Recitation 3: Various tests in IV-GMM framework

Seung-hun Lee

Columbia University
Introduction to Econometrics II Recitation

February 7th, 2022

Primer on Wald vs LM vs LR tests

Engle (1984): Wald uses unconstrained estimator and its distance

- Suppose we have the data y, parameter of interest β and the log likelihood function $L(y, \beta)$. The hypothesis we want to check is

$$H_0 : \beta = \beta_0 \text{ vs } H_1 : \beta \neq \beta_0$$

- Wald: This is a test based on unconstrained regression in the sense that it tests the composite alternative hypothesis against the null. The idea is to accept the null when the estimated value of β is reasonably close to β_0. The typical test statistics is

$$t_W = (\hat{\beta} - \beta_0)'(var(\hat{\beta}))^{-1}(\hat{\beta} - \beta_0) \sim \chi^2_{\dim(\beta)}$$

Engle (1984): LM uses constrained regressions

- LM: This is based on the constrained minimization problem of the likelihood function. We impose the null hypothesis with the following constrained optimization problem

  $$L(y, \beta) - \lambda'[\beta - \beta_0]$$

  with the first order conditions
  - With respect to β: $\frac{\partial L}{\partial \beta} = \lambda$
  - With respect to λ: $\beta = \beta_0$

  By complementary slackness conditions, we would have $\lambda \geq 0$ and $\lambda = 0$ if the constraint is not binding. If the shadow price is high (high $\lambda = s(y, \beta_0)$), then we would reject the constraint. Typically, the test statistics used has this form:

  $$t_{LM} = s(y, \beta_0)'(var(s))^{-1} s(y, \beta_0) \sim \chi^2_{\dim(\beta)}$$

Engle (1984): Use both restricted and unrestricted regressions

- LR: This is based on the difference between the maximum of the likelihood under the null vs under the alternative. Typically, the test statistics have the following form

$$t_{LR} = -2[L(y, \beta_0) - L(y, \hat{\beta})] \sim \chi^2_{\dim(\beta)}$$

- In short,

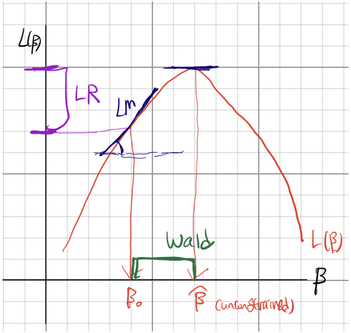

Figure: Wald vs LM vs LR

Tests of exongeneity and endogeneity

Hansen's test: Are the (overidentified) instruments really exogenous?

- Setup: k covariates x_i and l-vector instrument z_i ($l > k$)
- Hansen's J-test: $H_0 : E[g(w_i, \beta)] = 0$ vs. $H_1 : E[g(w_i, \beta)] \neq 0$ where $g(w_i, \beta) = z_i e_i$
  - Make use of a GMM criterion: $J(\beta) = n\bar{g}(\beta)'\widehat{\Omega}^{-1}\bar{g}(\beta)$.
  - Asymptotics:

$$J(\beta) = \underbrace{\left(\sqrt{n}\bar{g}(\beta)\right)'}_{\xrightarrow{d} N(0,\Omega)} \widehat{\Omega}^{-1} \underbrace{\left(\sqrt{n}\bar{g}(\beta)\right)}_{\xrightarrow{d} N(0,\Omega)} \xrightarrow{d} \chi^2_l$$

  - However, we need an estimate of e_i using $\hat{\beta}_{GMM}$.
  - We really do this with $J(\hat{\beta}_{GMM}) \sim \chi^2_{l-k}$. (We estimate k parameters)
  - For significance level α, compare $J(\hat{\beta}_{GMM}$ with $\chi^2_{l-k,\alpha}$ where $\Pr(\chi^2 \geq \chi^2_{l-k,\alpha}) = \alpha$
- The test introduced here also tests for other issues with the specification, so it is possible to reject this test for reasons other than instrument endogeneity
- This is robust to heteroskedasticity (We have not set any restrictions on $\Omega = E[z_i z_i' e_i^2]$)

Sargan's test: Hansen's test with homoskedasticity

- Assume

$$y_i = x_i'\beta + e_i \ (\dim(x_i) = k < \dim(z_i) = l)$$

- Construct a linear projection equation by projecting e_i onto z_i, obtaining

$$e_i = z_i'\delta + \epsilon_i \rightarrow \delta = E(z_i z_i')^{-1} E(z_i e_i)$$

- So we need to test $H_0 : \delta = 0$ vs. $H_1 : \delta \neq 0$
  - **Obtain** $\hat{e}_i$: This can be done using 2SLS estimates of β
  - **Obtain** $\hat{\delta}$: Replace e with $\hat{e}$ to get $\hat{\delta} = (Z'Z)^{-1}Z'\hat{e}$ **Sargan Test**: Assuming homoskedasticity, we use

$$S = \hat{\delta}'(var(\hat{\delta}))^{-1}\hat{\delta} = \frac{\hat{e}'Z(Z'Z)^{-1}Z'\hat{e}}{\hat{\sigma}^2} \xrightarrow{d} \chi^2_{l-k} \quad (\hat{\sigma}^2 = \frac{1}{n}\hat{e}'\hat{e})$$

- Same caution as before applies!

Subset exogeneity test: Some are certain, but others are not

- Partition z_i into two sets - $z_{ai} \in \mathbb{R}^{l_a}$ and $z_{bi} \in \mathbb{R}^{l_b}$.
- We are uncertain about z_{bi} and want to test

$$H_0 : E(z_{bi}e_i) = 0, \text{ vs. } H_1 : E(z_{bi}e_i) \neq 0$$

- Distance-test: Are GMM criterion values similar?
  - Estimate the model by the efficient GMM with only the z_{ai} set of instruments and obtain the GMM criterion J_a
  - Estimate the model with the full set of instruments and obtain a separate GMM criterion, denoted as $J_{a,b}$
  - Then, create a test statistic

$$C = J_{a,b} - J_a \xrightarrow{d} \chi^2_{l - l_a = l_b}$$

  - Find critical value $\chi^2_{l_b,\alpha}$ for a significance level α and reject the null hypothesis if $C > \chi^2_{l_b,\alpha}$

Other ways to do subset exogeneity test

- Amemiya-Lee-Newey: Use an auxiliary regression
  - Project e_i onto the ones we are certain about (z_a) and the ones we are not (z_b)

$$e_i = z_{ai}'\delta_a + z_{bi}'\delta_b + u_i$$

  - Test for the exogeneity of z_b instruments by checking $\delta_b = 0$
  - In practice: Obtain residuals $\hat{e}$ and run an auxiliary regression of this into the two sets of instruments and obtain estimates for δ_b
  - The test statistics used and its distribution under the null is

$$\hat{\delta}_b'(\text{var}(\hat{\delta}_b))^{-1}\hat{\delta}_b \sim \chi^2_{l_b}$$

Hausman test can be used here too!

- Using Hausman principle
  - Obtain two estimates of β - one using only the sure IVs and the other using all sets of IVs
  - Then, the idea is to check whether the two estimates are close to each other.
- To use this
  - Make sure z_a is always exogenous
  - We need to make sure that z_a is a consistent yardstick to compare the full set of instruments against

Hausman principles: Key is getting the right two estimators

- Hausmann Test can be utilized as a general test of specification.
- It is aimed at testing the consistency of an estimator that we are uncertain about relative to an estimator that is surely consistent.
- The general trick is that you need two types of estimators.
  - $\hat{\theta}_1$: Consistent and efficient under H_0, inconsistent under H_1
  - $\hat{\theta}_2$: Consistent in either H_0 or H_1. Inefficient under H_0.
- Then the difference between the two estimators have the following asymptotic distribution

$$\sqrt{n}(\hat{\theta}_1 - \hat{\theta}_2) \sim N(0, \text{var}(\hat{\theta}_2 - \hat{\theta}_1))$$

- The test statistic that is used is

$$H = (\hat{\theta}_1 - \hat{\theta}_2)'(\text{var}(\hat{\theta}_2 - \hat{\theta}_1))^{-1}(\hat{\theta}_1 - \hat{\theta}_2)$$

Endogeneity/Exogeneity of regressors x_i

- Assume that homoskedasticity is satisfied and the data generating process is

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + e_i$$

- We test $H_0 : E(x_{2i}e_i) = 0$ against $H_1 : E(x_{2i}e_i) \neq 0$
- Consider these properties of 2SLS and OLS estimators
  - $\hat{\beta}_{OLS}$: Consistent and minimal variance under H_0, inconsistent under H_1
  - $\hat{\beta}_{2SLS}$: Consistent in either H_0 or H_1. Inefficient under H_0.
- Under H_0, $\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})$ converges in distribution to $N(0, var(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}))$
- Hausman: $var(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) = var(\hat{\beta}_{2SLS}) - var(\hat{\beta}_{OLS})$ holds in homoskedasticity
- Thus the Hausman test statistic

$$(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})'(var(\hat{\beta}_{2SLS}) - var(\hat{\beta}_{OLS}))^{-1}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) \xrightarrow{d} \chi^2_{k_2}$$

Using Control function is also possible

- Assume the following setup

$$
\begin{aligned}
y_i &= x_{2i}'\beta_2 + e_i && (E[x_{2i}e_i] \neq 0) \\
x_{2i} &= z_i'\pi_2 + v_{2i} && (E[z_i e_i] = 0, E[z_i v_{2i}] = 0) \\
e_i &= v_{2i}'\rho + w_i \\
y_i &= x_{2i}'\beta_2 + v_{2i}\rho + w_i
\end{aligned}
$$

- Replace this with $\hat{v}_{2i}$ by running an OLS on the first stage regression and then write

$$
y_i = x_{2i}'\beta_2 + \hat{v}_{2i}\rho + w_i
$$

So what are we looking for?

- What we need to do is to check whether $\hat{\rho} = 0$ or not
- Recall that
$$E[x_{2i}e_i] = E[(z_i'\pi_2 + v_{2i})e_i] = E[v_{2i}e_i] \ (\because E[z_ie_i] = 0)$$
- From the projection of e_i onto v_{2i}
$$\rho = E[v_{2i}v_{2i}']^{-1}E[v_{2i}e_i] \rightarrow \hat{\rho} = \frac{1}{n}\sum_{i=1}^{n}(v_{2i}v_{2i}')\frac{1}{n}\sum_{i=1}^{n}(v_{2i}e_i)$$
- If x_2 is exogenous, we would get $\rho = 0$. So we would need to test whether $\hat{\rho}$ is close to zero or not.

Variable addition test and Hausman principle

- Suppose that you found an instrumental variable (or variables) for your endogenous X_2, and that you went ahead with your 2SLS.
- Then you may suddenly wonder whether it was worth getting an IV to begin with.
- One way of testing for the endogeneity of your X_2 is to see if the 2SLS and the OLS estimates differ drastically.
- The idea is that if there is not much endogeneity, the two estimates should be similar.
- Note that

$$\hat{\beta}_{2SLS} = (X'P_Z'X)^{-1}(X'P_Z'y)$$
$$\hat{\beta}_{OLS} = (X'X)^{-1}(X'y)$$

So how do we make use of the differences?

- The difference between the two is

$$
\begin{aligned}
\hat{\beta}_{2SLS} - \hat{\beta}_{OLS} &= (X'P_Z'X)^{-1}(X'P_Z'y) - (X'X)^{-1}(X'y) \\
&= (X'P_Z'X)^{-1}[(X'P_Z'y) - (X'P_Z'X)(X'X)^{-1}(X'y)] \\
&= (X'P_Z'X)^{-1}X'P_Z'[y - X(X'X)^{-1}(X'y)] \\
&= (X'P_Z'X)^{-1}X'P_Z'[y - P_X y] \\
&= (X'P_Z'X)^{-1}X'P_Z'[I - P_X]y \\
&= (X'P_Z'X)^{-1}X'P_Z'M_X y \\
&= (\widehat{X}'X)^{-1}\widehat{X}'\widehat{e}
\end{aligned}
$$

- The takeaway is that testing the endogeneity/exogeneity of X with respect to e is equivalent to testing for $X'P_Z'M_X y = 0$.

Building up the test statistics

- In the original equation where we had $y = X\beta + e$, we add a regressor $\widehat{X}$.
- Effectively, we are working with an auxiliary regression in this form

$$y = X\beta + \widehat{X}\gamma + \epsilon$$

- Note that the $\hat{\gamma}$ can be written as

$$\begin{aligned}\hat{\gamma} &= (\widehat{X}' M_X X)^{-1}(\widehat{X}' M_X y) \ (\because \text{Frisch-Waugh-Lovell}) \\ &= (X' P_Z' M_X X)^{-1}(X' P_Z' M_X y)\end{aligned}$$

- So when the X is exogenous, we have $\hat{\gamma} = 0$
- Since we are using two types of estimators, we are effectively doing a Hausman test.

Weak IV

Why is weak IV problematic?

- Set the equation as

$$y_i = x_i'\beta + e_i \ (E[x_i e_i] \neq 0)$$

and write the first stage equation as

$$x_i = \Gamma' z_i + v_i \ \ (x_i \in \mathbb{R}^k, z_i \in \mathbb{R}^l, \Gamma \in \mathbb{R}^{l \times k})$$

- Assume IV setting ($l = k$), $\hat{\beta}_{IV} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i'} = \beta + \frac{\frac{1}{n}\sum_{i=1}^n z_i e_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i'}$
- We assume that we have an irrelevant IV in the sense that $\frac{1}{n}\sum_{i=1}^n z_i x_i' \xrightarrow{p} 0$
- The asymptotics now look like

$$\hat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i e_i}{\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i x_i'}$$

- This is a Cauchy distribution, which has no moments. Our estimates are inconsistent

Sit back and relax: Local to zero framework

- Assume that the structural and reduced form equations are (we are working with a scalar regressors)

$$y_i = x_i\beta + e_i$$
$$x_i = z_i\gamma + u_i$$

- We say that there is a problem of weak instrument if $\gamma \simeq 0$. Specifically, let $\gamma = \frac{\mu}{\sqrt{n}}$.
- For simplicity, I will assume
  - $var\left(\begin{pmatrix} e_i \\ u_i \end{pmatrix} |z_i\right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \Sigma$
  - $E(z_i^2) = 1$
  - $\frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} z_i e_i \\ z_i u_i \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = N(0, \Sigma)$

Sit back and relax: OLS fails to be consistent

- OLS: Note that $\hat{\beta}_{OLS} - \beta$ can be written as $\frac{n^{-1}\sum_{i=1}^n x_i e_i}{n^{-1}\sum_{i=1}^n x_i^2}$, equivalent to

$$
\begin{aligned}
\frac{n^{-1}\sum_{i=1}^n x_i e_i}{n^{-1}\sum_{i=1}^n x_i^2} &= \frac{n^{-1}\sum_{i=1}^n (z_i\gamma + u_i)e_i}{n^{-1}\sum_{i=1}^n (z_i\gamma + u_i)^2} \\
&= \frac{n^{-1}\sum_{i=1}^n (z_i e_i\gamma + u_i e_i)}{n^{-1}\sum_{i=1}^n (z_i^2\gamma^2 + u_i^2 + 2z_i u_i\gamma)} \\
&= \frac{n^{-1}\sum_{i=1}^n u_i e_i}{n^{-1}\sum_{i=1}^n u_i^2} + o_p(1) \\
&\xrightarrow{p} E(u_i e_i)/E(u_i^2) = \rho
\end{aligned}
$$

Sit back and relax: IV estimator does not do any better

- IV: For $\hat{\beta}_{IV} - \beta$, we can write $\frac{n^{-1/2}\sum_{i=1}^{n} z_i e_i}{n^{-1/2}\sum_{i=1}^{n} z_i x_i}$ or equivalently

$$\begin{aligned}\hat{\beta}_{IV} - \beta &= \frac{n^{-1/2}\sum_{i=1}^{n} z_i e_i}{n^{-1/2}\sum_{i=1}^{n} z_i(z_i\gamma + u_i)} \\ &= \frac{n^{-1/2}\sum_{i=1}^{n} z_i e_i}{n^{-1/2}\sum_{i=1}^{n} z_i u_i + n^{-1}\sum_{i=1}^{n} z_i^2 \mu} \xrightarrow{d} \frac{\xi_1}{\xi_2 + \mu}\end{aligned}$$

  which is non-normal, making it inconsistent.
- Detection: Is the first stage F-statistic values sufficiently high? The rule of thumb part is to see if this value is greater than 10 or not (surprisingly many papers do this still)