STOCKHOLM UNIVERSITY
Department of Statistics
Johan Koskinen

WRITTEN RE-EXAMINATION, ECONOMETRICS I
2023-10-12

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)
We have data on $n = 601$ observations on the variables in Table 1

Table 1: Summary of variables

	Description
affairs	How often engaged in extramarital sexual intercourse during the past year? 0 = none, 1 = once, 2 = twice, 3 = 3 times, 7 = 4–10 times, 12 = monthly, 12 = weekly, 12 = daily.
gender	Dummy for female (=1) or male (=0)
age	variable coding age in years: 17.5 = under 20, 22 = 20–24, 27 = 25–29, 32 = 30–34, 37 = 35–39, 42 = 40–44, 47 = 45–49, 52 = 50–54, 57 = 55 or over.
yearsmarried	variable coding number of years married: 0.125 = 3 months or less, 0.417 = 4–6 months, 0.75 = 6 months–1 year, 1.5 = 1–2 years, 4 = 3–5 years, 7 = 6–8 years, 10 = 9–11 years, 15 = 12 or more years.
children	Are there children in the marriage? Dummy for yes (=1) or no (=0)
religiousness	variable coding religiousness: 1 = anti, 2 = not at all, 3 = slightly, 4 = somewhat, 5 = very
rating	variable coding self rating of marriage: 1 = very unhappy, 2 = somewhat unhappy, 3 = average, 4 = happier than average, 5 = very happy.

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

$$\begin{aligned} affairs = \beta_0 &+ \beta_1 gender + \beta_2 age + \beta_3 yearsmarried + \beta_4 children \\ &+ \beta_5 religiousness + \beta_6 rating + u \end{aligned}$$

Furthermore, another model, Model two, is investigated

$$\begin{aligned} affairs = \beta_0 &+ \beta_1 gender + \beta_2 age + \beta_3 yearsmarried + \beta_4 children \\ &+ \beta_5 religiousness + \beta_6 rating \\ &+ \beta_7 yearsmarried \times children + u \end{aligned}$$

To explore if there may be differences in Model 2 for men and women, Model 2 is estimated separately for men and for women. We refer to these models as Model 2 A and Model 2 B respectively.

The following results were obtained from `R` (some output is hidden):

```
Call:
lm(formula = affairs ~ gender + age + yearsmarried + children +
    religiousness + rating, data = Affairs)

Residuals:
    Min      1Q  Median      3Q     Max
-4.8737 -1.7511 -0.7853  0.1399 12.6164
```

```
Coefficients:
              Estimate Std. Error
(Intercept)    6.20979    0.84517
gender        -0.20277    0.26303
age           -0.04862    0.02250
yearsmarried   0.17146    0.04117
children      -0.22463    0.34354
religiousness -0.48500    0.11131
rating        -0.71164    0.11875
---

Residual standard error: 3.093 on 594 degrees of freedom
Multiple R-squared:  0.1296,Adjusted R-squared:  0.1208
F-statistic: 14.73 on 6 and 594 DF,  p-value: 9.922e-16

lm(formula = affairs ~ gender + age + yearsmarried + children +
    religiousness + rating + yearsmarried * children, data = Affairs)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2526 -1.6927 -0.7761  0.2759 11.9641

Coefficients:
                      Estimate Std. Error
(Intercept)            5.52403    0.87972
gender                -0.12332    0.26341
age                   -0.04640    0.02240
yearsmarried           0.33171    0.07297
children               0.63375    0.47059
religiousness         -0.47594    0.11080
rating                -0.69874    0.11825
yearsmarried:children -0.19640    0.07401
--

Residual standard error: 3.078 on 593 degrees of freedom
Multiple R-squared:  0.1398,Adjusted R-squared:  0.1296
F-statistic: 13.76 on 7 and 593 DF,  p-value: < 2.2e-16
```

For models Model 2 A and Model 2 B:

```
Call:
lm(formula = affairs ~ age + yearsmarried + children + religiousness +
    rating + yearsmarried * children, data = Affairs[Affairs$gender ==
    0, ])

Residuals:
```

```
    Min      1Q  Median      3Q     Max
-5.2732 -1.8160 -0.7969  0.4120 12.1843

Coefficients:
                      Estimate Std. Error
(Intercept)            5.43850    1.19029
age                   -0.04746    0.02893
yearsmarried           0.34093    0.09698
children               0.87475    0.67828
religiousness         -0.45812    0.15690
rating                -0.74184    0.17652
yearsmarried:children -0.19939    0.09783
---
Residual standard error: 3.074 on 279 degrees of freedom
Multiple R-squared:  0.1467,Adjusted R-squared:  0.1284
F-statistic: 7.995 on 6 and 279 DF,  p-value: 5.682e-08

Call:
lm(formula = affairs ~ age + yearsmarried + children + religiousness +
    rating + yearsmarried * children, data = Affairs[Affairs$gender ==
    1, ])

Residuals:
    Min      1Q  Median      3Q     Max
-4.5768 -1.6046 -0.8118  0.1386 11.1033

Coefficients:
                      Estimate Std. Error
(Intercept)            5.49789    1.21431
age                   -0.04679    0.03582
yearsmarried           0.34008    0.11467
children               0.50117    0.67095
religiousness         -0.49071    0.15827
rating                -0.68011    0.16173
yearsmarried:children -0.20871    0.11653
---

Residual standard error: 3.106 on 308 degrees of freedom
Multiple R-squared:  0.1358,Adjusted R-squared:  0.119
F-statistic:  8.07 on 6 and 308 DF,  p-value: 4.161e-08
```

(a) Is there evidence for men having more affairs than women? State clearly your assumptions and discuss your conclusion (in particular, consider the definition of variables).

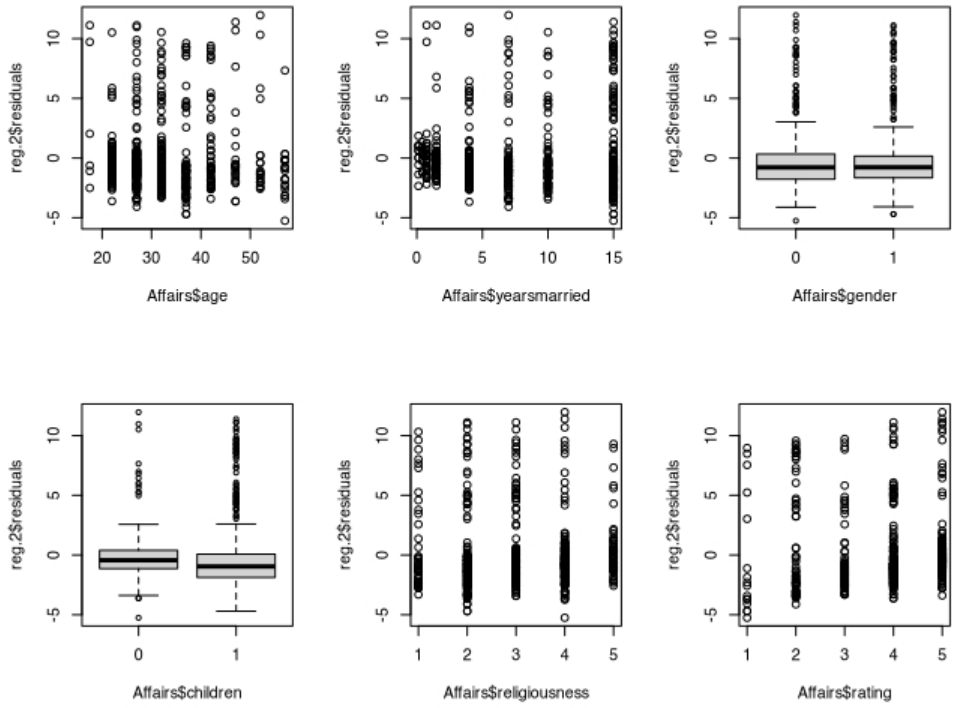
(b) What is the estimated difference in the expected number of affairs between someone who has been married for 1 year and 7, given that they both have 3 children in the marriage?

(c) For a male, who is 29, been married for 3 years, without children, and who is very happy with his marriage, what is the estimated probability that he has been faithful (had no affairs) according to the model? Make appropriate assumptions and motivate your answer.

(d) Is there any difference between males and females according to our models? 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 Model 2 in Problem 1 plotted against six of the predictors.



Some output for the regression of $\hat{u}_i^2$, on the fitted values $\hat{y}_i$ is provided below

```
Call:
lm(formula = reg.2$residuals^2 ~ reg.2$fitted.values + I(reg.2$fitted.values^2))

Residuals:
    Min      1Q  Median      3Q     Max
-26.979  -7.507  -5.187  -2.819 140.610

Coefficients:
                          Estimate Std. Error
(Intercept)                 2.4188     1.4060
reg.2$fitted.values         3.0450     1.4585
I(reg.2$fitted.values^2)    0.6853     0.3761
---

Residual standard error: 20.79 on 598 degrees of freedom
Multiple R-squared:  0.09752,Adjusted R-squared:  0.09451
```

The logarithm of the squared errors, $\log(g_i)$, where $g_i = \hat{u}_i^2$, are regressed on all of the predictors. Denote the predicted values from this equation $\hat{g}_i$.

(a) Based on the plots of the residuals, what assumptions, if any, are violated? Motivate your answer and provide the consequences of these violations for the tests in Problem 1.

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

(c) Write up a regression equation that uses some function of $\hat{g}_i$ to correct for the relevant violation of assumptions in Model 2.

Problem 3. (20 points)
Some researchers are interested in the causal effect of how well you adapt to circumstances, *adaptability*, on your job satisfaction, *satisfaction*. They collected data from 836 individuals on the variables in Table 2.

	mean	sd
gender	0.505	0.500
birthyear	1968.886	8.060
neuroticism	2.585	0.522
extraversion	3.417	0.492
conscientiousness	3.938	0.420
openess	3.482	0.510
agreeableness	3.416	0.297
adaptability	3.759	0.505
satisfaction	3.234	0.447

Table 2: Variables on job satisfaction, adaptability, personality, and relevant demographics

The variables `neuroticism`, `extraversion`, `conscientiousness`, `openess` (openess to new experiences), and `agreeableness`, are standard, psychological personality measures. Gender is one (1) for women and zero (0) for men. The variable `satisfaction` measures how satisfied you are with your job, from not at all (0), to completely (5). The variable `adaptability` measures how well you adapt to challenges, from not at all (0), to completely (5).

The researchers assume the structural model

$$\begin{aligned} jobsat &= \beta_0 + \beta_1 agreeableness + \beta_2 neuroticism + \beta_3 extraversion + \beta_4 gender \\ &+ \beta_5 birthyear + \beta_6 adaptability + u \end{aligned}$$

They want to estimate β_6 but suspect that the variable `adaptability` is endogenous. They believe that this variable can be instrumented using the exogenous variables `conscientiousness` and `openess`.

You are provided the following results.
Model 1:

```
Call:
lm(formula = satisfaction ~ agreeableness + neuroticism + extraversion +
    gender + birthyear + adaptability, data = jobsatdata)

Residuals:
     Min       1Q   Median       3Q      Max
-1.19599 -0.28682 -0.00672  0.29468  1.45400

Coefficients:
               Estimate Std. Error
```

```
(Intercept)    3.4427657  3.6503440
agreeableness  0.1035314  0.0502682
neuroticism   -0.1659116  0.0328864
extraversion   0.1211100  0.0344056
gender        -0.0110649  0.0304720
birthyear     -0.0003866  0.0018500
adaptability   0.0584554  0.0341877
---


Residual standard error: 0.4245 on 829 degrees of freedom
Multiple R-squared:  0.1033,Adjusted R-squared:  0.09684
```

Model 2:

```
Call:
lm(formula = satisfaction ~ agreeableness + neuroticism + extraversion +
    gender + birthyear + conscientiousness + openess + adaptability,
    data = jobsatdata)

Residuals:
     Min       1Q   Median       3Q      Max
-1.22059 -0.28040 -0.01496  0.28971  1.47309

Coefficients:
                   Estimate Std. Error
(Intercept)        2.9513065  3.6664211
agreeableness      0.1025376  0.0504772
neuroticism       -0.1656718  0.0334528
extraversion       0.1096429  0.0358156
gender            -0.0057088  0.0306543
birthyear         -0.0002453  0.0018561
conscientiousness  0.0681081  0.0403747
openess            0.0327082  0.0343371
adaptability       0.0239793  0.0396168
---

Residual standard error: 0.4242 on 827 degrees of freedom
Multiple R-squared:  0.1068,Adjusted R-squared:  0.0982
```

Model 3:

```
Call:
lm(formula = adaptability ~ agreeableness + neuroticism + extraversion +
```

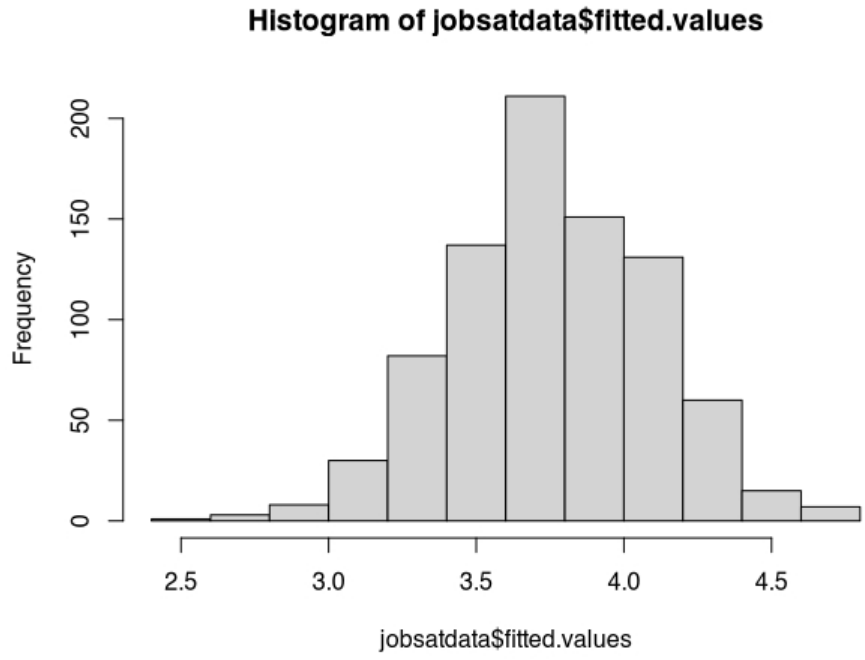
```
    gender + birthyear + conscientiousness + openess, data = jobsatdata)

Residuals:
     Min       1Q   Median       3Q      Max
-1.09647 -0.25786  0.00065  0.23706  1.44086

Coefficients:
                   Estimate Std. Error
(Intercept)       -2.897497   3.214658
agreeableness      0.010834   0.044278
neuroticism       -0.269763   0.027807
extraversion       0.116399   0.031156
gender             0.070808   0.026778
birthyear          0.002096   0.001627
conscientiousness  0.404771   0.032504
openess            0.333868   0.027797
---

Residual standard error: 0.3721 on 828 degrees of freedom
Multiple R-squared:  0.4615,Adjusted R-squared:  0.457
```

The fitted values, predictions, from this model (Model 3) are plotted in the figure below



Further, you have Model 4:

```
Call:
lm(formula = satisfaction ~ agreeableness + neuroticism + extraversion +
    gender + birthyear + fitted.values, data = jobsatdata)

Residuals:
     Min       1Q   Median       3Q      Max
-1.19447 -0.28498 -0.01231  0.29346  1.48521

Coefficients:
                Estimate Std. Error
(Intercept)    3.2548185  3.6462903
agreeableness  0.0985912  0.0502719
neuroticism   -0.1326843  0.0380969
extraversion   0.0910000  0.0385556
gender        -0.0147158  0.0304984
birthyear     -0.0004639  0.0018477
fitted.values  0.1584285  0.0674081
---


Residual standard error: 0.4238 on 829 degrees of freedom
Multiple R-squared:  0.1061,Adjusted R-squared:  0.09965
```

(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 *adaptability* on job satisfaction, judging by the structural model?

(d) Would our interpretation of the effect of *adaptability* on job satisfaction, be different if we did not treat *adaptability* as endogenous?

Problem 4. (20 points)

Consider the same dataset as in Problem 1, but define a new dependent variable as

$$y_i = \begin{cases} 1, & \text{if } affairs_i > 0 \\ 0, & \text{else} \end{cases}$$

You are provided the following results

```
Call:
lm(formula = y ~ gender + age + yearsmarried + children + religiousness +
    rating + yearsmarried * children, data = Affairs)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6108 -0.2715 -0.1622  0.1462  1.0514

Coefficients:
                      Estimate Std. Error
(Intercept)           0.797019   0.117472
gender               -0.053621   0.035174
age                  -0.006987   0.002991
yearsmarried          0.030344   0.009744
children              0.124748   0.062840
religiousness        -0.053844   0.014796
rating               -0.085211   0.015791
yearsmarried:children -0.017410   0.009883

Call:
glm(formula = y ~ gender + age + yearsmarried + children + religiousness +
    rating + yearsmarried * children, family = binomial(link = "logit"),
    data = Affairs)

Coefficients:
                     Estimate Std. Error
(Intercept)           1.66887    0.71222
gender               -0.33198    0.20932
age                  -0.04217    0.01812
yearsmarried          0.19432    0.05883
children              0.95768    0.41415
religiousness        -0.32787    0.08991
rating               -0.45306    0.08991
yearsmarried:children -0.11823    0.05850
---

Call:
```

```
glm(formula = y ~ gender + age + yearsmarried + children + religiousness +
    rating + yearsmarried * children, family = binomial(link = "probit"),
    data = Affairs)

Coefficients:
                      Estimate Std. Error
(Intercept)            0.93200    0.41031
gender                -0.19427    0.12179
age                   -0.02379    0.01049
yearsmarried           0.11407    0.03398
children               0.53762    0.23192
religiousness         -0.18567    0.05208
rating                -0.26362    0.05293
yearsmarried:children -0.07018    0.03399
---
```

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

	A	B
gender	1	1
age	34	34
yearsmarried	4	4
children	0	1
religiousness	3	3
rating	3	3

(a) For the three models, test if the effect of having children on the probability of having an affair changes with the number of years you have been married.

(b) What is the estimated difference in probability of having an affair for persons A and B according to the linear probability model?

(c) What is the estimated difference in probability of having an affair for persons A and B according to the linear Probit model?

(d) What is the estimated difference in probability of having an affair for persons A and B according to the Logit model