ECOM40006/90013 Econometrics 3

Week 10 Extras

Question 1: The Method of Moments

The Weak Law of Large Numbers allows us to connect 'sample moments' with 'population moments'. However, the concept of a moment can be quite confusing to some. After all, what is "moment matching" and why would we ever need something like that?

(a) Expected values are an example of what we call 'population' moments. These are theoretical expressions that we would see in a world where we had all the population data. However, in reality we only have a finite amount of data to work with. So why not modify our theory to take this into account? For the following expressions below, state whether or not it is a population or a sample moment.

(i.) $\mathbb{E}(X)$

(ii.) $\displaystyle\frac{1}{n}\sum_{i=1}^{n} x_i x_i'$

(iii.) $\mathbb{E}(x_i x_i')$

(iv.) $\mathbb{E}(x_i e_i)$

(v.) $\mathbb{E}(x_i x_i')^{-1}\mathbb{E}(x_i y_i)$

(vi.) $\displaystyle\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} x_i y_i$

(vii.) $\mathrm{Var}(X)$

For any sample moments you identify above, what are their population counterparts?

(b) Suppose that we had a linear model $y = X\beta + u$. Suppose that we want an estimator for β under the assumption that $\mathbb{E}(X'u) = 0$. Use this assumption to derive an estimator for β in terms of population moments.

(c) For the estimator in part (b), find its sample counterpart. Does the sample counterpart have a name? If so, what is it called?

Question 2: 2SLS and the Generalized Method of Moments

Consider the linear regression model

$$y = X\beta + u, \qquad \mathbb{E}(uu') = \Omega,$$

where as usual y and u are both $n \times 1$ column vectors and X is a $n \times k$ matrix of regressors. Suppose that you, as the econometrician, are concerned about potential endogeneity in your regressors: that is, for some subset of your regressors X, $\mathbb{E}(u|X) \neq 0$.

(a) Dimensions. The matrix X can be divided into two partitions:

$$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix},$$

where X_1 consists of k_1 *exogenous* regressors and X_2 consists of k_2 *endogenous* regressors. Furthermore, we have available a set of p instrumental variables Z (i.e. it is $n \times p$) with $p \geq k_2$.

(i.) Explain in instrumental variables (IV) terminology what $p \geq k_2$ represents.

(ii.) Determine the dimensions of the matrix of instrumental variables

$$W = \begin{bmatrix} X_1 & Z \end{bmatrix},$$

and discuss the differences between X_1 and W.

(b) Background. Explain intuitively what these assumptions represent, and whether you can re-express them in aggregate notation:

(i.) $\mathbb{E}(u_j|W_j) = 0$, $j = 1, 2, \ldots, n$, where W_j' is the j^{th} row of W.

(ii.) $\mathbb{E}(u_i u_j|W_i, W_j) = \omega_{ij}$, where ω_{ij} denotes the (i, j)-th element of Ω.

(c) Moment conditions. Consider the vector $W'u$. Derive expressions for $\mathbb{E}(W'u)$ and $\text{Var}(W'u)$. Show that in particular the expression $\mathbb{E}(W'u)$ represents a set of $k_1 + p$ equations, or *moment conditions*.

(d) Estimation. Suppose in this case that $p = k_2$. A *Method of Moments* estimator based on $\mathbb{E}(W'u) = 0$ is

$$\hat{\beta}_{IV} = (W'X)^{-1}W'y,$$

which is the IV estimator from the week 9 extras, but with different notation. Assume the following:

$$\frac{1}{n}W'X \xrightarrow{p} Q_{W'X} \text{ nonsingular, } \quad \text{and} \quad \frac{1}{n}W'\Omega W \xrightarrow{p} Q_{W'\Omega W} \text{ positive definite.}$$

Further assume that

$$\frac{1}{\sqrt{n}}W'u \xrightarrow{d} N(0, Q_{W'\Omega W}).$$

(i.) Show that $\hat{\beta}_{IV}$ is consistent for β.

(ii.) Derive an expression for the asymptotic distribution of $\hat{\beta}_{IV}$.

(e) Now, suppose that $p > k_2$. Is it still possible to use the IV estimator in part (d)? Explain your answer.

(f) It may seem somewhat absurd that an estimator would only allow us to use a limited number of instrumental variables. However, a method of implementing instrumental variables is via the 2SLS method, which circumvents this problem. Describe the 2SLS estimation process without using any equations (inline math is fine), and give intuition for your stages wherever you can.

(g) Now repeat part (f) above, but this time, give formal expressions of what we require from the first and second stages. Using these equations, give an explicit formula for the 2SLS estimator in terms of the data matrices X, W and y. You may potentially find it convenient to use the projection matrix shorthand $P_W = W(W'W)^{-1}W'$.

(h) To show consistency of the estimator in part (g), we will need to make a few assumptions:

$$\frac{1}{n}X'W \xrightarrow{p} Q_{X'W}, \qquad \frac{1}{n}W'X \xrightarrow{p} Q_{W'X} = Q'_{X'W}, \qquad \frac{1}{n}W'W \xrightarrow{p} Q_{W'W} \text{ full rank.}$$

Furthermore, $\frac{1}{n}W'u \xrightarrow{p} 0$. Show that the 2SLS estimator is consistent for β.

(i) On top of the assumptions in part (g), further assume (as before) that

$$\frac{1}{\sqrt{n}}W'u \xrightarrow{d} N(0, \sigma^2 Q_{W'\Omega W}).$$

Use this to derive the asymptotic distribution of the 2SLS estimator.

(j) The 2SLS estimator falls into a broader class of estimators and is technically a type of estimator called the *Generalized Method of Moments* (GMM) estimator.

When there are more equations than there are unknown variables in a system of linear equations, we can no longer solve the equation $W'u = 0$ so the IV estimator is not defined. For regression purposes, this is most definitely not acceptable. An alternative way is to use the *Generalized Method of Moments.*

One way to do it is this. Instead of solving a system explicitly for β, what we can do instead is construct a *quadratic form* and minimize that for β. The estimator we get from this is called the Generalized Method of Moments (GMM) estimator. Assume[1] $\mathbb{E}(W'u) = 0$ and define the sample counterpart to be

$$\hat{m} = \frac{1}{n}W'u = \frac{1}{n}W'(y - X\beta).$$

Suppose we consider the GMM minimization problem

$$\hat{\beta} = \arg\min_{\beta} \hat{m}'\hat{m} = \arg\min_{\beta} u'WW'u.$$

This expression is known as a GMM *criterion function.* Solve this system for $\hat{\beta}$ in terms of X, Z and y. You'll need the following rules of vector calculus:

$$\frac{\partial a\beta}{\partial \beta} = a', \qquad \frac{\partial \beta' A\beta}{\partial \beta} = 2A\beta, \text{ A symmetric.}$$

[1]Relevance is implicitly assumed.

(k) More generally, a GMM minimization problem will use a *weighting matrix* to scale the data accordingly. A common matrix to use is $\Lambda = (Z'Z)^{-1}$. Use this to solve the minimization problem

$$\hat{\beta}_{GMM} = \arg\min_{\beta} \hat{m}'\Lambda\hat{m}.$$

Hint: You may also want to use projection matrix shorthand: $P_W = W(W'W)^{-1}W'$.

Question 3 (bonus): The J-test of overidentifying restrictions

Those of you who have come from Econometrics 2 will recall that when we have more IVs than endogenous regressors, one way that we could get some statistical evidence on the exogeneity of the IVs used was to employ a test of *overidentifying restrictions.* Often known as the J-test, this test statistic is known to have an asymptotic distribution

$$J \overset{a}{\sim} \chi^2_{p-k_2} \quad \text{under } H_0,$$

where p is the number of IVs and k_2 the number of endogenous variables. We're going to derive this explicitly, using the GMM criterion function, which we'll denote as $f(\beta)$ for this exercise. Recall that for the linear model $y = X\beta + u$, the GMM objective function with weighting matrix $\Lambda = (W'W)^{-1}$ is

$$f(\beta) = u'W(W'W)^{-1}W'u = (y - X\beta)W(W'W)^{-1}W'(y - X\beta).$$

For reference, W is a $n \times m$ matrix of instruments with full rank m. X is a $n \times k$ matrix of data with full rank k. Here, $m > k$. The disturbance term u is i.i.d. such that $u \sim N(0, \sigma^2 I_n)$.

(a) Show that $W'W$ is positive definite.

(b) A positive definite matrix C has a decomposition $C^{-1/2}$ such that

$$C^{-1/2}C^{-1/2} = C^{-1}.$$

We will apply this to $(W'W)^{-1}$ to form the vector

$$z = (W'W)^{-1/2}W'u.$$

As a simplifying assumption, assume that W is non-stochastic. Show that $\dfrac{z}{\sigma} \xrightarrow{d} N(0, I_m)$.

(c) Let $\hat{\beta}_{GMM}$ be the GMM estimator

$$\hat{\beta}_{GMM} = (X'P_W X)^{-1}X'P_W y,$$

and also let $A = (W'W)^{-1/2}W'X$. Show that the GMM objective function evaluated at $\hat{\beta}_{GMM}$ can be written as

$$f(\hat{\beta}_{GMM}) = z'M_A z,$$

where $M_A = I_m - A(A'A)^{-1}A'$ is the residual maker corresponding to A.

(d) What is the rank of M_A? (Hint: the rank of an idempotent matrix is equal to its trace...)

(e) Finally, show that the GMM objective function weighted by $\hat{\sigma}^2$ follows a chi-squared distribution

$$\frac{f(\hat{\beta}_{GMM})}{\hat{\sigma}^2} \xrightarrow{d} \chi^2_{p-k_2},$$

where p is the number of IVs and k_2 the number of endogenous variables.[2]

[2] Behind the scenes, if we define k_1 to be the number of exogenous variables, then it follows that $m = p + k_1$ and $k = k_1 + k_2$. In other words, what you need to show is that the degrees of freedom of the chi-squared distribution is in fact $m - k$, from which $p - k_2$ follows.