Lectures on Evaluation Methods **Impact Evaluation Network**

Guido Imbens **October 2010, Miami**

Weak Instruments and Many Instruments

1. Introduction

In recent years a literature has emerged that has raised concerns with the quality of inferences based on conventional methods such as Two Stage Least Squares (TSLS) and Limited Information Maximum Likelihood (LIML) in instrumental variables settings when the instrument(s) is/are only weakly correlated with the endogenous regressor(s). Although earlier work had already established the poor quality of conventional normal approximations with weak or irrelevant instruments, the recent literature has been motivated by empirical work where *ex post* conventional large sample approximations were found to be misleading. The recent literature has aimed at developing better estimators and more reliable methods for inference.

There are two aspects of the problem. In the just-identified case (with the number of instruments equal to the number of endogenous regressors), or with low degrees of over-identification, the focus has largely been on the construction of confidence intervals that have good coverage properties even if the instruments are weak. Even with very weak, or completely irrelevant, instruments, conventional methods are rarely substantively misleading, unless the degree of endogeneity is higher than one typically encounters in studies using cross-section data. Conventional TSLS or LIML confidence intervals tend to be wide when the instrument is very weak, even if those intervals do not have the correct nominal coverage for all parts of the parameter space. In this case better estimators are generally not available. Improved methods for confidence intervals based on inverting test statistics have been developed although these do not have the simple form of an estimate plus or minus a constant times a standard error.

The second case of interest is that with a high degree of over-identification. These settings

often arise by interacting a set of basic instruments with exogenous covariates in order to improve precision. If there are many (weak) instruments, standard estimators can be severely biased, and conventional methods for inference can be misleading. In particular TSLS has been found to have very poor properties in these settings. Bootstrapping does not solve these problems. LIML is generally much better, although conventional LIML standard errors are too small. A simple to implement proportional adjustment to the LIML standard errors based on the Bekker many-instrument asymptotics or the Chamberlain-Imbens random coefficients argument appears to lead to substantial improvements in coverage rates.

2. Motivation

Much of the recent literature is motivated by a study by Angrist and Krueger (1991, AK). Subsequently Bound, Jaeger and Baker (1996, BJB) showed that for some specifications AK employed normal approximations were not appropriate despite very large sample sizes (over 300,000 observations).

2.1 The Angrist-Krueger Study

AK were interested in estimating the returns to years of education. Their basic specification is:

$$Y_i = \alpha + \beta \cdot E_i + \varepsilon_i,$$

where Y_i is log (yearly) earnings and E_i is years of education. Their concern, following a long literature in economics, e.g., Griliches, (1977), Card (2001), is that years of schooling may be endogenous, with pre-schooling levels of ability affecting both schooling choices and earnings given education levels. In an ingenuous attemp to address the endogeneity problem AK exploit variation in schooling levels that arise from differential impacts of compulsory schooling laws. School districts typically require a student to have turned six by January 1st of the year the student enters school. Since individuals are required to stay in school till they turn sixteen, individual born in the first quarter have lower required minimum schooling levels than individuals born in the last quarter. The cutoff dates and minimum

school dropout age differ a little bit by state and over time, so the full picture is more complicated but the basic point is that the compulsory schooling laws generate variation in schooling levels by quarter of birth that AK exploit.

One can argue that a more natural analysis of such data would be as a Regression Discontinuity (RD) design, where we focus on comparisons of individuals born close to the cutoff date. We will discuss such designs in a later lecture. However, in the census only quarter of birth is observed, not the actual date, so there is in fact little that can be done with the RD approach beyond what AK do. In addition, there are substantive arguments why quarter of birth need not be a valid instrument (e.g., seasonal patterns in births, or differential impacts of education by age at entering school). AK discuss many of the potential concerns. See also Bound, Jaeger and Baker (1996). We do not discuss these concerns here further.

Table 1 shows average years of education and average log earnings for individual born in the first and fourth quarter, using the 1990 census. This is a subset of the AK data.

Table 1: Summary Statistics Subset of AK Data

Variable	1st Quarter	4th Quarter	difference
Year of Education	12.688	12.840	0.151
Log Earnings	5.892	5.905	0.014
ratio			0.089

The sample size is 162,487. The last column gives the difference between the averages by quarter, and the last row the ratio of the difference in averages. The last number is the Wald estimate of the returns to education based on these data:

$$\hat{\beta} = \frac{\overline{Y}_4 - \overline{Y}_1}{\overline{E}_4 - \overline{E}_1} = 0.0893 \quad (0.0105),$$

where $\overline{Y}_t$ and $\overline{E}_t$ are the average level of log earnings and years of education for individuals born in the t-th quarter. This is also equal to the Two-Stage-Least-Squares (TSLS) and Limited-Information-Maximum-Likelihood (LIML) estimates because there is only a single instrument and a single endogenous regressor. The standard error here is based on the delta method and asymptotic joint normality of the numerator and denominator.

AK also present estimates based on additional instruments. They take the basic instrument and interact it with 50 state and 10 year of birth dummies. Here we take this a bit further, and following Chamberlain and Imbens (2004) we interact the single binary instrument with state times year of birth dummies to get 500 instruments. Also including the state times year of birth dummies as exogenous covariates leads to the following model:

$$Y_i = X_i'\beta + \varepsilon_i, \qquad \mathbb{E}[Z_i \cdot \varepsilon_i] = 0,$$

where X_i is the 501-dimensional vector with the 500 state/year dummies and years of education, and Z_i is the 1000-dimensional vector with 500 state/year dummies and the 500 state/year dummies multiplying the indicator for the fourth quarter of birth. Let $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{Z}$ be the $N \times 1$ vector of log earnings, the $N \times 501$ matrix of exogenous and endogenous regressors, and the $N \times 1000$ matrix of exogenous covariates and excluded instruments. Let $\mathbf{P}_Z = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$ be the projection matrix and $\mathbf{M}_Z = \mathbf{I} - \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$. The TSLS estimator for β is then

$$\hat{\beta}_{\text{TSLS}} = \left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{P}_Z\mathbf{Y}\right) = \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y}\right).$$

The conventional TSLS variance estimator is

$$V_{\text{TSLS}} = \hat{\sigma}^2 \cdot \left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1} = \hat{\sigma}^2 \cdot \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - X_i\hat{\beta}_{\text{TSLS}}\right)^2.$$

For these data this leads to

$$\hat{\beta}_{\rm TSLS} = 0.073 \quad (0.008).$$

The LIML estimator is based on maximization of the log likelihood function

$$L(\beta,\pi,\Omega) = \sum_{i=1}^{N} \left(-\frac{1}{2}\ln|\Omega| - \frac{1}{2} \left(\begin{array}{c} Y_i - \beta Z_i'\pi \\ E_i - Z_i'\pi \end{array} \right)' \Omega^{-1} \left(\begin{array}{c} Y_i - X_i'\beta \\ E_i - Z_i'\pi \end{array} \right) \right),$$

where Ω is the reduced form variance-covariance matrix. One can also write the liml estimator as a k-class estimator,

$$\hat{\beta}_{\rm liml} = \left(\mathbf{X}'\left(\mathbf{I} - k_{\rm liml}\cdot\mathbf{M}_Z\right)\mathbf{X}\right)^{-1}\left(\mathbf{X}'\left(\mathbf{I} - k_{\rm liml}\cdot\mathbf{M}_Z\right)\mathbf{Y}\right),$$

where $k_{\rm liml}$ is a minimum eigenvalue:

$$k_{\rm liml} = \min_{\beta} \frac{(\mathbf{Y}-\mathbf{X}\beta)'(\mathbf{Y}-\mathbf{X}\beta)}{(\mathbf{Y}-\mathbf{X}\beta)'\mathbf{M}_Z(\mathbf{Y}-\mathbf{X}\beta)}.$$

(TSLS also fits into this class with $k_{\rm tsls} = 1$.)

The conventional liml variance estimator is

$$V_{\rm liml} = \hat{\sigma}^2 \cdot \left(\mathbf{X}'\left(\mathbf{I} - k_{\rm liml}\cdot\mathbf{M}_Z\right)\mathbf{X}\right)^{-1}.$$

For this subset of the AK data we find, for the coefficient on years of education, with standard error,

$$\hat{\beta}_{\rm LIML} = 0.095 \quad (0.017).$$

In large samples the LIML and TSLS are equivalent under homoskedasticity.

2.2 The Bound-Jaeger-Baker critique

BJB found that are potential problems with the AK results. They suggested that despite the large samples used by AK large sample normal approximations may be very poor. The

reason is that the instruments are only very weakly correlated with the endogenous regressor. The most striking evidence for this is based on the following calculations, that are based on a suggestion by Alan Krueger. Take the AK data and re-calculate their estimates after replacing the actual quarter of birth dummies by random indicators with the same marginal distribution. In principle this means that the standard (gaussian) large sample approximations for TSLS and LIML are invalid since they rely on non-zero correlations between the instruments and the endogenous regressor. Doing these calculations once for the single and 500 instrument case, for both TSLS and LIML, leads to the results in Table 2.

Table 2: Real and Random QOB Estimates

	Single Instrument		500 Instruments			
			TSLS		LIML	
Real QOB	0.089	(0.011)	0.073	(0.008)	0.095	(0.017) [0.037]
Random QOB	0.181	(0.193)	0.059	(0.009)	-0.134	(0.065) [0.251]

With the single instrument the results are not so disconcertening. Although the confidence interval is obviously not valid, it is wide, and few researchers would be misled by the results. With many instruments the results are much more troubling. Although the instrument contains no information, the TSLS (and to a lesser extent LIML with the conventional standard errors) results suggest that the instruments can be used to infer precisely what the returns to education are. These results have provided the motivation for the recent weak instrument literature. Note that there is an earlier literature on small sample properies of IV estimators, e.g., Phillips (1984) Rothenberg (1984), but it is the BJB findings that got the attention of researchers doing empirical work.

2.3 Simulations with Weak Instruments and Varying Degrees of Endogeneity

Here we provide slightly more systematic simulation evidence of the weak instrument problems in the AK setting. We create 10,000 artificial data sets, all of size 160,000, designed to mimic the key features of the AK data. In each of these data sets half the units have quarter of birth (denoted by Q_i) equal to 0 and 1 respectively. Then we draw the two reduced form residuals ν_i and η_i from a joint normal distribution

$$\begin{pmatrix} \nu_i \\ \eta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.446 & \rho \cdot \sqrt{0.446} \cdot \sqrt{10.071} \\ \rho \cdot \sqrt{0.446} \cdot \sqrt{10.071} & 10.071 \end{pmatrix} \right).$$

The variances of the reduced form errors mimic those in the AK data. The correlation between the reduced form residuals in the AK data is 0.318. The implied OLS coefficient is $\rho \cdot \sqrt{0.446}/\sqrt{10.071}$. Then years of education is equal to

$$E_i = 12.688 + 0.151 \cdot Q_i + \eta_i,$$

and log earnings is equal to

$$Y_i = 5.892 + 0.014 \cdot Q_i + \nu_i.$$

Now we calculate the IV estimator and its standard error, using either the actual qob variable or a random qob variable as the instrument. We are interested in the size of tests of the null that coefficient on years of education is equal to $0.089 = 0.014/0.151$. We base the test on the t-statistic. Thus we reject the null if the ratio of the point estimate minus 0.089 and the standard error is greater than 1.96 in absolute value. We repeat this for 12 different values of the reduced form error correlation. In Table 3 we report the proportion of rejections and the median and 0.10 quantile of the width of the estimated 95% confidence intervals.

Table 3: Coverage Rates of Conv. TSLS CI by Degree of Endogeneity

ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
implied OLS	0.00	0.02	0.04	0.06	0.08	0.11	0.13	0.15	0.17	0.19	0.20	0.21
Real QOB	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.95	0.95	0.95	0.95
Med Width 95% CI	0.09	0.09	0.09	0.08	0.08	0.08	0.07	0.07	0.06	0.05	0.05	0.05
0.10 quant Width	0.08	0.08	0.08	0.07	0.07	0.07	0.06	0.06	0.05	0.04	0.04	0.04
Random QOB	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.92	0.82	0.53
Med Width 95% CI	1.82	1.81	1.78	1.73	1.66	1.57	1.45	1.30	1.09	0.79	0.57	0.26
0.10 quant Width	0.55	0.55	0.5403	0.53	0.51	0.48	0.42	0.40	0.33	0.24	0.17	0.08

In this example, unless the reduced form correlations are very high, e.g., at least 0.95, with irrelevant the conventional confidence intervals are wide and have good coverage. The amount of endogeneity that would be required for the conventional confidence intervals to be misleading is higher than one typically encounters in cross-section settings. It is likely that these results extend to cases with a low degree of over-identification, using either TSLS, or preferably LIML. Put differently, although formally conventional confidence intervals are not valid uniformly over the parameter space (e.g., Dufour, 1997), there are no examples we are aware of where they have substantively misleading in just-identified examples. This in contrast to the case with many weak instruments where especially TSLS can be misleading in empirically relevant settings.

3. Weak Instruments

Here we discuss the weak instrument problem in the case of a single instrument, a single endogenous regressor, and no additional exogenous regressors beyond the intercept. More generally the qualitative features of these results by and large apply to the case with a few weak instruments. We consider the model

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \varepsilon_i,$$

$$X_i = \pi_0 + \pi_1 \cdot Z_i + \eta_i,$$

with $(\varepsilon_i, \eta_i) \perp\!\!\!\perp Z_i$, and jointly normal with covariance matrix Σ. (The normality is mainly for some of the exact results, and it does not play an important role.) The reduced form for the first equation is

$$Y_i = \alpha_0 + \alpha_1 \cdot Z_i + \nu_i,$$

where the parameter of interest is $\beta_1 = \alpha_1/\pi_1$. Let

$$\Omega = \mathbb{E}\left[\begin{pmatrix} \nu_i \\ \eta_i \end{pmatrix} \cdot \begin{pmatrix} \nu_i \\ \eta_i \end{pmatrix}' \right], \qquad \text{and} \quad \Sigma = \mathbb{E}\left[\begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix}' \right],$$

be the covariance matrix of the reduced form and stuctural disturbances respectively. Many of the formal results in the literature are for the case of known Ω, and normal disturbances. This is largely innocuous, as Ω can be precisely estimated in typical data sets. Note that this it not the same as assuming that Σ is known, which is not innocuous since it depends on Ω and β, and cannot be precisely estimated in settings with weak instruments

$$\Sigma = \begin{pmatrix} \Omega_{11} - 2\beta\Omega_{12} + \beta^2\Omega_{22} & \Omega_{12} - \beta\Omega_{22} \\ \Omega_{12} - \beta\Omega_{22} & \Omega_{22} \end{pmatrix}.$$

The standard estimator for β_1 is

$$\hat{\beta}_1^{\rm IV} = \frac{\frac{1}{N}\sum_{i=1}^N \left(Y_i - \overline{Y}\right)\left(Z_i - \overline{Z}\right)}{\frac{1}{N}\sum_{i=1}^N \left(X_i - \overline{X}\right)\left(Z_i - \overline{Z}\right)},$$

where $\overline{Y} = \sum_i Y_i/N$, and similarly for $\overline{X}$ and $\overline{Z}$.

A simple interpretation of the weak instrument is that with the concentration parameter

$$\lambda = \pi_1^2 \cdot \sum_{i=1}^N \left(Z_i - \overline{Z}\right)^2 \Big/ \sigma_\eta^2.$$

close to zero, both the covariance in the numerator and the covariance in the denominator are close to zero. In reasonably large samples both are well approximated by normal distributions:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(Y_i-\overline{Y}\right)\left(Z_i-\overline{Z}\right)-\mathrm{Cov}(Y_i,Z_i)\right)\approx\mathcal{N}\left(0,V(Y_i\cdot Z_i)\right),$$

and

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(X_i-\overline{X}\right)\left(Z_i-\overline{Z}\right)-\mathrm{Cov}(X_i,Z_I)\right)\approx\mathcal{N}\left(0,V(X_i\cdot Z_i)\right).$$

These two normal approximations tend to be accurate in applications with reasonable sample sizes, irrespective of the population values of the covariances. If $\pi_1 \neq 0$, as the sample size gets large, then the ratio will eventually be well approximated by a normal distribution as well. However, if $\mathrm{Cov}(X_i, Z_i) \approx 0$, the ratio may be better approximated by a Cauchy distribution, as the ratio of two normals centered close to zero.

The weak instrument literature is concerned with inference for β_1 when the concentration parameter λ is too close to zero for the normal approximation to the ratio to be accurate.

Staiger and Stock (1997) formalize the problem by investigating the distribution of the standard IV estimator under an alternative asymptotic approximation. The standard asymptotics (strong instrument asymptotics in the Staiger-Stock terminology) is based on fixed parameters and the sample size getting large. In their alternative asymptotic sequence Staiger and Stock model π_1 as a function of the sample size, $\pi_{1N} = c/\sqrt{N}$, so that the concentration parameter converges to a constant:

$$\lambda \longrightarrow c^2 \cdot V(Z_i).$$

SS then compare coverage properties of various confidence intervals under this (weak instrument) asymptotic sequence.

The importance of the Staiger-Stock approach is not in the specific sequence. The concern is more that if a particular confidence interval does not have the appropriate coverage

asymptotically under the Staiger-Stock asymptotics, then there are values of the (nuisance) parameters in a potentially important part of the parameter space (namely around $\pi_i = 0$) such that the actual coverage is substantially away from the nominal coverage for any sample size. More recently the issue has therefore been reformulated as requiring confidence intervals to have asymptotically the correct coverage probabilities uniformly in the parameter space. See for a discussion from this perspective Mikusheva (2007). For estimation this perspective is not helpful: there cannot be estimators that are consistent for β^* uniformly in the parameter space, because if $\pi_1 = 0$, there are no consistent estimators for β_1. However, for testing there are generally confidence intervals that are uniformly valid, but they are not of the conventional form, that is, a point estimate plus or minus a constant times a standard error.

3.1 Tests and Confidence Intervals in the Just-identified Case

Let the instrument $\tilde{Z}_i = Z_i - \overline{Z}$ be measured in deviations from its mean. Then define the statistic

$$S(\beta_1) = \frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i \cdot (Y_i - \beta_1 \cdot X_i)\,.$$

Then, under the null hypothesis that $\beta_1 = \beta_1^*$, and conditional on the instruments, the statistic $\sqrt{N} \cdot S(\beta_1^*)$ has an exact normal distribution

$$\sqrt{N} \cdot S(\beta_1^*) \sim \mathcal{N}\left(0, \sum_{i=1}^{N} \tilde{Z}_i^2 \cdot \sigma_\varepsilon^2\right).$$

Importantly, this result does not depend on the strength of the instrument, that is, on the correlation between the instrument and the endogenous regressor. Anderson and Rubin (1949, AR) propose basing tests for the null hypothesis

$$H_0: \quad \beta_1 = \beta_1^0, \qquad \text{against the alternative hypothesis} \quad H_a: \quad \beta_1 \neq \beta_1^0,$$

on this idea, through the statistic

$$\text{AR}\left(\beta_1^0\right) = \frac{N \cdot S(\beta_1^0)^2}{\sum_{i=1}^N \tilde{Z}_i^2} \cdot \left(\begin{pmatrix} 1 & -\beta_1^0 \end{pmatrix} \Omega \begin{pmatrix} 1 \\ -\beta_1^0 \end{pmatrix} \right)^{-1}.$$

Under the null hypothesis this statistic has an exact chi-squared distribution with degrees of freedom equal to one. In practice, of course, one does not know the reduced form covariance matrix Ω, but substituting an estimated version of this matrix based on the average of the estimated reduced form residuals does not affect the large sample properties of the test.

A confidence interval can be based on this test statistic by inverting it. For example, for a 95% confidence interval for β_1, we would get

$$\text{CI}_{0.95}^{\beta_1} = \left\{ \beta_1 \,|\text{AR}(\beta_1) \leq 3.84 \right\}.$$

Note that this AR confidence interval cannot be empty, because at the standard IV estimator $\hat{\beta}_1^{\text{IV}}$ we have $\text{AR}(\hat{\beta}_1^{\text{IV}}) = 0$, and thus $\hat{\beta}_1^{\text{IV}}$ is always in the confidence interval. The confidence interval can be equal to the entire real line, if the correlation between the endogenous regressor and the instrument is close to zero. This is not surprising: in order to be valid even if $\pi_1 = 0$, the confidence interval must include all real values with probability 0.95.

3.3 Tests and Confidence Intervals in the Over-identified Case

The second case of interest is that with a single endogenous regressor and multiple instruments. We deal separately with the case where there are many (similar) instrument, so this really concerns the case where the instruments are qualitatively different. Let the number of instrumens be equal to K, so that the reduced form is

$$X_i = \pi_0 + \pi_1' Z_i + \eta_i,$$

with Z_i a k-dimensional column vector. There is still only a single endogenous regressor, and no exogenous regressors beyond the intercept. All the results generalize to the case with additional exogenous covariates at the expense of additional notatio. The AR approach can

be extended easily to this over-identified case, because the statistic $\sqrt{N}\cdot S(\beta_1^*)$ still has a normal distribution, but now a multivariate normal distribution. Hence one can base tests on the AR statistic

$$\text{AR}\left(\beta_1^0\right) = N\cdot S\left(\beta_1^0\right)'\left(\sum_{i=1}^N \tilde{Z}_i\cdot\tilde{Z}_i'\right)^{-1} S\left(\beta_1^0\right)\cdot\left(\left(\begin{array}{cc}1 & -\beta_1^0\end{array}\right)\Omega\left(\begin{array}{c}1\\ -\beta_1^0\end{array}\right)\right)^{-1}.$$

Under the same conditions as before this has an exact chi-squared distribution with degrees of freedom equal to the number of instruments, k. A practical problem arises if we wish to construct confidence intervals based on this statistic. Suppose we construct a confidence interval, analogously to the just-identified case, as

$$\text{CI}^{\beta_1}_{0.95} = \left\{\beta_1\,\middle|\,\text{AR}(\beta_1)\leq\mathcal{X}^2_{0.95}(K)\right\},$$

where $\mathcal{X}^2_{0.95}(k)$ is the 0.95 quantile of the chi-squared distribution with degrees of freedom equal to k. The problem is that this confidence interval can be empty. The interpretation is that the test does not only test whether $\beta_1=\beta_1^0$, but also tests whether the instruments are valid. However, one generally may not want to combine those hypotheses.

Kleibergen (2002) modifies the AR statistic and confidence interval construction. Instead of the statistic $S(\beta_1)$, he considers a statistic that looks at the correlation between a particular linear combination of the instruments (namely the estimated endogenous regressor) and the residual:

$$\tilde{S}\left(\beta_1^0\right) = \frac{1}{N}\sum_{i=1}^N\left(\tilde{Z}_i'\hat{\pi}_1(\beta_1^0)\right)\cdot\left(Y_i-\beta_1^0\cdot X_i\right),$$

where $\hat{\pi}$ is the maximum likelihood estimator for π_1 under the restriction $\beta_1=\beta_1^0$. The test is then based on the statistic

$$K\left(\beta_1^0\right) = \frac{N\cdot S(\beta_1^0)^2}{\sum_{i=1}^N\tilde{Z}_i^2}\cdot\left(\left(\begin{array}{cc}1 & -\beta_1^0\end{array}\right)\Omega\left(\begin{array}{c}1\\ -\beta_1^0\end{array}\right)\right)^{-1}.$$

This statistic has no longer an exact chi-squared distribution, but in large samples it still has an approximate chi-square distribution with degrees of freedom equal to one. Hence the test is straightforward to implement using standard methods.

Moreira (2003) proposes a method for adjusting the critical values that applies to a number of tests, including the Kleibergen test. His idea is to focus on *similar* tests, test that have the same rejection probability for all values of the nuisance parameter. The nuisance parameter is here the vector of reduced form coefficients π, since we assume the residual covariance matrix is known. The way to adjust the critical values is to consider the distribution of a statistic such as the Kleibergen statistic conditional on a complete sufficient statistic for the nuisance parameter. In this setting a complete sufficient statistic is readily available in the form of the maximum likelihood estimator under the null, $\hat{\pi}_1(\beta_1^0)$. Moreira's preferred test is based on the likelihood ratio. Let

$$LR\left(\beta_1^0\right) = 2 \cdot \left(L\left(\hat{\beta}_1, \hat{\pi}\right) - L\left(\beta_1^0, \hat{\pi}(\beta_1^0)\right)\right),$$

be the likelihood ratio. Then let $c_{LR}(p, 0.95)$, be the 0.95 quantile of the distribution of $LR(\beta_1^0)$ under the null hypothesis, conditional on $\hat{\pi}(\beta_1^0) = p$. The proposed test is to reject the null hypothesis at the 5% level if

$$LR\left(\beta_1^0\right) > c_{LR}(\hat{\pi}(\beta_1^0), 0.95),$$

where conventional test would use critical values from a chi-squared distribution with a single degree of freedom. This test can then be converted to construct a 95% confidence intervals. Calculation of the (large sample) critical values is simplified by the fact that they only depend on the number of instruments k, and a scaled version of the $\hat{\pi}(\beta_1^0)$. Tabulations of these critical values are in Moreira (2003) and have been programmed in STATA (See Moreira's website).

3.4 Conditioning on the First Stage

The AR, Kleibergen and Moreira proposals for confidence intervals are asymptotically

valid irrespective of the strength of the first stage (the value of π_1). However, they are not valid if one first inspects the first stage, and conditional on the strength of that, decides to proceed. Specifically, if in practice one first inspects the first stage, and decide to abandon the project if the first stage F-statistic is less than some fixed value, and otherwise proceed by calculating an AR, Kleibergen or Moreira confidence interval, the large sample coverage probabilities would not necessarily be the nominal ones. In practice researchers do tend to inspect and report the strength of the first stage. This is particularly true in recent instrumental variables literature where researchers argue extensively for the validity of the instrumental variables assumption. This typically involves detailed arguments supporting the alleged mechanism that leads to the correlation between the endogenous regressor and the instruments. For example, Section I in AK (page 981-994) is entirely devoted to discussing the reasons and evidence for the relation between their instruments (quarter of birth) and years of education. In such cases inference conditional on this may be more appropriate.

Chioda and Jansson (2006) propose a clever alternative way to construct a confidence interval that is valid conditional on the strength of the first stage. Their proposed confidence interval is based on inverting a test statistic similar to the AR statistic. It has a non-standard distribution conditional on the strength of the first stage, and they suggest a procedure that involves numerically approximating the critical values. A caveat is that because the first stage F-statistic, or the first stage estimates are not ancillary, conditioning on them involves loss of information, and as a result the Chioda-Jansson confidence intervals are wider than confidence intervals that are not valid conditional on the first stage.

4. Many Weak Instruments

In this section we discuss the case with many weak instruments. The problem is both the bias in the standard estimators, and the misleadingly small standard errors based on conventional procedures, leading to poor coverage rates for standard confidence intervals in many situations. The earlier simulations showed that especially TSLS, and to a much lesser extent LIML, have poor properties in this case. Note first that resampling methods such as bootstrapping do not solve these problems. In fact, if one uses the standard bootstrap with

TSLS in the AK data, one finds that the average of the bootstrap estimates is very close to the TSLS point estimat, and that the bootstrap variance is very close to the TSLS variance.

The literature has taken a number of approaches. Part of the literature has focused on alternative confidence intervals analogues to the single instrument case. In addition a variety of new point estimators have been proposed.

4.1 Bekker Asymptotics

In this setting alternative asymptotic approximations play a bigger role than in the single instrument case. In an important paper Bekker (1995) derives large sample approximations for TSLS and LIML based on sequences where the number of instruments increases proportionally to the sample size. He shows that TSLS is not consistent in that case. LIML is consistent, but the conventional LIML standard errors are not valid. Bekker then provides LIML standard errors that are valid under this asymptotic sequence. Even with relatively small numbers of instruments the differences between the Bekker and conventional asymptotics can be substantial. See also Chao and Swanson (2005), and Hansen, Hausman and Newey () for extensions.

Here we describe the Bekker correction to the standard errors for the model with a single endogenous regressors, allowing for the presence of exogenous regressors. We write the model as:

$$Y_i = \beta_1' X_{1i} + \beta_2' X_{2i} + \varepsilon_i = \beta' X_i + \varepsilon_i,$$

where the single endogenous variable X_{1i} satisfies:

$$X_{1i} = \pi_1' Z_{1i} + \pi_2' X_{2i} + \eta_i = \pi' Z_i + \eta_i.$$

Define the matrices $\mathbf{P}_Z$ and $\mathbf{M}_Z$ as:

$$\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}', \qquad \mathbf{M}_Z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'.$$

Under the standard, fixed number of instrument asymptotics, the asymptotic variance for LIML is identical to that for TSLS, and so in principle we can use the same estimator. In practice researchers typically estimate the variance for LIML as

$$V_{\text{liml}} = \hat{\sigma}^2 \cdot \left(\mathbf{X}'\left(\mathbf{I} - k_{\text{liml}} \cdot \mathbf{M}_Z\right)\mathbf{X}\right)^{-1},$$

To get Bekker's correction, we need a little more notation. Define

$$\Omega = \begin{pmatrix} \mathbf{Y} & \mathbf{X} \end{pmatrix}' \mathbf{P}_Z \begin{pmatrix} \mathbf{Y} & \mathbf{X} \end{pmatrix} / N = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix},$$

so that

$$\Omega_{11} = \mathbf{Y}'\mathbf{P}_Z\mathbf{Y}/N, \qquad \Omega_{12} = \mathbf{Y}'\mathbf{P}_Z\mathbf{X}/N, \qquad \text{and} \quad \Omega_{22} = \mathbf{X}'\mathbf{P}_Z\mathbf{X}/N.$$

Now define

$$\mathbf{A} = N \cdot \frac{\Omega_{12}'\Omega_{12} - \Omega_{22}\beta\Omega_{12} - \Omega_{12}'\beta'\Omega_{22} + \Omega_{22}\beta\beta'\Omega_{22}}{\Omega_{11} - 2\Omega_{12}\beta + \beta'\Omega_{22}\beta}.$$

Then:

$$V_{\text{bekker}} = \hat{\sigma}^2 \cdot \left(\mathbf{X}'\left(\mathbf{I} - k_{\text{liml}} \cdot \mathbf{M}_Z\right)\mathbf{X}\right)^{-1}$$

$$\times \left(\mathbf{X}'\mathbf{P}_Z\mathbf{X} - (k_{\text{liml}} - 1) \cdot \mathbf{A}\right) \cdot \left(\mathbf{X}'\left(\mathbf{I} - k_{\text{liml}} \cdot \mathbf{M}_Z\right)\mathbf{X}\right)^{-1}.$$

4.2 Random Effects Estimators

Chamberlain and Imbens (2004, CI) propose a random effects quasi maximum likelihood estimator. They propose modelling the first stage coefficients π_k, for $k = 1, \ldots, K$, in the regression

$$X_i = \pi_0 + \pi_1' Z_i + \eta_i = \pi_0 + \sum_{k=1}^{K} \pi_k \cdot Z_{ik} + \eta_i,$$

(after normalizing the instruments to have mean zero and unit variance,) as independent draws from a normal $\mathcal{N}(\mu_\pi, \sigma_\pi^2)$ distribution. (More generally CI allow for the possibility that only some of the first stage coefficients come from this common distribution, to take account of settings where some of the instruments are qualitatively different from the others.) The idea is partly that in most cases with many instruments, as for example in the AK study, the instruments arise from interacting a small set of distinct instruments with other covariates. Hence it may be natural to think of the coefficients on these instruments in the reduced form as exchangeable. This notion is captured by modelling the first stage coefficients as independent draws from the same distribution. In addition, this set up parametrizes the many-weak instrument problem in terms of a few parameters: the concern is that the values of both μ_π and σ_π^2 are close to zero.

Assuming also joint normality for (ε_i, η_i), one can derive the likelihood function

$$\mathcal{L}(\beta_0, \beta_1, \pi_0, \mu_\pi, \sigma_\pi^2, \Omega).$$

In contrast to the likelihood function in terms of the original parameters $(\beta_0, \beta_1, \pi_0, \pi_1, \Omega)$, this likelihood function depends on a small set of parameters, and a quadratic approximation to its logarithms is more likely to be accurate.

CI discuss some connections between the REQML estimator and LIML and TSLS in the context of this parametric set up. First they show that in large samples, with a large number of instruments, the TSLS estimator corresponds to the restricted maximum likelihood estimator where the variance of the first stage coefficients is fixed at a large number, or $\sigma_\pi^2 = \infty$:

$$\hat{\beta}_{\text{TSLS}} \approx \arg \max_{\beta_0, \beta_1, \pi_0, \mu_\pi} = L(\beta_0, \beta_1, \pi_0, \mu_\pi, \sigma_\pi^2 = \infty, \Omega).$$

From a Bayesian perspective, TSLS corresponds approximately to the posterior mode given a flat prior on all the parameters, and thus puts a large amount of prior mass on values of the parameter space where the instruments are jointly powerful.

In the same setting with a large number of instruments, no exogenous covariates, and a known reduced form covariance matrix, the LIML estimator corresponds approximately to the REQML estimator where we fix $\sigma_\pi^2 \cdot (1\ \beta_1)'\Omega^{-1}(1\ \beta_1)'$ at a large number. In the special case where we fix $\mu_\pi = 0$ and the random effects specification applies to all isntruments, CI show that the REQML estimator is identical to LIML. However, like the Bekker asymptotics, the REQML calculations suggests that the standard LIML variance is too small: the variance of the REQML estimator is approximately equal to the standard LIML variance times

$$1 + \sigma_\pi^{-2} \cdot \left(\begin{pmatrix} 1 \\ \beta_1 \end{pmatrix}' \Omega^{-1} \begin{pmatrix} 1 \\ \beta_1 \end{pmatrix} \right)^{-1}.$$

This is similar to the Bekker adjustment.

4.3 Choosing Subsets of the Instruments

In an interesting paper Donald and Newey (2001) consider the problem of choosing a subset of an infinite sequence of instruments. They assume the instruments are ordered, so that the choice is the number of instruments to use. Given the set of instruments they consider a variety of estimators including TSLS and LIML. The criterion they focus on is based on an approximation to the expected squared error. This criterion is not feasible because it depends on unknown parameters, but they show that using an estimated version of this leads to approximately the same expected squared error as using the infeasible criterion. Although in its current form not straightforward to implement, this is a very promising approach that can apply to many related problems such as generalized method of moments settings with many moments.

4.4 Other Estimators

Other estimators have also been investigated in the many weak instrument settings. Hansen, Hausman and Newey (2006), and Hausman, Newey and Woutersen (2007) look at Fuller's estimator, which is modification of LIML that has finite moments. Phillips and Hale (1977) (and later Angrist, Imbens and Krueger, 1999) suggest a jackknive estimator. Hahn, Hausman and Kuersteiner (2004) look at jackknife versions of TSLS.

4.5 Flores' simulations

Many simulations exercises have been carried out for evaluating the performance of testing procedures and point estimators. In general it is difficult to assess the evidence of these experiments. They are rarely tied to actual data sets, and so the choices for parameters, distributions, sample sizes, and number of instruments are typically arbitrary.

In one of the more extensive simulation studies Flores-Lagunes (2007) reports results comparing TSLS, LIML, Fuller, Bias corrected versions of TSLS, LIML and Fuller, a Jacknife version of TSLS (Hahn, Hausman and Kuersteiner, 2004), and the REQML estimator, in settings with 100 and 500 observations, and 5 and 30 instruments for the single endogenous variable. He looks at median bias, median absolute error, inter decile range, coverage rates, and He concludes that "our evidence indicates that the random-effects quasi-maximum likelihood estimator outperforms alternative estimators in terms of median point estimates and coverage rates."

References

Anderson, T., and H. Rubin, (1949), "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *Annals of Mathematical Statistics* 21, 570-582-.

Andrews, D., M. Moreira, and J. Stock, (2006), "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regresion," *Econometrica* 74, 715-752-.

Andrews, D., and J. Stock, (2007), "Inference with Weak Instruments," *Advances in Economics and Econometrics*, Vol III, Blundel,, Newey and Persson (eds.), 122-173.

Angrist, J., G. Imbens, and A. Krueger, (1999), "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics,* 14, 57-67.

Angrist, J., and A. Krueger, (1991), "Does Compulsory Schooling Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics* 106, 979-1014.

Bekker, P., (1994), "Alternative Approximations to the Distribution of Instrumental Variables Estimators," *Econometrica* 62, 657-681.

Bound, J., A. Jaeger, and R. Baker, (1996), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association* 90, 443-450.

Card, D., (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica* 69(5), 1127-1160.

Chamberlain, G., and G. Imbens, (2004), "Random Effects Estimators with Many Instrumental Variables," *Econometrica* 72(1), 295-306.

Chao, J., and N. Swanson, (2005), "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica* 73(5), 1673-1692.

Dufour, J.-M., (1997), "Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica* 65, 1365-1387.

Chioda, L., and M. Janson, (1998), "Optimal Conditional Inference for Instrumental Variables Regression," unpublished manuscript, department of economics, UC Berkeley.

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

Flores-Lagunes, A., (2007), "Finite Sample Evidence of IV Estimators Under Weak Instruments," *Journal of Applied Econometrics,* 22, 677-694.

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

Griliches, Z., (1977), "Estimating the Returns to Schooling – Some Econometric Problems," *Econometrica* 45(1), 1-22.

Hahn, J., and J. Hausman, (2003), "Weak Instruments: Diagnosis and Cures in Empirical Ecnometrics," *American Economic Review, Papers and Proceedings* 93, 118-115.

Hahn, J., J. Hausman, and G. Kuersteiner, (2004), "Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approxmiations," *Econometrics Journal*.

Hansen, C., J. Hausman, and W. Newey, (2006), "Estimation with Many Instrumental Variables," Unpublished Manuscript, Department of Economics, MIT.

Hausman, J., W. Newey, and T. Woutersen, (2006), "IV Estimation with Heteroskedasticity and Many Instruments," Unpublished Manuscript, MIT.

Kleibergen, F., (2002), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica* 70(5), 1781-1803.

Mikusheva, A., (2007), "Uniform Inferences in Econometrics," Chapter 3, PhD Thesis, Harvard University, Department of Economics.

Moreira, M., (2001), "Tests with Correct Size when Instruments can be Arbitrarily Weak," Unpublished Paper, Department of Economics, Harvard University.

Moreira, M., (2003), "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica* 71(4), 1027-1048.

Philips, P., (1984), "Exact Small Sample Theory in the Simultaneous Equations Model," *Handbook of Econometrics*, (Griliches and Intrilligator, eds), Vol 2, North Holland.

Phillips, G., and C. Hale, (1977), "The Bias of Instrumental Variables Estimators of Simultaneous Equations Systems," *International Economic Review*, 18, 219-228.

Rothenberg, T., (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," *Handbook of Econometrics*, (Griliches and Intrilligator, eds), Vol 2, Amsterdam, North Holland.

Staiger, D., and J. Stock, (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica* 68, 1055-1096.