

Econometrics Exams

Contents

1	Econometrics, 3. February, 2012	1
2	Econometrics, 31. July, 2009	3
3	Econometrics, 28. July 2007	7
4	Ökonometrie, 21. 7. 2007	9
5	Ökonometrie, 25. 7. 2008	11

1 Econometrics, 3. February, 2012

You have one hour to answer all questions. Please explain all your answers.

Question 1: You estimate the following model using OLS (log denotes the natural logarithm function):

$$Y_i = \beta_0 + \beta_X \log(X_i) + \beta_Y \log(Z_i) + \beta_{XY} \log(X_i) \log(Z_i) + u_i$$

You estimate your regression model in R and obtain the following output:

```
lm(formula = Y ~ log(X) * log(Z))

Residuals:
    Min       1Q   Median       3Q      Max
-28.049 -10.189  -1.381   8.635  51.267

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)      -800         1000   -0.8   0.4257
log(X)             200           200    1.0   0.3198
log(Z)             200           200    1.0   0.3198
log(X):log(Z)    -100            50   -2.0   0.0483 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.69 on 96 degrees of freedom
Multiple R-squared:  0.9388,    Adjusted R-squared:  0.9369
```

- Given the model (not the output of the regression) what is the marginal effect of X on Y?
- Assume $X = 10$ and $Z = 1$. Given the output of the regression, what is now the marginal effect of X on Y?

- Your level of significance is 5%. You still assume that $X = 10$ and $Z = 1$. Can you say whether the marginal effect of X on Y is significantly different from 0?
- Can you say whether the marginal effect of X on Y is significantly different from 100?

Question 2: You use an OLS model that explains wages as a function of several variables, one of them is gender. According to your regression results females obtain a salary which is 200€ smaller than the salary of males. The 95%-confidence interval ranges from 110€ to 290€. Can you conclude that your data provides strong evidence of gender discrimination in this labour market? Explain!

Question 3: A recent study finds that people who sleep longer than eight hours per night have (on average) a higher death rate than people who sleep only six to seven

hours. The study was conducted with Americans aged 30 to 102 which were each tracked for four years. The death rate for each group was calculated as the ratio of deaths in this group over this time of four years to the total number of participants on the study in this group. Would you recommend that people who currently sleep longer than eight hours should get up earlier if they want to live longer? Explain!

Question 4: Assume that you are estimating the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + u_i$$

When you estimate the above model you find that the correlation of X and Z is 0.7.

1. What are the advantages of leaving both X and Z in the regression?
2. What are the disadvantages?
3. What strategies might help you in deciding whether to keep both X and Z or whether to drop one of them?

Question 5: You have estimated the following model to explain how many days students are absent from school:

$$\widehat{\text{absence}}_i = 5 - 0.5 \cdot \text{male}_i + 0.2 \cdot \text{cold}_i + 0.5 \cdot \text{far}_i + 0.2 \cdot \text{cold}_i \cdot \text{far}_i - 0.2 \cdot \text{age}_i$$

(0.1) (0.1) (0.5) (0.3) (0.15)

Standard deviations are shown in parentheses below the coefficients.

- `absence` denotes the number of school days a student is absent from school within a school year.
- `male`=1 if the student is a boy and 0 otherwise.

- `cold` denotes the number of days within a school year where the average temperature falls below -10°C .
- `far`=1 if the student lives more than 10 km away from school and 0 otherwise.
- `age` denotes the student's age in years.

1. How many days of absence from school do you estimate for a 10 year-old boy who does not live more than 10 km away from his school if it has been colder than -10°C for 4 days within a school year?
2. How many days of absence from school do you estimate for a 12 year-old girl who lives more than 10 km from her school if it has been colder than -10°C for 5 days within a school year?
3. What is the marginal effect of gender (i.e. of being a male student)? Does this marginal effect depend on the distance to the school or on the number of cold days?
4. How would the coefficients of your estimated model look like if you used a dummy variable `female`=1 if the student is a girl and 0 otherwise instead of the dummy variable `male`?
5. How can you interpret the coefficient of `cold · far`? Why do you think was it added to the model? Do you have an explanation for this effect?
6. Is the model you estimated above a linear model? Explain!
7. You realise that 70% of all students are actually never absent from school. Can this be a problem if you estimate the above equation with OLS?

8. Another person is also working with the same dataset. Due to a technical mishap his dataset contains each observation twice, i.e. there are now 200 observations in his dataset (two times the first observation, two times the second observation,...), not 100, as in yours. With this data he estimates the same regression using OLS.

- In which way are the coefficients of his regression different?
- In which way are the estimated standard errors different?
- In which way does the R^2 change?
- How are p -values affected?

Quantiles of the t -distribution (you find quantiles for different degrees of freedom in the different rows of the table)

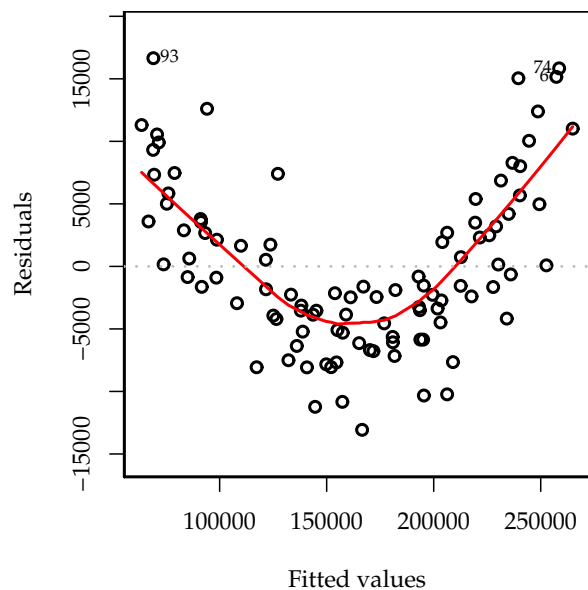
	0.9	0.95	0.975	0.99	0.995	0.999
50	1.299	1.676	2.009	2.403	2.678	3.261
90	1.291	1.662	1.987	2.368	2.632	3.183
91	1.291	1.662	1.986	2.368	2.631	3.182
92	1.291	1.662	1.986	2.368	2.630	3.181
93	1.291	1.661	1.986	2.367	2.630	3.180
94	1.291	1.661	1.986	2.367	2.629	3.179
95	1.291	1.661	1.985	2.366	2.629	3.178
96	1.290	1.661	1.985	2.366	2.628	3.177
97	1.290	1.661	1.985	2.365	2.627	3.176
98	1.290	1.661	1.984	2.365	2.627	3.175
99	1.290	1.660	1.984	2.365	2.626	3.175
100	1.290	1.660	1.984	2.364	2.626	3.174

2 Econometrics, 31. July, 2009

This exam was set for the Diplom in Econometrics at the Universität Jena

You have one hour. Please explain all your answers briefly.

Question 1: You estimate a linear regression and you obtain the following diagnostic plot:



```
> est <- lm(Y ~ X)
> plot(est, which = 1)
```

How do you interpret your results? Do you have to change anything with your regression?

Question 2: You compare the productivity of different economies. You assume that the production function can be described well with the help of a Cobb-Douglas function:

$$Y = \alpha K^\kappa L^\lambda M^\mu$$

Y is the GDP, K is the invested capital, L are the hours of labour, and M is the amount of invested raw material. α , κ , λ und μ are parameters of the production function.

1. Assume that you have data from 23 different countries, all from the year 1997. You want to estimate α , κ , λ and μ . What would you do?
2. A regression equation always has a noise term u . How and where does

this term appear in your estimated relationship?

3. You use the result of your estimation to predict the output of another economy. For this economy you know K , L , and M . If you substitute these values into your estimated equation, can you assume that your estimation is unbiased? Explain? If there is a bias, is it positive or negative?
4. Now you consider a different dataset for only one country but 23 successive years. Which assumption of the OLS model could be violated?

Question 3: The government wants to enhance the growth of firms with a support program. A first study with 12 firms shows the following result:

```
lm(growth ~ bigFirm * program)
```

```
z test of coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.2	0.6	-2.0	0.04550 .
bigFirm	5.2	0.8	6.5	0.00000 ***
program	4.5	0.75	6.0	0.00000 ***
bigFirm:program	-3.0	0.75	-4.0	0.00006 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R2=      0.97
```

bigFirm is a dummy which is 1 for firms with more than 100 employees and 0 otherwise. program is a dummy which is 1 for firms who are participants of the program. growth is the increase in turnover in percent.

1. How large is the estimated marginal effect of participating in the program for small firms with

- fewer than 100 employees?
- How large is the estimated marginal effect of participating in the program for larger firms?
 - Can you say anything about the average growth in turnover for large firms who participated in the program?
 - Can you say that the increase in growth for large firms is significantly different from zero? Explain your answer.
 - You want to calculate a confidence interval for the marginal effect with large firms. Your statistical software can calculate confidence intervals for estimated coefficients of a regression, but not confidence intervals for linear combinations of these coefficients. What regression equation do you use to estimate and how do you transform the estimation results to obtain the desired result?

Question 4: We still consider the problem of question 3. Now you use Huber's method of a robust regression.

Remember: all procedures that we considered minimise $\sum \rho(y_i - (\beta_0 + \beta_1 x_i))$. With OLS we have

$$\rho(x) = x^2.$$

With Huber's method we have

$$\rho(x) = \begin{cases} x^2/2 & \text{if } |x| \leq c \\ c|x| - c^2/2 & \text{otherwise} \end{cases}$$

c is an estimated value for σ_u .

You observe that the estimated coefficients of the robust regression are quite

different from the estimated coefficients of the OLS regression.

- What could explain these differences?
- What can you do to support your suspicion?
- Which result should you (under which conditions) prefer?
- Sometimes we obtain a better "fit" of the regression when variables enter the regression in a non-linear way (e.g. as polynomials, logarithms, etc.). Could this approach help here? Please explain.

Question 5: You carry out the above study with 12 firms from a different region. Your method is still OLS. You obtain the following result:

```
lm(growth ~ bigFirm * program)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.2	NA	NA	NA

bigFirm	3.0	NA	NA	NA
program	3.6	NA	NA	NA
bigFirm:program	-1.8	NA	NA	NA
R2=	0.81			

To your surprise you do not find any estimated standard errors in your output.

1. What is the likely reason?
2. You are still interested in the confidence interval of the marginal effect of participating in the support

program for large firms. This is difficult without proper standard errors. Can you be sure that without your data one can not determine a confidence interval? If you see (under certain conditions) a possibility, please explain what one could do?

Matrix algebra A matrix is a rectangular array of numbers $a_{ij} \in \mathbb{R} \forall i, j$

$$\mathbf{A} = (a_{ij}) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

This matrix is square for $m = n$.

1. Special matrices: identity matrix \mathbf{I} , zero matrix $\mathbf{0}$, diagonal matrix \mathbf{D} .

$$\mathbf{I} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}; \mathbf{0} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix};$$

$$\mathbf{D} = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{bmatrix}$$

2. Transpose of a matrix: $(a_{ij}) \rightarrow (a_{ji})$.

$$(\mathbf{A}_{m \times n})' = \mathbf{A}_{n \times m}$$

We have $\mathbf{A}'' = (\mathbf{A}')' = \mathbf{A}$; For a symmetric matrix $\mathbf{A}_{n \times n}$ we further have $\mathbf{A}' = \mathbf{A}$.

3. Addition/subtraction with $\mathbf{A} = \mathbf{A}_{m \times n}$, $\mathbf{B} = \mathbf{B}_{m \times n}$.

$$\mathbf{A} \pm \mathbf{B} \equiv \begin{bmatrix} a_{11} \pm b_{11} & \cdots & a_{1n} \pm b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & \cdots & a_{mn} \pm b_{mn} \end{bmatrix}$$

We have: $\mathbf{A} + \mathbf{0} = \mathbf{A}$; $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$; $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$; $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$

4. Scalar multiplication (α is a scalar)

$$\alpha \mathbf{A} = \begin{bmatrix} \alpha \cdot a_{11} & \cdots & \alpha \cdot a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha \cdot a_{m1} & \cdots & \alpha \cdot a_{mn} \end{bmatrix}.$$

$$\alpha \mathbf{A} = \mathbf{A} \alpha; (\alpha \mathbf{A})' = \alpha \mathbf{A}'; (\alpha + \beta) \mathbf{A} = \alpha \mathbf{A} + \beta \mathbf{A}$$

5. Multiplication.

$$\mathbf{A}_{n \times p} \cdot \mathbf{B}_{p \times m} = \mathbf{A} \mathbf{B} = (a_{ik}) \cdot (b_{kj}) = \left(\sum_{k=1}^p a_{ik} b_{kj} \right) = (c_{ij}) = \mathbf{C}_{n \times m}$$

\Leftrightarrow

$$\begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{p1} & \cdots & b_{pm} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^p a_{1k} b_{k1} & \cdots & \sum_{k=1}^p a_{1k} b_{km} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^p a_{nk} b_{k1} & \cdots & \sum_{k=1}^p a_{nk} b_{km} \end{bmatrix}$$

We have: $\mathbf{A} \mathbf{0} = \mathbf{0} \mathbf{A} = \mathbf{0}$; $\mathbf{A} \mathbf{I} = \mathbf{I} \mathbf{A} = \mathbf{A}$; $\alpha \mathbf{A} \mathbf{B} = \mathbf{A} (\alpha \mathbf{B}) = (\alpha \mathbf{A}) \mathbf{B}$; $\mathbf{A} \mathbf{B} \neq \mathbf{B} \mathbf{A}$; $\mathbf{A} \mathbf{B} + \mathbf{A} \mathbf{C} = \mathbf{A} (\mathbf{B} + \mathbf{C})$; $\mathbf{A} (\mathbf{B} \mathbf{C}) = (\mathbf{A} \mathbf{B}) \mathbf{C}$; $(\mathbf{A} \mathbf{B})' = \mathbf{B}' \mathbf{A}'$; $(\mathbf{A} \mathbf{B} \mathbf{C})' = \mathbf{C}' \mathbf{B}' \mathbf{A}'$

6. Taking derivatives.

$$\frac{\partial (\mathbf{A} \mathbf{x})}{\partial \mathbf{x}} = \mathbf{A}' \quad \left| \quad \frac{\partial (\mathbf{A} \mathbf{x})}{\partial \mathbf{x}'} = \mathbf{A} \right.$$

$$\frac{\partial (\mathbf{x}' \mathbf{A} \mathbf{x})}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}') \mathbf{x} \quad \left| \quad \frac{\partial^2 (\mathbf{x}' \mathbf{A} \mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} = \mathbf{A} + \mathbf{A}' \right.$$

7. Properties of specific matrices.

Symmetry: If \mathbf{A} is symmetric, i.e. $\mathbf{A} = \mathbf{A}_{n \times n}$ and $\mathbf{A}' = \mathbf{A}$, then also $\mathbf{A} \pm \mathbf{A}'$ and $\alpha \mathbf{A} + \beta \mathbf{A}$ are symmetric.

If A , B , C are symmetric, then also
 $A'BA$ are ACA' symmetric.

Inverse A^{-1} : $A^{-1}A = AA^{-1} = I$ mit

$$(A')^{-1} = (A^{-1})'$$

Orthogonal matrix: $A'A = AA' = I$ for
 $A = A_{n \times n}$

3 Econometrics, 28. July 2007

This exam was set for the Diplom in Econometrics at the Universität Jena

Please explain all your answers. You have one hour.

1. Please define or explain briefly:
 - a) Ordinary Least Squares
 - b) In which direction (+ or -), is the coefficient of the variable *experience* biased if you drop the the variable *age* in a regression equation which explain the wage of different workers.
 - c) What is a dummy variable?
 - d) What is an interaction term?
 - e) *t*-statistic
2. Statistical analysis:
 - a) What is "internal validity" of statistical analysis?
 - b) What problems regarding internal validity could one have?
 - c) What can be the consequences of these problems?
 - d) What possibilities do you see to avoid these problems?
 - e) What is external validity of statistical analysis?
3. You remember that in the simple regression model we estimated the coefficient β_1 as follows (variables have their usual interpretation):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Assume that we are studying the impact of different amounts of food on the weight of chicken. Be X the amount of food and Y the weight of a bird. There is a small problem in our dataset. We no longer have the data for individual birds, but only the data for pairs. I.e. we only know the the total amount of food and the sum of the weight for every two birds.

We again use the above formular to determine the marginal effect.

- a) What bias do we observe in our regression?
 - b) What can we say about the variance of this estimator?
 - c) Can you derive a better estimator?
4. You want to find out whether sex or nationality have an impact on the performance of students in a test on knowledge about economics.

In your regression model the test score S_i is explained as a function of a standardised IQ test I_i (with mean 100 and standard deviation 15), the average education of the parents (in years) A_i , a dummy variable G_i for sex (1 if female, 0 otherwise), and a dummy for nationality (1 if German, 0 otherwise).

Unfortunately you no longer have access to the raw data. You only see the results of the regression (standard errors in parentheses):

$$\hat{S}_i = 5.7 - 0.63 \cdot X_{1i} - 0.22 \cdot X_{2i} + 0.16 \cdot X_{3i} + 1.2 \cdot X_{4i}$$

$$\begin{array}{cccc} (0.63) & (0.88) & (0.08) & (0.1) \end{array}$$

The regression results are based on 24 observations, $R^2 = 0.54$.

- a) Find out which coefficient belongs to which variable. Explain your answers.
 - b) Based on your answers to question 4a determine reasonable hypotheses for your variables. Test your hypotheses on a significance level of 5%.
 - i. What test statistics do you calculate?
 - ii. Which distribution functions do you use?

If different assumptions would lead to different distribution functions, then explain in which case you would use which distribution function.
 - c) What is, in your opinion, the effect of sex and nationality on the test score S_i in this sample?
5. You are using R to analyse a hypothetical dataset.

Variables A , B , C , D , E and $year$ can be found in the first line of the file `file.csv`. This file contains 553 observations as `.csv` (comma separated values). Explain what the following commands do and explain which command you would choose.

- a) First you have to read the dataset.
 - i. `daten <- read.csv("file.csv",header=YES,sep=';')`
 - ii. `daten <- read.csv("file.csv",header=TRUE,sep=';')`
 - iii. `daten <- read.table("file.csv")`

- iv. `daten <- read.table("file.csv",header=YES,sep=',')`
- b) For a part of your analysis you only need observations from the year 1988. What do you have to do?
- i. `daten-1988 <- daten[if year==1988,]`
 - ii. `daten1988 <- daten[year==1988,]`
 - iii. `daten-1988 <- daten[daten$year==1988,]`
 - iv. `daten-1988 <- daten[year==1988,]`
- c) You start with an OLS regression where C is the dependent variable and A, B, D and E are explanatory variables.
- i. `result <- ols(C ~ A + B + D + E,data=daten)`
`summary(result)`
 - ii. `result <- ols(C ~ A + B + D + E,data=daten)`
 - iii. `foo <- lm(C ~ A + B + D + E,data=daten)`
`summary(foo)`
 - iv. `result <- reg(C ~ A + B + D + E,data=daten)`
`summary(result)`
- d) Furthermore, you want to know the correlation coefficients between B, C and D . How do you find them?
- i. `corr(B, C, D)`
 - ii. `corr(daten)`
 - iii. `cor(daten)`
 - iv. `cor(B, C, D)`

4 Ökonometrie, 21. 7. 2007

Diese Klausur war Teil der Diplomprüfung für Ökonometrie an der Universität Jena.

1. Geben Sie kurze Erklärungen bzw. Antworten für:
 - a) Signifikanzniveau
 - b) Die Elastizität von Y nach X in der Schätzgleichung:

$$\ln Y_i = \beta_0 + \beta_1 \cdot \ln X_i + \epsilon$$

- c) Die Richtung (+ oder -), in die der Koeffizient der Variablen Alter verzerrt wird, wenn man die Variable Erfahrung in einer Schätzgleichung, die das Gehalt verschiedener Arbeiter erklärt, weglässt.
- d) p -Wert
2. Multikollinearität ist eines der möglichen Probleme, das bei einer Schätzung auftreten kann.
- Beschreiben Sie, was man unter Multikollinearität versteht.
 - Was sind die Konsequenzen von Multikollinearität?
 - Was für Möglichkeiten gibt es, Multikollinearität festzustellen?
 - Was für Möglichkeiten gibt es, bestehende Multikollinearität zu vermeiden?
3. Das Produkt Z Ihrer Firma wurde in den letzten Jahren über Werbespots in zwei verschiedenen Fernsehsendern beworben: EURO1 und SAT5. Die Preise für Spots sind bei beiden Sendern gleich. Eine Datenstudie der letzten verfügbaren Perioden ergab die folgende Regression (Standardfehler in Klammern):

$$\hat{Y} = 300 + 10 \cdot X_1 + 20 \cdot X_2$$

(1,0) (2,5)

Ihnen stehen 44 Beobachtungen zur Verfügung, $R^2 = 0,9$. Dabei bezeichnen Y den Umsatz des Produkts Z (in Tausend €), X_1 die Werbeausgaben für Spots beim Sender EURO1 (in Tausend €), X_2 die Werbeausgaben für Spots beim Sender SAT5 (in Tausend €).

Welches Werbeinstrument sollten Sie gemäß den Regressionsdaten bevorzugt verwenden (gegeben alle anderen Faktoren sind konstant!)? Begründen Sie Ihre Antwort.

4. An Ihrer Universität wird die Anzahl der jährlichen Bewerbungen diskutiert. Sie werden um Rat gefragt, welche Einflussfaktoren hier eine Rolle spielen. Mit dem verfügbaren Datensatz wurden folgende Regressionsergebnisse geschätzt (Standardfehler in Klammern):

$$\hat{N}_t = 150 + 180 \cdot A_t + 1,5 \cdot \ln T_t + 30 \cdot P_t$$

(90) (1,5) (60)

$R^2 = 0,5$, $N = 22$. Es bezeichnet N_t die Anzahl der Studenten, die sich bewerben im Jahr t , A_t die Anzahl der Unimitarbeiter die Informationsveranstaltungen der Schulen besuchen, um das Programm der Uni vorzustellen, T_t die jährlichen Studiengebühren im Jahr t , P_t den Prozentsatz der unterrichtenden Fakultätsmitglieder die im Jahr t einen Dokortitel haben.

- a) Welche Vorzeichen würden Sie für die einzelnen Koeffizienten erwarten (mit Begründung)?
- b) Werden diese Erwartungen bestätigt durch die Regressionsergebnisse? Stellen Sie passende Nullhypothesen auf, berechnen Sie die t -Statistiken der Koeffizienten und wählen Sie ein passendes Signifikanzniveau.
- c) Wie ergibt sich der p -Wert aus der obigen t -Statistik. Welche Verteilungsfunktion müssen Sie verwenden?
- d) Welche Gründe könnte es geben, dass die geschätzten Vorzeichen nicht den erwarteten entsprechen?
- e) Diskutieren Sie, ob die funktionale Form für T_t (linear-log) angebracht ist.
- f) Machen Sie Vorschläge, wie die Spezifikation der Schätzgleichung verbessert werden könnte.

5.

- a) Beschreiben Sie die Bedeutung von Nicht-Linearität in den Variablen sowie Nicht-Linearität in den Parametern für die Schätzung mit OLS.
- b) Betrachten Sie folgende Gleichungen und bestimmen Sie, ob diese linear in den Variablen, linear in den Parametern, beides oder keins von beiden sind:

$$Y_i = \beta_0 + \beta_1 \cdot X_i^3 + \epsilon \quad (1)$$

$$Y_i = \beta_0 + \beta_1 \cdot \ln X_i + \epsilon \quad (2)$$

$$\ln Y_i = \beta_0 + \beta_1 \cdot \ln X_i + \epsilon \quad (3)$$

$$Y_i = \beta_0 + \beta_1 \cdot X_i^{\beta_2} + \epsilon \quad (4)$$

$$Y_i^{\beta_0} = \beta_1 + \beta_2 \cdot X_i^2 + \epsilon \quad (5)$$

5 Ökonometrie, 25. 7. 2008

Diese Klausur war Teil der Diplomprüfung für Ökonometrie an der Universität Jena. Bitte begründen Sie alle Ihre Antworten. Bearbeiten Sie die Klausur bitte in einer Stunde und ohne Hilfsmittel.

Viel Erfolg!

1. Sie planen, einen Imbissstand in Jena zu eröffnen. Sie wissen allerdings weder, wo Sie den Stand eröffnen wollen, noch welches Produkt Sie verkaufen wollen. Eine erste Studie mit einem mobilen Stand und wechselnden Produkten ergibt den folgenden durchschnittlichen Umsatz (jeweils pro Tag):

	Weißwurst	Bratwurst
Jena-West	3000 €	4000 €
Jena-Ost	2500 €	4500 €

Auf Basis des gleichen Datensatzes schätzen Sie auch eine Regression

$$Y = \beta_0 + \beta_1 \cdot d_B + \beta_2 \cdot d_O + \beta_3 \cdot d_B \cdot d_O + u$$

Dabei ist Y der Umsatz pro Tag, d_B ein Dummy der den Wert Eins annimmt, falls an diesem Stand gerade Bratwurst verkauft wird und sonst Null ist, und d_O ein Dummy, der den Wert Eins annimmt, falls sich der Stand in Jena-Ost befindet und sonst Null ist.

Welche Werte werden Sie für β_0 , β_1 , β_2 , und β_3 schätzen?

2. Sie führen eine weitere Untersuchung nur in Jena-West durch, diesmal untersuchen Sie allerdings drei verschiedene Speisenangebote: Weißwurst, Bratwurst, und Currywurst. Dazu führen Sie drei Dummyvariablen ein: d_W ist nur Eins am Weißwurststand, d_B ist nur Eins am Bratwurststand, und d_C ist nur Eins am Currywurststand. Ansonsten haben die Dummies den Wert Null. Sie schätzen die folgende Gleichung:

$$Y = \beta_0 + \beta_1 \cdot d_W + \beta_2 \cdot d_B + \beta_3 \cdot d_C + u$$

- a) Welches Problem tritt auf?
 - b) Was könnte man besser machen?
3. Nehmen Sie an, die drei Standardannahmen der einfachen OLS Regression seien erfüllt.
 - a) Unter welchen Voraussetzungen hat OLS die kleinste Varianz unter allen erwartungstreuen Schätzern, die linear in Y sind?
 - b) Unter welchen Voraussetzungen hat OLS die kleinste Varianz unter allen konsistenten Schätzern?
 - c) Unter welchen Voraussetzungen ist der OLS Schätzer erwartungstreu?
 - d) Unter welchen Voraussetzungen ist der OLS Schätzer konsistent?
 4. Sie schätzen die Gleichung $Y = \beta_0 + \beta_1 X + u$ und erhalten in R den folgenden Output:

```
lm(formula = Y ~ X)

Residuals:
    Min       1Q   Median       3Q      Max
-2.5734 -1.8240 -0.3968  1.4920  3.1480

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.0135     0.6539   9.196 3.66e-12 ***
X            2.7279     1.0787   2.529  0.0148 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.80 on 48 degrees of freedom
Multiple R-squared:  0.10,    Adjusted R-squared:  0.09919
F-statistic: 6.395 on 1 and 48 DF,  p-value: 0.01478
```

Sie gehen davon aus, dass die die Fehlerterme u homoskedastisch sind.

- Wie groß ist die Varianz der abhängigen Variablen Y in Ihrer Stichprobe?
 - Ihr Signifikanzniveau ist 1%. Ist der Koeffizient β_1 signifikant von Null verschieden?
 - Welche Berechnung müssen Sie durchführen, um ein 95% Konfidenzintervall zu berechnen? Geben Sie ein R-Kommando an, und verwenden Sie die Zahlen aus dem Output der Regression sowie den Befehl `qnorm()` der Quantile der Normalverteilung berechnet.
 - Was sagt Ihnen der p -Wert in der letzten Zeile des Regressionsoutputs?
5. In den folgenden Situationen sei der wahre Zusammenhang jeweils durch die Gleichung

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + u$$

gegeben. Sie schätzen jedoch

$$Y = \beta_0 + \beta_1 \cdot X_1 + u.$$

Betrachten Sie die folgenden Situationen. Welches Vorzeichen erwarten Sie für $\hat{\beta}_1$. Welches Vorzeichen hat die Differenz $E(\hat{\beta}_1) - \beta_1$. Von welchen Annahmen hängen Ihre Antworten jeweils ab?

- a) {
- Y_1 = Nachfrage nach Marmelade
 - X_1 = Preis für Marmelade
 - X_2 = verfügbares Einkommen
 - Datenbasis = Jährliche Nachfrage in den USA von 1952–2002. Preise und Einkommen sind nicht inflationsbereinigt.

- b) $\left\{ \begin{array}{l} Y_1 = \text{Produktionsmenge von Autos} \\ X_1 = \text{Arbeitseinsatz} \\ X_2 = \text{Kapital} \\ \text{Datenbasis} = 20 \text{ gr\u00f6\u00dft} \text{e KFZ Firmen in Europa 2008} \end{array} \right.$
- c) $\left\{ \begin{array}{l} Y_1 = \text{Einkommen von Arbeitern} \\ X_1 = \text{Arbeitserfahrung} \\ X_2 = \text{Alter des Arbeiters} \\ \text{Datenbasis} = \text{Alle m\u00e4nnlichen Arbeiter einer Marmeladen-} \\ \text{fabrik in Seattle, 1998} \end{array} \right.$
- d) $\left\{ \begin{array}{l} Y_1 = \text{Anzahl Zuschauer einer Open-Air Filmvorf\u00fchrung} \\ X_1 = \text{Wochenende-Dummy (= 1 falls Wochenende)} \\ X_2 = \text{Niederschlag in l/m}^2 \\ \text{Datenbasis} = \text{Alle Veranstaltungen eines Open-Air Kinos in} \\ \text{Duisburg im Sommer 2005} \end{array} \right.$