ECON 4160, Spring term 2013. Lecture 2

Ragnar Nymoen

University of Oslo

28 August 2013

References to Lecture 2

- Ch 3-5
- Ch 10.1-10.2, 10.5, and Lecture note 2.
- 6.1-6.3

Classical inference theory I

- Chapter 3-5 in D&M contain the **classical inference theory** for the parameters of in the conditional expectation function (regression model), with the use of matrix algebra. The basic result is, if the "classical assumptions" hold for the disturbances, test stastics and confidence intervals can be based on:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{1}$$

$$Var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{2}$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k} \tag{3}$$

Classical inference theory II

at least asymptotically. But remember also:

	Disturbances ε are:			
$\mathbf{X}$	heteroscedastic		autocorrelated	
	$\hat{\beta}$	$\widehat{Var}(\hat{\beta})$	$\hat{\beta}$	$\widehat{Var}(\hat{\beta})$
exogenous	unbiased consistent	wrong	unbiased consistent	wrong
predetermined	biased consistent	wrong	biased inconsistent	wrong

- Since there is nothing new here compared to an introductory course, read it as an advanced review.

Classical inference theory III

- Despite the change from scalar notation to matrices: Note the familiar role of **restricted** and **unrestricted** sum of squared residuals in many of the tests!
- Take care to note:
  - The importance of $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I} \mid \mathbf{X})$ assumption for the regression ,model disturbances for obtaining **exact tests**
  - and the importance of $\boldsymbol{\varepsilon} \sim IID(\mathbf{0}, \sigma^2\mathbf{I} \mid \mathbf{X})$ for the corresponding **asymptotic tests.**
- Note that the asymptotic tests are also valid for the case where the explanatory variables are **predetermined**, rather than **strictly exogenous**.

Classical inference theory IV

- Note also chapter 4.6, explaining the role of simulation as a method of assessing the relevance of the **asymptotic test** for the cases of known DGP and unknown DGP. We speak of **Monte Carlo** simulation in the first case and **Bootstrap simulation** in the second case.
- We will use Monte Carlo simulation, but bootstrap methods are becoming more common in practical research, and the book gives an introduction.
- Chapter 5 reviews in particular the role of **heteroscedasticity consistent covariance matrices** (ch 5.5), in situations where $\widehat{Var}(\hat{\boldsymbol{\beta}})$ above is "wrong". We use the PcGive version of these in seminar exercises.

The delta method I

- When we estimate a linear-in-parameter conditional expectation function, the purpose is sometimes to test hypotheses about **derived parameters** that are **non-linear functions** of the regression coefficients.
- The so called *delta-method* (Ch 5.6) is based on Taylor-expansions and relatively weak assumptions.
- It gives the asymptotically valid estimate of the variance of the derived parameter

The delta method II

Assume the simple regression model

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

and that we are interested in the derived parameter

$$\theta = \frac{\beta_2}{\beta_1}$$

The asymptotic variance is then

$$\widehat{Var(\hat{\theta})} \approx \left(\frac{1}{\hat{\beta}_1}\right)^2 \left[\widehat{Var(\hat{\beta}_2)} + \hat{\theta}^2 \widehat{Var(\hat{\beta}_1)} - 2\hat{\theta}\widehat{Cov(\hat{\beta}_1, \hat{\beta}_2)}\right] \quad (4)$$

which actually covers many applications in econometrics.

The delta method III

- Ch. 5.6 covers the case of $\theta = g(\frac{\beta_2}{\beta_1})$ where $g(\cdot)$ is a monotonic and differentiable function, as well as the vector case.
- But we shall see that in our course, slight modifications of (4) will cover most of our needs.

Maximum Likelihood Estimation (MLE) I

- We now jump briefly to Ch. 10 to review the third estimation principle that we will use: MLE.
- If the joint pdf $f(Y_i, X_i)$ is normal, and $(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)$ are independet pairs of variables, we can derive the conditional model of Y given X as the regression model:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i,\ i = 1, 2, \ldots, n \tag{5}$$

whith $\varepsilon_i \sim N(\mathbf{0}, \sigma^2)$. This is often called a Gaussian disturbance, and the model is also known as the Gaussian regression model.

Maximum Likelihood Estimation (MLE) II

- By straight-forward extension of the results from this simple Gaussian regression model, we know that the MLE estimator of $\boldsymbol{\beta}$ in

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{6}$$

$$\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I} \mid \mathbf{X}) \tag{7}$$

  is identical to the OLS and MM estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.
- But

$$\hat{\sigma}^2_{ML} = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n}$$

  is biased. So the unbiased estimator

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k}$$

  is used instead to construct test-statistics.

Likelihood ratio tests I

- When MLE estimates are inserted back into the (log)likelihood function, we obtain the maximized likelihood.
- For (6) and (7) this value is

$$l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_{ML}) = \text{constant} - \frac{n}{2} \ln SSR(\hat{\boldsymbol{\beta}}) \tag{8}$$

where $SSR(\hat{\boldsymbol{\beta}})$ is the minimized sum of squared residuals after OLS estimation, i.e.

$$SSR(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} \tag{9}$$

- We can now think of (8) and (9) as the outcome of **unrestricted** estimation.

Likelihood ratio tests II

- If we impose a number of r parameter restrictions on the model and estimate by ML, we obtain new, **restricted**, estimators $\tilde{\boldsymbol{\beta}}$ and $\tilde{\sigma}^2_{ML}$. The corresponding restricted log-likelihood cannot be larger than the unrestricted $l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_{ML})$ in (8).
- The **Likelihood-Ratio test statistic** for the regression model is:

$$LR = 2(l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_{MLE}) - l(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2_{MLE}) = \frac{n}{2}\left[\ln SSR(\tilde{\boldsymbol{\beta}}) - \ln SSR(\hat{\boldsymbol{\beta}})\right] \quad (10)$$

- LR is asymptotically distributed as $\chi^2(r)$.

Likelihood ratio tests III

- The LR statistic is closely connected to the F-statistic for r-restrictions that we know from an introductory course:

$$F = \frac{SSR(\tilde{\boldsymbol{\beta}}) - SSR(\hat{\boldsymbol{\beta}})}{SSR(\hat{\boldsymbol{\beta}})} \frac{(n-k)}{r} \sim F(r, n-k) \qquad (11)$$

- On page 421, DM show that:

$$LR \approxeq rF$$

- Since the "F-version" of the test basically corrects for degrees of freedom, it is advisable to use it in small samples (e.g., $n < 40$).

Likelihood ratio tests IV

- *LR* is one of three *test principles* in use in modern econometrics. The others are called **Wald-test** and **Lagrange-Multiplier test**.
- In principle a Wald-test is based on (only) the unrestricted estimation. Examples: Standard t-test and asymptotic t-test for $H_0{:}\theta = \theta_0$ using (4) above (an exercise to Seminar 2 will provide an example).
- A LM-test is in principle only based on the restricted estimation.
- Asymptotically LR, W and LM are equivalent.
- See Ch 10.5 for more about the W- and LM test principle. Or the separate Lecture note 2.

NLS estimation I

- So far the conditional expectation functions have been linear in parameters.
- As we know, this allows a great deal of flexibility (trough non-linear variable transformation) in the specification of the functional form.
- Nevertheless: Sometimes necessary or appealing to estimate a model which is non-linear in the parameters.
- The sum of squared residuals that we want to minimize is then

$$SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i(\hat{\boldsymbol{\beta}}))^2 \tag{12}$$

where $\mathbf{x}_i(\hat{\boldsymbol{\beta}})$ is the analogue to $\mathbf{x}_i\hat{\boldsymbol{\beta}}$ in the linear case.

NLS estimation II

- The NLS estimator is consistent under mild assumptions
- Minimization of (12) requires numerical optimization, for example Newton's method, see Ch 6.4.
- PcGive has a good algorithm for computing NLS estimates. Illustrate by estimation of Phillips curve natural rate:

$$INF_i = \beta_1(U_i - \beta_2) + \varepsilon_i,\ i = 1, 2, \ldots, n \qquad (13)$$

- A non-linear regression function. β_2 can be interpreted as the natural rate (a parameter), since

$$E(INF \mid U = \beta_2) = 0$$

- Estimate by PcGive with Norwegian annual data.