Seminar 2

Construction of Linear Regression Model (LRM)

1. Theoretical background

The traditional econometric approach might be divided into following steps that we have to deal with:

(i) Economic theory
(ii) Algebraic Economic model
(iii) Econometric Model
(iv) Economic data
(v) Estimation
(vi) Economic interpretation
(vii) Statistical verification
(viii) Application – policy analysis (ex-post analysis), simulation and forecasting

Econometric Model

Econometric model may contain two types of equations:

- Stochastic and
- Identity equations

The equations contain following variables:

- Endogenous variables – are generated within the model.
- Exogenous variables – are generated outside the model.
- Predetermined variables – are used to explain endogenous variables. Predetermined variables are: exogenous variables, lagged exogenous variables and endogenous variables.
- Error (Disturbance) terms.

2. Exercises

Data table

Year	C PM (kg/person/ year)	CP PM (CZK/kg)	CP BM (CZK/kg)	CP ChM (CZK/kg)	Income (CZK) *
Variable					
1995	8,04	84.2	94,81	52,32	55 578,0
1996	8,87	90,42	102,12	62,77	64 114,0
1997	8,74	92,11	104,82	70,64	70 968,0
1998	10,36	86,39	110,16	73,31	77 942,0
1999	9,78	80,47	107,80	56,51	80 771,0
2000	8,94	90,04	111,53	61,83	83 422,0
2001	9,05	101,66	112,56	71,28	90 167,0
2002	9,55	89,84	112,99	62,40	93 153,0
2003	10,14	82,74	108,02	60,67	98 102,0
2004	9,97	85,36	112,84	62,55	102 217,0
2005	11,18	85,30	117,73	62,73	116 573,5
Average	9,51	88,43	108,67	63,36	84 818,9

* Net income per household

Correlation matrix

Variable	C PM	CP PM	CP BM	CP ChM	ln
C PM	1				
CP PM	-0,369659	1			
CP BM	0,751629	0,165662	1		
CP ChM	0,209048	0,626810	0,416617	1	
ln	***0,814929***	-0,069005	***0,892922***	0,144044	1

Economic model

Pork meat consumption depends on pork meat consumer price, beef meat consumer price, chicken meat consumer price and income.

Assumptions (based on the economic theory):

- pork meat and chicken meat are substitutes;
- pork meat and beef meat are substitutes;
- income increase causes increase of pork meat consumption.

Tasks

1. Declare all variables in the data table (i.e. use symbols y and x with their relevant indexes).

2. Write values of the following variables:
   a. $y_{1,6}$ =
   b. $x_{2,10}$ =
   c. $x_{4,1}$ =

3. Write an algebraic form of the model defined above. Assume linear relationship among endogenous and exogenous variables.

4. Write the model based on the power function.

5. Modify the economic model (linear form of the model) to econometric model.

6. Explain term $u_{1t.}$

7. Write the procedure of $u_{1,5}$ calculation.

8. Check the multicollinearity occurrence and show possibilities of its elimination.

9. Data transformation:
    a. Include intercept term (constant) into model. How may it be included in data table?

    b. What is the problem of extreme values within the data set? What is the possible process of their detection? Propose any solution of the problem of extreme values occurrence.

    c. Which technique may be used to solve the problem of missing values?

    d. Modify data into form suitable for parameters' estimation in Microsoft Excel.

10. Define dynamic model (results should be written into the following table):
    a. Based on the time vector.

    b. Based on the lagged value (one period) of endogenous variable. What is the impact on the data table? Define vector of one period lagged values of variable y_{1t}, i.e. values of vector $y_{1(t-1)}$.

c. Based on the variables in form of 1st differences. Rewrite variables y_{1t}, x_{1t} a x_{2t} in form of 1st differences.

Table of other variables

Year	UV (intercept)	Time vector	$y_{1(t-1)}$	y_{1t} (differences)	x_{1t} (differences)	x_{2t} (differences)
Variable						
1995						
1996						
1997						
1998						
1999						
2000						
2001						
2002						
2003						
2004						
2005						
Average						

Modified data table

Year Variable	C PM (kg/person/ year)	CP PM (CZK/kg)	CP BM (CZK/kg)	CP ChM (CZK/kg)	Income (CZK) *
1996	8,87	90,42	102,12	62,77	8,536
1997	8,74	92,11	104,82	70,64	6,854
1998	10,36	86,39	110,16	73,31	6,974
1999	9,78	80,47	107,80	56,51	2,829
2000	8,94	90,04	111,53	61,83	2,651
2001	9,05	101,66	112,56	71,28	6,745
2002	9,55	89,84	112,99	62,40	2,986
2003	10,14	82,74	108,02	60,67	4,949
2004	9,97	85,36	112,84	62,55	4,115
2005	11,18	85,30	117,73	62,73	14,357
Average	9,66	88,43	110,06	64,47	6,100

* Net income per household

Modified correlation matrix

Variable	C PM	CP PM	CP BM	CP ChM	ln
C PM	1				
CP PM	-0,602642	1			
CP BM	0,614154	0,005568	1		
CP ChM	-0,163091	0,628116	-0,017416	1	
ln	0,456484	0,091175	0,220278	0,279075	1

Individual exercises

1. Based on the data table included in supplement 1 define one-equation linear model, which describes that households' expenditures depend on households' expenditures in previous period, inflation, interest rate and income. Define assumptions about the relations among the variables (based on the economic theory).

2. Declare all variables in the data table (i.e. use symbols y and x with their relevant indexes).

3. Write values of the following variables:
   a. $y_{1,4}$ =
   b. $x_{2,8}$ =
   c. $x_{3,2}$ =

4. Write an algebraic form of the model defined above. Assume linear relationship among endogenous and exogenous variables.

5. Write the model based on the power function.

6. Modify the economic model (linear form of the model) to econometric model.

7. Write the procedure of $u_{1,3}$ calculation.

8. Check the multicollinearity occurrence and show possibilities of its elimination.

9. Data transformation:
   a. Include intercept term (constant) into model.

   b. Modify data into form suitable for parameters' estimation in Microsoft Excel.

10. Check whether the model is static or dynamic. If the model is static, propose the possible ways to transform it into the dynamic form.