The Stata Journal

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

The Stata Journal (2007)
7, Number 4, pp. 465–506

Enhanced routines for instrumental variables/generalized method of moments estimation and testing

Christopher F. Baum
Department of Economics
Boston College
Chestnut Hill, MA
baum@bc.edu

Mark E. Schaffer
Economics Department
Heriot–Watt University
Edinburgh, UK
m.e.schaffer@hw.ac.uk

Steven Stillman
Motu Economic Public Policy Research
Wellington, New Zealand
stillman@motu.org.nz

Abstract. We extend our 2003 paper on instrumental variables and generalized method of moments estimation, and we test and describe enhanced routines that address heteroskedasticity- and autocorrelation-consistent standard errors, weak instruments, limited-information maximum likelihood and k-class estimation, tests for endogeneity and Ramsey's regression specification-error test, and autocorrelation tests for instrumental variable estimates and panel-data instrumental variable estimates.

Keywords: st0030_3, ivactest, ivendog, ivhettest, ivreg2, ivreset, overid, ranktest, instrumental variables, weak instruments, GMM, endogeneity, heteroskedasticity, serial correlation, HAC standard errors, LIML, CUE, overidentifying restrictions, Frisch–Waugh–Lovell theorem, RESET, Cumby–Huizinga test

1 Introduction

In Baum, Schaffer, and Stillman (2003), we discussed instrumental variables (IV) estimators in the context of generalized method of moments (GMM) estimation and presented Stata routines for estimation and testing consisting of the `ivreg2` suite. Since that time, those routines have been considerably enhanced and more routines have been added to the suite. This paper presents the analytical underpinnings of both basic IV/GMM estimation and these enhancements and describes the enhanced routines. Some of these features are now also available in Stata 10's `ivregress`, whereas others are not.

The additions include the following:

- Estimation and testing that is robust to and efficient in the presence of arbitrary serial correlation.

- A range of test statistics that allow the user to address the problems of underidentification or weak identification, including statistics that are robust in the presence of heteroskedasticity, autocorrelation, or clustering.

- Three additional IV/GMM estimators: the GMM continuously updated estimator (CUE) of Hansen, Heaton, and Yaron (1996); limited-information maximum likelihood (LIML); and k-class estimators.

- A more intuitive syntax for GMM estimation: the `gmm2s` option requests the two-step feasible efficient GMM (EGMM) estimator, which reduces to standard IV/2SLS if no robust covariance matrix estimator is also requested. The `cue` option requests the continuously updated GMM estimator, which reduces to standard LIML if no robust covariance matrix estimator is also requested.

- A more intuitive syntax for a "GMM distance" or C test of the endogeneity of regressors.

- An option that allows the user to "partial out" regressors: something that is particularly useful when the user has a rank-deficient estimate of the covariance matrix of orthogonality conditions (common with the `cluster()` option and singleton dummy variables).

- Several advanced options, including options that will speed up estimation using `ivreg2` by suppressing the calculation of various checks and statistics.

- A version of Ramsey's regression specification-error test (RESET), `ivreset`, that (unlike official Stata's `ovtest`) is appropriate for use in an IV context.

- A test for autocorrelation in time-series errors, `ivactest`, that (unlike official Stata's `estat bgodfrey`) is appropriate for use in an IV context.

We review the definitions of the method of IV and IV/GMM in the next section to set the stage. The following sections of the paper discuss each of these enhancements in turn. The last two sections provide a summary of `ivreg2` estimation options and syntax diagrams for all programs in the extended `ivreg2` suite.

2 IV and GMM estimation

GMM was introduced in Hansen (1982). It is now a mainstay of both econometric practice and econometrics textbooks. We limit our exposition here to the linear case, which is what `ivreg2` handles. The exposition here draws on Hayashi (2000). For more details and references, see also Baum, Schaffer, and Stillman (2003) and Baum (2006, chap. 8).

2.1 Setup

The equation to be estimated is, in matrix notation,

$$y = X\beta + u \tag{1}$$

with typical row

$$y_i = X_i\beta + u_i$$

The matrix of regressors X is $n \times K$, where n is the number of observations. Some of the regressors are endogenous, so that $E(X_i u_i) \neq 0$.

We partition the set of regressors into $[X_1 \;\; X_2]$, with the K_1 regressors X_1 assumed under the null to be endogenous and the $K_2 \equiv (K - K_1)$ remaining regressors X_2 assumed exogenous, giving us

$$y = [X_1 \;\; X_2][\beta_1' \;\; \beta_2']' + u$$

The set of IV is Z and is $n \times L$. This is the full set of variables that are assumed to be exogenous, i.e., $E(Z_i u_i) = 0$. We partition the instruments into $[Z_1 \; Z_2]$, where the L_1 instruments Z_1 are excluded instruments and the remaining $L_2 \equiv (L - L_1)$ instruments $Z_2 \equiv X_2$ are the included instruments/exogenous regressors:

$$\text{Regressors } X = [X_1 \;\; X_2] = [X_1 \;\; Z_2] = [\text{Endogenous} \quad \text{Exogenous}]$$

$$\text{Instruments } Z = [Z_1 \; Z_2] = [\text{Excluded} \quad \text{Included}]$$

The order condition for identification of the equation is $L \geq K$ implying there must be at least as many excluded instruments (L_1) as there are endogenous regressors (K_1) as Z_2 is common to both lists. If $L = K$, the equation is said to be *exactly identified* by the order condition; if $L > K$, the equation is *overidentified*. The order condition is necessary but not sufficient for identification; see section 7 for a full discussion.

2.2 GMM

The assumption that the instruments Z are exogenous can be expressed as $E(Z_i u_i) = 0$. We are considering linear GMM only, and here the L instruments give us a set of L moments

$$g_i(\beta) = Z_i' u_i = Z_i'(y_i - X_i\beta)$$

where g_i is $L \times 1$. The exogeneity of the instruments means that there are L moment conditions, or orthogonality conditions, that will be satisfied at the true value of β:

$$E[g_i(\beta)] = 0$$

Each of the L moment equations corresponds to a sample moment. For some given estimator $\widehat{\beta}$, we can write these L sample moments as

$$\overline{g}(\widehat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\widehat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} Z_i'(y_i - X_i\widehat{\beta}) = \frac{1}{n}Z'\widehat{u}$$

The intuition behind GMM is to choose an estimator for β that brings $\overline{g}(\widehat{\beta})$ as close to zero as possible. If the equation to be estimated is exactly identified, so that $L = K$, then we have as many equations—the L moment conditions—as we do unknowns: the K coefficients in $\widehat{\beta}$. Here it is possible to find a $\widehat{\beta}$ that solves $\overline{g}(\widehat{\beta}) = 0$, and this GMM estimator is in fact a special case of the IV estimator as we discuss below.

If the equation is overidentified, however, so that $L > K$, then we have more equations than we do unknowns. Generally, it will not be possible to find a $\widehat{\beta}$ that will set all L sample moment conditions exactly to zero. Here we take an $L \times L$ weighting matrix W and use it to construct a quadratic form in the moment conditions. This gives us the GMM objective function:

$$J(\widehat{\beta}) = n\overline{g}(\widehat{\beta})'W\overline{g}(\widehat{\beta})$$

A GMM estimator for β is the $\widehat{\beta}$ that minimizes $J(\widehat{\beta})$:

$$\widehat{\beta}_{\mathrm{GMM}} \equiv \arg\min_{\widehat{\beta}} J(\widehat{\beta}) = n\overline{g}(\widehat{\beta})'W\overline{g}(\widehat{\beta})$$

Linearly, we are considering, deriving, and solving the K first-order conditions $\{\partial J(\widehat{\beta})\}/(\partial\widehat{\beta}) = 0$ (treating W as a matrix of constants), which yields the GMM estimator:[1]

$$\widehat{\beta}_{\mathrm{GMM}} = (X'ZWZ'X)^{-1}X'ZWZ'y \tag{2}$$

The GMM estimator is consistent for any symmetric positive-definite weighting matrix W, and thus there are as many GMM estimators as there are choices of weighting matrix W. Efficiency is not guaranteed for an arbitrary W, so we refer to the estimator defined in (2) as the possibly *inefficient* GMM estimator.

1. The results of the minimization, and hence the GMM estimator, will be the same for weighting matrices that differ by a constant of proportionality.

We are particularly interested in *efficient* GMM estimators with minimum asymptotic variance. Moreover, for any GMM estimator to be useful, we must be able to conduct inference, and for that we need estimates of the variance of the estimator. Both require estimates of the covariance matrix of orthogonality conditions, a key concept in GMM estimation.

2.3 Inference, efficiency, and the covariance matrix of orthogonality conditions

Denoted by S, the asymptotic covariance matrix of the moment conditions g

$$S = \mathrm{AVar}\{\overline{g}(\beta)\} = \lim_{n\to\infty} \frac{1}{n} E(Z'uu'Z)$$

where S is an $L \times L$ matrix and $\overline{g}(\beta) = (1/n)Z'u$. That is, S is the variance of the limiting distribution of $\sqrt{n}\ \overline{g}$ (Hayashi 2000, 203).

The asymptotic distribution of the possibly inefficient GMM (IGMM) estimator can be written as follows. Let $Q_{XZ} \equiv E(X_i'Z_i)$. The asymptotic variance of the IGMM estimator defined by an arbitrary weighting matrix W is given by

$$V(\widehat{\beta}_{\mathrm{GMM}}) = (Q'_{XZ}WQ_{XZ})^{-1}(Q'_{XZ}WSWQ_{XZ})(Q'_{XZ}WQ_{XZ})^{-1} \tag{3}$$

Under standard assumptions (see Hayashi 2000, 202–203, 209) the IGMM estimator is $\sqrt{n}$-consistent; that is,

$$\sqrt{n}\ (\widehat{\beta}_{\mathrm{GMM}} - \beta) \to N\{0, V(\widehat{\beta}_{\mathrm{GMM}})\}$$

where $\to$ denotes convergence in distribution.

Strictly speaking, therefore, we should perform hypothesis tests on $\sqrt{n}\ \widehat{\beta}_{\mathrm{GMM}}$ by using (3) for the variance–covariance matrix. Standard practice, however, is to transform the variance–covariance matrix (3) rather than the coefficient vector (2). This is done by normalizing $V(\widehat{\beta}_{\mathrm{GMM}})$ by $1/n$, so that the variance–covariance matrix reported by statistical packages such as Stata is in fact

$$V\left(\frac{1}{\sqrt{n}}\ \widehat{\beta}_{\mathrm{GMM}}\right) = \frac{1}{n}(Q'_{XZ}WQ_{XZ})^{-1}(Q'_{XZ}WSWQ_{XZ})(Q'_{XZ}WQ_{XZ})^{-1} \tag{4}$$

The EGMM estimator makes use of an estimator with an optimal weighting matrix W, which minimizes the asymptotic variance of the estimator. This is achieved by choosing $W = S^{-1}$. If we substitute this into (2) and (4), we obtain the EGMM estimator

$$\widehat{\beta}_{\mathrm{EGMM}} = (X'ZS^{-1}Z'X)^{-1}X'ZS^{-1}Z'y \tag{5}$$

with asymptotic variance

$$V(\widehat{\beta}_{\text{EGMM}}) = (Q'_{XZ} S^{-1} Q_{XZ})^{-1}$$

Similarly,

$$\sqrt{n}\ (\widehat{\beta}_{\text{EGMM}} - \beta) \to N\{0, V(\widehat{\beta}_{\text{EGMM}})\}$$

and we perform inference on $\sqrt{n}\ \widehat{\beta}_{\text{EGMM}}$ by using

$$V\left(\frac{1}{\sqrt{n}}\ \widehat{\beta}_{\text{EGMM}}\right) = \frac{1}{n}(Q'_{XZ} S^{-1} Q_{XZ})^{-1} \tag{6}$$

as the variance–covariance matrix for $\widehat{\beta}_{\text{EGMM}}$.

Obtaining an estimate of Q_{XZ} is straightforward: we simply use the sample analog

$$\frac{1}{n}\sum_{i=1}^{n} X'_i Z_i = \frac{1}{n} X'Z$$

If we have an estimate of S, therefore, we can conduct asymptotically correct inference for any GMM estimator, efficient or inefficient.

An estimate of S also makes the EGMM estimator a feasible estimator. In two-step feasible EGMM estimation an estimate of S is obtained in the first step, and we calculate the estimator and its asymptotic variance by using (5) and (6) in the second step.

2.4 Estimating the covariance matrix of orthogonality conditions

The first-step estimation of the matrix S requires the residuals of a consistent GMM estimator $\widetilde{\beta}$. Efficiency is not required in the first step of two-step GMM estimation, which simplifies the task considerably. But to obtain an estimate of S, we must make some further assumptions.

We illustrate this by using the case of independent but possibly heteroskedastic disturbances. If the errors are independent, $E(g_i g'_j) = 0$ for $i \neq j$, and so

$$S = \text{AVar}(\bar{g}) = E(g_i g'_i) = E(u_i^2 Z'_i Z_i)$$

This matrix can be consistently estimated by an Eicker–Huber–White robust covariance estimator

$$\widehat{S} = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2 Z_i' Z_i = \frac{1}{n}(Z'\widehat{\Omega}Z) \tag{7}$$

where $\widehat{\Omega}$ is the diagonal matrix of squared residuals $\widehat{u}_i^2$ from $\widetilde{\beta}$, the consistent but not necessarily efficient first-step GMM estimator. In the `ivreg2` implementation of two-step EGMM, the first-step estimator is $\widehat{\beta}_{\mathrm{IV}}$, the IV estimator.

The resulting estimate $\widehat{S}$ can be used to conduct consistent inference for the first-step estimator using (3), or it can be used to obtain and conduct inference for the EGMM estimator using (5) and (6).

In the next section, we discuss how the two-step GMM estimator can be applied when the errors are serially correlated.

2.5 Using ivreg2 for GMM estimation

The `ivreg2` command is included in the electronic supplement to this issue. The latest version of `ivreg2` can also be downloaded from the SSC archive with the command `ssc describe ivreg2`. We summarize the command's syntax and options in sections 11 and 12, respectively. The commands below illustrate how to use `ivreg2` to obtain the coefficient and variance–covariance estimators discussed above. The example uses the dataset provided in Wooldridge (2003).

The first command requests a standard IV/2SLS estimator and a variance–covariance matrix, which assumes conditionally homoskedastic and independent errors. In this case, IV/2SLS is the EGMM estimator. The second requests the IV/2SLS estimator and a variance–covariance estimator that is robust to heteroskedasticity based on an estimate of $\widehat{S}$ as in (7); here IV/2SLS is an IGMM estimator. The third command requests the two-step feasible EGMM estimator and corresponding variance–covariance matrix. $\widehat{S}$ is again based on (7). The fourth command is equivalent to the first, illustrating that the two-step EGMM estimator reduces to two-stage least squares when the disturbance is assumed to be independently and identically distributed (i.i.d.) and S can be consistently estimated by a classical nonrobust covariance matrix estimator.

1. `ivreg2 lwage exper expersq (educ=age kidslt6 kidsge6)`
2. `ivreg2 lwage exper expersq (educ=age kidslt6 kidsge6), robust`
3. `ivreg2 lwage exper expersq (educ=age kidslt6 kidsge6), gmm2s robust`
4. `ivreg2 lwage exper expersq (educ=age kidslt6 kidsge6), gmm2s`

3 GMM and HAC standard errors

Equation (7) illustrates how the asymptotic covariance matrix of the GMM estimator could be derived in the presence of conditional heteroskedasticity. We now further extend the estimator to handle the case of nonindependent errors in a time-series context. We correspondingly change our notation so that observations are indexed by t and s rather than i. In the presence of serial correlation, $E(g_t g_s') \neq 0, t \neq s$. To derive consistent estimates of S, we define $\Gamma_j = E(g_t g_{t-j}')$ as the autocovariance matrix for lag j. We may then write the long-run covariance matrix

$$S = \text{AVar}(\bar{g}) = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j') \tag{8}$$

which may be seen as a generalization of (7), with $\Gamma_0 = E(g_i g_i')$ and

$$\Gamma_j = E(g_t g_{t-j}'),\ j = \pm 1, \pm 2, \dots$$

As g_t is defined as the product of Z_t and u_t, the autocovariance matrices may be expressed as

$$\Gamma_j = E(u_t u_{t-j} Z_t' Z_{t-j})$$

As usual, we replace the u_t, u_{t-j} by consistent residuals from first-stage estimation to compute the sample autocovariance matrices $\widehat{\Gamma}_j$, defined as

$$\widehat{\Gamma}_j = \frac{1}{n} \sum_{t=1}^{n-j} \widehat{g}_t \widehat{g}_{t-j} = \frac{1}{n} \sum_{t=1}^{n-j} Z_t' \widehat{u}_t \widehat{u}_{t-j} Z_{t-j}$$

We obviously do not have an infinite number of sample autocovariances to insert into the infinite sum in (8). Less obviously, we also cannot simply insert all the autocovariances from 1 to n, because this would imply that the number of sample orthogonality conditions $\widehat{g}_i$ is going to infinity with the sample size, which precludes obtaining a consistent estimate of S.[2] The autocovariances must converge to zero asymptotically as n increases.

The usual way this is handled in practice is for the summation to be truncated at a specified lag q. Thus the S matrix can be estimated by

$$\widehat{S} = \widehat{\Gamma}_0 + \sum_{j=1}^{q} \kappa \left(\frac{j}{q_n} \right) (\widehat{\Gamma}_j + \widehat{\Gamma}_j')$$

2. Although a consistent estimate cannot be obtained with bandwidth equal to sample size, Hall (2005, 305–310) points out that it is possible to develop an asymptotic framework providing inference about the parameters.

where u_t, u_{t-j} are replaced by consistent estimates from first-stage estimation. The *kernel* function, $\kappa(j/q_n)$, applies appropriate weights to the terms of the summation, with q_n defined as the *bandwidth* of the kernel (possibly as a function of n).[3] In many kernels, consistency is obtained by having the weight fall to zero after a certain number of lags.

The best-known approach to this problem in econometrics is that of Newey and West (1987b), which generates $\widehat{S}$ by using the Bartlett kernel function and a user-specified value of q. For the Bartlett kernel, $\kappa(\cdot) = (1 - j/q_n)$ if $j \leq q_n - 1$, 0 otherwise. These estimates are said to be heteroskedasticity- and autocorrelation-consistent (HAC), as they incorporate the standard sandwich formula (7) in computing Γ_0.

HAC estimates can be calculated by using `ivreg2` with the `robust` and `bw()` options with the kernel function's bandwidth (the `bw()` option) set to q.[4] The bandwidth may also be chosen optimally by specifying `bw(auto)` by using the automatic bandwidth selection criterion of Newey and West (1994).[5,6] By default, `ivreg2` uses the Bartlett kernel function.[7] If the equation contains endogenous regressors, these options will cause the IV estimates to be HAC. If the equation is overidentified and the `robust`, `gmm2s`, and `bw()` options are specified, the resulting GMM estimates will be both HAC and more efficient than those produced by IV.

The Newey–West (Bartlett kernel function) specification is only one of many feasible HAC estimators of the covariance matrix. Andrews (1991) shows that in the class of positive semidefinite kernels, the rate of convergence of $\widehat{S} \rightarrow S$ depends on the choice of kernel and bandwidth. The Bartlett kernel's performance is bettered by those in a subset of this class, including the quadratic spectral kernel. Accordingly, `ivreg2` provides a menu of kernel choices, including (abbreviations in parentheses): quadratic spectral (`qua` or `qs`), truncated (`tru`), Parzen (`par`), Tukey–Hanning (`thann`), Tukey–Hamming (`thamm`), Daniell (`dan`), and Tent (`ten`). For the Bartlett, Parzen, and Tukey–Hanning/Hamming kernels, the number of lags used to construct the kernel estimate equals the bandwidth (`bw()`) minus one.[8] If the kernels above are used with `bw(1)`, no lags are used and `ivreg2` will report the usual Eicker–Huber–White "sandwich" heteroskedasticity-robust variance estimates. Most, but not all, of these kernels guarantee that the estimated $\widehat{S}$ is positive definite and therefore always invertible; the truncated kernel, for example, was proposed in the early literature in this area but is now rarely used because it can generate an noninvertible $\widehat{S}$. For a survey covering various kernel estimators and their properties, see Cushing and McGarvey (1999) and Hall (2005, 75–86).

3. For more detail on this GMM estimator, see Hayashi (2000, 406–417).
4. For the special case of ordinary least squares (OLS), Newey–West standard errors are available from [TS] **newey** with the maximum lag $(q-1)$ specified by `newey`'s `lag()` option.
5. This implementation is identical to that provided by Stata's `ivregress` command; see [R] **ivregress**.
6. Automatic bandwidth selection is only available for the Bartlett, Parzen, and quadratic spectral kernels; see below.
7. A common choice of bandwidth for the Bartlett kernel function is $T^{1/3}$.
8. A common choice of bandwidth for these kernels is $(q-1) \approx T^{1/4}$ (Greene 2008, 643). A value related to the periodicity of the data (4 for quarterly, 12 for monthly, etc.) is often chosen.

Under conditional homoskedasticity the expression for the autocovariance matrix simplifies:

$$\Gamma_j \quad = \quad E(u_t u_{t-j} Z_t' Z_{t-j}) = E(u_t u_{t-j}) E(Z_t' Z_{t-j})$$

and the calculations of the corresponding kernel estimators also simplify; see Hayashi (2000, 413–414). These estimators may perform better than their heteroskedasticity-robust counterparts in finite samples. If the researcher is satisfied with the assumption of homoskedasticity but wants to deal with autocorrelation of unknown form, the researcher should use the AC correction without the H correction for arbitrary heteroskedasticity by omitting the `robust` option. `ivreg2` allows selection of H, AC, or HAC VCEs by combining the `robust`, `bw()`, and `kernel()` options. Thus both `robust` and `bw()` must be specified to calculate a HAC VCE of the Newey–West type, using the default Bartlett kernel. [9]

To illustrate the use of HAC standard errors, we fit a quarterly time-series model relating the change in the U.S. inflation rate (`D.inf`) to the unemployment rate (`UR`) for 1960q3–1999q4. As instruments, we use the second lag of quarterly GDP growth and the lagged values of the Treasury bill rate, the trade-weighted exchange rate, and the Treasury medium-term bond rate.[10] We first estimate the equation with standard IV under the assumption of i.i.d. errors.

```
. use http://fmwww.bc.edu/ec-p/data/stockwatson/macrodat
. generate inf = 100 * log( CPI / L4.CPI )
(4 missing values generated)
. generate ggdp = 100 * log( GDP / L4.GDP )
(10 missing values generated)
```

9. Stata's official `newey` command (see [TS] **newey**) does not allow gaps in time-series data. As there is no difficulty in computing HAC estimates with gaps in a regularly spaced time series, `ivreg2` handles this case properly.
10. These data accompany Stock and Watson (2003).

```
. ivreg2 D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON)
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only
                                                      Number of obs =      158
                                                      F(  1,   156) =    10.16
                                                      Prob > F      =   0.0017
Total (centered) SS     =  60.04747699                Centered R2   =   0.1914
Total (uncentered) SS   =  60.05149156                Uncentered R2 =   0.1915
Residual SS             =  48.55290564                Root MSE      =    .5543

------------------------------------------------------------------------------
       D.inf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          UR |   -.155009   .0483252    -3.21   0.001    -.2497246   -.0602933
       _cons |   .9380705   .2942031     3.19   0.001      .361443    1.514698
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):          58.656
                                                   Chi-sq(4) P-val =    0.0000
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):               22.584
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias    16.85
                                         10% maximal IV relative bias    10.27
                                         20% maximal IV relative bias     6.71
                                         30% maximal IV relative bias     5.34
                                         10% maximal IV size             24.58
                                         15% maximal IV size             13.96
                                         20% maximal IV size             10.26
                                         25% maximal IV size              8.31
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):          5.851
                                                   Chi-sq(3) P-val =    0.1191
------------------------------------------------------------------------------
Instrumented:         UR
Excluded instruments: L2.ggdp L.TBILL L.ER L.TBON
------------------------------------------------------------------------------
```

In these estimates, the negative coefficient on the unemployment rate is consistent with macroeconomic theories of the natural rate. In that context, lowering unemployment below the natural rate will cause an acceleration of price inflation. The Sargan statistic implies that the test of overidentifying restrictions cannot reject its null hypothesis.

An absence of autocorrelation in the error process is unusual in time-series analysis, so we test the equation by using `ivactest`, as discussed in section 10. By using the default value of one lag, we consider whether the error process exhibits AR(1) behavior. The test statistic implies that the errors do not exhibit serial independence:

(*Continued on next page*)

```
. ivactest

Cumby-Huizinga test with H0: errors nonautocorrelated at order 1
Test statistic:  25.909524
Under H0, Chi-sq(1) with p-value:  3.578e-07
```

Given this strong rejection of the null of independence, we reestimate the equation with HAC standard errors, choosing a bandwidth (`bw()`) of 5 (roughly $T^{1/3}$) and the `robust` option. By default, the Bartlett kernel is used, so that these are Newey–West two-step EGMM estimates.

```
. ivreg2 D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON), gmm2s robust bw(5)
2-Step GMM estimation
---------------------

Estimates efficient for arbitrary heteroskedasticity and autocorrelation
Statistics robust to heteroskedasticity and autocorrelation
  kernel=Bartlett; bandwidth=5
  time variable (t):  date
                                                      Number of obs =      158
                                                      F(  1,   156) =     2.46
                                                      Prob > F      =   0.1185
Total (centered) SS     =  60.04747699                Centered R2   =   0.1548
Total (uncentered) SS   =  60.05149156                Uncentered R2 =   0.1548
Residual SS             =  50.75430293                Root MSE      =    .5668

------------------------------------------------------------------------------
             |               Robust
       D.inf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          UR |  -.1002374    .0634562    -1.58   0.114    -.2246092    .0241344
       _cons |   .5850796     .372403     1.57   0.116     -.144817    1.314976
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):              7.954
                                                   Chi-sq(4) P-val =    0.0933
------------------------------------------------------------------------------
Weak identification test (Kleibergen-Paap rk Wald F statistic):          7.362
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias    16.85
                                         10% maximal IV relative bias    10.27
                                         20% maximal IV relative bias     6.71
                                         30% maximal IV relative bias     5.34
                                         10% maximal IV size             24.58
                                         15% maximal IV size             13.96
                                         20% maximal IV size             10.26
                                         25% maximal IV size              8.31
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):         3.569
                                                   Chi-sq(3) P-val =    0.3119
------------------------------------------------------------------------------
Instrumented:         UR
Excluded instruments: L2.ggdp L.TBILL L.ER L.TBON
------------------------------------------------------------------------------
```

It appears that by generating HAC estimates of the covariance matrix, the statistical significance of the unemployment rate in this equation is now questioned. One important statistic is also altered: the test for overidentification, denoted as the Sargan test in the former estimates, is on the borderline of rejecting its null hypothesis at the 90% level. When we reestimate the equation with HAC standard errors, various summary statistics are “robustified” as well: here the test of overidentifying restrictions, now denoted Hansen’s J. That statistic is now far from rejection of its null, giving us greater confidence that our instrument set is appropriate.

4 CUE, LIML, and k-class estimation

4.1 CUE and LIML

Again consider the two-step feasible EGMM estimator. In the first step, a consistent but IGMM estimator, $\widetilde{\beta}$, is used to estimate S, the covariance matrix of orthogonality conditions. In the second step, the GMM objective function is maximized by using S^{-1} as the weighting matrix. If we write S as a function of the first-step estimator $\widetilde{\beta}$, the minimization problem in the second step of two-step EGMM estimation that defines the estimator is

$$\widehat{\beta}_{2S\mathrm{EGMM}} \equiv \arg\min_{\widehat{\beta}} J(\widehat{\beta}) = n\overline{g}(\widehat{\beta})'\{S(\widetilde{\beta})\}^{-1}\overline{g}(\widehat{\beta})$$

As noted earlier, the second-step minimization treats the weighting matrix $W = \{S(\widetilde{\beta})\}^{-1}$ as a constant matrix. Thus the residuals in the estimate of S are the first-stage residuals defined by $\widetilde{\beta}$, whereas the residuals in the orthogonality conditions $\overline{g}$ are the second-stage residuals defined by $\widehat{\beta}$.

The minimization problem that defines the GMM/CUE of Hansen, Heaton, and Yaron (1996) is, by contrast,

$$\widehat{\beta}_{\mathrm{CUE}} \equiv \arg\min_{\widehat{\beta}} J(\widehat{\beta}) = n\overline{g}(\widehat{\beta})'\{S(\widehat{\beta})\}^{-1}\overline{g}(\widehat{\beta})$$

Here the weighting matrix is a function of the β being estimated. The residuals in S are the same residuals that are in $\overline{g}$, and estimation of S is done simultaneously with the estimation of β. Generally, solving this minimization problem requires numerical methods.

Both the two-step EGMM and CUE/GMM procedures reduce to familiar estimators under linearity and conditional homoskedasticity. Here $S = E(g_i g_i') = E(u_i^2 Z_i' Z_i) = E(u_i^2)E(Z_i' Z_i) = \sigma^2 Q_{ZZ}$. Q_{ZZ} is estimated by its sample counterpart $(1/n)Z'Z$. In two-step EGMM under homoskedasticity, the minimization becomes

$$\widehat{\beta}_{\mathrm{IV}} \equiv \arg\min_{\widehat{\beta}} J(\widehat{\beta}) = \frac{\widehat{u}(\widehat{\beta})' P_Z \widehat{u}(\widehat{\beta})}{\widehat{\sigma}^2} \tag{9}$$

where $\widehat{u}(\widehat{\beta}) \equiv (y - X\widehat{\beta})$ and $P_Z \equiv Z(Z'Z)^{-1}Z'$ is the projection matrix. In the minimization, the error variance $\widehat{\sigma}^2$ is treated as a constant and hence does not require first-step estimation, and the $\widehat{\beta}$ that solves (9) is the IV estimator $\beta_{\rm IV} = (X'P_ZX)^{-1}X'P_Zy$.[11]

With CUE/GMM under conditional homoskedasticity, the estimated error variance is a function of the residuals

$$\widehat{\sigma}^2 = \widehat{u}'(\widehat{\beta})\widehat{u}(\widehat{\beta})/n$$

and the minimization becomes

$$\widehat{\beta}_{\rm LIML} \equiv \arg\min_{\widehat{\beta}} J(\widehat{\beta}) = \frac{\widehat{u}(\widehat{\beta})'P_Z\widehat{u}(\widehat{\beta})}{\widehat{u}(\widehat{\beta})'\widehat{u}(\widehat{\beta})/n} \tag{10}$$

The $\widehat{\beta}$ that solves (10) is defined as the LIML estimator.

Unlike CUE estimators in general, the LIML estimator can be derived analytically and does not require numerical methods. This derivation is the solution to an eigenvalue problem (see Davidson and MacKinnon 1993, 644–649). The LIML estimator was first derived by Anderson and Rubin (1949), who also provided the first test of overidentifying restrictions for estimation of an equation with endogenous regressors. This Anderson–Rubin statistic (not to be confused with the test discussed below under "weak identification") follows naturally from the solution to the eigenvalue problem. If we denote the minimum eigenvalue by λ, the Anderson–Rubin likelihood-ratio test statistic for the validity of the overidentifying restrictions (orthogonality conditions) is $n\ \log(\lambda)$. Since LIML is also an EGMM estimator, the value J of the minimized GMM objective function also provides a test of overidentifying restrictions. The J test of the same overidentifying restrictions is closely related to the Anderson–Rubin test; the minimized value of the LIML GMM objective function is in fact $J = n(1/1 - \lambda)$. Of course, $n\ \log(\lambda) \approx n(1/1 - \lambda)$.

Although CUE and LIML provide no asymptotic efficiency gains over two-step GMM and IV, recent research suggests that their finite-sample performance may be superior. In particular, there is evidence suggesting that CUE and LIML perform better than IV/GMM in the presence of weak instruments (Hahn, Hausman, and Kuersteiner 2004). This is reflected, for example, in the critical values for the Stock–Yogo weak instruments test discussed in section 7.3.[12] The disadvantage of CUE in general is that it requires numerical optimization; LIML does not but does require the often rather strong assumption of i.i.d. disturbances. In `ivreg2`, the `cue` option combined with the `robust`, `cluster()`, and/or `bw()` options generates coefficient estimates that are efficient in the presence of the corresponding deviations from i.i.d. disturbances. Specifying `cue` with no other options is equivalent to the combination of the options `liml` and `coviv` ("covariance-IV"; see below).

11. The error variance $\widehat{\sigma}^2$, required for inference, is calculated at the end using the IV residuals.
12. With one endogenous regressor and four excluded instruments, the critical value for the Cragg–Donald statistic for 10% maximal size distortion is 24.58 in the case of IV but only 5.44 for LIML.

The implementation of the CUE estimator in `ivreg2` uses Stata's `ml` routine to minimize the objective function. The starting values are either IV or two-step EGMM coefficient estimates. These can be overridden with the `cueinit()` option, which takes a matrix of starting values of the coefficient vector β as its argument. The `cueoptions()` option passes its contents to Stata's `ml` command. Estimation with the `cue` option can be slow and problematic when the number of parameters to be estimated is substantial, and it should be used with caution.

4.2 k-class estimators

LIML, IV, and OLS (but not CUE or two-step GMM) are examples of k-class estimators. A k-class estimator can be written as (Davidson and MacKinnon 1993, 649)

$$\beta_k = \{X'(I - kM_Z)X\}^{-1}X'(I - kM_Z)y$$

where M denotes the annihilation matrix $I - P$. LIML is a k-class estimator with k=λ, the LIML eigenvalue; IV is a k-class estimator with k=1; and OLS is a k-class estimator with k=0. Estimators based on other values of k have been proposed. Fuller's modified LIML (available with the `fuller(#)` option) sets $k = \lambda - \{\alpha/(N - L)\}$, where λ is the LIML eigenvalue, L = number of instruments (included and excluded), and the Fuller parameter α is a user-specified positive constant. The value of $\alpha = 1$ has been suggested as a good choice; see Fuller (1977) or Davidson and MacKinnon (1993, 649–650). Nagar's bias-adjusted 2SLS estimator can be obtained with the `kclass(#)` option by setting $k = 1 + (L - K)/N$, where $(L - K)$ is the number of overidentifying restrictions and N is the sample size; see Nagar (1959). Research suggests that both of these k-class estimators have a better finite-sample performance than IV in the presence of weak instruments, although neither estimator is robust to violations of the i.i.d. assumption. `ivreg2` also provides Stock–Yogo critical values for the Fuller version of LIML.

The default covariance matrix reported by `ivreg2` for the LIML and general k-class estimators is (Davidson and MacKinnon 1993, 650):

$$\widehat{\sigma}^2\{X'(I - kM_Z)X\}^{-1}$$

In fact, the usual IV-type covariance matrix

$$\widehat{\sigma}^2\{X'(I - M_Z)X\}^{-1} = \widehat{\sigma}^2(X'P_ZX)^{-1}$$

is also valid and can be obtained with the `coviv` option. With `coviv`, the covariance matrix for LIML and the other general k-class estimators will differ from that for the IV estimator only because the estimate of the error variance $\widehat{\sigma}^2$ will differ.

4.3 Example of CUE/LIML estimation

We illustrate the use of CUE/LIML estimation using the same equation we used in our discussion of HAC standard errors.

```
. ivreg2 D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON ), cue robust bw(5)
initial:       neg GMM obj function -J =  -3.285175
rescale:       neg GMM obj function -J = -2.8716146
Iteration 0:   neg GMM obj function -J = -2.8716146
Iteration 1:   neg GMM obj function -J =  -2.793201
Iteration 2:   neg GMM obj function -J = -2.7931805
Iteration 3:   neg GMM obj function -J = -2.7931798
Iteration 4:   neg GMM obj function -J = -2.7931798
CUE estimation
--------------

Estimates efficient for arbitrary heteroskedasticity and autocorrelation
Statistics robust to heteroskedasticity and autocorrelation
  kernel=Bartlett; bandwidth=5
  time variable (t):  date
                                                     Number of obs =      158
                                                     F(  1,   156) =     0.55
                                                     Prob > F      =   0.4577
Total (centered) SS     =  60.04747699               Centered R2   =   0.0901
Total (uncentered) SS   =  60.05149156               Uncentered R2 =   0.0901
Residual SS             =   54.6384785               Root MSE      =    .5881

------------------------------------------------------------------------------
             |               Robust
       D.inf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          UR |  -.0483119   .0644743    -0.75   0.454    -.1746792    .0780555
       _cons |   .2978451   .3804607     0.78   0.434    -.4478442    1.043534
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):              7.954
                                                   Chi-sq(4) P-val =    0.0933
------------------------------------------------------------------------------
Weak identification test (Kleibergen-Paap rk Wald F statistic):          7.362
Stock-Yogo weak ID test critical values: 10% maximal LIML size            5.44
                                         15% maximal LIML size            3.87
                                         20% maximal LIML size            3.30
                                         25% maximal LIML size            2.98
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):         2.793
                                                   Chi-sq(3) P-val =    0.4246
------------------------------------------------------------------------------
Instrumented:         UR
Excluded instruments: L2.ggdp L.TBILL L.ER L.TBON
------------------------------------------------------------------------------
```

When this estimator is used, the magnitude of the point estimate of the `UR` coefficient falls yet farther, and it is no longer significantly different from zero at any reasonable level of significance.

5 GMM distance tests of endogeneity and exogeneity

The value J of the GMM objective function evaluated at the EGMM estimator $\widehat{\beta}_{\text{EGMM}}$ is distributed as χ^2 with $(L-K)$ degrees of freedom under the null hypothesis that the full set of orthogonality conditions are valid. This is variously known as the *Sargan statistic*, *Hansen J statistic*, *Sargan–Hansen J test*, or simply a test of overidentifying restrictions.[13]

A C or GMM distance test can be used to test the validity of a subset of orthogonality conditions. Say that the investigator wants to test the validity of L_B orthogonality conditions. Denote J as the value of the GMM objective function for the EGMM estimator that uses the full set of orthogonality conditions and J_A as the value of the EGMM estimator that uses only the $L_A = L - L_B$ orthogonality conditions that the investigator is not questioning. Then under the null that the L_B suspect orthogonality conditions are actually satisfied, the test statistic $(J - J_A) \sim \chi^2$ with L_B degrees of freedom. If the $\widehat{S}$ matrix from the estimation using the full set of orthogonality conditions is used to calculate both GMM estimators, the test statistic is guaranteed to be nonnegative in finite samples.

Baum, Schaffer, and Stillman (2003) discuss how `ivreg2`'s `orthog()` option can be used to conduct a C test of the exogeneity of one or more regressors or instruments. To recapitulate, the `orthog()` option takes as its argument the list of exogenous variables Z_B whose exogeneity is called into question. If the exogenous variable being tested is an instrument, the EGMM estimator that does not use the corresponding orthogonality condition simply drops the instrument. This is illustrated in the following pair of estimations where the second regression is the estimation implied by the `orthog()` option in the first:

```
. ivreg2 y x1 x2 (x3 = z1 z2 z3 z4), orthog(z4)
. ivreg2 y x1 x2 (x3 = z1 z2 z3)
```

If the exogenous variable that is being tested is a regressor, the efficient GMM estimator that does not use the corresponding orthogonality condition treats the regressor as endogenous, as below; again, the second estimation is implied by the use of `orthog()` in the former equation:

```
. ivreg2 y x1 x2 (x3 = z1 z2 z3 z4), orthog(x2)
. ivreg2 y x1 (x2 x3 = z1 z2 z3)
```

Sometimes the researcher wants to test whether an endogenous regressor can be treated as exogenous. This is commonly termed an "endogeneity test", but as we discussed in our earlier paper (Baum, Schaffer, and Stillman 2003, 24–27), it is equivalent to estimating the same regression but treating the regressor as exogenous, and then testing the corresponding orthogonality condition using the `orthog()` option. Although the

13. If the test statistic is required for an IGMM estimator (e.g., an overidentifying restrictions test for the IV estimator that is robust to heteroskedasticity), `ivreg2` reports the J statistic for the corresponding EGMM estimator; see Baum, Schaffer, and Stillman (2003). This J statistic is identical to that produced by `estat overid` following official Stata's `ivregress gmm` command.

procedure described there is appropriate, it is not intuitive. To address this, we have added a new `endogtest()` option to `ivreg2` to conduct endogeneity tests of one or more endogenous regressors. Under the null hypothesis that the specified endogenous regressors can actually be treated as exogenous, the test statistic is distributed as χ^2 with degrees of freedom equal to the number of regressors tested. Thus, in the following estimation,

```
. ivreg2 y x1 x2 (x3 = z1 z2 z3 z4), endogtest(x3)
```

the test statistic reported for the endogeneity of `x3` is numerically equal to the test statistic reported for the `orthog()` option in

```
. ivreg2 y x1 x2 x3 ( = z1 z2 z3 z4), orthog(x3)
```

The `endogtest()` option is both easier to understand and more convenient to use.

Under conditional homoskedasticity, this endogeneity test statistic is numerically equal to a Hausman test statistic; see Hayashi (2000, 233–234) and Baum, Schaffer, and Stillman (2003, 19–22). The endogeneity test statistic can also be calculated after `ivregress` or `ivreg2` by the command `ivendog`. Unlike the Durbin–Wu–Hausman versions of the endogeneity test reported by `ivendog`, the `endogtest()` option of `ivreg2` can report test statistics that are robust to various violations of conditional homoskedasticity. The `ivendog` option unavailable in `ivreg2` is the Wu–Hausman F-test version of the endogeneity test.

To illustrate this option, we use a dataset provided in Wooldridge (2003). We estimate the log of females' wages as a function of the worker's experience, experience-squared, and years of education. If the education variable is considered endogenous, it is instrumented with the worker's age and counts of the number of preschool children and older children in the household. We test whether the `educ` variable need be considered endogenous in this equation with the `endogtest()` option:

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/mroz.dta
. ivreg2 lwage exper expersq (educ=age kidslt6 kidsge6), endogtest(educ)
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only
                                                      Number of obs =      428
                                                      F(  3,   424) =     7.49
                                                      Prob > F      =   0.0001
Total (centered) SS     =  223.3274513                Centered R2   =   0.1556
Total (uncentered) SS   =   829.594813                Uncentered R2 =   0.7727
Residual SS             =  188.5780571                Root MSE      =    .6638

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0964002   .0814278     1.18   0.236    -.0631952    .2559957
       exper |    .042193   .0138831     3.04   0.002     .0149827    .0694033
     expersq |  -.0008323   .0004204    -1.98   0.048    -.0016563   -8.33e-06
       _cons |  -.3848718   1.011551    -0.38   0.704    -2.367476    1.597732
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):          12.816
                                                   Chi-sq(3) P-val =    0.0051
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                4.342
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias    13.91
                                         10% maximal IV relative bias     9.08
                                         20% maximal IV relative bias     6.46
                                         30% maximal IV relative bias     5.39
                                         10% maximal IV size             22.30
                                         15% maximal IV size             12.83
                                         20% maximal IV size              9.54
                                         25% maximal IV size              7.80
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):           0.702
                                                   Chi-sq(2) P-val =    0.7042
-endog- option:
Endogeneity test of endogenous regressors:                               0.019
                                                   Chi-sq(1) P-val =    0.8899
Regressors tested:    educ
------------------------------------------------------------------------------
Instrumented:         educ
Included instruments: exper expersq
Excluded instruments: age kidslt6 kidsge6
------------------------------------------------------------------------------
```

In this context, we estimate the equation treating `educ` as endogenous, and merely name it in the `endogtest()` varlist to perform the C (GMM distance) test. The test cannot reject its null that `educ` may be treated as exogenous. In contrast, we may calculate this same test statistic with the earlier `orthog()` option:

```
. ivreg2 lwage exper expersq educ (=age kidslt6 kidsge6), orthog(educ)
```

By using `orthog()`, we again list `educ` in the option's varlist, but we must estimate the equation with that variable treated as exogenous: an equivalent but perhaps a less intuitive way to perform the test.

6 The FWL theorem and a rank-deficient S matrix

According to the Frisch–Waugh–Lovell (FWL) theorem (Frisch and Waugh 1933, Lovell 1963), the coefficients estimated for a regression in which some exogenous regressors, say, X_{2A}, are partialled out from the dependent variable y; the endogenous regressors X_1; the other exogenous regressors X_{2B}; and the excluded instruments Z_1 will be the same as the coefficients estimated for the original model for certain estimators. That is, if we denote a partialled-out variable with a tilde so that $\widetilde{y} \equiv M_{2A}y$, the coefficients estimated for the partialled-out version of the model

$$\widetilde{y} = [\widetilde{X}_1 \;\; \widetilde{X}_{2B}][\beta_1' \;\; \beta_{2B}']' + \widetilde{u}$$

with instruments $\widetilde{Z}_1$ and $\widetilde{X}_{2B}$ will be the same as the shared coefficients fitted for the original model

$$y = [X_1 \;\; X_2][\beta_1' \;\; \beta_2']' + u$$

with instruments Z_1 and X_2. It is even possible to partial-out the full set of included exogenous variables X_2, so that the partialled-out version of the model becomes

$$\widetilde{y} = \widetilde{X}_1\beta_1 + \widetilde{u}$$

with no exogenous regressors and only excluded instruments $\widetilde{Z}_1$, and the estimated $\widehat{\beta}_1$ will be the same as that obtained when estimating the full set of regressors.

The FWL theorem is implemented in `ivreg2` by the new `partial(`*varlist*`)` option, which requests that the exogenous regressors in the *varlist* should be partialled out from all the other variables (other regressors and excluded instruments) in the estimation. If the equation includes a constant, it is automatically partialled out as well.

The `partial()` option is most useful when the covariance matrix of orthogonality conditions S is not of full rank. When this is the case, EGMM and overidentification tests are infeasible as the optimal GMM weighting matrix $W = S^{-1}$ cannot be calculated. Sometimes partialling-out enough exogenous regressors can make the covariance matrix of the remaining orthogonality conditions full rank, and EGMM becomes feasible.

The invariance of the estimation results to partialling-out applies to one- and two-step estimators such as OLS, IV, LIML, and two-step GMM, but not to CUE or to GMM iterated more than two steps. The reason is that the latter estimators update the estimated S matrix. An updated S implies different estimates of the coefficients on the partialled-out variables, which imply different residuals, which in turn produce a different estimated S. Intuitively, partialling-out uses OLS estimates of the coefficients on the partialled-out variables to generate the S matrix, whereas CUE would use more efficient heteroskedastic OLS (HOLS) estimates.[14] Partialling-out exogenous regressors that are not of interest may still be desirable with CUE estimation, however, because reducing the number of parameters estimated makes the CUE numerical optimization faster and more reliable.

14. We are grateful to Manuel Arellano for helpful discussions on this point. For information on HOLS, see Baum, Schaffer, and Stillman (2003).

One common case calling for partialling-out arises when using `cluster()` and the number of clusters is less than L, the number of (exogenous regressors + excluded instruments). This causes the matrix S to be rank deficient (Baum, Schaffer, and Stillman 2003, 9–10). The problem can be addressed by using `partial()` to remove enough exogenous regressors for S to have full rank. A similar problem arises if a robust covariance matrix is requested when the regressors include a variable that is a singleton dummy, i.e., a variable with one value of 1 and $(N-1)$ values of zero or vice versa. The singleton dummy causes the robust covariance matrix estimator to be less than full rank. Here partialling-out the variable with the singleton dummy solves the problem.

The `partial()` option has two limitations: it cannot be used with time-series operators, and postestimation command `predict` can be used only to generate residuals.

7 Underidentification, weak identification, and instrument relevance

7.1 Identification and the rank condition

For (1) to be estimable, it must be *identified*. The order condition $L \geq K$ is necessary but not sufficient; the rank condition must also be satisfied. The rank condition states that the matrix $Q_{XZ} \equiv E(X_i'Z_i)$ is of full column rank, i.e., Q_{XZ} must have rank K. Since $X_2 \equiv Z_2$, we can simplify by partialling them out from X_1 and Z_1, and the rank condition becomes $\rho(Q_{\widetilde{X}_1\widetilde{Z}_1}) = K_1$. There are several ways of interpreting this condition.

One interpretation is in terms of *correlations*: the excluded instruments must be correlated with the endogenous regressors. In the simplest possible case of an endogenous regressor, an excluded instrument, and partialling-out any exogenous regressors including the constant, $L_1 = K_1 = 1$ and $Q_{\widetilde{X}_1\widetilde{Z}_1}$ is a scalar. As the constant has been partialled-out, $E(X_i) = E(Z_i) = 0$ and $Q_{\widetilde{X}_1\widetilde{Z}_1}$ is a covariance. The rank condition in this simple case requires that the correlation or covariance between $\widetilde{X}_1$ and $\widetilde{Z}_1$ is nonzero.

This interpretation can be extended to the general case of $L_1, K_1 \geq 1$ by using *canonical correlations* (Anderson 1984, ch. 12; Hall, Rudebusch, and Wilcox 1996, 287; [MV] **canon**). The canonical correlations r_i between $\widetilde{X}_1$ and $\widetilde{Z}_1$, $i = 1, \ldots, K_1$ represent the correlations between linear combinations of the K_1 columns of $\widetilde{X}_1$ and linear combinations of the L_1 columns of $\widetilde{Z}_1$.[15] In the special case of $L_1 = K_1 = 1$ (and partialling-out the constant), the canonical correlation between $\widetilde{X}_1$ and $\widetilde{Z}_1$ is the usual Pearson correlation coefficient. In the slightly more general case of $L_1 \geq 1$ and $K_1 = 1$, the canonical correlation between $\widetilde{X}_1$ and $\widetilde{Z}_1$ is simply R: the square root of R^2 in a regression of $\widetilde{X}$ on $\widetilde{Z}$. In the general case of $L_1, K_1 \geq 1$, the squared canonical corre-

15. As $X_2 \equiv Z_2$, these variables are perfectly correlated with each other. The canonical correlations between X and Z before partialling out would also include the $L_2 \equiv K_2$ correlations that are equal to unity.

lations may be calculated as the eigenvalues of $(\widetilde{X}_1'\widetilde{X}_1)^{-1}(\widetilde{X}_1'\widetilde{Z}_1)(\widetilde{Z}_1'\widetilde{Z}_1)^{-1}(\widetilde{Z}_1'\widetilde{X}_1)$. The rank condition can then be interpreted as the requirement that all K_1 of the canonical correlations must be significantly different from zero. If one or more of the canonical correlations is zero, the model is *underidentified* or *unidentified*.

An alternative and useful interpretation of the rank condition is to use the reduced form. Write the set of reduced-form (first stage) equations for the regressors X as

$$X = Z\Pi + v$$

By using our partitioning of X and Z, we can rewrite this as

$$X_1 = [Z_1 \;\; Z_2] \; [\Pi_{11}' \;\; \Pi_{12}']' + v_1 \tag{11}$$

$$X_2 = [Z_1 \;\; Z_2] \; [\Pi_{21}' \;\; \Pi_{22}']' + v_2$$

The equation for X_2 is not interesting because $X_2 \equiv Z_2$ follows that $\Pi_{21} = 0$ and $\Pi_{22} = I$. The rank condition for identification comes from the equation for the endogenous regressors X_1. The $L \times K_1$ matrix Π_{11} must be of full column rank ($\rho(\Pi_{11}) = K_1$). If $\rho(\Pi_{11}) < K_1$, the model is again unidentified.

The consequence of utilizing excluded instruments that are uncorrelated with the endogenous regressors is increased bias in the estimated IV coefficients (Hahn and Hausman 2002) and worsening of the large-sample approximations to the finite-sample distributions. Here the bias of the IV estimator is the same as that of the OLS estimator and IV becomes inconsistent (ibid.). Here instrumenting only aggravates the problem, as IV and OLS share the same bias but IV has a larger mean squared error (MSE) by virtue of its larger variance. Serious problems also arise if the correlations between the excluded instruments and endogenous regressors are nonzero but "weak". Standard IV/GMM methods of estimating β_1 suffer from serious finite sample bias problems and alternative methods should be considered.

In the rest of this section, we show how to use `ivreg2` to conduct tests for underidentification and weak identification and how `ivreg2` provides a procedure for inference that is robust to weak identification.

7.2 Testing for underidentification and instrument redundancy

Of course, we do not observe the true Q_{XZ} or Π_{11} matrices; these matrices must be estimated. Testing whether or not the rank condition is satisfied therefore amounts to testing the rank of a matrix. Do the data enable the researcher to reject the null hypothesis that the equation is underidentified, i.e., that $\rho(\widehat{\Pi}_{11}) = (K_1 - 1)$, or, equivalently, $\rho(\widehat{Q}_{\widetilde{X}\widetilde{Z}}) = (K_1 - 1)$? Rejection of the null implies full rank and identification; failure to reject the null implies the matrix is rank-deficient and the equation is underidentified.

If the reduced-form errors v are i.i.d., two approaches are available for testing the rank of $Q_{\widetilde{X}\widetilde{Z}}$: Anderson's (1951) canonical correlations test and the related test of Cragg and Donald (1993). In Anderson's approach, $H_0: \rho(\widehat{Q}_{\widetilde{X}\widetilde{Z}}) = (K_1 - 1)$ is equivalent to the null hypothesis that the smallest canonical correlation r_{K_1} is zero. A large sample test statistic for this is simply $nr^2_{K_1}$. Under the null, the test statistic is distributed χ^2 with $(L - K + 1)$ degrees of freedom, so that it may be calculated even for an exactly identified equation. A failure to reject the null hypothesis suggests that the model is unidentified. Not surprisingly given its "$N \times R^2$" form this test can be interpreted as an LM test.[16]

The Cragg–Donald (1993) statistic is an alternative and closely related test for the rank of a matrix that can also be used to test for underidentification. Whereas the Anderson test is an LM test, the Cragg–Donald test is a Wald test, also derived from an eigenvalue problem. Poskitt and Skeels (2002) show that in fact the Cragg–Donald test statistic can be stated in terms of canonical correlations as $nr^2_{K_1}/(1 - r^2_{K_1})$ (see Poskitt and Skeels 2002, 17). It is also distributed as $\chi^2(L - K + 1)$.

Both these tests require the assumption of i.i.d. errors and hence are reported if `ivreg2` is invoked without the `robust`, `cluster()`, or `bw()` options. The Anderson LM χ^2 statistic is reported by `ivreg2` in the main regression output whereas both the Anderson LM and Cragg–Donald Wald χ^2 statistics are reported when the `first` option is specified.

If the errors are heteroskedastic or serially correlated, the Anderson and Cragg–Donald statistics are not valid. This is an important shortcoming, because these violations of the i.i.d. assumption would typically be expected to cause the null of underidentification to be rejected too often. Researchers would face the danger of interpreting a rejection of the null as evidence of a well-specified model that is adequately identified, when in fact it was both underidentified and misspecified.

Recently, several robust statistics for testing the rank of a matrix have been proposed. Kleibergen and Paap (2006) have proposed the rk statistic for this purpose. Their rk test statistic is reported by `ivreg2` if the user requests any sort of robust covariance estimator. The LM version of the Kleibergen–Paap rk statistic can be considered as a generalization of the Anderson canonical correlation rank statistic to the non-i.i.d. case. Similarly, the Wald version of the rk statistic reduces to the Cragg–Donald statistic when the errors are i.i.d. The rk test is implemented in Stata by the `ranktest` command of Kleibergen and Schaffer (2007), which `ivreg2` uses to calculate the rk statistic. If `ivreg2` is invoked with the `robust`, `bw()`, or `cluster()` options, the tests of underidentification reported by `ivreg2` are based on the rk statistic and will be correspondingly robust to heteroskedasticity, autocorrelation, or clustering. For a full discussion of the rk statistic, see Kleibergen and Paap (2006).

16. Earlier versions of `ivreg2` reported an LR version of this test, where the test statistic is $-n \log(1 - r^2_{K_1})$. This LR test has the same asymptotic distribution as the LM form. See Anderson (1984, 497–498).

In the special case of one endogenous regressor, the Anderson, Cragg–Donald, and Kleibergen–Paap statistics reduce to familiar statistics available from OLS estimation of the single reduced-form equation with an appropriate choice of VCE estimator. Thus the Cragg–Donald Wald statistic can be calculated by estimating (11) and testing the joint significance of the coefficients Π_{11} on the excluded instruments Z_1 by using a standard Wald test and a traditional nonrobust covariance estimator. The Anderson LM statistic can be obtained by calculating an LM test of the same joint hypothesis.[17] The Kleibergen–Paap rk statistics can be obtained by performing the same tests with the desired robust covariance estimator. For example, estimating (11) using OLS and testing the joint significance of Z_1 using a heteroskedasticity-robust covariance estimator yields the heteroskedastic-robust Kleibergen–Paap rk Wald statistic.[18]

The same framework may also be used to test a set of instruments for *redundancy* as shown by Breusch et al. (1999). In an overidentified context with $L \geq K$, if some of the instruments are redundant then the large-sample efficiency of the estimation is not improved by including them. It is well known, moreover, that using several instruments or moment conditions can cause the estimator to have poor finite-sample performance. Dropping redundant instruments may therefore lead to more reliable estimation.

The intuition behind a test for instrument redundancy is straightforward. As above, assume that we have partialled-out any exogenous regressors X_2. Partition the excluded instruments $\widetilde{Z}_1$ into $[\ \widetilde{Z}_{1A}\ \ \widetilde{Z}_{1B}\ ]$, where $\widetilde{Z}_{1B}$ is the set of possibly redundant instruments after X_2 has been partialled-out. Breusch et al. (1999, 106) show that the redundancy of $\widetilde{Z}_{1B}$ can be stated in several ways: (a) $\text{plim}(1/n)\widetilde{Z}'_{1B}M_{\widetilde{Z}_{1A}}\widetilde{X}_1 = 0$; (b) the correlations between $\widetilde{Z}_{1B}$ and $\widetilde{X}_1$ (given $\widetilde{Z}_{1A}$) are zero; (c) in a regression of $\widetilde{X}_1$ on the full set of excluded instruments $\widetilde{Z}_1$, the coefficients on $\widetilde{Z}_{1B}$ are zero. It is easy to see that the FWL theorem can be used to restate this last condition without the partialling-out of X_2: (d) in a regression of X_1 on the full set of included and excluded instruments Z, i.e., the reduced form (11), the coefficients on Z_{1B} are zero. As Hall and Peixe (2003) point out, redundancy is a conditional concept. Z_{1B} either is or is not redundant conditional on Z_{1A}.

The above suggests a straightforward test of redundancy: simply estimate (11) using OLS and test the significance of Z_{1B} by using a large-sample LM, Wald, or LR test. For example, the redundancy test proposed by Hall and Peixe (2003) is the LR version of this test. These test statistics are all distributed as χ^2 with degrees of freedom equal to the number of endogenous regressors times the number of instruments tested. As usual, implementing this test is easy for the case of an endogenous variable, as only

17. This can be done simply in Stata using `ivreg2` by estimating (11) with only Z_2 as regressors, Z_1 as excluded instruments, and an empty list of endogenous regressors. The Sargan statistic reported by `ivreg2` will be the Anderson LM statistic. See Baum, Schaffer, and Stillman (2003) for further discussion.

18. See the online help for `ranktest` for examples. These test statistics are large-sample χ^2 tests and can be obtained from OLS regression using `ivreg2`. Stata's `regress` command reports finite-sample t tests. Also the robust rk LM statistic can be obtained as described in the preceding footnote. Invoke `ivreg2` with X_1 as the dependent variable, Z_2 as regressors, Z_1 as excluded instruments and no endogenous regressors. With the `robust` option, the reported Hansen J statistic is the robust rk statistic.

an OLS estimation is necessary. The tests of the coefficients can be made robust to various violations of i.i.d. errors in the usual way. However, this procedure is more laborious (though still straightforward) if $K_1 > 1$ as it is then necessary to jointly estimate multiple reduced-form equations.

Fortunately, a simpler procedure is available that will generate numerically equivalent test statistics for redundancy. Define a matrix $\breve{X}$ as X with both X_2 and Z_{1A} partialled-out. Then condition (a) can be restated as (e) $\text{plim}(1/n)\breve{Z}_{1B}'\breve{X}_1 = 0$ or (f) that the correlations between $\breve{Z}_{1B}$ and $\breve{X}_1$ (given Z_{1A} and Z_2) are zero. The redundancy of Z_{1B} can be evaluated by using the `ranktest` command to test the null hypothesis that the rank of $Q_{\breve{X}\breve{Z}}$ is zero. Rejection of the null indicates that the instruments are not redundant. The LM version of the Anderson canonical correlations test is reported if the user indicates that the errors are i.i.d. Here the LM test statistic is n times the sum of the squared canonical correlations between $\breve{Z}_{1B}$ and $\breve{X}_1$. If the user estimates the equation with `robust`, `bw()`, or `cluster()`, an LM version of the Kleibergen–Paap rk statistic is reported that is correspondingly robust to heteroskedasticity, autocorrelation, or clustering.

7.3 Testing for weak identification

The *weak-instruments* problem arises when the correlations between the endogenous regressors and the excluded instruments are nonzero but small. In the past 10–15 years, much attention in the econometrics literature has been devoted to this topic. What is surprising is that, as Bound, Jaeger, and Baker (1995), Staiger and Stock (1997), and others have shown, the weak-instruments problem can arise even when the correlations between X and Z are significant at conventional levels (5% or 1%) and the researcher is using a large sample. For more detailed discussion of the weak-instruments problem, see Staiger and Stock (1997), Stock, Wright, and Yogo (2002), or Dufour (2003). Thus rejecting the null of underidentification using the tests in the previous section and conventional significance levels is not enough; you must call for other methods.

One approach that has been advanced by Stock and Yogo (2005) is to test for the presence of weak instruments. The difference between this approach and the aforementioned underidentification tests is not in the basic statistic used, but in the finite-sample adjustments and critical values and in the null hypothesis being tested. Moreover, the critical values for a weak-instruments test are different for different estimators because the estimators are not affected to the same degree by weak instruments. Specifically, the LIML and CUE estimators are more robust to the presence of weak instruments than are IV and two-step GMM.

The test statistic proposed by Stock and Yogo (2005) is the F-statistic form of the Cragg and Donald (1993) statistic, $\{(N-L)/L_2\}\{r_{K_1}^2/(1-r_{K_1}^2)\}$. `ivreg2` will report this statistic for an estimation that assumes i.i.d. disturbances. The null hypothesis being tested is that the estimator is weakly identified in the sense that it is subject to bias that the investigator finds unacceptably large. The Stock–Yogo weak-instruments tests come in two types: maximal relative bias and maximal size, where the null is that the instruments do not suffer from the specified bias. Rejection of their null hypothesis represents the absence of a weak-instruments problem. The first type is based on the ratio of the bias of the estimator to the bias of OLS. The null is that instruments are weak, where weak instruments are defined as instruments that can lead to an asymptotic relative bias greater than some value b. Because this test uses the finite-sample distribution of the IV estimator, it cannot be calculated in certain cases. This is because the m^{th} moment of the IV estimator exists if and only if $m < (L - K + 1)$.[19]

The second type of the Stock–Yogo tests is based on the performance of the Wald test statistic for β_1. Under weak identification, the Wald test rejects too often. The test statistic is based on the rejection rate r (10%, 20%, etc.) that the researcher is willing to tolerate if the true rejection rate should be the standard 5%. Weak instruments are defined as instruments that will lead to a rejection rate of r when the true rejection rate is 5%.

Stock and Yogo (2005) have tabulated critical values for their two weak-identification tests for the IV estimator, the LIML estimator, and Fuller's modified LIML estimator. The weak-instruments bias in the IV estimator is larger than that of the LIML estimators, and hence the critical values for the null that instruments are weak are also larger. The Stock–Yogo critical values are available for a range of possible circumstances (up to 3 endogenous regressors and 100 excluded instruments).

The weak-identification test that uses the Cragg–Donald F statistic, like the corresponding underidentification test, requires an assumption of i.i.d. errors. This is a potentially serious problem, for the same reason as given earlier: if the test statistic is large simply because the disturbances are not i.i.d., the researcher will commit a type-I error and incorrectly conclude that the model is adequately identified.

If the user specifies the `robust`, `cluster()`, or `bw()` options in `ivreg2`, the reported weak-instruments test statistic is a Wald F statistic based on the Kleibergen–Paap rk statistic. We are not aware of any studies on testing for weak instruments in the presence of non-i.i.d. errors. In our view, however, the use of the rk Wald statistic, as the robust analog of the Cragg–Donald statistic, is a sensible choice and clearly superior to the use of the latter in the presence of heteroskedasticity, autocorrelation, or clustering. We suggest, however, that when using the rk statistic to test for weak identification, users either apply with caution the critical values compiled by Stock and Yogo (2005) for the i.i.d. case or refer to the older "rule of thumb" of Staiger and Stock (1997), which says that the F statistic should be at least 10 for weak identification not to be considered a problem.

`ivreg2` will report in the main regression output the relevant Stock and Yogo (2005) critical values for IV, LIML, and Fuller-LIML estimates if they are available. The reported test statistic will be the Cragg–Donald statistic if the traditional covariance estimator is used or the rk statistic if a robust covariance estimator is requested. If the user requests two-step GMM estimation, `ivreg2` will report an rk statistic and the IV critical values. If the user requests the CUE estimator, `ivreg2` will report an rk statistic and

19. See Davidson and MacKinnon (1993, 221–222).

the LIML critical values. The justification for this is that IV and LIML are special cases of two-step GMM and CUE, respectively. The similarities carry over to weak instruments: the literature suggests that IV and two-step GMM are less robust to weak instruments than LIML and CUE. However, users of `ivreg2` may again wish to exercise some caution in applying the Stock–Yogo critical values in these cases.

7.4 Inference robust to weak identification: the Anderson–Rubin test

The first-stage `ivreg2` output also includes the Anderson and Rubin (1949) test of the significance of the endogenous regressors in the structural equation being estimated (not to be confused with the Anderson and Rubin (1949) overidentification test discussed earlier). In the form reported by `ivreg2`, the null hypothesis tested is that the coefficients β_1 of the endogenous regressors X_1 in the structural equation are jointly equal to zero. It is easily extended to testing the equality of the coefficients of X_1 to other values, but this is not supported explicitly by `ivreg2`; see the next section for further discussion.

The development of this Anderson and Rubin (1949) test is straightforward. Substitute the reduced-form expression (11) for the endogenous regressors X_1 into the main equation of the model

$$y = X\beta + u = X_1\beta_1 + Z_2\beta_2 + u = ([Z_1 \;\; Z_2]\,[\Pi_{11}' \;\; \Pi_{12}']' + v_1)\beta_1 + Z_2\beta_2 + u$$

and rearrange to obtain

$$y = Z_1\Pi_{11}\beta_1 + Z_2(\Pi_{12}\beta_1 + \beta_2) + (v_1\beta_1 + u)$$

Now consider estimating a reduced-form equation for y with the full set of instruments as regressors:

$$y = Z_1\gamma_1 + Z_2\gamma_2 + \eta$$

If the null $H_0 : \beta_1 = 0$ is correct, $\Pi_{11}\beta_1 = 0$, and therefore $\gamma_1 = 0$. Thus the Anderson and Rubin (1949) test of the null $H_0 : \beta_1 = 0$ is obtained by estimating the reduced form for y and testing that the coefficients γ_1 of the excluded instruments Z_1 are jointly equal to zero. If we fail to reject $\gamma_1 = 0$, then we also fail to reject $\beta_1 = 0$.

The Anderson–Rubin statistic is robust to the presence of weak instruments. As instruments become weak, the elements of Π_{11} become smaller and hence so does $\Pi_{11}\beta_1$: the null $H_0 : \gamma_1 = 0$ is less likely to be rejected. That is, as instruments become weak, the power of the test declines, an intuitively appealing feature: weak instruments come at a price. `ivreg2` reports both the χ^2 version of the Anderson–Rubin statistic (distributed with L_1 degrees of freedom) and the F statistic version of the test. `ivreg2` also reports the closely related Stock and Wright (2000) S statistic. The S statistic tests the same null hypothesis as the Anderson–Rubin statistic and has the same distribution under the null. It is given by the value of the CUE objective function (with the exogenous regressors partialled out). Whereas the Anderson–Rubin statistic provides a Wald test, the S statistic provides an LM or GMM distance test of the same hypothesis.

More importantly, if the model is fitted with a robust covariance matrix estimator, both the Anderson–Rubin statistic and the S statistic reported by `ivreg2` are correspondingly robust. See Dufour (2003) and Chernozhukov and Hansen (2005) for further discussion of the Anderson–Rubin approach. For related alternative test statistics that are also robust to weak instruments (but not violations of the i.i.d. assumption), see the `condivreg` and `condtest` commands available from Moreira and Poi (2003) and Mikusheva and Poi (2006).

7.5 An example of estimation with weak instruments using ivreg2

We illustrate the weak-instruments problem with a variation on a log wage equation illustrated in Hayashi (2000). The explanatory variables are `s` (completed years of schooling), `expr` (years of work experience), `tenure` in the current job in years, `rns` (a dummy for residency in the southern U.S.), `smsa` (a dummy for urban workers), the worker's `iq` score, and a set of year dummies. Instruments include the worker's `age` and `mrt` (marital status: 1 = married) as instruments.

```
. use http://www.stata-press.com/data/imeus/griliches, clear
(Wages of Very Young Men, Zvi Griliches, J.Pol.Ec. 1976)

. ivreg2 lw s expr tenure rns smsa _I* (iq = age mrt), ffirst robust
> redundant(mrt)

Summary results for first-stage regressions
-------------------------------------------

Variable    | Shea Partial R2 |   Partial R2    |  F(  2,   744)    P-value
iq          |     0.0073      |     0.0073      |        2.93       0.0539

NB: first-stage F-stat heteroskedasticity-robust

Underidentification tests
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Kleibergen-Paap rk LM statistic            Chi-sq(2)=5.90     P-val=0.0524
Kleibergen-Paap rk Wald statistic          Chi-sq(2)=5.98     P-val=0.0504

Weak identification test
Ho: equation is weakly identified
Kleibergen-Paap Wald rk F statistic                2.93
See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and overidentifying restrictions are valid
Anderson-Rubin Wald test     F(2,744)= 46.95     P-val=0.0000
Anderson-Rubin Wald test     Chi-sq(2)=95.66     P-val=0.0000
Stock-Wright LM S statistic  Chi-sq(2)=69.37     P-val=0.0000

NB: Underidentification, weak identification and weak-identification-robust
    test statistics heteroskedasticity-robust

Number of observations              N  =       758
Number of regressors                K  =        13
Number of instruments               L  =        14
Number of excluded instruments      L1 =         2
```

```
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity
                                                      Number of obs =      758
                                                      F( 12,   745) =     4.42
                                                      Prob > F      =   0.0000
Total (centered) SS     =  139.2861498                Centered R2   =  -6.4195
Total (uncentered) SS   =  24652.24662                Uncentered R2 =   0.9581
Residual SS             =  1033.432656                Root MSE      =    1.168

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |  -.0948902   .0418904    -2.27   0.024    -.1769939   -.0127865
           s |   .3397121   .1183267     2.87   0.004     .1077959    .5716282
        expr |   -.006604   .0292551    -0.23   0.821    -.0639429     .050735
      tenure |   .0848854   .0306682     2.77   0.006     .0247768     .144994
         rns |  -.3769393   .1559971    -2.42   0.016     -.682688   -.0711906
        smsa |   .2181191   .1031119     2.12   0.034     .0160236    .4202146
   _Iyear_67 |   .0077748   .1663252     0.05   0.963    -.3182166    .3337662
   _Iyear_68 |   .0377993   .1523585     0.25   0.804    -.2608179    .3364165
   _Iyear_69 |   .3347027   .1637992     2.04   0.041     .0136622    .6557432
   _Iyear_70 |   .6286425   .2468458     2.55   0.011     .1448336    1.112451
   _Iyear_71 |   .4446099   .1861877     2.39   0.017     .0796887     .809531
   _Iyear_73 |    .439027   .1668657     2.63   0.009     .1119763    .7660778
       _cons |   10.55096   2.781762     3.79   0.000     5.098812    16.00312
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):              5.897
                                                   Chi-sq(2) P-val =    0.0524
-redundant- option:
IV redundancy test (LM test of redundancy of specified instruments):     0.002
                                                   Chi-sq(1) P-val =    0.9665
Instruments tested:   mrt
------------------------------------------------------------------------------
Weak identification test (Kleibergen-Paap rk Wald F statistic):          2.932
Stock-Yogo weak ID test critical values: 10% maximal IV size             19.93
                                         15% maximal IV size             11.59
                                         20% maximal IV size              8.75
                                         25% maximal IV size              7.25
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):         1.564
                                                   Chi-sq(1) P-val =    0.2111
------------------------------------------------------------------------------
Instrumented:         iq
Included instruments: s expr tenure rns smsa _Iyear_67 _Iyear_68 _Iyear_69
                      _Iyear_70 _Iyear_71 _Iyear_73
Excluded instruments: age mrt
------------------------------------------------------------------------------
```

In the first-stage regression results, the Kleibergen–Paap underidentification LM and Wald tests fail to reject their null hypotheses at the 95% level, suggesting that even for overidentification with the order condition, the instruments may be inadequate to identify the equation. The Anderson–Rubin Wald test and Stock–Wright LM test readily reject their null hypothesis and indicate that the endogenous regressors are relevant. However, given that those null hypotheses are joint tests of irrelevant regressors and appropriate overidentifying restrictions, the evidence is not so promising. In the main equation output, the `redundant(mrt)` option indicates that `mrt` provides no useful information to identify the equation. This equation may be exactly identified at best.

7.6 The relationship between inference robust to weak identification and overidentification tests

The Anderson–Rubin test that is robust to weak identification (and its related alternatives) relies heavily on the orthogonality of the excluded instruments Z_1. If the orthogonality conditions are violated, the Anderson–Rubin test will tend to reject the null $H_0: \beta_1 = 0$ even if the true $\beta_1 = 0$. The reason is easy to see: if Z_1 is correlated with the disturbance u, it will therefore also be correlated with the reduced-form error η, and so the estimated $\widehat{\gamma}_1$ will be biased away from zero even if in reality $\beta_1 = 0$.

Generally, in a test of overidentification, the maintained hypothesis is that the model is identified, so that a rejection means rejecting the orthogonality conditions. In the β_1 test that is robust to weak identification, the maintained hypothesis is that the instruments are valid, so that a rejection means rejecting the null that β_1 equals the hypothesized value.

This relationship between weak identification and overidentification tests can be stated precisely in the case of CUE or LIML estimation. We have been careful in the above to state that the two Anderson–Rubin tests should not be confused, but they are, in a sense, based on the same statistic. Assume that the exogenous regressors X_2, if any, have been partialled-out so that $\beta_1 \equiv \beta$. The value of the CUE/GMM objective function at $\widehat{\beta}_{CUE}$ provides a test of the orthogonality conditions; the LIML LR version of this test is the Anderson–Rubin overidentifying restrictions test. The value of the CUE/GMM objective function at some other, hypothesized $\widetilde{\beta}$ provides a test $H_0: \beta = \widetilde{\beta}$. This is the Stock and Wright (2000) S statistic, which is a Lagrange multiplier (LM) version of the Anderson–Rubin weak-instruments-robust test.

This can be illustrated using the Hayashi–Griliches example below. We assume conditional homoskedasticity and estimate using LIML. The Anderson–Rubin LR overidentification statistic (distributed with one degree of freedom) is small, as is the Sargan–Hansen J statistic, suggesting that the orthogonality conditions are valid:

```
. use http://www.stata-press.com/data/imeus/griliches, clear
(Wages of Very Young Men, Zvi Griliches, J.Pol.Ec. 1976)

. qui ivreg2 lw s expr tenure rns smsa _I* (iq = age mrt),
> partial(s expr tenure rns smsa _I*) liml

. di e(arubin)
1.1263807

. di e(j)
1.1255442
```

The Anderson–Rubin test of $H_0: \beta_{IQ} = 0$ is calculated automatically by `ivreg2` with the `ffirst` option and is equivalent to estimating the reduced form for `lw` and testing the joint significance of the excluded instruments `age` and `mrt`:

```
. qui ivreg2 lw s expr tenure rns smsa _I* (iq = age mrt), liml ffirst
. di e(archi2)
89.313862
. qui ivreg2 lw s expr tenure rns smsa _I* age mrt
. test age mrt
 ( 1)  age = 0
 ( 2)  mrt = 0
           chi2(  2) =   89.31
         Prob > chi2 =    0.0000
```

The Stock–Wright S statistic is an LM or GMM distance test of the same hypothesis. This LM version of the Anderson–Rubin Wald test of `age` and `mrt` using the reduced-form estimation above is asymptotically equivalent to an LM test of the same hypothesis, available by using `ivreg2` and specifying these as excluded instruments (see Baum, Schaffer, and Stillman 2003 for further discussion). It is this LM version of the Anderson–Rubin weak-instruments-robust test that is numerically identical to the value of the GMM objective function at the hypothesized value $\beta_{IQ} = 0$:

```
. qui ivreg2 lw s expr tenure rns smsa _I* (=age mrt)
. di e(j)
79.899445
. mat b[1,1]=0
. qui ivreg2 lw s expr tenure rns smsa _I* (iq = age mrt),
> partial(s expr tenure rns smsa _I*) b0(b)
. di e(j)
79.899445
```

For $J(\beta_0)$ to be the appropriate test statistic, it is necessary for the exogenous regressors to be partialled out with the `partial()` option.

7.7 More first-stage options

To aid in the diagnosis of weak instruments, the `savefirst` option requests that the individual first-stage regressions be saved for later access by using the `estimates` command; see [R] **estimates**. If saved, they can also be displayed using `first` or `ffirst` and the `ivreg2` replay syntax. The regressions are saved with the prefix `_ivreg2_` unless the user specifies an alternative prefix with the `savefprefix(`*prefix*`)` option. The saved estimation results may be made the active set with `estimates restore`, allowing commands such as `test`, `lincom`, and `testparm` to be used.

The `rf` option requests that the reduced-form estimation of the equation be displayed. The `saverf` option requests that the reduced-form estimation is saved for later access by using the `estimates` command. If saved, it can also be displayed by using the `rf` and `ivreg2` replay syntax. The regression is saved with the prefix `_ivreg2_` unless the user specifies an alternative prefix with the `saverfprefix(`*prefix*`)` option.

8 Advanced ivreg2 options

Two options are available for speeding `ivreg2` execution. `nocollin` specifies that the collinearity checks not be performed. This option should be used with caution. `noid` suspends calculating and reporting of the underidentification and weak identification statistics in the main output.

The `b0(`*matrix*`)` option allows the user to specify that the GMM objective function, J, should be calculated for an arbitrary parameter vector. The parameter vector must be given as a matrix with appropriate row and column labels. The `b0()` option is most useful if the user wishes to conduct a weak-instruments-robust test of $H_0 : \beta_1 = b_0$, where b_0 is specified by the user. For example, in the illustration given in section 7.6, the null hypothesis that the coefficient on `iq` is 0.05 can be tested simply by replacing the line `mat b=J(1,1,0)` with `mat b=J(1,1,0.05)`. A heteroskedastic-robust S statistic can be obtained by specifying `robust` along with `b0(b)`. To construct a weak-instruments-robust confidence interval, the user can simply conduct a grid search over the relevant range for β_1.[20]

Two options have been added to `ivreg2` for special handling of the GMM estimation process. The `wmatrix(`*matrix*`)` option allows the user to specify a weighting matrix rather than computing the optimal weighting matrix. Estimation with the `wmatrix()` option yields a possibly inefficient GMM estimator. `ivreg2` will use this inefficient estimator as the first-step GMM estimator in two-step EGMM when combined with the `gmm2s` option; otherwise, `ivreg2` reports this IGMM estimator.

The `smatrix(`*matrix*`)` option allows the user to directly specify the matrix `S`, the covariance matrix of orthogonality conditions. `ivreg2` will use this matrix in the calculation of the variance–covariance matrix of the estimator, the J statistic, and if the `gmm2s` option is specified, the two-step EGMM coefficients. The `smatrix()` option can be useful for guaranteeing a positive test statistic in user-specified GMM-distance tests as described in section 5.

As Ahn (1997) shows, Hansen's J test has an LM interpretation but can also be calculated as the result of a Wald test. This is an application of the Newey and West (1987a) results on the equivalence of LM, Wald, and GMM distance tests. In the context of an overidentified model, the J statistic will be identical to a Wald χ^2 test statistic from an exactly identified model in which more instruments are included as regressors as long as the same estimate of S is used in both estimated equations. As an example:

```
. use http://www.stata-press.com/data/imeus/griliches, clear
(Wages of Very Young Men, Zvi Griliches, J.Pol.Ec. 1976)
. qui ivreg2 lw (iq=med kww age), gmm2s
. di e(sargan)
102.10909
```

20. It is important to note that an Anderson–Rubin confidence region need not be finite nor connected. The test provided in `condivreg` (Moreira and Poi 2003, Mikusheva and Poi 2006) is uniformly most powerful in the situation where there is one endogenous regressor and i.i.d. errors. The Anderson–Rubin test provided by `ivreg2` is a simple and preferable alternative when errors are not i.i.d. or there is more than one endogenous regressor.

```
. mat S0 = e(S)
. qui ivreg2 lw med age (iq=kww), gmm2s smatrix(S0)
. test med age
 ( 1)  med = 0
 ( 2)  age = 0
           chi2(  2) =  102.11
         Prob > chi2 =    0.0000
. qui ivreg2 lw kww age (iq=med), gmm2s smatrix(S0)
. test kww age
 ( 1)  kww = 0
 ( 2)  age = 0
           chi2(  2) =  102.11
         Prob > chi2 =    0.0000
. qui ivreg2 lw med kww (iq=age), gmm2s smatrix(S0)
. test med kww
 ( 1)  med = 0
 ( 2)  kww = 0
           chi2(  2) =  102.11
         Prob > chi2 =    0.0000
```

9 RESET in the IV context

The `ivreset` command performs various flavors of RESET as adapted by Pesaran and Taylor (1999) and Pagan and Hall (1983) for IV estimation. RESET is sometimes called an omitted-variables test (as in official Stata's `ovtest`) but probably is best interpreted as a test of neglected nonlinearities in the choice of functional form (Wooldridge 2002, 124–125). Under the null hypothesis that there are no neglected nonlinearities, the residuals should be uncorrelated with low-order polynomials in $\widehat{y}$, where the $\widehat{y}$'s are predicted values of the dependent variable. In the `ivreset` implementation of the test, an equation of the form $y = X\beta + Y\gamma + v$ is estimated by IV, where the Ys are powers of $\widehat{y}$, the fitted value of the dependent variable y. Under the null hypothesis that there are no neglected nonlinearities and the equation is otherwise well specified, γ should not be significantly different from zero.

As Pesaran and Taylor (1999) and Pagan and Hall (1983) point out, however, RESET for an IV regression cannot use the standard IV predicted values $\widehat{y} \equiv X\widehat{\beta}$ because X includes endogenous regressors that are correlated with u. Instead, RESET must be implemented using "forecast values" of y that are functions of the instruments (exogenous variables) only. In the Pagan–Hall version of the test, the forecast values $\widehat{y}$ are the reduced-form predicted values of y, i.e., the predicted values from a regression of y on the instruments Z. In the Pesaran–Taylor version of the test, the forecast values $\widehat{y}$ are the "optimal forecast" values. The optimal forecast (predictor) $\widehat{y}$ is defined as $\widehat{X}\widehat{\beta}$, where $\widehat{\beta}$ is the IV estimate of the coefficients and $\widehat{X} \equiv [Z\widehat{\Pi}\ Z_2]$, i.e., the reduced-form predicted values of the endogenous regressors plus the exogenous regressors. If the equation is exactly identified, the optimal forecasts and reduced-form forecasts coincide, and the Pesaran–Taylor and Pagan–Hall tests are identical.

The `ivreset` test types vary according to the polynomial terms (square, cube, fourth power of $\widehat{y}$), the choice of forecast values (Pesaran–Taylor optimal forecasts or Pagan–Hall reduced-form forecasts), test statistic (Wald or GMM-distance), and large- versus small-sample statistic (χ^2 or F statistic). The test statistic is distributed with degrees of freedom equal to the number of polynomial terms. The default is the Pesaran–Taylor version using the square of the optimal forecast of y and a χ^2 Wald statistic with one degree of freedom.

If the original `ivreg2` estimation was heteroskedasticity-robust, cluster–robust, AC, or HAC, the reported RESET will be as well. The `ivreset` command can also be used after OLS regression with `regress` (see [R] **regress**) or `ivreg2` when there are no endogenous regressors. Then either a standard RESET using fitted values of y or a robust test corresponding to the specification of the original regression is reported.

We illustrate use of `ivreset` using a model fitted to the Griliches data:

```
. use http://fmwww.bc.edu/ec-p/data/hayashi/griliches76.dta
(Wages of Very Young Men, Zvi Griliches, J.Pol.Ec. 1976)
. quietly ivreg2 lw s expr tenure rns smsa (iq=med kww), robust
. ivreset
Ramsey/Pesaran-Taylor RESET test
Test uses square of fitted value of y (X-hat*beta-hat)
Ho: E(y|X) is linear in X
Wald test statistic:             Chi-sq(1) =  4.53   P-value = 0.0332
Test is heteroskedastic-robust
. ivreset, poly(4) rf small
Ramsey/Pagan-Hall RESET test
Test uses square, cube and 4th power of reduced form prediction of y
Ho: E(y|X) is linear in X
Wald test statistic:             F(3,748) =  1.72    P-value = 0.1616
Test is heteroskedastic-robust
```

The first `ivreset` takes all the defaults and corresponds to a second-order polynomial in $\widehat{y}$ with the Pesaran–Smith optimal forecast and a Wald χ^2 test statistic that rejects the null at better than 95%. The second uses a fourth-order polynomial and requests the Pagan–Hall reduced-form forecast with a Wald F statistic, falling short of the 90% level of significance.

10 A test for autocorrelated errors in the IV context

The `ivactest` command performs the Cumby and Huizinga (1992) generalization of a test proposed by Sargan (1988) for serial independence of the regression errors, which in turn generalizes the test proposed by Breusch and Godfrey (`estat bgodfrey`) applicable to OLS regressions. Sargan's extension of the Breusch–Godfrey test to the IV context, the serial correlation (SC) test, is described as a "general misspecification chi-squared statistic" by Pesaran and Taylor (1999, 260). The SC test statistic is based on the residuals of the IV regression and its conventional VCE. Cumby and Huizinga extend Sargan's test to cases in which the IV VCE was estimated as heteroskedasticity-robust, autocorrelation-robust, or HAC.

Cumby and Huizinga (1992) state that the null hypothesis of the test is "that the regression error is a moving average of known order $q \geq 0$ against the general alternative that autocorrelations of the regression error are nonzero at lags greater than q. The test ...is thus general enough to test the hypothesis that the regression error has no serial correlation ($q = 0$) or the null hypothesis that serial correlation in the regression error exists, but dies out at a known finite lag ($q > 0$)" (p. 185).

The Cumby–Huizinga test is especially attractive because it can be used in three frequently encountered cases where alternatives such as the Box–Pierce test ([TS] **wntestq**), Durbin's h test (`estat durbinalt`), and the Breusch–Godfrey test (`estat bgodfrey`) are not applicable. One of these cases is the presence of endogenous regressors, which renders each of these tests invalid. A second case involves the overlapping data commonly encountered in financial markets where the observation interval is shorter than the holding period, which requires the estimation of the induced moving-average (MA) process. The Cumby–Huizinga test avoids estimation of the MA process by using only the sample autocorrelations of the residuals and a consistent estimate of their asymptotic covariance matrix. The third case involves conditional heteroskedasticity of the regression error term, which is also handled without difficulty by the Cumby–Huizinga test.

If the prior estimation command estimated a VCE under the assumption of i.i.d. errors, the Cumby–Huizinga statistic becomes the Breusch–Godfrey statistic for the same number of autocorrelations and will return the same result as `estat bgodfrey`. That special case of the test was that proposed by Sargan in an unpublished working paper in 1976 (reprinted in Sargan 1988).

Two parameters may be specified in `ivactest`: `s()`, the number of lag orders to be tested, and `q()`, the lowest lag order to be tested.[21] By default, `ivactest` takes `s=1` and `q=0` and produces a test for AR(1). A test for AR(p) may be produced with `s=`p. Under the null hypothesis of serial independence for lags $q - (q + s)$, the Cumby–Huizinga test statistic is distributed χ^2 with s degrees of freedom.

We illustrated the use of `ivactest` in section 3.

11 Syntax

These syntax diagrams describe all the programs in the `ivreg2` suite, including those that have not been substantially modified since their documentation in Baum, Schaffer, and Stillman (2003).

21. If the previous command estimated a VCE under the assumption of i.i.d. errors, `q()` must be 0.

`ivreg2` *depvar* [*varlist1*] (*varlist2*=*varlist_iv*) [*if*] [*in*] [*weight*] [, `gmm` `gmm2s` `bw(`*#*|`auto)` `kernel(`*string*`)` `liml` `fuller(`*#*`)` `kclass(`*#*`)` `coviv` `cue` `cueinit(`*matrix*`)` `cueoptions(`*string*`)` `b0(`*matrix*`)` `robust` `cluster(`*varname*`)` `orthog(`*varlist_ex*`)` `endogtest(`*varlist_en*`)` `redundant(`*varlist_ex*`)` `partial(`*varlist_ex*`)` `small` `noconstant` `smatrix(`*matrix*`)` `wmatrix(`*matrix*`)` `first` `ffirst` `savefirst` `savefprefix(`*string*`)` `rf` `saverf` `saverfprefix(`*string*`)` `nocollin` `noid` `level(`*#*`)` `noheader` `nofooter` `eform(`*string*`)` `depname(`*varname*`)` `plus`]

`overid` [, `chi2` `dfr` `f` `all` `depvar(`*varname*`)`]

`ivhettest` [*varlist*] [, `ivlev` `ivsq` `fitlev` `fitsq` `ph` `phnorm` `nr2` `bpg` `all`]

`ivendog` [*varlist*]

`ivreset` [, `polynomial(`*#*`)` `rform` `cstat` `small`]

`ivactest` [, `s(`*#*`)` `q(`*#*`)`]

12 A summary of ivreg2 estimation options

The version of `ivreg2` accompanying this paper uses a different syntax for specifying the type of estimator to be employed. In previous versions of the software (Baum, Schaffer, and Stillman 2003; 2004; 2005), the `gmm` option implied a heteroskedasticity-robust estimator. When the `gmm` option was combined with the `bw()` option, estimates were autocorrelation-robust but not heteroskedasticity-robust. This version of `ivreg2` uses a new taxonomy of estimation options, summarized below. The `gmm2s` option by itself produces the IV/2SLS estimator, as described in section 2.5. One of the options—`robust`, `cluster()`, or `bw()`—must be added to generate two-step EGMM estimates.

Table 1 summarizes the estimator and the properties of its point and interval estimates for each combination of estimation options.

Table 1: Summary of ivreg2 estimation options

Estimator option	Covariance matrix option(s): (none)	Covariance matrix option(s): robust, cluster(), bw(), kernel()
(none)	IV/2SLS SEs consistent under homoskedasticity	IV/2SLS with robust SEs
liml	LIML SEs consistent under homoskedasticity	LIML with robust SEs
gmm2s	IV/2SLS SEs consistent under homoskedasticity	Two-step GMM with robust SEs
cue	LIML SEs consistent under homoskedasticity	CUE/GMM with robust SEs
kclass()	k-class estimator SEs consistent under homoskedasticity	k-class estimator with robust SEs
wmatrix()	possibly IGMM SEs consistent under homoskedasticity	IGMM with robust SEs
gmm2s + wmatrix()	Two-step GMM with user-specified first step SEs consistent under homoskedasticity	Two-step GMM with robust SEs

12.1 ivreg2 versus ivregress

Stata's official ivregress command, available in Stata 10 and later, provides an LIML and GMM estimator in addition to two-stage least squares. The GMM estimator can produce HAC estimates, as discussed in section 3 but cannot produce AC estimates. The ivregress command does not support the general k-class estimator nor GMM/CUE but provides an "iterative GMM" estimator. Overidentification tests and first-stage statistics are available as estat subcommands. ivreg2's ability to partial-out regressors with the partial() option is not available in ivregress.

Several tests performed by ivreg2 are also not available with ivregress. These include the GMM distance tests of endogeneity/exogeneity discussed in section 5, the general underidentification/weak-identification test of Kleibergen and Paap (2006) discussed in section 7, and tests that permit inference robust to weak instruments. In diagnosing potentially weak instruments, ivreg2's ability to save the first-stage regressions is also unique.

The default behavior of `ivregress gmm` with the `vce(robust)` option produces coefficients that match those of `ivreg2, gmm2s robust` but with different standard errors. Whereas `ivreg2` uses the expression for the VCE of the efficient GMM estimator (6), `ivregress gmm` calculates the VCE as if the estimator was not efficient using (4) from a new estimate of the asymptotic covariance matrix S based on the second-step GMM residuals. To replicate this behavior of `ivregress gmm` and generate identical standard errors, the `wmatrix()` option of `ivreg2` can be used

```
. webuse abdata, clear
. qui ivreg2 n (w = k ys), gmm2s robust
. mat W_GMM2S = e(W)
. ivregress gmm n (w =k ys), vce(robust)
. ivreg2 n (w = k ys), wmatrix(W_GMM2S) robust
```

Both programs' methodology for calculation of the variance–covariance matrix yield consistent estimates.

13 Acknowledgments

We are grateful to many members of the Stata user community who have assisted in the identification of useful features in the `ivreg2` suite and helped identify problems with the programs. We thank (without implicating) Austin Nichols for his suggestions on this draft, Frank Kleibergen for discussions about testing for identification, and Manuel Arellano and Graham Elliott for helpful discussions of GMM/CUE. Some portions of the discussion of weak instruments in section 7.3 are taken from chapter 8 of Baum (2006).

14 References

Ahn, S. C. 1997. Orthogonality tests in linear models. *Oxford Bulletin of Economics and Statistics* 59: 183–186.

Anderson, T. W. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22: 327–351.

———. 1984. *Introduction to Multivariate Statistical Analysis*. 2nd ed. New York: Wiley.

Anderson, T. W., and H. Rubin. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20: 46–63.

Andrews, D. W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817–858.

Baum, C. F. 2006. *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press.

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.

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.

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski. 1999. Redundancy of moment conditions. *Journal of Econometrics* 91: 89–111.

Chernozhukov, V., and C. Hansen. 2005. The reduced form: A simple approach to inference with weak instruments. Working Paper, University of Chicago, Graduate School of Business.

Cragg, J. G., and S. G. Donald. 1993. Testing identifiability and specification in instrumental variables models. *Econometric Theory* 9: 222–240.

Cumby, R. E., and J. Huizinga. 1992. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica* 60: 185–195.

Cushing, M., and M. McGarvey. 1999. Covariance matrix estimation. In *Generalized Methods of Moments Estimation*, ed. L. Matyas. Cambridge: Cambridge University Press.

Davidson, R., and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. 2nd ed. New York: Oxford University Press.

Dufour, J.-M. 2003. Identification, weak instruments and statistical inference in econometrics. Working Paper 2003s-49, CIRANO.

Frisch, R., and F. V. Waugh. 1933. Partial time regressions as compared with individual trends. *Econometrica* 1: 387–401.

Fuller, W. A. 1977. Some properties of a modification of the limited information estimator. *Econometrica* 45: 939–53.

Greene, W. H. 2008. *Econometric Analysis*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Hahn, J., and J. Hausman. 2002. Notes on bias in estimators for simultaneous equation models. *Economics Letters* 75: 237–41.

Hahn, J., J. Hausman, and G. Kuersteiner. 2004. Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *Econometrics Journal* 7: 272–306.

Hall, A. R. 2005. *Generalized Method of Moments*. Oxford: Oxford University Press.

Hall, A. R., and F. P. M. Peixe. 2003. A consistent method for the selection of relevant instruments. *Econometric Reviews* 22: 269–287.

Hall, A. R., G. D. Rudebusch, and D. W. Wilcox. 1996. Judging instrument relevance in instrumental variables estimation. *International Economic Review* 37: 283–298.

Hansen, L. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.

Hansen, L., J. Heaton, and A. Yaron. 1996. Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14: 262–280.

Hayashi, F. 2000. *Econometrics*. Princeton, NJ: Princeton University Press.

Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular-value decomposition. *Journal of Econometrics* 127: 97–126.

Kleibergen, F., and M. Schaffer. 2007. ranktest: Stata module to test the rank of a matrix using the Kleibergen–Paap rk statistic. Boston College Department of Economics, Statistical Software Components S456865. Downloadable from http://ideas.repec.org/c/boc/bocode/s456865.html.

Lovell, M. 1963. Seasonal adjustment of economic time series. *Journal of the American Statistical Association* 58: 993–1010.

Mikusheva, A., and B. P. Poi. 2006. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* 6: 335–347.

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.

Newey, W. K., and K. D. West. 1987a. Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28: 777–787.

———. 1987b. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–708.

———. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61: 631–653.

Pagan, A. R., and D. Hall. 1983. Diagnostic tests as residual analysis. *Econometric Reviews* 2: 159–218.

Pesaran, M. H., and L. W. Taylor. 1999. Diagnostics for IV regressions. *Oxford Bulletin of Economics & Statistics* 61: 255–281.

Poskitt, D., and C. Skeels. 2002. Assessing instrumental variable relevance: An alternative measure and some exact finite sample theory. Technical report, Monash University and University of Melbourne.

Sargan, J. 1988. Testing for misspecification after estimation using instrumental variables. In *Contributions to econometrics: John Denis Sargan*, ed. E. Maasoumi, vol. 1. Cambridge: Cambridge University Press.

Staiger, D., and J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65: 557–86.

Stock, J., and M. Watson. 2003. *Introduction to Econometrics*. Reading, MA: Addison–Wesley.

Stock, J. H., and J. H. Wright. 2000. GMM with weak identification. *Econometrica* 68: 1055–1096.

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.

Stock, J. H., and M. Yogo. 2005. Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D. W. K. Andrews and J. H. Stock, 80–108. Cambridge: Cambridge University Press.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

———. 2003. *Introductory Econometrics: A Modern Approach*. 2nd ed. New York: Thomson Learning.

About the authors

Christopher F. Baum is associate professor of economics at Boston College. He is an associate editor of *Computational Economics*, the *Stata Journal*, and the *Journal of Statistical Software*. Baum founded and manages the Boston College Statistical Software Components (SSC) archive at RePEc (*http://repec.org*). His recent research has focused on the effects of uncertainty on international trade flows, bank lending behavior, and firms' cash holdings.

Mark E. Schaffer is professor of economics and Director of the Centre for Economic Reform and Transformation (CERT) at Heriot–Watt University, Edinburgh, Scotland. He is also a Research Fellow at the Centre for Economic Policy Research (CEPR), the Institute for the Study of Labor (IZA), and the William Davidson Institute. His research interests include various aspects of firm and household behavior in the transition countries of Eastern Europe, the former USSR, and East Asia.

Steven Stillman is a senior fellow at Motu Economic and Public Policy Research Trust. He is also an affiliated research fellow at the Institute for the Study of Labour (IZA), the Centre

for Research and Analysis of Migration (CReAM), and the William Davidson Institute. Recent work includes analyzing the effect of suffering injuries on future labor market outcomes in New Zealand, the impact of migration to New Zealand on the income and health of Pacific Islanders, and the effect of economic shocks in Russia on household consumption, individual nutrition, and overall living standards.