JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **14**: 57–67 (1999)

JACKKNIFE INSTRUMENTAL VARIABLES ESTIMATION

J. D. ANGRIST,[a] G. W. IMBENS[b]* AND A. B. KRUEGER[c]
[a]*Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b]*Department of Economics, UCLA, Los Angeles, CA 90095-1477, USA*
[c]*Department of Economics, Princeton University, Princeton, NJ 08544, USA*

SUMMARY

Two-stage-least-squares (2SLS) estimates are biased towards the probability limit of OLS estimates. This bias grows with the degree of over-identification and can generate highly misleading results. In this paper we propose two simple alternatives to 2SLS and limited-information-maximum-likelihood (LIML) estimators for models with more instruments than endogenous regressors. These estimators can be interpreted as instrumental variables procedures using an instrument that is independent of disturbances even in finite samples. Independence is achieved by using a 'leave-one-out' jackknife-type fitted value in place of the usual first stage equation. The new estimators are first order equivalent to 2SLS but with finite-sample properties superior, in terms of bias and coverage rate of confidence intervals, compared to those of 2SLS and similar to those of LIML, when there are many instruments.

1. INTRODUCTION

This paper develops two simple alternatives to two-stage-least-squares (2SLS) and limited-information-maximum-likelihood (LIML) estimators for models with more instruments than endogenous regressors. The new estimators can be interpreted as instrumental variable estimators based on an asymptotically optimal instrument constructed in a manner that ensures that even in finite samples it is independent of the disturbance in the equation of interest. The key insight is that the bias associated with 2SLS is due to the use of the ith observation in constructing the optimal instrument for the ith observation. Our proposed estimators remove this dependence in a jackknife, leave-one-out, approach similar to the SSIV and USSIV estimators developed by Angrist and Krueger (1995) but without requiring an arbitrary sample split. The computation required is of the order of that for weighted least squares estimation. Both of the new Jackknife Instrumental Variables Estimators (JIVE) estimators are simple to implement in standard packages and are first-order equivalent to 2SLS and LIML. These estimators, which we proposed in Angrist, Imbens and Krueger (1993), have also been derived by Blomquist and Dahlberg (1994).

The finite sample properties of JIVE are superior, in terms of bias and coverage rates of normal-distribution-based confidence intervals, to those of 2SLS and similar to those of LIML in the case of many instruments which are only weakly correlated with the endogenous regressor. This case has received considerable attention in the recent literature on instrumental variables

* Correspondence to: Dr. G. W. Imbens, Department of Economics, UCLA, 405 Hilgard Avenue, Los Angeles, CA 90095-1477 USA. Email: imbens@econ.ucla.edu

Contract grant sponsor: Alfred P. Sloan Foundation.
Contract grant sponsor: NSF.
Contract grant number 95-11718.

CCC 0883–7252/99/010057–11$17·50

Received 5 August 1996
Revised 15 April 1997

estimation (See, e.g., Phillips, 1983; Nelson and Starz, 1990; Maddala and Jeong, 1992; Buse, 1992; Bekker, 1994; Staiger and Stock, 1994; Angrist and Krueger, 1995; Bound, Jaeger and Baker, 1995), partly in reaction to some recent applications (Angrist, 1990; Angrist and Krueger, 1991, 1992). The problem is discussed from a Bayesian perspective in Chamberlain and Imbens (1995), who find that the implicit prior distribution in 2SLS puts large probability mass on the event that the instruments combined are very informative and suggest that a prior distribution that puts substantial mass on the event that the instruments combined are not very informative might be more appropriate in many cases.

Like 2SLS, JIVE1 AND JIVE2 can be interpreted as instrumental variables estimators with a constructed instrument of the same dimension as the endogenous regressor. For 2SLS, as well as for JIVE1 and JIVE2, this constructed instrument converges to the best linear prediction of the endogenous regressor given the instruments. The probability limit of the new estimators and their first-order asymptotic distributions are therefore identical to those of 2SLS even under general misspecification. This is important because if the model is misspecified, LIML and 2SLS can have very different properties. While neither dominates the other, Fisher (1966, 1967) suggests that in misspecified models 2SLS (and therefore by implication JIVE1 and JIVE2) may be preferable to LIML.

2. THE BIAS OF TWO-STAGE LEAST SQUARES

The basic model we consider has two equations. The first, describing the relation of interest between a scalar endogenous variable Y_i and a row vector of potentially endogenous regressors X_i is:

$$Y_i = X_i\beta + \varepsilon_i.$$

The second equation captures the relation between the endogenous regressors and the instruments Z_i:

$$X_i = Z_i\pi + \eta_i.$$

Here, X_i is an L-dimensional row vector and the instrument Z_i is a K-dimensional row vector, with $K \geqslant L$. The number of overidentifying restrictions is $K - L$. In matrix notation we can write this model as

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \tag{1}$$

$$\mathbf{X} = \mathbf{Z}\pi + \eta, \tag{2}$$

where $\mathbf{Y}$ and ε_i are N vectors with typical element Y_i and ε_i, and $\mathbf{X}$, $\mathbf{Z}$ and η are $N \times L$, $N \times K$, and $N \times L$-dimensional matrices with typical row X_i, Z_i, and η_i respectively. If there are M common elements in the vector of regressors and the vector of instruments, then M columns of the $N \times L$ matrix η are identically zero.

We assume that conditional on the instruments Z_i the disturbance ε_i has expectation zero and variance σ^2. We also assume that $E[\eta \mid \mathbf{Z}] = 0$ and $E[\eta_i'\eta_i \mid \mathbf{Z}] = \Sigma_\eta$, with rank $L - M$. The expectation $E[\varepsilon_i\eta_i' \mid \mathbf{Z}]$ is equal to the L-dimensional column vector $\sigma_{\varepsilon\eta}$ and the probability limits of $\mathbf{Z}'\mathbf{Z}/N$ and $\mathbf{X}'\mathbf{X}/N$ are denoted by Σ_Z and Σ_X respectively. Finally we assume that all observations of the triple (Y_i, X_i, Z_i) are independent and identically distributed.

As is well known, the OLS estimator for β,

$$\hat{\beta}_{\text{ols}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}),$$

is not consistent if $\sigma_{\varepsilon\eta}$ differs from zero. Its probability limit equals $\beta + (\pi'\Sigma_Z\pi + \Sigma_\eta)^{-1}\sigma_{\varepsilon\eta}$. The optimal instrumental variables (IV) estimator, using the optimal instrument of lowest dimension, $\mathbf{Z}\pi$, is:

$$\hat{\beta}_{\text{opt}} = ((\mathbf{Z}\pi)'\mathbf{X})^{-1}((\mathbf{Z}\pi)'\mathbf{Y}).$$

This estimator is not feasible because π is unknown.

The 2SLS estimator, a feasible version of the optimal IV estimator, is:

$$\hat{\beta}_{\text{2sls}} = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}).$$

It is useful to work with a characterization of $\hat{\beta}_{\text{2sls}}$ as an instrumental variable estimator using the constructed instrument $\mathbf{Z}\hat{\pi}$ where $\hat{\pi} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})$, so we can write

$$\hat{\beta}_{\text{2sls}} = ((\mathbf{Z}\hat{\pi})'\mathbf{X})^{-1}(\mathbf{Z}\hat{\pi})'\mathbf{Y}). \tag{3}$$

The limiting distributions of both $\sqrt{N}(\hat{\beta}_{\text{2sls}} - \beta)$ and $\sqrt{N}(\hat{\beta}_{\text{opt}} - \beta)$ are normal with mean zero and variance $(\pi'\Sigma_z\pi)^{-1}\sigma_\varepsilon^2$.

The motivation for our approach begins with the observation that $\hat{\beta}_{\text{opt}}$ has much better small sample properties than $\hat{\beta}_{\text{2sls}}$ in the presence of many instruments, even though the two estimators have the same asymptotic normal distribution. This follows directly from the Nagar (1959) bias formula, which shows that keeping the explanatory power of the instruments constant while increasing the number of instruments increases the bias of $\hat{\beta}_{\text{2sls}}$, while obviously increasing the number of instruments with $Z_i\pi$ fixed does not affect the properties of $\hat{\beta}_{\text{opt}}$. The intuition for the former is that the first-stage fitted values, $\mathbf{Z}\hat{\pi}$, can be written as $P_Z\mathbf{X} = \mathbf{Z}\pi + P_Z\eta$ where P_Z is the projection matrix $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. The second component of these fitted values, $P_Z\eta$, is correlated with η and hence with ε. Formally,

$$E[\varepsilon_i Z_i\hat{\pi}] = E[E[\varepsilon_i Z_i\hat{\pi} \mid \mathbf{Z}]] = E[Z_i(\mathbf{Z}'\mathbf{Z})^{-1}Z_i' \cdot E[\varepsilon_i\eta_i \mid \mathbf{Z}]] = E[Z_i(\mathbf{Z}'\mathbf{Z})^{-1}Z_i' \cdot \sigma_{\varepsilon\eta}'] = (K/N) \cdot \sigma_{\varepsilon\eta}'$$

Even though this correlation vanishes in large samples as $K/N \to 0$, it increases with the number of instruments for fixed sample size and fixed $\sigma_{\varepsilon\eta}$.

The previous discussion suggests that the bias of $\hat{\beta}_{\text{2sls}}$ towards $\hat{\beta}_{\text{ols}}$ is related to the difference between the estimated instrument $Z_i\hat{\pi}$ and the optimal instrument $Z_i\pi$. This leads us to develop new estimators of β based on different estimates of the optimal instrument $Z_i\pi$. The key feature of our approach is that these alternative estimates of the optimal instrument are independent of ε_i even in finite samples, unlike the standard estimate $Z_i\hat{\pi}$ which is only asymptotically independent of ε_i. Although these estimated instruments differ from $Z_i\hat{\pi}$ and $Z_i\pi$ in finite samples, the difference in variance goes to zero fast enough to give the resulting estimators the same first-order asymptotic properties as both $\hat{\beta}_{\text{opt}}$ and $\hat{\beta}_{\text{1sls}}$. The resulting bias reduction is such that in models with many instruments the associated estimators of β are superior to 2SLS.

3. JACKKNIFE INSTRUMENTAL VARIABLES ESTIMATION

Note that the *i*th row of the estimated instrument $\mathbf{Z}\hat{\pi}$ underlying 2SLS can be written as:

$$Z_i\hat{\pi} = Z_i(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X}). \tag{4}$$

Let $\mathbf{Z}(i)$ and $\mathbf{X}(i)$ denote $(N-1)\times K$ and $(N-1)\times L$-dimensional matrices equal to $\mathbf{Z}$ and $\mathbf{X}$ respectively with the *i*th row removed. JIVE1 removes the dependence of the constructed instrument $Z_i'\hat{\pi}$ on the endogenous regressor for observation i by using

$$\tilde{\pi}(i) = (\mathbf{Z}(i)'\mathbf{Z}(i))^{-1}(\mathbf{Z}(i)'\mathbf{X}(i)),$$

as an estimate of π rather than $\hat{\pi} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})$. The estimate of the optimal instrument is

$$Z_i\tilde{\pi}(i) = Z_i(\mathbf{Z}(i)'\mathbf{Z}(i))^{-1}(\mathbf{Z}(i)'\mathbf{X}(i)).$$

Because ε_i is independent of X_j if $j \neq i$, it follows that

$$E[\varepsilon_i Z_i\tilde{\pi}(i)] = E[Z_i(\mathbf{Z}(i)'\mathbf{Z}(i))^{-1}(\mathbf{Z}(i)')E[\mathbf{X}(i)\varepsilon_i \mid \mathbf{Z}]] = 0,$$

implying $E[\hat{\mathbf{X}}'_{\text{jive1}}\mathbf{X}/N] = E[(\mathbf{Z}\pi)'X/N]$ and $E[\hat{X}'_{\text{jive1}}\mathbf{Y} \mid N] = E[(\mathbf{Z}\pi)'Y/N]$, where $\hat{\mathbf{X}}_{\text{jive1}}$ is the $N \times L$-dimensional matrix with *i*th row $Z_i\tilde{\pi}(i)$. The associated estimator for β, denoted by JIVE1, is equal to:

$$\hat{\beta}_{\text{jive1}} = (\hat{\mathbf{X}}'_{\text{jive1}}\mathbf{X})^{-1}(\hat{\mathbf{X}}'_{\text{jive1}}\mathbf{Y}).$$

To calculate this estimator the researcher is not required to calculate the N regression coefficients $\tilde{\pi}(i)$. Rather, the calculation only requires the evaluation of the constructed instrument, $Z_i\tilde{\pi}(i)$, which can be calculated as

$$Z_i\tilde{\pi}(i) = Z_i\frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{1 - Z_i(\mathbf{Z}'\mathbf{Z})^{-1}Z_i'}(\mathbf{Z}'\mathbf{X} - Z_i'X_i) = \frac{Z_i\hat{\pi} - h_iX_i}{1 - h_i}, \tag{5}$$

where $h_i = Z_i(\mathbf{Z}'\mathbf{Z})^{-1}Z_i'$. Given $Z_i\tilde{\pi}(i)$, calculation of $\hat{\beta}_{\text{jive1}}$ is straightforward. This calculation requires only two passes through the data, one to calculate regular fitted first-stage values and the leverage, h_i, (see Cook, 1979), and a second using $\tilde{\pi}(i)Z_i$ as an instrument.

An alternative estimator, denoted by JIVE2, also based on the idea of eliminating the correlation between the estimate of $Z_i\pi$ and X_i, adjusts only the $\mathbf{Z}'\mathbf{X}$ component of $\hat{\pi} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})$. Define

$$\tilde{\pi}(i) = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}(i)'\mathbf{X}(i)) \cdot (N/(N-1)) = (N/N-1)) \cdot (\hat{\pi} - (\mathbf{Z}'\mathbf{Z})^{-1}Z_i'X_i),$$

as the associated first-stage parameter for observation i. A formulation similar to equation (5) is:

$$Z_i\tilde{\pi}(i) = Z_i\frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{1 - 1/N}(\mathbf{Z}'\mathbf{X} - Z_i'X_i) = \frac{Z_i\hat{\pi} - h_iX_i}{1 - 1/N}, \tag{6}$$

with the same definition for h_i. Again this estimate of the optimal instrument is independent of ε_i in finite samples:

$$E[\varepsilon_i Z_i \tilde{\pi}(i)] = E[Z_i(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}(i)')E[\mathbf{X}(i)\varepsilon_i \mid \mathbf{Z}] = 0.$$

The resulting estimator for β is:

$$\hat{\beta}_{\text{jive2}} = (\hat{\mathbf{X}}'_{\text{jive2}}\mathbf{X})^{-1}(\hat{\mathbf{X}}'_{\text{jive2}}\mathbf{Y}),$$

where the $N \times L$-dimensional matrix $\hat{\mathbf{X}}'_{\text{jive2}}$ has ith row equal to $Z_i\tilde{\pi}(i)$. Note that the only difference between JIVE1 and JIVE2 is the difference between $1 - Z_i(\mathbf{Z}'\mathbf{Z})^{-1}Z_i'$ and $1 - 1/N$ in the denominator of equations (5) and (6). The second estimator also requires only two passes through the data.

In both cases the estimator for $Z_i\pi$ is consistent. The probability limit of the estimators of β and their first-order asymptotic distribution are therefore the same as those of $\hat{\beta}_{\text{opt}}$ and $\hat{\beta}_{\text{2sls}}$.

4. A MONTE CARLO STUDY

This section reports evidence on the finite sample behaviour of the JIVE estimators, focusing on robust measures of bias. In particular, we report quantiles of the Monte Carlo sampling distribution along with the median absolute error. Mean squared error is not likely to be as useful a standard for comparison because neither LIML or the JIVE estimators necessarily have first or second moments. For additional theoretical discussion of the bias of JIVE, 2SLS amd LIML, see our working paper (Angrist, Imbens and Krueger, 1995). An important conclusion of this discussion and earlier work on the finite-sample properties of IV estimators is that LIML is less biased than 2SLS in models with many instruments.

In addition to measures of bias, we also report coverage rates for 95% confidence intervals computed using the usual asymptotic approximation to the distribution of OLS, 2SLS, and LIML (i.e. the estimate plus or minus 1·96 times the asymptotic standard error). For the JIVE estimators, we report coverage rates based on asymptotic standard errors for a just-identified IV estimator using $\hat{\mathbf{X}}'_{\text{jive1}}$ and $\hat{\mathbf{X}}'_{\text{jive2}}$ as instruments. The justification for this is pragmatic: if the usual approximation works in the sense of providing accurate coverage for the approximately unbiased LIML and JIVE estimators, there would seem to be little reason to report more sophisticated approximations such as those developed by Bekker (1994) and Staiger and Stock (1994). In fact Bekker (1994) finds that some theoretically more accurate approximations to the limiting distribution of LIML based on group-asymptotics provide little or no improvement over the usual asymptotic approximation in cases with a linear first stage. This is not true for 2SLS, however. The results of our simulations confirm and extend this: asymptotic confidence intervals for LIML and JIVE estimators turn out to be remarkably accurate while in contrast conventional asymptotic confidence intervals for 2SLS are quite poor.

We begin with a model where there is a single overidentifying restriction. The second model is similar, with the modification that there are a large number of instruments relative to the number of regressors. In both of these first two models, the errors are homoscedastic and the first-stage regression is linear, so that LIML is the maximum likelihood estimator. In the third model, the first stage is non-linear and heteroscedastic. Here there is less reason to expect LIML to have good small sample properties since it is no longer the maximum likelihood estimator. In both the second and third models, 2SLS should be badly biased because of the large number of overidentifying restrictions. The fourth model sets the true reduced-form coefficients to zero for all instruments in an attempt to ascertain how misleading the estimators might be in this non-identified case. In the final experiment we introduce some misspecification where one of the instruments has incorrectly been left out of the main regression. The models and results are as follows:

Model 1

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \varepsilon_i,$$

$$X_{i1} = \pi_0 + \sum_{j=1}^{2} \pi_j \cdot Z_{ij} + \eta_i,$$

with $\beta_1 = 1$, $\beta_0 = 0$, $\pi_0 = 0$, $\pi_1 = 0{\cdot}3$, and $\pi_2 = 0$. Here, $K = 3$ and $L = 2$, and

$$\begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \Bigg| \mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0{\cdot}25 & 0{\cdot}20 \\ 0{\cdot}20 & 0{\cdot}25 \end{pmatrix}\right).$$

All Z_{ij} are independent, normally distributed random variables with mean zero and unit variance.

The first panel of Table I presents quantiles of the sampling distributions of the estimators, as well as the median absolute error and coverage rates. In this set of simulations, LIML, JIVE1 and JIVE2 all have median absolute error close to that of 2SLS, which is the estimator with the minimum median absolute error. But confidence interval coverage is actually more accurate for JIVE and LIML than for 2SLS. It is not surprising that LIML does very well, however, since it is the maximum likelihood estimator under normality and in this example the disturbances are in fact normal. Note that confidence interval coverage for JIVE is as good as that for LIML, in spite of some asymmetry in the Monte Carlo sampling distribution of JIVE.

Model 2

Model 2 adds 18 worthless instruments to the design in Model 1. This is a situation where we expect the performance of 2SLS to deteriorate, as in the experiments reported in Bound, Jaeger and Baker (1995).

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \varepsilon_i,$$

$$X_{i1} = \pi_0 + \sum_{j=1}^{20} \pi_j \cdot Z_{ij} + \eta_i,$$

with $\beta_1 = 1$, $\beta_0 = 0$, $\pi_0 = 0$, $\pi_1 = 0{\cdot}3$, and $\pi_j = 0$ for $j = 2, 3, \ldots, 20$. Here, $K = 21$ and $L = 2$, and

$$\begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \Bigg| \mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0{\cdot}25 & 0{\cdot}20 \\ 0{\cdot}20 & 0{\cdot}25 \end{pmatrix}\right).$$

All Z_{ij} are independent, normally distributed random variables with mean zero and unit variance.

Table I. Simulations

Estimator	Quantiles around β_1 0·10	0·25	0·50	0·75	0·90	Median Absolute error	Coverage rate 95% conf. interval
Model 1: $N = 100$, $L = 2$, $K = 3$; 5000 replications							
OLS	0·50	0·55	0·59	0·64	0·67	0·59	0·00
2SLS	−0·19	−0·06	0·04	0·14	0·22	0·11	0·91
LIML	−0·26	−0·13	0·00	0·11	0·19	0·12	0·96
JIVE1	−0·40	−0·20	−0·05	0·07	0·17	0·13	0·96
JIVE2	−0·40	−0·20	−0·05	0·07	0·17	0·13	0·96
Model 2: $N = 100$, $L = 2$, $K = 21$; 5000 replications							
OLS	0·51	0·55	0·59	0·63	0·67	0·59	0·00
2SLS	0·14	0·21	0·28	0·35	0·41	0·28	0·31
LIML	−0·31	−0·14	0·00	0·11	0·20	0·13	0·94
JIVE1	−0·61	−0·28	−0·04	0·12	0·23	0·17	0·94
JIVE2	−0·63	−0·29	−0·04	0·11	0·23	0·17	0·94
Model 3: $N = 100$, $L = 2$, $K = 21$; 5000 replications							
OLS	0·12	0·14	0·17	0·20	0·23	0·17	0·03
2SLS	0·04	0·10	0·16	0·22	0·27	0·16	0·57
LIML	−0·59	−0·15	0·10	0·32	0·80	0·25	0·97
JIVE1	−0·69	−0·13	0·16	0·43	0·95	0·32	0·97
JIVE2	−0·41	−0·13	0·04	0·16	0·33	0·15	0·95
Model 4: $N = 100$, $L = 2$, $K = 21$; 5000 replications							
OLS	0·72	0·76	0·80	0·84	0·87	0·80	0·00
2SLS	0·62	0·71	0·80	0·89	0·97	0·80	0·00
LIML	−1·14	0·18	0·81	1·42	2·69	1·01	0·71
JIVE1	−0·40	0·41	0·80	1·21	2·07	0·88	0·71
JIVE2	−0·35	0·41	0·80	1·20	2·05	0·88	0·71
Model 5: $N = 100$, $L = 2$, $K = 21$; 5000 replications							
OLS	0·50	0·54	0·59	0·64	0·68	0·59	0·00
2SLS	0·10	0·19	0·28	0·37	0·45	0·28	0·38
LIML	−1·13	−0·69	−0·41	−0·21	−0·06	0·41	0·93
JIVE1	−0·66	−0·28	−0·04	0·14	0·28	0·20	0·93
JIVE2	−0·67	−0·28	−0·05	0·14	0·28	0·20	0·94

The second panel of Table I presents Monte Carlo statistics for this model. In this set of simulations, LIML, JIVE1 and JIVE2 are all superior to 2SLS and OLS in terms of median absolute error. Unlike 2SLS and OLS, the three other estimators are essentially median unbiased and the asymptotic confidence intervals have very good coverage. LIML is less dispersed than both JIVE1 and JIVE2 with the latter having thick tails. The asymptotic coverage for 2SLS is poor. Again, it is not surprising that LIML does very well here since in this example the disturbances are normally distributed.

Model 3

The third model has the same basic structure as before, except that the relationship between X_i and Z_i is non-linear and heteroscedastic. As in Model 2, the model is estimated with 20 linear instruments so that non-linearities in the first stage are ignored in the estimation.

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \varepsilon_i,$$

$$X_{i1} = \pi_0 + \sum_{j=1}^{20} \pi_j \cdot Z_{ij} + 0{\cdot}3 \cdot \sum_{j=2}^{20} Z_{ij}^2 + \eta_{i0} \cdot \sum_{j=2}^{20} Z_{ij}^2/19,$$

with $\beta_1 = 1$, $\beta_0 = 0$, $\pi_0 = 0$, $\pi_1 = 0{\cdot}3$, and $\pi_j = 0$ for $j = 2, 3, \ldots, 20$. Here, $K = 21$ and $L = 2$, and

$$\begin{pmatrix} \varepsilon_i \\ \eta_{i0} \end{pmatrix} \Big| \mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1{\cdot}0 & 0{\cdot}8 \\ 0{\cdot}8 & 1{\cdot}0 \end{pmatrix} \right).$$

The third panel of Table I presents Monte Carlo statistics for this model. As expected, OLS and 2SLS are still biased, as evidenced by the fact that almost all probability is concentrated on one side of the true value of β_1 for these estimators. Moreover, in spite of the low median absolute error of 2SLS in this case, the asymptotic coverage of 2SLS is very poor.

JIVE1 and LIML do not do as well in Model 3 as in Models 1 and 2. But JIVE2 is the best estimator in terms of median-bias and median absolute error. It is clearly superior to LIML and even to JIVE1 in this model, both in terms of bias and (slightly) in terms of asymptotic coverage. The medians of JIVE1, LIML, and 2SLS are all similar. The large difference in spread between JIVE1 and JIVE2 is surprising and only in this type of non-linear example have we seen such a difference. It is important to note, however, that in contrast with 2SLS, even the highly dispersed JIVE1 generates an asymptotic confidence interval with reasonably accurate coverage. The lack of dispersion in 2SLS, reflected in the 2SLS asymptotic standard errors, actually leads to highly misleading inferences.

Model 4

The fourth model has the same basic structure as Model 2 but all coefficients in the reduced form are set to zero.

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \varepsilon_i,$$

$$X_{i1} = \pi_0 + \sum_{j=1}^{20} \pi_j \cdot Z_{ij} + \eta_i,$$

with $\beta_1 = 1$, $\beta_0 = 0$, $\pi_j = 0$, for all j. Again, $L = 2$ and $K = 21$, and

$$\begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \Big| \mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0{\cdot}25 & 0{\cdot}20 \\ 0{\cdot}20 & 0{\cdot}25 \end{pmatrix} \right).$$

The fourth panel of Table I presents Monte Carlo statistics for this model. The two JIVE estimators and LIML are much more dispersed than either OLS or 2SLS in this case, suggesting that a researcher would not be misled by JIVE or LIML estimates into thinking that the instruments generate reliable inferences regarding the coefficient of interest. It is also interesting to note that the correlation between JIVE and LIML in this model is very low, unlike in models where the instruments are valid. This suggests that a comparison of JIVE and LIML could provide a useful check on the validity of inferences in applications with weak instruments.

Model 5

The fifth model has the same basic structure as Model 2 but one of the instruments that has a zero coefficient in the second regression has a non-zero coefficient in the first regression from which it is inappropriately left out.

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \varepsilon_i,$$

$$X_{i1} = \pi_0 + \sum_{j=1}^{20} \pi_j \cdot Z_{ij} + \eta_i,$$

with $\beta_1 = 1$, $\beta_0 = 0$, $\pi_1 = 0{\cdot}3$, and $\pi_j = 0$ for $j > 1$. Again, $L = 2$ and $K = 21$, but now $\varepsilon_i = 0{\cdot}2 \cdot Z_{i2} + \nu_i$, with

$$\begin{pmatrix} \nu_i \\ \eta_i \end{pmatrix} \Bigg| \mathbf{Z} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0{\cdot}25 & 0{\cdot}20 \\ 0{\cdot}20 & 0{\cdot}25 \end{pmatrix} \right).$$

Because Y_i directly depends on Z_{i2}, Z_{i2} is not a valid instrument. However, because X_i does not depend on Z_{i2}, the estimated instrument does not depend on Z_{i2}, and therefore 2SLS, as well as JIVE1 and JIVE2 are not affected by this form of specification, and all three are consistent. LIML is affected by this direct effect of Z_{i2} on Y_i and is not consistent.

The fifth panel of Table I presents Monte Carlo statistics for this model. The two new estimators, JIVE1 and JIVE2, are superior to the other estimators considered. While 2SLS is still consistent, because of the many instruments it is badly biased towards OLS. LIML is clearly also biased.

5. RETURNS TO EDUCATION USING QUARTER OF BIRTH AS INSTRUMENT

In this section, we return to the Angrist and Krueger (1991) application that has motivated some of the recent literature on instrumental variables estimates with many weak instruments. Angrist and Krueger (1991) estimated schooling coefficients using quarter of birth as an instrument in a sample of 329,500 men born 1930–39 from the 1980 census. The dependent variable is the log weekly wage. In one version of this model, there are 30 instruments created by interacting quarter and year of birth. In a second version there are 180 instruments constructed by adding interactions of 50 state and quarter of birth dummies to the 30 original instruments. The appendix to Angrist and Krueger (1991) provides a detailed description of the data.

Table II reports schooling coefficients generated by different estimators applied to the Angrist and Krueger data. Exogenous covariates are listed in the table (these are either state effects or state and year effects). Table II shows that all IV estimators give similar results. This is important because Bound, Jaeger and Baker (1995) and Angrist and Krueger (1995) note that if the

Table II. Angrist–Krueger data

No. of instr.	State effects	Year effect	OLS	2SLS	LIML	JIVE1	JIVE2
30	No	Yes	0·071	0·089	0·093	0·096	0·096
			(0·0003)	(0·016)	(0·018)	(0·022)	(0·022)
180	Yes	Yes	0·067	0·093	0·106	0·121	0·121
			(0·0003)	(0·009)	(0·012)	(0·020)	(0·020)

instruments are in fact uncorrelated with schooling (as in the second model in the Monte Carlo section), 2SLS could still give results very close to OLS. In contrast, the two JIVE estimators and LIML would not be expected to give similar or statistically significant estimates in such circumstances. Also worth noting is the fact that the standard errors of the two JIVE estimators are considerably lower than the standard errors for the USSIV estimators discussed in Angrist and Krueger (1995).

6. CONCLUSION

In this paper we present two alternatives to 2SLS, LIML and other k-class estimators for models with endogenous regressors. In models with many weak instruments these estimators perform much better than 2SLS, and have finite sample properties similar to those of LIML. Moreover, under certain forms of misspecification the JIVE estimators may have less bias than LIML. The JIVE estimators therefore seem to provide useful alternatives in applications where there is concern about the number of instruments, although the question of when JIVE estimators should be preferred to LIML remains open.

Instrumental variables is one special case in a larger class of generalized method of moments estimators where a weight matrix is estimated in an initial stage and a weighted set of restrictions is imposed in a second stage. In some cases, using the same data set to estimate the weight matrix and to impose the moment restrictions leads to poor small sample properties. In this context, Altonji and Segal (1996) discuss a sample splitting approach similar to that used by Angrist and Krueger (1995). The jackknife idea developed here for instrumental variables would also appear to extend to estimators such as those considered by Altonji and Segal.

ACKNOWLEDGEMENTS

We are grateful for comments by Gary Chamberlain and Whitney Newey. Imbens thanks the Alfred P. Sloan Foundation for financial support through a Research Fellowship and the NSF for grant 95-11718. Also, we thank seminar participants at the Harvard-MIT Econometrics workshop, the Duke-UNC-SAS Econometrics workshop, Tel Aviv University, and Hebrew University for helpful comments.

REFERENCES

Altonji, J. and L. Segal (1996), 'Small-sample bias in GMM estimation of covariance structures', *Journal of Business and Economic Statistics*, June.

Angrist, J. (1990), 'Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records', *American Economic Review*, **80**, 313–335.

Angrist, J. and A. Krueger (1991), 'Does compulsory school attendance affect schooling and earnings', *Quarterly Journal of Economics*, **106**, 979–1014.

Angrist, J. and A. Krueger (1992). 'The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples', *Journal of the American Statistical Association*, **87**, June.

Angrist, J. D. and A. Krueger (1995). 'Split sample instrumental variables estimates of the return to schooling', *Journal of Business and Economic Statistics*, April.

Angrist, J. W., G. W. Imbens and A. Krueger (1993), 'Inference with instrumental variables', National Science Foundation proposal, August.

Angrist, J., G. Imbens, and A. Krueger, (1995) 'Jackknife Instrumental Variables Estimation', NBER technical working paper.

Bekker, P. (1994), 'Alternative approximations to the distributions of instrumental variables estimators', *Econometrica*, **62**, 657–682.

Blomquist, S. and M. Dahlberg (1994), 'Small sample properties of jackknife instrumental variables estimators: experiments with weak instruments', mimeo, Uppsala University.

Bound. J., D. Jaeger and R. Baker (1995), 'Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association*, June, 443–450.

Buse, A. (1992), 'The bias of instrumental variables estimators', *Econometrica*, **60**, 173–180.

Chamberlain, G. and G. Imbens (1995), 'Hierarchical Bayes models with many instruments', mimeo, Department of Economics, Harvard University.

Cook, R. D., (1979), 'Influential Observations in Linear Regressions', *Journal of the American Statistical Association*, **74**, 169–174.

Fisher, F. (1966), 'The relative sensitivity to specification error of different k-class estimators', *Journal of the American Statistical Association*, **61**, 345–356.

Fisher, F. (1967), 'Approximate specification and the choice of a k-class estimator, *Journal of the American Statistical Association*, **62**, 1265–1276.

Maddala, G. S. and J. Jeong (1992), 'On the exact small sample distribution of the instrumental variables estimator', *Econometrica*, **60**, 181–183.

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.

Nelson, C. and R. Starz (1990), 'The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one', *Journal of Business*.

Phillips, P. C. B. (1983), 'Exact small sample properties in the simultaneous equations model', Chapter 8 in Z. Griliches and M. Intriligator (eds), *The Handbook of Econometrics*, North-Holland, Amsterdam.

Staiger, D. and J. Stock (1994), 'Instrumental variables regression with weak instruments', NBER Technical Working Paper, No. 151, January.

