14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
Novemeber 13, 2007
corrected September 2012

Lecture 7-8

Weak IV.

This lecture extensively uses lectures given by Jim Stock as a part of mini-course at the NBER Summer Institute.

1. What are Weak Instruments?

Consider the simplest classical homoskedastic IV model:

$$y_t = \beta x_t + u_t$$
$$x_t = Z_t \pi + v_t,$$

where y_t are one-dimensional, x_t is $n \times 1$, Z_t is $k \times 1$ and $Eu_t Z_t = 0$, one observes i.i.d. data $\{y_t, x_t, Z_t\}$. Assume that $n \le k$. In general u_t and v_t are correlated, and thus, x_t is an endogenous regressor. Z_t is exogenous (since we assumed that $Eu_t Z_t = 0$), if it also relevant ($EZ'x$ has rank n), then it can serve as instrument and the model is identified. The usual TSLS will be $\hat{\beta}_{TSLS} = (x' P_Z x)^{-1} x' P_Z y$, where $P_Z = Z(Z'Z)^{-1}Z'$.

The problem of weak identification arises when moment conditions are not very informative about the parameter of interest, that is, when the rank of the matrix $\frac{\partial g_0(\theta_0)}{\partial \theta}|_{\theta=\theta_0} = R(\theta_0) = EZ'x$ is n, but at the same time it is very close to a reduced rank matrix (for example, the smallest eigenvalue of $n \times n$ matrix $x'Z(Z'Z)^{-1}Z'x$ is very close to zero). In the case of 1 instrument and 1 endogenous regressor weak identification corresponds to a weak correlation between the instrument and the regressor.

To explain the essence of weak IV problem we start with a toy example of totally irrelevant instruments.

Not relevant instrument. Imagine a situation when one has 1 endogenous regressor and 1 instrument which is independent of everything (totally irrelevant, $\pi = 0$). That is, the instrument is not valid and β is not identified. The question is how $\hat{\beta}_{TSLS}$ behaves? This should explain what we see in Bounder, Jaeger, Baker's (1995) "random quarter of birth" exercise:

$$\hat{\beta}_{TSLS} - \beta_0 = \frac{\sum Z_t u_t}{\sum Z_t v_t} = \frac{\frac{1}{\sqrt{T}} \sum Z_t u_t}{\frac{1}{\sqrt{T}} \sum Z_t v_t} \Rightarrow \frac{\xi_u}{\xi_v},$$

where $(\xi_u, \xi_v)' \sim N(0, \Sigma)$, $\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$. Let $\delta = \sigma_{uv}/\sigma_v^2$, then $\xi_u = \delta \xi_v + \xi$, and $\hat{\beta}_{TSLS} - \beta_0 \Rightarrow \delta + \frac{\xi}{\xi_v}$.

Conclusions:

- $\hat{\beta}_{TSLS}$ is inconsistent (as expected, since β is not identified).
- $\hat{\beta}_{TSLS}$ is centered around $\beta_0 + \delta$ (since $\frac{\xi}{\xi_v}$ has symmetric distribution), which is the limit of OLS.
- Asymptotically $\hat{\beta}_{TSLS}$ has heavy tails (since $\frac{\xi}{\xi_v}$ has Cauchy distribution)

Non-uniform asymptotics If the instrument is relevant $EZ_t x_t \neq 0$, then as the sample size ($T \to \infty$) increases $\hat{\beta}_{TSLS}$ is consistent and asymptotically normal. If $EZ_t x_t = 0$, as shown before, the asymptotics breaks down. So, the correlation equals zero is a point of discontinuity of asymptotics. That is, the limit of $\sqrt{T}(\hat{\beta}_{TSLS} - \beta_0)$ depends on the value of $EZ_t x_t$, which is a *nuisance parameter* (parameter that we do not care about per se, but which affects the distribution) in this case. It means that the convergence of $\sqrt{T}(\hat{\beta}_{TSLS} - \beta_0)$ to normal distribution is not uniform with respect to the nuisance parameter. That is, if $EZ_t x_t \neq 0$ but is very small, the convergence is slow and it requires a larger sample to allow for normal approximation to be accurate. One may hope that another asymptotic embedding will provide better asymptotic approximation.

Concentration parameter

Consider the same IV model as before

$$y_t = \beta x_t + u_t$$
$$x_t = Z_t \pi + v_t,$$

but now assume one endogenous regressor ($n = 1$) and several instruments $k \geq 1$. Introduce a concentration parameter $\mu^2 = \pi' Z' Z \pi / \sigma_v^2$. Then

$$\hat{\beta}_{TSLS} - \beta_0 = \frac{x' P_Z u}{x' P_Z x} = \frac{(Z\pi + v)' P_Z u}{(Z\pi + v)' Z (Z'Z)^{-1} Z (Z\pi + v)} = \frac{\pi' Z u + v' P_Z u}{\mu^2 \sigma_v^2 + 2\pi' Z v + v' P_Z v}.$$

Let's assume that instruments are fixed and errors are normals. Let us introduce $\xi_u = \frac{\pi' Z u}{\sqrt{\pi' Z' Z \pi \sigma_u}}$, $\xi_v = \frac{\pi' Z v}{\sqrt{\pi' Z' Z \pi \sigma_v}}$, $S_{vv} = \frac{v' P_Z v}{\sigma_v^2}$, and $S_{uv} = \frac{u' P_Z v}{\sigma_u \sigma_v}$. Then ξ_u and ξ_v are standard normal, and distribution of S_{vv} and S_{uv} does not depend on sample size (chi-squared). Finally, we have:

$$\mu(\hat{\beta}_{TSLS} - \beta_0) = \frac{\sigma_u}{\sigma_v} \frac{\xi_u + S_{vu}/\mu}{1 + 2\xi_v/\mu + S_{vv}/\mu^2}.$$

Notice, that in this expression μ plays the role of the sample size! If μ is large, $\mu(\hat{\beta}_{TSLS} - \beta_0)$ will be approximately normal, if μ is small, then the distribution is non-standard. That is, μ is an effective nuisance parameter here, and it measures the amount of information data have about the parameter β.

Weak instruments asymptotics

Weak instrument asymptotics is the name for asymptotic embedding modeling correlation as converging to zero at speed $\sqrt{T}$. It is the same as modeling μ being constant. So, assume that $\pi = C/\sqrt{T}$. Then

$$\hat{\beta}_{TSLS} - \beta_0 \Rightarrow \frac{(\lambda + z_v)' z_u}{(\lambda + z_v)'(\lambda + z_v)},$$

where $(z_u, z_v) \sim N(0, \Sigma)$, $\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$. $\lambda = C' Q_{ZZ}^{1/2}$, $Q_{ZZ} = EZ_t Z_t'$.

What is also important here is that under this nesting (weak instrument asymptotics) the first-stage F-statistic (for testing all coefficients on instruments are zeros) converges in distribution to a non-central χ_k^2 with non-centrality parameter μ^2/k.

2. Detecting Weak Instruments

There are several approaches to detect whether one has a weak instrument problem:

(1) Compare the first-stage F statistic with a cut-off (Stock, Yogo). Assume we have a homoskedastic IV model with one endogenous variable x_t and some exogenous variables W_t

$$y_t = \beta x_t + \gamma W_t + u_t$$

where y_t and x_t are one-dimensional. $EW_t u_t = 0$. The data you observe is i.i.d. Let Z_t be a $k \times 1$ instrument, in particular, $Eu_t Z_t = 0$. The first-stage regression in this case is

$$x_t = Z_t \pi + \delta W_t + v_t,$$

and the relevance condition means that $\pi \neq 0$. The weak instrument problem arises when $\pi \approx 0$. Stock and Yogo showed that the first-stage F-statistic is distributed as a non-central χ^2 with a non-centrality parameter directly related to the concentration parameter μ. As a result, the first-stage F-statistic can serve as an indicator of the value of μ.

Idea: look at the first-stage F-statistics since it is an indicator of μ^2 and choose a cut-off that would guarantee either relative bias less than 10% (for estimation) or the size of 5% test being less than 10% (for testing and confidence sets). By relative bias we mean the following: the maximum bias of TSLS is no more than 10% of the bias of OLS.

Realization:

- then the first stage F-statistic is the statistic for testing $\pi = 0$ in the first stage regression $x_t = \pi Z_t + \delta W_t + v_t$.
- We know that $EF = 1 + \mu^2/k$, so we can estimate μ^2/k as $F - 1$.
- Compare the obtained μ^2/k with cut-off (tables in Stock, Wright and Yogo). Note that the cut-offs are far higher than the critical values for the F-test (weak instruments are more often than what would be detected by pre-test $\pi = 0$)
- Rule of thumb $F < 10$ indicates weak instruments.

The procedure described above works only for a single endogenous variable x_t. If the regression has more than one endogenous (instrumented) regressor, then the analog of the F-test will be the first-stage matrix and a test for rank of this matrix. See Cragg and Donald (1993) for more details.

Caution! This test for weak IV assumes a homoskedastic setting! What to do in the heteroskedastic case or when one has autocorrelation is **an open question.**

(2) The Hahn-Hausman test of the null of strong instruments. The idea is that if instruments are strong then the regression and the reverse regression should give estimates of β and $1/\beta$, which are consistent with each other. Think about the "Elasticity of inter-temporal substitution" example from the last lecture. The test is based on comparing them. Problem: it tests the null of strong identification and does not control for the probability of a type-II error (mistake of not-detecting weak IV when it is present). The test may experience power problems as well.

(3) Do not test for weak-strong instruments, but rather use methods robust towards weak instruments.

3. Inference methods robust towards weak instruments

Inferences include tests and confidence sets. We concentrate mainly on tests. Tests robust towards weak instruments are supposed to maintain the correct size no matter whether instruments are weak or strong. These can be achieved in two ways: using statistics whose distribution do not depend on μ or using conditioning on sufficient statistics for μ. The problem of robust inferences is fully solved for the case of one endogenous variable. It is still an open question for the case of more than one endogenous variable.

3.1 Case of one endogenous variable.

There are two widely known statistics whose distributions do not depend on μ: Anderson-Rubin (AR) and Lagrange Multiplier (LM).

AR test Consider our model

$$y = X\beta + u,$$
$$X = Z\Pi + v,$$

where X is one-dimensional and test for hypothesis $H_0 : \beta = \beta_0$. Under the null, vector $y - X\beta$ is equal to the error u_t and is uncorrelated with Z (due to exogeneity of instruments). The suggested statistics is

$$AR(\beta_0) = \frac{(y - X\beta)' P_Z (y - X\beta)}{(y - X\beta)' M_Z (y - X\beta)/(T - k)}.$$

here $P_Z = Z(Z'Z)^{-1}Z', M_Z = I - P_z$. The distribution of AR does not depend on μ asymptotically $AR \to \chi^2_k/k$. The formula may remind you of the J-test for over-identifying restrictions. It would be a J-test if one were to plugs in $\hat{\beta}_{TSLS}$.

In a more general situation of more than one endogenous variable and/or included exogenous regressors AR statistic is F-statistic testing that all coefficients on Z are zero in the regression of $y - \beta_0 X$ on Z and W. Note, that one tests all coefficients β simultaneously (as a set) in a case of more than one endogenous regressor.

AR confidence set One can construct a confidence set robust towards weak instruments based on the AR test by inverting it. That is, by finding all β which are not rejected by the data. In this case, it is the set :

$$\text{Conf. set} = \{\beta_0 : AR(\beta_0) < \chi^2_{k,1-\alpha}\}.$$

The nice thing about this procedure is that solving for the confidence set is equivalent to solving a quadratic inequality. This confidence set can be empty with positive probability (caution!).

LM test The LM test formula can be found in Kleibergen (2002). It has a χ^2_1 distribution irrespective of the strength of the instruments. The problem with this test, though, is that it has non-monotonic power and tends to produce wider confidence sets than the CLR test described below.

Conditional tests The idea comes from Moreira (2003). He suggested that one consider any test statistic conditional on a sufficient statistic for μ be called Q_T. By definition of sufficient statistic, the conditional (on Q_T) distribution of any variable does not depend on μ. So, instead of using fixed critical values, one would use critical values depending on realization of Q_T (that is, random) $q_{1-\alpha}(Q_T)$. Moreira also showed that any test that has exact size α for all values of (nuisance) parameter μ is a conditional test. The CLR (the conditional likelihood ratio test) is a conditional test based on the LR statistic. CLR is preferable since it possesses some optimality properties.

Confidence sets If one has a robust test, s/he can produce a robust confidence set by inverting the test. Namely, test all potential hypotheses $H_0 : \beta = \beta_0$ and consider the set of all β_0 for which the hypothesis is accepted. This set will be a valid confidence set. In general the procedure can be implemented as a grid testing (testing on a fine enough grid of values of β_0). For the homoskedastic case with one endogenous variable, the inversion of the three tests mentioned above can be done analytically. The CLR test and confidence set (as well as those for AR and LM) are implemented in Stata (command `condivreg`). For more than one endogenous regressor, the numerical complication rises dramatically.

Confidence sets robust to weak identification may be (and often will be) unbounded. Think about why!

3.2 More than one endogenous regressor

Consider the following IV regression:

$$y_t = \beta x_t + \gamma x_t^* + u_t,$$

where both x_t and x_t^* may be endogenous. Assume that one has an $k \times 1$ instrument Z_t ($k \geq 2$), which is exogenous. The potential problem is that it may be weakly relevant. That usually means that a 2×2 matrix $X'Z(Z'Z)^{-1}Z'X$ (we stack both x_t and x_t^* in X) has at least one small eigenvalue (but perhaps both eigenvalues are small).

The problem that has been somewhat solved is what to do if one wants to test a hypothesis of the form $H_0 : \beta = \beta_0, \gamma = \gamma_0$ (that is, a joint hypothesis about both parameters) or he wants to construct a joint confidence set for (β, γ). There exists a generalization of all three tests (AR, LM, CLR) to such a situation. For example

$$AR(\beta_0, \gamma_0) = \frac{(y - \beta_0 x - \gamma_0 x^*)' P_Z (y - \beta_0 x - \gamma_0 x^*)}{(y - \beta_0 x - \gamma_0 x^*)' M_Z (y - \beta_0 x - \gamma_0 x^*)/(T-k)},$$

it has a χ^2_k asymptotic distribution if hypothesis $H_0 : \beta = \beta_0, \gamma = \gamma_0$ is true. By inverting this test one can obtain a joint confidence set for (β, γ).

The problem is that in applied research we are often interested in separate confidence sets for β and γ, and tests of the form $H_0 : \beta = \beta_0$ (γ is a nuisance parameter for such a hypothesis). One way of obtaining a confidence set for β only is to do a projection. That is,

$$Conf.\ \ set(\beta) = \{\beta : \exists \gamma \ \ s.t.\ \ (\beta, \gamma) \in joint\ \ conf.\ \ set\}$$

It is equivalent to using test statistic $\min_\gamma AR(\beta_0, \gamma)$ with the critical values χ^2_k. However, this always leads to conservative inferences. In particular, if the instruments are in fact strong, the confidence set for the subset of parameters will be wider than the classical (valid in this case) OLS confidence set. There is a statement, that if γ is strongly identified then

$$\min_\gamma AR(\beta_0, \gamma) \Rightarrow \chi^2_{k-1},$$

and the limit is substantially smaller than χ^2_k. The higher the dimensionality of γ, the higher the potential loss of power. Kleibergen and Mavroiedis recently showed that if the model is homoskedastic, and γ is weakly identified then statistic $\min_\gamma AR(\beta_0, \gamma)$ is asymptotically stochastically dominated by χ^2_{k-1}. So, smaller critical values can be used here as well.

The question of good inferences on a subset of parameters remains open.

4. What about estimation?

This is also a mostly unsolved question. We need something with bias or MSE less than OLS for weak instruments. Consider $k-$class estimators:

$$\hat\beta(k) = [X'(I - kM_Z)X]^{-1} X'(I - kM_Z)y$$

$k = 1$ corresponds to TSLS, $k = k_{LIML}$ =smallest root of det $X^{\perp'} X^{\perp} - k X^{\perp'} M_Z X^{\perp}$ gives LIML. If $k = k_{LIML} - \frac{c}{T-K-r}$ where $r = rank(W)$, then the estimator is called Fuller.

The comparison between the estimators is not that easy: LIML is first order median-unbiased but does not have finite moments, and thus it can give unexpected high values. $\hat\beta_{LIML} = \arg\min AR(\beta)$.

Fuller has some optimality properties for $c = 1$ and performs well in simulations.

The last word of caution

Weak instruments is an asymptotic problem, or better to say, a problem of non-uniformity of classical GMM asymptotics. As a result, bootstrap, Edgeworth expansion, and subsampling are not appropriate solutions.

14.384 Time Series Analysis
Fall 2013