Anna Mikusheva

Survey on statistical inferences in weakly-identified instrumental variable models

This paper provides a brief review of the current state of knowledge on the topic of weakly-identified instrumental variable regression. We describe the essence of the problem of weak identification, possible methods for detecting it in applied work as well as methods robust to weak identification. Special attention is devoted to the question of hypothesis testing in the presence of weak identification.

Keywords: instrumental variable regression; weak instruments; robustness; uniform asymptotics.

JEL classification: C26; C36; C12.

1. Introduction

Instrumental variable (IV) regression is a very popular way of estimating the causal effect of a potentially endogenous regressor X on variable Y. Classical ordinary least squares (OLS) regression results in biased estimators and invalid inferences when the regressor X is endogenous, that is, correlated with the error term in the structural equation. This arises in many practically relevant situations when the correlation between X and Y does not correctly reflect the causation from X to Y, because, for example, some variables that influence both X and Y are omitted from the regression, or because there is reverse causality from Y to X. The idea behind IV regression is to use some exogenous variables Z (that is, variables not correlated with the error term) to disentangle some part of the variation in X that is exogenous and to estimate the causal effect of this part on Y using classical methods.

The typical requirements for the validity of the IV regression are twofold: the instruments Z are required to be exogenous (not correlated with the error term) and relevant. The last requirement loosely means that Z should be correlated with X. The problem of weak identification arises when this latter requirement of relevance is close to being violated. As we will see below the problem of weak identification manifests itself when an IV estimator is very biased and when classical IV inferences are unreliable.

To fix the ideas let us assume that one wants to estimate and make inferences about a $k \times 1$-dimensional coefficient β in the regression

$$Y_i = \beta'X_i + \gamma'W_i + e_i, \tag{1}$$

where X_i is a $k \times 1$ regressor potentially correlated with the error term e_i. We assume that $p \times 1$-dimensional regressors W_i are exogenous and that the coefficient γ is not of interest by itself. Since there may be a non-zero correlation between X_i and e_i, the OLS estimator of coefficient β is biased and asymptotically inconsistent, while all statistical inferences using it, such as OLS confidence sets and OLS tests based on *t*-statistics provide coverage (size) that is asymptotically wrong.

The IV regression approach assumes that one has $r \times 1$-dimensional variable Z_i which satisfies two conditions: (i) exogeneity ($EZ_ie_i = 0$) and (ii) relevance; that is, the rank of matrix

$$E\left(\begin{bmatrix} X_i \\ W_i \end{bmatrix}[Z_i'; W_i']\right) \text{ is } k + p.$$

The estimation procedure often used in the IV setting is the so-called Two-Stage Least-Squares (TSLS) estimator, which employs two steps. First, it disentangles an exogenous variation in X_i which is due to variation in Z_i; for this one uses the OLS regression of X_i on exogenous variables Z_i and W_i. In the second stage coefficients β and γ are estimated via the OLS regression of Y_i on W_i and the exogenous part of X_i obtained during the first stage. For a classical treatment of TSLS the reader may refer any modern econometrics textbook (for example, chapter 3 in (Hayashi, 2000) and chapter 8 in (Greene, 2012)). It has been shown that under assumptions of exogeneity and relevance of instruments Z_i, the TSLS estimator of β is consistent and asymptotically normal. Asymptotically valid testing procedures as well as procedures for the construction of a confidence set for β can be based on the TSLS t-statistics.

The problem of weak instruments arises when instrument Z_i is exogenous but the relevance condition is close to being violated. In such a case classical asymptotic approximations work poorly, and inferences based on the TSLS t-statistics become unreliable and are often misleading.

Example: return to education. What follows is one of the most widely known empirical examples of weak IV regression. For an initial empirical study we refer to Angrist and Krueger (1991), and for a discussion of the weakness of the used instruments to Bound et al. (1995). The empirical question of interest is the estimation of the causal effect of years of education on the lifetime earnings of a person. This question is known to be difficult to answer because the years of education attained is an endogenous variable, since there exist some forces that both affect the educational level as well as the earnings of a person. Many cite «innate ability» as one such force. Indeed, an innately more talented person tends to remain in school longer, and, at the same time, s/he is more likely to earn more money, everything else being equal.

Angrist and Krueger (1991) suggested the use of the «quarter of birth» (quarter in which the person was born) as an instrument. They argue that the season in which a person is born likely will not have a direct effect on his earnings, while it may have an indirect effect (through the education attained). The argument here is that most states have compulsory education laws. These laws typically state that a student can be admitted to a public school only if s/he is at least six years old by September 1. Most states also require that a student stay in school at least until he or she turns 16 years old. In this way a person born on August 31 will have a year more of education than the person born on September 2 by the time they both reach the age of 16, when they have the option of dropping out of school. Thus, the quarter of birth is arguably correlated with the years of education attained.

Even though the instrument (quarter of birth) is arguably relevant in this example, that is, correlated with the regressor (years of education), we may suspect that this correlation is weak. At the time Angrist and Krueger (1991) was written, it was known that weak correlation between the instrument and the regressor could lead to significant finite-sample bias, but for a long time this was considered to be a theoretical peculiarity rather than an empirically-relevant phenomenon. There existed several beliefs at that time. One of them was that the weak correlation between the instrument and the regressor would be reflected in large standard errors of the TSLS estimator, and they would tell an empirical researcher that the instrument was not informative. Another be-

Anna Mikusheva

lief was that the bias of the TSLS estimator was a finite-sample phenomenon, and that empirical studies with a huge number of observations were immune to such a problem. Bound et al. (1995) showed that these beliefs were incorrect.

Bound et al. (1995) used the data from Angrist and Krueger's (1991) study, but instead of using the actual quarter of birth, they randomly assigned a quarter of birth to each observation. This «randomly assigned quarter of birth» is obviously an exogenous variable, but it is totally irrelevant, as it is not correlated with education. Thus, the IV regression with a randomly-assigned instrument cannot identify the true causal effect. However, Bound et al. (1995) obtained, by running TSLS with randomly-assigned instruments, results very similar to those of Angrist and Krueger (1991). What is especially interesting in this experiment is that the TSLS standard errors for a regression with invalid instruments were not much different from those of Angrist and Krueger (1991). That is, just by looking at the TSLS standard errors, the researcher cannot detect a problem. Another amazing aspect of this exercise was that the initial study described in (Angrist, Krueger, 1991) had a humongous number of observations (exceeding 300000), but nevertheless revealed significant bias in the TSLS estimator.

In what follows we will discuss the asymptotic foundations of weak identification, how one can detect weak instruments in practice and tests robust to weak identification.

There are several great surveys available on weak instruments: they include Andrews and Stock (2005), Dufour (2004) and Stock et al. (2002) among others. I also draw the reader's attention to a lecture on weak instruments given by Jim Stock as a part of a mini-course at the NBER Summer Institute in 2008[1].

2. What are weak instruments?

To explain the problem that arises from the presence of only weak correlation between the instruments and the regressor, we consider the highly simplified case of a homoskedastic IV model with one endogenous regressor and no controls. Even though this example is artificial, it illustrates well all the difficulties associated with weak instruments.

Assume that we are interested in inferences about coefficient β in the following regression model

$$Y_i = \beta X_i + e_i, \tag{2}$$

where Y_i and X_i are one-dimensional random variables. We employ TSLS estimation with instruments Z_i and the first-stage regression is

$$X_i = Z_i \pi + v_i, \tag{3}$$

where Z_i is an $r \times 1$-fixed exogenous instrument; e_i and v_i are mean zero random error terms. In general, error terms e_i and v_i are correlated, and thus, X_i is an endogenous regressor. If the unknown coefficient π is not zero, then the instrument Z_i is relevant, and the coefficient β is point-identified. The usual TSLS estimator is $\hat{\beta}_{TSLS} = \dfrac{X'P_Z Y}{X'P_Z X}$, where $P_Z = Z(Z'Z)^{-1}Z'$, and all observations

---

[1]   Available on http://www.nber.org/minicourse_2008.html.

are stacked in matrices Y, X and Z according to the usual conventions. Let us make the additional assumption that error terms (e_i, v_i) are independently drawn from a normal distribution with variances σ_e^2 and σ_v^2 and correlation ρ.

Let us introduce a concentration parameter $\mu^2 = \pi' Z' Z \pi / \sigma_v^2$ and random variables

$$\xi_e = \frac{\pi' Z e}{\sqrt{\pi' Z' Z \pi}\, \sigma_e}, \ \xi_v = \frac{\pi' Z v}{\sqrt{\pi' Z' Z \pi}\, \sigma_v}, \ S_{vv} = \frac{v' P_Z v}{\sigma_v^2}, \text{ and } S_{ev} = \frac{e' P_Z v}{\sigma_e \sigma_v}.$$

It is easy to see that ξ_e and ξ_v are standard normal Gaussian variables with correlation ρ, while S_{vv} and S_{ev} are quadratic forms of normal random variables with respect to the idempotent matrix P_Z. One can show that the joint distribution of $(\xi_e, \xi_v, S_{vv}, S_{ve})$ is known, depends on ρ, and does not depend on sample size or π. Under assumptions stated above, Rothenberg (1984) (another important result on the distribution of TSLS is (Nelson, Startz, 1990)) derived the following exact finite-sample distribution of the TSLS estimator $\hat{\beta}_{TSLS}$:

$$\mu(\hat{\beta}_{TSLS} - \beta_0) = \frac{\sigma_e}{\sigma_v} \cdot \frac{\xi_e + S_{ve} / \mu}{1 + 2\xi_v / \mu + S_{vv} / \mu^2}, \tag{4}$$

where β_0 is the true value of β. Notice that in this expression μ^2 plays the role of sample size. If μ^2 is large, in particular, if $\mu \to \infty$, then $\mu(\hat{\beta}_{TSLS} - \beta_0)$ asymptotically converges to a normal distribution, while if μ is small, then the finite-sample distribution of $\hat{\beta}_{TSLS}$ is non-standard and may be far from normal. From this perspective μ measures the amount of information data have about the parameter β.
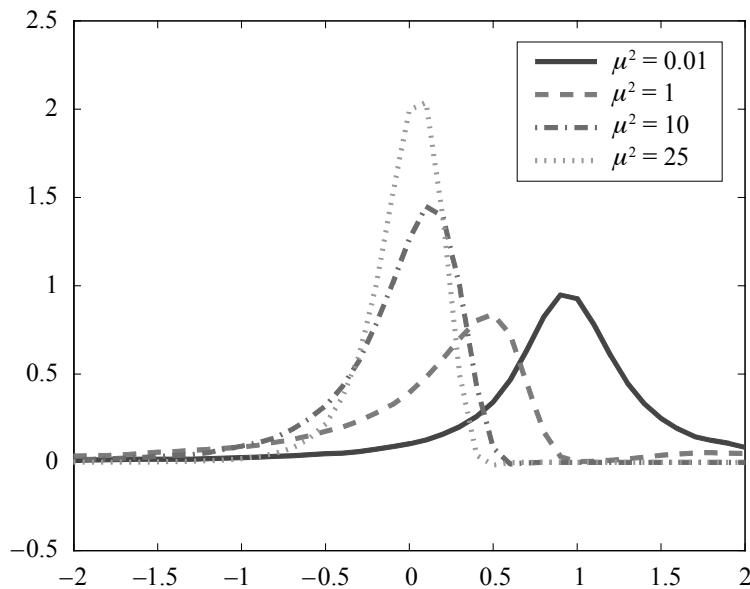

Fig. 1. Finite-sample distribution of the TSLS estimator given by formula (4)

for different values of the concentration parameter $\rho = 0.95$, $\dfrac{\sigma_e}{\sigma_v} = 1$, $\beta_0 = 0$, $r = 1$

Anna Mikusheva

In Figure 1 we depict the finite-sample distribution of the TSLS estimator based on equation (4) for different values of the concentration parameter. The degree of endogeneity is characterized by the correlation between errors ρ. For Figure 1 we used $\rho = 0.95$. The true value of β is chosen to be zero. What we can see is that for extremely small values of μ^2 the TSLS estimator is very biased towards the OLS estimator. It is easy to show that for $\mu^2 = 0$ the distribution of the TSLS is centered around the limit of the OLS estimator, which in this case is equal to 0.95. The bias becomes smaller as μ increases, but the distribution is still skewed and quite non-normal. For large μ^2 ($\mu^2 = 25$) the estimator has nearly no bias, and the distribution is quite close to normal. The behavior of the finite-sample distribution of the *t*-statistic is very similar to that of the distribution of the TSLS.

Looking at the definition of the concentration parameter we notice that μ can be small if π is small, that is, if the correlation between the instrument and the regressor is weak. The weaker the correlation, the further away the finite sample distribution of $\hat{\beta}_{TSLS}$ is from normality.

However, we may notice that the concept of «weak» correlation depends in a significant way on the sample size n. Indeed, let us look again at the expression for the concentration parameter $\mu^2 = \pi' \left(\sum_{i=1}^{n} Z_i'Z_i \right) \pi / \sigma_v^2$. It is a customary assumption in classical econometrics that

$\frac{1}{n} \sum_{i=1}^{n} Z_i'Z_i \to Q_{ZZ}$ as n becomes large. So, we can see that to get the same value of the concentration parameter, which measures the quality of the normal approximation, we may have different combinations of π and n. The weaker the correlation π, the larger the number of observations we need to guarantee the same quality of asymptotic normal approximation. The exact trade-off can be expressed if the coefficient π changes with the sample size, namely, $\pi_n = C / \sqrt{n}$, where C is a constant non-zero vector. In such setting, as the sample size increases, μ^2 converges to a constant value of $\mu_\infty^2 = C'Q_{ZZ}C / \sigma_v^2$. This asymptotic embedding is referred to as «weak instrument asymptotics» and was first introduced in (Staiger, Stock, 1997).

Staiger and Stock (1997) also proved that if one has a more general setting, allowing for random (rather than fixed) instruments, non-normal error terms and additional exogenous controls, and consider a sequence of models with $\pi_n = C / \sqrt{n}$, then under quite general assumptions $\mu(\hat{\beta}_{TSLS} - \beta_0)$ asymptotically converges (as the sample size increases to infinity) to the right-hand side of equation (4).

We know that if the instrument is relevant, that is, if $\mathrm{E}Z_iX_i \neq 0$ is fixed, then as the sample size increases ($n \to \infty$) the concentration coefficient μ^2 increases as well, and as a result, $\hat{\beta}_{TSLS}$ is consistent and asymptotically normal. From this perspective some believe that weak instruments are a finite-sample problem, and if one has a larger sample the problem will disappear. We argue here that this is neither a useful nor a constructive way to consider the problem; an applied researcher in economics usually does not have the luxury of choosing the sample size he or she would most prefer. As Staiger and Stock (1997) showed for each sample size (even for a very large one) there will exist some values of correlation between the instrument and the regressor such that the quality of normal approximation is poor. From this perspective it is better to treat the problem of weak instruments as an issue of the non-uniformity of asymptotics in the sense defined by Mikusheva (2007). Namely, as the sample size goes to infinity and the correlation between X_i and

Z_i is non zero, the convergence of $\sqrt{n}(\hat{\beta}_{TSLS} - \beta_0)$ to a normal distribution is not uniform with respect to this correlation. If the correlation is small the convergence is slow, and it will require a larger sample to allow for the normal approximation to be accurate. One may hope that another asymptotic embedding will provide better asymptotic approximation. Andrews and Guggenberger (2010) proved that the weak-instrument asymptotic of Staiger and Stock (1997) results in the uniform asymptotic approximation.

3. Detecting weak instruments

The weak-instrument problem arises when the correlation between the instruments and the regressor is too small for a given sample size and leads to several failures. First, the TSLS estimator is significantly biased towards the inconsistent OLS estimator. Second, tests and confidence sets based on the TSLS *t*-statistics violate size (coverage) requirements. The formal test that allows one to detect the weak-instrument problem has been developed by Stock and Yogo (2005).

Stock and Yogo's (2005) test of weak instruments is based on so-called first-stage *F*-statistics. Assume that we wish to run regression (1) with instruments Z_i. Then the first-stage regression is:

$$X_i = \pi'Z_i + \delta'W_i + v_i. \tag{5}$$

Consider the OLS *F*-statistic for testing hypothesis $\mathbf{H}_0 : \pi = 0$ in the first-stage regression. Stock and Yogo (2005) demonstrated that there is a direct relation between the concentration parameter and the value of the *F*-statistic, and in particular, the low value of an *F*-statistic indicates the presence of weak instruments.

Stock and Yogo (2005) suggested two criteria for determining the cut-offs for the value of the first-stage *F*-statistic such that if the value of the *F*-statistic falls above the cut-off, then a researcher can safely assume that he can use the TSLS method. The first criterion is to choose the cut-off in such a way that the bias of the TSLS estimator does not exceed 10% of the bias of the OLS estimator. The second criterion guarantees that if the value of the *F*-statistic is above the cut-off then the 5%-size test based on the TSLS *t*-statistic for β is not of a size exceeding 15%. Stock and Yogo (2005) provided the tables with cut-offs for different numbers of instruments, *r*, for both criteria. These tables resulted in a more rough, but commonly used, rule of thumb, that a first stage *F*-statistic below 10 indicates the presence of weak instruments. Stock and Yogo (2005) also established a generalization of this result to the case when the regressor X_i is multi-dimensional, and in such a case one ought to consider the first-stage matrix and a test for rank of this matrix (see (Cragg, Donald, 1993) for more details).

At this juncture, I want to voice a word of caution. The logic behind the detection of weak instruments through the first-stage *F*-statistic relies heavily on the assumption that the model is homoskedastic. To the best of my knowledge the problem of detecting weak instruments in models with heteroskedasticity or autocorrelation of error terms remains unsolved.

An alternative approach to detect weak instruments is Hahn and Hausman's (2002) test. It tests the null hypothesis that the instruments are strong and thus the rejection of such a hypothesis indicates the presence of weak instruments. Unfortunately, the power of this test is low for some alternatives (see (Hausman et al., 2005)), and many cases of weak instruments may slip through the cracks.

Anna Mikusheva

4. Inference methods robust towards weak instruments

In this section we discuss statistical inferences, that is, testing procedures and confidence set construction procedures that are robust to weak instruments. Tests (confidence sets) robust towards weak instruments are supposed to maintain the correct size (coverage) no matter whether the instruments are weak or strong.

The problems of testing and confidence set construction are dual problems. If one has a robust test, s/he can produce a robust confidence set simply by inverting the test. Namely, in order to construct a confidence set for β she should test all hypotheses of the form $\mathbf{H}_0 : \beta = \beta_0$ for different values of β_0 and then examine the set of all β_0 for which the hypothesis is accepted. This «acceptance set» will be a valid confidence set. In general the procedure can be implemented via grid testing (testing on a fine enough grid of values of β_0). Because of the duality from now on we will mainly restrict our attention to the problem of robust testing.

One dramatic observation about tests (confidence sets) robust to weak instruments was made by Dufour (1997), whose statement was closely related to an earlier observation by Gleser and Hwang (1987). Dufour (1997) showed that if one allows the strength of the instruments to be arbitrarily weak, that is, the correlation between the instrument and the regressor are arbitrarily close to zero, then any robust testing procedure must produce confidence sets of infinite length with positive probability. This statement has a relatively simple intuition. If the instruments are not correlated with the regressor, i.e. they are irrelevant, then parameter β is not identified, and any value of β is consistent with data. A valid confidence set in such a case must be infinite at least with probability equal to the coverage. Dufour (1997) spells out a continuity argument if the correlation can approach zero arbitrarily closely. Dufour's (1997) result implies that the classical TSLT t-test which compares the t-statistic with quantiles of the standard-normal distribution cannot be robust to weak instruments, since the corresponding confidence set is finite with probability one.

The main difficulty of performing inferences robust to weak instruments may be formulated in the following way. The distributions of the TSLS estimator and the TSLS t-statistic depend on the value of the concentration parameter μ^2, and from this perspective it can be called a nuisance parameter. Unfortunately, in weak-instrument asymptotics the value of the concentration parameter μ^2 cannot be consistently estimated.

Current literature contains several ideas of how to construct inferences robust to weak identification. Among them are the idea of using a statistic the distribution of which does not depend on μ, the idea to perform inferences conditionally on the sufficient statistics for μ, and the idea of the projection method. We spell out these ideas one by one in more detail below. Currently the most progress has been achieved in the case of a single endogenous regressor (that is, when X_i is one-dimensional). The inferences in the case when X_i is multi-dimensional mostly constitute an open econometric problem. I will discuss difficulties of this case in a separate section later on.

5. Case of one endogenous variable

Assume that data $\{Y_i, X_i, Z_i\}$ satisfy structural equation (2) and first-stage equation (3). Assume that Y_i and X_i are both one-dimensional, while Z_i is an $r \times 1$ vector. Assume also that the error terms in both equations are conditionally homoskedastic. We are interested in testing the null hypothesis $\mathbf{H}_0 : \beta = \beta_0$.

All of the tests in this section can easily be generalized to include exogenous controls, that is, if we have structural equation (1) and first-stage equation (5). In such a case consider variables $Y^{\perp} = (I - P_W)Y$, $X^{\perp} = (I - P_W)X$ and $Z^{\perp} = (I - P_W)Z$, where $P_W = W(W'W)^{-1}W'$, then data on $\{Y_i^{\perp}, X_i^{\perp}, Z_i^{\perp}\}$ satisfy the system of equations (2) and (3).

One approach to accurately perform inferences robust to weak instruments is to find statistics whose distributions do not depend on the value of the concentration parameter μ^2. We are aware of two such statistics: the Anderson–Rubin (AR) statistic introduced by Anderson and Rubin (1949) and the Lagrange Multiplier (LM) statistic whose robust properties were pointed out in (Kleibergen, 2002; Moreira, 2002).

The AR statistic is defined in the following way:

$$AR(\beta_0) = \frac{(Y - \beta_0 X)' P_Z (Y - \beta_0 X) / r}{(Y - \beta_0 X)' M_Z (Y - \beta_0 X) / (n - r)},$$

here $P_Z = Z(Z'Z)^{-1}Z'$, $M_Z = I - P_Z$, n is the sample size and r is the number of instruments. Under quite general assumptions the asymptotic distribution of the AR statistic does not depend on μ^2 either in classical or in weak-instrument asymptotics and converges in large samples to χ_r^2 / r. Large values of the AR statistics indicate violations of the null hypothesis.

To introduce the LM test let us consider the reduced-form for IV regression. For this plug equation (3) into equation (2), and obtain:

$$Y_i = \beta \pi' Z_i + w_i,$$

where $w_i = e_i + \beta v_i$. Let Ω be the covariance matrix of error terms $(w_i, v_i)'$. A natural estimator of Ω is $\hat{\Omega} = \tilde{Y}' M_Z \tilde{Y} / (n - r)$, where $\tilde{Y} = [Y, X]$. Let us now introduce the following statistics (first used in (Moreira, 2002)):

$$\hat{S} = \frac{(Z'Z)^{-1/2} Z' \tilde{Y} b_0}{\sqrt{b_0' \hat{\Omega} b_0}}, \quad \hat{T} = \frac{(Z'Z)^{-1/2} Z' \tilde{Y} \hat{\Omega}^{-1} a_0}{\sqrt{a_0' \hat{\Omega}^{-1} a_0}}, \tag{6}$$

where $b_0 = [1, -\beta_0]'$ and $a_0 = [\beta_0, 1]'$. The LM statistic is of the following form

$$LM(\beta_0) = \frac{(\hat{S}'\hat{T})^2}{(\hat{T}'\hat{T})}.$$

Asymptotically, the LM statistic has a χ_1^2 distribution in both classical and in weak-instrument asymptotics (independently from the value of μ^2). A high value of the LM statistics indicates violations of the null hypothesis.

Both the AR statistics and the LM statistics when paired with the quantiles of the corresponding χ^2 distributions can be used to form weak-instrument robust testing procedures known as the AR and the LM tests.

Moreira (2003) came up with a different, new idea of how to perform testing in a manner that is robust to the weak-instrument problem. Moreira (2003) considered a model like that described by equations (2) and (3) with the additional assumptions that instruments Z_i are fixed, error terms e_i and v_i are jointly i.i.d. normal, and the covariance matrix of reduced-form error terms Ω is known. Consider statistics S and T which are defined as in equation (6) and use Ω in place of $\hat{\Omega}$.

Anna Mikusheva

Moreira (2003) showed that S and T are sufficient statistics for the model considered, and $T'T$ is the sufficient statistic for the concentration parameter. In particular, if one considers a distribution of any test statistic R conditional on random variable $T'T$, $F_{R|T'T}(x\,|\,t) = P\{R \leq x\,|\,T'T = t\}$, then this distribution does not depend on μ^2. So, instead of using fixed critical values, Moreira (2003) suggested the use of critical values that depend on the realization of $T'T$, that is, random critical values that are quantiles of conditional distribution $F_{R|T'T}(x\,|\,t) = P\{R \leq x\,|\,T'T = t\}$ evaluated at $t = T'T$. Moreira (2003) also demonstrated that any test that has exact size α for all values of (nuisance) parameter μ, a so-called «similar test», is a conditional test on the statistic $T'T$.

Any test can be corrected to be robust to weak instruments in this setting using the conditioning idea. There are two conditional tests usually considered: the conditional Wald test (corrected squared t-test) and the conditional likelihood ratio test (CLR).

The conditional Wald test uses a statistic equal to the square of the TSLS t-statistic and a critical value dependant on the realization of $t = T'T$, which are quantiles of the conditional distribution $P\{Wald \leq x\,|\,T'T = t\}$ evaluated at $t = T'T$. Conditional quantiles are calculated using Monte Carlo simulations of the conditional distribution. Andrews et al. (2007) discuss the details of this testing procedure. They also showed that the power of the conditional Wald test is much lower than the power of alternatively available tests such as the AR, the LM and the CLR tests, and recommended that researchers not employ the conditional Wald test in practice.

The CLR test was introduced in (Moreira, 2003) and is based on the likelihood ratio (LR) statistic paired with conditional on $T'T$ critical values. Below is the definition of the LR statistic in this case

$$LR = \frac{1}{2}\left(S'S - T'T + \sqrt{(S'S + T'T)^2 - 4\left((S'S)(T'T) - (S'T)^2\right)}\right). \tag{7}$$

If the instruments are strong, then the LR statistic has asymptotically χ_1^2 distribution. But under weak instruments this approximation is poor and we use instead the conditioning argument. Critical values can be calculated by Monte Carlo simulations of the conditional distribution, but this is numerically a very time-consuming procedure. A more accurate and quick way of arriving at conditional critical values was suggested in (Andrews et al., 2007). If one wishes to get rid of the assumptions of fixed instruments, the normality of error terms and that Ω is known, one should use the formulation of the LR statistic similar to that stated in equation (7) but with $\hat{S}$ and $\hat{T}$ in place of S and T. Mikusheva (2010) showed that under quite general assumptions the resulting test is asymptotically valid uniformly over all values of the concentration parameter.

Andrews et al. (2006) examined the question of how to construct a test with optimal power properties while keeping it robust to weak instruments. They considered a model with fixed instruments, normal errors and known Ω. They produced a power envelope for a class of similar two-sided tests invariant to any orthogonal rotation of the instruments. They showed that the power functions of the CLR test in simulations cannot be distinguished from the power envelope in all the cases they considered. Based on this observation they claimed that the CLR is «nearly uniformly most powerful» in this class and recommended the CLR for practical use.

About confidence set construction. As was mentioned in the beginning of this section, the problem of constructing a confidence set is dual to the problem of testing. Since we have several tests robust to weak instruments (the AR, the LM and the CLR) we can invert them and come up with the corresponding robust confidence sets. Apparently this can be done analytically for the AR and the LM statistics, and using a fast and accurate numerical algorithm for the CLR statistic. The algorithm for the inversion of the CLR test was suggested in (Mikusheva, 2010).

Inference procedures robust to weak instruments in the case of one endogenous regressor are implemented in the software known as STATA (command condivreg). For more detail about the use of this command in empirical studies consult Mikusheva and Poi (2006).

6. Multiple endogenous regressors

If the regression has more than one endogenous regressor for which we use instrumental variables, the situation becomes much more complicated, and econometric theory currently has many lacunae pertaining to this case.

Let us consider the following IV regression:

$$Y_i = \beta X_i + \alpha X_i^* + e_i,$$

where both one-dimensional regressors X_i and X_i^* may be endogenous, and we need instruments for both of them. The assumption that X_i and X_i^* are one-dimensional is inessential and needed only for notational simplicity. Assume that one has an $r \times 1$ instrument Z_i ($r \geq 2$), which is exogenous. We assume that the first-stage regressions are

$$X_i = Z\pi_1 + v_{1i},$$
$$X_i^* = Z\pi_2 + v_{2i}.$$

Here potential problem is that the instruments may be weakly relevant, that is, the $r \times 2$ matrix $[\pi_1, \pi_2]$ is close to having rank 1 or 0. In such a case the classical normal approximations for the TSLS estimator and the TSLS t-statistics both fail to provide good accuracy.

There are a number of ways to asymptotically model weak identification which correspond to different-weak instrument asymptotic embeddings. For example, we may assume that π_1 is fixed, $\pi_2 = C / \sqrt{n}$ where π_1 and C are both $r \times 1$ fixed vectors and $[\pi_1, C]$ has rank 2. In such a case we say that β is strongly identified, while coefficient α is weakly identified (the degree of weak identification is 1). If we assume that $[\pi_1, \pi_2] = C / \sqrt{n}$, where C is an $r \times 2$ matrix of rank 2, then both β and α are weakly identified (the degree of weak identification is 2). In practice, however, one is more likely to encounter a situation where some linear combination of β and α is weakly identified, while another linear combination of them is strongly identified. This corresponds to the degree of weak identification being 1, and the case reduces to the first one after some rotation of the regressors.

We consider now two different testing (confidence set construction) problems: the one when we are interested in testing all structural coefficients jointly ($\mathbf{H}_0 : \beta = \beta_0, \ \alpha = \alpha_0$) and the one when we want to test a subset of structural coefficients ($\mathbf{H}_0 : \beta = \beta_0$). The literature at its current stage has some good answers for the former problem and contains many open questions for the latter.

6.1. Testing all structural coefficients jointly

Assume we want to test a null hypothesis $\mathbf{H}_0 : \beta = \beta_0, \ \alpha = \alpha_0$ about both structural parameters β and α simultaneously. Kleibergen (2007) provided a generalization of weak-instrument robust AR, the LM and the CLR tests for the joint hypothesis.

Anna Mikusheva

The idea here is to consider IV estimation problem to be a generalized method-of-moments (GMM) moment condition:

$$\mathrm{E}\Big[Z_i'(Y_i - \beta_0 X_i - \alpha_0 X_i^*) \Big] = 0$$

and its implied objective function, which, if evaluated at postulated (β_0, α_0), is called the AR statistic, following Stock and Wright (2000):

$$AR(\beta_0, \alpha_0) = \frac{(Y - \beta_0 X - \alpha_0 X^*)' P_Z (Y - \beta_0 X - \alpha_0 X^*)}{(Y - \beta_0 X - \alpha_0 X^*)' M_Z (Y - \beta_0 X - \alpha_0 X^*)/(n-r)}.$$

Under quite general conditions the AR statistic has a χ_r^2-asymptotic distribution if hypothesis $\mathbf{H}_0 : \beta = \beta_0$, $\alpha = \alpha_0$ is true. The convergence holds if identification is strong and if it is weak (under the full variety of weak-instrument asymptotic embeddings discussed above).

Kleibergen (2007) also contains a generalization of the LM test, known as the KLM test, which is robust to weak instruments. This test compares statistic $KLM(\beta_0, \alpha_0)$ with χ_2^2 critical values. Kleibergen (2007) also introduced a new statistic, called J-statistic, $J(\beta_0, \alpha_0) = AR(\beta_0, \alpha_0) - KLM(\beta_0, \alpha_0)$ and showed that it is asymptotically independent from $KLM(\beta_0, \alpha_0)$ and has asymptotic distribution χ_{r-2}^2 for all possible weak-instrument embeddings.

There are several generalizations of the CLR test to the case with multiple endogenous regressors. Kleibergen (2007) called these generalizations the quasi-likelihood ratio (QLR) test and defined it as:

$$QLR(\beta_0, \alpha_0) = \frac{1}{2}\Big[AR - rk + \sqrt{(AR + rk)^2 - 4J \cdot rk} \Big],$$

where $AR = AR(\beta_0, \alpha_0)$ and $J = J(\beta_0, \alpha_0)$ are the AR and J-statistics defined above, while $rk = rk(\beta_0, \alpha_0)$ is the so-called rank statistic that measures the strength of identification. There exist several potential choices for the rank statistic, among them statistics introduced in (Cragg, Donald, 1993; Robin, Smith, 2000; Kleibergen, Paap, 2006). The QLR statistic should be compared with conditional critical values that are quantiles of the conditional distribution of the QLR statistic given statistic $rk(\beta_0, \alpha_0)$. The conditional distribution can be simulated using the following fact. Conditionally on $rk(\beta_0, \alpha_0)$, statistics $KLM(\beta_0, \alpha_0)$ and $J(\beta_0, \alpha_0)$ are independent and have χ_2^2 and χ_{k-2}^2 distributions correspondingly, while $AR = KLM + J$.

Kleibergen (2007) also showed that the AR, KLM and QLR tests are robust to the weak-instrument problem, and they maintain good size properties. However, the power comparison between these three tests remains unclear, the optimal choice of the rank statistic for the QLR test remains unknown as well.

The robust tests can be inverted in order to obtain weak-instrument robust confidence sets. We should note that as a result of such an inversion one would end up with a joint (2-dimensional) confidence set for β and α.

6.2. Testing a subset of parameters

In applied research we are often interested in testing a hypothesis about β only, that is, $\mathbf{H}_0 : \beta = \beta_0$, or in constructing a confidence set for β while treating α as a free unknown parameter (the so-called nuisance parameter). This problem is widely known to be challenging from

a theoretical perspective, and solutions to it heavily rely on our willingness to make additional assumptions.

If α is strongly identified. Assume that parameter α is strongly identified while β may be weakly identified, namely, $\pi_1 = C / \sqrt{n}$ where $[\pi_2, C]$ is a fixed matrix of rank 2, i.e., instruments are weakly correlated with X_i while strongly correlated with X_i^*. In such a case one can show that under the null hypothesis there exists a consistent estimator of α, in particular, if the null hypothesis $\mathbf{H}_0 : \beta = \beta_0$ holds true the continuously updating estimator

$$\hat{\alpha}(\beta_0) = \arg\min_{\alpha} AR(\beta_0, \alpha)$$

is a consistent estimator of α. One also can arrive at the asymptotic distribution of this estimator. Kleibergen (2004) showed that this estimator can be used to construct valid tests about the coefficient β. In particular, if we evaluate the AR statistic at a value of α equal to $\hat{\alpha}(\beta_0)$, in other words consider

$$AR(\beta_0) = AR(\beta_0, \hat{\alpha}(\beta_0)) = \min_{\alpha} AR(\beta_0, \alpha),$$

then if the null holds we have $AR(\beta_0) \Rightarrow \chi_{r-1}^2$. Notice that for the joint test of both β and α with the AR statistic we used a χ_r^2 distribution. We have a reduction in the degrees of freedom in the case of a subset of parameter tests due to the estimation of α.

Kleibergen (2004) provided formulas for the KLM and QLR tests for testing $\mathbf{H}_0 : \beta = \beta_0$ under the assumption that α is strongly identified. The corresponding statistics are equal to the statistics for the joint test evaluated at $\alpha = \hat{\alpha}(\beta_0)$, while the limit distributions are corrected for the degrees of freedom.

No assumptions about strength of identification of α. Unfortunately, the assumption that α is strongly identified is in general questionable, and as of now we do not have a viable way of checking it. Hence, we need a method of testing that would be robust to the weak identification of α as well as the weak identification of β. The current literature contains two competing approaches.

The first approach is the so-called projection method popularised by Dufour and Taamouti (2005, 2007). It is based on the following observation. Imagine that we have a test of a 5% size for testing the hypothesis $\mathbf{H}_0 : \beta = \beta_0, \alpha = \alpha_0$, and the test compares statistic $R(\beta_0, \alpha_0)$ with the critical value q and accepts if $R < q$. Then a test which accepts if $\min_{\alpha} R(\beta_0, \alpha) < q$ is a test of the hypothesis $\mathbf{H}_0 : \beta = \beta_0$ with a size not exceeding 5%. Indeed, if the null $\mathbf{H}_0 : \beta = \beta_0$ then there exists α^* such that (β_0, α^*) are the true parameters of the model. We always have $\min_{\alpha} R(\beta_0, \alpha) \le R(\beta_0, \alpha^*)$, while the right side of the inequality does not exceed q with probability 95%. To translate this approach into confidence set construction assume that we have a valid joint confidence set for β and α with coverage of 95%. Then the projection of this set on the β axis constitutes a confidence set for β with coverage of not less than 95%. Note that for the projection method to work no assumptions about the identification of α are necessary. This projection-method technique can be applied to any valid test of the joint hypothesis.

By applying this approach to the AR test we end up with the AR-projection test. To test $\mathbf{H}_0 : \beta = \beta_0$ we compare statistic $\min_{\alpha} AR(\beta_0, \alpha)$ with quantile of χ_r^2. Notice that under the assumption that α is strongly identified, Kleibergen (2004) uses the same statistic $\min_{\alpha} AR(\beta_0, \alpha)$ but compares it to a smaller quantile of χ_{r-1}^2. This loss of power by the projection method is the

Anna Mikusheva

price we pay for being robust to the weak identification of α. In general, the projection method is known to be conservative. Chaudhuri and Zivot (2008) created a procedure which improves upon the projection method by switching to a larger critical value when we have strong empirical evidence of α being strongly identified.

An alternative to the projection method was recently suggested in (Guggenberger et al., 2012), where the authors considered an IV model with more than one endogenous regressor. Guggenberger et al. (2012) showed that if errors are homoskedastic and the hypothesis $\mathbf{H}_0 : \beta = \beta_0$ holds then statistic $AR(\beta_0) = \min_{\alpha} AR(\beta_0, \alpha)$ is asymptotically stochastically dominated by a χ^2_{r-1}-distribution if α is weakly identified. Quantiles of χ^2_{r-1} can be used as critical values both with and without the assumption that α is strongly identified. This provides significant power improvement over the AR-projection method. However, we do not know if this result is generalizable to the case of heteroskedasticity or to any other statistic. Guggenberger et al. (2012) noted that a direct generalization of their AR result to the LM statistic does not hold.

7. Conclusions

This paper discusses recent advances in the theory of making statistical inferences in IV regression with potentially weak instruments. Weak instrument theory is currently an area of active research. It has experienced some successes such as a good understanding of how to make inferences in the case of a single endogenous regressor. At the same time there remain many open questions. Among them: how to test for weak identification under heteroskedasticity, what the optimal tests are (in terms of power) in a model with multiple endogenous regressors, and how to find similar tests for hypotheses about a subset of parameters.

There are many areas close to the main theme of this paper that we do not discuss. Among them are the problems of finding an estimator with some optimal properties for a weak-IV model, the problem of many instruments, and the generalization of a weak-instrument problem to the non-linear context known as the weakly-identified GMM problem.

References

Anderson T., Rubin H. (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., Guggenberger P. (2010). Asymptotic size and a problem with subsampling and with the *m* out of *n* bootstrap. *Econometric Theory*, 26, 426–468.

Andrews D. W. K., Moreira M., Stock J. (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica*, 74 (3), 715–752.

Andrews D. W. K., Moreira M., Stock J. (2007). Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics*, 139 (1), 116–132.

Andrews D. W. K., Stock J. H. (2005). Inference with weak instruments. Unpublished manuscript, Yale University.

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

Bound J., Jaeger D. A., Baker R. (1995). 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.

Chaudhuri S., Zivot E. (2008). A new method of projection-based inference in GMM with weakly identified parameters. *Working paper*, University of North Carolina, Chapel Hill. http://www.unc.edu/~maguilar/UNCNCSU/sc-ez-08.pdf.

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

Dufour J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65, 1365–1388.

Dufour J.-M. (2004). Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics*, 36 (4), 767–808.

Dufour J.-M., Taamouti M. (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, 73, 1351–1365.

Dufour J.-M., Taamouti M. (2007). Further results on projection-based inference in iv regressions with weak, collinear or missing instruments. *Journal of Econometrics*, 139, 133–153.

Gleser L., Hwang J. (1987). The non-existence of $100(1 - \alpha)$% confidence sets of finite expected diameter in error-in-variables and related models. *Annals of Statistics*, 15, 1351–1362.

Greene W. H. (2012). *Econometric Analysis*, 7th edition. *Pearson*.

Guggenberger P., Kleibergen F., Mavroeidis S., Chen L. (2012). On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression. *Econometrica*, 80 (6), 2649–2666.

Hahn J., Hausman J. (2002). A new specification test for the validity of instrumental variables. *Econometrica*, 70, 163–189.

Hausman J., Stock J. H., Yogo M. (2005). Asymptotic properties of the Hahn–Hausman test for weak instruments. *Economics Letters*, 89 (3), 333–342.

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

Kleibergen F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70, 1781–1803.

Kleibergen F. (2004). Testing subsets of structural parameters in the instrumental variables regression model. *Review of Economic Studies*, 86 (1), 418–423.

Kleibergen F. (2007). Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, 139, 181–216.

Kleibergen F., Paap R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133 (1), 97–126.

Mikusheva A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75 (5), 1411–1452.

Mikusheva A. (2010). Robust confidence sets in the presence of weak instruments. *Journal of Econometrics*, 157, 236–247.

Mikusheva A., Poi B. (2006). Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal*, 6 (3), 335–347.

Moreira M. (2002). Tests with correct size in the simultaneous equations model. *PhD Thesis*, UC Berkeley.

Anna Mikusheva

Moreira M. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71 (4), 1027–1048.

Nelson C., Startz R. (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica*, 58, 967–976.

Rothenberg T. J. (1984). Approximating the distributions of econometric estimators of test statistics. In: *Handbook of Econometrics*, Vol. II, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland, 881–936.

Robin J.-M., Smith R. J. (2000). Tests of rank. *Econometric Theory*, 16, 151–175.

Staiger D., Stock J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65 (3), 557–586.

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

Stock J. H., Yogo M. (2005). Testing for weak instruments in linear IV regression. In: *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg*, D. W. K. Andrews and J. H. Stock (eds.). Cambridge, UK: Cambridge University Press.

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