Econ 534: Section Note 04

JJ Chen[*]

September 19, 2014

Bugs might exist. In this week's review session, we will talk about:

- Sampling
- Sample Analog
- Stata Demo: Monte Carlo simulation

  - Programming basics: macros and loop
  - Write a simple program
  - Simulation and the unbiased results

1   Review of the Week

This passing week we extended the original GM model to the modified GM model. Given the modified assumptions:

$$\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

$$\mathrm{rk}(\boldsymbol{X}) = k, \tag{2}$$

$$\mathrm{E}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{0}, \tag{3}$$

$$\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}, \tag{4}$$

$$\boldsymbol{X} \text{ stochastic}, \tag{5}$$

[*]jchen215@uic.edu

we have the following finite sample properties of the OLS estimator:

$$\mathrm{E}(\hat{\beta} \mid \boldsymbol{X}) = \boldsymbol{\beta}, \ \mathrm{E}(\hat{\beta}) = \boldsymbol{\beta}, \tag{6}$$

$$\mathrm{V}(\hat{\beta} \mid \boldsymbol{X}) = \sigma^2 (\boldsymbol{X'X})^{-1}, \ \mathrm{V}(\hat{\beta}) = \sigma^2 \, \mathrm{E}[(\boldsymbol{X'X})^{-1}], \tag{7}$$

$$\mathrm{V}(\hat{\beta} \mid \boldsymbol{X}) \leq \mathrm{V}(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}), \ \mathrm{V}(\hat{\beta}) \leq \mathrm{V}(\tilde{\boldsymbol{\beta}}), \tag{8}$$

where $\tilde{\boldsymbol{\beta}}$ is other linear unbiased estimator.

Note that there is another parameter σ^2 in the GM model, and its estimator $\hat{\sigma}^2 = \boldsymbol{e'e}/(n-k)$ has the following properties:

$$\mathrm{E}(\hat{\sigma}^2) = \sigma^2, \tag{9}$$

$$\mathrm{E}(\hat{\mathrm{V}}(\hat{\beta})) = \mathrm{V}(\hat{\beta}). \tag{10}$$

2   Sampling

Combined with a random sample from a multivariate population, the modified GM model can be used to investigate **statistical replationship** of interest. The model contains a population specification, and the sampling scheme supports the model so that we can use both of them to estimate population parameters.

Knowing a little bit about random sample scheme can help us seeing both the strength and weakness of the modified GM model. In this week's review session, we will discuss only the most widely adopted scheme: i.i.d. sampling. Note that this is not the only scheme that can support our model.

To begin with, I will repeat the modified GM assumptions in a slightly different way:

$$\mathrm{E}\left(\boldsymbol{y} \mid \boldsymbol{X}\right) = \boldsymbol{X\beta}, \tag{11}$$

$$\mathrm{V}\left(\boldsymbol{y} \mid \boldsymbol{X}\right) = \sigma^2 \boldsymbol{I}, \tag{12}$$

$$\mathrm{rk}(\boldsymbol{X}) = k, \tag{13}$$

$$\boldsymbol{X} \text{ stochastic}. \tag{14}$$

This set of assumptions is mathematically identical to the assumptions we mentioned in last section. Both sets of assumptions have its merits and there are some subtleties behind them. We will talk about it when we have the language of causal inference.

Now let's turn to the population of interest.

(A note on the notations: In matrix algebra, we reserve bold-italic capital letters for matrices and bold italic lowercase letters for vectors. Elements of matrices and vectors are in italic lowercase. In statistics, we use capital letters for random variables (r.v.) and lowercase letters for their realizations. This creates inconsistency and confusions. I have no good solution to it. Here I will respect matrix algebra and use italic lowercase letters for r.v. A vector of r.v. is in bold-italic lowercase. Realization of a r.v. would be mentioned explicitly.)

2.1 A Multivariate Population

Consider a set of k r.v. $(y, x_2, \cdots, x_k)$ with pdf $f(y, x_2, \cdots, x_k)$ and conditional pdf $g(y \mid x_2, \cdots, x_k)$. The first and second moments of these r.v.'s are: $\mathrm{E}(y)$, $\mathrm{V}(y)$, $\mathrm{Cov}(x_h, x_j)$, and $\mathrm{Cov}(x_h, y)$.

Suppose the **statistical relationship** we are interested in is the CEF of y given the x's:

$$\mathrm{E}\left(y \mid 1, x_2, \cdots, x_k\right) = \mathrm{E}\left(y \mid \boldsymbol{x}\right).$$

Suppose further that we can approximate and parametrize the CEF by using a linear function, so that $\mathrm{E}\left(y \mid \boldsymbol{x}\right) = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k = \boldsymbol{x}'\beta$. We assume the CVF is constant: $\mathrm{V}(y \mid \boldsymbol{x}) = \sigma^2$.

To sum, the population we specified has:

$$
\begin{aligned}
\mathrm{E}\left(y \mid \boldsymbol{x}\right) &= \boldsymbol{x}'\beta \\
\mathrm{V}(y \mid \boldsymbol{x}) &= \sigma^2
\end{aligned}
\tag{15}
$$

2.2 The result of identical distribution

When we sample randomly from the population, every drawing $(y_i, \boldsymbol{x}_i')$ comes from an identical multivariate distribution. Before actually drawing (a realization), an

"intention" of a single draw $(y_i, \boldsymbol{x}_i')$ can be considered as a random vector that has the same pdf and cdf of the population: $f(y, x_2, \cdots, x_k)$ and $g(y \mid x_2, \cdots, x_k)$.

Hence, the result of identical distributed sample is

$$
\begin{aligned}
\mathrm{E}\left(y_i \mid \boldsymbol{x}_i\right) &= \boldsymbol{x}_i'\boldsymbol{\beta} \ (i = 1, \cdots, n), \\
\mathrm{V}\left(y_i \mid \boldsymbol{x}_i\right) &= \sigma^2 \ (i = 1, \cdots, n).
\end{aligned}
\tag{16}
$$

Note that the idea of (16) is different from that of (15).

2.3 The result of independent distribution

When we sample randomly from the population, say we sample observation i and j for simplicity, the pdf of the observation i on y conditional on the observation i and j on their k characteristics has the following result:

$$
\begin{aligned}
g^*(y_i \mid \boldsymbol{x}_i, \boldsymbol{x}_j) &= \frac{f^*(y_i, \boldsymbol{x}_i, \boldsymbol{x}_j)}{h^*(\boldsymbol{x}_i, \boldsymbol{x}_j)} \\
&= \frac{f(y_i, \boldsymbol{x}_i)h(\boldsymbol{x}_j)}{h(\boldsymbol{x}_i)h(\boldsymbol{x}_j)} \\
&= \frac{f(y_i, \boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \\
&= g(y_i \mid \boldsymbol{x}_i) \\
\implies \mathrm{E}\left(y_i \mid \boldsymbol{x}_i, \boldsymbol{x}_j\right) &= \mathrm{E}\left(y_i \mid \boldsymbol{x}_i\right) = \boldsymbol{x}_i'\boldsymbol{\beta} \\
\implies \mathrm{V}\left(y_i \mid \boldsymbol{x}_i, \boldsymbol{x}_j\right) &= \mathrm{V}\left(y_i \mid \boldsymbol{x}_i\right) = \sigma^2
\end{aligned}
$$

The second equality comes from the fact that observation i and j are drew independently.

Now consider the joint pdf of y_i and y_j:

$$
\begin{aligned}
g^{**}(y_i, y_j \mid \boldsymbol{x}_i, \boldsymbol{x}_j) &= \frac{f^{**}(y_i, \boldsymbol{x}_i, y_j, \boldsymbol{x}_j)}{h^*(\boldsymbol{x}_i, \boldsymbol{x}_j)} \\
&= \frac{f(y_i, \boldsymbol{x}_i) f(y_j, \boldsymbol{x}_j)}{h(\boldsymbol{x}_i) h(\boldsymbol{x}_j)} \\
&= \left(\frac{f(y_i, \boldsymbol{x}_i)}{h(\boldsymbol{x}_i)}\right)\left(\frac{f(y_j, \boldsymbol{x}_j)}{h(\boldsymbol{x}_j)}\right) \\
&= g(y_i \mid \boldsymbol{x}_i) g(y_j \mid \boldsymbol{x}_j) \\
\implies \operatorname{Cov}(y_i, y_j \mid x_i, x_j) &= 0
\end{aligned}
$$

The second equality again is the independent result.

Extend the above results from observation i and j to any observations, we have the result of independent distribution:

$$
\begin{aligned}
\mathrm{E}\left(y_i \mid \boldsymbol{X}\right) &= \mathrm{E}\left(y_i \mid \boldsymbol{x}_i\right) = \boldsymbol{x}_i'\boldsymbol{\beta}, \\
\mathrm{V}\left(y_i \mid \boldsymbol{X}\right) &= \mathrm{V}\left(y_i \mid \boldsymbol{x}_i\right) = \sigma^2, \\
\operatorname{Cov}(y_i, y_j \mid \boldsymbol{X}) &= 0, i \neq j
\end{aligned}
\tag{17}
$$

Note that the idea of (17) is different from that of (16).

2.4 The Trivial Final Step

The final step is to stack n lines of (17) and write them compactly in algebra matrix notations:

$$
\begin{aligned}
\mathrm{E}\left(\boldsymbol{y} \mid \boldsymbol{X}\right) &= \boldsymbol{X}\boldsymbol{\beta}, \\
\mathrm{V}\left(\boldsymbol{y} \mid \boldsymbol{X}\right) &= \sigma^2 \boldsymbol{I}.
\end{aligned}
\tag{18}
$$

We reach the key GM assumptions. Hence, i.i.d. sampling supports the modified GM model. Question: Can you think of an important violation of the modified GM assumptions? (For example, think of situations when $\mathrm{E}\left(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{x}_j\right) \neq \mathrm{E}\left(\boldsymbol{y}_i \mid \boldsymbol{x}_i\right)$.)

3   Sample Analog

Now that we see a little bit clearer about the sampling scheme and the modified GM model (hopefully), it would be a great chance to introduce Methods of Moments and reinterpret OLS estimate as a sample analog.

Recall that if we specify the population, $\boldsymbol{\beta}$ can be considered as a population parameter that uses all the information from the population. We have

$$
\begin{aligned}
\boldsymbol{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= [(1/n)(\boldsymbol{X}'\boldsymbol{X})]^{-1}[(1/n)(\boldsymbol{X}'\boldsymbol{y})] \\
&= \mathrm{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\mathrm{E}\left(\boldsymbol{x}_iy_i\right),
\end{aligned}
\tag{19}
$$

where $\mathrm{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}$ and $\mathrm{E}\left(\boldsymbol{x}_iy_i\right)$ is the population moments of the distribution.

When we have a random sample at hand (a realization), we can replace the population moments with sample moments and get the OLS estimate as a sample analog:

$$
\hat{\beta} = (\Sigma_i\boldsymbol{x}_i\boldsymbol{x}_i')^{-1}(\Sigma_i\boldsymbol{x}_iy_i).
\tag{20}
$$

In the bivariate case, we have $\beta_1 = \frac{\mathrm{Cov}(x_i,y_i)}{\mathrm{V}(x_i)}$ and $\hat{\beta}_1 = \frac{\sigma_{xy}}{\sigma_x^2}$, where σ_{xy} and σ_x^2 are sample covariance and sample variance.

4   Monte Carlo Simulation

Usually we only have one sample, and thus one OLS estimate. Simulation can generate as many samples as we want, and thus it is useful to study to behavior of an estimator. Please see the session do-file for simulation in Stata.