Lecture 11
Weak IV

Instrument exogeneity and instrument relevance are two crucial requirements in empirical analysis using GMM. It now appears that in many applications of GMM and IV regressions, instruments are only weakly correlated with included endogenous variables. It has been noted in the literature that

1. If instruments are weak, the sampling distribution for GMM and IV statistics are in general nonnormal and standard inference is not reliable.
2. Features that make an instrument plausibly exogenous also make it weak.
3. Weak Instrument is not a small sample problem. (Bound, Jaeger and Baker (1995) demonstrates one example with 329,000 observations.)
4. There are more robust methods than conventional GMM.

In this lecture note, we first describe the effect of weak IV, and then introduce some more robust alternatives of 2SLS.

To focus on the main idea, we consider a linear regression model with fixed exogenous regressor (Z) and normal error.

$$y_1 = \beta y_2 + u \quad with \quad y_2 = Z\Pi + v_2$$

where y_1 and y_2 are $n\times 1$ vectors and Z is $n\times k$ matrix of instruments, and $(u_i, v_{2,i})' \sim N(0, [\sigma_u^2, \sigma_{uv}; \sigma_{uv}, \sigma_v^2])$. The reduced form of the structural equation is

$$y_1 = Z\Pi\beta + v_1, \quad and \quad y_2 = Z\Pi + v_2, \tag{1}$$

where $v_1 = \beta v_2 + u$. Then $(v_{1,i}, v_{2,i}) \sim N(0, [\sigma_1^2, \rho\sigma_1\sigma_2; \rho\sigma_1\sigma_2, \sigma_2^2])$, where $\sigma_1 = \beta^2\sigma_v^2 + \sigma_u^2 + 2\beta\sigma_{uv}$, $\rho\sigma_1\sigma_2 = \beta\sigma_v^2 + \sigma_{uv}$ and $\sigma_2^2 = \sigma_v^2$.

The 2SLS estimator is

$$\begin{aligned}
\hat{\beta}^{2SLS} &= \frac{y_2' Z(Z'Z)^{-1}Z'y_1}{y_2'Z(Z'Z)^{-1}Z'y_2} \\
&= \frac{(\Pi'Z' + v_2')P_Z(\beta Z\Pi + v_1)}{(\Pi'Z' + v_2')P_Z(Z\Pi + v_2)} \\
&= \frac{\Pi'(Z'Z)\Pi\beta + \Pi'Z'v_1 + v_2'Z\Pi\beta + v_2'P_Z v_1}{\Pi'Z'Z\Pi + 2\Pi'Z'v_2 + v_2'P_Z v_2} \\
&= \beta + \frac{\Pi'Z'(\sigma_{uv}v_2/\sigma_v^2 + \sigma_1\sqrt{1-\rho^2}v_0) + v_2'P_Z(\sigma_{uv}v_2/\sigma_v^2 + \sigma_1\sqrt{1-\rho^2}v_0)}{\Pi'Z'Z\Pi + 2\Pi'Z'v_2 + v_2'P_Z v_2}
\end{aligned} \tag{2}$$

Clearly, the 2SLS estimator is not unbiased, due to the noise v_2. In the extremum case that $\Pi = 0$,

$$\begin{aligned}\hat{\beta}^{2SLS} &= \frac{v_2' P_Z v_1}{v_2' P_Z v_2} = \frac{v_2' P_Z(\rho v_2/\sigma_2 + \sqrt{1-\rho^2} v_0)\sigma_1}{v_2' P_Z v_2} \\ &= \rho\sigma_1/\sigma_2 + \sigma_1\sqrt{1-\rho^2} \cdot \frac{v_2' P_Z v_0}{v_2' P_Z v_2} \\ &= \beta + \sigma_{uv}/\sigma_v^2 + \sigma\sqrt{1-\rho^2} \cdot \frac{v_2' P_Z v_0}{v_2' P_Z v_2}. \end{aligned} \tag{3}$$

$v_{1,i} = \sigma_1(\rho v_{2,i}/\sigma_2 + \sqrt{1-\rho^2} v_0)$, where $v_0 \sim N(0,1)$ and v_0 is independent of v_2.

The OLS estimator is

$$\hat{\beta}^{OLS} = \frac{y_2' y_1}{y_2' y_2} = \frac{(\Pi' Z' + v_2')(\beta Z\Pi + v_1)}{(\Pi' Z' + v_2')(Z\Pi + v_2)} = \frac{\Pi' Z' Z\Pi\beta + \Pi' Z' v_1 + v_2'\beta Z\Pi + v_2' v_1}{\Pi' Z' Z\Pi + 2\Pi' Z' v_2 + v_2' v_2} \tag{4}$$

In the extreme case that $\Pi = 0$,

$$\hat{\beta}^{OLS} = \frac{v_2' v_1}{v_2' v_2} = \beta + \sigma_{uv}/\sigma_v^2 + \sigma_1\sqrt{1-\rho^2} \cdot \frac{v_2' v_0}{v_2' v_2}. \tag{5}$$

Therefore, in the extreme case, the OLS estimator and the 2SLS estimator have the same bias. But they clearly do not have the same distributional feature. The OLS estimator converges to $\rho\sigma_2/\sigma_1$, but the 2SLS estimator does not converge to a deterministic limit as $n \to \infty$.

The 2SLS estimator is biased and nonnormal. In principle, under the normality and fixed Z assumption, one can study the exact bias and exact distribution of the 2SLS estimator, but it is obviously very cumbersome to do. It is easier to use asymptotics.

1 Weak IV Asymptotics

The weak IV asymptotics was introduced by Staiger and Stock (1997). To reflect the instrument weakness, Staiger and Stock (1997) propose to derive the asymptotic distribution under a drifting sequence of DGPs such that $\Pi = C/\sqrt{n}$, where C is a finite $k \times 1$ vector. Under such a sequence, we shall have

$$\begin{aligned}\hat{\beta}^{2SLS} - \beta &\to_d \frac{(C' Q_{ZZ}^{1/2} + Z_{v_2}' Q_{ZZ}^{-1/2}) Q_{ZZ}^{-1/2} Z_u}{(C' Q_{ZZ}^{1/2} + Z_{v_2}' Q_{ZZ}^{-1/2})(C' Q_{ZZ}^{1/2} + Z_{v_2}' Q_{ZZ}^{-1/2})'} \\ &= \frac{(\lambda + W_{v_2})' W_u}{(\lambda + W_{v_2})'(\lambda + W_{v_2})}, \end{aligned} \tag{6}$$

where $(W_{v_2}', W_{v_u}')' \sim N(0, \Sigma \otimes I_k)/\sigma_v$ and $\lambda = C' Q_{ZZ}^{1/2}/\sigma_v^2$. Clearly, the relative magnitude of $\lambda'\lambda$ and $W_{v_2}' W_{v_2}$ determines how normal the right hand side is. In this sense, the strength of the instruments is measured by $\lambda'\lambda$.

The distribution of the right hand side of the above display is still difficult to study. However, it simplifies if we then let the number of instruments k goes to infinity, and keep $\lambda'\lambda/K \to \Lambda$, then the right-hand-side of the above equation converges to

$$\frac{\sigma_{uv}}{(\Lambda+1)\sigma_v^2}.$$

This can be considered the asymptotic bias of the 2SLS estimator. Note that when $\Lambda = 0$, we are close to the extreme case in the finite sample, and the bias of 2SLS is the same as the OLS estimator.[1]

If Λ could be consistently estimated, one could evaluate the relative bias of 2SLS and OLS estimator: $\sigma_{uv}/((\Lambda+1)\sigma_v^2)$. However, Λ cannot be consistently estimated. Yet, one could conduct a test for the hypothesis $H_0 : \sigma_{uv}/((\Lambda+1)\sigma_v^2) = c$ versus $H_1 : \sigma_{uv}/((\Lambda+1)\sigma_v^2) < c$ using the first stage F-statistic. This is the idea proposed in Staiger and Stock (1997) and developed in Stock and Yogo (2005).

Even though the F-statistic is used, the usual F-critical value for overall significance of the first-stage model is too small because now the null hypothesis is not $\Pi = 0$, but $\Pi'Z'Z\Pi/k$ equals some positive constant that depends on the relative bias level that one can tolerate. Stock and Yogo (2005) provides a critical value table for a 5% test for $c^2 = 5\%, 10\%, 15\%$ and 20% and for different k. The critical value for the tolerance level 5% ranges from 13 to 21 as k ranges from 3 to 30.

Besides bias, one also might want to know how far the usual t-statistic based on 2SLS is from $N(0,1)$. This determines how much over-rejection the 2SLS-based t-test has. Using similar logic, Stock and Yogo also proposed a F-statistic based test for the null hypothesis H_0: a 5% t-test over rejects by x%, versus the alternative that it over rejects more. The critical values for such a hypothesis is also tablized in Stock and Yogo (2005). For $x = 5$, the critical value ranges from 16 to 86, (!) when k ranges from 1 to 30. (This probably means 2SLS-based t-test usually doesn't work...)

More robust estimators There are more robust estimators than 2SLS. These estimators can be summarized as the k-class estimators as they have similar shape and only differ from each other by

[1] The "sequential asymptotics" that we describe here is not rigorous. To be more rigorous, one should take n and k to infinity simultaneously while specifying a relative diverging rate between the two. Our result here, however, is correct, under the conditions that $k = o(n)$, $k \to \infty$ and $\Pi'Z'Z\Pi/k \to \Lambda$. See Bekker (1994) or Chao and Swanson (2005). Under different conditions on the relative divergence rate of k and n and on the limiting behavior of $\Pi'Z'Z\Pi$, one may obtain different limits. See Hahn and Hausman (2005).

Though this limit is obtained under large k asymptotics, it is argued that it characterizes the high-order bias of the estimator well even when k is not very large. See Bekker (1994).

an adjusting factor k. The k-class estimator is of the shape

$$\hat{\beta}^{k-class} = \frac{y_2'(I_n - kM_Z)y_1}{y_2'(I_n - kM_Z)y_2},$$

where $M_Z = I_n - Z(Z'Z)^{-1}Z'$. Several famous k-class estimators include:

1. LIML estimator. k is the smallest root of the equation $det(\underline{Y}'\underline{Y} - k\underline{Y}'M_Z\underline{Y}) = 0$, where $\underline{Y} = [y_1, y_2]$.

2. Fuller-k estimator. $k = k_{LIML} - c/(n-k)$, where c is an adjustment constant.

The LIML and Fuller-k estimators are less biased than 2SLS, especially for large k. The LIML, Fuller-based t-tests are also closer to N(0,1) under the null than the 2SLS based t-test, especially for large k. One can see their superiority in Stock and Yogo (2005), where the F statistic-based weak instrument tests are also designed for LIML and Fuller estimators. For some reason, the LIML is only tested for over-rejection, while the Fuller is only tested for bias. For LIML, for H_0: a 5% t-test over rejects by 5%, the critical value ranges from 16 to 4, when k ranges from 1 to 30. For Fuller (with c=1), the critical value for the tolerance level 5% ranges from 24 to 2 as k ranges from 1 to 30. This suggests that whenever k is greater than 2 and the instruments seem weak, one should not use the 2SLS estimator, but should try LIML or Fuller.

2 Fully Robust Inference

Even though LIML or Fuller estimator are more robust to weak instruments than 2SLS, they are not fully robust. They are biased and their t-statistic nonnormal if the strength of the instruments are below the weak instrument threshold as tablized in Stock and Yogo (2005).

One question is: is there any fully robust estimator? That is, is there an estimator $\hat{\beta}_n$ that is consistent regardless of instrument strength, and a t-test based on it has null distribution $N(0,1)$ regardless of instrument strength? It is quite easy to see that there cannot be such an estimator. This is because when the instrument strength is zero ($\Pi = 0$), the β parameter is not identified. A non-identified parameter cannot be consistently estimated. Without consistency, there can also be no valid t-test.

However, there are fully robust procedures to test the hypothesis $H_0 : \beta = \beta_0$. One such test is the Anderson-Rubin test (AR). The AR test considers the following regression model

$$y_1 - \beta_0 y_2 = Z\kappa + u$$

and write H_0 as $H_0 : \kappa = 0$, and use the statistic

$$AR = \frac{(y_1 - \beta_0 y_2)' P_Z (y_1 - \beta_0 y_2)}{(y_1 - \beta_0 y_2)' M_Z (y_1 - \beta_0 y_2)/(n - k - 1)}.$$

Under H_0,

$$AR = \frac{u' P_Z u}{(u'u - u' P_Z u)/(n - k - 1)} \to_d \chi^2(k).$$

Thus, the AR test uses the $1-\alpha$ quantile of $\chi^2(k)$ as critical value and rejects H_0 iff $AR > \chi^2_{1-\alpha}(k)$.

The AR test is fully robust to weak instrument because the AR statistic's null distribution has nothing to do with Π. Assuming that u_i is normal with mean zero and homoskedastic variance, and assuing Z is fixed, the AR statistic is actually pivotal under the null – its distribution is fully known under the null.

Under the alternative $\beta \neq \beta_0$,

$$AR = \frac{(u + (\beta - \beta_0) Z\Pi)' P_Z ((\beta - \beta_0) Z\Pi + u)}{(u'u - u' P_Z u)/(n - k - 1)},$$

thus, the power of the AR test depends on $(\beta - \beta_0)^2 \Pi' Z' Z \Pi / \sigma_u^2$. When $\Pi = 0$, that is, the instruments are completely irrelevant, AR test has no power, which is evidence that AR is robust because in that case no test should have power. This is a feature for all the fully robust tests discussed below.

A fully robust confidence interval for β can be constructed by inverting the AR test.

3 Conditional Test

The AR test has very good power, and is in fact shown to be uniformly most powerfu unbiasedl <u>when $k = 1$</u> in Moreira (2001). However, its power deteriorates quickly when k increases because the critical value increases quickly with k. Better tests are developed using the conditional test idea proposed in Moreira (2001, 2003).

To introduce the conditional test, we first review the concept of sufficient statistics.

Consider a parametric family of distributions $\{f(W, \theta) : \theta \in \Theta\}$ of the observable W.

Example: the number of successes is a sufficient statistic for the success rate given the i.i.d. binomial sample $X_1, ..., X_n$.

Definition A statistic $T(W_1, ..., W_n)$ is a sufficient statistic for a parameter θ if the distribution of $W_1, ..., W_n$ given T does not depend on θ.

Factorization Theorem: Let $f_W(w; \theta)$denote the pdf of a random variable (vector) W. A statistic $T(X)$ is a sufficient statistic for θ if and only if there are functions $g(t)$ and $h(x)$ such that the pdf can be written as $f_W(w;\theta) = g(t(w);\theta) \cdot h(x)$ for all values of θ.

Now consider the reduced form model

$$y_1 = Z\Pi\beta + v_1, \quad and \quad y_2 = Z\Pi + v_2, \tag{7}$$

where $v_1 = \beta v_2 + u$. Then $(v_{1,i}, v_{2,i}) \sim N(0,\Omega) = N(0, [\sigma_1^2, \rho\sigma_1\sigma_2; \rho\sigma_1\sigma_2, \sigma_2^2])$, where $\sigma_1 = \beta^2\sigma_v^2 + \sigma_u^2 + 2\beta\sigma_{uv}$, $\rho\sigma_1\sigma_2 = \beta\sigma_v^2 + \sigma_{uv}$ and $\sigma_2^2 = \sigma_v^2$.

Suppose that Ω is known and Z is nonrandom. The likelihood of this model (which is also the joint density of y_1 and y_2) is

$$f(y_1, y_2; \theta, \Pi) = (2\pi)^{n/2}|\Omega|^{n/2} \exp\left(-\left(\sum_{i=1}^n Y_i\Omega^{-1}Y_i' - 2\Pi' Z' Y\Omega^{-1}a + a'\Omega^{-1}\Pi' Z' Z\Pi\right)/2\right), \tag{8}$$

where $Y_i = [y_{1,i}, y_{2,i}]$ and $Y = [y_1, y_2]$ and $a = [\beta, 1]'$.

By the Factorization Theorem, $Z'Y$ is a sufficient statistic for $(\beta, \Pi')'$. Let $D = [b_0, \Omega^{-1}a_0]$, where $b_0 = [1, -\beta_0]$ and $a_0 = [\beta_0, 1]$. Then D is invertible, and thus $Z'YD \equiv [Z'(y_1 - \beta_0 y_2), Z'Y\Omega^{-1}a_0]$ is also a sufficient statistic. And it can be shown that $S := Z'(y_1 - \beta_0 y_2)$ and $T := Z'Y\Omega^{-1}a_0$ are independent normal random vectors. Under the null, the first vector is simply $Z'u$ and its distribution does not depend on Π. The second vector has distribution that depends on Π.

According to the Rao-Blackwell thoerem (e.g. Theorem 7.3.1, Hogg, McKean and Craig, 2005), there is no loss of information by only considering test statistics that are functions of the sufficient statistics.

For any test statistic that is a function of S, T, say $\psi(S, T; \beta_0)$, the null distribution of it is not pivotal – it depends on the nuisance parameter Π because T's distribution depends on Π. However, due to the independence between S and T, the conditional null distribution of $\psi(S, T; \beta_0)$ given $T = t$ is simply the distribution of $\psi(S, t; \beta_0)$ and is pivotal – it does not depend on Π. The distribution of $\psi(S, t; \beta_0)$ is known because Z is fixed and u's distribution is assumed to be known. Thus, its $(1-\alpha)$ quantile is also known. Let the quantile be $c(S, t, 1-\alpha; \beta_0)$. If the distribution of $\psi(S, t; \beta_0)$ is continuous and strictly increasing at its $1-\alpha$ quantile, then we have

$$\Pr(\psi(S, t; \beta_0) > c(S, t, 1-\alpha; \beta_0)) = \alpha, \tag{9}$$

for any t. This implies that $\Pr(\psi(S, T; \beta_0) > c(T, 1-\alpha; \beta_0)) = \alpha$. This clearly holds regardless of Π.

This powerful technique allows us to construct a fully robust test from any test statistic that is a function of S and T.

A few examples of test statistics that are functions of S and T are as follows:

1. the AR test statistic with known Ω:

$$AR = \frac{(y_1 - \beta_0 y_2)' P_Z (y_1 - \beta_0 y_2)}{\sigma_u^2} = S'(Z'Z)^{-1}S/\sigma_u^2 \sim \chi^2(k).$$

   AR does not depend on T. Thus, $c_{AR}(T, 1-\alpha; \beta_0) = \chi^2_{1-\alpha}(k)$.

2. the 2SLS-based t-statistic.

$$t_n = \frac{y_2' Z(Z'Z)^{-1}Z'(y_1 - \beta_0 y_2)}{\sigma_u \sqrt{y_2' Z(Z'Z)^{-1}Z' y_2}} = \frac{(T' - c_1 S')(Z'Z)^{-1}S}{\sigma_u \sqrt{(T' - c_1 S')(Z'Z)^{-1}Z'(T - c_1 S)/c_2}},$$

   for some constants c_1, c_2 that depends on Ω and β_0. Thus, the critical value $c_{t_n}(t, 1-\alpha; \beta_0)$ can be the the quantile of

$$\frac{(t' - c_1 u'Z)(Z'Z)^{-1}Z'u}{\sigma_u \sqrt{(t' - c_1 u'Z)(Z'Z)^{-1}Z'(t - c_1 Zu)/c_2}},$$

   where $Z'u \sim N(0, Z'Z\sigma_u^2)$. The critical value can be easily simulated.

3. the LR statistic. The usual LR statistic based on the likelihood we wrote down above can be written into

$$LR = (1/2)(S'S - T'T + \sqrt{(S'S - T'T)^2 + 4(S'T)^2}),$$

   The critical value is the $1-\alpha$ quantile of $(1/2)(S'S - t't + \sqrt{(S'S - t't)^2 + 4(S't)^2})$ evaluated at $t = T$. It can also be easily simulated because $S \sim N(0, Z'Z\sigma_u^2)$ and $\sigma_u^2 = b_0'\Omega b_0$.

4. the LM statistic:

$$LM = (S'T)^2/(T'Z'ZT \cdot \sigma_u^2)$$

   The critical value is $\chi^2_{1-\alpha}(1)$ because $LM|T = t \sim \chi^2(1)$.

Now that we have a number of fully-robust tests once we use the conditional technique. Which test is the best? Andrews and Stock (2005) recommends the conditional LR test because numerical evidence show that show that the CLR test's power function is close to a power envelope derived in Andrews, Moreira and Stock for all two-sided invariant similar tests.