ECON 3150/4150: Seminars spring semester 2013

Ragnar Nymoen*

24 January 2013.

Exercise set to seminar 1 (week 6, 4-8 Feb)

The exercises to the first seminar invite you to review important results, and to train skills, from the first couple of lectures, where we worked with the least squares method of fitting a straight line to a given data set ("regression without statistics").

At the seminar it is important that you aid the seminar leader, for example by expressing clear views about what can be "checked out" quickly, and which exercises require more time and discussion. To do that you need to prepare!

Question A

1. For a given set of data observations X_i $(i = 1, 2, \ldots, n)$, and Y_i $(i = 1, 2, \ldots, n)$, show that:

(1) $$\sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

(2) $$\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n}(X_i - \overline{X})Y_i$$

(3) $$\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n}X_i(Y_i - \bar{Y})$$

where $\overline{X}$ and $\bar{Y}$ denote the arithmetic means.

2. The least squares estimation principle entails that we chose the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the following criterion function;

(4) $$S(\beta_0, \beta_1) = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

where $\{X_i, Y_i;\ i = 1, 2, \ldots, n\}$ are interpreted as given numbers (data).

Show that the first order conditions for a minimum of $S(\beta_0, \beta_1)$ are:

(5) $$\overline{Y} - \hat{\beta}_0 - \hat{\beta}_1\overline{X} = 0$$

(6) $$\sum_{i=1}^{n}X_iY_i - \hat{\beta}_0\sum_{i=1}^{n}X_i - \hat{\beta}_1\sum_{i=1}^{n}X_i^2 = 0.$$

*Thanks to Herman Kruse, Eivind Hammersmark Olsen and Erling Skancke for suggestions to, and discussion of, this exercise set.

3. Solve (5) and (6) for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Question B

1. Show that the alternative formulation of the criterion function

$$S(\alpha, \beta_1) = \sum_{i=1}^{n} (Y_i - \alpha - \beta_1(X_i - \overline{X}))^2 \tag{7}$$

gives rise to the same OLS estimate $\hat{\beta}_1$ as in Question A.

2. What is the algebraic relationship between the two OLS estimates $\hat{\beta}_0$ and $\hat{\alpha}$?

3. Assume that the empirical correlation coefficient between X and Y, $r_{X,Y}$, is 0.5. Does this imply that the regression coefficient $\hat{\beta}_1$ in the regression where X is the regressor is positive? Explain.

4. Consider the ("inverse") regression where X is the regressand and Y is the regressor. What is the expression for the OLS estimate of the slope coefficient (call it $\hat{\beta}_1'$) in that regression? Are the two regression coefficients the same?

5. Assume that you run the two regressions on two sub-samples with time series data and that you observe the following for the OLS estimate $\hat{\beta}_1$ (in the regression with X as the regressor) and the squared empirical correlation coefficient:

	$\hat{\beta}_1$	$r^2_{X,Y}$
First sample	0.5	$0.5^2 = 0.25$
Second sample	0.5	$0.9^2 = 0.81$

What are the two estimates for $\hat{\beta}_1'$ in the inverse regression (where Y is the regressor)?

6. How does the invariance of $\hat{\beta}_1$ with respect to the break in the correlation structure between the two sample periods affect your thinking about a possible causal relationship between X and Y?

Question C

Download the Stata data set *ConsInc.dta* from the course web page. (An Excel version of the data set has also been made available). This file contains annual observations for the period 1960 to 2006. There are two variables, dubbed I and C. The data is computer generated (artificial data) and we interpret I as real disposable income (in million kroner) and C as real private consumption expenditure (in million kroner).

1. Calculate the means $\bar{C}$ and $\bar{I}$ of private consumption and disposable income over the full sample 1960-2006. Then estimate a linear consumption function with I as regressor and C as regressand and find that the OLS estimated regression coefficient of 0.805. Use the estimated equation and the mean of income ($\bar{I}$), to calculate "the fitted mean" of C which we can denote $\widehat{\bar{C}}$. Compare $\widehat{\bar{C}}$ to $\bar{C}$. What do you find? Can you give an interpretation in terms of a scatter-plot and a regression line?

2. Calculate the two variables $c = C - \bar{C}$ and $i = I - \bar{I}$. Estimate two linear equations between c and i with the use of OLS: One with, and another without an intercept. What do you find? Can you explain your findings in terms of OLS algebra?

3. Divide C and I by 1000 so that the the unit of measurement becomes billion kroner instead of million kroner. How does this affect the OLS estimate of the intercept and of the regression coefficient?

4. Construct the two log-transformed variables $LC = \ln(C)$ and $LI = \ln(I)$. Regress LC on LI. What is the economic interpretation of the regression coefficient in this case?

Extra exercise

An answer proposal to this exercise will be posted on the course web page in Week 6.

Define the OLS fitted values for Y_i by

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, 2, \ldots, n$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the OLS estimates given by (5) and (6) above.

1. Show that, equivalently, $\hat{Y}_i$ can be defined by

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 (X_i - \overline{X}), \quad i = 1, 2, \ldots, n$$

2. Define the OLS residuals by

(8) $$\hat{\varepsilon}_i = Y_i - \widehat{Y}_i, \; i = 1, 2, ..., n$$

and show that

(9) $$\overline{\hat{\varepsilon}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i = 0$$

(10) $$\frac{1}{n} \sum_{i=1}^{n} (\hat{\varepsilon}_i - \overline{\hat{\varepsilon}})(X_i - \overline{X}) = 0$$

(11) $$\overline{\hat{Y}} = \frac{1}{n} \sum_{i=1}^{n} \hat{Y}_i = \overline{Y}$$

3. Explain intuitively why the following decomposition of the total variation in Y (the regressand) is true:

$$\underbrace{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}_{\text{total variation}} = \underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}_{\text{residual variation}} + \underbrace{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{\widehat{Y}})^2}_{\text{explained variation}}. \tag{12}$$

4. What is the relationship between the empirical correlation coefficient $r_{X,Y}$ between X and Y and the coefficient of determination (R^2)?

5. Do (9), (10) and (12) hold if the fitted values are instead from an OLS estimated linear relationship with no intercept (i.e. $\hat{\beta}_0$ is dropped)? What does this imply for the conventional R^2 statistic?

6. If we scale the variables Y_i and X_i in the way explained in Lecture 2, prior to OLS estimation. Will R^2 be the same as in the "unscaled case"?

Exercise set to seminar 2 (week 7, 11-15 Feb)

Question A

Consider the statistical regression model ("RM1" in the lectures)

(1) $$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1, 2, ..., n$$

for the case of a deterministic (non-stochastic) regressor where we assume that the random errors ε_i $(i = 1, 2, ..., n)$ have classical properties. In particular, we have that $Var(\varepsilon_i) = \sigma^2$ for all i $(\forall\ i)$.

1. With reference to the lectures, the re-parameterized model is

   $$Y_i = \alpha + \beta_1(X_i - \bar{X}) + \varepsilon_i,\ i = 1, 2, ..., n$$

   Find expressions for $Var(\hat{\alpha})$ and $Var(\hat{\beta}_1)$ by using the assumption of the model, and rules for the variance of a sum of stochastic variables, e.g. $Z_1, Z_2, \ldots, Z_n$;

   $$Var\left(\sum_{i=1}^{n} a_i Z_i\right) = a_i^2 \sum_{i=1}^{n} Var(Z_i) + 2\sum_{i<j}^{n} a_i a_j Cov(Z_i, Z_j)$$

2. With reference to the lectures, we have that $Cov(\hat{\alpha}, \hat{\beta}_1) = 0$. What are the expressions for $Var(\hat{\beta}_0)$ and $Cov(\hat{\beta}_0, \hat{\beta}_1)$?

Question B

Assume that Y_i is explained by a single dummy (or indicator) variable X_i that takes either the value 0 or 1. For concreteness think of Y_i as the number of recorded damages to the eye during New Year celebrations in year i, and define X_i as

$$X_i = \begin{cases} 1, \text{ if year } i \text{ is with the law against rockets in private fireworks} \\ 0, \text{ if not} \end{cases}$$

1. What are the algebraic expressions for the OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ in this case?

   Remark: In econometric books (as in HGL page 277) the terminology *Difference estimator* is often used.

2. In Norway, the use of rockets in private fireworks became illegal in 2008. Before this policy intervention the number of damages during New Year festivities averaged 20. The numbers of damages during the New Year celebrations after rockets were prohibited have been: 10, 10,17,16,17. What is the estimated values of $\hat{\beta}_0$ and $\hat{\beta}_1$ based on these data?[1]

3. Using the model above: What is your predicted number of damages for the next New Year celebration in Norway? What do you regard to be the main sources of forecast uncertainty in this case? (No calculations expected/required!)

[1] http://www.helse-bergen.no/aktuelt/nyheter/Sider/17-personar-med-alvorlege-augeskadar–.aspx

Question C

Download the data set *ManualMc_detX_seminar.dta*. This data file contains 100 observations from 10 Monte Carlo experiments (see section 2 and Appendix 2G in HGL for a brief explanation of Monte Carlo simulation, BN kap 5, side 108-110). Hence we have (Yj_i, X_i), where $j = 1, 2, ..., 10$ and $i = 1, 2, ..., 100$.

Since the observations for the regressor X are the same in all 10 data sets, we can interpret this as the case with deterministic regressor.

Estimate the 10 regressions between Yj and X, with X as the regressor and using the full sample size in each regression. Let a_j and b_j denote the OLS estimates from regression number j ($j = 1, 2, \ldots, 10$). Calculate the averages of the 10 estimates for the constant terms and the regression coefficients. These averages are Monte Carlo estimates of the true mathematical expectations $E(\hat{\beta}_0)$ and $E(\hat{\beta}_1)$ of the OLS estimators of the parameters β_0 and β_1 in the regression model (1) when X_i is interpreted as a deterministic variable. The true values of the parameters are $\beta_0 = 0.1$ and $\beta_1 = 0.75$. What are the (estimated) biases?

Exercise set to seminar 3 (week 9, 25 Feb-1 Mar)

Question A

Consider the two models

(1) $$Y_i = \beta_0 + \varepsilon_i,\ i = 1, 2, ..., n$$

and

(2) $$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1, 2, ..., n$$

where we for both models assume that the disturbances ε_i have classical properties and that they are normally distributed with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. In (2), X_i is interpreted as deterministic.

From statistical theory, we have the important results

(3) $$\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sigma^2} \sim \chi^2(n-1)$$

for the OLS residuals $\hat{\varepsilon}_i$ from model (1), and

(4) $$\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sigma^2} \sim \chi^2(n-2)$$

for the OLS residuals from model (2).

1. The degrees of freedom (*df*) can heuristically be defined as the number of observations (here, this is the same as the number of residual variables $\hat{\varepsilon}_i$), minus the number of independent linear restrictions between the variables. Can you give an intuitive explanation for the difference in *df* between (3) and (4)?

2. Explain briefly why the test statistic that we use when we make inference about β_1 has a t-distribution (under the assumptions given above)

Question B

1. Download the data set *BNPcap.xls/dta*. The file contains time series data for the variable *BNPcap* which is Norwegian GDP per capita (GDP is measured in fixed 2000 prices, Million NOK). The data is annual and the sample period 1830-2010. There is also a deterministic time trend in the data set (*Trend).*

2. Create a new variable which is the natural logarithm of *BNPcap*. Call it *LBNPcap*. Create also the difference *DLBNPcap=LBNPcap-LBNPcap_1*. Finally create the exact GDP per capita growth rate. Call it *BNPcapgrowth* for example. Plot *DLBNPcap* and *BNPcapgrowth* against time. Explain why *DLBNPcap* is an approximation of *BNPcapgrowth,* and explain also when the approximation in good, and when it is less satisfactory.

3. Show how you can estimate the parameter γ in the non-linear relationship

(5) $$BNPcap_t = A\exp(\gamma Trend_t + \varepsilon_t)$$

   by OLS, and by using the transformed data series *LBNPcap*.

4. What is the OLS estimate of γ when you use *LBNPcap* and the sample is 1947-2010.

5. Show that an alternative OLS estimate of γ can be based on a regression with *DLBNPcap* as the regressand

6. Assume that the correlation coefficient between ε_t and ε_{t-1} in (5) is close to $+1$. Which of the two models (i.e. in Q4 and Q5) would you then use for forming a 95 % confidence interval for the growth rate parameter γ? Calculate the confidence interval of your choice.

7. Assume instead that the correlation coefficient between ε_t and ε_{t-1} in (5) is close to zero, and use the model in Q4 to calculate a 95 % prediction interval for *LBNPcap* for 2011.

8. Based on your model, what is your predicted value of *BNPcap* in 2011? How accurate is this prediction when you compare with data that Statistics Norway has published for the year 2011?

Exercise set to seminar 4 (week 10, 4-8 Mar)

Question A

1. Consider the three stochastic variables X, Y and Z. The three variables are connected by the linear function

$$Y = a + bX + Z$$

where a and b are parameters. Assume that $E(Z) = 0$ and $Cov(X, Z) = 0$ and show that the parameter b can be written as

$$b = \frac{Cov(X, Y)}{Var(X)}$$

Note: This is an example of a population regression. An interpretation of the regression model with stochastic regressor in econometrics is that is a method that by conditioning allows us to obtain probabilistic knowledge about the population slope parameter, written as b here, without having prior knowledge about $Cov(X, Y)$ and $Var(X)$

2. Assume that X and Y are two stochastic variables. The law of iterated expectations (also called law of double expectation) states that

(1) $$E\left[E(Y \mid X)\right] = E(Y)$$

   (a) Interpret the two expectations operators on the left hand side of equality sign.

   (b) Try to prove (1) for the case where X and Y are discrete stochastic variables.

3. Consider the two stochastic variables X and Y. The conditional expectation $E(Y \mid x)$ is deterministic for any given value of $X = x$. But we can also regard the expectation of Y for all possible values of X. In this interpretation $E(Y \mid X)$ is a stochastic variable with realization $E(Y \mid x)$ when $X = x$. Hence, the conditional expectation function $E(Y \mid X)$ is a function of the stochastic variable X so that we can write $E(Y \mid X) = g_X(X)$ where $g_X(X)$ is a function.

Consider the discrete distribution function in table 1.

Table 1: A discrete probability distribution for X and Y

		X			
		-8	0	8	$f_Y(y_i)$
	-2	0.1	0.5	0.1	0.7
Y	6	0	0.2	0.1	0.3
	$f_X(x_i)$	0.1	0.7	0.2	

Note that the marginal probabilities are in the last row (X) and in the column on the right right (Y) of the table.

Table 2: Conditional distribution for Y given X, based on table 1.

		X		
		-8	0	8
Y	-2	$\frac{0,1}{0,1}$	$\frac{0.5}{0.7}$	$\frac{0.1}{0.2}$
	6	$\frac{0}{0.1}$	$\frac{0.2}{0.7}$	$\frac{0.1}{0.2}$

(a) Obtain the conditional distribution for Y shown in table 2.

(b) Show that the conditional expectation function $E(Y \mid X)$ becomes:

$$E(Y \mid X = -8) = -2$$
$$E(Y \mid X = 0) = \frac{2}{7}$$
$$E(Y \mid X = 8) = 2$$

(c) Calculate $E\left[E(Y \mid X)\right]$ by using the law of iterated expectation, i.e., by summing the three conditional expectations multiplied by the marginal probabilities $f_x(x_i)$.

(d) Use the marginal distribution for Y to show that $E(Y) = E\left[E(Y \mid X)\right]$, as a check of the law of iterated expectations.

4. Consider the stochastic variables X and Y and the conditional expectations function $E(Y \mid X)$. Let the stochastic variable ε be implicitly defined by the equation

$$Y = E(Y \mid X) + \varepsilon$$

(a) Show that $E(\varepsilon \mid X) = 0$.

In the lectures, we show that $Cov(X, \varepsilon) = E(X\varepsilon) = 0$ is a consequence of $E(\varepsilon \mid X) = 0$.

(b) Assume that $E(Y \mid X)$ is linear: $E(Y \mid X) = \beta_0 + \beta_1 X$. What are $E(\varepsilon \mid X)$ and $Cov(X, \varepsilon)$ for the stochastic variable ε which is defined by

$$Y = \beta_0 + \beta_1 X + \varepsilon?$$

5. Assume that Y_i and X_i are stochastic variables, $i = 1, 2, \ldots, n$. Assume that the conditional expectations function $E(Y_i \mid x_i)$ is linear: $E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$. Show that the OLS estimator $\hat{\beta}_1$ is unbiased:

$$E(\hat{\beta}_1 - \beta_1) = 0$$

Question B

Download the data set *ManualMc_stochX_seminar.dta*. The only difference from the data in *ManualMc_detX_seminar.dta* that we analyzed in Seminar 2, is that we now have random variation in the regressor, so the 10 data sets are (Yj_i, Xj_i), where $j = 1, 2, ..., 10$ and $i = 1, 2, ..., 100$.

Use the data set to obtain a Monte Carlo estimate of the biases $E(\hat{\beta}_0) - \beta_0$ and $E(\hat{\beta}_1) - \beta_1$ in the case where X_i is a stochastic regressor and $\beta_0 = 0.1$ and $\beta_1 = 0.75$. Compare the differences/similarities of results from Seminar 2. (Remember that the data generating process has disturbances with classical properties in both data sets).

Exercise set to seminar 5 (week 11, 11-15 Mar)

Question A

Let μ_{at} denote the total required rate of return on an asset in period t. The capital asset pricing model (CAPM) states that

$$\mu_{at} = r_t + \phi_t \sigma_a \rho_{am}$$

where r_t is a riskless interest rate in time period t (usually the return on government bonds). ϕ_t is a an aggregated market parameter (the market price of risk). σ_a and ρ_{am} are parameters of the joint probability density function $f(r_{at}, r_{mt})$ of the return on an individual asset (r_{at}) and the stock market return (r_{mt}). σ_a is the standard deviation of the real return on the asset, and ρ_{am} is the correlation coefficient between r_{at} and r_{mt}.

1. Show that, with the use of

   $$\phi_t = \frac{E(r_{mt}) - r_t}{\sigma_m}$$

   where σ_m is the standard deviation of the market real return, the CAPM relationship can be re-expressed as:

   $$\mu_{at} = r_t + (E(r_{mt}) - r_t)\frac{\sigma_{am}}{\sigma_m^2}$$

   where σ_{am} is the covariance between r_{at} and r_{mt}.

2. In the finance literature, a key parameter of interest is the so called BETA for the asset:

   $$\beta_a^{CAPM} \equiv \frac{\sigma_{am}}{\sigma_m^2}$$

   What is the economic interpretation of this parameter (just briefly)?

3. Under which (sufficient) assumption(s) is β_a^{CAPM} a parameter in the conditional expectation function $E(r_{at} \mid r_{mt})$?

4. *Oilvol.xls/dta* contains weekly data for oil prices in dollar per barrel (Oljespotpris) and the Standard & Poor 500 stock exchange index (S&P500_stock index). Each observation is an average of daily observations. The start of the data set is week 1 in 2005, and the last observation is from the last week of June in 2008.

   Specify an econometric model, and estimate the BETA β_a^{CAPM} for the oil production real option. Use the data set *Oilvol.xls/dta* and construct the return on oil (r_{at}) as the weekly change in the log of the oil price in the data set, and construct r_{mt} and the weekly change in the log of the stock exchange index in the data set.

5. Test H_0: $\beta_a^{CAPM} = 0$ against H_1: $\beta_a^{CAPM} \neq 0$.

Question B

Download the data set *Multreg1.xls/dta*. The data set contains 100 observations ($n = 100$) of three stochastic variables *Ya*, *Za* and *Zb*.

1. Regress *Ya* on *Zb* and show that the OLS estimate of the (simple) regression coefficient is 0.58 (with two decimals). Take care to include an intercept, and explain why!

2. Let $\widetilde{Ya}_i = E(Ya_i \mid Za_i, Zb_i)$ denote the conditional expectation. Assuming that you are interested in estimating the partial derivative $\partial\widetilde{Ya}_i/\partial Zb_i$, how relevant is the point estimate 0.58?

3. Can you calculate a more relevant point estimate for $\partial\widetilde{Ya}_i/\partial Zb_i$ by using the data set, but without estimating a multiple (i.e., bivariate) regression?

4. Check your results in QB3 by estimating the bivariate regression with Ya_i as regressand, and Za_i and Zb_i as regressors.

5. State the necessary assumption about the properties of the disturbances of the model and test the following hypotheses:

   (a) The partial derivative of $\widetilde{Ya}_i$ with respect to Za_i is statistically significant different from zero

   (b) The partial derivative of $\widetilde{Ya}_i$ with respect to Zb_i is statistically significant different from zero

6. Are the two potential omitted variable biases statistically significant?

7. Investigate the degree of multicollinearity in this model. Would you say that multicollinearity is a problem here?

Exercise set to seminar 6 (week 12, 18-22 Mar)

Question A

1. Complete the econometric specification of the linear relationship

   (1) $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, i = 1, 2, ..., n$$

   so that it becomes a regression model with stochastic regressors (i.e., an example of *RM2)*

2. Assume that you construct OLS residuals $e_{Y|X_2i}$ and $e_{X_1|X_2i}$ and that you regress $e_{Y|X_2i}$ on $e_{X_1|X_2i}$. How will that estimate compare to the OLS estimate $\hat{\beta}_1$ from (1)? Explain briefly. What about the respective estimated standard errors?

3. Under the assumption of *RM2*, we know that $E(\hat{\beta}_1) = \beta_1$. Can you say anything about the expectation of the estimator in the model with $e_{Y|X_2i}$ as regressand and $e_{X_1|X_2i}$ as regressor?

4. Explain what is meant by partial empirical correlation coefficient in association with the estimation of (1).

Question B

Exercise 4.13 in HGL (as in the Lecture plan, the reference is to the fourth edition of the book).

Exercise set to seminar 7 (week 15, 8-13 Apr)

Exercise 5.3, 5.4. 5.14 in HGL

Exercise set to seminar 8 (week 16, 15-19 Apr)

Exercise 6.1, 6.2, 6.3, 6.5 in HGL

Exercise set to seminar 9 (week 17, 22-26 Apr)

1. Exercise 7.1, 8.1, 8.2 in HGL

2. With reference to Seminar 3, Question B, investigate the hypothesis of no autocorrelation in the disturbances of the two regression models used in Q4 and Q5

3. With reference to Seminar 3, Question B, test the hypothesis that there was a structural break in the growth parameter γ after the end of WW-II. Use one sample that starts in 1905 and ends in 1946, and a second that starts in 1947 and ends in 2012 . Can you think of other potential structural breaks in this parameter?

4. Consider an dynamic relationship *LBNPcap*:

$$\text{(1)} \quad LBNPcap_t = \beta_0 + \beta_1 LBNPcap_{t-1} + \beta_2 trend_t + \beta_3 trend_t^2 + \beta_4 D_{47,t} + \beta_5 D_{58,t} + \beta_6 D_{82,t} + \beta_7 D_{88,t} + \beta_8 D_{09,t} + \varepsilon_t$$

   where t is the deterministic time trend and $D_{year,t}$ are indicator variables for the years 1947, 1958, 1982, 1988 and 2009.

   (a) Complete the specification of (1) as a dynamic regression model.

   (b) Estimate the model on the post WW-II sample and investigate residual autocorrelation for this model.

   (c) Give an (asymptotically valid) 60 % confidence interval for the parameter β_1.

   (d) Can you test whether the two models that you used in Seminar 3, Question B4 and Question B5 are valid regression models?

   (e) What do the estimation results for (1) imply for the future of Norwegian GDP per capita in the future? No formal analysis required here!

Exercise set to seminar 10 (week 19, 6-10 May)

1. HGL (4th edition): Exercise 9.7.

2. Assume that the stochastic variable Y satisfies:

$$(1) \qquad \mathsf{E}(Y) = e^{\mu+\frac{1}{2}\sigma^2},$$

$$(2) \qquad \mathsf{E}[\ln(Y)] = \mu,$$

where μ and σ^2 are unknown parameters.

   (a) Show that $\ln[\mathsf{E}(Y)] > \mathsf{E}[\ln(Y)]$ always holds.

   (b) Explain how you,by using the Method of Moments and (1) and (2), would estimate the parameters μ and σ^2, from a sample of n observations: $Y_1, \dots, Y_n$.

   (c) Assume that you in addition to (1) and (2) know that Y follows a log-normal distribution (see Appendix 4C in the HGL textbook), which implies

$$(3) \qquad Var[\ln(Y)] = \sigma^2,$$

   Use this to propose a second estimator of σ^2 based on the Method of Moments principle. Does the existence of two estimators for the same parameter σ^2 represent a problem? Explain briefly.

3. Assume that the three observable variables (Y, X, Z) satisfy the relationship

$$(4) \qquad Y = \alpha + \beta_X X + \beta_Z Z + u,$$

where u is a disturbance.

   (a) Show, by using algebra for first- and second-order moments, that (4) implies

$$(5) \quad Var(Y) = \beta_X^2 Var(X) + \beta_Z^2 Var(Z) + Var(u) + 2\beta_X\beta_Z Cov(X,Z) + 2\beta_X Cov(u,X) + 2\beta_Z Cov(u,Z),$$

(6)
$$Cov(Y,X) = \beta_X Var(X) + \beta_Z Cov(Z,X) + Cov(u,X),$$

(7)
$$Cov(Y,Z) = \beta_X Cov(X,Z) + \beta_Z Var(Z) + Cov(u,Z).$$

   Assume that the theoretical variances and covariances
   $\sigma_Y^2 = Var(Y),\ \sigma_X^2 = Var(X),\ \sigma_Z^2 = Var(Z)$
   $\sigma_{YX} = Cov(Y,X),\ \sigma_{YZ} = Cov(Y,Z),\ \sigma_{XZ} = Cov(X,Z)$
   are unknown. They can be estimated consistently from the corresponding empirical variances and covariances, $S_Y^2, S_X^2, S_Z^2, S_{YX}, S_{YZ}, S_{XZ}$, because $\text{plim}(S_Y^2) = Cov(Y)$, $\text{plim}(S_{YX}) = Cov(Y,X)$ etc. (convergence in second-order moments) – by utilizing the Method of Moments (MM)

(b) Assume that we have an economic theory and a statistical hypothesis that together imply $Cov(u, X) = Cov(u, Z) = 0$. Derive from (6) and (7) expressions that give β_X, β_Z as functions of the theoretical variances and covariances of the three observable variables.

(c) Explain how the expressions β_X, β_Z that you have derived are related to the probability limits of the OLS estimators $\hat{\beta}_X, \hat{\beta}_Z$ of the parameters β_X and β_Z in (4) (assuming that you have n observations of (Y, X, Z)).

(d) Assume that we have an alternative economic theory and a statistical hypothesis that together imply: $\beta_Z = 0$, $Cov(u, Z) = 0$. However,$Cov(u, X) \neq 0$ and unknown. Derive from (6) and (7) – using these three assumptions – expressions for β_X and $Cov(u, X)$ as functions of the theoretical variances and covariances of the observable variables, when you assume that $Cov(X, Z) \neq 0$. Why is the last assumption essential? Explain next how the expression you have derived for β_X is related to the instrument variable estimator for β_X with Z serving as instrument for X. Still assume that you have n observations on (Y, X, Z).

(e) Change the assumptions made in Q3c) by instead of $\beta_Z = 0$, $Cov(u, Z) = 0$ assume only that $Cov(u, Z) = 0$ while β_Z is *a priori* unknown. Could you then use the procedure elaborated under Q3d) to estimate β_X? Explain briefly.