ECON 3150/4150, Spring term 2013. Lecture 6

Review of theoretical statistics for econometric modelling (II)

Ragnar Nymoen

University of Oslo

31 January 2013

References to Lecture 3 and 6

Lecture 3 reviewed the statistical theory used in Lecture 4 and 5 (Regression model with determinsitc regressor (RM1))
Lecture 6 extends the statistical theory that is used in the Regression model with stochastic regressor (RM2), that will be the main “working-horse” for the rest of the course

- **HGL:** Probability Primer, Appendix B1-B4, C2-C6
- **BN:** Kap 4

Definition of conditional expectation I

- To save some time and space we concentrate on the continuous random variable case.
- There is seminar exercise about the discrete variable case.
- Using the concepts that we reviewed in Lecture 3: The *conditional probability density function* (pdf) for Y given $X = x$ is

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)} \tag{1}$$

where $f_{XY}(x, y)$ is the *joint pdf* for the two random variables X and Y, and $f_X(x)$ is the *marginal pdf* for X.

Definition of conditional expectation II

Definition (Conditional expectation)

Let Y be the random variable with conditional pdf $f_{Y|X}(y \mid x)$. The conditional expectation of Y is

$$E(Y \mid x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\, dy = \mu_{Y|x}.$$

Conditional expectation function I

- For a given value of $X = x$ the conditional expectation $E(Y \mid x)$ is deterministic, it is a number.
- We can however consider the expectation of Y for the whole value set of X. In this interpretation, $E(Y \mid X)$ is a random variable with $E(Y \mid x)$ as a value for $X = x$.
- This line of reasoning motivates that the *conditional expectation function* $E(Y \mid X)$ is a function of the random variable X:

$$E(Y \mid X) = g_X(X)$$

Conditional expectation function II

- If the *conditional expectation function* is linear, we can write it as
$$E(Y \mid X) = \mu_{Y|X} = \beta_0 + \beta_1 X \tag{2}$$
much like the "systematic" or "explanatory part" of a regression model.
- It is this idea that we will utilize when we formulate the regression model for stochastic regressors in Lecture 7 and onwards.
- In this lecture first show the specification of $E(Y \mid X) = g_X(X)$ when the joint pdf of X and Y is normal.
- This is a special case, but it is a relevant one, since the normal distribution in often assumed in econometric models

Conditional expectation function III

- We next discuss important properties of conditional expectation functions more generally. For example:
  - The law of iterated expectations
  - Linear independence of X and Y when $E(Y \mid X) = \textit{constant}$
- The second part of the lecture reviews some results from asymptotic theory that become relevant for the model with regressors that are random variables.

Bivariate normal distribution

- Let the marginal expectations of X and Y be $E(X) = \mu_X$ and $E(Y) = \mu_Y$.
- The variances are σ_X , σ_Y, and σ_{XY} is the covariance
- Correlation coefficient: $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_Y \sigma_X}$

Using the "standardized" notation $z_Y = (y - \mu_Y)/\sigma_Y$ and $z_X = (x - \mu_X)/\sigma_X$, the bivariate normal *pdf* is can be written as:

$$f_{XY}(y, x) = \frac{1}{\sigma_Y \sigma_X 2\pi \sqrt{(1 - \rho_{XY}^2)}} \times \exp\left[-\frac{1}{2} \frac{(z_Y^2 - 2\rho_{XY} z_Y z_X + z_X^2)}{(1 - \rho_{XY}^2)} \right]$$

From the joint density $f_{XY}(y, x)$ we can obtain the marginal pdf $f_X(x)$ (see Lect 3). In this case it becomes:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]$$

Inserting in (1) and simplifying, $f_{Y|X}(y \mid x)$ can be written as

$$f_{Y|X}(y \mid x) = \frac{1}{\sqrt{2\pi\sigma_Y^2\left(1-\rho_{XY}^2\right)}} \times \exp\left\{-\frac{\left(\frac{y-\mu_Y}{\sigma_Y} - \frac{\sigma_{XY}}{\sigma_Y\sigma_X}\frac{x-\mu_X}{\sigma_X}\right)^2}{2\left(1-\rho_{XY}^2\right)}\right\}$$

$$= \frac{1}{\sqrt{2\pi\left(\sigma_Y^2 - \sigma_Y^2\frac{(\sigma_{XY})^2}{\sigma_Y^2\sigma_X^2}\right)}} \times \exp\left\{-\frac{1}{2}\frac{\left[y-\left(\mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X + \frac{\sigma_{XY}}{\sigma_X^2}x\right)\right]^2}{\sigma_Y^2 - \sigma_Y^2\frac{(\sigma_{XY})^2}{\sigma_Y^2\sigma_X^2}}\right\}.$$

We can write

$$y - \left(\mu_Y - \frac{\sigma_{YX}}{\sigma_X^2}\mu_X + \frac{\sigma_{YX}}{\sigma_X^2}x\right)$$

as

$$y - \mu_{Y|X}$$

where

$$\begin{aligned}\mu_{Y|X} &= \underbrace{\mu_Y - \frac{\sigma_{YX}}{\sigma_X^2}\mu_X}_{\beta_0} + \underbrace{\frac{\sigma_{YX}}{\sigma_X^2}}_{\beta_1}x \\ &= \beta_0 + \beta_1 X \qquad (3)\end{aligned}$$

Finally, define the conditional variance as

$$\sigma_{Y|X}^2 = \sigma_Y^2(1 - \frac{\sigma_{YX}^2}{\sigma_Y^2\sigma_X^2}) \qquad (4)$$

Summing up:

- The conditional pdf for Y is

$$f_{Y|X}(y \mid x) = \frac{1}{\sqrt{2\pi\sigma^2_{Y|X}}} \times \exp\left\{-\frac{1}{2}\frac{\left[y-\mu_{Y|X}\right]^2}{\sigma^2_{Y|X}}\right\} \quad (5)$$

  (5) is a normal pdf with two parameters: **the expectation** $\mu_{Y|X}$ in (3) and the variance $\sigma^2_{Y|X}$ in (4).
- The *conditional expectation function* for Y given X can be written as in (2):

$$\mu_{Y|X} = E(Y \mid X) = \beta_0 + \beta_1 X \text{ with coefficients:}$$
$$\beta_0 = \mu_Y - \frac{\sigma_{YX}}{\sigma^2_X}\mu_X$$
$$\beta_1 = \frac{\sigma_{YX}}{\sigma^2_X}$$

  when the variables X and Y are jointly normally distributed.

The law of iterated expectations I

- Let Y and X be two random variables and let $E\left(Y \mid X\right)$ be a conditional expectation function (not necessarily linear)
- the *Law of iterated (or double) expectations* says that:

$$E\left[E\left(Y \mid X\right)\right] = E\left(Y\right). \tag{6}$$

- In HGL, this law is presented in Appendix B 1.7 and B.2.4
- In BN you give a proof by solving exercise 4.12
- The *interpretation* of (6) is that if we take the expectation over all the values that we first condition on, we obtain the unconditional expectation.
- Heuristically: If we use the probabilities of all the values that X can take, it does not matter what the value of X is.

Linearity of conditional expectations

- For all operations where we condition on X, we treat X as if it was a deterministic number.
  - For example: $E[X \mid X] = X$ and $E\left[\sqrt{X} \mid X\right] = \sqrt{X}$
- This motivates the linearity property of conditional expectations: For any deterministic function $h(X)$:

$$E[h(X)Y \mid X] = h(X)E(Y \mid X) \quad (7)$$

Combined with the Law of iterated expectations, this gives a powerful result:

$$E[h(X)Y] = \underbrace{E\{E[h(X)Y \mid X]\}}_{\text{Law of Itr Exp}} = E[h(X)E(Y \mid X)] \quad (8)$$

Linear independence between variables

If the conditional expectation function is constant, and therefore independent of the conditioning variable, so $E(Y \mid X) = E(Y)$, then $Cov(X, Y) = 0$.

The proof is by use of (8):

$$Cov(X, Y) = E(XY) - E(X)E(Y) = \underbrace{E\left[XE\left(Y \mid X\right)\right]}_{(8)} - E(X)E(Y)$$

$Cov(Y, X) = 0$ if $E(Y \mid X) = E(Y)$.

Conditional expectation in model form I

Often (although not always) the random variable Y can be written as

$$Y = E(Y \mid X) + \varepsilon \tag{9}$$

where $E(\varepsilon \mid X) = E(\varepsilon) = 0$,and therefore $E(X\varepsilon) = 0$.

- (9) can be interpreted as using the conditional expectation of Y given X as a model of Y.
- This results will be an important reference for the econometric regression models with random variables as regressors (RM2)
- The linear function form of $E(Y \mid X)$ is not general, i.e. $E(Y \mid X)$ in (9) may well be a non-linear function.
- Nevertheless, we will concentrate on linear $E(Y \mid X)$ functions in our Introductory Econometrics course

Conditional expectation in model form II

- *Variable transformations* can be interpreted as an effort to "prepare the data" for the use of a linear conditional expectation function at the modelling stage
- This is popular among economists, since the regression coefficients for the transformed variables can usually be given economic interpretation
  (Refer back to Lecture 2, and HGL Ch 4 and BN kap 2)

Motivation I

- Already in the discussion of RM1, we have encountered questions about the statistical properties of the estimators when "n grows towards infinity"
- A precise answer requires asymptotic analysis
- An alternative is to simulate the asymptotic properties if of estimators and test statistics by Monte Carlo simulation.
- In this last part of Lecture 6 we review a few of the elementary concepts and theorems.
- The exposition is not complete, even for elementary asymptotic analysis, but might serve as a reference point for further studies.

Probability limit I

Definition (Convergence in probability)

Let $\{Z_n\}$ be an infinite sequence of random variables. If for all $\epsilon > 0$

$$\lim_{n\to\infty} P\left(|Z_n - Z| > \epsilon\right) = 0$$

Z_n converge in probability to the random variable Z. Convergence in probability is written

$$Z_n \xrightarrow{p} Z.$$

The random variable Z is called the **probability limit** of Z_n. A much used notation is:

$$\text{plim}\left(Z_n\right) = Z.$$

Consistency of estimators

We are often interest in situations where Z_n converge to a number c_Z, so that

$$Z_n \xrightarrow{p} c_Z.$$

If an estimator converges in probability to the true parameter value, it is a *consistent estimator*.
Assume that $\hat{\theta}_n$ is an estimator of θ from a sample of n observations. Let n grow towards infinity. The sequence of estimators $\hat{\theta}_n$ is a converging sequence if

$$\operatorname{plim} \hat{\theta}_n = \theta \tag{10}$$

and (10) then defines $\hat{\theta}_n$ as a consistent estimator of θ.

Rules for the probability limit (Slutsky's theorem)

Let the infinite sequences Z_n and W_n converge in probability to the constants c_Z and c_W. The following rules then hold

$$\begin{aligned} \text{plim}\,(Z_n + W_n) &= \text{plim}\, Z_n + \text{plim}\, W_n = c_Z + c_W \\ \text{plim}\,(Z_n W_n) &= \text{plim}\, Z_n \times \text{plim}\, W_n = c_Z c_W \\ \text{plim}\left(\frac{Z_n}{W_n}\right) &= \frac{\text{plim}\, Z_n}{\text{plim}\, W_n} = \frac{c_Z}{c_W}. \end{aligned}$$

- In econometrics these rules are much used, because **empirical moments**, averages and empirical (co)variances can be shown to converge in probability to their theoretical counterparts, expectation and covariance.
- In this way consistency, or inconsistency, can often be shown for a given estimator, and for a given model specification

Law of large numbers I

Theorem (Weak law of large numbers)

Let X_i be independent and identically distributed variables with $E(X_i) = \mu_X$ and $0 < \sigma_X^2 < \infty$. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges in probability to μ_X

$$\text{plim}\left(\bar{X}_n\right) = \mu_X.$$

Law of large numbers II

Application: In RM1, the OLS estimator for α is

$$\hat{\alpha}_n = \alpha + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i = \alpha + \bar{\varepsilon}_n$$

If the classical assumptions of the model hold, then

$$\operatorname{plim}\left(\bar{\varepsilon}_n\right) = 0$$

by the Law of large numbers and the OLS estimator $\hat{\alpha}_n$ is consistent:

$$\operatorname{plim}\left(\hat{\alpha}_n\right) = \alpha$$

Convergence in distribution I

- In regression models with stochastic regressors, where the classical assumptions for the disturbances hold, but where we don't invoke the assumption about normally distributed disturbances, it is possible to show that the distribution functions of the random variable $\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right)$ converge to a cumulative normal probability distribution.
- Results about convergence in distribution often make use of the Central Limit Theorem

Convergence in distribution II

Theorem (Central Limit Theorem)

Let X_i be independent and identically distributed variables with $E(X_i) = \mu_X$ and $0 < \sigma_X^2 < \infty$. The distribution function of the sequence of standardized averages $Z_n = \frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}}$ converges to the cumulative distribution function of the standard normal distribution, so that $\{Z_n\}$ will converge to the standard distributed random variable $Z_n \xrightarrow{d} Z \sim N(0,1)$.

Remark: The notation $Z_n \xrightarrow{d} Z \sim N(0,1)$ means "converge in distribution" and should not be confused with:

$$Z_n \xrightarrow{p} c_Z.$$

Convergence in distribution III

Application: For the OLS estimator $\hat{\alpha}$ of α in RM1,

$$Z_n = \sqrt{n}\frac{(\hat{\alpha}_n - \alpha)}{\sigma} = \sqrt{n}\frac{\bar{\varepsilon}_n}{\sigma}$$

This means that

$$\sqrt{n}\frac{(\hat{\alpha}_n - \alpha)}{\sigma} \xrightarrow{d} N(0,1)$$

since the classical assumptions part of RM1 satisfy the conditions of the Central Limit Theorem.

We then also have:

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{d} N(0,\sigma^2)$$

Why do we consider convergence of $\sqrt{n}(\hat{\alpha}_n - \alpha)$ and not $(\hat{\alpha}_n - \alpha)$?