STOCKHOLM UNIVERSITY
Department of Statistics
Econometrics I, Regression analysis, ST223G
Autumn semester 2020

Written Examination in Econometrics I

Date	2020-12-02
Hour:	9.00-15.00
Examiner:	Jörgen Säve-Söderbergh
Allowed tools:	1) Textbook: Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*, Cengage.
	2) Pocket calculator
	3) Notes written in the text book are allowed.

- Note that no formula sheet is provided.
- Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description.
- The maximum number of points for each problem is stated after each question. If not indicated otherwise, to obtain the maximum number of points on each problem, detailed and clear solutions are required. Answers may be given in English or Swedish.

For questions about the content of the exam, contact the course coordinator on jorgen.save-soderbergh@stat.su.se. Incoming e-mail questions are answered continuously during the exam.

If the course coordinator needs to send out information to all students during the exam, this is done to your registered email address. Therefore, check your email during the exam.

Please note that practical help is only available during the first hour of the exam by email expedition@stat.su.se. Carefully read the enclosed instructions for exam submission. There you find all the necessary information about submission, anonymous code, etc. If you, despite the instructions have problems submitting the exam, email the exam to tenta@stat.su.se. However, this is only done in exceptional cases.

Good luck!

The Data Material

Data on air pollution in 41 U.S. cities are given below. There is a single dependent variable, so2, the annual mean concentration of sulphur dioxide (svaveldioxid) and six explanatory variables, two which concern human ecology and four climate.

so2 Annual mean concentration of sulphur dioxide in micrograms per cubic meter (sulphur dioxide *sv. svaveldioxid*)

temp Average annual temperature in Fahrenheit

manuf Number of manufacturing enterprises employing 20 or more workers

pop Population size (1970 census) in thousands

wind Average annual wind speed in miles per hour

precip Average annual precipitation in inches (precipitation *sv. nederbörd*)

days Average number of days with precipitation

north Indicator for low average temperature (=1 if $temp < 54.8$; =0 if $temp > 54.8$)

1. The following equations were estimated using the above data. The first equation is for cities with an average temperature less than 54.8 Fahrenheit (presumably northern cities) and the second for cities with an average temperature above 54.8

$$\widehat{\text{so2}} = \beta_0 + \beta_1\text{temp} + \beta_2\text{manuf} + \beta_3\text{pop} + \beta_4\text{wind} + \beta_5\text{precip}$$

where $n = 20, R^2 = 0.4792$ and $SSR = 5014.321$.

$$\widehat{\text{so2}} = \beta_0 + \beta_1\text{temp} + \beta_2\text{manuf} + \beta_3\text{pop} + \beta_4\text{wind} + \beta_5\text{precip}$$

where $n = 20, R^2 = 0.6335$ and $SSR = 1289.055$.

The third equation includes all cities regardless of their average temperature.

$$\widehat{\text{so2}} = \beta_0 + \beta_1\text{temp} + \beta_2\text{manuf} + \beta_3\text{pop} + \beta_4\text{wind} + \beta_5\text{precip} + \beta_6\text{north}$$

where $n = 40, R^2 = 0.5283$ and $SSR = 7304.498$.

Compute the usual Chow test for testing the null hypothesis that the regression equations are the same for cities with an average temperature below 54.8 and for cities with an average temperature above 54.8. (30 p)

2. A binary response model is given by

$$P\left(y = 1 \,|\, x_1, x_2, \ldots, x_n\right) = G\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\right),$$

see equation[17.2] on page 560. This question concerns the interpretation of the coefficients β_j.

(a) Calculate the partial derivative of $P\left(y = 1 \,|\, x_1, x_2, \ldots, x_n\right)$ with respect to x_j, where we assume that x_j is continuous. Analyze the equation that you get; what sign will the derivative assume if $\beta_j > 0$. (10 p)

(b) Let us define a dummy variable for the data material above based on the variable *so*2. Define

$$\text{LowSo2} = \begin{cases} 1 & \text{if } so2 < 25 \\ 0 & \textit{otherwise} \end{cases}$$

that is, we consider cities with a low so2. Next, we estimate the logistic regression model

$$P(y = 1 \mid \cdot) = G(\beta_0 + \beta_1 temp + \beta_2 manuf + \beta_3 pop + \beta_4 wind + \beta_5 precip).$$

This resulted in

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.8368524  0.8761629  -3.238 0.002689 **
temp         0.0482298  0.0120722   3.995 0.000328 ***
manuf       -0.0002049  0.0005101  -0.402 0.690476
pop         -0.0001186  0.0004366  -0.272 0.787544
wind         0.1331483  0.0513941   2.591 0.014006 *
precip      -0.0127349  0.0062563  -2.036 0.049651 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Which variables seem to affect the probability of having low so2? (10 p)

(c) The average partial effects for the low so2 model was also estimated using R

```
Marginal Effects:
             dF/dx   Std. Err.        z   P>|z|
temp    0.05119661  0.02806054  1.8245 0.06808 .
manuf  -0.00014748  0.00045185 -0.3264 0.74412
pop    -0.00012788  0.00039549 -0.3234 0.74643
wind    0.13195843  0.08217819  1.6058 0.10833
precip -0.01387561  0.00960666 -1.4444 0.14863
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Interpret the average partial effects of temp, wind and precip. What general conclusion seems our result suggest? (10 p)

3. Assume that

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

is a model that you are interested in estimating. A professor suggested that you should test the null hypothesis $H_0 : \beta_1 = \beta_2$. Show the professor how this can be done by rewriting the model so that an ordinary t-test can be utilized. (20 p)

4. Consider the simple linear regression $y = \beta_0 + \beta_1 x + u$, where the assumptions SLR.1 - SLR.5 are assumed to be true. Let the OLS estimator of β_1 be denoted (as usual) by $\hat{\beta}_1$. Show that

$$\mathrm{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

(20 p)