STOCKHOLM UNIVERSITY
Department of Statistics
Johan Koskinen

WRITTEN RE-EXAMINATION, ECONOMETRICS I
2024-02-29

Time for examination: 14.00-19.00

Allowed tools: Pocket calculator, own formula sheet (1 double-sided A4 page), Course text-book: Wooldridge, J.M. *Introductory Econometrics - a Modern Approach (any edition)*
Note that no formula sheet will be provided.

The exam consists of 4 independent problems. Well motivated and clear solutions are required for full scoring on a problem. Don't forget to state any necessary assumptions or conditions where needed.
Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description. Answers may be given in English or Swedish.

Good luck!

Problem 1. (35 points)
The central fish market can be a hectic place, where restauranteurs and wholesalers bid for the best fresh fish. Other than species and supply, the price of fish is largely guided by weight. From a fish market we have data on $n = 158$ observations on fish, for which we have the variables in Table 1

Table 1: Summary of numerical variables

	Mean	SD	Description
Weight	400.847	357.698	Weight in gram
Length	26.293	10.011	Length in cm
Height	8.987	4.295	Height in cm
Width	4.424	1.689	Width in cm

In addition we have the species of each fish as per Table 2

Table 2: Summary of Species

j	1	2	3	4	5	6	7
Species	Bream	Parkki	Perch	Pike	Roach	Smelt	Whitefish
Number	35	11	56	17	19	14	6

For this data we first assume a model that we call Model one:

$$\log(Weight) = \beta_0 + \beta_1 Length + \beta_2 Height + \beta_3 Width + u$$

Furthermore, another model, Model two, is investigated

$$\begin{aligned}\log(Weight) &= \beta_0 + \beta_1 Length + \beta_2 Height + \beta_3 Width \\ &+ \sum_{j=2}^{7} \gamma_j D_j + u\end{aligned}$$

Where the dummies

$$D_j = \begin{cases} 1, & \text{if fish belongs to Species } j \\ 0, & \text{else} \end{cases}.$$

for $j = 1, \ldots, 7$. The following results were obtained from `R` (some output is hidden):

```
Call:
lm(formula = log(Weight) ~ Length + Height + Width, data = Fish)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1961 -0.1857  0.1237  0.2763  0.5771

Coefficients:
            Estimate Std. Error
```

```
(Intercept) 1.978759   0.085186
Length      0.046696   0.006014
Height      0.079145   0.011464
Width       0.337274   0.045647
---

Residual standard error: 0.3684 on 154 degrees of freedom
Multiple R-squared:  0.9243,Adjusted R-squared:  0.9228
F-statistic:   627 on 3 and 154 DF,  p-value: < 2.2e-16
```

A researcher confirms that the difference in size of fish depend on their species by running a regression of Height

```
Call:
lm(formula = Height ~ Species, data = Fish)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7499 -1.4647 -0.2457  1.2040  4.9383

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       15.1832     0.3654  41.547  < 2e-16 ***
SpeciesParkki     -6.2208     0.7473  -8.324 4.71e-14 ***
SpeciesPerch      -7.3213     0.4658 -15.716  < 2e-16 ***
SpeciesPike       -7.4694     0.6391 -11.687  < 2e-16 ***
SpeciesRoach      -8.4769     0.6161 -13.759  < 2e-16 ***
SpeciesSmelt     -12.9738     0.6837 -18.976  < 2e-16 ***
SpeciesWhitefish  -5.1560     0.9553  -5.397 2.57e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.162 on 151 degrees of freedom
Multiple R-squared:  0.7563,Adjusted R-squared:  0.7466
F-statistic: 78.11 on 6 and 151 DF,  p-value: < 2.2e-16
```

A model with dummy variables for species is subsequently estimated, yielding the following results from `R`

```
Call:
lm(formula = log(Weight) ~ Length + Height + Width + Species,
    data = Fish)

Residuals:
     Min       1Q   Median       3Q      Max
-1.63406 -0.07113  0.03562  0.12319  0.40443
```

```
Coefficients:
                  Estimate Std. Error
(Intercept)       2.600474   0.192595
Length            0.053665   0.008936
Height            0.067765   0.029558
Width             0.204504   0.054552
SpeciesParkki     0.043578   0.109118
SpeciesPerch     -0.025003   0.185400
SpeciesPike      -0.073067   0.281650
SpeciesRoach      0.030908   0.174519
SpeciesSmelt     -1.269232   0.206414
SpeciesWhitefish  0.189300   0.177535
---
Residual standard error: 0.2159 on 148 degrees of freedom
```

(a) Test if the coefficient for Height is statistically significantly different from 0 on the 5%-level. State and reflect on necessary assumptions.

(b) What is the average height of a Perch?

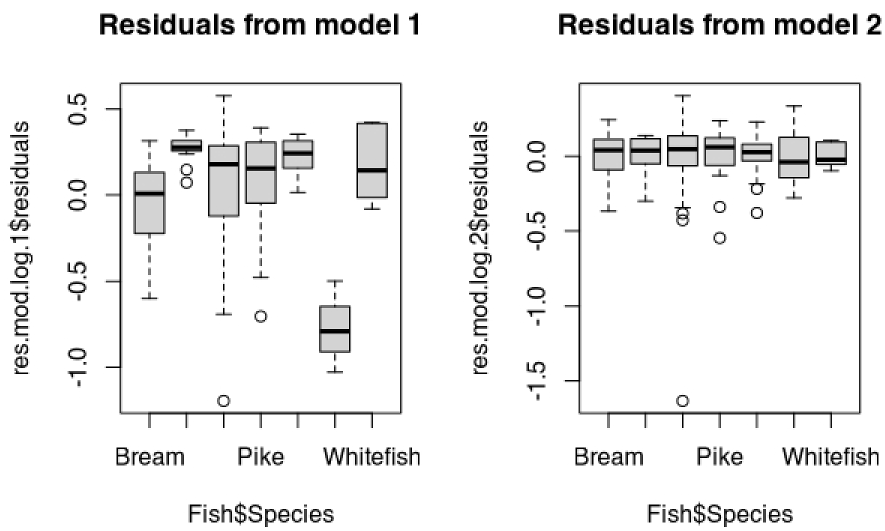
(c) Does the hight of a fish affect weight given that you have controlled for the species?

(d) Using the second model, what is the percentage increase in weight for a Perch if it grows one cm in length but otherwise retains the same height and width?

(e) Test if the type of species of a fish significantly explains weight? Note that `Residual standard error` in the `R` output corresponds to $\hat{\sigma}$

Problem 2. (25 points)

In the figure below, you will find the residuals from Models 1 and 2 in Problem 1 plotted against species.



Let $\hat{u}_i = Weight_i - \widehat{Weight}_i$, where $\widehat{Weight}_i$ are the predictions from Model 2. The output for the regression of $\hat{u}_i^2$ (`residuals`), on the fitted values $\hat{y}_i$ (`fitted`) is provided below

```
Call:
lm(formula = residuals^2 ~ fitted + I(fitted^2))

Residuals:
     Min       1Q   Median       3Q      Max
-0.13287 -0.03377 -0.02158 -0.00778  2.58405

Coefficients:
             Estimate Std. Error
(Intercept)  0.276264   0.182093
fitted      -0.074594   0.075363
I(fitted^2)  0.005519   0.007573

Residual standard error: 0.2138 on 155 degrees of freedom
Multiple R-squared:  0.01901,Adjusted R-squared:  0.006355
```

(a) Based on the plots of the residuals, can you determine if Model 2 suffers from heteroskedasticity?

(b) Based on the regression output provided in Problem 2, for the relevant assumption, perform a formal test.

(c) If the coefficients of Model 2 (of Problem 1) had been estimated by OLS, what would the coefficients of a regression of $\hat{u}_i$ on the predictors have been? Motivate your answer in as much details as you can.

Problem 3. (20 points)
An agricultural economist is looking at crop yield and how this depends on rainfall and temperature and estimates the structural model

$$Yield = \beta_0 + \beta_1 Mintemp + \beta_2 Rain + u,$$

where `Mintemp` is the minimum temperature, with OLS:

```
Call:
lm(formula = Yield ~ Mintemp + Rain, data = Crops)

Residuals:
     Min       1Q   Median       3Q      Max
-1154.10  -327.21   -57.71   254.32  1514.35

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -21.25     587.96  -0.036 0.971197
Mintemp        89.56      31.78   2.818 0.005108 **
Rain          -36.37      10.77  -3.376 0.000821 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 451.8 on 341 degrees of freedom
Multiple R-squared:  0.06329,Adjusted R-squared:  0.05779
F-statistic: 11.52 on 2 and 341 DF,  p-value: 1.442e-05
```

The data is collected for $n = 344$ areas over a large geographical area. A fellow economist argues that the minimum temperature is endogenous and proposes that this can be instrumented with geographical location operationalised in terms of longitude and latitude. Consequently they estimate the reduced form equation

```
Call:
lm(formula = Mintemp ~ Rain + Latitude + Longitude, data = Crops)

Coefficients:
             Estimate Std. Error
(Intercept) -6.978019   2.169064
Rain        -0.005848   0.015763
Latitude     0.001647   0.027018
Longitude    0.269425   0.021425
---


Residual standard error: 0.6368 on 340 degrees of freedom
Multiple R-squared:  0.3359,Adjusted R-squared:  0.3301
```

The predictions `fitted.values` from this regression are appended to the dataset (`Crops`) and together they estimate the following model

```
Call:
lm(formula = Yield ~ Rain + fitted.values, data = Crops)

Coefficients:
              Estimate Std. Error
(Intercept)    4157.05    1040.28
Rain            -48.91      11.11
fitted.values  -137.99      56.51
---

Residual standard error: 453.1 on 341 degrees of freedom
Multiple R-squared:  0.05794,Adjusted R-squared:  0.05241
```

(a) What assumptions need to be satisfied for the instruments? Comment on the plausibility of these insofar possible

(b) Perform a formal test for the IV assumption that is testable

(c) Is there a causal effect of the minimal temperature on the crop yield, judging by the structural model?

Problem 4. (20 points)
An oceanographer is interested in determining the origin of salmon in Alaskan fishing camps. She collects 100 sample fish

	mean	sd	Description
Alaska	0.500		Origin of the fish: a dichotomous variable with 1 for Alaskan, and 0 for Canadian
Male	0.480		Male (1) or Female (0)
Fresh	117.920	26.001	diameter of rings for the first-year freshwater growth (hundrets of an inch)
Ocean	398.140	46.240	diameter of rings for the first-year marine growth (hundrets of an inch)

You are provided the following results for three models

```
Call:
lm(formula = Alaska ~ Male + Fresh + Ocean, data = Salmon)

Residuals:
     Min       1Q   Median       3Q      Max
-0.72731 -0.19651 -0.02904  0.17117  0.81220

Coefficients:
              Estimate Std. Error
(Intercept)  0.0499651  0.4048482
Male        -0.0191219  0.0578986
Fresh       -0.0105047  0.0013272
Ocean        0.0042646  0.0007469
---

Residual standard error: 0.289 on 96 degrees of freedom
Multiple R-squared:  0.6794,Adjusted R-squared:  0.6693
F-statistic:  67.8 on 3 and 96 DF,  p-value: < 2.2e-16

Call:
glm(formula = Alaska ~ Male + Fresh + Ocean, family = binomial(link = "logit"),
    data = Salmon)

Coefficients:
            Estimate Std. Error
(Intercept) -3.78657    6.29358
Male        -0.28156    0.83383
Fresh       -0.12642    0.03570
Ocean        0.04865    0.01457
---
```

```
    Null deviance: 138.629  on 99  degrees of freedom
Residual deviance:  38.674  on 96  degrees of freedom
AIC: 46.674

Number of Fisher Scoring iterations: 7

Call:
glm(formula = Alaska ~ Male + Fresh + Ocean, family = binomial(link = "probit"),
    data = Salmon)

Coefficients:
             Estimate Std. Error
(Intercept) -2.789536   3.438482
Male        -0.268494   0.456793
Fresh       -0.066092   0.016536
Ocean        0.027595   0.007782
---

    Null deviance: 138.629  on 99  degrees of freedom
Residual deviance:  38.785  on 96  degrees of freedom
AIC: 46.785

Number of Fisher Scoring iterations: 8
```

There are two individuals, A and B, with the following covariate values

	A	B
gender	1	1
Fresh	140	118
rating	350	400

(a) For the three models, test if the diameter of rings for the first-year freshwater growth has an effect on the probability a salmon born in Alaska.

(b) What is the estimated difference in probability of the salmons A and B being from Alaska according to the linear probability model?

(c) What is the estimated difference in probability of the salmons A and B being from Alaska according to the Probit model?

(d) What is the estimated difference in probability of the salmons A and B being from Alaska according to the Logit model