ECON 3150/4150, Spring term 2014. Lecture 2

Ragnar Nymoen

University of Oslo

16 January 2014

References to Lecture 2 (this slide set)

- Stock and Watson (SW)
  - Ch 2.5 and 3-3.6; 3.7 is saved for Lecture 3, as a bridge and motivation for Ch 4: Linear regression with One Regressor
- Bårdsen and Nymoen (BN)
  - Kap 4.5-4.9

Sum and averages of variable I

- We continue of tour into the statistical theory part of the combined discipline called econometric modelling.
- Lecture 1 noted several important results. For example,
- A linear combination of n independent normal variables $\{X_i; i = 1, 2, \ldots, n\}$

$$Y = a + \sum_{i=1}^{n} b_i X_i \tag{1}$$

has the normal distribution:

$$Y \sim N\left(a + \sum_{i=1}^{n} b_i \mu_{Xi}, \sum_{i=1}^{n} b_i^2 \sigma_{Xi}^2\right) \tag{2}$$

Sum and averages of variable II

- It follows that the average of n identically distributed independent (i..i.d) normal variables, $X_i \sim IIN(\mu_X, \sigma_X)$, is distributed:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) \tag{3}$$

which can be standardized:

$$\frac{\bar{X} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} \sim N(0, 1). \tag{4}$$

- See **Key concept 2.5** in SW about *i.i.d Random Variables* and the assumptions of *Simple Random Sampling*.

Sum and averages of variable III

- The results in (3) and (4) simply add the assumption that the distribution of each of the n i.i.d. variables is standard normal $N(0, 1)$.
- Note, that results (2.46)-(2.48) on page 88 in SW follow directly from (3) with suitable choice of symbols.
- However, it is not always that we can realistically make the normality assumption.
- We therefore need results that tell us when the normal distribution is a valid approximation

The Central Limit Theorem,CLT I

Theorem (Central Limit Theorem)

Let X_i $i = 1, 2, , ..., n$ *be independent and identically distributed random variables with* $E\left(X_i\right) = \mu_X$ *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:* $Z_n \xrightarrow{d} Z \sim N\left(0, 1\right)$.

- This is the same as in *Key Concept 2.7* in SW.
- It is an important theorem about "convergence in distribution".

The Central Limit Theorem,CLT II

- Note that nothing is said about the common distribution of the X_i variables (it can be any distribution), compare Figure 2.10 for an illustration.
- Therefore, CLT is all important for being able to perform statistical testing in cases where we, as researchers, cannot control (or design) an exact distribution for each X_i.

A special case (de Moivre's theorem)

Theorem (de Moivres theorem)

'If X has a binominal (Bernoulli) distribution with $E(X) = np$ and $var(X) = np(1-p)$, the standardized variable $Z = \frac{X - E(X)}{\sqrt{var(X)}}$ converges to $Z \sim N(0,1)$ when $n \to \infty$.

- The approximation is good when $np(1-p) > 10$

A generalization of CLT to functions I

- Due to a theorem called the continuous function theorem, it is possible to extend CLT to random variables that are functions of variables for which CLT holds.
- For example, if CLT holds, then

$$Z_n^2 = \left(\frac{\bar{X}_n - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} \right)^2 \xrightarrow{d} \chi^2(1)$$

which is the basis for the large sample (asymptotic) counterparts to the results about the χ^2, t and F distributed random variables in *slide 23-35 in Lecture 1*.

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:

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

Law of large numbers

Theorem (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

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

- The proof is with reference to the famous *Chebychev's-inequality*. SW page 715-716: BN page 92.
- Law of large Numbers is found in *Key Concept 2.6 in SW*

Consistency of estimators

- We are often interest in situations where Z_n converge to a number c_Z, so that
$$Z_n \xrightarrow{p} c_Z.$$
as in the case of the Law of Large Numbers.
- 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
$$\operatorname{plim} \hat{\theta}_n = \theta \tag{5}$$
and (5) then defines $\hat{\theta}_n$ as a consistent estimator of θ.
- A sufficient condition for consistency is that the estimator is unbiased for every value of n, $E(\hat{\theta}_n) = \theta$, and that $var(\hat{\theta}_n) \to 0$ as $n \to \infty$.

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

- A population *parameter* is a numerical aspect of a statistical distribution. Expectation and variance are examples
- The goal of statistical inference is to obtain valid conclusions about the population parameters with the use of the data in a given sample.
- Let θ denote the unknown parameter (it is a number!) we are interested in. Statistical methodology lets us formulate a function $\hat{\theta}$ of the observed random variables. $\hat{\theta}$ is called an *estimator*.
- The *estimator* $\hat{\theta}$ is itself a stochastic variable that has a distribution that follow from the assumptions first made about the distribution function of the variables
- A realization of the random variable $\hat{\theta}$ is called an *estimate*.

- It is common to let denote $\hat{\theta}$ both the estimator (a random variable) and the estimate (a number).
- Statistical inference has many aspects:
  - Parameter estimation—we want an estimator of $\hat{\theta}$ of θ to have "good properties"!
  - Hypothesis testing
  - Confidence interval construction
- Here: Review the theory for testing a hypothesis about the population mean in an i.i.d experiment

Model formulation I

- Assume that we have n independent and identically normally distributed ε_i variables

$$\varepsilon_i \sim IID\left(0, \sigma_\varepsilon^2\right) \quad i = 1, 2, \ldots, n \tag{6}$$

  and that the variable Y_i is defined by the function:

$$Y_i = \mu_Y + \varepsilon_i \tag{7}$$

  where μ_Y is a parameter of the i.i.d. sequence $\{Y_i; i = 1, 2, ..., n\}$
- (6) and (7) define a *statistical model*
- The rules for making logically *valid inference* about μ_Y is based on the model.

Model formulation II

- If the assumptions of the model are wrong (in this case the i.i.d. assumption) the inference may no be *reliable* (cf. "black swan")
- Both valid and reliable statistical inference is model-based. Conclusions follow from assumptions, here as elsewhere.
- There is nothing wrong about making bold assumptions in order to establish a model. But in econometrics, a critical attitude to the assumptions of the statistical model is essential.

Finding a good estimator I

- Consider $\bar{Y}$ as an estimator of μ_Y:

$$\hat{\mu}_Y = \bar{Y} \tag{8}$$

- $\hat{\mu}_Y$ is clearly a random variable, because $\bar{Y}$ is a random variable (because Y_i is, because...)
- What is the expectation and variance of $\hat{\mu}_Y$?

Finding a good estimator II

- We have already the answer in (3), since by applying rules of E and var on

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n}(\mu_Y + \varepsilon_i) = \mu_Y + \bar{\varepsilon} \tag{9}$$

we get

$$E(\hat{\mu}_Y) = \mu_Y \tag{10}$$

$$Var(\hat{\mu}_Y) = \frac{\sigma_\varepsilon^2}{n} \tag{11}$$

  - (10) says that $\hat{\mu}_Y$ is a good guess on μ_Y on average. $\hat{\mu}_Y$ is an *unbiased estimator* (“*forventningsrett*”)
  - (11) shows that $Var(\hat{\mu}_Y) \rightarrow 0$ when n.

Finding a good estimator III

- $\hat{\mu}_Y$ is thefore a *consistent estimator:*

$$\text{plim}\,\hat{\mu}_Y = \mu_Y$$

- Also BLUE (but we come back to that)

Large sample test I

- The test situation with one-sided alternative:

$$H_0\colon\ \mu_Y = \mu_Y^0 \quad \text{against} \quad H_1\colon\ \mu_Y > \mu_Y^0.$$

- Intuitively, we reject H_0 when $\bar{Y}$ is larger than we can expect if H_0 is true.
- But how "large is large"? We need a decision rule that allows for "randomly high" $\bar{Y}s$ even when H_0 holds.

Large sample test II

- From CLT we have:

$$\frac{\hat{\mu}_Y - \mu_Y^0}{\sqrt{Var(\hat{\mu}_Y)}} \xrightarrow{d} N(0,1) \tag{12}$$

  under H_0, meaning that we can obtain the *critical values* (see Key Concept 3.5) that we need to perform an asymptotic test from the $N(0,1)$ distribution.

- The standardized random variable $\frac{\hat{\mu}_Y - \mu_Y^0}{\sqrt{Var(\hat{\mu}_Y)}}$ is usually called the *t-ratio* (or the *t-statistic*), when it it $N(0,1)$ distributed.

Large sample test III

- When σ_ε^2 is unknown, we estimate it by:

$$\hat{\sigma}_\varepsilon^2 = \frac{S^2}{n-1}$$

where S^2 is the sum of squares for the Y_is (as in Lecture 1 slide 28, with an obvious change in notation).

- $\hat{\sigma}_\varepsilon^2$ is an consistent estimator of σ_ε^2, because

$$E(\hat{\sigma}_\varepsilon^2) = E\left(\frac{S^2}{n-1}\right) = E\left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2}\frac{S^2}{n-1}\right) = \frac{\sigma_\varepsilon^2}{n-1}E\left(\frac{S^2}{\sigma_\varepsilon^2}\right) = \sigma_\varepsilon^2$$

$$var(\hat{\sigma}_\varepsilon^2) = \frac{\sigma_\varepsilon^4}{(n-1)}var\left(\frac{S^2}{\sigma_\varepsilon^2}\right) = \frac{\sigma_\varepsilon^4}{(n-1)^2}2(n-1) = \frac{2\sigma_\varepsilon^4}{(n-1)}$$

Large sample test IV

- It follows that the critical values obtained from the $N(0, 1)$ are reliable to use, also when the variance of Y_i is unknown and has to be estimated.

Finite sample test I

- If we can replace the $\varepsilon_i \sim IID\left(0, \sigma_\varepsilon^2\right)$ with the stronger assumption $\varepsilon_i \sim IIN\left(0, \sigma_\varepsilon^2\right)$, we can obtain an exact test for any given sample size n.(See Section 3.6 in SW)
- We use the same estimator of $\sqrt{Var(\hat{\mu}_Y)}$ as above:

$$\sqrt{\widehat{Var(\hat{\beta}_0)}} = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{n}}$$

but the t-ratio is now distributed under H_0 :

$$\frac{\hat{\mu}_Y - \mu_Y^0}{\sqrt{\widehat{Var(\hat{\mu}_Y)}}} \sim t(n-1) \qquad (13)$$

Finite sample test II

- The proof is by showing that by re-arrangement, the random variable can be written as:

$$\frac{\frac{\hat{\mu}_Y - \mu_Y^0}{\frac{\sigma_\varepsilon}{\sqrt{n}}}}{\frac{\sqrt{\frac{S^2}{\sigma_\varepsilon^2}}}{\sqrt{n-1}}} \sim t(n-1)$$

to conform with (10) in Lecture 1, and noting that the numerator is $N(0,1)$ and the denominator is $\chi^2(n-1)$.

Finite sample test III

- To perform the test, we use the test statistic

$$t = \frac{\hat{\mu}_Y - \mu_Y^0}{\sqrt{\widehat{Var(\hat{\mu}_Y)}}} \tag{14}$$

which is observable under H_0. The estimated $\hat{\mu}_Y$ and $\sqrt{\widehat{Var(\hat{\mu}_Y)}}$ from the sample are used to calculate the value ("score") of the t-statistic (14), for example 2.5. The H_0 is rejected if $t = 2.5$ is higher than the critical value corresponding to the chosen *significance level* (controlling *Type-I error)*

- Alternatively: Reject when the *p-value* is lower than a chosen significance level.

Take care to review I

1. Calculation of critical values, and p-value when H_1 is *two-sided*
2. The calculation and estimation of a *confidence interval* for μ_Y and for the difference of two population means.
3. In class, we review the interesting Section 3.5 in SW about the estimation and testing of causal effect using difference of means.