Higher Order MSE of Jackknife 2SLS

Jinyong Hahn	Jerry Hausman	Guido Kuersteiner
Brown University	MIT	MIT

April, 2001

Abstract

In this paper we consider parameter estimation in a simple linear simultaneous equations model. It is well known that two stage least squares (2SLS) estimators 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. Monte Carlo simulations show that even in relatively large samples the MSE of LIML and Nagar can be substantially larger than for Jackknife 2SLS.

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

JEL C13,C21,C31,C51

1 Introduction

There has been a 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 (2000).

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 (1998) 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. One might conjecture that such results may well generalize to situations other than the simultaneous equations models. In other words, one may conjecture that bias reduced version of an estimator would in general have a better risk property than the original estimator. Donald and Newey (1999) and Newey and Smith (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 contribute to this approach by considering 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 calculation of the Jackknife estimator has in general 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 such type. 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. Our proof strategy can in principle be generalized to non-IV estimators. Second, the Jackknife 2SLS may prove to be a reasonable competitor to the LIML or Nagar's estimator. It is well-known that the LIML and Nagar 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). LIML and Nagar's estimator have better higher order risk properties than 2SLS, as shown by Rothenberg (1979) or Donald and Newey (1998). The moment problem would not pose any practical concern if the problem were concentrated in the extreme end of the tails. Unfortunately, in Hahn and Hausman (2000) we found in Monte Carlo that LIML and Nagar's estimator tend to have bigger spread (measured in terms of interquartile range, etc) than 2SLS for some parameter combinations. It seems possible that the moment problem, which in principle should be a mere object of theoretical curiosity, presents itself in the form of undesirable finite sample risk properties despite the prediction based on higher order expansions. On the other hand, it can be shown that Jackknife 2SLS is known to have moments up to the degree of overidentification. If Jackknife 2SLS has a higher order MSE comparable to LIML or Nagar's estimator, we can then conjecture that its actual finite sample properties may be more stable.

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

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

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 $f_i \equiv E\left[x_i | z_i\right] = z_i'\pi$, 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 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 (1998) modification of Nagar's (1959) estimator.

Donald and Newey (1998) computed higher order mean squared errors (MSE) of the k-class estimators. They showed that n times the MSE of 2SLS, LIML, and B2SLS are approximately equal to

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

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 $\left(\frac{K\sigma_{u\varepsilon}}{\sqrt{n}H}\right)^2=\frac{K^2}{n}\frac{\sigma_{u\varepsilon}^2}{H^2}$ in the higher order MSE calculation. 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 would suggest that LIML and B2SLS should be preferred to 2SLS. Unfortunately, it is well-known that the LIML and Nagar 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 found in Monte Carlo that LIML and B2SLS tend to have bigger spread (measured in terms of interquartile range, etc) than 2SLS for some parameter combinations. 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

$$\frac{n-1}{n}\sum_i\left(\frac{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_jy_j}{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_jx_j}-\frac{\widehat{\pi}'\sum_i z_iy_i}{\widehat{\pi}'\sum_i z_ix_i}\right)=\frac{n-1}{n}\sum_i\left(\frac{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_j\varepsilon_j}{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_jx_j}-\frac{x'P\varepsilon}{x'Px}\right)\tag{1}$$

and the corresponding jackknife estimator is given by

$$\begin{aligned} b_J &= \frac{\widehat{\pi}'\sum_i z_iy_i}{\widehat{\pi}'\sum_i z_ix_i}-\frac{n-1}{n}\sum_i\left(\frac{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_jy_j}{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_jx_j}-\frac{\widehat{\pi}'\sum_i z_iy_i}{\widehat{\pi}'\sum_i z_ix_i}\right)\\ &= n\frac{\widehat{\pi}'\sum_i z_iy_i}{\widehat{\pi}'\sum_i z_ix_i}-\frac{n-1}{n}\sum_i\frac{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_jy_j}{\widehat{\pi}_{(i)}'\sum_{j\neq i}z_jx_j}\end{aligned}$$

Here, $\widehat{\pi}$ denotes the OLS estimator of the first stage coefficient π, and $\widehat{\pi}_{(i)}$ denotes such 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 would have 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 (1998), we let $h \equiv \frac{f'\varepsilon}{n}$. We impose following assumptions. First, we assume normality[1]:

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)* $\frac{1}{n} \sum_i f_i^6 = O(1)$.[2]

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

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

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

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

[1] 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.

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

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 Lemma 2 in 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 can ignore it from our further computation.

We now examine the resultant bias corrected estimator (1) ignoring R_n:

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

Theorem 1 below is obtained by squaring and taking expectation of the RHS of (2):

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 (1998). Therefore, the Jackknife does not increase variance too much. Although it has long been known that Jackknife does reduce the bias, the literature has been hesitant in recommending its use primarily because of the concern that variance may increase too much due to 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. In Hahn and Hausman (2000), we demonstrated that LIML is approximately equivalent to the optimal linear combination of two Nagar's 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.

3 Monte Carlo

We generated

$$y_i = x_i\beta + \varepsilon_i, \quad x_i = z_i'\pi + u_i \qquad i = 1, \ldots, n$$

such that $z_i \sim N(0, I_K)$, $\mathrm{Var}(\varepsilon_i) = \mathrm{Var}(u_i) = 1$, and $\beta = 0$. We let $\pi = (\overline{\pi}, \ldots, \overline{\pi})$, such that

$$\mathsf{R}^2 \equiv \frac{\pi' E[z_i z_i']\pi}{\pi' E[z_i z_i']\pi + \mathrm{Var}(u_i)} = \frac{K\overline{\pi}^2}{K\overline{\pi}^2 + 1}$$

Here, R^2 denotes the theoretical R^2 in the first stage regression. We considered combinations of the following parameters:

$$\begin{aligned} n &= 100, \quad 500, \quad 1000 \\ K &= 5, \quad 10, \quad 30 \\ \mathsf{R}^2 &= .001, \quad .1, \quad .3 \\ \rho &= \mathrm{Cov}(\varepsilon_i, u_i) = 0, \quad .5, \quad .9 \end{aligned}$$

Results based 5000 Monte Carlo runs are summarized in Tables 1 - 3.

We first discuss the sample size 100 case in Table 1. In the upper panel the "moment problem" appears for both LIML and the Nagar estimator with both the mean and RMSE considerably larger than for J2SLS. However, for low $\mathsf{R}^2 = .001$, which corresponds to the weak instrument setup, J2SLS does considerably better than either LIML or Nagar. The interquartile range for J2SLS is about $\frac{1}{2}$ as large as for the other estimators. As the R^2 increases, the superiority of J2SLS is not as great. However, it is typically better than the other estimators for the interquartile range. When LIML does better than J2SLS, it is only by a very small amount. In the middle and lower panels of Table 1 as the number of instruments increases which exacerbates the weak instrument problem, the superiority of J2SLS increases with respect to the interquartile range. Now for the low R^2 situation, its interquartile range is approximately $\frac{1}{4}$ as large as LIML or the Nagar estimator. However, the most interesting finding may be that the "classical" 2SLS estimator typically does the best of any of the second order unbiased estimators in terms of the interquartile range. Thus, while LIML and the Nagar estimator demonstrate their expected superiority in terms of lower median bias, the finite sample performance of 2SLS in terms of the interquartile range is striking.

In Table 2 we increase the sample size to 500 while the other parameters remain the same. In terms of the interquartile range we again find that J2SLS is often superior to LIML and the Nagar estimator. In no situation does LIML have a significant superiority to J2SLS although it is slightly better in a few cases. Once again, classical 2SLS does better than the other 3 estimators in terms of interquartile range, especially when R^2 is very low. Thus, in the weak instrument situations of low R^2 and high K, regular 2SLS has much to recommend it. Lastly, in Table 3 we increase the sample size to 1000, again keeping the other parameters constant. Now, only in the low R^2 case does J2SLS do better than LIML or the Nagar estimator. In the other situations, LIML does as well as J2SLS or slightly better. However, LIML never demonstrates a marked superiority in terms of the interquartile range. Once again, regular 2SLS does best in terms of the interquartile range.[3][4]

Summing up, even for sample sizes of 1000 the superior performance of LIML with respect to median unbiasedness is counteracted by the "moment" problem. The moment problem often leads to high RMSE and a large interquartile range, especially when R^2 is low, the number of instruments is high, or the correlation between the two equations stochastic disturbances is large. All of these situations are characteristic of the weak instrument situation as discussed by Hahn and Hausman (2000). Thus, we suggest caution in using either LIML or the Nagar estimator in the weak instrument situation. J2SLS or regular 2SLS may offer better properties depending on the (implicit) finite sample risk function in use. We also recommend the use of the Hahn-Hausman (2000) specification test as a means of ascertaining the degree of reliance appropriate for the large sample approximations being used.

[3] In order to confirm that our result is not specific to the $\beta = 0$ case, we considered the case where $\beta = -1$, $n = 100$, $K = 30$, $R^2 = .001$, $\rho = .5$. We found that the mean biases of LIML, Nagar, 2SLS, and J2SLS are equal to 0.5544, 2.3186, 1.3002, 1.2965. The median biases were equal to 1.2729, 1.3111, 1.3005, 1.2948, and the RMSE were equal to 42.3779 60.4197 1.3103 1.3450. Finally, interquartile ranges were equal to 1.7120, 1.1807, 0.2144, 0.4365. We repeated the Monte Carlo for the case where $\beta = -1.8$, $n = 100$, $K = 30$, $R^2 = .001$, $\rho = .9$. The mean biases were equal to 8.0161, 1.2714, 0.8995, 0.8976, and the median biases were equal to 0.8920, 0.9024, 0.8984, 0.8978. RMSE were equal to 514.8346, 27.0408, 0.9033, 0.9162, and the interquartile ranges were equal to 0.8811, 0.6009, 0.1080, 0.2190. These numbers suggest that our results quite generally hold.

[4] We also examined properties of Angrist, Imbens, and Krueger's (1995) JIVE by a small scale Monte Carlo experiment. For the case where $\beta = 0$, $n = 100$, $K = 30$, $R^2 = .001$, $\rho = .9$, similar to Table 1, we found that the JIVE has a mean bias equal to 0.6183, median bias equal to 0.8970, RMSE equal to 17.9269, and interquartile range equal to 0.6346. For the case where $\beta = 0$, $n = 100$, $K = 30$, $R^2 = .1$, $\rho = .5$, we found that the JIVE has a mean bias equal to 0.2938, median bias equal to 0.1639, RMSE equal to 13.0949, and interquartile range equal to 0.8962. These numbers suggest that the JIVE may well have a "moment" problem similar to LIML and the Nagar estimator.

Appendix

A Higher Order Expansion

We first present two Lemmas:

Lemma 1 *Let* υ_i *be a smample of* n *independent random variables with* $\max_i E[|\upsilon_i|^r] < c^r < \infty$ *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

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

The conclusion follows by Markov inequality. ∎

Lemma 2 *Assume that Conditions 2 and 3 are satisfied. Further assume that* $E[|\varepsilon_i|^r] < \infty$ *and* $E[|u_i|^r] < \infty$ *for* r *sufficiently large* $(r > 3)$*. 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

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

We have $(\max|f_i| + \max|u_i|) \cdot \max|\varepsilon_i| = O_p\left(n^{2/r}\right)$ by Lemma 1. Because $\max|(Pu)_i|^2 \leq \max P_{ii} \cdot u'u$, and $\max P_{ii} = O\left(\frac{1}{n}\right)$, we also have $\max|(Pu)_i| = O_p(1)$. Similarly, $\max|(P\varepsilon)_i| = O_p(1)$. 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 (1998) that we have the following expansion[5]:

$$H\sqrt{n}\frac{x'P\varepsilon}{x'Px} = \sum_{j=1}^{7} T_j + o_p\left(\frac{K}{n}\right), \tag{3}$$

[5] 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

$$\frac{H}{\frac{1}{n}x'Px} = \frac{H}{H + W_3 + W_4} = 1 - \frac{1}{H}W_3 - \frac{1}{H}W_4 + \frac{1}{H^2}W_3^2 + o_p\left(\frac{K}{n}\right), \tag{4}$$

$$\left(\frac{H}{\frac{1}{n}x'Px}\right)^2 = 1 - 2\frac{1}{H}W_3 - 2\frac{1}{H}W_4 + 3\frac{1}{H^2}W_3^2 + o_p\left(\frac{K}{n}\right) \tag{5}$$

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

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_i \delta_{1i} &= -\frac{1}{\sqrt{n}}\sum_i x_i\varepsilon_i + \frac{1}{\sqrt{n}}\sum_i (1 - P_{ii})^{-1}(Mx)_i (M\varepsilon)_i \\ &= -h - \frac{1}{\sqrt{n}}u'\varepsilon + \frac{1}{\sqrt{n}}(Mu)'\left(I - \widetilde{P}\right)^{-1}(M\varepsilon) \\ &= -h - \frac{1}{\sqrt{n}}u'\varepsilon + \frac{1}{\sqrt{n}}u'M\varepsilon + \frac{1}{\sqrt{n}}(Mu)'\overline{P}(M\varepsilon) \\ &= -h - \frac{1}{\sqrt{n}}u'P\varepsilon + \frac{1}{\sqrt{n}}u'\overline{P}\varepsilon - \frac{1}{\sqrt{n}}u'P\overline{P}\varepsilon - \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon + \frac{1}{\sqrt{n}}u'P\overline{P}P\varepsilon \\ &= -h - \frac{1}{\sqrt{n}}u'C'\varepsilon - \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon + \frac{1}{\sqrt{n}}u'P\overline{P}P\varepsilon, \end{aligned} \tag{6}$$

where, as in Donald and Newey (1998), 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

$$\left|\frac{u'\overline{P}P\varepsilon}{\sqrt{n}}\right| \leq \frac{1}{\sqrt{n}}\sqrt{u'u}\sqrt{\varepsilon'P\overline{P}^2P\varepsilon} = \frac{1}{\sqrt{n}}\sqrt{O_p(n)}\sqrt{O\left(\frac{1}{n^2}\right)O_p(K)} = O_p\left(\frac{\sqrt{K}}{n}\right),$$

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

To conclude, we can write

$$\frac{1}{\sqrt{n}}\sum_i \delta_{1i} = -h + W_5 + W_6 + o_p\left(\frac{K}{n}\right), \tag{7}$$

where

$$W_5 \equiv -\frac{1}{\sqrt{n}}u'C'\varepsilon = O_p\left(\frac{\sqrt{K}}{\sqrt{n}}\right)$$

$$W_6 \equiv \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon = O_p\left(\frac{\sqrt{K}}{n}\right).$$

We now expand $\left(\frac{H}{\frac{1}{n}x'Px}\right)\left(\frac{1}{\sqrt{n}}\sum_i \delta_{1i}\right)$ using (4) and (7):

$$\begin{aligned}
&\left(\frac{H}{\frac{1}{n}x'Px}\right)\left(\frac{1}{\sqrt{n}}\sum_i \delta_{1i}\right) \\
&= \left(1-\frac{1}{H}W_3-\frac{1}{H}W_4+\frac{1}{H^2}W_3^2\right)(-h+W_5+W_6)+o_p\left(\frac{K}{n}\right) \\
&= -h+W_3\frac{1}{H}h+W_4\frac{1}{H}h-W_3^2\frac{1}{H^2}h+W_5+W_6-\frac{1}{H}W_3W_5+o_p\left(\frac{K}{n}\right) \\
&= -T_1-T_3-T_5-T_7+T_8+T_9+T_{10}+o_p\left(\frac{K}{n}\right)
\end{aligned} \tag{8}$$

where

$$T_8 \equiv W_5 = -\frac{1}{\sqrt{n}}u'C'\varepsilon = O_p\left(\frac{\sqrt{K}}{\sqrt{n}}\right),$$

$$T_9 \equiv W_6 = \frac{1}{\sqrt{n}}u'\overline{P}P\varepsilon = O_p\left(\frac{\sqrt{K}}{n}\right),$$

$$T_{10} \equiv -\frac{1}{H}W_3W_5 = W_3\frac{1}{H}\frac{1}{\sqrt{n}}u'C'\varepsilon = O_p\left(\frac{\sqrt{K}}{n}\right).$$

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 (6), 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

$$\left|u'P\overline{P}Pu\right| \le \max\left(\frac{P_{ii}}{1-P_{ii}}\right)\cdot u'Pu = O_p\left(\frac{K}{n}\right),$$

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

we may write

$$\frac{1}{n}\sum_i \delta_{2i} = -H - W_3 - W_7 + o_p\left(\frac{K}{n}\right) \tag{9}$$

where

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

Combining (4) and (9), 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 (3), yields

$$\begin{aligned}
\frac{H}{\frac{1}{n}x'Px}\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n}\sum_i \delta_{2i}\right) &= \frac{1}{H}\left(\sum_{j=1}^{7} T_j\right)(-H + W_4 - W_7) + o_p\left(\frac{K}{n}\right) \\
&= -\sum_{j=1}^{7} T_j + W_4\frac{1}{H}h - W_7\frac{1}{H}h + o_p\left(\frac{K}{n}\right) \\
&= -\sum_{j=1}^{7} T_j - T_5 + T_{11} + o_p\left(\frac{K}{n}\right) \qquad (10)
\end{aligned}$$

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

$$\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{1}{n\sqrt{n}}\sum_i \delta_{1i}\delta_{2i}\right) = T_{12} + o_p\left(\frac{K}{n}\right) \tag{11}$$

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

$$\frac{1}{H}\left(\frac{H}{\frac{1}{n}x'Px}\right)^2\left(\frac{\frac{1}{\sqrt{n}}x'P\varepsilon}{\frac{1}{n}x'Px}\right)\left(\frac{1}{n^2}\sum_i \delta_{2i}^2\right) = T_{14} + o_p\left(\frac{K}{n}\right) \tag{12}$$

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 (2), (3), (8), (10), (11), and (12), we obtain

$$\begin{aligned} & H\sqrt{n}\left(\frac{x'P\varepsilon}{x'Px} - \frac{n-1}{n}\sum_i\left(\frac{x'P\varepsilon+\delta_{1i}}{x'Px+\delta_{2i}} - \frac{x'P\varepsilon}{x'Px}\right) + R_n\right) \\ = \ & T_1 + T_3 + T_7 - T_8 - T_9 - T_{10} + T_{11} + T_{12} - T_{14} + o_p\left(\frac{K}{n}\right). \end{aligned} \tag{13}$$

B Approximate MSE Calculation

In computing the (approximate) mean squared error, we keep terms up to $O_p\left(\frac{1}{n}\right)$. From (13), we can see that the MSE of the jackknife estimator approximately equal to

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

Combining (14) with (15), (16), (17), (18), (19), (20), (21), (22), (23), (26), (37), and (38) in the next two subsections, it can shown that the approximate MSE up to $O_p\left(\frac{1}{n}\right)$ is given by

$$\sigma_\varepsilon^2 H + \frac{K}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + \sigma_{u\varepsilon}^2\right) + \left(-\frac{12}{H}\right)\frac{\sigma_u^2\sigma_\varepsilon^2}{n} + 20\frac{\sigma_{u\varepsilon}^2}{n} + \frac{12}{H}\frac{\sigma_u^4\sigma_\varepsilon^2}{n},$$

which proves Theorem 1.

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

From Donald and Newey (1998), we can see that

$$
\begin{aligned}
E\left[T_1^2\right] &= \sigma_\varepsilon^2 H && (15)\\
E\left[T_3^2\right] &= \frac{4}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right) + o\left(\frac{1}{n}\right) && (16)\\
E\left[T_1T_3\right] &= 0 && (17)\\
E\left[T_1T_7\right] &= \frac{4}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right) + o\left(\frac{1}{n}\right) && (18)\\
E\left[T_8^2\right] &= \frac{K}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + \sigma_{u\varepsilon}^2\right) + o\left(\frac{K}{n}\sup P_{ii}\right) && (19)
\end{aligned}
$$

Also, by symmetry, we have

$$
\begin{aligned}
E\left[T_1T_8\right] &= 0 && (20)\\
E\left[T_1T_9\right] &= 0. && (21)
\end{aligned}
$$

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 (1998), we obtain

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

Therefore, we have

$$
\begin{aligned}
E\left[T_1T_{10}\right] &= E\left[T_3T_8\right] = \frac{2}{nH}\frac{\sigma_u^2\sigma_\varepsilon^2 f'C'f + \sigma_{u\varepsilon}^2 f'Cf}{n}\\
&= \frac{2}{nH}\left(\sigma_u^2\sigma_\varepsilon^2 H + \sigma_{u\varepsilon}^2 H + o\left(\frac{1}{n}\right)\right) = \frac{2}{n}\left(\sigma_u^2\sigma_\varepsilon^2 + \sigma_{u\varepsilon}^2\right) + o\left(\frac{1}{n^2}\right), && (22)
\end{aligned}
$$

where the second equality is based on equation (20) of Donald and Newey (1998).

Now, note that

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

and

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

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

$$E[T_1T_{11}] = -2\frac{\sigma_{u\varepsilon}^2}{n}. \tag{23}$$

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

Note that

$$\begin{aligned} \delta_{1i}\delta_{2i} &= x_i^3\varepsilon_i + (1-P_{ii})^{-2}(Mu)_i^3(M\varepsilon)_i \\ &\quad - (1-P_{ii})^{-1}x_i\varepsilon_i(Mu)_i^2 - (1-P_{ii})^{-1}x_i^2(Mu)_i(M\varepsilon)_i \\ &= f_i^3\varepsilon_i + 3f_i^2u_i\varepsilon_i + 3f_iu_i^2\varepsilon_i + u_i^3\varepsilon_i \\ &\quad + (1-P_{ii})^{-2}(Mu)_i^3(M\varepsilon)_i - (1-P_{ii})^{-1}f_i\varepsilon_i(Mu)_i^2 \\ &\quad - (1-P_{ii})^{-1}u_i\varepsilon_i(Mu)_i^2 - (1-P_{ii})^{-1}f_i^2(Mu)_i(M\varepsilon)_i \\ &\quad -2(1-P_{ii})^{-1}f_iu_i(Mu)_i(M\varepsilon)_i - (1-P_{ii})^{-1}u_i^2(Mu)_i(M\varepsilon)_i \end{aligned} \tag{24}$$

and

$$\begin{aligned} \delta_{2i}^2 &= \left(-f_i^2 - 2f_iu_i - u_i^2 + (1-P_{ii})^{-1}(Mu)_i^2\right)^2 \\ &= f_i^4 + 6f_i^2u_i^2 + u_i^4 + (1-P_{ii})^{-2}(Mu)_i^4 \\ &\quad +4f_i^3u_i - 2f_i^2(1-P_{ii})^{-1}(Mu)_i^2 + 4f_iu_i^3 \\ &\quad -4f_iu_i(1-P_{ii})^{-1}(Mu)_i^2 - 2(1-P_{ii})^{-1}u_i^2(Mu)_i^2 \end{aligned} \tag{25}$$

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

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

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

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 $\frac{10}{n^3}\sum_i (1-P_{ii})^{-4} E\left[(Mu)_i^6 (M\varepsilon)_i^2\right]$. 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

$$
\frac{10}{n^3}\sum_i (1-P_{ii})^{-4} E\left[(Mu)_i^6 (M\varepsilon)_i^2\right] = \frac{10}{n^3}\sum_i \left(15\sigma_\varepsilon^2\sigma_u^6 + 90\sigma_{u\varepsilon}^2\sigma_u^4\right) = o\left(\frac{1}{n}\right).
$$

We now compute $\frac{10}{n^3}\sum_i (1-P_{ii})^{-2} f_i^2 E\left[\varepsilon_i^2 (Mu)_i^4\right]$. Because

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

we have

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

We now compute $\frac{10}{n^3}\sum_i (1-P_{ii})^{-2} E\left[u_i^2\varepsilon_i^2 (Mu)_i^4\right]$. 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

$$\frac{10}{n^3}\sum_i (1-P_{ii})^{-2} E\left[u_i^2\varepsilon_i^2 (Mu)_i^4\right] = o\left(\frac{1}{n}\right).$$

We now compute $\frac{10}{n^3}\sum_i (1-P_{ii})^{-2} f_i^4 E\left[(Mu)_i^2 (M\varepsilon)_i^2\right]$. 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

$$\frac{10}{n^3}\sum_i (1-P_{ii})^{-2} f_i^4 E\left[(Mu)_i^2 (M\varepsilon)_i^2\right] = o\left(\frac{1}{n}\right).$$

We now compute $\frac{10}{n^3}\sum_i 4(1-P_{ii})^{-2} f_i^2 E\left[u_i^2 (Mu)_i^2 (M\varepsilon)_i^2\right]$. 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

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

We finally compute $\frac{10}{n^3}\sum_i (1-P_{ii})^{-2} E\left[u_i^4 (Mu)_i^2 (M\varepsilon)_i^2\right]$. Because

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

it easily follows that

$$\frac{10}{n^3}\sum_i (1-P_{ii})^{-2} E\left[u_i^4 (Mu)_i^2 (M\varepsilon)_i^2\right] = o\left(\frac{1}{n}\right).$$

To summarize, we have

$$E\left[T_{12}^2\right] = o\left(\frac{1}{n}\right). \tag{26}$$

B.2.2 $E[T_1T_{12}]$

We now compute $E[T_1T_{12}]$. We compute the expectation of the product of each term on the right side of (24) with $f'\varepsilon$.

$$E\left[(f'\varepsilon)\left(f_i^3\varepsilon_i\right)\right] = f_i^4\sigma_\varepsilon^2 \tag{27}$$

$$E\left[(f'\varepsilon)\left(3f_i^2 u_i\varepsilon_i\right)\right] = 0 \tag{28}$$

$$E\left[(f'\varepsilon)\left(3f_i u_i^2\varepsilon_i\right)\right] = 3f_i^2\left(\sigma_u^2\sigma_\varepsilon^2 + 2\sigma_{u\varepsilon}^2\right) \tag{29}$$

$$E\left[(f'\varepsilon)\left(u_i^3\varepsilon_i\right)\right] = 0 \tag{30}$$

Now note that

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

which implies independence. Therefore, we have

$$E\left[\left(f'\varepsilon\right)\cdot\left(1-P_{ii}\right)^{-2}\left(Mu\right)_i^3\left(M\varepsilon\right)_i\right]=0 \tag{31}$$

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 $E\left[A^2BC\right]=\text{Cov}\left(B,C\right)\text{Var}\left(A\right)$.*

Proof. Write

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

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

Using Lemma 3, we obtain

$$\begin{aligned} E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}f_i\varepsilon_i\left(Mu\right)_i^2\right)\right] &= -\left(1-P_{ii}\right)^{-1}\text{Cov}\left(f'\varepsilon,f_i\varepsilon_i\right)\text{Var}\left(\left(Mu\right)_i\right) \\ &= -\left(1-P_{ii}\right)^{-1}f_i^2\sigma_\varepsilon^2\left(1-P_{ii}\right)\sigma_u^2 \\ &= -f_i^2\sigma_\varepsilon^2\sigma_u^2 \end{aligned} \tag{32}$$

Symmetry implies

$$E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}u_i\varepsilon_i\left(Mu\right)_i^2\right)\right]=0 \tag{33}$$

and

$$E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}f_i^2\left(Mu\right)_i\left(M\varepsilon\right)_i\right)\right]=0 \tag{34}$$

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

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

$$\begin{aligned} &E\left[\left(f'\varepsilon\right)\cdot\left(-2\left(1-P_{ii}\right)^{-1}f_iu_i\left(Mu\right)_i\left(M\varepsilon\right)_i\right)\right] \\ &= -2\left(1-P_{ii}\right)^{-1}\text{Cov}\left(\left(Mu\right)_i,\left(M\varepsilon\right)_i\right)\text{Cov}\left(f'\varepsilon,f_iu_i\right) \\ &= -2\left(1-P_{ii}\right)^{-1}\left(1-P_{ii}\right)\sigma_{u\varepsilon}f_i^2\sigma_{u\varepsilon} \\ &= -2\sigma_{u\varepsilon}^2f_i^2 \end{aligned} \tag{35}$$

Finally, using symmetry again, we obtain

$$E\left[\left(f'\varepsilon\right)\cdot\left(-\left(1-P_{ii}\right)^{-1}u_i^2\left(Mu\right)_i\left(M\varepsilon\right)_i\right)\right]=0 \tag{36}$$

Combining (27) - (36), 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

$$E\left[T_1T_{12}\right]=\frac{1}{H}\frac{1}{n^2}\sum_i E\left[\left(f'\varepsilon\right)\cdot\left(\delta_{1i}\delta_{2i}\right)\right]=\frac{1}{n}\frac{\sigma_\varepsilon^2}{H}\left(\frac{1}{n}\sum_i f_i^4\right)+2\frac{\sigma_u^2\sigma_\varepsilon^2+2\sigma_{u\varepsilon}^2}{n}. \tag{37}$$

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 (25), 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 (25) with $(f'\varepsilon)^2$, noting independence between $(Mu)_i$ and $f'\varepsilon$. We have

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

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

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

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

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

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

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

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

Therefore,

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

and therefore, we have

$$\begin{aligned}E[T_1T_{14}] &= \frac{1}{H^2n^3}E\left[(f'\varepsilon)^2\sum_{i=1}^n\delta_{2i}^2\right]\\ &= \frac{1}{n}\frac{1}{H}\sigma_\varepsilon^2\left(\frac{1}{n}\sum_{i=1}^n f_i^4\right)+\frac{1}{n}\left(\frac{6}{H}\sigma_\varepsilon^2\sigma_u^2+4\sigma_\varepsilon^2\sigma_u^2-\frac{6}{H}\sigma_\varepsilon^2\sigma_u^4\right)+o\left(\frac{1}{n}\right). \qquad (38)\end{aligned}$$

References

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

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

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

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

[5] Donald, S. G. and W. K. Newey (1998), "Choosing the Number of Instruments", *mimeo*.

[6] Donald, S.G., and W.K. Newey (1999), "A Jackknife Interpretation of the Continuous Updating Estimator", *mimeo*.

[7] Hahn, J., and J. Hausman (2000), "A New Specification Test of the Validity of Instrumental Variables", forthcoming in *Econometrica*.

[8] Maddala, G.S., and J. Jeong (1992), "On the Exact Small Sample Distribution of the Instrumental Variables Estimator", *Econometrica* 60, 181 - 183.

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

[10] 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.

[11] 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.

[12] Nagar, A.L. (1959), "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", *Econometrica* 27, 575 - 595.

[13] Newey, W.K., and R.J. Smith (2000), "Asymptotic Equivalence of Empirical Likelihood and the Continuous Updating Estimator", *mimeo*.

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

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

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

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

Table 1: Finite Sample Comparison of IV Estimators

DGP				Mean				RMSE				Median				InterQuartile Range			
n	K	R^2	ρ	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife
100	5	0.001	0	1.8670	-0.2704	-0.0037	-0.0238	188.5526	29.9396	0.5778	2.2778	-0.0024	0.0227	0.0064	0.0089	1.9474	1.3153	0.6328	1.0899
100	5	0.1	0	-0.0488	-1.3125	-0.0020	0.0001	3.7974	94.3252	0.2874	0.4239	-0.0037	-0.0019	-0.0009	-0.0031	0.4841	0.4411	0.3442	0.4206
100	5	0.3	0	-0.0007	-0.0010	-0.0009	-0.0010	0.1717	0.1645	0.1515	0.1640	-0.0002	0.0004	0.0015	0.0011	0.2124	0.2051	0.1924	0.2045
100	5	0.001	0.5	0.6353	0.4666	0.4927	0.5335	27.5698	21.9502	0.7016	2.9508	0.4583	0.4681	0.4860	0.4799	1.6976	1.1589	0.5541	0.9409
100	5	0.1	0.5	-0.0528	-0.0084	0.1197	0.0067	2.1640	2.1109	0.2923	0.5745	0.0030	0.0470	0.1344	0.0615	0.4691	0.4417	0.3220	0.4119
100	5	0.3	0.5	-0.0149	-0.0024	0.0337	-0.0008	0.1709	0.1682	0.1522	0.1664	-0.0007	0.0110	0.0427	0.0119	0.2086	0.2080	0.1898	0.2088
100	5	0.001	0.9	2.5881	-1.6707	0.8785	0.8533	110.5458	185.5735	0.9163	1.6862	0.8423	0.8749	0.8766	0.8638	0.9801	0.6294	0.2971	0.5026
100	5	0.1	0.9	-0.2033	-0.0854	0.2207	0.0340	2.8753	45.1669	0.3119	0.4124	-0.0010	0.0910	0.2483	0.1193	0.4250	0.4215	0.2575	0.3745
100	5	0.3	0.9	-0.0244	-0.0032	0.0617	-0.0004	0.1726	0.1752	0.1535	0.1709	0.0001	0.0224	0.0781	0.0226	0.2032	0.2100	0.1802	0.2062
100	10	0.001	0	0.8233	0.5163	0.0001	0.0080	56.4218	17.7069	0.3478	0.8226	-0.0106	-0.0136	-0.0038	-0.0035	1.9874	1.3891	0.4307	0.8164
100	10	0.1	0	0.0948	-0.0187	0.0021	0.0081	4.8792	4.3281	0.2401	0.3899	0.0092	0.0075	0.0041	0.0071	0.5822	0.5268	0.3009	0.4290
100	10	0.3	0	0.0045	0.0034	0.0016	0.0029	0.2022	0.1811	0.1451	0.1737	0.0060	0.0061	0.0041	0.0065	0.2336	0.2259	0.1893	0.2205
100	10	0.001	0.5	0.2197	1.5822	0.4956	0.4890	25.3111	103.4363	0.5827	0.8730	0.4878	0.5078	0.4944	0.4903	1.7190	1.2252	0.3918	0.7249
100	10	0.1	0.5	0.1976	0.1746	0.2251	0.0932	12.0014	8.3742	0.3148	0.3926	0.0220	0.0787	0.2347	0.1330	0.5444	0.5185	0.2795	0.4074
100	10	0.3	0.5	-0.0130	-0.0018	0.0840	0.0122	0.1858	0.1906	0.1624	0.1768	0.0060	0.0177	0.0924	0.0283	0.2279	0.2327	0.1819	0.2233
100	10	0.001	0.9	0.9984	1.3864	0.8909	0.8757	42.7934	41.8263	0.9049	0.9690	0.8578	0.8942	0.8904	0.8815	0.9046	0.6179	0.1968	0.3726
100	10	0.1	0.9	-0.0327	-0.9260	0.4044	0.1665	8.7910	85.9999	0.4359	0.3721	0.0124	0.1349	0.4155	0.2278	0.4592	0.5205	0.1998	0.3432
100	10	0.3	0.9	-0.0221	-0.0062	0.1500	0.0197	0.1780	0.2104	0.1930	0.1794	0.0032	0.0290	0.1622	0.0452	0.2139	0.2359	0.1581	0.2201
100	30	0.001	0	-0.3916	0.5787	-0.0020	-0.0020	33.4457	59.7482	0.1856	0.4095	-0.0161	0.0020	0.0026	-0.0027	1.9727	1.3446	0.2467	0.5040
100	30	0.1	0	-1.5909	0.2781	-0.0008	-0.0009	194.0282	12.3847	0.1592	0.3046	-0.0050	0.0050	0.0008	0.0017	0.8120	0.7160	0.2093	0.3636
100	30	0.3	0	-0.0092	-0.0023	0.0005	0.0012	5.1629	0.3048	0.1189	0.1810	0.0017	0.0022	0.0018	0.0031	0.2856	0.2691	0.1552	0.2221
100	30	0.001	0.5	-0.3880	0.7288	0.4944	0.4922	44.3700	20.7709	0.5208	0.6101	0.4618	0.4858	0.4943	0.4894	1.7442	1.2106	0.2120	0.4391
100	30	0.1	0.5	0.0990	-0.6021	0.3619	0.2635	13.4288	79.5736	0.3902	0.3840	0.0635	0.1698	0.3597	0.2695	0.7444	0.7090	0.1878	0.3358
100	30	0.3	0.5	-0.0281	-0.0794	0.2027	0.0814	0.5345	2.7292	0.2315	0.1933	0.0032	0.0250	0.2036	0.0907	0.2645	0.2860	0.1452	0.2185
100	30	0.001	0.9	1.6061	0.8467	0.8959	0.8931	25.6643	22.1187	0.8998	0.9121	0.8889	0.8890	0.8960	0.8935	0.9178	0.5887	0.1087	0.2209
100	30	0.1	0.9	-0.1476	-0.1871	0.6533	0.4739	6.3475	24.1535	0.6604	0.5116	0.0205	0.2731	0.6535	0.4900	0.5207	0.6709	0.1248	0.2313
100	30	0.3	0.9	-0.0263	-0.0054	0.3644	0.1443	0.1931	3.4821	0.3746	0.2138	0.0037	0.0439	0.3658	0.1593	0.2205	0.3184	0.1114	0.1968

Note: n denotes the sample size. K denotes the number of instruments. R^2 denotes the *theoretical* R^2 of the first stage regression. ρ denotes the covariance between the first stage and the second stage error terms. All results are based on 5000 Monte Carlo runs.

Table 2: Finite Sample Comparison of IV Estimators

DGP				Mean				RMSE				Median				InterQuartile Range			
n	K	R^2	ρ	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife
500	5	0.001	0	4.8289	3.8475	0.0000	0.0101	382.1974	266.0531	0.5562	2.2054	0.0109	0.0014	0.0054	0.0076	1.7504	1.2623	0.6223	1.0081
500	5	0.1	0	0.0001	0.0004	0.0004	0.0004	0.1421	0.1386	0.1307	0.1382	0.0018	0.0015	0.0015	0.0006	0.1910	0.1874	0.1778	0.1866
500	5	0.3	0	0.0002	0.0003	0.0003	0.0003	0.0686	0.0682	0.0673	0.0682	0.0011	0.0010	0.0013	0.0013	0.0948	0.0946	0.0930	0.0942
500	5	0.001	0.5	-3.8430	1.4358	0.4489	0.3598	177.6126	87.4602	0.6689	2.5949	0.3780	0.4259	0.4524	0.4277	1.5844	1.0907	0.5474	0.8936
500	5	0.1	0.5	-0.0097	-0.0001	0.0272	0.0014	0.1428	0.1396	0.1306	0.1387	0.0021	0.0118	0.0362	0.0123	0.1894	0.1869	0.1726	0.1863
500	5	0.3	0.5	-0.0021	0.0003	0.0073	0.0004	0.0688	0.0682	0.0671	0.0682	0.0008	0.0032	0.0100	0.0032	0.0939	0.0936	0.0916	0.0941
500	5	0.001	0.9	1.2619	0.6209	0.8142	0.7581	55.2251	31.0760	0.8617	1.4129	0.6407	0.7775	0.8112	0.7638	1.0935	0.6506	0.3077	0.5118
500	5	0.1	0.9	-0.0166	-0.0002	0.0490	0.0025	0.1444	0.1438	0.1323	0.1418	0.0016	0.0186	0.0629	0.0200	0.1853	0.1850	0.1623	0.1831
500	5	0.3	0.9	-0.0039	0.0004	0.0130	0.0005	0.0691	0.0686	0.0673	0.0685	0.0007	0.0045	0.0172	0.0048	0.0935	0.0924	0.0892	0.0923
500	10	0.001	0	22.8209	-2.0032	-0.0164	-0.0380	1712.2477	138.4924	0.3506	0.8410	-0.0376	-0.0049	-0.0132	-0.0218	1.8932	1.3601	0.4402	0.8029
500	10	0.1	0	-0.0043	-0.0038	-0.0033	-0.0037	0.1546	0.1503	0.1287	0.1470	-0.0034	-0.0030	-0.0030	-0.0037	0.1992	0.1931	0.1700	0.1902
500	10	0.3	0	-0.0013	-0.0012	-0.0012	-0.0012	0.0707	0.0703	0.0677	0.0702	-0.0019	-0.0018	-0.0015	-0.0015	0.0942	0.0932	0.0904	0.0934
500	10	0.001	0.5	1.2878	1.0645	0.4752	0.4469	59.3740	51.4618	0.5641	0.8706	0.4096	0.4793	0.4740	0.4545	1.7331	1.2004	0.3855	0.6966
500	10	0.1	0.5	-0.0150	-0.0049	0.0631	0.0047	0.1541	0.1544	0.1377	0.1473	-0.0038	0.0081	0.0692	0.0158	0.1962	0.1931	0.1596	0.1871
500	10	0.3	0.5	-0.0036	-0.0009	0.0174	-0.0003	0.0706	0.0704	0.0688	0.0703	-0.0016	0.0015	0.0191	0.0019	0.0931	0.0937	0.0880	0.0931
500	10	0.001	0.9	0.1546	4.8027	0.8638	0.8336	40.3894	272.5208	0.8800	0.9305	0.7270	0.8393	0.8636	0.8380	1.1346	0.6598	0.2090	0.3854
500	10	0.1	0.9	-0.0202	-0.0043	0.1173	0.0126	0.1487	0.1613	0.1605	0.1491	-0.0031	0.0178	0.1279	0.0308	0.1848	0.1951	0.1418	0.1853
500	10	0.3	0.9	-0.0052	-0.0003	0.0326	0.0007	0.0701	0.0711	0.0722	0.0707	-0.0012	0.0040	0.0360	0.0051	0.0924	0.0936	0.0848	0.0932
500	30	0.001	0	-4.8498	3.4596	0.0008	0.0005	431.8658	229.8122	0.1911	0.3980	0.0209	0.0215	0.0012	0.0046	1.9408	1.4264	0.2536	0.4925
500	30	0.1	0	-0.0005	0.0021	0.0014	0.0017	0.1909	0.1758	0.1100	0.1496	0.0044	0.0042	0.0019	0.0027	0.2309	0.2223	0.1498	0.1984
500	30	0.3	0	0.0009	0.0010	0.0010	0.0010	0.0730	0.0726	0.0640	0.0716	0.0013	0.0013	0.0017	0.0012	0.0994	0.0988	0.0867	0.0970
500	30	0.001	0.5	0.6264	1.1437	0.4934	0.4860	28.4382	67.7444	0.5197	0.5940	0.4724	0.5190	0.4961	0.4912	1.6492	1.1580	0.2163	0.4126
500	30	0.1	0.5	-0.0132	-0.0073	0.1717	0.0583	0.1738	0.1904	0.1990	0.1562	0.0031	0.0157	0.1742	0.0676	0.2185	0.2242	0.1318	0.1868
500	30	0.3	0.5	-0.0016	0.0010	0.0594	0.0076	0.0720	0.0737	0.0855	0.0722	0.0015	0.0052	0.0617	0.0110	0.0979	0.0976	0.0814	0.0956
500	30	0.001	0.9	0.7776	0.8171	0.8853	0.8709	10.0471	69.8382	0.8895	0.8891	0.7802	0.8710	0.8854	0.8709	1.0063	0.6385	0.1143	0.2243
500	30	0.1	0.9	-0.0167	-0.0165	0.3080	0.1038	0.1517	0.2542	0.3176	0.1711	0.0030	0.0252	0.3111	0.1194	0.1931	0.2398	0.1010	0.1685
500	30	0.3	0.9	-0.0032	0.0011	0.1062	0.0129	0.0697	0.0767	0.1201	0.0738	0.0017	0.0077	0.1092	0.0190	0.0936	0.0993	0.0736	0.0941

Note: n denotes the sample size. K denotes the number of instruments. R^2 denotes the *theoretical* R^2 of the first stage regression. ρ denotes the covariance between the first stage and the second stage error terms. All results are based on 5000 Monte Carlo runs.

Table 3: Finite Sample Comparison of IV Estimators

DGP				Mean				RMSE				Median				InterQuartile Range			
n	K	R^2	ρ	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife	LIML	Nagar	2SLS	Jackknife
1000	5	0.001	0	113.1191	-0.1273	0.0061	0.0275	7952.3330	13.5684	0.5161	2.8438	0.0119	0.0052	0.0041	0.0110	1.5684	1.1697	0.5791	0.9193
1000	5	0.1	0	0.0004	0.0004	0.0004	0.0004	0.0981	0.0971	0.0944	0.0971	-0.0017	-0.0018	-0.0019	-0.0017	0.1273	0.1263	0.1229	0.1268
1000	5	0.3	0	0.0001	0.0001	0.0001	0.0001	0.0486	0.0485	0.0481	0.0485	-0.0008	-0.0008	-0.0007	-0.0007	0.0644	0.0643	0.0638	0.0643
1000	5	0.001	0.5	-0.4428	-1.1049	0.4107	0.3243	36.4286	81.0943	0.6134	2.2080	0.2910	0.3862	0.4131	0.3663	1.3928	1.0468	0.5163	0.8099
1000	5	0.1	0.5	-0.0044	0.0003	0.0139	0.0007	0.0980	0.0975	0.0945	0.0974	-0.0020	0.0034	0.0168	0.0040	0.1278	0.1277	0.1238	0.1278
1000	5	0.3	0.5	-0.0010	0.0002	0.0037	0.0002	0.0486	0.0485	0.0481	0.0485	-0.0007	0.0005	0.0040	0.0006	0.0644	0.0647	0.0640	0.0646
1000	5	0.001	0.9	-1.1842	0.8814	0.7419	0.6154	100.6250	18.2819	0.7995	1.3749	0.4344	0.6858	0.7401	0.6552	1.0140	0.7044	0.3186	0.5006
1000	5	0.1	0.9	-0.0080	0.0002	0.0247	0.0008	0.0981	0.0987	0.0950	0.0985	-0.0015	0.0073	0.0307	0.0079	0.1267	0.1284	0.1196	0.1277
1000	5	0.3	0.9	-0.0020	0.0002	0.0065	0.0002	0.0485	0.0486	0.0481	0.0486	-0.0007	0.0019	0.0080	0.0018	0.0640	0.0649	0.0636	0.0650
1000	10	0.001	0	0.0625	-0.0754	0.0054	0.0070	61.9989	20.3426	0.3361	0.7572	0.0238	0.0208	0.0074	0.0133	1.6553	1.3018	0.4189	0.7589
1000	10	0.1	0	0.0008	0.0008	0.0007	0.0008	0.0999	0.0989	0.0919	0.0984	0.0020	0.0019	0.0018	0.0018	0.1319	0.1309	0.1222	0.1307
1000	10	0.3	0	0.0002	0.0002	0.0002	0.0002	0.0488	0.0487	0.0478	0.0487	0.0011	0.0011	0.0009	0.0009	0.0649	0.0645	0.0636	0.0645
1000	10	0.001	0.5	0.2726	0.3051	0.4512	0.4036	30.6185	40.4696	0.5393	0.7832	0.3435	0.4513	0.4535	0.4193	1.5017	1.1287	0.3605	0.6498
1000	10	0.1	0.5	-0.0042	0.0000	0.0347	0.0024	0.0994	0.1002	0.0964	0.0993	0.0021	0.0061	0.0390	0.0077	0.1301	0.1293	0.1172	0.1287
1000	10	0.3	0.5	-0.0010	0.0001	0.0094	0.0003	0.0487	0.0489	0.0484	0.0488	0.0009	0.0019	0.0109	0.0019	0.0646	0.0645	0.0626	0.0642
1000	10	0.001	0.9	-1.2805	0.5336	0.8143	0.7356	92.9170	13.2870	0.8321	0.8510	0.5220	0.7746	0.8149	0.7557	1.0928	0.6951	0.2117	0.3782
1000	10	0.1	0.9	-0.0081	-0.0009	0.0616	0.0034	0.0987	0.1033	0.1053	0.1013	0.0018	0.0081	0.0679	0.0120	0.1272	0.1315	0.1103	0.1295
1000	10	0.3	0.9	-0.0021	0.0000	0.0166	0.0002	0.0487	0.0493	0.0497	0.0492	0.0007	0.0024	0.0187	0.0024	0.0639	0.0648	0.0617	0.0647
1000	30	0.001	0	1.0738	2.1929	-0.0005	-0.0007	104.1162	191.0581	0.1869	0.3852	-0.0021	-0.0058	-0.0007	-0.0007	1.7626	1.3678	0.2470	0.4760
1000	30	0.1	0	-0.0006	-0.0006	-0.0005	-0.0006	0.1095	0.1082	0.0849	0.1031	-0.0015	-0.0019	-0.0013	-0.0018	0.1409	0.1405	0.1120	0.1339
1000	30	0.3	0	-0.0003	-0.0003	-0.0003	-0.0003	0.0496	0.0495	0.0464	0.0493	-0.0006	-0.0006	-0.0006	-0.0008	0.0661	0.0659	0.0620	0.0656
1000	30	0.001	0.5	7.0566	0.7395	0.4819	0.4649	571.6937	18.2415	0.5082	0.5720	0.3813	0.4732	0.4807	0.4649	1.5671	1.1775	0.2126	0.4049
1000	30	0.1	0.5	-0.0062	-0.0029	0.1016	0.0191	0.1067	0.1116	0.1292	0.1042	-0.0014	0.0021	0.1024	0.0224	0.1380	0.1476	0.1075	0.1368
1000	30	0.3	0.5	-0.0015	-0.0004	0.0306	0.0014	0.0493	0.0499	0.0548	0.0496	-0.0004	0.0010	0.0313	0.0024	0.0655	0.0662	0.0606	0.0660
1000	30	0.001	0.9	1.8229	0.1098	0.8706	0.8420	49.6333	58.7159	0.8749	0.8605	0.6427	0.8464	0.8696	0.8412	1.1197	0.6737	0.1129	0.2164
1000	30	0.1	0.9	-0.0091	-0.0046	0.1832	0.0350	0.0998	0.1194	0.1953	0.1072	-0.0008	0.0075	0.1857	0.0437	0.1291	0.1547	0.0925	0.1354
1000	30	0.3	0.9	-0.0024	-0.0004	0.0553	0.0029	0.0483	0.0508	0.0701	0.0501	-0.0008	0.0014	0.0566	0.0044	0.0637	0.0688	0.0589	0.0676

Note: n denotes the sample size. K denotes the number of instruments. R^2 denotes the *theoretical* R^2 of the first stage regression. ρ denotes the covariance between the first stage and the second stage error terms. All results are based on 5000 Monte Carlo runs.