Exam ECON5100, ECON9100 – 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 headache some choose to take a painkiller while others do not. More specifically assume people can: i) take a paracetamol (P), ii) take aspirin (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 headache 2 hours later.

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

   Answer hint: H_i^P, H_i^A, H_i^N, namely whether the person would have headache two hours later if she would have taken P, A or done nothing.

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

   Answer hint: $H_i = 1(M_i = P) * H_i^P + 1(M_i = A) * H_i^A + 1(M_i = N) * H_i^N$

   You are interested in the effectiveness of painkillers to treat headaches and randomly assign paracetamol (P), aspirin (A), or a placebo (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.

   Answer hint: SUTVA means that every person gets the same treatment. In this context it means that all the pills are identical, and that there are also no spillovers: a person's potential outcomes are unaffected by whether someone else takes a given painkiller (or not).

   (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.

Answer hint: $E[H|T=N]=E[H|T=O]$ implies $E[H^N]=E[H^O]$, i.e. no placebo effects. note we need a new potential outcome

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

Answer hint: $E[H_i^k - H_i^l]$ for all pairs l,k (P vs A, P vs O, A vs O). (and strictly speaking weighted averages of these.)

(f) Suppose that when they have a headache, 30% of the people tend to use aspirin, 50% uses paracetamol 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.

Answer hint: $E[H|T=O]-E[H|T=A]*0.3-E[H|T=P]*0.5-E[H|T=O]*0.2$

(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)~~ (f)?

Answer hint: With selection on gains $E[H^l-H^k|M=l]>E[H^l-H^k]$. This means that (d) will underestimate the increase in headaches.

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

Answer hint: $E[H|T=O]-E[H|T=A,M=A]*0.3-E[H|T=P,M=P]*0.5-E[H|T=O,M=N]*0.2$.

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.

   Answer hint: use the general Moulton factor

   $$\frac{V_c(\hat{\beta})}{V_s(\hat{\beta})} = 1 + \left(\frac{V(n_g)}{\overline{n_g}} + \overline{n_g} - 1 \right) \rho_u \rho_x$$

   to judge this.
   while it is likely that $\rho_u > 0$ within villages, note that $\rho_x = 0$ because of the random assignment of the treatment. therefore **no clustering** at the village (or other) level.

   (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.

   Answer hint: it is likely that $\rho_u > 0$ within villages, also note that $\rho_x = 1$ because of the random assignment of the treatment to village. therefore clustering at the village level. similar reasoning because of the multiple observations at the farmer level suggests that we also need to cluster at the individual level. because this cluster is nested in the larger village cluster, **clustering at the village level** will do.

   (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.

   Answer hint: it is likely that $\rho_u > 0$ within persons, $\rho_x = 1$ because of the random assignment of the treatment to people. therefore **clustering at the individual level.**

3. (30%) You are interested in estimating the effect of speaking Non-Oslo 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.

Answer hint: One can come up with many different types of omitted variable bias in this simple bivariate regression. I would suspect that dialect speakers are from more rural areas and are likely to have less or different types of schooling, labor markets are different etc. In most cases I would think that dialect speaking correlates *positively* with things that affect earnings *negatively* and this result in the estimated β_1 being too low. Consistent and well argued stories that use the omitted variable bias formula are rewarded here.

(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?

Answer hint: Here one would like to control for X's which correlate with dialect, but are not caused by it such as examples from (a): schooling, labor market conditions, information about parents, etc. Now we assume that dialect is exogenous (independent of the error terms/potential outcomes) conditional on our covariates. The implementation could be a simple regression

$$earnings_i = \beta_0 + \beta_1 dialect_i + x_i'\gamma + u_i$$

or (better) less restrictive estimations such as a a regression that interacts the coefficients on x_i with $dialect_i$ (and then backing out the effect by calculating the APE, or a semiparametric procedure such as matching).

(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?

Answer hint: Tell the colleague that this is what Angrist and Pischke call a "bad control"; it is at least partially caused by dialect our endogenous variable of interest and therefore endogenous itself. We would now need at least 2 instruments to separately estimate the effects of dialect and hours worked.

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

Answer hint: Matching has the advantage i) of relaxing the functional form specification in our OLS regression, and ii) highlighting support issues and not use the functional form of the regression to extrapolate to treated units for who we have no untreated units with the same Xs. The main disadvantage lies in more involved implementation, and in particular in a loss of precision (the matching estimator is less efficient than OLS).

You think that you have a valid instrument, namely distance to Oslo ($distance_i$), 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?

Answer hint: We need

i. exogeneity: distance should be independent of potential outcomes $Y^{dialect,distance}$.

ii. exclusion: distance may not have an effect over and above dialect speaking. since distance will probably correlated with urbanity and thus labor markets i would expect that at least unconditionally it will violate exclusion.

iii. monotonicity: increasing distance from oslo should increase (and not decrease) the likelihood of speaking non-oslo dialect. probably reasonable.

iv. relevance: does distance affect dialect speaking. this we can check in a first stage.

(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?

Answer hint: The exclusion restriction involves multiple counterfactuals and only 1 is observed. This assumption (like any counterfactual assumption) is therefore untestable with observational data.

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

Answer hint: the effect of dialect speaking on earnings for those who would speak dialect living far away from Oslo, but not otherwise. (Note that this make the continuous instrument binary. With a continuous instrument, IV estimates some weighted average of local compliers where local is now defined as the value of the instrument).

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.

Answer hint: variation within state over time in deviation of yearly averages.

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

Answer hint: wolfers needs strict exogeneity = conditional on the fixed effects $unilateral_{st}$ is independent $\varepsilon_{s,t'}$ for all t' . f.e. a positive shock in $\varepsilon_{i,t}$ may not lead to the introduction of a law in period t (or $t+1$ etc). Note that this, in addition to the fixed effects, is part of the common trend assumption.
It essentially assumes random timing of these laws. To judge this requires good institutional knowledge but below we'll see that, perhaps unsurprisingly, there may be issues with this assumption.

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

Answer hint: FE are attractive because it allows for arbitrary correlation between unilateral and the error terms. the RE model assumes that unilateral is exogenous, a much stronger and less attractive assumption.

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

Answer hint: Weighting changes the population for which you will be estimating the effect. The unweighted regression estimates the effect for an average states, the weighted regression for an average American. (I would argue that the last one is more interesting.) With homogenous effects weighting obviously does not matter, but otherwise it will: for example, if effects are different in large states vs small states.

(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.

Answer hint: Column (1) shows an essentially and non-significant point effect of divorce laws on divorce rates.
The introduction of divorce laws may correlate with pre-existing trends in the divorce rate. This would be a violation of the strict-exogeneity/common trend assumption. Column (2) addresses this concern and includes states specific time trends. Now we see a statistically significant estimate of 0.43; an increase in 4.3 divorcees per 1000 habitants. Apparently states where the divorce rate was increasing were more likely to introduced unilateral divorce laws.
Allowing for more flexible trends in column (3) does not change this conclusion.

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.