UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Postponed exam: **ECON5100 – Advanced Econometrics**

Date of exam: Friday, January 13, 2017

Time for exam: 09:00 a.m. – 12:00 noon

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

Resources allowed:

- Open book exam, where all written and printed resources, as well as calculator, is allowed

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

Postponed exam ECON5100 – Fall 2016

IMPORTANT: Answers should show knowledge and understanding of the concepts taught in the course. Therefore always explain answers: yes/no answers do not get credit. Be to the point: We value correct answers, not long answers.

Subquestions are weighted equally.

---

1. (35%) When people have a stomach pain some choose to take a painkiller while others do not. Assume that people can: i) take P, ii) take A, iii) do nothing (N). Let $M_i \in \{P, A, N\}$ denote person i's usual choice, and $H_i \in \{0, 1\}$ whether that person is still suffering from a stomach pain 2 hours later.

   (a) What potential outcomes can be defined in this context?

   (b) Write down the function that links the observed outcome to the potential outcomes.

   You are interested in the effectiveness of painkillers to treat stomach pain and randomly assign P, A, or a placebo, denoted by O to a representative sample of the population in a double blind trial. Let the variable $T_i \in \{P, A, O\}$ define person i's treatment assignment. There is perfect compliance with the treatment.

   (c) Explain SUTVA in this context.

   (d) In addition to a placebo one can imagine a treatment that assigns nothing (N). Given the randomization of treatment T_i explain what

   $$E[H_i|T_i = N] = E[H_i|T_i = O]$$

   implies in potential outcomes and words.

   (e) What average treatment effects can you estimate in your experiment using a random sample with information on H_i and T_i?

   (f) Suppose that when they have a stomach pain, 30% of the people tend to use A, 50% uses P while 20% does nothing. Assume effects are homogenous. Explain how you would estimate the effect of abolishing painkillers using your experimental data on H_i and T_i.

   (g) Suppose now that effects are heterogeneous and that in practice there is selection on gains with respect to taking of painkillers. Does this have implications for your answer in (d)?

   (h) How would you estimate the effect in (d) using existing data on H_i, M_i and T_i?

2. (15%) In each of the following cases explain whether and at what level(s) you ideally should cluster your standard errors

   (a) You have randomly assigned a microcredit program to poor farmers across 100 villages. You will estimate how access to this credit affects crop yields at the next harvest.

   (b) You have randomly selected 50 villages where poor farmers will get access to a microcredit program, and 50 villages where poor farmers do not get access. You will estimate how access to this credit affects crop yields. You analyze a panel dataset with crop yields from 3 harvests.

   (c) You have randomly assigned a microcredit program to poor farmers across 100 villages. You will estimate how access to this credit affects crop yields. You analyze a panel dataset with crop yields from 3 harvests.

3. (30%) You are interested in estimating the effect of speaking rural dialect ($dialect_i$) on the log of earnings ($earnings_i$) (β_1) by running the following regression

$$earnings_i = \beta_0 + \beta_1 dialect_i + e_i$$

but worry about omitted variable bias.

(a) Discuss potential sources of omitted variable bias and how (ie in what direction) this may bias your estimate of β_1.

(b) You have a dataset with a rich set of controls X_i. Explain how you would estimate β_1 using linear regression techniques. What is your identifying assumption here?

(c) A helpful colleague tells you that you forgot to control for hours worked in (b). What is the best way to address this suggestion?

(d) Discuss potential advantages of matching over OLS. Is there a potential disadvantage?

You think that you have a valid instrument, namely distance to London ($distance_i$), for $dialect_i$ and now plan to implement an instrumental variables (IV) strategy.

(e) Discuss instrument validity, and make sure to explain the difference between instrument exogeneity and the exclusion restriction in the model above. Do you want to add additional controls to your IV model?

(f) Someone suggests to test the exclusion restriction by estimating the following regression

$$earnings_i = \beta_0 + \beta_1 dialect_i + \beta_2 distance_i + u_i$$

and test $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$. What to you reply?

(g) What is the local average treatment effect interpretation of β_1.

4. (20%) Wolfers (2006, AER) investigates how unilateral divorce laws (the possibility to file for divorce without the consent of your spouse) affect divorce rates. He exploits a state (s), year (t) level panel dataset with information on the nr of divorces per 1000 inhabitants ($divorce_{st}$) and whether the state s had a unilateral divorce law in year t ($unilateral_{st}$), and estimates the following regression

$$divorce_{st} = \beta unilateral_{st} + \eta_s + \tau_t + \varepsilon_{st}$$

where η_s are state fixed effects and τ_t year fixed effects.

(a) What variation in $unilateral_{st}$ is used in the equation above to estimate β.

(b) Explain what Wolfers needs to assume about ε_{st} for consistent estimation of β, and why this is or is not reasonable.

(c) Instead of using state fixed effects, Wolfers could have used random effects estimation instead. Would this have been more attractive? Explain.

(d) Should the regression above be weighted with state population weights? Discuss.

(e) Table 1 shows the results from the specification above in column (1), as well as from two specifications that control for state specific trends. Discuss these results.

Figure 1: Table 1 from Wolfers (2006, AER)

	(1) Basic specification	(2) State-specific trends linear	(3) State-specific trends quadratic
Unilateral	0.000	0.431	0.435
	(0.057)	(0.051)	(0.055)
Year effects	$F = 89.3$	$F = 95.3$	$F = 9.0$
State effects	$F = 216.5$	$F = 191.6$	$F = 129.1$
State trend, linear	No	$F = 24.4$	$F = 9.3$
State trend, quadratic	No	No	$F = 6.6$
Adjusted R^2	0.946	0.976	0.981

Notes: Sample: 1968–1988, $n = 1043$ (unbalanced panel). Estimated using state population weights. Standard errors in parentheses.