UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Exam: **ECON4136 – Applied statistical analysis for the social sciences**

Date of exam: Tuesday, November 25, 2014 **Grades are given: December 15, 2014**

Time for exam: 2.30 p.m. – 5.30 p.m.

The problem set covers 6 pages (incl. cover sheet)

Resources allowed:

- All written and printed resources, as well as calculator, are allowed.

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

Exam ECON4136, ECON5100, ECON9100 – Fall 2014

IMPORTANT: Always explain answers. Answers should show knowledge and understanding of the concepts taught in the course. Each subquestion is weighted equally in the final grade.

Buser, T. (2014), The effect of income on religiousness (*American Economic Journal: Applied Economics*) investigates whether income affects religiousness using *survey* data from Ecuador. The sample consists of relatively poor households. Assume that they spend all their (monthly) income, i.e. $income = expenditures$. Religiousness is measured both by self-assessment on a scale from 0 to 10 (-*religiousness*-), and by the number of religious services that are attended in a month (-*attendance*-). Below you see some sample statistics, and the results from a regression of attendance on log of expenditures (-*logexp*-), log of household size (-*loghhs*-), and the age (-*age*-) and schooling (-*edu*-) of the respondent (both in years):

```
. sum attendance logexp edu age loghhs
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
  attendance |      2645    4.586579    6.406292          0         30
      logexp |      2638    5.578003    .5088621   2.079442   7.438384
         edu |      2630    7.446388    3.686763          0         18
         age |      2645    42.71871     11.0413          0         90
      loghhs |      2645    1.403439    .4428044          0   3.091042

. reg attendance logexp edu age loghh, robust
Linear regression                                      Number of obs =    2623
                                                       F(  4,  2618) =    7.67
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0119
                                                       Root MSE      =  6.3378
------------------------------------------------------------------------------
             |               Robust
  attendance |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      logexp |   .1589818   .2896088     0.55   0.583    -.4089035     .726867
         edu |   .0492528   .0390654     1.26   0.208    -.0273494    .1258551
         age |   .0694903   .0126477     5.49   0.000     .0446899    .0942908
      loghhs |   .1777673   .3421433     0.52   0.603    -.4931314     .848666
       _cons |   .0901014   1.652156     0.05   0.957    -3.149562    3.329765
------------------------------------------------------------------------------

. mat l e(V)
            logexp         edu         age      loghhs       _cons
logexp   .08387323
   edu  -.00004765   .00152611
   age   .00022973   .00016474   .00015996
loghhs  -.04446538  -.00180977   .00058075   .11706204
 _cons   -.4139303  -.01526259  -.00986856   .07526461   2.7296187
```

1. Answer the following questions in detail

   (a) Explain when the coefficient on -*logexp*- in the regression above can be given a causal interpretation.

   (b) Give examples that lead to the coefficient on -*logexp*- being biased upwards. Give examples that can cause the coefficient on -*logexp*- to be biased downwards.

   Assume that the regression above can be given a causal interpretation.

   (c) What is the estimated average effect of a 12 percent increase in monthly income on the number of religious services attended? What is the standard error of this estimate?

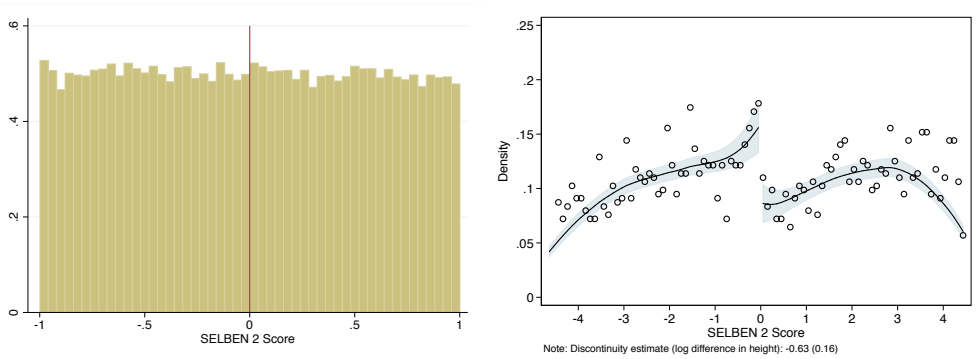
   (d) Consider a policy that increases the income of households by 12 percent. You are interested in the effect of this policy on religious attendance. Suppose that in a sample of size 2,600, half randomly receive the treatment, while the other half is the control group. You will estimate the treatment effect by the difference in mean attendance between the treated and controls. What is the statistical power of this design?

   (e) What sample size is necessary to achieve a power of 0.8 in (d)?

2. To estimate the effect of income on religious attendance, Buser (2014) exploits a feature of a cash transfer program: Families who are below a threshold on a poverty index are eligible for a cash transfer. The index variable is called -*selben2*- and the threshold is at 0. The variable -*eligible*- equals one if a family's index is below the threshold, and is zero otherwise. On average the size of the transfer is about 12 percent of family income.

   (a) The estimation approach consists of comparing people around the threshold. Explain what conditions need to hold for this approach to give causal estimates of the cash transfer.

   (b) The graph to the left shows the distribution of people around the threshold in Ecuador reported in Buser (2014). The graph to the right shows the same distribution, but in the sample used for estimation. Discuss possible explanations for the observed differences, and the potential implications of these differences for the analysis.



(c) Explain for each of the variables below whether they are good candidates for a test of covariate balance around the cutoff.

```
. reg selben2 rel edu loghhs age, noheader
------------------------------------------------------------------------------
     selben2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
religiousness |  -.0317039   .0210593    -1.51   0.132    -.0729984    .0095907
         edu |   .0481182   .0152145     3.16   0.002     .0182845    .0779518
      loghhs |  -.2263135   .1144582    -1.98   0.048     -.450751   -.0018761
         age |   .0080028   .0051462     1.56   0.120    -.0020883    .0180939
       _cons |  -.0591586    .367655    -0.16   0.872    -.7800816    .6617644
------------------------------------------------------------------------------

. reg attendance rel edu loghhs age, noheader
------------------------------------------------------------------------------
  attendance |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
religiousness |   .6664174    .050627    13.16   0.000     .5671446    .7656902
         edu |   .0443575    .036576     1.21   0.225    -.0273632    .1160781
      loghhs |   .3528093   .2751594     1.28   0.200    -.1867419    .8923605
         age |   .0560195   .0123716     4.53   0.000     .0317603    .0802786
       _cons |  -3.191777   .8838487    -3.61   0.000    -4.924888   -1.458666
------------------------------------------------------------------------------
```

(d) Consider the regression

$$collect = \delta_0 + \delta_1 eligible + \delta_2 selben2$$

where -*collect*- equals one if families collect the cash transfer and is zero otherwise. What are the implications of a so-called sharp design for the coefficients in this regression?

(e) Consider the OLS regression

$$attendance = \beta_0 + \beta_1 eligible + \beta_2 selben2 + \beta_3 selben2 * eligible$$

on the subsample where $selben2 \in [-1, 1]$. Is the estimate $\hat{\beta}_1$ from this regression equivalent to the local linear regression estimate of eligibility using a uniform kernel and

bandwidth of 1? Explain your answer.

3. Most people, about 83 percent, attend at most one service a week. Define a dummy variable -*dattend*- which equals one if people attend a religious service more than once a week, and which is zero otherwise. You are interested in estimating

$$\Pr(dattend = 1|eligible) \tag{1}$$

   (a) What Stata code would you use to estimate (1) using a linear probability model?

   You decide to estimate (1) using the non-linear Logit model.

   (b) Will this give the same estimate of (1) as the linear probability model?

   Suppose the results of the Logit estimation are:

```
. logit dattend eligible, noheader

------------------------------------------------------------------------------
     dattend |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    eligible |   .2256065   .1053454     2.14   0.032     .0191333    .4320797
       _cons |  -1.741551    .077833   -22.38   0.000    -1.894101   -1.589001
------------------------------------------------------------------------------

. mat l e(V)

                    dattend:    dattend:
                   eligible       _cons
dattend:eligible   .01109765
   dattend:_cons  -.00605798   .00605798
```

   (c) What is the estimated sample average effect of eligibility on the probability to attend at least one religious service per week?

   (d) What is the standard error of the sample average effect of eligibility in (c)?

4. Not everybody who is eligible for the cash transfer in Buser (2014), actually collected the money. Let -*collect*- be a dummy variable that equals 1 if a family collected the cash transfer, and is zero otherwise.

   Consider the following regression:

```
. gen selben2_1 = selben2 * eligible
. gen selben2_0 = selben2 * (1 - eligible)
. reg collect eligible  selben2_1 selben2_0, robust noheader
------------------------------------------------------------------------------
             |               Robust
     collect |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    eligible |   .8045451    .022037    36.51   0.000     .7613336    .8477566
  selben2_0  |  -.0015438   .0041907    -0.37   0.713    -.0097612    .0066736
  selben2_1  |  -.0091181   .0075935    -1.20   0.230    -.0240079    .0057717
       _cons |   .0328388   .0111414     2.95   0.003     .0109921    .0546855
------------------------------------------------------------------------------
```

(a) How do you interpret the coefficient on -*eligible*-?

(b) Explain how you would estimate the causal effect of receiving the cash-benefit on -*attendance*-.

(c) Suppose the point estimate you obtain in (b) is 1.8. Interpret this estimate, also in light of 1(c) above.

(d) Explain how you would estimate the causal effect of the cash-benefit policy on -*attendance*-, and discuss its interpretation.

(e) Consider the following expression

$$\delta = E[attendance \times collect|eligible = 1] - E[attendance \times collect|eligible = 0]$$

Derive δ in terms of potential outcomes and potential treatments, assuming eligibility is random, and has no independent effect on -attendance-. Also assume that eligibility can only affect people's tendency to collect money in one direction.