UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Postponed exam: ECON3150/4150 – Introductory Econometrics

Date of exam: Friday, 17 June, 2022

Time for exam: 09.00 – 12.00 (3 hours)

The problem set covers 5 pages (incl. cover sheet)

Resources allowed:

- All written and printed resources – as well as two alternative calculators - are allowed

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

ECON3150/4150: Introductory Econometrics – Postponed Exam Spring 2022

1. (80%) Suppose you have data from from the 1977–1978 Australian Health Survey. Descriptive statistics of your data stored in data frame `df` are as follows:

```
##              mean        SD  min   max    N
## visits  0.3007239 0.7931689 0.00  9.00 5111
## gender  0.5188808 0.4996923 0.00  1.00 5111
## age     0.4075954 0.2050066 0.19  0.72 5111
## income  0.5921737 0.3644960 0.01  1.50 5111
## private 0.4423792 0.4967173 0.00  1.00 5111
## health  1.2105263 2.1181110 0.00 12.00 5111
```

   where

   1. **visits** Number of doctor visits in the past 2 weeks.
   2. **gender** indicates gender (1=female, 0=male)
   3. **age** Age in years divided by 100.
   4. **income** Monthly income in tens of thousands of dollars.
   5. **private** Does the individual have private health insurance? (1=yes, 0=no)4
   6. **health** General health questionnaire score (a higher score implies worse health)

   You perform the following analysis:

```
df$age = df$age - 0.4064
reg = feols(visits ~ private + age + log(income), df, vcov="hetero")
reg
```

```
## OLS estimation, Dep. Var.: visits
## Observations: 5,111
## Standard-errors: Heteroskedasticity-robust
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)  0.250317   0.020171 12.40992  < 2.2e-16 ***
## private      0.023836   0.022646  1.05256 2.9259e-01
## age          0.445872   0.057003  7.82190 6.2774e-15 ***
## log(income) -0.052624   0.018486 -2.84677 4.4342e-03 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## RMSE: 0.785756   Adj. R2: 0.017835
```

```
vcov(reg)
```

```
##                (Intercept)       private          age   log(income)
## (Intercept)  0.0004068594 -2.855744e-04 0.0002570139  2.582298e-04
## private     -0.0002855744  5.128520e-04 0.0001087971 -9.223615e-05
## age          0.0002570139  1.087971e-04 0.0032493452  1.513678e-04
## log(income)  0.0002582298 -9.223615e-05 0.0001513678  3.417155e-04
```

a. Interpret the estimated coefficient on `log(income)`.

b. Construct (briefly explain your steps) and interpret the 86 percent confidence interval for the estimate in 1.a.

c. Can we give the estimate in 1.a a causal interpretation? Motivate your answer.

d. What is the interpretation of the Intercept?

e. Predict the outcome for a 60 year old female with private health insurance who earns 5,000.

f. Compute the marginal effect of income:

$$\partial E[visits|gender, age, income]/\partial income$$

when income=2,000.

g. Compute the standard error of the estimate in 1.f.

A friend is concerned that the estimate of having private health insurance on the number of doctor visits suffers from omitted variable bias. She suggests to use instrumental variable estimation instead with gender as the instrumental variable, arguing that women are more risk averse than men and therefore more likely to take out private health insurance.

Your friend provides the following OLS regressions:

```
feols(health ~ gender + age + log(income), df, vcov="hetero")
```

```
## OLS estimation, Dep. Var.: health
## Observations: 5,111
## Standard-errors: Heteroskedasticity-robust
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)  0.939807   0.048328 19.44635  < 2.2e-16 ***
## gender       0.173493   0.062528  2.77462 5.5466e-03 **
## age         -0.168304   0.154712 -1.08785 2.7671e-01
## log(income) -0.242052   0.046601 -5.19412 2.1368e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.10766   Adj. R2: 0.009073
```

```
feols(visits ~ gender + age + log(income), df, vcov="hetero")
```

```
## OLS estimation, Dep. Var.: visits
## Observations: 5,111
## Standard-errors: Heteroskedasticity-robust
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)  0.235551   0.017337 13.58641  < 2.2e-16 ***
## gender       0.065540   0.024525  2.67238 7.5555e-03 **
## age          0.408414   0.059960  6.81148 1.0775e-11 ***
## log(income) -0.041048   0.018700 -2.19507 2.8203e-02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.78523   Adj. R2: 0.019149
```

```
feols(private ~ gender + age + log(income), df, vcov="hetero")
```

```
## OLS estimation, Dep. Var.: private
## Observations: 5,111
## Standard-errors: Heteroskedasticity-robust
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)  0.516371   0.011696 44.15045 < 2.2e-16 ***
## gender       0.125924   0.014002  8.99360 < 2.2e-16 ***
## age         -0.048729   0.034591 -1.40870   0.15898
## log(income)  0.186356   0.010227 18.22188 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.476583   Adj. R2: 0.078707
```

i. Compute and interpret the IV estimate of the effect of having private health insurance on the number of doctor visits in the past 2 weeks.

j. What are the assumptions that are necessary for the IV estimator in 1.i to be internally valid and explain **why** they need to hold.

k. Do you believe that the assumptions in 1.j hold in this specific application? Explain your answer.

2. (20%) Briefly outline and motivate the analysis you would perform in the following cases. Make sure to explain the assumption(s) that need to hold for internal validity in the context of the application.

   a. From 2007 workers in the municipality of Mandal could self-certify their sickness, meaning that they did not need a doctor's certificate to be absent from work. You want to estimate the causal effect of the self-certification policy on sickness absence. You have a dataset for the years 2000 to 2015. In the data you have information on workers' annual sickness absence, their municipality of residence, education, age and gender.

   b. A number of secondary schools in the Netherlands run a program for gifted students. Students are graded on a 0 to 10 point scale, and those who score at least an 8 participate in the program. You are interested in estimating the causal impact of participating in the program on student's final exam score in mathematics. You have a dataset with students' average grade used to determine their eligibility for the program, their gender, parents' education, and final exam score in mathematics.