14.383

Fall 2003

Handout # 6

(Testing; GMM, NLIV)

T.A.: Alexis León
aleon@mit.edu
E52-351

TESTING

There have been questions concerning what each of the tests presented is actually testing. Let's clarify.

Hausman Test:

This can be used for a variety of "tests". It requires three things for estimators $\underset{k\times 1}{\hat{\delta}_0}, \underset{k\times 1}{\hat{\delta}_1}$:

1) $\hat{\delta}_0, \hat{\delta}_1$ have the same plim under H_0
2) $\hat{\delta}_0, \hat{\delta}_1$ have different plims under H_1
3) One estimator (WLOG, assume $\hat{\delta}_1$) is efficient under H_0

Then the estimator is given by

$$H = (\hat{\delta}_0 - \hat{\delta}_1)'\left[\widehat{Var}(\hat{\delta}_0) - \widehat{Var}(\hat{\delta}_1)\right]^{-1}(\hat{\delta}_0 - \hat{\delta}_1) \xrightarrow{d} \chi^2_k$$

* Note, $\widehat{Var}(\hat{\delta}_0)$ is the estimated variance of $\hat{\delta}_0$.

If we wish to use $\widehat{AVar}(\hat{\delta}_0)$ where $\sqrt{T}(\hat{\delta}_0 - \delta_0) \longrightarrow N(0, AVar(\hat{\delta}_0))$, then

$$H = T(\hat{\delta}_0 - \hat{\delta}_1)'\left[\widehat{AVar}(\hat{\delta}_0) - \widehat{AVar}(\hat{\delta}_1)\right]^{-1}(\hat{\delta}_0 - \hat{\delta}_1) \xrightarrow{d} \chi^2_k$$

Hausman Test (Cont):

What can we test with the Hausman test?
In general, we test if something is correctly specified.

Example 1: $\underset{L}{Z} = [\underset{L_0}{Z_0} \quad \underset{L_1}{Z_1}]$ $\quad L = L_0 + L_1$

We are confident Z_1 are valid instruments but are not sure about Z_0.

$$\underset{T\times 1}{Y} = \underset{T\times k}{X}\,\underset{k\times 1}{\delta} + \underset{T\times 1}{\epsilon} \qquad \text{Assume } L_1 \geq K$$

$$H_0: \begin{cases} E(Z_{0t}'\epsilon_t) = 0 \\ E(Z_{1t}'\epsilon_t) = 0 \end{cases} \quad \text{vs.} \quad H_1: \begin{cases} E(Z_{0t}'\epsilon_t) \neq 0 \\ E(Z_{1t}'\epsilon_t) = 0 \end{cases}$$

Compute $\hat{\delta}_{0,OIV}$ using Z_1 as set of instruments
$\hat{\delta}_{1,OIV}$ using Z as set of instruments

- Both are consistent under H_0 => same plim under H_0
- $\hat{\delta}_0$ consistent under H_1, $\hat{\delta}_1$ inconsistent under H_1
  => diff. plims under H_1
- $\hat{\delta}_1$ efficient under H_0

=> test H_0 with

$$H = (\hat{\delta}_0 - \hat{\delta}_1)'[\hat{Var}(\hat{\delta}_0) - \hat{Var}(\hat{\delta}_1)]^{-1}(\hat{\delta}_0 - \hat{\delta}_1) \xrightarrow{d} \chi^2_k$$

Hausman Test (Cont)

Suppose in the full information setting, we are unsure that one or more equations is correctly specified.

Example 2:

$$y_1 = Y_1 \beta_1 + Z_1 \gamma_1 + u_1 = \underset{k_1}{X_1} \delta_1 + u_1$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \underset{\left(\sum_{i=1}^{5} k_i \times 1\right)}{\delta} = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_5 \end{pmatrix}$$

$$y_5 = Y_5 \beta_5 + Z_5 \gamma_3 + u_5 = \underset{k_5}{X_5} \delta_5 + u_5$$

We fear that equations 4, 5 are incorrectly specified.

Let $K = k_1 + k_2 + k_3$

1) Let $\hat{\delta}_1$ be a Full Information estimator of δ. For example, 3SLS or FIML. Under H_0, that all equations are correctly specified, $\hat{\delta}_1$ is consistent and efficient. Therefore, $[\hat{\delta}_1]_K$ (first K elements of $\hat{\delta}_1$) is efficient and consistent for $\begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix}$. Under H_1, that eq.s 4, 5 are incorrectly specified, $[\hat{\delta}_1]_K$ is inconsistent.

Hausman Test (Cont)

Example 2 (Cont):

2) Let $\hat{\delta}_0$ be a consistent estimate of $\begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix}$ under both the null and alternative.

Choices include: (a) 2SLS eqn. by eqn on Equations 1, 2, 3

(b) 3SLS or FIML on the system composed of Eqns. 1, 2, 3.

Then $\hat{\delta}_0$, $\hat{\delta}_1$ satisfy the requirements of a Hausman Test:

$$H = (\hat{\delta}_0 - \hat{\delta}_1)' [\hat{Var}(\hat{\delta}_0) - \hat{Var}(\hat{\delta}_1)]^{-1} (\hat{\delta}_0 - \hat{\delta}_1) \xrightarrow{d} \chi^2_k$$

* When choosing $\hat{\delta}_0$, choosing (b) gives a more efficient $\hat{\delta}_0$ (lower variance) and improves the power of your Hausman test.

Hausman Test (Cont)

Now we give a Hausman test that parallels the Omnibus Test. Suppose we want to check the exogeneity of included endog. variables and the validity of the instruments for an equation.

Example 3:

$$y_1 = Y_1 \beta_1 + Z_1 \gamma_1 + u_1 = \underset{\overset{\leftrightarrow}{k}}{X} \delta + u_1$$

H_0: Instruments and errors uncorrelated, $E(Z_t' u_t) = 0$

H_1: Istruments correlated with errors, $E(Z_t' u_t) \neq 0$.

This calls for the Hausman Specification test.

Under correct specification (H_0),

plim $\hat{\delta}_{2SLS}$ = plim $\hat{\delta}_{OIV} = \delta$ and $\hat{\delta}_{OIV}$ is efficient.

Under wrong specification (H_1), $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{OIV}$ are inconsistent with different plims.

$$H = (\hat{\delta}_{2SLS} - \hat{\delta}_{OIV})' [\hat{V}(\hat{\delta}_{2SLS}) - \hat{V}(\hat{\delta}_{OIV})]^{-1} (\hat{\delta}_{2SLS} - \hat{\delta}_{OIV}) \xrightarrow{d} \chi^2_K$$

The specification tested in Example 3 can also be addressed using:

The Omnibus Test:

The Omnibus Test, like the Hausman test described in Example 3, rejects if the instruments are correlated with the errors.

$$y_1 = Y_1\beta_1 + Z_1\gamma_1 + u_1 = \underset{K}{X}\delta + u_1$$

$$\underset{L}{Z} = (Z_1 \;\; Z_2) \qquad L \geq K$$

Degree of overidentification given by $L-K$
This will give the degrees of freedom in our Chi-squared estimator.

Let $Z'u/\sqrt{T} \xrightarrow{d} N(0, V)$. Then the estimator is given by

$$W = \left(\frac{Z'\hat{u}}{\sqrt{T}}\right)' \hat{V}^{-1} \left(\frac{Z'\hat{u}}{\sqrt{T}}\right) \xrightarrow{d} \chi^2_{L-K}$$

* For a proof of this result, please see Appendix of Section 3 Notes.

Omnibus Test (Cont)

In the case that u is spherical,

$$V = \sigma^2 \operatorname{plim}(Z'Z/T), \quad \hat{V} = \hat{\sigma}^2\left(\frac{Z'Z}{T}\right) = \left(\frac{\hat{u}'\hat{u}}{T}\right)\left(\frac{Z'Z}{T}\right)$$

$$\Rightarrow W = T\,\frac{\hat{u}'P_Z\hat{u}}{\hat{u}'\hat{u}} \xrightarrow{d} \chi^2_{L-k}$$

An equivalent way of computing this statistic is to first regress $\hat{u}$ on Z, obtain the R^2 of the regression, and let $TR^2 \xrightarrow{d} \chi^2_{L-k}$ be your statistic.

The omnibus Test presented thus far is for the single equation (Limited Information) case. An analogous statistic exists for the system. Write

$$\bar{y} = \bar{X}\bar{\delta} + \bar{u} \qquad \text{stacked model}$$

$$\bar{X} = \begin{pmatrix} \underset{\leftrightarrow k_1}{X_1} & & & \\ & \underset{\leftrightarrow k_2}{X_2} & & \\ & & \ddots & \\ & & & \underset{\leftrightarrow k_m}{X_m} \end{pmatrix}, \quad \underset{\leftrightarrow L}{Z}, \quad p_i = L - k_i, \quad p = \sum_{i=1}^{m} p_i$$

$$W = \hat{\bar{u}}'(\hat{\Sigma}^{-1} \otimes P_Z)\hat{\bar{u}} \xrightarrow{d} \chi^2_{\underset{\sim}{p}}$$

deg. of overidentification in the system

Generalized Method of Moments : GMM

Recall the method of moments estimator.

Suppose for a random variable y_i, we have $E(y_i) = \mu$. How can we estimate μ?

Write the moment condition as $E(y_i - \mu) = 0$.

The sample analog of the moment condition:

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\mu}) = 0$$

Now, solve for $\hat{\mu}$: $\hat{\mu} = \frac{1}{n} \sum y_i$

This is the simplest example of method of mome.

It has one equation with one parameter to be estimated. Suppose y_i has the additional property that $E(y_i^3) = \mu = E(y_i)$. Now we have two moment conditions giving two sample analogs and solving for one parameter

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\mu}) = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} (y_i^3 - \hat{\mu}) = 0$$

How do we solve for $\hat{\mu}$? This is where GMM comes in.

9

GMM (Cont)

Combine the sample analogs:

$$\frac{1}{n}\sum_{i=1}^{n}\left[\begin{pmatrix} y_i \\ y_i^3 \end{pmatrix} - \begin{pmatrix} \hat{\mu} \\ \hat{\mu} \end{pmatrix}\right] = \underset{2\times 1}{0}$$

We can find the GMM estimator of μ by

$$\hat{\mu}_{GMM} = \underset{\mu}{\text{argmin}}\ \underset{1\times 2}{\hat{g}(\mu)'}\ \underset{2\times 2}{\hat{A}}\ \underset{2\times 1}{\hat{g}(\mu)}$$

where $\underset{2\times 1}{\hat{g}(\mu)} = \frac{1}{n}\sum_{i=1}^{n} g_i(\mu)$, $\underset{2\times 1}{g_i(\mu)} = \begin{pmatrix} y_i \\ y_i^3 \end{pmatrix} - \begin{pmatrix} \hat{\mu} \\ \hat{\mu} \end{pmatrix}$

and $\hat{A}$ is some weighting matrix.

The optimal $\hat{A}$ is given by $\left[\hat{Var}(\hat{g}(\mu))\right]^{-1}$.

The general result for this GMM estimator is that for $\hat{A}$ positive definite and $\text{plim}\ \hat{g}(\mu) = 0$, $\hat{\mu}_{GMM}$ is consistent. The asymptotic covariance matrix of the GMM estimator is given by

$$V_{GMM} = \left[G'\hat{A}G\right]^{-1} = \left[G'\hat{Var}(\hat{g}(\mu))^{-1}G\right]^{-1} = \frac{1}{n}\left[G'\hat{W}^{-1}G\right]$$

where $G = \text{plim}\frac{\partial \hat{g}(\mu)}{\partial \mu'}$, $\hat{W} = A\hat{Var}(\hat{g}(\mu))$

Applying the CLT, we have

$$\sqrt{n}(\hat{\mu}_{GMM} - \mu) \xrightarrow{d} N(0, (\hat{G}'\hat{W}^{-1}\hat{G})^{-1})$$

GMM (Cont) - IV

Now lets apply GMM to our IV framework.

Suppose $y_t = h(X_t, \beta) + \epsilon_t$, $\underset{T \times L}{Z}$, $\underset{K \times 1}{\beta}$, $L \geq K$

$E(\epsilon) = 0$, $E(\epsilon\epsilon') = \Sigma$.

In the Newey handout, $h(X_t, \beta) = X_t'\beta$.

Some of our RHS variables are endogenous and we instrument using Z. The moment conditions come from: $E(Z_t \epsilon_t) = 0$.

Write $g_t(\beta) = \underset{L \times 1}{Z_t}(y_t - h(X_t, \beta))$

$$\hat{g}(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\beta)$$

Then the GMM estimator of β is given by

$$\hat{\beta}_{GMM} = \underset{\beta}{\operatorname{argmin}} \; \underset{1 \times L}{\hat{g}(\beta)'} \underset{L \times L}{\hat{A}} \underset{L \times 1}{\hat{g}(\beta)}$$

where $\hat{A} = \hat{Var}(\hat{g}(\beta))^{-1}$

Since $Var(\hat{g}(\beta)) = \frac{1}{T} AVar(\hat{g}(\beta))$, this can be written:

$$\hat{\beta}_{GMM} = \underset{\beta}{\operatorname{argmin}} \; T \cdot \hat{g}(\beta)' \left[A\hat{Var}(\hat{g}(\beta))\right]^{-1} \hat{g}(\beta)$$

GMM - IV (Cont)

The Newey handout outlines what $\widehat{AVar}(\hat{g}(\beta))$ is in the cases of

1) no homoskedasticity and no autocorrelation

$$\widehat{AVar}(\hat{g}(\beta)) : \sqrt{T}\,\hat{g}(\beta) = \frac{Z'\epsilon}{\sqrt{T}} \rightarrow N\left(0, \sigma^2\left(\frac{Z'Z}{T}\right)\right)$$

$$\Rightarrow \widehat{AVar}(\hat{g}(\beta)) = \left(\frac{Z'Z}{T}\right) \quad \text{(dropping the constant)}$$

2) just homoskedasticity
3) just autocorrelation
4) both

$\Big\}$ See handout, Pg. 3

Again we have $\sqrt{T}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, (\hat{G}'\hat{W}^{-1}\hat{G})^{-1})$

where $G = \text{plim} \frac{\partial g(\beta)}{\partial \beta}$, $\hat{W} = \widehat{AVar}(\hat{g}(\beta))$

GMM - System IV

So far, we have been dealing with only one equation. Suppose we have M equations.

Then we have $E(Z_t \epsilon_{it}) = 0$, $t=1,..,T$, $i=1,..,M$

The sample analog to this is:

$$\frac{1}{T} \sum_{t=1}^{T} Z_t (y_{it} - h_i(\bar{X}, \bar{\beta})) = 0 \quad , i=1,..,M$$

If we use this result for each equation i, this would be GMM equation by equation.

But this neglects correlation between disturbances, and has the additional efficiency loss of not imposing cross-equation constraints.

Define $\underset{L \times L}{\frac{1}{T} Z' \Sigma_{ij} Z} = E[Z' \epsilon_i \epsilon_j' Z]$

Then the GMM estimator of $\bar{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$ is

$$\hat{\bar{\beta}}_{GMM} = \underset{\bar{\beta}}{\text{argmin}} \sum_{i=1}^{M} \sum_{j=1}^{M} [\epsilon_i(\bar{\beta})' Z][Z' \Sigma_{ij} Z]^{-1} [Z' \epsilon_j(\bar{\beta})]$$

where $\epsilon_i(\bar{\beta}) = y_i - h_i(\bar{X}, \bar{\beta})$