of $(v_2 - \kappa\rho)/(v_1 - \kappa)$ by generating y, x, Z according to (1) for sample sizes $n = (100, 500, 1000)$. We chose $r = \{.01, .1, .3\}$ and $K = \{5, 10, 30\}$. We use $\left(n\left(\widetilde{\phi} - 1\right), n^{-1/2}Z'\varepsilon, n^{-1/2}Z'u\right)$ as an approximation to $\left(\overline{\phi}, \Psi_{Z\varepsilon}, \Psi_{Zu}\right)$ and compute

$$\beta_i^* = \frac{\left(\lambda + n^{-1/2}Z'u\right)' n^{-1/2}Z'\varepsilon - n(k-1)\rho}{\left(\lambda + n^{-1/2}Z'u\right)'\left(\lambda + n^{-1/2}Z'u\right) - n(k-1)}$$

for each Monte Carlo replication. We compute the median of the empirical distribution of $\beta_i^*, i = 1, ..., S$ as an approximation to the median of $\left(v_2 - \overline{\phi}\rho\right)/\left(v_1 - \overline{\phi}\right)$. The results for 10,000 replications are summarized in Table 11. The simulation results indicate that the approximation of Staiger and Stock (1997) implies the presence of a median bias for LIML when the concentration parameter λ is small. In those instances the median bias is also an increasing function of the number of instruments K. The numerical evaluation of the Staiger and Stock (1997) local to zero approximation is quite close to the actual finite sample distribution of LIML obtained by Monte Carlo procedures as far as median bias properties are concerned.

5.4 An Empirical Exploration

Our last set of empirical analysis consists of regressing the log of the MSE on the logs of the determinants of the MSE: n, K, ρ, and $\frac{1-R^2}{R^2} = Ratio$. The results are given in Table 10.

The effects of the number or observations n and R^2 have the expected magnitude and are estimated quite precisely-the first order effect dominates as we would expect. However, the second order effects of the correlation coefficient (squared) and the number of instruments are considerably less important. While the number of instruments is most important for 2SLS as the second order MSE formulae predict, the estimated coefficient is far below 2.0, which is the theoretical prediction. Perhaps the most important finding is that the effect of the number of instruments is considerably less than expected. Thus, "number of instruments pessimism" that arises from the second order formulae on the asymptotic bias seems to be overdone. This finding is consistent with our results that 2SLS does better than expected in many situations.

Lastly, we run a regression with the same RHS variables as controls but with the LHS side variable the log of MSE and additional RHS variables as indicator variables. Thus, we run a "horse race" among the different estimators. For the log of MSE estimators we find the "no moments" estimators to have significantly higher log MSE than the baseline estimator, 2SLS. We find 2SLS significantly better than all of the other estimators except JN2SLS. JN2SLS has a smaller log MSE than 2SLS. Both estimators are significantly better than Fuller (4) which, in turn, is significantly better than the "no moments" estimators. For log IQR we find that 2SLS is insignificantly better than F(4), which in turn is insignificantly better than JN2SLS. No significant difference exist for the three estimators with respect to log IQR. The "no moments" estimators do significantly worse than these three estimators. Thus, the choice of estimator may depend on whether the researcher is interested more in the entire distribution as given by the MSE or in the IQR. The overall finding is that the F(4) and JN2SLS should be used along with 2SLS in the weak instruments situation.

Appendix

A Higher Order Expansion

We first present two Lemmas:

Lemma 1 *Let υ_i be a sample of n independent random variables with $\max_i E\left[|\upsilon_i|^r\right] < c^r < \infty$ for some constant $0 < c < \infty$ and some $1 < r < \infty$. Then $\max_i |\upsilon_i| = O_p\left(n^{1/r}\right)$.*

Proof. By Jensen's inequality, we have

$$
\begin{aligned}
E\left[\max_i |\upsilon_i|\right] &\leq \left(E\left[\max_i |\upsilon_i|^r\right]\right)^{1/r} \leq \left(\sum_i E\left[|\upsilon_i|^r\right]\right)^{1/r} \\
&\leq \left(n \max_i E\left[|\upsilon_i|^r\right]\right)^{1/r} = n^{1/r}\left(\max_i E\left[|\upsilon_i|^r\right]\right)^{1/r} \leq n^{1/r} c
\end{aligned}
$$

The conclusion follows by Markov inequality. ■

Lemma 2 *Assume that Conditions 2 and 3 are satisfied. Further assume that $E\left[|\varepsilon_i|^r\right] < \infty$ and $E\left[|u_i|^r\right] < \infty$ for r sufficiently large ($r > 12$). We then have (i) $n^{-1/6} \max|\delta_{1i}| = o_p(1)$ and $n^{-1/6} \max|\delta_{2i}| = o_p(1)$; and (ii) $\frac{1}{n\sqrt{n}} \sum_i \left|\delta_{1i}\delta_{2i}^2\right| = o_p(1)$ and $\frac{1}{n\sqrt{n}} \sum_i \left|\delta_{2i}^3\right| = o_p(1)$.*

Proof. Note that

$$
\begin{aligned}
\max|\delta_{1i}| \leq\ & (\max|f_i| + \max|u_i|) \cdot \max|\varepsilon_i| \\
& + \max(1 - P_{ii})^{-1} \cdot \left(\max|u_i| + \max\left|(Pu)_i\right|\right) \cdot \left(\max|\varepsilon_i| + \max\left|(P\varepsilon)_i\right|\right),
\end{aligned}
$$

We have $(\max|f_i| + \max|u_i|) \cdot \max|\varepsilon_i| = O_p\left(n^{2/r}\right)$ by Lemma 1. Because $\max\left|(Pu)_i\right|^2 \leq \max P_{ii} \cdot u'u$, and $\max P_{ii} = O\left(\frac{1}{n}\right)$, we also have $\max\left|(Pu)_i\right| = O_p(1)$. Similarly, $\max\left|(P\varepsilon)_i\right| = O_p(1)$. Therefore, we obtain we obtain $\max|\delta_{1i}| = o_p\left(n^{1/6}\right)$. That $\max|\delta_{2i}| = o_p\left(n^{1/6}\right)$ can be established similarly. It then easily follows that $\frac{1}{n\sqrt{n}} \sum_i \left|\delta_{1i}\delta_{2i}^2\right| \leq \frac{1}{\sqrt{n}} \max|\delta_{1i}| \max|\delta_{2i}|^2 = o_p(1)$, and $\frac{1}{n\sqrt{n}} \sum_i \left|\delta_{2i}^3\right| \leq \frac{1}{\sqrt{n}} \max|\delta_{2i}|^3 = o_p(1)$. ■

We note from Donald and Newey (2001) that we have the following expansion[7]:

$$
H\sqrt{n}\frac{x'P\varepsilon}{x'Px} = \sum_{j=1}^{7} T_j + o_p\left(\frac{K}{n}\right), \tag{10}
$$

[7] Our representation of Donald and Newey's result reflects our simplifying assumption that the first stage is correctly specified.

where

$$T_1 = h = O_p(1), \qquad T_2 = W_1 = O_p\left(\frac{K}{\sqrt{n}}\right), \qquad T_3 = -W_3 \frac{1}{H} h = O_p\left(\frac{1}{\sqrt{n}}\right),$$
$$T_4 = 0, \qquad T_5 = -W_4 \frac{1}{H} h = O_p\left(\frac{K}{n}\right), \qquad T_6 = -W_3 \frac{1}{H} W_1 = O_p\left(\frac{K}{n}\right),$$
$$T_7 = W_3^2 \frac{1}{H^2} h = O_p\left(\frac{1}{n}\right),$$

and

$$\begin{aligned} h &= \frac{f'\varepsilon}{\sqrt{n}} = O_p(1), \qquad W_1 = \frac{u'P\varepsilon}{\sqrt{n}} = O_p\left(\frac{K}{\sqrt{n}}\right), \\ W_3 &= 2\frac{f'u}{n} = O_p\left(\frac{1}{\sqrt{n}}\right), \qquad W_4 = \frac{u'Pu}{n} = O_p\left(\frac{K}{n}\right). \end{aligned}$$

We now expand $\frac{H}{\frac{1}{n}x'Px}$ and $\left(\frac{H}{\frac{1}{n}x'Px}\right)^2$ up to $O_p\left(\frac{1}{n}\right)$. Because $\frac{1}{n}x'Px = H + W_3 + W_4$, we have

$$\frac{H}{\frac{1}{n}x'Px} = \frac{H}{H + W_3 + W_4} = 1 - \frac{1}{H}W_3 - \frac{1}{H}W_4 + \frac{1}{H^2}W_3^2 + o_p\left(\frac{K}{n}\right), \tag{11}$$

$$\left(\frac{H}{\frac{1}{n}x'Px}\right)^2 = 1 - 2\frac{1}{H}W_3 - 2\frac{1}{H}W_4 + 3\frac{1}{H^2}W_3^2 + o_p\left(\frac{K}{n}\right) \tag{12}$$

We now expand $\frac{1}{\sqrt{n}}\sum_i \delta_{1i}$. Observe that

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_i \delta_{1i} &= -\frac{1}{\sqrt{n}}\sum_i x_i\varepsilon_i + \frac{1}{\sqrt{n}}\sum_i (1 - P_{ii})^{-1}(Mx)_i (M\varepsilon)_i \\ &= -h - \frac{1}{\sqrt{n}}u'\varepsilon + \frac{1}{\sqrt{n}}(Mu)'\left(I - \widetilde{P}\right)^{-1}(M\varepsilon) \\ &= -h - \frac{1}{\sqrt{n}}u'\varepsilon + \frac{1}{\sqrt{n}}u'M\varepsilon + \frac{1}{\sqrt{n}}(Mu)'\overline{P}(M\varepsilon) \\ &= -h - \frac{1}{\sqrt{n}}u'P\varepsilon + \frac{1}{\sqrt{n}}u'\overline{P}\varepsilon - \frac{1}{\sqrt{n}}u'P\overline{P}\varepsilon - \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon + \frac{1}{\sqrt{n}}u'P\overline{P}P\varepsilon \\ &= -h - \frac{1}{\sqrt{n}}u'C'\varepsilon - \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon + \frac{1}{\sqrt{n}}u'P\overline{P}P\varepsilon, \end{aligned} \tag{13}$$

where, as in Donald and Newey (2001), we let

$$C \equiv P - \overline{P}(I - P) = P - \overline{P}M, \quad \overline{P} \equiv \widetilde{P}\left(I - \widetilde{P}\right)^{-1},$$

and $\widetilde{P}$ is a diagonal matrix with element P_{ii} on the diagonal. Now, note that, by Cauchy-Schwartz, $\left|u'\overline{P}P\varepsilon\right| \leq \sqrt{u'u}\sqrt{\varepsilon'P\overline{P}^2P\varepsilon}$. Because $u'u = O_p(n)$, and $\varepsilon'P\overline{P}^2P\varepsilon \leq \max\left(\frac{P_{ii}}{1-P_{ii}}\right)^2 \varepsilon'P\varepsilon = O\left(\frac{1}{n^2}\right)O_p(K)$, we obtain

$$\begin{aligned} \left|\frac{u'\overline{P}P\varepsilon}{\sqrt{n}}\right| &\leq \frac{1}{\sqrt{n}}\sqrt{u'u}\sqrt{\varepsilon'P\overline{P}^2P\varepsilon} = \frac{1}{\sqrt{n}}\sqrt{O_p(n)}\sqrt{O\left(\frac{1}{n^2}\right)O_p(K)} = O_p\left(\frac{\sqrt{K}}{n}\right), \\ \left|\frac{u'P\overline{P}P\varepsilon}{\sqrt{n}}\right| &\leq \frac{1}{\sqrt{n}}\sqrt{u'Pu}\sqrt{\varepsilon'P\overline{P}^2P\varepsilon} = \frac{1}{\sqrt{n}}\sqrt{O_p(K)}\sqrt{O\left(\frac{1}{n^2}\right)O_p(K)} = O_p\left(\frac{K}{n^{3/2}}\right) = o_p\left(\frac{K}{n}\right). \end{aligned}$$

To conclude, we can write

$$\frac{1}{\sqrt{n}}\sum_i \delta_{1i} = -h + W_5 + W_6 + o_p\left(\frac{K}{n}\right), \tag{14}$$

where

$$\begin{aligned} W_5 &\equiv -\frac{1}{\sqrt{n}}u'C'\varepsilon = O_p\left(\frac{\sqrt{K}}{\sqrt{n}}\right) \\ W_6 &\equiv \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon = O_p\left(\frac{\sqrt{K}}{n}\right). \end{aligned}$$

We now expand $\left(\frac{H}{\frac{1}{n}x'Px}\right)\left(\frac{1}{\sqrt{n}}\sum_i \delta_{1i}\right)$ using (11) and (14):

$$\begin{aligned} &\left(\frac{H}{\frac{1}{n}x'Px}\right)\left(\frac{1}{\sqrt{n}}\sum_i \delta_{1i}\right) \\ &= \left(1 - \frac{1}{H}W_3 - \frac{1}{H}W_4 + \frac{1}{H^2}W_3^2\right)(-h + W_5 + W_6) + o_p\left(\frac{K}{n}\right) \\ &= -h + W_3\frac{1}{H}h + W_4\frac{1}{H}h - W_3^2\frac{1}{H^2}h + W_5 + W_6 - \frac{1}{H}W_3W_5 + o_p\left(\frac{K}{n}\right) \\ &= -T_1 - T_3 - T_5 - T_7 + T_8 + T_9 + T_{10} + o_p\left(\frac{K}{n}\right) \end{aligned} \tag{15}$$

where

$$\begin{aligned} T_8 &\equiv W_5 = -\frac{1}{\sqrt{n}}u'C'\varepsilon = O_p\left(\frac{\sqrt{K}}{\sqrt{n}}\right), \\ T_9 &\equiv W_6 = \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon = O_p\left(\frac{\sqrt{K}}{n}\right), \\ T_{10} &\equiv -\frac{1}{H}W_3W_5 = W_3\frac{1}{H}\frac{1}{\sqrt{n}}u'C'\varepsilon = O_p\left(\frac{\sqrt{K}}{n}\right). \end{aligned}$$

We now expand $\frac{H}{\frac{1}{n}x'Px}\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n}\sum_i \delta_{2i}\right)$. We begin with expansion of $\frac{1}{n}\sum_i \delta_{2i}$. As in (13), we can show that

$$\frac{1}{n}\sum_i \delta_{2i} = -H - \frac{2}{n}f'u - \frac{1}{n}u'C'u - \frac{1}{n}u'\overline{P}Pu + \frac{1}{n}u'P\overline{P}Pu$$

Because

$$\begin{aligned} \left|u'P\overline{P}Pu\right| &\leq \max\left(\frac{P_{ii}}{1-P_{ii}}\right)\cdot u'Pu = O_p\left(\frac{K}{n}\right), \\ \left|u'\overline{P}Pu\right| &\leq \sqrt{u'u}\sqrt{u'P\overline{P}^2Pu} \leq \sqrt{O_p(n)}\sqrt{\max\left(\frac{P_{ii}}{1-P_{ii}}\right)^2\cdot u'Pu} = O_p\left(\sqrt{\frac{K}{n}}\right), \end{aligned}$$

we may write

$$\frac{1}{n}\sum_i \delta_{2i} = -H - W_3 - W_7 + o_p\left(\frac{K}{n}\right) \tag{16}$$

where

$$W_7 \equiv \frac{1}{n}u'C'u = O_p\left(\frac{\sqrt{K}}{n}\right).$$

Combining (11) and (16), we obtain

$$\begin{aligned}
\frac{H}{\frac{1}{n}x'Px}\left(\frac{1}{n}\sum_i \delta_{2i}\right) &= \left(1 - \frac{1}{H}W_3 - \frac{1}{H}W_4 + \frac{1}{H^2}W_3^2\right)(-H - W_3 - W_7) + o_p\left(\frac{K}{n}\right) \\
&= -H + W_3 + W_4 - \frac{1}{H}W_3^2 - W_3 + \frac{1}{H}W_3^2 - W_7 + o_p\left(\frac{K}{n}\right) \\
&= -H + W_4 - W_7 + o_p\left(\frac{K}{n}\right)
\end{aligned}$$

which, combined with (10), yields

$$\begin{aligned}
\frac{H}{\frac{1}{n}x'Px}\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n}\sum_i \delta_{2i}\right) &= \frac{1}{H}\left(\sum_{j=1}^{7} T_j\right)(-H + W_4 - W_7) + o_p\left(\frac{K}{n}\right) \\
&= -\sum_{j=1}^{7} T_j + W_4\frac{1}{H}h - W_7\frac{1}{H}h + o_p\left(\frac{K}{n}\right) \\
&= -\sum_{j=1}^{7} T_j - T_5 + T_{11} + o_p\left(\frac{K}{n}\right) \qquad (17)
\end{aligned}$$

where

$$T_{11} \equiv -W_7\frac{1}{H}h = O_p\left(\frac{\sqrt{K}}{n}\right).$$

We now examine $\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{1}{n\sqrt{n}}\sum_i \delta_{1i}\delta_{2i}\right)$. Later in Section B.2.1, it is shown that

$$\frac{1}{n\sqrt{n}}\sum_i \delta_{1i}\delta_{2i} = O_p\left(\frac{1}{\sqrt{n}}\right)$$

Therefore, we should have

$$\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{1}{n\sqrt{n}}\sum_i \delta_{1i}\delta_{2i}\right) = T_{12} + o_p\left(\frac{K}{n}\right) \tag{18}$$

where

$$T_{12} \equiv \frac{1}{H}\frac{1}{n\sqrt{n}}\sum_i \delta_{1i}\delta_{2i} = O_p\left(\frac{1}{\sqrt{n}}\right)$$

Now, we examine $\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n^2}\sum_i \delta_{2i}^2\right)$. Later in Section B.2.3, it is shown that

$$\frac{1}{n^2}\sum_i \delta_{2i}^2 = O_p\left(\frac{1}{n}\right).$$

Therefore, we have

$$\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n^2}\sum_i \delta_{2i}^2\right) = T_{14} + o_p\left(\frac{K}{n}\right) \tag{19}$$

where

$$T_{14} \equiv \frac{1}{H^2}h\frac{1}{n^2}\sum_i \delta_{2i}^2 = O_p\left(\frac{1}{n}\right)$$

Combining (3), (10), (15), (17), (18), and (19), we obtain

$$\begin{aligned} & H\sqrt{n}\left(\frac{x'P\varepsilon}{x'Px} - \frac{n-1}{n}\sum_i\left(\frac{x'P\varepsilon+\delta_{1i}}{x'Px+\delta_{2i}} - \frac{x'P\varepsilon}{x'Px}\right) + R_n\right) \\ = \; & T_1 + T_3 + T_7 - T_8 - T_9 - T_{10} + T_{11} + T_{12} - T_{14} + o_p\left(\frac{K}{n}\right). \end{aligned} \tag{20}$$

B Approximate MSE Calculation

In computing the (approximate) mean squared error, we keep terms up to $O_p\left(\frac{1}{n}\right)$. From (20), we can see that the MSE of the jackknife estimator approximately equal to

$$\begin{aligned} & E\left[T_1^2\right] + E\left[T_3^2\right] + E\left[T_8^2\right] + E\left[T_{12}^2\right] \\ & +2E\left[T_1T_3\right] + 2E\left[T_1T_7\right] - 2E\left[T_1T_8\right] - 2E\left[T_1T_9\right] - 2E\left[T_1T_{10}\right] + 2E\left[T_1T_{11}\right] \\ & +2E\left[T_1T_{12}\right] - 2E\left[T_1T_{14}\right] - 2E\left[T_3T_8\right] \end{aligned} \tag{21}$$

Combining (21) with (22), (23), (24), (25), (26), (27), (28), (29), (30), (33), (44), and (45) in the next two subsections, it can be shown that the approximate MSE up to $O_p\left(\frac{1}{n}\right)$ is given by

$$\sigma_\varepsilon^2 H + \frac{K}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + \sigma_{u\varepsilon}^2\right) + O\left(\frac{1}{n}\right),$$

which proves Theorem 1.

B.1 Approximate MSE Calculation: Intermediate Results That Only Require Symmetry

From Donald and Newey (2001), we can see that

$$
\begin{aligned}
E\left[T_1^2\right] &= \sigma_\varepsilon^2 H && (22)\\
E\left[T_3^2\right] &= \frac{4}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right) + o\left(\frac{1}{n}\right) && (23)\\
E\left[T_1T_3\right] &= 0 && (24)\\
E\left[T_1T_7\right] &= \frac{4}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right) + o\left(\frac{1}{n}\right) && (25)\\
E\left[T_8^2\right] &= \frac{K}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + \sigma_{u\varepsilon}^2\right) + o\left(\frac{K}{n}\sup P_{ii}\right) && (26)
\end{aligned}
$$

Also, by symmetry, we have

$$
\begin{aligned}
E\left[T_1T_8\right] &= 0 && (27)\\
E\left[T_1T_9\right] &= 0. && (28)
\end{aligned}
$$

It remains to compute $E\left[T_{12}^2\right]$, $E\left[T_1T_{10}\right]$, $E\left[T_1T_{11}\right]$, $E\left[T_1T_{12}\right]$, $E\left[T_1T_{14}\right]$, and $E\left[T_3T_8\right]$. We will take care of $E\left[T_{12}^2\right]$, $E\left[T_1T_{12}\right]$, and $E\left[T_1T_{14}\right]$ in the next section.

Note that

$$
E\left[T_1T_{10}\right] = E\left[T_3T_8\right] = E\left[2\frac{f'u}{n}\frac{1}{H}\frac{1}{\sqrt{n}}u'C'\varepsilon\frac{f'\varepsilon}{\sqrt{n}}\right] = \frac{2}{n^2H}E\left[u'f'f\varepsilon\cdot u'C'\varepsilon\right]
$$

Using equation (18) of Donald and Newey (2001), we obtain

$$
\begin{aligned}
E\left[u'f'f\varepsilon\cdot u'C'\varepsilon\right] &= \sum_{i=1}^{n}E\left[u_i^2\varepsilon_i^2f_i^2C'_{ii}\right] + \sum_{i=1}^{n}\sum_{j\neq i}E\left[u_i\varepsilon_iu_j\varepsilon_jf_i^2C'_{jj}\right]\\
&\quad + \sum_{i=1}^{n}\sum_{j\neq i}E\left[u_i^2\varepsilon_j^2f_if_jC'_{ij}\right] + \sum_{i=1}^{n}\sum_{j\neq i}E\left[u_i\varepsilon_ju_j\varepsilon_if_if_jC'_{ji}\right]\\
&= \sigma_u^2\sigma_\varepsilon^2\sum_{i=1}^{n}\sum_{j\neq i}f_if_jC'_{ij} + \sigma_{u\varepsilon}^2\sum_{i=1}^{n}\sum_{j\neq i}f_if_jC'_{ji}\\
&= \sigma_u^2\sigma_\varepsilon^2f'C'f + \sigma_{u\varepsilon}^2f'Cf
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
E\left[T_1T_{10}\right] &= E\left[T_3T_8\right] = \frac{2}{nH}\frac{\sigma_u^2\sigma_\varepsilon^2f'C'f + \sigma_{u\varepsilon}^2f'Cf}{n}\\
&= \frac{2}{nH}\left(\sigma_u^2\sigma_\varepsilon^2H + \sigma_{u\varepsilon}^2H + o\left(\frac{1}{n}\right)\right) = \frac{2}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + \sigma_{u\varepsilon}^2\right) + o\left(\frac{1}{n^2}\right), && (29)
\end{aligned}
$$

where the second equality is based on equation (20) of Donald and Newey (2001).

Now, note that

$$E[T_1T_{11}] = -\frac{1}{n^2H}E[u'Cu \cdot \varepsilon' ff'\varepsilon]$$

and

$$\begin{aligned} E[\varepsilon' ff'\varepsilon \cdot u'Cu] &= \sum_{i=1}^{n} E[u_i^2\varepsilon_i^2 f_i^2 C_{ii}] + \sum_{i=1}^{n}\sum_{j\neq i} E[\varepsilon_i^2 u_j^2 f_i^2 C_{jj}] \\ &\quad + \sum_{i=1}^{n}\sum_{j\neq i} E[\varepsilon_i\varepsilon_j u_i u_j f_i f_j C_{ij}] + \sum_{i=1}^{n}\sum_{j\neq i} E[\varepsilon_i\varepsilon_j u_j u_i f_i f_j C_{ji}] \\ &= \sigma_{u\varepsilon}^2 f'C'f + \sigma_{u\varepsilon}^2 f'Cf \end{aligned}$$

Because $Cf = Pf - \overline{P}(I-P)f = PZ\pi - \overline{P}(I-P)Z\pi = Z\pi = f$, we obtain

$$E[T_1T_{11}] = -2\frac{\sigma_{u\varepsilon}^2}{n}. \tag{30}$$

B.2 Approximate MSE Calculation: Intermediate Results Based On Normality

Note that

$$\begin{aligned} \delta_{1i}\delta_{2i} &= x_i^3\varepsilon_i + (1-P_{ii})^{-2}(Mu)_i^3(M\varepsilon)_i \\ &\quad - (1-P_{ii})^{-1} x_i\varepsilon_i (Mu)_i^2 - (1-P_{ii})^{-1} x_i^2 (Mu)_i (M\varepsilon)_i \\ &= f_i^3\varepsilon_i + 3f_i^2 u_i\varepsilon_i + 3f_i u_i^2\varepsilon_i + u_i^3\varepsilon_i \\ &\quad + (1-P_{ii})^{-2}(Mu)_i^3(M\varepsilon)_i - (1-P_{ii})^{-1} f_i\varepsilon_i (Mu)_i^2 \\ &\quad - (1-P_{ii})^{-1} u_i\varepsilon_i (Mu)_i^2 - (1-P_{ii})^{-1} f_i^2 (Mu)_i (M\varepsilon)_i \\ &\quad -2(1-P_{ii})^{-1} f_i u_i (Mu)_i (M\varepsilon)_i - (1-P_{ii})^{-1} u_i^2 (Mu)_i (M\varepsilon)_i \end{aligned} \tag{31}$$

and

$$\begin{aligned} \delta_{2i}^2 &= \left(-f_i^2 - 2f_i u_i - u_i^2 + (1-P_{ii})^{-1}(Mu)_i^2\right)^2 \\ &= f_i^4 + 6f_i^2u_i^2 + u_i^4 + (1-P_{ii})^{-2}(Mu)_i^4 \\ &\quad +4f_i^3u_i - 2f_i^2(1-P_{ii})^{-1}(Mu)_i^2 + 4f_iu_i^3 \\ &\quad -4f_iu_i(1-P_{ii})^{-1}(Mu)_i^2 - 2(1-P_{ii})^{-1}u_i^2(Mu)_i^2 \end{aligned} \tag{32}$$

B.2.1 $E\left[T_{12}^2\right]$

We first compute $E\left[T_{12}^2\right]$ noting that

$$\begin{aligned}
H^2 E\left[T_{12}^2\right] \leq\; & \frac{10}{n^2}\sum_i f_i^6 E\left[(\varepsilon_i)^2\right] + \frac{10}{n^2}\sum_i 9 f_i^4 E\left[(u_i\varepsilon_i)^2\right] \\
& + \frac{10}{n^2}\sum_i 9 f_i^2 E\left[\left(u_i^2\varepsilon_i\right)^2\right] + \frac{10}{n^2}\sum_i E\left[\left(u_i^3\varepsilon_i\right)^2\right] \\
& + \frac{10}{n^2}\sum_i (1-P_{ii})^{-4} E\left[(Mu)_i^6 (M\varepsilon)_i^2\right] \\
& + \frac{10}{n^2}\sum_i (1-P_{ii})^{-2} f_i^2 E\left[\varepsilon_i^2 (Mu)_i^4\right] \\
& + \frac{10}{n^2}\sum_i (1-P_{ii})^{-2} E\left[u_i^2\varepsilon_i^2 (Mu)_i^4\right] \\
& + \frac{10}{n^2}\sum_i (1-P_{ii})^{-2} f_i^4 E\left[(Mu)_i^2 (M\varepsilon)_i^2\right] \\
& + \frac{10}{n^2}\sum_i 4(1-P_{ii})^{-2} f_i^2 E\left[u_i^2 (Mu)_i^2 (M\varepsilon)_i^2\right] \\
& + \frac{10}{n^2}\sum_i (1-P_{ii})^{-2} E\left[u_i^4 (Mu)_i^2 (M\varepsilon)_i^2\right]
\end{aligned}$$

Under the assumption that $\frac{1}{n}\sum_i f_i^6 = O(1)$, the first four terms are all $O\left(\frac{1}{n}\right)$. Below, we characterize orders of the rest of the terms.

We now compute $\frac{10}{n^2}\sum_i (1-P_{ii})^{-4} E\left[(Mu)_i^6 (M\varepsilon)_i^2\right]$. We write

$$\varepsilon_i \equiv \frac{\sigma_{u\varepsilon}}{\sigma_u^2} u_i + v_i,$$

where v_i is independent of u_i. Because

$$\begin{aligned}
& (1-P_{ii})^{-4} E\left[(Mu)_i^6 (M\varepsilon)_i^2\right] \\
&= (1-P_{ii})^{-4}\left(\frac{\sigma_{u\varepsilon}^2}{\sigma_u^4} 105\,\mathrm{Var}\left((Mu)_i\right)^4 + 15\,\mathrm{Var}\left((Mu)_i\right)^3 \mathrm{Var}\left((Mv)_i\right)\right) \\
&= (1-P_{ii})^{-4}\left(105\frac{\sigma_{u\varepsilon}^2}{\sigma_u^4}(1-P_{ii})^4\sigma_u^8 + 15(1-P_{ii})^3\sigma_u^6(1-P_{ii})\left(\sigma_\varepsilon^2 - \frac{\sigma_{u\varepsilon}^2}{\sigma_u^2}\right)\right) \\
&= 15\sigma_\varepsilon^2\sigma_u^6 + 90\sigma_{u\varepsilon}^2\sigma_u^4,
\end{aligned}$$

we have

$$\frac{10}{n^2}\sum_i (1-P_{ii})^{-4} E\left[(Mu)_i^6 (M\varepsilon)_i^2\right] = \frac{10}{n^2}\sum_i \left(15\sigma_\varepsilon^2\sigma_u^6 + 90\sigma_{u\varepsilon}^2\sigma_u^4\right) = O\left(\frac{1}{n}\right).$$

We now compute $\frac{10}{n^2}\sum_i (1-P_{ii})^{-2} f_i^2 E\left[\varepsilon_i^2 (Mu)_i^4\right]$. Because

$$\begin{aligned}
(1-P_{ii})^{-2} E\left[\varepsilon_i^2 (Mu)_i^4\right] &= (1-P_{ii})^{-2} E\left[(Mu)_i^4 \left((P\varepsilon)_i^2 + (M\varepsilon)_i^2\right)\right] \\
&= (1-P_{ii})^{-2} \cdot 3 \operatorname{Var}\left((Mu)_i\right)^2 \cdot \operatorname{Var}\left((P\varepsilon)_i\right) \\
&\quad + (1-P_{ii})^{-2} \left(\frac{\sigma_{u\varepsilon}^2}{\sigma_u^4} 15 \operatorname{Var}\left((Mu)_i\right)^3 + 3 \operatorname{Var}\left((Mu)_i\right)^2 \operatorname{Var}\left((Mv)_i\right)\right) \\
&= 3P_{ii}\sigma_\varepsilon^2\sigma_u^4 + 15(1-P_{ii})\sigma_{u\varepsilon}^2\sigma_u^2 + 3(1-P_{ii})\left(\sigma_\varepsilon^2\sigma_u^4 - \sigma_{u\varepsilon}^2\sigma_u^2\right) \\
&= 3\sigma_\varepsilon^2\sigma_u^4 + 12(1-P_{ii})\sigma_{u\varepsilon}^2\sigma_u^2,
\end{aligned}$$

we have

$$\frac{10}{n^2}\sum_i (1-P_{ii})^{-2} f_i^2 E\left[\varepsilon_i^2 (Mu)_i^4\right] \le \left(3\sigma_\varepsilon^2\sigma_u^4 + 12\sigma_{u\varepsilon}^2\sigma_u^2\right)\frac{10}{n^2}\sum_i f_i^2 = O\left(\frac{1}{n}\right)$$

We now compute $\frac{10}{n^2}\sum_i (1-P_{ii})^{-2} E\left[u_i^2\varepsilon_i^2 (Mu)_i^4\right]$. Because

$$\begin{aligned}
E\left[u_i^2\varepsilon_i^2 (Mu)_i^4\right] &= E\left[(Mu)_i^4 \left((Pu)_i^2 + (Mu)_i^2\right)\left((P\varepsilon)_i^2 + (M\varepsilon)_i^2\right)\right] \\
&= E\left[(Mu)_i^4\right] E\left[(Pu)_i^2 (P\varepsilon)_i^2\right] + E\left[(Mu)_i^6\right] E\left[(P\varepsilon)_i^2\right] \\
&\quad + E\left[(Mu)_i^4 (M\varepsilon)_i^2\right] E\left[(Pu)_i^2\right] + E\left[(Mu)_i^6 (M\varepsilon)_i^2\right] \\
&= 3(1-P_{ii})^2 P_{ii}^2\sigma_u^4\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right) \\
&\quad + 15(1-P_{ii})^3 P_{ii}\sigma_u^6\sigma_\varepsilon^2 \\
&\quad + (1-P_{ii})^3 P_{ii}\left(12\sigma_{u\varepsilon}^2\sigma_u^2 + 3\sigma_\varepsilon^2\sigma_u^4\right)\sigma_u^2 \\
&\quad + (1-P_{ii})^4\left(15\sigma_\varepsilon^2\sigma_u^6 + 90\sigma_{u\varepsilon}^2\sigma_u^4\right),
\end{aligned}$$

it easily follows that

$$\frac{10}{n^2}\sum_i (1-P_{ii})^{-2} E\left[u_i^2\varepsilon_i^2 (Mu)_i^4\right] = O\left(\frac{1}{n}\right).$$

We now compute $\frac{10}{n^2}\sum_i (1-P_{ii})^{-2} f_i^4 E\left[(Mu)_i^2 (M\varepsilon)_i^2\right]$. Because

$$E\left[(Mu)_i^2 (M\varepsilon)_i^2\right] = (1-P_{ii})^2\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right),$$

it easily follows that

$$\frac{10}{n^2}\sum_i (1-P_{ii})^{-2} f_i^4 E\left[(Mu)_i^2 (M\varepsilon)_i^2\right] = O\left(\frac{1}{n}\right).$$

We now compute $\frac{10}{n^2}\sum_i 4(1-P_{ii})^{-2} f_i^2 E\left[u_i^2 (Mu)_i^2 (M\varepsilon)_i^2\right]$. Because

$$\begin{aligned}
E\left[u_i^2 (Mu)_i^2 (M\varepsilon)_i^2\right] &= E\left[\left((Mu)_i^2 + (Pu)_i^2\right)(Mu)_i^2 (M\varepsilon)_i^2\right] \\
&= (1-P_{ii})^3\left(12\sigma_{u\varepsilon}^2\sigma_u^2 + 3\sigma_\varepsilon^2\sigma_u^4\right) + P_{ii}(1-P_{ii})^2\sigma_u^2\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right),
\end{aligned}$$

it easily follows that

$$
\begin{aligned}
&\frac{10}{n^2}\sum_i 4\left(1-P_{ii}\right)^{-2} f_i^2 E\left[u_i^2\left(Mu\right)_i^2\left(M\varepsilon\right)_i^2\right] \\
&= \left(12\sigma_{u\varepsilon}^2\sigma_u^2+3\sigma_\varepsilon^2\sigma_u^4\right)\frac{40}{n^2}\sum_i f_i^2\left(1-P_{ii}\right)+\sigma_u^2\left(\sigma_u^2\sigma_\varepsilon^2+2\sigma_{u\varepsilon}^2\right)\frac{40}{n^2}\sum_i f_i^2 P_{ii}\left(1-P_{ii}\right)^2 \\
&= O\left(\frac{1}{n}\right).
\end{aligned}
$$

We finally compute $\frac{10}{n^2}\sum_i\left(1-P_{ii}\right)^{-2}E\left[u_i^4\left(Mu\right)_i^2\left(M\varepsilon\right)_i^2\right]$. Because

$$
\begin{aligned}
&E\left[u_i^4\left(Mu\right)_i^2\left(M\varepsilon\right)_i^2\right] \\
&= E\left[\left(\left(Mu\right)_i^4+2\left(Mu\right)_i^2\left(Pu\right)_i^2+\left(Pu\right)_i^4\right)\left(Mu\right)_i^2\left(M\varepsilon\right)_i^2\right] \\
&= E\left[\left(Mu\right)_i^6\left(M\varepsilon\right)_i^2\right]+2E\left[\left(Pu\right)_i^2\right]E\left[\left(Mu\right)_i^4\left(M\varepsilon\right)_i^2\right] \\
&\quad +E\left[\left(Pu\right)_i^4\right]E\left[\left(Mu\right)_i^2\left(M\varepsilon\right)_i^2\right] \\
&= \left(1-P_{ii}\right)^4\left(15\sigma_\varepsilon^2\sigma_u^6+90\sigma_{u\varepsilon}^2\sigma_u^4\right)+2P_{ii}\left(1-P_{ii}\right)^3\sigma_u^2\left(12\sigma_{u\varepsilon}^2\sigma_u^2+3\sigma_\varepsilon^2\sigma_u^4\right) \\
&\quad +3P_{ii}^2\left(1-P_{ii}\right)^2\sigma_u^4\left(\sigma_u^2\sigma_\varepsilon^2+2\sigma_{u\varepsilon}^2\right),
\end{aligned}
$$

it easily follows that

$$\frac{10}{n^2}\sum_i\left(1-P_{ii}\right)^{-2}E\left[u_i^4\left(Mu\right)_i^2\left(M\varepsilon\right)_i^2\right]=O\left(\frac{1}{n}\right).$$

To summarize, we have

$$E\left[T_{12}^2\right]=O\left(\frac{1}{n}\right). \tag{33}$$

B.2.2 $E\left[T_1T_{12}\right]$

We now compute $E\left[T_1T_{12}\right]$. We compute the expectation of the product of each term on the right side of (31) with $f'\varepsilon$.

$$E\left[\left(f'\varepsilon\right)\left(f_i^3\varepsilon_i\right)\right] = f_i^4\sigma_\varepsilon^2 \tag{34}$$

$$E\left[\left(f'\varepsilon\right)\left(3f_i^2u_i\varepsilon_i\right)\right] = 0 \tag{35}$$

$$E\left[\left(f'\varepsilon\right)\left(3f_iu_i^2\varepsilon_i\right)\right] = 3f_i^2\left(\sigma_u^2\sigma_\varepsilon^2+2\sigma_{u\varepsilon}^2\right) \tag{36}$$

$$E\left[\left(f'\varepsilon\right)\left(u_i^3\varepsilon_i\right)\right] = 0 \tag{37}$$

Now note that

$$
\begin{aligned}
E\left[Mu\left(f'u\right)\right] &= \sigma_u^2Mf=0, \qquad E\left[Mu\left(f'\varepsilon\right)\right]=\sigma_{u\varepsilon}Mf=0, \\
E\left[M\varepsilon\left(f'u\right)\right] &= \sigma_{u\varepsilon}Mf=0, \qquad E\left[M\varepsilon\left(f'\varepsilon\right)\right]=\sigma_\varepsilon^2Mf=0,
\end{aligned}
$$

which implies independence. Therefore, we have

$$E\left[\left(f'\varepsilon\right)\cdot\left(1-P_{ii}\right)^{-2}\left(Mu\right)_i^3\left(M\varepsilon\right)_i\right]=0 \tag{38}$$

Lemma 3 *Suppose that A, B, C are zero mean normal random variables. Also suppose that A and B are independent of each other. Then $E\left[A^2BC\right]=\operatorname{Cov}\left(B,C\right)\operatorname{Var}\left(A\right)$.*

Proof. Write

$$C=\frac{\operatorname{Cov}\left(A,C\right)}{\operatorname{Var}\left(A\right)}A+\frac{\operatorname{Cov}\left(B,C\right)}{\operatorname{Var}\left(B\right)}B+v$$

where v is independent of A and B. Conclusion easily follows. ■

Using Lemma 3, we obtain

$$\begin{aligned}E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}f_i\varepsilon_i\left(Mu\right)_i^2\right)\right] &= -\left(1-P_{ii}\right)^{-1}\operatorname{Cov}\left(f'\varepsilon,f_i\varepsilon_i\right)\operatorname{Var}\left(\left(Mu\right)_i\right)\\ &= -\left(1-P_{ii}\right)^{-1}f_i^2\sigma_\varepsilon^2\left(1-P_{ii}\right)\sigma_u^2\\ &= -f_i^2\sigma_\varepsilon^2\sigma_u^2 \end{aligned} \tag{39}$$

Symmetry implies

$$E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}u_i\varepsilon_i\left(Mu\right)_i^2\right)\right]=0 \tag{40}$$

and

$$E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}f_i^2\left(Mu\right)_i\left(M\varepsilon\right)_i\right)\right]=0 \tag{41}$$

Lemma 4 *Suppose that A, B, C, D are zero mean normal random variables. Also suppose that (A,B) and C are independent of each other. Then $E\left[ABCD\right]=\operatorname{Cov}\left(A,B\right)\operatorname{Cov}\left(C,D\right)$*

Proof. Write $D=\xi_1A+\xi_2B+\xi_3C+v$, where ξs denote regression coefficients. Note that $\xi_3=\operatorname{Cov}\left(C,D\right)/\operatorname{Var}\left(C\right)$ by independence. We then have

$$ABCD=\xi_1A^2BC+\xi_2AB^2C+\xi_3ABC^2+ABCv$$

from which the conclusion follows. ■

Using Lemma 4, we obtain

$$\begin{aligned}&E\left[\left(f'\varepsilon\right)\cdot\left(-2\left(1-P_{ii}\right)^{-1}f_iu_i\left(Mu\right)_i\left(M\varepsilon\right)_i\right)\right]\\ &= -2\left(1-P_{ii}\right)^{-1}\operatorname{Cov}\left(\left(Mu\right)_i,\left(M\varepsilon\right)_i\right)\operatorname{Cov}\left(f'\varepsilon,f_iu_i\right)\\ &= -2\left(1-P_{ii}\right)^{-1}\left(1-P_{ii}\right)\sigma_{u\varepsilon}f_i^2\sigma_{u\varepsilon}\\ &= -2\sigma_{u\varepsilon}^2f_i^2\end{aligned} \tag{42}$$

Finally, using symmetry again, we obtain

$$E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}u_i^2\left(Mu\right)_i\left(M\varepsilon\right)_i\right)\right]=0 \tag{43}$$

Combining (34) - (43), we obtain

$$E\left[\left(f'\varepsilon\right)\cdot\left(\delta_{1i}\delta_{2i}\right)\right]=f_i^4\sigma_\varepsilon^2+2f_i^2\left(\sigma_u^2\sigma_\varepsilon^2+2\sigma_{u\varepsilon}^2\right),$$

from which we obtain

$$E\left[T_1T_{12}\right]=\frac{1}{H}\frac{1}{n^2}\sum_i E\left[\left(f'\varepsilon\right)\cdot\left(\delta_{1i}\delta_{2i}\right)\right]=\frac{1}{n}\frac{\sigma_\varepsilon^2}{H}\left(\frac{1}{n}\sum_i f_i^4\right)+2\frac{\sigma_u^2\sigma_\varepsilon^2+2\sigma_{u\varepsilon}^2}{n}. \tag{44}$$

B.2.3 $\frac{1}{n^2}\sum_{i=1}^n\delta_{2i}^2$

We compute $E\left[\frac{1}{n^2}\sum_{i=1}^n\delta_{2i}^2\right]$ and characterize its order of magnitude. From (32), we can obtain

$$E\left[\delta_{2i}^2\right]=f_i^4+4f_i^2\sigma_u^2+4P_{ii}\sigma_u^4,$$

and hence, it follows that

$$E\left[\frac{1}{n^2}\sum_{i=1}^n\delta_{2i}^2\right]=\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^n f_i^4\right)+\frac{4H\sigma_u^2}{n}+o\left(\frac{K}{n}\right).$$

B.2.4 $E\left[T_1T_{14}\right]$

We compute the expectation of the product of each term on the right hand side of (32) with $\left(f'\varepsilon\right)^2$, noting independence between $\left(Mu\right)_i$ and $f'\varepsilon$. We have

$$E\left[\left(f'\varepsilon\right)^2\cdot f_i^4\right]=f'f\sigma_\varepsilon^2f_i^4=nH\sigma_\varepsilon^2f_i^4,$$

$$\begin{aligned}E\left[\left(f'\varepsilon\right)^2\cdot 6f_i^2u_i^2\right]&=6f_i^2\left(\left(f'f\sigma_\varepsilon^2\right)\sigma_u^2+2\left(f_i\sigma_{u\varepsilon}\right)^2\right)\\&=6nH\sigma_\varepsilon^2\sigma_u^2f_i^2+12\sigma_{u\varepsilon}^2f_i^4,\end{aligned}$$

$$\begin{aligned}E\left[\left(f'\varepsilon\right)^2\cdot u_i^4\right]&=12f_i^2\sigma_{u\varepsilon}^2\sigma_u^2+3f'f\sigma_\varepsilon^2\sigma_u^4\\&=3nH\sigma_\varepsilon^2\sigma_u^4+12f_i^2\sigma_{u\varepsilon}^2\sigma_u^2,\end{aligned}$$

$$E\left[\left(f'\varepsilon\right)^2\cdot\left(1-P_{ii}\right)^{-2}\left(Mu\right)_i^4\right]=\left(f'f\sigma_\varepsilon^2\right)\cdot 3\sigma_u^4=3nH\sigma_\varepsilon^2\sigma_u^2,$$

$$E\left[\left(f'\varepsilon\right)^2\cdot\left(4f_i^3u_i\right)\right]=0,$$

$$E\left[\left(f'\varepsilon\right)^2\cdot\left(-2f_i^2\left(1-P_{ii}\right)^{-1}\left(Mu\right)_i^2\right)\right]=-2f_i^2f'f\sigma_\varepsilon^2\sigma_u^2=-2nHf_i^2\sigma_\varepsilon^2\sigma_u^2,$$

$$E\left[\left(f'\varepsilon\right)^2\cdot\left(4f_iu_i^3\right)\right]=0,$$

$$E\left[\left(f'\varepsilon\right)^2\cdot\left(-4f_iu_i\left(1-P_{ii}\right)^{-1}\left(Mu\right)_i^2\right)\right]=0,$$

$$E\left[\left(f'\varepsilon\right)^2\cdot\left(-2\left(1-P_{ii}\right)^{-1}u_i^2\left(Mu\right)_i^2\right)\right]=-4f_i^2\sigma_{u\varepsilon}^2\sigma_u^2-4\left(1-P_{ii}\right)nH\sigma_\varepsilon^2\sigma_u^4-2nH\sigma_\varepsilon^2\sigma_u^4.$$

Therefore,

$$\begin{aligned}E\left[\left(f'\varepsilon\right)^2\sum_{i=1}^n\delta_{2i}^2\right] &= n^2H\sigma_\varepsilon^2\left(\frac{1}{n}\sum_{i=1}^n f_i^4\right)+6n^2H^2\sigma_\varepsilon^2\sigma_u^2+12n\sigma_{u\varepsilon}^2\left(\frac{1}{n}\sum_{i=1}^n f_i^4\right)\\ &\quad+3n^2H\sigma_\varepsilon^2\sigma_u^2+12nH\sigma_{u\varepsilon}^2\sigma_u^2+3n^2H\sigma_\varepsilon^2\sigma_u^2-2n^2H^2\sigma_\varepsilon^2\sigma_u^2\\ &\quad-4nH\sigma_{u\varepsilon}^2\sigma_u^2-4\left(n-K\right)nH\sigma_\varepsilon^2\sigma_u^4-2n^2H\sigma_\varepsilon^2\sigma_u^4,\end{aligned}$$

and therefore, we have

$$\begin{aligned}E\left[T_1T_{14}\right] &= \frac{1}{H^2n^3}E\left[\left(f'\varepsilon\right)^2\sum_{i=1}^n\delta_{2i}^2\right]\\ &= \frac{1}{n}\frac{1}{H}\sigma_\varepsilon^2\left(\frac{1}{n}\sum_{i=1}^n f_i^4\right)+\frac{1}{n}\left(\frac{6}{H}\sigma_\varepsilon^2\sigma_u^2+4\sigma_\varepsilon^2\sigma_u^2-\frac{6}{H}\sigma_\varepsilon^2\sigma_u^4\right)+o\left(\frac{1}{n}\right). \qquad (45)\end{aligned}$$

C Higher Order Bias and MSE of Fuller

The results in this section was derived using Donald and Newey's (2001) adaptation of Rothenberg (1983). Under the normalization where $\sigma_\varepsilon^2=\sigma_u^2=1$ and $\sigma_{\varepsilon u}=\rho$, we have the following result. For F(1), the higher order mean bias and the higher order MSE are equal to

$$0$$

and

$$\frac{1-R^2}{nR^2}+\rho^2\frac{-K+2}{n^2}\left(\frac{1-R^2}{R^2}\right)^2+\frac{K}{n^2}\left(\frac{1-R^2}{R^2}\right)^2$$

Here, R^2 denotes the theoretical R^2 as defined in (5). For F(4), they are equal to

$$\rho\frac{3}{n}\frac{1-R^2}{R^2}$$

and

$$\frac{1-R^2}{nR^2} + \rho^2 \frac{-1-K}{n^2} \left(\frac{1-R^2}{R^2}\right)^2 + \frac{K-6}{n^2} \left(\frac{1-R^2}{R^2}\right)^2$$

For Fuller(Optimal), they are equal to

$$\rho \frac{2+\frac{1}{\rho^2}}{n} \frac{1-R^2}{R^2}$$

and

$$\frac{1-R^2}{nR^2} + \frac{1}{n^2} \left(\left(1-\rho^2\right) K - \frac{1}{\rho^2} - 2\rho^2 - 4\right) \left(\frac{1-R^2}{R^2}\right)^2$$

References

[1] Akahira, M. (1983), "Asymptotic Deficiency of the Jackknife Estimator", *Australian Journal of Statistics* 25, 123 - 129.

[2] Anderson, T.W., K. Morimune and T. Sawa (1983), "The Numerical Values of Some Key Parameters in Econometric Models," *Journal of Econometrics* 21, 229-243.

[3] Anderson, T.W., N. Kunitomo and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator", *Econometrica* 50, 1009-1028.

[4] Anderson, T.W. and T. Sawa (1973), "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions", *Econometrica* 41, 683-714.

[5] Anderson, T.W. and T. Sawa (1979), "Evaluation of the Distribution Function of the Two Stage Least Squares Estimator", *Econometrica* 47, 163 - 182.

[6] Angrist, J.D., G.W. Imbens, and A. Krueger (1995), "Jackknife Instrumental Variables Estimation", NBER Technical Working Paper No. 172.

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

[8] Bekker, P.A. (1994), "Alternative Approximations to the Distributions of Instrumental Variable Estimators", *Econometrica* 92, 657 – 681.

[9] Davidson, R., and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

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

[11] Donald, S.G., and W.K. Newey (2000), "A Jackknife Interpretation of the Continuous Updating Estimator", *Economics Letters* 67, 239-243.

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

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

[14] HAHN, J., AND J. HAUSMAN (2001), "Notes on Bias in Estimators for Simultaneous Equation Models", forthcoming in *Economics Letters.*

[15] HAUSMAN, J. (1983), "Specification and Estimation of Simultaneous Equation Models", in GRILICHES, Z., AND M.D. INTRILIGATOR *eds Handbook of Econometrics*, Vol. I. Amsterdam: Elsevier.

[16] HOLLY, A., AND P.C.B. PHILLIPS (1979), "A Saddlepoint Approximation to the Distribution of the k-Class Estimator of a Coefficient in a Simultaneous System", *Econometrica* 47, 1527-1548.

[17] KUERSTEINER, G.M., 2000, "Mean Squared Error Reduction for GMM Estimators of Linear Time Series Models" *unpulished manuscript.*

[18] MADDALA, G.S., AND J. JEONG (1992), "On the Exact Small Sample Distribution of the Instrumental Variables Estimator", *Econometrica* 60, 181 - 183.

[19] MARIANO, R. S (1972), "The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators", *Econometrica* 40, 643 - 652.

[20] MARIANO, R.S. AND T. SAWA (1972), "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables", *Journal of the American Statistical Association* 67, 159 – 163.

[21] MORIMUNE, K. (1983), "Approximate Distributions of the k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size", *Econometrica* 51, 821 - 841.

[22] 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.

[23] NEWEY, W.K., AND R.J. SMITH (2001), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *mimeo.*

[24] OBERHELMAN, T.W., AND K.R. KADIYALA (1983), "Asymptotic Probability Concentrations and Finite Sample Properties of Modified LIML Estimators for Equations with more than Two Endogenous Variables," *Journal of Econometrics* 98, 163-185.

[25] PHILLIPS, G.D.A., AND C. HALE (1977), "The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems", *International Economic Review* 18, 219-228.

[26] PHILLIPS, P.C.B. (1980), "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing n+1 Endogenous Variables", *Econometrica* 48, 861-878.

[27] Rothenberg, T.J. (1983), "Asymptotic Properties of Some Estimators In Structural Models", in *Studies in Econometrics, Time Series, and Multivariate Statistics*.

[28] Rothenberg, T.J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics", in *Handbook of Econometrics* 2, Amsterdam: North-Holland.

[29] Sawa, T. (1972), "Finite-Sample Properties of the k-Class Estimators", *Econometrica* 40, 653-680.

[30] Shao, J., and D. Tu (1995), *The Jackknife and Bootstrap*. New York: Springer-Verlag.

[31] Staiger, D., and J. Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica* 65, 557 - 586.

Table 1: Homoscedastic Design, ρ=.5

			Median Bias									RMSE									InterQuartile Range								
n	K	R^2	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS
100	5	0.01	0.29	0.36	0.43	0.45	0.37	0.41	0.36	0.44	0.49	238.41	0.68	0.52	0.50	52.40	0.63	2.08	60.74	0.50	1.41	0.78	0.38	0.26	1.05	0.51	0.84	1.20	0.11
100	10	0.01	0.35	0.39	0.44	0.46	0.44	0.45	0.42	0.47	0.50	37.53	0.79	0.56	0.52	33.93	0.54	0.79	234.05	0.50	1.50	0.96	0.50	0.35	1.16	0.38	0.67	1.28	0.12
100	30	0.01	0.38	0.40	0.43	0.45	0.47	0.48	0.46	0.47	0.49	60.60	0.97	0.65	0.58	31.46	0.51	0.58	98.83	0.50	1.61	1.18	0.70	0.51	1.15	0.21	0.43	1.28	0.11
500	5	0.01	0.06	0.13	0.25	0.31	0.14	0.23	0.15	0.13	0.50	35031.25	0.47	0.37	0.38	28.85	0.40	0.78	12.51	0.50	0.78	0.56	0.34	0.26	0.65	0.40	0.56	0.92	0.05
500	10	0.01	0.06	0.13	0.25	0.31	0.18	0.32	0.23	0.19	0.49	11.76	0.59	0.42	0.40	9.62	0.41	0.55	49.87	0.50	0.91	0.67	0.41	0.31	0.83	0.32	0.52	1.08	0.05
500	30	0.01	0.17	0.21	0.29	0.33	0.33	0.43	0.37	0.38	0.50	125.65	0.78	0.53	0.48	68.54	0.45	0.47	11.63	0.50	1.09	0.86	0.58	0.45	1.00	0.20	0.37	1.13	0.05
1000	5	0.01	0.00	0.05	0.16	0.22	0.06	0.15	0.07	-0.02	0.49	1.55	0.36	0.28	0.29	18.30	0.31	0.49	9.07	0.50	0.50	0.42	0.30	0.24	0.46	0.33	0.42	0.61	0.04
1000	10	0.01	0.02	0.07	0.16	0.22	0.08	0.24	0.14	0.03	0.50	33.44	0.42	0.32	0.31	71.57	0.32	0.39	20.02	0.50	0.56	0.47	0.34	0.27	0.55	0.27	0.40	0.71	0.04
1000	30	0.01	0.04	0.08	0.17	0.23	0.14	0.37	0.28	0.13	0.50	35.56	0.61	0.41	0.38	25.91	0.40	0.38	16.23	0.50	0.73	0.62	0.45	0.37	0.76	0.19	0.33	0.92	0.04
100	5	0.1	0.00	0.05	0.14	0.21	0.05	0.13	0.06	-0.02	0.45	2.16	0.35	0.27	0.28	2.11	0.29	0.57	14.36	0.46	0.47	0.40	0.29	0.24	0.44	0.32	0.41	0.58	0.11
100	10	0.1	0.02	0.06	0.16	0.22	0.08	0.23	0.13	0.03	0.45	12.00	0.42	0.31	0.30	8.37	0.31	0.39	44.83	0.46	0.54	0.46	0.33	0.27	0.52	0.28	0.41	0.67	0.11
100	30	0.1	0.06	0.10	0.18	0.23	0.17	0.36	0.27	0.16	0.45	13.43	0.64	0.43	0.39	79.57	0.39	0.38	13.09	0.45	0.74	0.64	0.47	0.37	0.71	0.19	0.34	0.90	0.11
500	5	0.1	0.00	0.01	0.04	0.06	0.01	0.04	0.01	-0.01	0.45	0.14	0.14	0.13	0.13	0.14	0.13	0.14	0.15	0.45	0.19	0.18	0.17	0.16	0.19	0.17	0.19	0.20	0.05
500	10	0.1	0.00	0.00	0.03	0.05	0.01	0.07	0.02	-0.01	0.45	0.15	0.15	0.14	0.13	0.15	0.14	0.15	0.17	0.45	0.20	0.19	0.18	0.16	0.19	0.16	0.19	0.20	0.05
500	30	0.1	0.00	0.01	0.04	0.06	0.02	0.17	0.07	0.00	0.45	0.17	0.17	0.15	0.15	0.19	0.20	0.16	0.21	0.45	0.22	0.21	0.20	0.18	0.22	0.13	0.19	0.24	0.05
1000	5	0.1	0.00	0.00	0.02	0.03	0.00	0.02	0.00	-0.01	0.45	0.10	0.10	0.09	0.09	0.10	0.09	0.10	0.10	0.45	0.13	0.13	0.12	0.12	0.13	0.12	0.13	0.13	0.04
1000	10	0.1	0.00	0.01	0.02	0.03	0.01	0.04	0.01	0.00	0.45	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.10	0.45	0.13	0.13	0.12	0.12	0.13	0.12	0.13	0.13	0.04
1000	30	0.1	0.00	0.00	0.02	0.03	0.00	0.10	0.02	-0.01	0.45	0.11	0.10	0.10	0.10	0.11	0.13	0.10	0.12	0.45	0.14	0.14	0.13	0.13	0.15	0.11	0.14	0.15	0.04
100	5	0.3	0.00	0.01	0.04	0.07	0.01	0.04	0.01	-0.01	0.35	0.17	0.16	0.15	0.15	0.17	0.15	0.17	0.19	0.36	0.21	0.20	0.18	0.17	0.21	0.19	0.21	0.23	0.10
100	10	0.3	0.01	0.02	0.05	0.08	0.02	0.09	0.03	-0.01	0.35	0.19	0.18	0.16	0.16	0.19	0.16	0.18	0.23	0.36	0.23	0.22	0.20	0.18	0.23	0.18	0.22	0.25	0.10
100	30	0.3	0.00	0.02	0.05	0.07	0.03	0.20	0.09	-0.01	0.35	0.53	0.25	0.20	0.18	2.73	0.23	0.19	0.87	0.36	0.26	0.25	0.22	0.20	0.29	0.15	0.22	0.32	0.10
500	5	0.3	0.00	0.00	0.01	0.02	0.00	0.01	0.00	0.00	0.35	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.35	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.05
500	10	0.3	0.00	0.00	0.01	0.01	0.00	0.02	0.00	0.00	0.35	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.35	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.05
500	30	0.3	0.00	0.00	0.01	0.02	0.01	0.06	0.01	0.00	0.35	0.07	0.07	0.07	0.07	0.07	0.09	0.07	0.08	0.35	0.10	0.10	0.10	0.09	0.10	0.08	0.10	0.10	0.05
1000	5	0.3	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.35	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.35	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.03
1000	10	0.3	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.35	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.35	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.03
1000	30	0.3	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.35	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.35	0.07	0.07	0.06	0.06	0.07	0.06	0.07	0.07	0.03

Table 2: Homoscedastic Design, ρ=.9

			Median Bias									RMSE									InterQuartile Range								
n	K	R^2	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS
100	5	0.01	0.43	0.62	0.76	0.77	0.68	0.73	0.65	0.80	0.89	34.03	0.78	0.80	0.81	49.23	0.79	1.61	37.91	0.89	1.08	0.49	0.25	0.24	0.72	0.33	0.53	0.92	0.06
100	10	0.01	0.54	0.66	0.77	0.77	0.77	0.81	0.75	0.84	0.89	42.69	0.83	0.82	0.83	42.72	0.83	0.85	295.13	0.89	1.08	0.58	0.29	0.28	0.68	0.22	0.39	0.79	0.06
100	30	0.01	0.69	0.74	0.80	0.80	0.84	0.87	0.84	0.88	0.89	383.17	0.93	0.86	0.86	46.63	0.87	0.86	22.03	0.89	1.07	0.72	0.40	0.39	0.64	0.11	0.23	0.67	0.06
500	5	0.01	0.03	0.19	0.44	0.46	0.23	0.42	0.27	0.15	0.89	12.44	0.36	0.48	0.49	40.22	0.48	1.03	76.60	0.89	0.69	0.35	0.17	0.17	0.57	0.28	0.44	0.90	0.03
500	10	0.01	0.03	0.20	0.45	0.46	0.33	0.59	0.43	0.32	0.89	24.58	0.43	0.51	0.52	58.06	0.61	0.54	117.56	0.89	0.74	0.39	0.19	0.19	0.71	0.21	0.37	1.06	0.03
500	30	0.01	0.11	0.26	0.47	0.48	0.53	0.77	0.66	0.56	0.89	25.67	0.55	0.57	0.57	29.99	0.77	0.68	83.35	0.89	0.83	0.45	0.24	0.24	0.79	0.12	0.23	1.10	0.03
1000	5	0.01	0.00	0.08	0.27	0.29	0.10	0.26	0.13	-0.03	0.89	2.56	0.27	0.30	0.31	17.96	0.33	0.56	8.38	0.89	0.45	0.34	0.17	0.17	0.47	0.26	0.40	0.69	0.02
1000	10	0.01	0.01	0.09	0.28	0.29	0.13	0.43	0.25	0.02	0.89	1.86	0.28	0.31	0.32	16.18	0.45	0.38	15.07	0.89	0.47	0.34	0.18	0.17	0.56	0.20	0.34	0.79	0.02
1000	30	0.01	0.00	0.09	0.29	0.30	0.24	0.67	0.51	0.21	0.89	65.18	0.35	0.35	0.35	78.14	0.68	0.53	12.14	0.89	0.55	0.40	0.21	0.20	0.72	0.12	0.22	0.92	0.02
100	5	0.1	0.00	0.08	0.26	0.27	0.09	0.25	0.12	-0.04	0.81	2.88	0.26	0.29	0.30	45.16	0.31	0.41	23.53	0.81	0.43	0.33	0.18	0.17	0.42	0.26	0.37	0.64	0.07
100	10	0.1	0.01	0.09	0.26	0.28	0.13	0.42	0.23	0.04	0.81	8.79	0.29	0.30	0.31	85.99	0.44	0.37	19.51	0.81	0.46	0.35	0.18	0.18	0.52	0.20	0.34	0.72	0.07
100	30	0.1	0.02	0.10	0.28	0.29	0.27	0.65	0.49	0.25	0.81	6.35	0.39	0.35	0.36	24.15	0.66	0.51	215.21	0.81	0.52	0.39	0.22	0.21	0.67	0.12	0.23	0.88	0.07
500	5	0.1	0.00	0.02	0.06	0.07	0.02	0.06	0.02	-0.01	0.81	0.14	0.14	0.13	0.13	0.14	0.13	0.14	0.17	0.81	0.19	0.18	0.15	0.15	0.19	0.16	0.18	0.20	0.03
500	10	0.1	0.00	0.01	0.06	0.06	0.02	0.13	0.03	-0.02	0.81	0.15	0.14	0.13	0.13	0.16	0.16	0.15	0.19	0.81	0.18	0.18	0.15	0.15	0.20	0.14	0.19	0.22	0.03
500	30	0.1	0.00	0.02	0.06	0.07	0.03	0.31	0.12	-0.01	0.81	0.15	0.14	0.13	0.13	0.25	0.32	0.17	0.59	0.81	0.19	0.18	0.16	0.16	0.24	0.10	0.17	0.26	0.03
1000	5	0.1	0.00	0.01	0.03	0.03	0.01	0.03	0.01	-0.01	0.81	0.10	0.10	0.09	0.09	0.10	0.09	0.10	0.11	0.81	0.13	0.12	0.11	0.11	0.13	0.12	0.13	0.13	0.02
1000	10	0.1	0.00	0.01	0.03	0.03	0.01	0.07	0.01	-0.01	0.81	0.10	0.10	0.09	0.09	0.10	0.11	0.10	0.11	0.81	0.13	0.12	0.12	0.12	0.13	0.11	0.13	0.14	0.02
1000	30	0.1	0.00	0.01	0.03	0.03	0.01	0.19	0.04	-0.01	0.81	0.10	0.10	0.09	0.09	0.12	0.20	0.11	0.13	0.81	0.13	0.13	0.12	0.12	0.15	0.09	0.14	0.16	0.02
100	5	0.3	0.00	0.02	0.08	0.08	0.02	0.08	0.02	-0.02	0.63	0.17	0.16	0.14	0.14	0.18	0.15	0.17	0.22	0.63	0.20	0.19	0.16	0.16	0.21	0.18	0.21	0.24	0.08
100	10	0.3	0.00	0.02	0.08	0.09	0.03	0.16	0.05	-0.01	0.63	0.18	0.16	0.15	0.15	0.21	0.19	0.18	4.82	0.63	0.21	0.20	0.17	0.17	0.24	0.16	0.22	0.27	0.08
100	30	0.3	0.00	0.02	0.08	0.08	0.04	0.37	0.16	-0.01	0.63	0.19	0.17	0.15	0.15	3.48	0.37	0.21	1.74	0.63	0.22	0.21	0.17	0.17	0.32	0.11	0.20	0.37	0.08
500	5	0.3	0.00	0.00	0.02	0.02	0.00	0.02	0.00	0.00	0.63	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.63	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.03
500	10	0.3	0.00	0.00	0.02	0.02	0.00	0.04	0.01	0.00	0.63	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.63	0.09	0.09	0.09	0.09	0.09	0.08	0.09	0.10	0.03
500	30	0.3	0.00	0.01	0.02	0.02	0.01	0.11	0.02	0.00	0.63	0.07	0.07	0.07	0.07	0.08	0.12	0.07	0.08	0.63	0.09	0.09	0.09	0.09	0.10	0.07	0.09	0.10	0.03
1000	5	0.3	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.63	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.63	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.02
1000	10	0.3	0.00	0.00	0.01	0.01	0.00	0.02	0.00	0.00	0.63	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.63	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.02
1000	30	0.3	0.00	0.00	0.01	0.01	0.00	0.06	0.00	0.00	0.63	0.05	0.05	0.05	0.05	0.05	0.07	0.05	0.05	0.63	0.06	0.06	0.06	0.06	0.07	0.06	0.07	0.07	0.02

Table 3: Heteroscedastic Design, ρ=.5

			Median Bias									RMSE									InterQuartile Range								
n	K	R^2	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS
100	5	0.01	0.44	0.46	0.46	0.46	0.46	0.46	0.41	0.49	0.47	1049.14	0.85	0.58	0.53	46.73	0.72	2.25	61.53	0.48	2.23	1.05	0.48	0.32	1.21	0.61	0.99	1.42	0.12
100	10	0.01	0.50	0.49	0.48	0.49	0.50	0.48	0.45	0.50	0.48	112.18	0.92	0.62	0.55	33.92	0.58	0.86	110.03	0.49	2.07	1.18	0.59	0.40	1.25	0.42	0.74	1.40	0.12
100	30	0.01	0.47	0.47	0.48	0.48	0.49	0.49	0.47	0.48	0.49	62.98	1.05	0.69	0.60	24.16	0.51	0.60	100.32	0.50	1.97	1.36	0.77	0.55	1.16	0.22	0.44	1.31	0.12
500	5	0.01	0.05	0.13	0.26	0.32	0.21	0.27	0.18	0.15	0.47	32.96	0.64	0.44	0.41	65.55	0.48	1.00	15.33	0.47	1.14	0.77	0.46	0.34	0.77	0.48	0.67	1.10	0.05
500	10	0.01	0.05	0.14	0.26	0.32	0.25	0.35	0.25	0.22	0.48	173.77	0.72	0.48	0.44	10.33	0.45	0.61	55.87	0.48	1.30	0.88	0.51	0.38	0.87	0.35	0.57	1.20	0.05
500	30	0.01	0.19	0.23	0.30	0.35	0.37	0.44	0.38	0.38	0.49	173.73	0.88	0.58	0.51	82.41	0.47	0.49	12.52	0.49	1.36	1.04	0.65	0.49	0.99	0.21	0.38	1.18	0.05
1000	5	0.01	-0.03	0.03	0.14	0.21	0.09	0.17	0.08	-0.02	0.47	53.80	0.50	0.34	0.33	18.00	0.37	0.58	10.35	0.47	0.72	0.58	0.40	0.32	0.55	0.40	0.51	0.74	0.04
1000	10	0.01	-0.01	0.04	0.15	0.22	0.11	0.26	0.15	0.03	0.48	11.55	0.53	0.36	0.34	72.84	0.35	0.42	22.52	0.48	0.72	0.59	0.40	0.32	0.59	0.30	0.44	0.78	0.04
1000	30	0.01	0.02	0.07	0.17	0.23	0.18	0.38	0.29	0.14	0.49	20.87	0.69	0.45	0.41	23.61	0.41	0.39	16.79	0.49	0.88	0.72	0.51	0.41	0.77	0.20	0.34	0.94	0.04
100	5	0.1	-0.03	0.03	0.13	0.20	0.08	0.16	0.08	-0.02	0.43	61.57	0.47	0.32	0.31	2.48	0.34	0.81	13.35	0.44	0.63	0.52	0.37	0.30	0.50	0.38	0.48	0.66	0.12
100	10	0.1	0.00	0.05	0.15	0.21	0.11	0.25	0.14	0.03	0.44	82.90	0.51	0.35	0.33	8.23	0.34	0.42	53.30	0.45	0.69	0.57	0.40	0.32	0.55	0.31	0.44	0.72	0.11
100	30	0.1	0.05	0.09	0.17	0.23	0.20	0.37	0.28	0.17	0.44	10.80	0.71	0.46	0.41	72.63	0.40	0.39	14.29	0.45	0.88	0.75	0.53	0.42	0.72	0.19	0.35	0.94	0.11
500	5	0.1	0.00	0.00	0.03	0.05	0.02	0.04	0.02	-0.01	0.43	0.18	0.17	0.16	0.15	0.16	0.15	0.16	0.18	0.43	0.23	0.23	0.21	0.20	0.22	0.20	0.22	0.23	0.05
500	10	0.1	-0.01	0.00	0.02	0.05	0.02	0.07	0.02	-0.01	0.44	0.18	0.17	0.15	0.15	0.17	0.15	0.16	0.18	0.44	0.22	0.21	0.20	0.18	0.21	0.18	0.20	0.22	0.05
500	30	0.1	0.00	0.00	0.03	0.05	0.02	0.18	0.07	0.00	0.45	0.19	0.18	0.16	0.15	0.19	0.20	0.16	0.22	0.45	0.23	0.23	0.21	0.19	0.23	0.14	0.19	0.25	0.05
1000	5	0.1	0.00	0.00	0.01	0.02	0.01	0.02	0.00	-0.01	0.43	0.12	0.12	0.11	0.11	0.12	0.11	0.12	0.12	0.43	0.16	0.16	0.15	0.15	0.15	0.15	0.15	0.16	0.04
1000	10	0.1	0.00	0.00	0.02	0.03	0.01	0.04	0.01	0.00	0.44	0.11	0.11	0.10	0.10	0.11	0.11	0.11	0.11	0.44	0.14	0.14	0.14	0.13	0.14	0.13	0.14	0.15	0.04
1000	30	0.1	-0.01	0.00	0.01	0.03	0.01	0.10	0.02	-0.01	0.45	0.11	0.11	0.11	0.10	0.11	0.13	0.11	0.12	0.45	0.15	0.14	0.14	0.13	0.15	0.11	0.14	0.16	0.04
100	5	0.3	-0.01	0.00	0.03	0.06	0.02	0.05	0.01	-0.01	0.33	0.21	0.20	0.18	0.17	0.19	0.18	0.20	0.22	0.34	0.26	0.25	0.22	0.21	0.24	0.22	0.25	0.26	0.11
100	10	0.3	0.00	0.01	0.04	0.07	0.03	0.10	0.03	-0.01	0.34	0.22	0.20	0.18	0.17	0.20	0.18	0.19	0.25	0.35	0.26	0.25	0.23	0.21	0.25	0.20	0.24	0.27	0.11
100	30	0.3	-0.01	0.00	0.04	0.07	0.04	0.21	0.09	0.00	0.35	0.56	0.27	0.21	0.19	2.89	0.24	0.20	0.91	0.35	0.29	0.28	0.25	0.22	0.29	0.15	0.23	0.34	0.10
500	5	0.3	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.33	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.33	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.05
500	10	0.3	0.00	0.00	0.01	0.01	0.00	0.02	0.00	0.00	0.34	0.08	0.08	0.08	0.07	0.08	0.08	0.08	0.08	0.34	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.05
500	30	0.3	0.00	0.00	0.01	0.02	0.01	0.06	0.01	0.00	0.35	0.08	0.07	0.07	0.07	0.08	0.09	0.07	0.08	0.35	0.10	0.10	0.10	0.10	0.10	0.08	0.10	0.10	0.05
1000	5	0.3	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.33	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.33	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.03
1000	10	0.3	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.34	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.34	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.03
1000	30	0.3	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.35	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.35	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.07	0.03

Table 4: Heteroscedastic Design, ρ=.9

			Median Bias									RMSE									InterQuartile Range								
n	K	R^2	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	OLS
100	5	0.01	0.78	0.81	0.83	0.83	0.84	0.83	0.74	0.90	0.85	106.05	0.98	0.88	0.87	34.48	0.91	1.65	105.15	0.85	2.02	0.81	0.35	0.33	0.83	0.41	0.68	1.11	0.09
100	10	0.01	0.85	0.86	0.86	0.86	0.87	0.86	0.81	0.90	0.87	276.61	1.01	0.90	0.90	84.60	0.89	0.93	253.52	0.87	1.72	0.85	0.38	0.37	0.70	0.25	0.44	0.89	0.07
100	30	0.01	0.88	0.88	0.88	0.88	0.88	0.88	0.86	0.90	0.88	64.83	1.04	0.93	0.92	44.76	0.89	0.88	22.63	0.88	1.39	0.85	0.46	0.44	0.63	0.12	0.24	0.71	0.06
500	5	0.01	-0.05	0.18	0.45	0.46	0.34	0.47	0.31	0.17	0.85	372.81	0.51	0.53	0.54	38.42	0.56	1.28	92.98	0.85	1.07	0.51	0.26	0.26	0.63	0.35	0.55	1.07	0.04
500	10	0.01	-0.05	0.19	0.47	0.48	0.44	0.64	0.46	0.35	0.87	22.11	0.57	0.57	0.58	54.64	0.66	0.59	132.65	0.87	1.15	0.51	0.27	0.26	0.70	0.24	0.43	1.18	0.03
500	30	0.01	0.06	0.25	0.49	0.50	0.61	0.79	0.68	0.57	0.88	33.45	0.68	0.63	0.64	28.43	0.79	0.70	87.75	0.88	1.12	0.56	0.33	0.32	0.73	0.13	0.24	1.13	0.03
1000	5	0.01	-0.07	0.03	0.25	0.26	0.17	0.31	0.16	-0.03	0.85	3.17	0.37	0.32	0.33	18.54	0.39	0.63	10.00	0.85	0.65	0.46	0.24	0.24	0.52	0.33	0.47	0.81	0.03
1000	10	0.01	-0.07	0.04	0.26	0.27	0.19	0.46	0.26	0.02	0.87	3.96	0.35	0.32	0.33	15.38	0.49	0.41	16.42	0.87	0.62	0.43	0.22	0.21	0.57	0.22	0.38	0.87	0.02
1000	30	0.01	-0.08	0.04	0.28	0.29	0.31	0.69	0.52	0.21	0.88	53.85	0.42	0.37	0.38	73.77	0.69	0.54	12.30	0.88	0.71	0.47	0.24	0.23	0.69	0.13	0.23	0.95	0.02
100	5	0.1	-0.07	0.03	0.23	0.24	0.15	0.28	0.13	-0.04	0.77	14.36	0.35	0.30	0.31	47.44	0.36	0.49	26.74	0.77	0.59	0.44	0.25	0.24	0.47	0.31	0.44	0.75	0.09
100	10	0.1	-0.05	0.04	0.25	0.26	0.20	0.44	0.25	0.04	0.79	42.14	0.35	0.31	0.31	35.77	0.46	0.41	20.39	0.79	0.60	0.44	0.23	0.22	0.53	0.22	0.38	0.78	0.08
100	30	0.1	-0.05	0.05	0.26	0.27	0.33	0.66	0.50	0.26	0.80	577.96	0.45	0.37	0.38	21.37	0.67	0.53	237.42	0.80	0.66	0.46	0.24	0.23	0.66	0.13	0.24	0.92	0.07
500	5	0.1	-0.01	0.00	0.05	0.05	0.03	0.07	0.02	-0.02	0.77	0.18	0.17	0.15	0.15	0.17	0.16	0.17	0.19	0.77	0.23	0.22	0.19	0.19	0.21	0.19	0.21	0.24	0.04
500	10	0.1	-0.02	0.00	0.05	0.05	0.03	0.14	0.03	-0.01	0.79	0.17	0.16	0.14	0.14	0.17	0.17	0.16	0.20	0.79	0.21	0.20	0.17	0.17	0.21	0.16	0.20	0.24	0.03
500	30	0.1	-0.01	0.00	0.05	0.05	0.04	0.32	0.12	-0.01	0.80	0.17	0.15	0.13	0.13	0.26	0.33	0.18	0.62	0.80	0.21	0.20	0.17	0.17	0.24	0.10	0.18	0.27	0.03
1000	5	0.1	-0.01	0.00	0.02	0.03	0.01	0.04	0.01	-0.01	0.77	0.12	0.12	0.11	0.11	0.12	0.11	0.12	0.12	0.77	0.16	0.15	0.14	0.14	0.16	0.15	0.16	0.16	0.03
1000	10	0.1	-0.01	0.00	0.03	0.03	0.02	0.07	0.01	-0.01	0.79	0.11	0.11	0.10	0.10	0.11	0.11	0.11	0.12	0.79	0.14	0.14	0.13	0.13	0.14	0.12	0.14	0.15	0.02
1000	30	0.1	-0.01	0.00	0.02	0.03	0.01	0.19	0.04	-0.01	0.80	0.11	0.10	0.10	0.10	0.12	0.20	0.11	0.13	0.80	0.14	0.13	0.12	0.12	0.16	0.09	0.14	0.17	0.02
100	5	0.3	-0.02	0.00	0.06	0.07	0.04	0.09	0.03	-0.02	0.60	0.21	0.19	0.16	0.16	0.20	0.18	0.20	0.25	0.60	0.26	0.24	0.21	0.20	0.24	0.21	0.24	0.28	0.08
100	10	0.3	-0.01	0.01	0.07	0.07	0.05	0.17	0.05	-0.02	0.61	0.21	0.19	0.16	0.16	0.22	0.21	0.20	5.31	0.62	0.25	0.23	0.20	0.19	0.25	0.17	0.24	0.29	0.08
100	30	0.3	-0.02	0.01	0.06	0.07	0.06	0.37	0.16	-0.01	0.62	0.22	0.19	0.15	0.15	3.15	0.38	0.22	1.78	0.63	0.24	0.23	0.19	0.19	0.32	0.12	0.20	0.39	0.08
500	5	0.3	0.00	0.00	0.01	0.01	0.01	0.02	0.01	0.00	0.60	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.60	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.04
500	10	0.3	0.00	0.00	0.01	0.01	0.01	0.04	0.01	0.00	0.61	0.08	0.08	0.07	0.07	0.08	0.08	0.08	0.08	0.61	0.10	0.10	0.10	0.10	0.10	0.09	0.10	0.10	0.03
500	30	0.3	0.00	0.00	0.01	0.01	0.01	0.11	0.02	0.00	0.62	0.07	0.07	0.07	0.07	0.08	0.12	0.08	0.08	0.63	0.10	0.10	0.09	0.09	0.10	0.08	0.10	0.11	0.03
1000	5	0.3	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.60	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.60	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.03
1000	10	0.3	0.00	0.00	0.01	0.01	0.00	0.02	0.00	0.00	0.61	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.61	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.02
1000	30	0.3	0.00	0.00	0.01	0.01	0.00	0.06	0.00	0.00	0.63	0.05	0.05	0.05	0.05	0.05	0.07	0.05	0.05	0.63	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.07	0.02

Table 5: Actual 95%-5% Range Divided by Asymptotic Range for Homoscedastic Design

n	K	R2	rho	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE	n	K	R2	rho	LIML	F(1)	F(4)	F(opt)	Nagar	2SLS	JN2SLS	JIVE
100	5	0.01	0.5	2.87	0.57	0.28	0.20	1.85	0.44	0.91	2.36	100	5	0.01	0.9	2.31	0.39	0.18	0.18	1.37	0.29	0.62	1.72
100	10	0.01	0.5	2.77	0.67	0.33	0.24	2.00	0.29	0.62	2.46	100	10	0.01	0.9	2.29	0.42	0.21	0.20	1.27	0.17	0.37	1.58
100	30	0.01	0.5	2.88	0.89	0.45	0.33	2.15	0.16	0.36	2.35	100	30	0.01	0.9	2.23	0.51	0.26	0.25	1.28	0.09	0.19	1.34
500	5	0.01	0.5	2.34	1.03	0.59	0.43	1.86	0.74	1.20	3.67	500	5	0.01	0.9	1.96	0.58	0.33	0.32	2.04	0.55	1.03	5.71
500	10	0.01	0.5	3.24	1.30	0.71	0.53	3.07	0.59	1.07	4.95	500	10	0.01	0.9	2.24	0.69	0.40	0.39	3.47	0.39	0.75	5.94
500	30	0.01	0.5	4.48	1.78	0.95	0.71	4.31	0.35	0.67	5.07	500	30	0.01	0.9	3.82	1.12	0.57	0.55	4.16	0.21	0.41	4.68
1000	5	0.01	0.5	1.54	1.11	0.74	0.59	1.47	0.87	1.20	2.53	1000	5	0.01	0.9	1.39	0.82	0.40	0.39	1.62	0.69	1.13	3.15
1000	10	0.01	0.5	1.93	1.30	0.84	0.66	1.97	0.69	1.11	3.59	1000	10	0.01	0.9	1.53	0.85	0.41	0.40	2.28	0.49	0.93	5.44
1000	30	0.01	0.5	3.64	1.97	1.18	0.91	4.08	0.46	0.83	5.79	1000	30	0.01	0.9	1.95	0.97	0.51	0.49	6.25	0.30	0.58	8.09
100	5	0.1	0.5	1.60	1.14	0.75	0.61	1.40	0.86	1.17	2.33	100	5	0.1	0.9	1.40	0.86	0.45	0.44	1.52	0.70	1.13	3.04
100	10	0.1	0.5	2.01	1.36	0.89	0.70	1.99	0.73	1.16	3.58	100	10	0.1	0.9	1.56	0.91	0.47	0.45	2.49	0.53	1.00	5.92
100	30	0.1	0.5	4.15	2.19	1.29	0.97	4.56	0.49	0.91	5.80	100	30	0.1	0.9	2.14	1.05	0.57	0.56	5.43	0.32	0.63	8.08
500	5	0.1	0.5	1.04	1.00	0.93	0.86	1.02	0.94	1.02	1.09	500	5	0.1	0.9	1.03	0.98	0.84	0.83	1.04	0.90	1.03	1.16
500	10	0.1	0.5	1.12	1.08	1.00	0.92	1.14	0.92	1.09	1.22	500	10	0.1	0.9	1.08	1.03	0.88	0.86	1.17	0.81	1.08	1.30
500	30	0.1	0.5	1.24	1.20	1.10	1.01	1.36	0.74	1.07	1.46	500	30	0.1	0.9	1.08	1.03	0.88	0.87	1.50	0.58	1.00	1.68
1000	5	0.1	0.5	1.03	1.02	0.97	0.94	1.03	0.98	1.02	1.06	1000	5	0.1	0.9	1.03	1.00	0.93	0.92	1.03	0.95	1.02	1.08
1000	10	0.1	0.5	1.05	1.03	0.99	0.95	1.06	0.95	1.05	1.10	1000	10	0.1	0.9	1.03	1.00	0.94	0.93	1.09	0.90	1.07	1.15
1000	30	0.1	0.5	1.11	1.09	1.04	1.01	1.17	0.83	1.07	1.21	1000	30	0.1	0.9	1.03	1.00	0.93	0.93	1.23	0.70	1.05	1.29
100	5	0.3	0.5	1.10	1.06	0.95	0.85	1.09	0.97	1.09	1.19	100	5	0.3	0.9	1.10	1.02	0.83	0.82	1.12	0.91	1.09	1.30
100	10	0.3	0.5	1.19	1.14	1.02	0.92	1.19	0.92	1.13	1.32	100	10	0.3	0.9	1.13	1.04	0.85	0.84	1.24	0.78	1.12	1.46
100	30	0.3	0.5	1.57	1.49	1.25	1.11	1.69	0.73	1.13	1.95	100	30	0.3	0.9	1.21	1.11	0.89	0.88	1.93	0.57	1.01	2.34
500	5	0.3	0.5	0.99	0.99	0.96	0.95	0.99	0.97	0.99	1.00	500	5	0.3	0.9	1.00	0.99	0.95	0.94	0.99	0.96	0.99	1.02
500	10	0.3	0.5	1.04	1.03	1.01	0.98	1.04	0.98	1.04	1.06	500	10	0.3	0.9	1.03	1.02	0.98	0.97	1.05	0.95	1.04	1.08
500	30	0.3	0.5	1.05	1.04	1.02	0.99	1.08	0.90	1.05	1.10	500	30	0.3	0.9	1.02	1.00	0.96	0.96	1.12	0.83	1.07	1.15
1000	5	0.3	0.5	1.01	1.01	1.00	0.98	1.00	0.99	1.01	1.01	1000	5	0.3	0.9	1.01	1.00	0.98	0.98	1.00	0.99	1.00	1.02
1000	10	0.3	0.5	1.01	1.01	1.00	0.99	1.01	0.98	1.01	1.02	1000	10	0.3	0.9	1.01	1.01	0.99	0.99	1.03	0.97	1.02	1.04
1000	30	0.3	0.5	1.01	1.01	1.00	0.99	1.02	0.94	1.01	1.03	1000	30	0.3	0.9	0.99	0.99	0.97	0.97	1.04	0.89	1.03	1.05

Table 6: Definitions of k-class estimators

k-class estimator	$\frac{x'Py - \kappa \cdot x'My}{x'Px - \kappa \cdot x'Mx}$
LIML	$\kappa = \phi$
F(1)	$\kappa = \phi - \frac{1}{n-K}$
F(4)	$\kappa = \phi - \frac{4}{n-K}$
F(opt)	$\kappa = \phi - \frac{3+1/\rho^2}{n-K}$
Nagar	$\kappa = \frac{K-2}{n} \Big/ \left(1 - \frac{K-2}{n}\right)$
2SLS	$\kappa = 0$

ϕ is equal to the smallest eigenvalue of the matrix $W'PW\,(W'MW)^{-1}$, where $W \equiv [y, x]$.

Table 7: Concentration Parameters

R^2	n	δ^2
0.01	100	1.01
0.1	100	11.11
0.3	100	42.86
0.01	500	5.05
0.1	500	55.56
0.3	500	214.29
0.01	1000	10.10
0.1	1000	111.11
0.3	1000	428.57

Table 8: Regression on Asymptotic MSE Formulae

Estimator	Constant	Intercept	Regression R^2
JN2SLS	0.153	0.015	0.101
	(0.061)	(0.0061)	
F(4)	0.056	0.065	0.244
	(0.019)	(0.016)	
2SLS	0.091	0.00091	0.281
	(0.020)	(0.0002)	

Table 9: Regression on Asymptotic MSE Formulae

Estimator	C1	C2	Regression R^2
JN2SLS	1.74	-.030	.665
	(.168)	(.006)	
F(4)	.807	-.173	.957
	(.023)	(.009)	
2SLS	.384	.0003	.407
	(.064)	(.0002)	

Table 10: Regression of Log MSE

RHS Variables	JN2SLS	F(4)	2SLS
$\log n$	-.955	-1.01	-.999
	(.037)	(.025)	(.036)
$\log \rho^2$	.570	.568	.961
	(.108)	(.074)	(.106)
$\log K$	-.060	.098	.389
	(.078)	(.053)	(.077)
$\log Ratio$	1.10	.930	.880
	(.041)	(.028)	(.040)
Regression R^2	.950	.969	.939

Table 11: Median Bias Approximation for LIML Based on Weak Instrument Limit Distributions

		Median Bias					
		n=100		n=500		n=1000	
K	R^2	$\rho = .5$	$\rho = .9$	$\rho = .5$	$\rho = .9$	$\rho = .5$	$\rho = .9$
5	.01	0.323	0.538	0.061	0.057	0.008	-0.004
10	.01	0.398	0.667	0.094	0.090	0.018	-0.002
30	.01	0.496	0.919	0.226	0.261	0.067	0.018
5	.1	0.011	-0.012	-0.001	-0.001	0.001	-0.001
10	.1	0.003	-0.047	-0.003	-0.002	-0.002	0.001
30	.1	0.320	0.967	-0.013	-0.030	-0.002	-0.008
5	.3	-0.004	-0.009	0.001	0.000	0.001	0.001
10	.3	-0.017	-0.025	-0.002	0.001	0.000	-0.001
30	.3	-0.169	-0.326	-0.005	-0.009	-0.002	-0.001