Section 3: Asymptotics for Least Squares

Maria Carolina Caetano

February 2, 2007

Contents

1 Asymptotics for Least Squares

As you should know from Powell's first class, this course is about relaxing as many of the assumptions of the OLS regression model as we can. Whenever we intend to relax a certain assumption, we have to ask ourselves two questions:

1. What for was this assumption used in the first place? (This way we know what could potentially change in the model if we didn't make this assumption)

2. How can we remedy that? What can we do in order to still have a tractable model?

Very well, today we would like to relax the "normality hypothesis", which said: $\varepsilon \sim N(0, \sigma^2 I)$, and now we ask the two questions above, as we will do over and over again during the semester, and the answers are:

1. The normality assumption served to derive the distribution of the OLS estimate. As seen last semester,

$$\varepsilon \sim N(0, \sigma^2 I), \quad y = X\beta + \varepsilon \Longrightarrow y \sim N(X\beta, \sigma^2 I)$$

so, as a result,

$$\widehat{\beta} = (X'X)^{-1}X'y \sim N((X'X)^{-1}X'X\beta, Var(\widehat{\beta}/X)) = N(\beta, Var(\widehat{\beta}/X))$$

and we calculated $Var(\widehat{\beta}/X)$ by doing:

$$\begin{aligned}&E((\widehat{\beta}-\beta)(\widehat{\beta}-\beta)/X) = E((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}/X) = \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} = (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} = \\ &= \sigma^2(X'X)^{-1}XX(X'X)^{-1} = \sigma^2(X'X)^{-1}\end{aligned}$$

Hence, could always use:

$$\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$$

which was very useful for hypothesis testing.

2. Without the normality assumption, we can't derive the distribution of $\widehat{\beta}$. The solution for this problem is to use asymptotic theory to derive *asymptotic* distributions of this estimator as we increase the sample size. Pay special attention to the word "asymptotic". This means that we can't really say anything about our estimates, but, if the sample is large enough, we can trust that we must be somehow close to the asymptotic distribution we calculated.

1.1 Particular case: X of dimension 1

We will start with the model:

$$(1) \quad y = x\beta + \varepsilon$$

Where x is just a scalar. We'll go over the whole procedure using this particular case, then in the next section we will talk about the vector case. We assume that a random sample was taken from the true population, and indexed by i. By *random sample* we mean that the drawings are i.i.d.

We would like to derive the asymptotic (sometimes called "limiting") distribution of $\sqrt{n}\left(\widehat{\beta}-\beta\right)$, when β is the true value given in (1). We will need to do some manipulations:

$\sqrt{n}\left(\widehat{\beta}-\beta\right) = \sqrt{n}((X'X)^{-1}X'y-\beta)$, where X is the column vector $(x_1, x_2, ..., x_n)'$.

$$\begin{aligned}&= \sqrt{n}((X'X)^{-1}X'y - (X'X)^{-1}X'X\beta) \\ &= \sqrt{n}((X'X)^{-1}X'(y-X\beta)) \\ &= \sqrt{n}(X'X)^{-1}X'\varepsilon \\ &= \sqrt{n}(\tfrac{X'X}{n})^{-1}\tfrac{X'\varepsilon}{n} \qquad (2)\end{aligned}$$

Now, let's write this in sum notation, instead of in vector notation.

$$(2) = \underbrace{(\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{-1}}_{A}\underbrace{(\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i)}_{B}$$

1.1.1 Asymptotic Distribution of A:

We assumed that the sample was random, so the values x_i are i.i.d. Hence, so are the values of x_i^2.

By the Law of Large numbers,
$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \longrightarrow_p E(x_i^2) \overset{def}{=} D$

Observe that here we are assuming (a somewhat hidden assumption) that $E(x_i^2) < \infty$.

Now, define $g : \mathbb{R}^{++} \longrightarrow \mathbb{R}$; $g(z) = \frac{1}{z}$. Then g is continuous as long as $E(x_i^2) \neq 0$ (another hidden assumption). Therefore we can use the Continuous Mapping Theorem, so
$\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{-1} = g(\frac{1}{n}\sum_{i=1}^{n} x_i^2) \longrightarrow_p g(E(x_i^2)) = (E(x_i^2))^{-1} = D^{-1}$

1.1.2 Asymptotic Distribution of B:

We will use the Central Limit Theorem, so we need to check its conditions:

1. The sequence $x_i\varepsilon_i$ is iid, because the sample is a random draw from the population.

2. $E(x_i\varepsilon_i) = 0$ because we are still using the same hypothesis as before, except for the normality hypothesis, so, in particular, X and ε are independent.

3. $Var(x_i\varepsilon_i) = E(x_i^2{\varepsilon_i}^2) = E(x_i^2)E({\varepsilon_i}^2)$ by independence of X and ε,
$= \sigma^2 E(x_i^2) < \infty$

Observe that here we are assuming (another hidden assumption) that $E(x_i^2) < \infty$ or in other words, that it exists. However, we already assumed it in order to show A, so this condition is automatically satisfied.

We will call $C = \sigma^2 E(x_i^2)$

Hence, we are ready to apply the CLT, and therefore,
$B = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i \longrightarrow_d N(0, C)$

1.1.3 Limiting Distribution of Least Squares

Now we will use Slutsky Theorem:
$A \longrightarrow_p D^{-1}$
$B \longrightarrow_d N(0, C)$
So,
$AB \longrightarrow_d D^{-1}N(0, C) = N(0, D^{-1}CD^{-1})$
Translating into the terms that we know:

$\sqrt{n}\left(\widehat{\beta} - \beta\right) = (\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{-1}(\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i) \longrightarrow_d N(0, E(x_i^2)^{-1}\sigma^2 E(x_i^2)E(x_i^2)^{-1}) = N(0, \sigma^2 E(x_i^2)^{-1})$

And the final result is:

$\sqrt{n}\left(\widehat{\beta} - \beta\right) \longrightarrow_d N(0, \sigma^2 E(x_i^2)^{-1})$

1.1.4 How can we do hypothesis testing now?

To do this the most formal (and correct) way, we should proceed as follows:

We know that $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \longrightarrow_p E(x_i^2)$, so again, by the Continuous Mapping Theorem:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \longrightarrow_p \sqrt{E(x_i^2)}$$

So we pre-multiply $\sqrt{n}\left(\widehat{\beta}-\beta\right)$ by $\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$ and, by Slutsky,

$$\begin{aligned}\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}(\sqrt{n}\left(\widehat{\beta}-\beta\right)) &\longrightarrow_d \sqrt{E(x_i^2)}N(0,\sigma^2 E(x_i^2)^{-1}) \\ &= N(0,\left(\sqrt{E(x_i^2)}\right)^2 \sigma^2 E(x_i^2)^{-1}) = N(0,\sigma^2)\end{aligned}$$

so:

$$\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right) \longrightarrow_d N(0,\sigma^2)$$

Now we have a distribution that we can deal with, and we have two possibilities:

<u>If we know σ^2:</u>

Then we use the statistics $\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right)$, and we reject the null hypothesis that $\widehat{\beta}=\beta$ if

$\left|\frac{\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right)}{\sigma}\right| \geq 1.96$ at level of significance of 5%.

<u>If we don't know σ^2:</u>

We know that the variance estimator $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - x_i\beta)^2$ is consistent, ($s^2 \longrightarrow_p \sigma^2$). So, by CMT, $s^{-1} = (s^2)^{-\frac{1}{2}} \longrightarrow_p (\sigma^2)^{-\frac{1}{2}} = \sigma^{-1}$

Now, we apply Slutsky's theorem AGAIN (you'd better learn it by now), and we have:

$$\frac{\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right)}{s} \longrightarrow_d \sigma^{-1}N(0,\sigma^2) = N(0,(\sigma^{-1})^2\sigma^2) = N(0,1)$$

Now we have a statistic: $\frac{\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right)}{s}$ which is asymptotically distributed with a tabulated distribution, and this is the way we should proceed when doing hypothesis testing[1] , so, for instance, we reject the null hypothesis that $\widehat{\beta}=\beta$, if:

$$\left|\frac{\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right)}{s}\right| \geq 1.96 \text{ at level of significance of 5\%.}$$

---

[1] Meaning that the test statistics isn't based in any unknown parameter but the ones we want to test, and its limiting distribution is well known (and therefore tabulated).

1.1.5 A small remark: consistent variance estimation

The remark is that the "formal" way of setting up the statistic for hypothesis testing is not so distant from the way we are used to. By a certain amount of manipulations, we can arrive to the form of the asymptotic distribution that is most well known. Our result is:

$$\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right) \longrightarrow_d N(0,\sigma^2)$$

Now, suppose we assume that with n observations, we are close enough to the asymptotic distribution, so we treat the statistic as if it was distributed like that in n. We denote:

$$\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right) \overset{A}{\sim} N(0,\sigma^2)$$

And if we treat it as if it were the actual distribution, we can do:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}(\sqrt{n}\left(\widehat{\beta}-\beta\right)) \overset{A}{\sim} N(0,\sigma^2)$$

$$\Longrightarrow \sqrt{n}\left(\widehat{\beta}-\beta\right) \overset{A}{\sim} N(0,\sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{-1}) = N(0,\sigma^2(\frac{X'X}{n})^{-1})$$

which is the result that we usually see.

Now, if σ^2 is unknown (as it always is), we do:

$$\frac{\sqrt{\sum_{i=1}^{n} x_i^2}\left(\widehat{\beta}-\beta\right)}{s} \overset{A}{\sim} N(0,1)$$

$$\Longrightarrow \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}(\sqrt{n}\left(\widehat{\beta}-\beta\right))}{s} \overset{A}{\sim} N(0,1)$$

$$\Longrightarrow \sqrt{n}\left(\widehat{\beta}-\beta\right) \overset{A}{\sim} N(0,s^2(\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{-1})$$

which is the way we usually see it.

1.2 Vector Case

The case where x_i is actually a vector os dimension k is exactly similar to the scalar case. And I mean it: although the manipulations become now matrix manipulations, all the theorems, assumptions and results are still the same. In the notes prof. Powell wrote, you can find everything we just saw but for vectors.

Remember that all his explanation is intended to find the limiting distribution of

$$\underbrace{(\frac{1}{n}\sum_{i=1}^{n} x_i x_i')^{-1}}_{A}\underbrace{(\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i)}_{B}$$

,

where $x_i = (x_{i1}, x_{i2}, ..., x_{ik})'$