The Stata Journal

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press, and Stata is a registered trademark of StataCorp LP.

The Stata Journal (2006)
6, Number 3, pp. 364–376

Jackknife instrumental variables estimation in Stata

Brian P. Poi
StataCorp
College Station, TX
bpoi@stata.com

Abstract. The two-stage least-squares (2SLS) instrumental variables estimator is commonly used to address endogeneity. However, the estimator suffers from bias that is exacerbated when the instruments are only weakly correlated with the endogenous variables and when many instruments are used. In this article, I discuss jackknife instrumental variables estimation as an alternative to 2SLS. Monte Carlo simulations comparing the jackknife instrument variables estimators to 2SLS and limited information maximum likelihood (LIML) show that two of the four variants perform remarkably well even when 2SLS does not. In a weak-instrument experiment, the two best performing jackknife estimators also outperform LIML.

Keywords: st0108, jive, 2SLS, LIML, JIVE, instrumental variables, endogeneity, weak instruments

1 Introduction

The two-stage least-squares (2SLS) estimator is perhaps the most common instrumental variables estimator used to address endogeneity in econometric applications. However, its use has come under increasing scrutiny in recent decades because of potential finite-sample and asymptotic problems. Much of the focus has been on the weak-instrument problem and the use of many instruments.

For example, Angrist and Krueger (1991) fitted an earnings equation in which education was treated as an endogenous variable. They argued that the season in which a person was born would be correlated with educational attainment. States typically require students to remain in school until their 16th birthday. Because the school year starts in autumn, students born early in the year begin school at a later age than students born later in the year and thus can drop out of school having spent less time in the classroom. Angrist and Krueger therefore used season-of-birth indicators (and interactions with year of birth) as instruments for their education variable.

Although Angrist and Krueger (1991) had a sample of more than 300,000 students, later work by Bound, Jaeger, and Baker (1995) and others showed that Angrist and Krueger's estimates were biased because the correlation between their education variable and their instruments was weak. By definition an instrument must be uncorrelated with the structural error term to be valid, though in practice that condition may not always be met. Bound, Jaeger, and Baker (1995) showed that if the instruments are only weakly correlated with the endogenous variable, then any such correlation between the instruments and the structural error term can have a profound impact on the consistency of the 2SLS estimator. Moreover, the 2SLS estimator is biased toward the ordinary least-squares (OLS) estimator, and that bias becomes more severe as the correlation between the endogenous regressor and the instruments approach zero. Regressions of education on Angrist and Krueger's seasonal dummy instruments resulted in R^2 values as low as 0.0001, suggesting that the bias of 2SLS was severe. Also see Nelson and Startz (1990) and Stock, Wright, and Yogo (2002) for introductions to using weak instruments.

Even when the instruments are relevant in the sense that they are sufficiently correlated with the endogenous variable, the 2SLS estimator still exhibits a bias that increases as more instruments are used. Researchers often use many instruments under the presumption that doing so will make up for the instruments' being weak. However, that logic is faulty, because the bias of the 2SLS estimator increases with the number of instruments; and, to the extent that the instruments are correlated with one another, using more of them may not aid in identification.

Recently, Angrist, Imbens, and Krueger (1999) and Blomquist and Dahlberg (1999) proposed estimators that attempt to eliminate the finite-sample bias of 2SLS. Because they are based on a "leave one out" approach reminiscent of the jackknife, they are known as the jackknife instrumental variables estimators (JIVEs). Monte Carlo simulations show that they often work well for bias and coverage probabilities, even when the conventional 2SLS estimator does not. However, JIVEs are not a panacea, for their distributions are much more dispersed than the distribution of the 2SLS estimator.

The rest of this paper is organized as follows. Section 2 illustrates the source of the finite-sample bias of the 2SLS estimator. Section 3 presents four variants of the JIVE and shows how they circumvent the source of the finite-sample bias discussed in section 2. Section 4 presents the syntax for the Stata command `jive`, which implements these estimators. Section 5 provides Monte Carlo evidence, and section 6 concludes.

2 Finite-sample bias of 2SLS

The model we consider is

$$\begin{aligned} y_i &= \mathbf{x}_i\beta + \epsilon_i \\ \mathbf{x}_i &= \mathbf{z}_i\pi + \nu_i \end{aligned}$$

where $\mathbf{x}_i$ and ν_i are $1 \times L$, $\mathbf{z}_i$ is $1 \times K$, β is $L \times 1$, and π is $K \times L$. By assumption, some elements of $\mathbf{x}_i$ are correlated with ϵ_i, though all elements of $\mathbf{z}_i$ are uncorrelated with ν_i. Let $\sigma_{\epsilon\nu}$ denote the $1 \times L$ vector of covariances between ϵ_i and ν_i. Let $\mathbf{y}$ denote the $N \times 1$ result of stacking y_i for $i = 1, \ldots, N$, where N is the sample size; and similarly define $\mathbf{X}$, $\mathbf{Z}$, ϵ, and ν.

If the jth column of $\mathbf{x}_i$ is exogenous, then a corresponding column of $\mathbf{z}_i$ is equal to x_{ij}, ν_{ij} is equal to 0, and one of the rows of π has one element equal to one with the remaining elements zero. Identification of β requires that $K \geq L$.

If π were known, then the optimal instrumental variables estimator of β would be

$$\widehat{\beta}_{\mathrm{opt}} = \left(\pi'\mathbf{Z}'\mathbf{X}\right)^{-1}\pi'\mathbf{Z}'\mathbf{y}$$

Obviously, $E(\widehat{\beta}_{\mathrm{opt}}) = \beta$; and $\mathbf{Z}\pi$ is therefore known as the optimal-instrument matrix. However, π is not known, and the 2SLS estimator uses the OLS estimator of π:

$$\begin{aligned}\widehat{\beta}_{\mathrm{2SLS}} &= \left(\widehat{\pi}'\mathbf{Z}'\mathbf{X}\right)^{-1}\widehat{\pi}'\mathbf{Z}'\mathbf{y}\\ &= \left\{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right\}^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}\end{aligned}$$

Taking expectations of that expression, we see that the bias of $\widehat{\beta}_{\mathrm{2SLS}}$ depends on $E\left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\epsilon\right)$, which is equal to $E\left(\widehat{\pi}\mathbf{Z}'\epsilon\right)$. For any observation i,

$$\begin{aligned}E\left(\widehat{\pi}'\mathbf{z}_i'\epsilon_i\right) &= E\left\{E\left(\widehat{\pi}'\mathbf{z}_i'\epsilon_i|\mathbf{Z}\right)\right\}\\ &= E\left[E\left\{\left(\pi'\mathbf{z}_i'\epsilon_i + \nu_i'\mathbf{z}_i(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_i'\epsilon_i\right)|\mathbf{Z}\right\}\right]\\ &= E\left\{\mathbf{z}_i(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_i'\cdot E\left(\nu_i\epsilon_i\right)\right\}\\ &= \frac{K}{N}\,\sigma_{\epsilon\nu}\end{aligned} \tag{1}$$

where the last equality follows from the expected value of the ith diagonal element of the hat matrix (K/N).

Equation (1) shows that the bias of the 2SLS estimator arises from the correlation of the fitted value from the first-stage regression for observation i with ϵ_i. Moreover, this bias persists even if the instruments $\mathbf{z}_i$ are uncorrelated with ϵ_i (as valid instruments must be). See Nagar (1959) for the seminal work on the finite-sample bias of k-class estimators, of which 2SLS is a particular case.

3 JIVE

The bias of the 2SLS estimator arises from the correlation between the OLS estimate of the optimal instrument matrix $\mathbf{z}_i\widehat{\pi}$ and the residual ϵ_i. Thus what is needed is an alternative estimator of $\mathbf{z}_i\pi$ that does not suffer from such correlation.

3.1 UJIVE1

Angrist, Imbens, and Krueger (1999) and Blomquist and Dahlberg (1999), building on the work of Phillips and Hale (1977), suggest using all observations except observation i to estimate the parameter matrix π and then using this estimate along with $\mathbf{z}_i$ to compute the fitted value of the instrument for observation i. This process is repeated for each $i = 1, \ldots, N$. That is, let

$$\widehat{\pi}_{-i} = (\mathbf{Z}'_{-i}\mathbf{Z}_{-i})^{-1}\mathbf{Z}'_{-i}\mathbf{X}_{-i} \tag{2}$$

where $\mathbf{Z}_{-i}$ denotes the $(N-1)\times K$ matrix consisting of all rows of $\mathbf{Z}$ except the ith row and similarly for $\mathbf{X}_{-i}$. The ith row of the optimal instrument matrix is estimated by $\mathbf{z}_i\widehat{\pi}_{-i}$. Notice that

$$\begin{aligned} E\left(\widehat{\pi}'_{-i}\mathbf{z}'_i\epsilon_i\right) &= E\left\{\left(\mathbf{X}'_{-i}\mathbf{Z}_{-i}(\mathbf{Z}'_{-i}\mathbf{Z}_{-i})^{-1}\mathbf{z}'_i\epsilon_i\right)|\mathbf{Z}\right\} \\ &= E\left\{E\left(\mathbf{X}'_{-i}\epsilon_i|\mathbf{Z}\right)\mathbf{Z}_{-i}(\mathbf{Z}'_{-i}\mathbf{Z}_{-i})^{-1}\mathbf{z}_i\right\} \\ &= \mathbf{0} \end{aligned}$$

because observations are assumed to be independent. Therefore, the estimator we will call UJIVE1 defined by

$$\widehat{\beta}_{\text{UJIVE1}} = (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'\mathbf{y}$$

where the ith row of $\widehat{\mathbf{X}}$ is defined as $\mathbf{z}_i\widehat{\pi}_{-i}$, does not suffer from the finite-sample bias of 2SLS. Moreover, since $\widehat{\pi}_{-i}$ is a consistent estimator of π, $\widehat{\beta}_{\text{UJIVE1}}$ is a consistent estimator of β. The variance of $\widehat{\beta}_{\text{UJIVE1}}$ is given by

$$\text{Var}\left(\widehat{\beta}_{\text{UJIVE1}}\right) = (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'\epsilon\epsilon'\widehat{\mathbf{X}}(\mathbf{X}'\widehat{\mathbf{X}})^{-1}$$

If we assume that ϵ_i is homoskedastic, then an estimator of the covariance matrix is simply

$$\widehat{\text{Var}}\left(\widehat{\beta}_{\text{UJIVE1}}\right) = \widehat{\sigma^2}(\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'\widehat{\mathbf{X}}(\mathbf{X}'\widehat{\mathbf{X}})^{-1}$$

where

$$\widehat{\sigma^2} = \frac{1}{N}\sum_i\left(y_i - \mathbf{x}_i\widehat{\beta}_{\text{UJIVE1}}\right)^2$$

A heteroskedasticity-robust estimator is[1]

$$\widehat{\text{Var}}\left(\widehat{\beta}_{\text{UJIVE1}}\right) = (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\sum_i\widehat{\epsilon_i}^2\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i{}'(\mathbf{X}'\widehat{\mathbf{X}})^{-1}$$

where

$$\widehat{\epsilon_i}^2 = \left(y_i - \mathbf{x}_i\widehat{\beta}_{\text{UJIVE1}}\right)^2$$

3.2 UJIVE2

Angrist, Imbens, and Krueger (1999) also proposed adjusting only the $\mathbf{Z}'\mathbf{X}$ component of $\widehat{\pi}$, which we will call UJIVE2. The only difference between UJIVE1 and UJIVE2 is that for UJIVE2 we redefine $\widehat{\pi}_{-i}$ as

$$\widehat{\pi}_{-i} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'_{-i}\mathbf{X}_{-i} \tag{3}$$

1. Chao and Swanson (2004) derive the asymptotic distribution of JIVE with heteroskedastic errors.

The formulas for the estimator and its covariance matrix are otherwise identical to those for UJIVE1.

Whether UJIVE1 or UJIVE2 is in any sense better than the other will be explored using Monte Carlo simulation in the next section.

3.3 JIVE1 and JIVE2

Blomquist and Dahlberg (1999) also proposed the estimator

$$\widehat{\beta}_{\text{JIVE1}} = (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\mathbf{y}$$

using $\widehat{\pi}_{-i}$ as in (2) that has the (modest) advantage of requiring just OLS regression once $\widehat{\mathbf{X}}$ has been computed; we will call this estimator JIVE1. Using (3) instead of (2) for $\widehat{\pi}_{-i}$ results in what we will call the JIVE2 estimator.

Although JIVE1 and JIVE2 are consistent estimators of β, unlike UJIVE1 and UJIVE2, they are not unbiased. Moreover, while apparently Blomquist and Dahlberg (1999) use the covariance matrix estimator

$$\widehat{\text{Var}}\left(\widehat{\beta}_{\text{JIVE1}}\right) = \widehat{\sigma^2}(\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}$$

that is only an approximation to the true covariance matrix even with the assumption of homoskedasticity. In the Stata implementation, we also offer the heteroskedasticity-robust covariance matrix

$$\widehat{\text{Var}}\left(\widehat{\beta}_{\text{JIVE1}}\right) = (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}\sum_i \widehat{\epsilon}_i^{\,2}\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i{}'(\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}$$

4 Stata implementation

4.1 Syntax

`jive` *depvar* [*varlist1*] `(`*varlist2* `=` *varlistiv*`)` [*if*] [*in*] [`,` *options*]

options	description
`ujive1`	use Angrist et al. (unbiased) JIVE1 estimator
`ujive2`	use Angrist et al. (unbiased) JIVE2 estimator
`jive1`	use Blomquist and Dahlberg JIVE1 estimator
`jive2`	use Blomquist and Dahlberg JIVE2 estimator
`robust`	compute heteroskedasticity-consistent standard errors
`level(#)`	set confidence level; default is `level(95)`

`by`, `rolling`, `statsby`, and `xi` are allowed; see [U] **11.1.10 Prefix commands**.

4.2 Options

`ujive1`, the default, requests Angrist, Imbens, and Krueger's (1999) UJIVE1 estimator, which adjusts both the $\mathbf{Z}'\mathbf{Z}$ and $\mathbf{Z}'\mathbf{X}$ terms of the first-stage regression.

`ujive2` requests Angrist, Imbens, and Krueger's (1999) UJIVE2 estimator, which adjusts only the $\mathbf{Z}'\mathbf{Z}$ term of the first-stage regression.

`jive1` requests Blomquist and Dahlberg's (1999) JIVE1 estimator, which adjusts both the $\mathbf{Z}'\mathbf{Z}$ and $\mathbf{Z}'\mathbf{X}$ terms of the first-stage regression and uses OLS regression in the second stage.

`jive2` requests Blomquist and Dahlberg's (1999) JIVE2 estimator, which adjusts only the $\mathbf{Z}'\mathbf{Z}$ term of the first-stage regression and uses OLS regression in the second stage.

`robust` requests that the Huber/White/sandwich heteroskedasticity-consistent covariance matrix be used in place of the traditional calculation. See [U] **20.14 Obtaining robust variance estimates**.

`level(#)` sets the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`. See [U] **20.6 Specifying the width of confidence levels**.

4.3 Example

The basic syntax of `jive` is the same as that for `ivreg`; see [R] **ivreg**. To fit a regression of `rent` on `pcturban` and `hsngval` using the UJIVE2 estimator, treating `hsngval` as endogenous and using `faminc` and region dummies as instruments, we type

```
. use http://www.stata-press.com/data/r9/hsng2
(1980 Census housing data)
. jive rent pcturban (hsngval = faminc reg2-reg4), ujive2
Jackknife instrumental variables regression (UJIVE2)
First-stage summary                                Number of obs   =        50
------------------------------------------         F(   2,    47)  =     34.99
F(   4,    44)  =    13.30                         Prob > F        =    0.0000
Prob > F        =   0.0000                         R-squared       =    0.6638
R-squared       =   0.6908                         Adj R-squared   =    0.6495
                                                   Root MSE        =   20.9304

------------------------------------------------------------------------------
        rent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     hsngval |   .0017197   .0003812     4.51   0.000     .0009529    .0024865
    pcturban |   .4020523   .3134261     1.28   0.206    -.2284796    1.032584
       _cons |   124.4641    14.4686     8.60   0.000     95.35705    153.5712
------------------------------------------------------------------------------
Instrumented:  hsngval
Instruments:   pcturban faminc reg2 reg3 reg4
------------------------------------------------------------------------------
```

See the left-hand column in the output header. When the model contains one right-hand-side endogenous variable, `jive` lists the first-stage regression's F statistic and

R^2. Stock, Wright, and Yogo (2002) indicate that the first-stage F statistic typically must exceed 10 for inference based on the 2SLS estimator to be reliable. That paper also includes an alternative to the F statistic to use when the number of endogenous regressors is greater than one, though the alternative is not implemented in `jive`. The first-stage R^2 is also often used to gauge the validity of instruments. Although there is no universally accepted test for or against weak instruments, the first-stage F statistic and R^2 both have the benefit of intuitive appeal.

4.4 Saved results

`jive` saves the following in `e()`:

Scalars			
`e(N)`	number of observations	`e(F1)`	first-stage F statistic
`e(rmse)`	root mean squared error	`e(df_m_F1)`	first-stage model degrees of freedom
`e(F)`	model F statistic		
`e(df_m)`	model degrees of freedom	`e(df_r_F1)`	first-stage residual degrees of freedom
`e(df_r)`	residual degrees of freedom		
`e(r2)`	R^2	`e(r2_1)`	first-stage R^2
`e(r2_a)`	adjusted R^2		
Macros			
`e(model)`	`UJIVE1`, `UJIVE2`, `JIVE1`, or `JIVE2`	`e(instd)`	instrumented variables
		`e(insts)`	instruments
`e(title)`	title in estimation output	`e(properties)`	`b V`
`e(depvar)`	name of dependent variable		
Matrices			
`e(b)`	coefficient vector	`e(V)`	variance–covariance matrix of the estimator
Functions			
`e(sample)`	marks estimation sample		

5 Monte Carlo simulation

5.1 Model 1

Our first simulation model is identical to the first one analyzed by Angrist, Imbens, and Krueger (1999):

$$
\begin{aligned}
y_i &= 0 + 1 \cdot x_i + \epsilon_i \\
x_i &= 0 + 0.3 \cdot z_{i1} + 0 \cdot z_{i2} + \nu_i
\end{aligned}
$$

where z_{i1} and z_{i2} are independently and identically distributed standard normal and

$$
\begin{pmatrix} \epsilon_i \\ \nu_i \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix} \right\}
$$

The population coefficient on z_{i2} in the equation for x_i is zero. In this and all subsequent models, I used a sample size of $N = 100$ and 5,000 simulations.

Table 1: Results from model 1 simulations

Estimator	Percentile					95% CI	
	10th	25th	50th	75th	90th	Standard	Robust
UJIVE1	0.612	0.799	0.947	1.068	1.158	0.964	0.957
UJIVE2	0.608	0.798	0.946	1.068	1.158	0.964	0.958
JIVE1	0.525	0.710	0.866	0.992	1.085	0.965	0.946
JIVE2	0.545	0.737	0.901	1.032	1.129	0.963	0.950
2SLS	0.766	0.903	1.021	1.118	1.200	0.939	0.931
LIML	0.730	0.874	0.995	1.098	1.183	0.948	0.945

Table 1 shows the 10th, 25th, 50th, 75th, and 90th percentiles of the observed distribution, as well as the coverage rate of the 95% confidence interval constructed using the standard and heteroskedasticity-consistent covariance matrices.

Because Stata's official `ivreg` command does not implement the LIML estimator, we used the `ivreg2` command of Baum, Schaffer, and Stillman (2003, 2004, 2005). The `condivreg` command of Moreira and Poi (2003) and Mikusheva and Poi (2006) can also perform LIML estimation when the model contains one endogenous regressor.

Here the 2SLS estimator performed well, since we had only two instruments (one relevant, one irrelevant) and one endogenous variable. The confidence intervals had coverage rates near their nominal sizes. All four jackknife estimators exhibited right-skewed distributions (more mass to the left of the true value), and the distributions were more dispersed than the ones for 2SLS and limited information maximum likelihood (LIML). The JIVEs produced relatively large standard errors, as the confidence intervals performed well despite the skewed distributions. The UJIVEs performed slightly better than the JIVEs in terms of the amount of skew in the distributions. Which definition of $\widehat{\pi}_{-i}$ was used had a modest impact on the JIVEs but not the UJIVEs. Our results are similar to those reported by Angrist, Imbens, and Krueger (1999).

5.2 Model 2

Our second model is the same as the second model of Angrist, Imbens, and Krueger (1999) and is similar to model 1 except that there are 19 irrelevant instruments instead of just one:

$$x_i = 0 + 0.3 \cdot z_{i1} + \sum_{j=2}^{j=20} 0 \cdot z_{ij} + \nu_i$$

Table 2: Results from model 2 simulations

Estimator	Percentile					Coverage rate	
	10th	25th	50th	75th	90th	Standard	Robust
UJIVE1	0.393	0.720	0.948	1.109	1.220	0.948	0.939
UJIVE2	0.395	0.718	0.946	1.106	1.222	0.947	0.940
JIVE1	0.142	0.334	0.521	0.687	0.807	0.231	0.239
JIVE2	0.182	0.424	0.663	0.872	1.029	0.652	0.635
2SLS	1.137	1.205	1.278	1.347	1.408	0.318	0.319
LIML	0.702	0.854	0.996	1.113	1.203	0.928	0.953

Table 2 shows that the 2SLS estimator is clearly biased upward, and the coverage rates of the confidence intervals are low. This finding is not surprising given the analytical result of section 1. The JIVEs are severely right skewed, and their confidence intervals perform poorly as well. The UJIVEs exhibit right skew, but the median estimate is off by only about 5%. Moreover, the confidence intervals have excellent coverage rates. The LIML estimator also has good coverage, and it does not suffer from nearly as much right skew as the UJIVEs.

5.3 Model 3

Model 3 is similar to model 1, except that the error term in the equation for y_i is conditionally heteroskedastic:

$$y_i = 0 + 1 \cdot x_i + z_{1i}^2 \epsilon_i$$

Here we are concerned mainly with testing the `robust` option of our `jive` command.

Table 3: Results from model 3 simulations

Estimator	Percentile					Coverage rate	
	10th	25th	50th	75th	90th	Standard	Robust
UJIVE1	−0.200	0.418	0.906	1.303	1.644	0.697	0.942
UJIVE2	−0.164	0.430	0.907	1.293	1.625	0.712	0.943
JIVE1	−0.174	0.371	0.828	1.206	1.540	0.658	0.946
JIVE2	−0.145	0.399	0.858	1.237	1.578	0.679	0.944
2SLS	0.125	0.598	1.017	1.376	1.690	0.676	0.930
LIML	0.057	0.556	0.990	1.366	1.690	0.667	0.931

Table 3 shows that, just as with model 1, the 2SLS and LIML estimators provide good results in that they are approximately median unbiased. As expected, the standard 95% confidence interval has relatively poor coverage, but the heteroskedasticity-robust variant has coverage near the nominal size of the confidence interval. The UJIVEs are again right skewed, and the median estimates are roughly 9% too small. However, the robust covariance matrices do result in confidence intervals with good coverage. The JIVEs are slightly less dispersed than the UJIVEs, but they show more negative bias. Nevertheless, the confidence intervals based on the robust covariance matrix perform well.

5.4 Model 4

The fourth model we consider contains nonlinearities and heteroskedasticity in the first-stage equation for x_i; the full model is

$$
\begin{aligned}
y_i &= 0 + 1 \cdot x_i + \epsilon_i \\
x_i &= 0 + 0.3 \cdot z_{i1} + \sum_{j=2}^{j=20} 0 \cdot z_{ij} + 0.3 \sum_{j=2}^{j=20} z_{ij}^2 + \nu_i \sum_{j=2}^{j=20} z_{ij}^2/19
\end{aligned}
$$

The error structure is

$$
\begin{pmatrix} \epsilon_i \\ \nu_i \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{pmatrix} \right\}
$$

Table 4: Results from model 4 simulations

Estimator	Percentile					Coverage rate	
	10th	25th	50th	75th	90th	Standard	Robust
UJIVE1	0.223	0.865	1.155	1.425	1.986	0.969	0.966
UJIVE2	0.621	0.894	1.049	1.176	1.350	0.954	0.944
JIVE1	−0.508	−0.248	−0.001	0.210	0.376	1.000	1.000
JIVE2	−1.084	−0.816	−0.521	−0.247	−0.012	1.000	1.000
2SLS	1.032	1.091	1.151	1.211	1.264	0.601	0.585
LIML	0.311	0.820	1.082	1.301	1.736	0.813	0.944

As table 4 shows, the nonlinear relationship between x_i and the z_i's is apparently causing the 2SLS estimates to exhibit an upward bias; and, because the heteroskedasticity affects x_i and not y_i, the use of a robust covariance matrix does not improve the coverage rate of the confidence interval. The UJIVE1 estimates are widely dispersed and slightly biased upward, though the confidence intervals do provide good coverage.

The UJIVE2 results show much less dispersion than the UJIVE1 ones; Angrist, Imbens, and Krueger (1999) report that only in nonlinear data-generating processes like this one have they seen such a disparity between UJIVE1 and UJIVE2. The JIVE1 and JIVE2 estimates are severely biased downward. The confidence interval coverage rates are virtually 100%, because the standard errors were so large. The covariance matrix used for JIVE1 and JIVE2 is only an approximation to the true covariance matrix, and that approximation clearly fails when the relationship between x_i and the z_i's is nonlinear. The LIML estimator performs reasonably well, though UJIVE2 clearly does better here.

5.5 Model 5

The average of the first-stage R^2s in the simulations of model 1 was 0.2725, with 90% of the simulations' R^2s lying between 0.1525 and 0.4013. Moreover, the first-stage F statistic averaged 18.17, with 90% of the values lying between 8.73 and 32.51. From the assertion of Stock, Wright, and Yogo (2002) that the first-stage F statistic should be at least 10 for 2SLS inference to be reliable, there does not appear to be a weak-instrument problem, as our Monte Carlo results confirmed.

Here we use the setup of model 2, except that we reduce the correlation between x_i and the instruments:

$$x_i = 0 + 0.03 \cdot z_{i1} + \sum_{j=2}^{j=20} 0 \cdot z_{ij} + \nu_i$$

This setup allows us to see whether UJIVE and JIVE provide any safeguards against the common use of many weak instruments. With this model, the first-stage F statistic averaged only 0.98, with 99% of the 5,000 simulations yielding $F < 2.11$. The average first-stage R^2 was 0.1987, however.

Table 5: Results from model 5 simulations

	Percentile					Coverage rate	
Estimator	10th	25th	50th	75th	90th	Standard	Robust
UJIVE1	0.424	1.372	1.800	2.209	3.011	0.734	0.717
UJIVE2	0.497	1.383	1.807	2.224	3.082	0.731	0.718
JIVE1	−0.903	−0.500	−0.103	0.236	0.473	0.084	0.084
JIVE2	−1.165	−0.650	−0.137	0.296	0.604	0.320	0.305
2SLS	1.606	1.691	1.784	1.879	1.959	0.004	0.004
LIML	−0.125	1.097	1.727	2.342	3.659	0.541	0.728

Table 5 shows that the 2SLS estimator shows a substantial upward bias, and the nominal 95% confidence interval has a coverage rate of virtually zero. The UJIVEs show a median bias approximately equal to that of the 2SLS estimator. However, the distribution is much more dispersed; although the confidence intervals do not have coverage rates equal to their nominal sizes, they do perform notably better than all the other estimators except LIML. The JIVEs exhibit extreme negative bias, and this model provides no compelling reason to prefer them over 2SLS. The LIML estimator is slightly less biased in its median than the UJIVEs, but its distribution is more dispersed.

6 Conclusion

The 2SLS estimator is known to be biased, and instruments that are only weakly correlated with the endogenous regressor compound the problem. In this article, I have presented a new Stata command for fitting models by using the JIVE, and I have provided Monte Carlo evidence showing that two variants of the estimator, UJIVE1 and UJIVE2, yield good results even when the usual 2SLS estimator does not. My simulations also show LIML to be a good alternative to 2SLS. Two other variants, JIVE1 and JIVE2, appear to offer no compelling advantages over 2SLS based on the simulations, though they are implemented in the Stata command `jive` for those who wish to try other models.

7 References

Angrist, J. D., G. W. Imbens, and A. B. Krueger. 1999. Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14: 57–67.

Angrist, J. D., and A. B. Krueger. 1991. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106: 979–1014.

Baum, C. F., M. E. Schaffer, and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3: 1–31.

———. 2004. Software updates: Instrumental variables and GMM: Estimation and testing. *Stata Journal* 4: 224.

———. 2005. Software updates: Instrumental variables and GMM: Estimation and testing. *Stata Journal* 5: 607.

Blomquist, S., and M. Dahlberg. 1999. Small sample properties of LIML and jackknife IV estimators: Experiments with weak instruments. *Journal of Applied Econometrics* 14: 69–88.

Bound, J., D. Jaeger, and R. Baker. 1995. Problems with instrumental variable estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association* 90: 443–450.

Chao, J. C., and N. R. Swanson. 2004. Asymptotic distribution of JIVE in a heteroskedastic IV regression with many weak instruments. http://www.econ.vt.edu/Seminars/Seminarpapers/Chao_B_jive2.pdf.

Mikusheva, A., and B. P. Poi. 2006. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* 6. 51–63.

Moreira, M. J., and B. P. Poi. 2003. Implementing tests with correct size in the simultaneous equations model. *Stata Journal* 3: 57–70.

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. R., and R. Startz. 1990. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business* 63: 5125–5140.

Phillips, G. D. A., and C. Hale. 1977. The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review* 18: 219–228.

Stock, J. H., J. H. Wright, and M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20: 518–529.

About the author

Brian Poi is a senior economist at StataCorp.