Introduction to Econometrics 2: Recitation 5

Seung-hun Lee

Columbia University

February 19th, 2020

Wald Test Statistics

- Assume that you have found any kind of a GMM estimator
- he GMM estimator $\hat{\beta}_{GMM}$ that you found has a limiting distribution that can be characterized as

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, V_\beta)$$

- Define a fuction $r(\beta) : \mathbb{R}^k \xrightarrow{p} \Theta \in \mathbb{R}^q$ that characterizes the type of limitations we put on our parameter of interest β
- For $\theta = r(\beta)$, the GMM estimator of θ would naturally be $\hat{\theta}_{GMM} = r(\hat{\beta}_{GMM})$
- By delta method, we can characterize the limiting distribution of $\hat{\theta}_{GMM}$ as

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta) \xrightarrow{d} R' N(0, V_\beta) = N(0, \underbrace{R' V_\beta R}_{=V_\theta})$$

where $R = \frac{\partial r(\beta)'}{\partial \beta} \in \mathbb{R}^{k \times q}$

Wald Test Statistics

- The hypothesis test can be set up in the following manner

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

- We can use the following Wald statistic

$$W \equiv n(\hat{\theta} - \theta)' \widehat{V}_\theta^{-1} (\hat{\theta} - \theta)$$

  where $W \xrightarrow{d} \chi^2_q$ under H_0
- If we conduct a test with a significance level α, we need to find a critical value C such that $\alpha = 1 - F(C)$ where F is the CDF for χ^2_l.
- We reject H_0 if $W > C$ and do not reject if otherwise.

Topics on GMM

Restricted GMM (Constrained GMM)

- Consider $r(\beta) = 0$ as a constraint on β
- Finding a restricted GMM (or constrained GMM) is not too different from the unconstrained GMM .
    - The only difference is that we are now solving a constrained minimization problem.
- In math, CGMM estimator is the solution to the following problem

$$\hat{\beta}_{CGMM} = \arg\min_{\beta} J_n(\beta) \text{ s.t. } r(\beta) = 0$$

- Example: Consider a linear constraint on β coefficients in the form of $R'\beta = c$. We can write this out in a Lagrangian form.

$$n\bar{g}(\beta)' W \bar{g}(\beta) + \lambda'[R'\beta - c]$$

where $\lambda \in \mathbb{R}^q$ is the vector of Lagrange multipiers.

Restricted GMM (Constrained GMM)

- Suppose that the moment condition given is $E(z_i e_i) = 0$. Then

$$\bar{g}(\beta) = \left(\frac{Z'y}{n} - \frac{Z'X\beta}{n} \right)$$

- To find the optimal β, we differentiate the objective function with respect to β. This will become

$$\frac{y'ZWz'Y}{n} - \frac{y'ZWZX\beta}{n} - \frac{\beta'X'ZWZ'y}{n} + \frac{\beta'X'ZWZ'X\beta}{n} + \lambda'R'\beta - \lambda'C$$
$$\xrightarrow{\partial\beta} -2\frac{X'ZWZ'y}{n} + 2\frac{X'ZWZ'X\beta}{n} + R\lambda = 0$$

- Unconstrained GMM can be written as $\left(\frac{X'ZWZ'X}{n}\right)^{-1} \frac{X'ZWZ'y}{n}$.
- Using this, we pre-multiply $-R'\left(\frac{X'ZWZ'X}{n}\right)^{-1}$

Topics on GMM

Restricted GMM (Constrained GMM)

- Thus the FOC becomes

$$2R'\hat{\beta}_{GMM} - 2R'\beta - R'\left(\frac{X'ZWZ'X}{n}\right)^{-1}R\lambda = 0$$

- We can find what λ is

$$\lambda = 2\left(R'\left(\frac{X'ZWZ'X}{n}\right)^{-1}R\right)^{-1}R'(\hat{\beta}_{GMM} - \beta)$$

- Then put this back into the FOC

$$-2\frac{X'ZWZ'y}{n} + 2\frac{X'ZWZ'X\beta}{n} + 2R\left(R'\left(\frac{X'ZWZ'X}{n}\right)^{-1}R\right)^{-1}R'(\hat{\beta}_{GMM} - \beta) = 0$$

- We can further rewrite the above as

$$(X'ZWZ'X)\beta - R(R'(X'ZWZ'X)^{-1}R)^{-1}\underbrace{R'\beta}_{=c} = X'ZWZ'y - R(R'(X'ZWZ'X)^{-1}R)^{-1}$$

Restricted GMM (Constrained GMM)

- Or we can write

$$
\begin{aligned}
(X'ZWZ'X)\beta &= X'ZWZ'y - R(R'(X'ZWZ'X)^{-1}R)^{-1}(R'\hat{\beta}_{GMM} - c) \\
\implies \hat{\beta}_{CGMM} &= \hat{\beta}_{GMM} - (X'ZWZ'X)^{-1}R(R'(X'ZWZ'X)^{-1}R)^{-1}(R'\hat{\beta}_{GMM} - c)
\end{aligned}
$$

- Note that if we premultiply R' to both sides, we can get

$$
\begin{aligned}
R'\hat{\beta}_{CGMM} &= R'\hat{\beta}_{GMM} - R'(X'ZWZ'X)^{-1}R(R'(X'ZWZ'X)^{-1}R)^{-1}(R'\hat{\beta}_{GMM} - c) \\
&= R'\hat{\beta}_{GMM} - R'\hat{\beta}_{GMM} + c = c
\end{aligned}
$$

  which shows that the CGMM satisfies the constraint.

Topics on GMM

Distance Test

- If we are working with a possibly nonlinear form of $r(\beta)$, we can use an alternative criterion-based statistic.
- The basic idea of distance test is to compare unrestricted and restricted estimators by contrasting the criterion functions
- Define
$$J(\beta) = n\bar{g}_n(\beta)'\widehat{\Omega}^{-1}\bar{g}_n(\beta)$$
where $\bar{g}_n(\beta)$ is $\frac{1}{n}\sum_{i=1}^{n} g(w_i, \beta)$ and $\widehat{\Omega}$ is the efficient weight matrix
- With the unconstrained estimator and the constrained estimator, we can write
$$J(\hat{\beta}_{GMM}) = n\bar{g}_n(\hat{\beta}_{GMM})'\widehat{\Omega}^{-1}\bar{g}_n(\hat{\beta}_{GMM})$$
$$J(\hat{\beta}_{CGMM}) = n\bar{g}_n(\hat{\beta}_{CGMM})'\widehat{\Omega}^{-1}\bar{g}_n(\hat{\beta}_{CGMM})$$
- The distance statistic D, is defined as
$$D \equiv J(\hat{\beta}_{CGMM}) - J(\hat{\beta}_{GMM}) \geq 0$$

Distance Test

- As we did for hypothesis testing on Wald test statistic, we can use D for hypothesis tests of the following setup

$$H_0 : r(\beta) = \theta, \quad H_1 : r(\beta) \neq \theta$$

- Under H_0, D converges in distribution to χ^2_q.
- We can find a critical value c s.t. $\alpha = 1 - F(c)$, where F is the CDF for χ^2_q and reject H_0 if $D > c$.
- The idea is that if H_0 is true, then imposing the restriction does not alter the moment equations greatly, making it a sensible restriction. Otherwise, the moment equation changes greatly, making it unreasonable constraint.
- In fact, we can show that when $r(\beta)$ is linear, D becomes identical to the Wald Test Statistic W (after a long algebra)

Topics on GMM

Overidentification Test

- In this section, we generalize the overidentification test we learned in 2SLS setup to a GMM setting.
- We can allow for heteroskedasticity (Sargan's test relied on homoskedasticity)
- If $\dim(g_i) = l > k = \dim(\beta)$, it is possible that there exists no β s.t. $E[g(w_i, \beta)] = 0$ is satisfied. Therefore, the overidentifying restrictions become testable.
- Effectively, we are testing the hypothesis

$$H_0 : E[g(w_i, \beta)] = 0 \text{ vs. } H_1 : E[g(w_i, \beta)] \neq 0$$

- Let β_0 be the true value for the parameter of interest. Then $\bar{g}_n(\beta_0) = \frac{Z'y - Z'X\beta_0}{n} = \frac{Z'e}{n}$ has a limiting distribution characterized by

$$\sqrt{n}\bar{g}_n(\beta_0) \xrightarrow{d} N(0, \Omega), \quad \Omega \in \mathbb{R}^{l \times l}$$

  and thus $J(\beta_0) \xrightarrow{d} \chi^2_l$.

Overidentification Test

- Use a GMM estimator to build our test statistic $J = J(\hat{\beta}_{GMM})$, which has a limiting distribution χ^2_{l-k}
- Like before, we can find c s.t. $\alpha = 1 - F(c)$, and then reject H_0 when $J > c$.
- It should be noted that when the H_0 is rejected, we only know that some moment condition is violated. We cannot pick out which. Nevertheless, rejection of the H_0 is a bad sign

Topics on GMM

Overidentification Test

- We can apply overidentification test on a (strict) subset of instruments whose validity is uncertain.
- To do this, we can 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

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

- Test statistic
  - Estimate the model by the efficient GMM with only the z_{ai} set of instruments and obtain the GMM criterion. This will be denoted as J_a
  - Then, estimate the model with the full set of instruments and obtain a separate GMM criterion, denoted as $J_{a,b}$
  - Create a test statistic

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

  - Then, we find a critical value c for a significance level α and reject the null hypothesis if $C > c$.

Overidentification Test (Endogeneity)

- One example of a subset overidentification test is an endogeneity test.
- Assume a following data generating process

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

  where x_{1i} is exogenous but x_{2i} may not be

- Then, we want to test

$$H_0 : E(x_{2i}e_i) = 0, \text{ vs. } H_1 : E(x_{2i}e_i) \neq 0$$

- Let $z_i = (x_{1i} \quad z_{2i})' \in \mathbb{R}^l$ and assume that $E(z_i e_i) = 0$'

Overidentification Test (Endogeneity)

- Create a test statistic
  - Estimate the efficient GMM by using (x_{1i}, z_{2i}) to instrument (x_{1i}, x_{2i}). Then obtain the GMM criterion $\tilde{J}$.
  - Work with a larger set by using (x_{1i}, x_{2i}, z_{2i}) to instrument (x_{1i}, x_{2i}). After this, we can get the GMM criterion $\widehat{J}$.
  - The relevant test statistic is

$$C = \widehat{J} - \tilde{J} \xrightarrow{d} \chi^2_{k_2}$$

  - we find a relevant critical value c for a significance level α and reject H_0 of exogeneity for x_2 if $C > c$
- We know that x_{1i} and z_{2i} are exogenous. If x_{2i} is also exogenous, then the GMM criterion obtained by using all three should not be too different from GMM criterion from variables x_{1i} and z_{2i}

Conditional Moment Conditions

- Assume that the data generating process and conditional moment condition of interest is
$$y_i = m(x_i, \beta^0) + e_i \quad E(e_i(\beta)|z_i) = 0$$
- One way to address is to use the idea that $E[e_i|z_i] = 0 \implies E[h(z_i)e_i] = 0$ to construct some number of instruments and solve a GMM with unconditional moments
- The other is to construct an optimal instrument
$$R(z_i) = E\left[\frac{\partial e_i(\beta^0)}{\partial \beta} \mid z_i\right] \in \mathbb{R}^k, \sigma^2(z_i) = E[e_i(\beta^0)^2|z_i]$$
Then, the optimal instrument is defined by
$$A_i = -\frac{R(z_i)}{\sigma^2(z_i)} \in \mathbb{R}^k$$
yielding the optimal moment
$$g_i^*(\beta) = A_i e_i(\beta)$$

Panel Data

Motivation

- Let y and $x = (x_1 1, x_2, ..., x_k)$ be the observable factors.
- Denote α as an unobservable random variable that is incorporated into the data generating process additively.
- Then, we can write $E[y|x, \alpha]$ as

$$E[y|x, \alpha] = x'\beta + \alpha$$

  where β is the coefficient of interest
- If α is independent from x, the it is not different from the idiosyncratic error
- If otherwise, then we cannot find a consistent estimator for β
- If α is fixed across time for an individual and we have access to panel data, we can address this issue

Panel Data

Framework

- The data now has two dimensions - dimensions across different unit of observation i and across time t. In maths,

$$(y_{it}, x_{it}) \text{ where } i = 1, ..., n, \text{ and } t = 1, ..., T$$

- To be more concrete, we can write the data generating process as

$$y_{it} = x_{it}'\beta + \alpha_i + u_{it}$$

  where x_{it} is an observable variable that varies across individuals and time periods. α_i is the **individual (fixed) effect** that is unobservable.
- We can define $v_{it} = \alpha_i + u_{it}$ and rewrite the data generating process as

$$y_{it} = x_{it}'\beta + v_{it}$$

Panel Data

A word of caution

- Depending on whether α_i is correlated with x_{it} or not, we can categorize α_i into the following
  - **Random Effects**: There is no correlation between the observables and α_i
  - **Fixed Effects**: The correlation between x_{it} and α_i is nonzero.
- Caveat: *In case you are looking into some old textbooks, the categorization is slightly different. If α_i is considered to be a parameter to be estimated, the old textbooks refers to this as fixed effects. If α_i is a random variable, it was called a random effect. Note that in the above discussion, α_i is random variable in both fixed and random effects.*

Estimation

- The following assumption is required to show whether the panel estimates are consistent or not

Strict Exogeneity

We say that the regressor x_{it} is **strictly exogenous** with respect to u_{it} if

$$E[y_{it}|x_{i1},..,x_{iT},\alpha_i] = E[y_{it}|x_{it},\alpha_i]$$

which boils down to

$$E[u_{it}|x_{i1},..,x_{iT},\alpha_i] = E[u_{it}|x_{it},\alpha_i] = 0$$

- The above condition also implies that

$$E[x_{is}u_{it}] = 0 \ (s,t) = 1,..,T$$

Panel Data

Pooled OLS

- Given that our model is

$$y_{it} = x_{it}'\beta + \underbrace{\alpha_i + u_{it}}_{=v_{it}}$$

  The estimator can be written as

$$\hat{\beta}_{POLS} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} x_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} y_{it}$$

  or

$$\hat{\beta}_{POLS} - \beta = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} x_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} v_{it}$$

- Since I can write $E[x_{it}(\alpha_i + u_{it})]$ instead, it can be seen that unless $cov(x_{it}, \alpha_i) = 0$, the $E[x_{it}\alpha_i]$ term will remain. Thus, POLS estimator is not consistent.

First Difference

- To obtain the first difference estimator, we need to subtract the original data generating process by one lag of y_{it}. Or

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta u_{it} \quad (i = 1, ..., n \text{ and } t = 2, .., T)$$

- Notice that since α_i is same across $t = 1, .., T$ (but different for each i), it vanishes.
- By taking an OLS, we can obtain

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta y_{it}$$

Panel Data

First Difference

- We can show that this is a consistent estimator. Write

$$\hat{\beta}_{FD} - \beta = \left(\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta u_{it}$$

- We need to show that $E[\Delta x_{it} \Delta u_{it}] = 0$, which can be written

$$\begin{aligned} E[\Delta x_{it} \Delta u_{it}] &= E[(x_{it} - x_{i,t-1})(u_{it} - u_{i,t-1})] \\ &= E[x_{it} u_{it}] - E[x_{it} u_{i,t-1}] - E[x_{i,t-1} u_{it}] + E[x_{i,t-1} u_{i,t-1}] \\ &= 0 - 0 - 0 + 0 = 0 \end{aligned}$$

  Therefore, $\hat{\beta}_{FD}$ is consistent.

- Another requirement for this to be defined is that $\left(\sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}' \right)$ should be a full column matrix so that the inverse matrix is defined. This would effectively rule out time-constant regressors.

Panel Data

Within Estimator

- Write

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

- By average over time across all variables, we can get a cross-sectional equation

$$\bar{y}_i = \bar{x}_i'\beta + \alpha_i + \bar{u}_i$$

- So subtract the cross-sectional equation from the original data generating process to get

$$\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{u}_{it}, \quad (i = 1, .., n, \text{ and } t = 1, .., T)$$

- The within estimator is obtained by taking an OLS to above equation

$$\hat{\beta}_{WE} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{y}_{it}$$

Panel Data

Within Estimator

- To show consistency, rewrite the above as

$$\hat{\beta}_{WE} - \beta = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{u}_{it}$$

- $E[\tilde{x}_{it} \tilde{u}_{it}]$ can be written as

$$E[\tilde{x}_{it} \tilde{u}_{it}] = E[x_{it} u_{it}] - E[x_{it} \bar{u}_i] - E[\bar{x}_i u_{it}] + E[\bar{x}_i \bar{u}_i]$$

- Since $\bar{x}_i, \bar{u}_i$ incorporates regressors and errors from all time periods, applying strict exogeneity (and strict exogeneity only) reduces the above equation to 0
- In addition, we need that $E\left[\tilde{x}_{it} \tilde{x}_{it}'\right]$ be a full column rank for its inverse to be defined. Time-constant regressors are ruled out.

Least Square Dummy Variables

- Let Dk_i be the dummy variable that equals 1 if $i = k$ and 0 otherwise. The idea is to put a total of $N - 1$ of such dummy variables into the regression
- Therefore, we work with

$$y_{it} = x_{it}'\beta + D1_i\alpha_1 + ... + D(n-1)_i\alpha_{n-1} + u_{it}$$

- Each individual has his/her own constant term.
    - For the n'th individual, the constant term is represented by the β_0, the coefficient on the column vector of x_{it}. For $k(\neq n)$'th individual, the intercept term is $\beta_0 + \alpha_k$.

Panel Data

Some Interesting Topics:**WE vs FD**

- When $T = 2$, it can be shown that they are numerically equal.
- When $T \geq 3$, they are no longer equal
  - If u_{it} is free from serial correlation (or IID), then taking a first difference would introduce serial correlation. This is because

    $$\begin{aligned} cov(\Delta u_{it}, \Delta u_{i,t-1}) &= E[u_{it}u_{it-1}] - E[u_{it}u_{it-2}] - E[u_{it-1}u_{it-1}] + E[u_{it-1}u_{it-2}] \\ &= 0 - 0 - var(u_{it-1}) + 0 \neq 0 \end{aligned}$$

    As such, first difference in this situation suffers from inconsistency problem
  - There may be a case when Δu_{it} is serially uncorrelated. For instance, u_{it} could be a random walk process in the sense that

    $$u_{it} = u_{it-1} + \eta_{it} \quad (E[\eta_{it}] = 0, E[\eta_{it}\eta_{is}] = 0 (s \neq t), var(u_{it}) = \sigma^2)$$

    If we use a first difference estimator here, we get to obtain the most efficient estimator.

Some Interesting Topics:**WE=LSDV**

- With the knowledge of Kronecker product and bunch of algebra, we can show that these are numerically equal
- I have the details in the recitation note. Derivation took 90 minutes.