ECON 4160: Seminars autumn semester 2015—FIRST SEMINAR

André K. Anundsen, Claudia Foroni and Ragnar Nymoen

August 26, 2015

Exercise set to seminar 1 (week 36, 31 Aug & 2 Sep)

This exercise set is longer than the later ones. In Question A we review PcGive use from the first computer class as well as some important results from an introductory course in econometrics. Question B gives some training in the use of the matrix algebra and the theory from Lecture 2 and 3.

Question A

1. Downloadthe pdf document *A first regression in OxMetrics-PcGive* from the course web page (*Computer Class* section) and the zippet data file the zip file *KonsDataSim* (from the *Data Sets* section). Follow the step-by-step instructions and become acquainted with simple regression analysis in OxMetrics-PcGive. (This is basically a review of some of the things from the first computer class).

2. Since the data set consists of time series, there is a natural ordering of the data (from "oldest" to "newest"). Therefore: estimate the same relationship using recursive estimation (see the note *A first regression in OxMetrics-PcGive*)**.** In the **Model-Test** menu choose **Recursive graphics** and then *Beta Coefficient* $\pm 2SE$. This should produce graphs with the sequences of point estimates as a function of the sample, both for the constant and for the regression coefficient, with $\pm$ 2 estimated coefficient standard errors.

   - The graphs can be said to "contain" the sequences of approximate 95 % confidence intervals. Why?
   - The graphs show that the confidence intervals are wider for shorter samples than for the longer sample. Can you briefly explain this feature?
   - The estimates of the coefficients are quite unstable at the start but the variability becomes less as the sample becomes longer. Is there an intuitive explanation for this?

3. Write the model we have estimated in Question A2 as

$$C_t = \beta_0 + \beta_1 I_t + \varepsilon_t,\ t = 1, 2, \ldots, T \qquad (1)$$

   with C_t for consumption and I_t for income.

(a) With reference to your introductory classes in econometrics and statistics, we let the expression

$$\hat{\sigma}_{CI} = \hat{\sigma}_{IC} = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \bar{C}\right)\left(I_t - \bar{I}\right)$$

denote the empirical covariance between the variables C and I . Likewise, we define the variances of I and C as $\hat{\sigma}_{II} = \hat{\sigma}_I^2 = \frac{1}{T}\sum_{t=1}^{T}\left(I_t - \bar{I}\right)^2$ and $\hat{\sigma}_{CC} = \hat{\sigma}_C^2 = \frac{1}{T}\sum_{t=1}^{T}\left(C_i - \bar{C}\right)^2$.
Show that the expression for the empirical covariance between C and I can be written in in the 3 alternative ways shown in equation (**??**):

$$\hat{\sigma}_{CI} = \frac{1}{T}\sum_{t=1}^{T}\left(C_t - \bar{C}\right)\left(I_t - \bar{I}\right) = \frac{1}{T}\sum_{t=1}^{T}C_t\left(I_t - \bar{I}\right) = \frac{1}{T}\sum_{t=1}^{T}I_t\left(C_t - \bar{C}\right)$$

(b) Now, let $\hat{\beta}_1$ denote the OLS estimator of β_1 in (1) and show that it can be expressed as:

$$\hat{\beta}_1 = r_{CI}\frac{\hat{\sigma}_C}{\hat{\sigma}_I}$$

where $r_{CI} = \frac{\hat{\sigma}_{CI}}{\hat{\sigma}_C\hat{\sigma}_I}$ is the correlation coefficient and $\hat{\sigma}_C$ and $\hat{\sigma}_I$ are the two variables' empirical standard deviations.

(c) Consider the "inverse regression"

$$I_t = \beta_0^{'} + \beta_1^{'}C_t + \varepsilon_t^{'} \tag{2}$$

and show that

$$\hat{\beta}_1 = \hat{\beta}_1^{'}\frac{\hat{\sigma}_C^2}{\hat{\sigma}_I^2}$$

where $\hat{\beta}_1^{'}$ is the OLS estimator for the "inverse regression".

(d) Assume that the sequence of recursive $\hat{\beta}_1$ estimates supports the interpretation that β_1 in (1) is a parameter which is *stable* over time. Assume next that the ratio $\hat{\sigma}_C^2/\hat{\sigma}_I^2$ is *unstable* over time. We may call this a *regime-shift* in the *system* that determines C_t and I_t, where σ_C^2 and σ_I^2 are parameters. Can $\hat{\beta}_1^{'}$ be recursively stable in this case?

(e) With reference to section §3.2 in Hendry's and Nilsen's book: Is the analysis you did in Q3e, relevant for the question whether (1) may have a structural interpretation?

4. Download the zip file *KonsData1Nor* from the course web page.

(a) Use the data for Norwegian consumption and income to estimate a log-linear "consumption function". Use the data series *CP* and *RCa* (click on the variable names to see a short description of the data) and transform to logs before estimating the consumption function. The data is quarterly and sasonally unadjusted, as a plot of the data will show, so include three seasonal dummies in the model . You do not have to create the dummies, just add *Seasonal* from the **Formulate** menu and Seasonal, Seasonal_1 and Seasonal_2 will be added to the model, representing dummies for the first, second and third quarter each year. Use the sample, 1970(1)-2012(1).

(b) Why do we only include 3 seasonal dummies when there are four quarters in a year? Would your answer be different if you had omitted the constant from the model? Explain briefly.

(c) What is the estimated elasticity of consumption with respect to income?

(d) Define the variable s_t as $s_t = \ln(CP_t) - \ln(RCa_t)$. Explain why this variable is approximately equal to (minus) the savings rate.

(e) Regress s_t on $\ln(RCa_t)$, the constant and the three seasonal dummies. Compared to the regression in 4(a), why has R^2 changed, while the estimated standard error of the regression ($\hat{\sigma}$) is unchanged?

(f) If you were asked to test the hypothesis that the savings rate is independent of income, what would the conclusion be when apply for example a 5 % significance level? Would this inference be reliable?

Question B

Let $\mathbf{X}$ be a $n \times k$ matrix with the regressors of the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{y}$ is $n \times 1$and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector with disturbances and the parameter vector $\boldsymbol{\beta}$ is $k \times 1$.

1. Define the residual vector $\hat{\boldsymbol{\varepsilon}}$

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ is a random vector (an estimator). Show that by requiring that $\boldsymbol{X}'\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$, an $\hat{\boldsymbol{\beta}}$ estimator of $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{5}$$

2. Why is the estimator in (5) the OLS estimator of $\boldsymbol{\beta}$?

3. Under which assumption(s) is it also the ML estimator?

4. Show that the elements in $n^{-1}\mathbf{X}'\mathbf{X}$ and $n^{-1}\mathbf{X}'\mathbf{y}$ are (uncentered) second order empirical moments.

5. Assume that the first column in $\mathbf{X}$ has the number 1 in each position. Partition $\mathbf{X}$ as $\mathbf{X} = \begin{bmatrix} \boldsymbol{\iota} & \mathbf{X}_2 \end{bmatrix}$ where $\boldsymbol{\iota}$ is the $n \times 1$ column vector with ones and $\mathbf{X}_2$ is the $n \times (k-1)$ matrix with random variables (and deterministic) variables $X_2, \ldots X_k$. Partition $\boldsymbol{\beta}$ accordingly as $\boldsymbol{\beta}' = \begin{pmatrix} \beta_1 & \boldsymbol{\beta}_2' \end{pmatrix}$ where β_1 is a scalar and $\boldsymbol{\beta}_2$ is $(k-1) \times 1$. Show that you can write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \boldsymbol{\iota}\alpha + (\mathbf{X}_2 - \bar{\mathbf{X}}_2)\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

where $\bar{\mathbf{X}}_2$ is the $n \times (k-1)$ matrix with the means of each variable in each column and

$$\alpha = \beta_1 + \bar{\mathbf{x}}_2' \boldsymbol{\beta}_2$$

where $\bar{\mathbf{x}}_2$ is the $(k-1) \times 1$ vector with the means of each of the $k-1$ variables $X_2, \ldots X_k$.

6. Explain why the OLS estimator $\hat{\boldsymbol{\beta}}_2$ is found as

$$\hat{\boldsymbol{\beta}}_2 = \left[(\mathbf{X}_2 - \bar{\mathbf{X}}_2)'(\mathbf{X}_2 - \bar{\mathbf{X}}_2)\right]^{-1} (\mathbf{X}_2 - \bar{\mathbf{X}}_2)'\mathbf{y} \tag{6}$$

and why the OLS estimator of α is

$$\hat{\alpha} = (\boldsymbol{\iota}'\boldsymbol{\iota})^{-1} \boldsymbol{\iota}'\mathbf{y} = \bar{Y} \tag{7}$$

where $\bar{Y}$ is the mean of the variables $\{Y_i;\, i = 1, 2, \ldots, n\}$ in the vector $\mathbf{y}$.

7. Is the first element in $\hat{\boldsymbol{\beta}}_2$ identical to the second element in $\hat{\boldsymbol{\beta}}$; the second in $\hat{\boldsymbol{\beta}}_2$ identical to the third in $\hat{\boldsymbol{\beta}}$, and so on? Explain your answer.

8. Use scalar notation to express $\hat{\boldsymbol{\beta}}_2$ for the case of $k-1 = 2$ and show that the existence of these estimators depends on both the absence of perfect collinarity and the existence of variability in each of the two variables