Econ 2120: Section 7

Part 2: System Estimation with Orthogonality Conditions aka Linear GMM

Ashesh Rambachan

Fall 2018

Outline

Outline

Motivation

We have a linear model

$$\underset{n\times 1}{Y_i} = \underset{n\times K}{R_i}\underset{K\times 1}{\gamma} + V_i.$$

Without saying anything about V_i, this is completely vacuous.

For any γ, define $V_i = Y_i - R_i\gamma$ and it holds.

One way to give the model meaning is specify the **moment conditions** or **orthogonality conditions**.

Motivation

Lecture Note 5: Suppose that V_i is orthogonal to R_i. Then, γ is the coefficient from the best linear predictor.

What's different now? What if there is omitted variables bias? Our model is now

$$Y_i = R_i\gamma + Q_i\delta + U_i,$$

where V_i is orthogonal to R_i, Q_i. Q_i is unobserved. Can we do anything?

Yes! We can do something. We need to find some other variable B_i that is orthogonal to Q_i, U_i but is correlated with R_i – an "instrument."

The instrument B_i will give us moment conditions that we can use to identify γ.

Want more detail?

Hayashi, Chapter 3-4 covers everything you could want to know about Linear GMM.

Outline

Set-Up

Consider the model

$$
\begin{aligned}
Y_i &= R_i\gamma + V_i, \\
E[B_i V_i] &= 0.
\end{aligned}
$$

We observe (Y_i, R_i, B_i) for $i = 1, \ldots, n$ and assume that there is random sampling (Y_i, R_i, B_i) are i.i.d.

Dimensions:

$$
\begin{aligned}
Y_i &\text{ is } M \times 1, \\
R_i &\text{ is } M \times K, \\
V_i &\text{ is } M \times 1, \\
\gamma &\text{ is } K \times 1, \\
B_i &\text{ is } L \times M.
\end{aligned}
$$

Set-Up

We use the orthogonality conditions

$$E[B_i V_i] = 0$$

to estimate γ. Note that $B_i V_i$ is $L \times 1$ and so, we have L orthogonality conditions to estimate K parameters.

We need to have that $L \geq K$ for this to work.

Example: Best linear predictor

$M = 1$, $R_i = X_i'$, $\gamma = \beta$, $B_i = X_i$. Then,

$$Y_i = X_i'\beta + V_i, \quad E[X_i V_i] = 0.$$

This is the same as writting

$$E^*[Y_i | X_i] = X_i'\beta.$$

Example: Generalized linear predictor

Let $\Phi = I$, $\gamma = \beta$, $B_i = R_i$. Then,

$$Y_i = R_i\beta + V_i, \quad E[R_i' V_i] = 0.$$

This is the same as writing

$$E_I^* = R_i'\beta.$$

Outline

Moment Function

The key to this will be the orthogonality condition:

$$E[B_i V_i] = 0.$$

We define an associated **moment function**

$$\psi(W_i, a) = B_i(Y_i - R_i a),$$

where $W_i = (Y_i, R_i, B_i)$ is the data for the i-th observation and a is $K \times 1$. So, as a function of a,

$$\psi : \mathbb{R}^K \to \mathbb{R}^L.$$

The moment function satisfies: At $a = \gamma$,

$$E[\psi(W_i, \gamma)] = 0.$$

We typically assume that γ is unique value such that this holds.

Sample Moment Function

From the moment function, we define the **sample function moment**

$$\hat{\psi}_n(a) = \frac{1}{n} \sum_{i=1}^{n} \psi(W_i, a).$$

Our proposed estimator will be the one that gets the sample moment function "close" to zero. That is, our estimator $\hat{\gamma}$ is given by

$$\hat{\psi}_n(\hat{\gamma}) \approx 0.$$

Sample Moment Function

In the case of a linear moment function $\psi(W_i, a) = B_i(Y_i - R_i a)$, the sample moment function is

$$\begin{aligned}\hat{\psi}_n(a) &= \frac{1}{n}\sum_{i=1}^{n} B_i(Y_i - R_i a) \\ &= \frac{1}{n}\sum_{i=1}^{n} B_i Y_i - \left(\frac{1}{n} B_i R_i\right) a \\ &= S_{BY} - S_{BR} a,\end{aligned}$$

where

$$S_{BY} = \frac{1}{n}\sum_{i=1}^{n} B_i Y_i, \quad S_{BR} = \frac{1}{n}\sum_{i=1} B_i R_i.$$

We want to choose our estimator $\hat{\gamma}$ so that

$$S_{BY} - S_{BR}\hat{\gamma} \approx 0.$$

This is L equations with K unknowns so we need to be careful.

Outline

Just vs. Over-identified

If $L = K$, we are **just-identified** and S_{BR} is a square matrix. If it is invertible (which we'll assume), then

$$\hat{\gamma} = S_{BR}^{-1} S_{BY}.$$

If $L > K$, we are **over-identified**. We introduce a weight matrix, $\hat{D}$ that is $K \times L$ and satisfies

$$\hat{D} S_{BR} \textit{ invertible}$$
$$\hat{D} \xrightarrow{p} D, \text{ with } DE[B_i R_i] \text{ invertible.}$$

We multiply the sample moment equation by $\hat{D}$ and then, consider the estimator $\hat{\gamma}$ with

$$\hat{D}\psi_n(\hat{\gamma}) = 0 \implies \hat{D} S_{BY} - \hat{D} S_{BR}\hat{\gamma} = 0 \implies \hat{\gamma} = (\hat{D} S_{BR})^{-1}(\hat{D} S_{BY}).$$

What is the weight matrix $\hat{D}$?

When you are over-identified, you have more moment conditions than parameters. In general, you may not be able to satisfy them all exactly.

Even if the moment conditions are satisfied in the population, you may not be able to satisfy them *in-sample*.

What if the population does not satisfy the moment conditions? TBD.

The weight matrix $\hat{D}$ is introduced to govern how you trade off between your moment conditions in sample.

What does this mean? We can set-up an associated minimum norm problem

$$\hat{\gamma} = \arg\min_a \hat{\psi}_n(a)' \hat{C} \hat{\psi}_n(a),$$

for some positive-definite, symmetric matrix $\hat{C}$. In the linear case,

$$\hat{\gamma} = \arg\min_a (S_{BY} - S_{BR}a)' \hat{C} (S_{BY} - S_{BR}a).$$

What is the weight matrix $\hat{D}$?

Take the first-order condition and set it equal to zero. You see that the solution is

$$\hat{\gamma} = (S'_{BR}\hat{C}S_{BR})^{-1}S'_{BR}\hat{C}S_{BY}.$$

Matching to our original expression, we see that this is the same as setting

$$\hat{D} = S'_{BR}\hat{C}.$$

We will have a lot to say about how you choose $\hat{C}$ (or $\hat{D}$) later.

Outline

Outline

Consistent Estimation

Suppose that $\hat{D} \xrightarrow{p} D$, a $K \times L$ nonrandom matrix and that $DE[B_i R_i]$ is non-singular.

Claim: As $n \to \infty$, $\hat{\gamma} \to \gamma$.

By LLN, Slutsky

$$\hat{\gamma} \xrightarrow{p} ([DE[B_i R_i])^{-1} DE[B_i Y_i].$$

We have $Y_i = R_i \gamma + V_i$. So, premultiplying by DB_i gives

$$DB_i Y_i = DB_i R_i \gamma + DB_i V_i \implies DE[B_i Y_i] = DE[B_i R_i]\gamma.$$

Substituting this into the probability limit of $\hat{\gamma}$ gives

$$([DE[B_i R_i])^{-1} DE[B_i Y_i] = ([DE[B_i R_i])^{-1} DE[B_i R_i]\gamma = \gamma.$$

Outline

Limit Distribution

Claim: $\sqrt{n}(\hat{\gamma}-\gamma) \xrightarrow{d} N(0, \alpha\Sigma\alpha')$, where

$$\alpha = (DE[B_iR_i])^{-1}, \quad \Sigma = E[B_iV_iV_i'B_i'].$$

We have

$$Y_i = R_i\gamma + V_i$$

Premultiply by B_i and average across i to get

$$\frac{1}{n}\sum_{i=1}^{n} B_iY_i = \frac{1}{n}\sum_{i=1}^{n} B_iR_i\gamma + \frac{1}{n}\sum_{i=1}^{n} B_iV_i \implies S_{BY} = S_{BR}\gamma + S_{BV}$$

Then substitute this into $\hat{\gamma} = (\hat{D}S_{BR})^{-1}\hat{D}S_{BY}$ to get

$$\begin{aligned}\hat{\gamma} &= (\hat{D}S_{BR})^{-1}\hat{D}(S_{BR}\gamma + S_{BV})\\ &= \gamma + (\hat{D}S_{BR})^{-1}\hat{D}S_{BV}\end{aligned}$$

Limit Distribution

Rearrange and get that

$$\sqrt{n}(\hat{\gamma} - \gamma) = (\hat{D} S_{BR})^{-1} \hat{D} (\sqrt{n} S_{BV}).$$

By the CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} B_i V_i \xrightarrow{d} N(0, \Sigma)$$

By LLN and Slutsky,

$$(\hat{D} S_{BR})^{-1} \hat{D} \xrightarrow{p} (DE[B_i R_i])^{-1} D = \alpha.$$

Apply Slutsky to the product and you're done.

Limit Distribution

As usual, we use the limit distribution to construct asymptotically valid confidence intervals. See the notes for details.

Outline

Additional Assumptions

If we additionally assume that

$$E[V_i|B_i] = 0, \quad Cov(V_i|B_i) = E[V_i'V_i'|B_i] = \Omega,$$

then this simplifies our asymptotic covariance matrix. We get that

$$\begin{aligned} \Sigma &= E[B_i V_i V_i' B_i'] \\ &= E[B_i E[V_i V_i'|B_i] B_i'] = E[B_i \Omega B_i']. \end{aligned}$$

We can estimate Ω with

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i \hat{V}_i'$$

and then consistently estimate Σ.

Panel Data Application

Let's apply this to our panel data model. We had

$$Y_{it} = \gamma Z_{it} + A_i + U_{it}, \quad E[U_{it} | Z_{i1}, \ldots, Z_{iT}, A_i] = 0.$$

Applying the within transformation, we get

$$E[Y_{it} - \bar{Y}_i | Z_{i1}, \ldots, Z_{iT}] = \gamma(Z_{it} - \bar{Z}_i)$$

for $t = 1, \ldots, T$. Define

$$\tilde{Y}_i = \begin{pmatrix} Y_{i1} - \bar{Y}_i \\ \vdots \\ Y_{iT} - \bar{Y}_i \end{pmatrix}, R_i = \begin{pmatrix} Z_i 1 - \bar{Z}_i \\ \vdots \\ Z_{iT} - \bar{Z}_i \end{pmatrix}, V_i = \begin{pmatrix} U_{i1} - \bar{U}_i \\ \vdots \\ U_{iT} - \bar{U}_i \end{pmatrix}.$$

Panel Data Application

Setting $B_i = R_i'$, we get that

$$\tilde{Y}_i = R_i\gamma + V_i, \quad E[B_i V_i] = 0,$$

where $E[V_i|B_i] = 0$ because of strict exogeneity and

$$\Omega = E[V_i V_i']$$

is not diagonal because of serial correlation in the errors U_{it}. So, when we estimate

$$\hat{\Omega}_{st} = \frac{1}{n}\sum_{i=1}^{n} \hat{V}_{is}\hat{V}_{it}$$

we are allowing for serial correlation in our errors.

Outline

The model

The structural regression model is

$$E[Y_{it} | Y_{i1}, \ldots, Y_{i,t-1}, A_i] = \gamma Y_{i,t-1} + A_i.$$

A_i is unobserved.

Define the prediction error

$$U_{it} = Y_{it} - E[Y_{it} | Y_i^{(t-1)}, A_i],$$

where

$$Y_i^{(s)} = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{is} \end{pmatrix}.$$

So, we write

$$Y_{it} = \gamma Y_{i,t-1} + A_i + U_{it}, \quad E[U_{it} | Y_i^{(t-1)}, A_i] = 0$$

for $t = 2, \ldots, T$.

First-differences

We can eliminate A_i by first differences. We get that

$$Y_{i,t} - Y_{i,t-1} = \gamma(Y_{i,t-1} - Y_{i,t-2}) + (U_{i,t} - U_{i,t-1})$$

for $T = 3, \ldots, T$.

We can map this into our linear GMM framework. Consider

$$\tilde{Y}_i = \begin{pmatrix} Y_{i3} - Y_{i2} \\ \vdots \\ Y_{iT} - Y_{i,T-1} \end{pmatrix}, R_i = \begin{pmatrix} Y_{i2} - Y_{i1} \\ \vdots \\ Y_{i,T-1} - Y_{i,T-2} \end{pmatrix}, V_i = \begin{pmatrix} U_{i3} - U_{i2} \\ \vdots \\ U_{i,T} - U_{i,T-1} \end{pmatrix}.$$

So, we can write

$$\tilde{Y}_i = R_i\gamma + V_i.$$

How do we estimate γ? Clearly the elements of R_i are correlated with V_i.

Constructing an instrument

Let

$$B_i = \begin{pmatrix} Y_i^{(1)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Y_i^{(T-2)}. \end{pmatrix}$$

Then, note that

$$B_i V_i = \begin{pmatrix} Y_i^{(1)}(U_{i3} - U_{i2}) \\ \vdots \\ Y_i^{(T-2)}(U_{iT} - U_{i,T-1}) \end{pmatrix}$$

and so, because $E[U_{it}|Y_i^{(t-1)}, A_i] = 0$,

$$E[B_i V_i] = 0$$

B_i provides us with a set of valid instruments.