LuTK, OyKKK, KaTK, TTK, TST ja BMTK (Linnanmaan tentit)
SCI, OBS, OMS, TECH, ITEE and BMM (Linnanmaa campus)

UNIVERSITY OF OULU

YLIOPISTOTENTTI - UNIVERSITY EXAM

Opiskelijan nimi / Student name:	**Opiskelijanumero / Student number:**

Opettaja täyttää / Lecturer fills in:

Opintojakson koodi and nimi / The code and the name of the course:

Koodi / Code 721066S

Tentin nimi / Exam name Principles of Econometrics

Tiedekunta / Faculty: Oulun yliopiston kauppakorkeakoulu / Oulu Business School

Tentin pvm / Date of exam: 5.2.2018	**Tentin kesto tunteina / Exam in hours: 3 h**
Tentaattori(t) / Examiner(s):	**Opintopistemäärä / Credit units: 6**
Sanna Huikari	**Sisäinen postios. / Internal address:** **6 OyKKK**

Sallitut apuvälineet / The devices allowed in the exam:

☒ Funktiolaskin / Scientific calculator

☒ Ohjelmoitava laskin / Programmable calculator

☒ Muu materiaali, tarkennettu alla / Other material, specified below: **Two-sided hand-written A4-sheet, which, NB!, must be included into the answer sheet**

Tenttiin vastaaminen / Please answer the questions:

☒ Suomeksi / in Finnish ☒ Englanniksi / in English

Suomenkielisessä tutkinto-ohjelmassa olevalla opiskelijalla on oikeus käyttää arvioitavassa opintosuorituksessa suomen kieltä, vaikka opintojakson opetuskieli olisi englanti. Tämä ei koske vieraan kielen opintoja. (Kts. Koulutuksen johtosääntö 18 §)

In a Finnish degree programme a student has a right to use Finnish language for their study attainment, even though the language of instruction is English, (excluding language studies) even when the language of instruction is other than Finnish. (See the Education Regulations 18 §)

Kysymyspaperi on palautettava / Paper with exam questions must be returned:

☐ Kyllä / Yes ☒ Ei / No

1. (31 p.) Diekmann et al. (2008)[1] studied used tractor prices. They modelled the price of a used tractor as a function of the attributes of that tractor and the time of the sale. Table 1 contains the definitions of the variables used in replicating Diekmann's regressions. The dependent variable is the natural logarithm of the price of a used farm tractor that was sold at auction in the United States between June 1, 2011 and May 31, 2012. The number of observations is 276. Table 1 shows has the mean or median values of each variable in the dataset. Table 2 contains results from six estimated variations of Diekmann's model using the dataset. All estimations were made by using the heteroskedasticity-robust standard errors. The values of standard errors are presented in the table in parentheses below the values of the estimates. Statistical tables can be found on page 5.

Table 1. Variable definitions and their mean/median values in the dataset of the used tractor price model (N=276).

Variable	Description	Mean/Median
$lnsaleprice_i$	natural logarithm of the price paid for the tractor i	$\bar{x} = 9.5$
$saleprice_i$	price paid for the tractor i in dollars	$\bar{x} = 20\,747$
Tractor Specifications:		
$horsepower_i$	horsepower of the tractor i engine	$\bar{x} = 101$
age_i	number of years since tractor i was manufactured	$\bar{x} = 16$
$enghours_i$	number of hours of use recorded on tractor i	$\bar{x} = 3530$
$diesel_i$	a dummy variable = 1 if tractor i runs on diesel fuel, 0 otherwise	Md = 1
fwd_i	a dummy variable = 1 if tractor i has four-wheel drive, 0 otherwise	Md = 1
$manual_i$	a dummy variable = 1 if tractor i transmission is manual, 0 otherwise	Md = 1
$johndeere_i$	a dummy variable = 1 if tractor i is manufactured by John Deere, 0 otherwise	Md = 0
cab_i	a dummy variable = 1 if tractor i has an enclosed cab, 0 otherwise	Md = 1
Time of year:		
$spring_i$	a dummy variable = 1 if tractor i sold in April or May, 0 otherwise	Md = 0
$summer_i$	a dummy variable = 1 if tractor i sold June-September, 0 otherwise	Md = 0
$winter_i$	a dummy variable = 1 if tractor i sold December-March, 0 otherwise	Md = 0

[1] Diekmann, F., Roe, B.E., & Batte, M.T. (2008). Tractors on eBay: Differences between internet and in-person auctions. *American Journal of Agricultural Economics* 90(2), 306-320.

Answer the following questions based on the results in Table 2. Remember to justify your conclusion / explain your thinking! No credit for answers if you do not justify your conclusion / explain your thinking!

a) State the precise meaning of the coefficient on *age* in regression (1). Use both hypothesis testing and confidence intervals in determining whether the estimated coefficient is statistically significant at the 5% significance level. (3 p)

b) Carefully interpret the coefficient of *johndeere* in regression (1). What does it mean in real-world terms? (2 p)

c) Do the variables included in the regression (1) account a large fraction of the variance in the natural logarithm of the price of a used farm tractor? (2 p)

d) Does the time of year of the sale have an effect on the price of the tractors? (2 p)

e) Regression (2): Is there evidence that *horsepower* has a nonlinear effect on *lnsaleprice*? What do the signs of estimated coefficients of *horsepower* and *horsepower*^{2} (= horsepower raised to the power of two) suggest regarding the relationship between horsepower and the price of a used tractor? (3 p)

f) Regression (2): If horsepower increases from 50 to 60, how is the price of a used tractor expected to change? If horsepower increases from 100 to 110, how is the price of a used tractor expected to change? (3 p)

g) Which equation do you prefer between regression (1) and regression (2)? Why? (2 p)

h) When comparing the results of regression (2) and regression (3), does the regression (2) seem to suffer from omitted variable bias? (2 p)

i) What does the regression (3) predict for the value of *lnsaleprice*, when the values of all the variables are at their means or medians? (2 p)

j) Is there any evidence that the effect of additional hour of use affects the value of a diesel tractor differently than for a gasoline tractor? (2 p)

k) Regression (5): State the precise meaning of the coefficient on *lnage* (= natural logarithm of number of years since tractor was manufactured). (2 p)

l) Regression (5): If *age* increases from 10 to 11, how is the price of a used tractor expected to change? If *age* increases from 20 to 21, how is the price of a used tractor expected to change? (3 p)

m) Regression (6): How is the coefficient *lnage* changed compared to the regression (5)? Why? Briefly explain the reasons behind the magnitude and direction of the change. (3 p)

Table 2. Regression results.

Explanatory variables	Dependent variable: *lnsaleprice*					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
horsepower	0.007 (0.001)	0.015 (0.001)	0.011 (0.001)	0.007 (0.001)	0.010 (0.001)	0.010 (0.001)
horsepower^2	–	-0.00002 (2.2×10^{-6})	-0.00001 (2.3×10^{-6})	–	-0.00001 (2.3×10^{-6})	-0.00001 (2.3×10^{-6})
age	-0.028 (0.004)	-0.032 (0.004)	-0.032 (0.004)	-0.028 (0.004)	–	–
lnage	–	–	–	–	-0.389 (0.048)	-0.485 (0.041)
enghours	-0.00002 (0.00001)	-0.00003 (0.00001)	-0.00004 (0.00001)	-0.00003 (0.00003)	-0.00004 (0.00001)	–
diesel	0.499 (0.120)	0.279 (0.106)	0.215 (0.099)	0.467 (0.148)	0.212 (0.101)	0.212 (0.103)
diesel × *enghours*	–	–	–	0.00001 (0.00003)	–	–
fwd	0.357 (0.073)	0.331 (0.063)	0.275 (0.059)	0.358 (0.073)	0.337 (0.058)	0.338 (0.059)
manual	-0.122 (0.077)	-0.124 (0.067)	-0.153 (0.062)	-0.123 (0.078)	-0.170 (0.063)	-0.174 (0.065)
johndeere	0.173 (0.089)	0.243 (0.078)	0.310 (0.072)	0.180 (0.092)	0.285 (0.074)	0.266 (0.075)
cab	–	–	0.478 (0.070)	–	0.490 (0.072)	0.469 (0.073)
spring	-0.032 (0.081)	-0.049 (0.070)	-0.049 (0.065)	-0.030 (0.081)	-0.054 (0.066)	-0.057 (0.068)
summer	-0.119 (0.079)	-0.084 (0.069)	-0.057 (0.064)	-0.117 (0.080)	-0.046 (0.065)	-0.047 (0.067)
winter	0.040 (0.089)	0.052 (0.077)	0.046 (0.071)	0.039 (0.089)	0.029 (0.073)	0.038 (0.075)
Intercept	8.77 (0.14)	8.59 (0.12)	8.73 (0.11)	8.80 (0.16)	9.22 (0.16)	9.41 (0.15)
F-statistic and p-values on joint hypotheses:						
spring, summer, and winter	1.09 (0.35)	–	–	–	–	–
Measures of fit						
R^2	0.70	0.77	0.81	0.70	0.80	0.79
$\overline{R}^2$	0.69	0.76	0.80	0.68	0.79	0.78

Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients, and *p*-values are given in parentheses under *F*-statistics. N=276 in all regressions.

2. (9 p.) Internal and external validity. Question 2 consists of 6 multiple choice questions. Each question has only one correct (or clearly best) answer. You get +1.5 points for a right answer, zero points for no answer, and -0.75 point for a wrong answer. Maximum points from this question are therefore 9 points, minimum is 0. Write the letters that indicate your answers in answering sheet.

1) The reliability of a study using multiple regression analysis depends on all of the following with the exception of
   A) omitted variable bias.
   B) errors-in-variables.
   C) presence of homoskedasticity in the error term.
   D) external validity.

2) Threats to internal validity lead to
   A) perfect multicollinearity
   B) the inability to transfer data sets into your statistical package
   C) failures of one or more of the least squares assumptions
   D) a false generalization to the population of interest

3) The analysis is externally valid if
   A) the statistical inferences about causal effects are valid for the population being studied.
   B) the study has passed a double blind refereeing process for a journal.
   C) its inferences and conclusions can be generalized from the population and setting studied to other populations and settings.
   D) some committee outside the author's department has validated the findings.

4) Sample selection bias
   A) occurs when a selection process influences the availability of data and that process is related to the dependent variable.
   B) is only important for finite sample results.
   C) results in the OLS estimator being biased, although it is still consistent.
   D) is more important for nonlinear least squares estimation than for OLS.

5) Errors-in-variables bias
   A) is only a problem in small samples.
   B) arises from error in the measurement of the independent variable.
   C) becomes larger as the variance in the explanatory variable increases relative to the error variance.
   D) is particularly severe when the source is an error in the measurement of the dependent variable.

6) Simultaneous causality bias
   A) is also called sample selection bias.
   B) happens in complicated systems of equations called block recursive systems.
   C) results in biased estimators if there is heteroskedasticity in the error term.
   D) arises in a regression of Y on X when, in addition to the causal link of interest from X to Y, there is a causal link from Y to X.

Statistical tables:

Large-Sample Critical Values for the t-statistic from the Standard Normal Distibution	Significance Level		
	10%	5%	1%
2-Sided Test (≠)			
Reject if t is greater than	1.64	1.96	2.58
1-Sided Test (>)			
Reject if t is greater than	1.28	1.64	2.33
1-Sided Test (<)			
Reject if t is less than	−1.28	−1.64	−2.33

TABLE 3 Critical Values for the χ^2 Distribution

Degrees of Freedom	Significance Level		
	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48

Large-Sample Critical Values for the F-statistic from the $F_{m,\infty}$ Distribution

Reject if $F >$ Critical Value

Degrees of Freedom (m)	Significance Level		
	10%	5%	1%
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64