ECON4137 - Postponed exam - Spring 2022

1. Let Y_i^d be the potential outcome of treatment $d \in (0,1)$ for individual i. Let the variable D_i equal 1 if treated and zero otherwise.

   (a) The observed outcome is

   $$Y_i = Y_i^0 + D_i(Y_i^1 - Y_i^0)$$

   Explain the equation. Which variables are observed under which conditions?

   (b) Interpret each of the following 4 differences in expectations as treatment effects and explain how they differ. Which of these differences can be given a causal interpretation? Make sure you address all 4 expectations.

   i. $E[Y_i^1 - Y_i^0]$
   ii. $E[Y_i^1 - Y_i^0 | D_i = 1]$
   iii. $E[Y_i^1 - Y_i^0 | D_i = 0]$
   iv. $E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 0]$

   (c) Assume that $D_i \perp \{Y_i^0, Y_i^1\}$. Explain the content of this assumption. Under which empirical conditions may this assumption be credible?

   (d) Explain how you would estimate the treatment effects in 1b given the assumption in 1c? Which of the treatment effects discussed in question 1b would you be able to estimate consistently? How do they differ in expectation?

   (e) Assume now that $D_i \perp \{Y_i^0, Y_i^1\} | X_i$ where X_i is a vector (or scalar) of individual specific predetermined variables. What is this assumption called? Is this assumption stronger than the assumption in question 1c? Give an example of how to motivate such an assumption in practice.

(f) Relative to your answer in question 1c, how would you change your approach to estimating the average treatment effect.

(g) Assume that you estimate the average treatment effect under the assumption in question 1e but that is in reality the assumption in question 1c is true. Would your results be biased?

2. Bell, Costa & Machin (2016, Economics of Education Review) investigates the link between schooling and crime. The authors use commuting zone level data from the US (Commuting zones are the travel-to-work areas of a city where at least 15 % of employed residents are working in a city). Consider the following relationship between crime and schooling on commuting zone level s at time t:

$$Crime_{st} = \beta Schooling_{st} + X_{st}\delta + \varepsilon_{st},$$

where $Crime_{st}$ is the crime rate and $Schooling_{st}$ the average years of completed schooling in commuting zone s and year t, X_{st} is a set of control variables and ε_i is an error term.

(a) The authors explain that schooling is unlikely to be exogenous. Explain why this might be true. Explain why exogeneity is important.

(b) The authors have additional information on compulsory schooling laws that regulate the minimum required number of years required to be in school. Denote CSL_{st} as the minimum number of year in school for commuting zone s in year t. They estimate the following model:

$$Crime_{st} = \theta CSL_{st} + X_{st}\gamma + u_{st}.$$

Under what assumption is θ causal? What is the interpretation of θ. Explain when policy-makers would be interested in this parameter?

(c) The authors also estimate the following model:

$$Schooling_{st} = \alpha CSL_{st} + X_{st}\lambda + e_{st}.$$

Under what assumption is α causal? What is the interpretation of α. Could policy-makers be interested in this parameter?

(d) The authors use CSL_{st} as an instrument for $Schooling_{st}$ in a 2-stage-least-squares (2SLS) regression of $Crime$ on $Schooling$. Assuming constant β what are the required assumptions for the instrument to be valid?

(e) Assume that CSL_{st} is correlated with $Schooling_{st}$ conditional on X_{st}. What is the relationship between β, θ and α ?

(f) The authors argue that if *"there is a direct impact of school leaving laws on crime, to the extent that such laws force juveniles to be in a supervised environment rather than roaming the streets, then this can be interpreted as a straightforward incapacitation effect that reduces crime."* Would such a causal link invalidate the use of compulsory schooling age as a instrument for $Schooling_{st}$ to estimate β ? Would this change your interpretation of θ ? Discuss.

(g) Assume that you estimate the following model by 2SLS:

$$Crime_{st} = \beta Schooling_{st} + \gamma X_{st} + \sum_{l} D_s^l + \varepsilon_{st},$$

once again using CSL_{st} as an instrument for $Schooling_{st}$ and where $D_s^l = \mathbf{1}(s = l)$ are location specific dummies (one for each commuting zone). How does this change the source of variation that you are using to estimate β? Discuss the possible advantages of including these dummies. Does this specification change which variables may be included in X_{st}?

3. Assume consumer i has utility of consuming a good: $V - p_i + \varepsilon_i$, where V is the same for all consumers, p_i is the price of the good which varies over consumers and ε_i is an idiosyncratic component, i.e. ε_i also varies over consumers. You as a researcher do not observe V or ε_i but you observe the price facing consumer i, p_i. The consumer buys the product if utility is positive. If the consumer does not buy the good, she receives utility of zero. Assume that the idiosyncratic term is distributed according to the logistic distribution:

$$\Pr(\varepsilon < x) = F(x) = \frac{e^x}{1+e^x}.$$

(a) Derive the probability that consumer i buys the good given the price they face. (Hint: $F(x)$ is symmetric around zero)

(b) As a researcher you do not observe V and therefore need to estimate it. Let y_i be a dummy-variable which takes the value of 1 if consumer i buys the good (and 0 otherwise). Write down the log-likelihood function and derive the first-order condition with respect to V. Give an interpretation to the first-order condition in relation to the aggregate probability of buying the good. (Hint: V can be interpreted as the parameter on a constant term and the model assumes the parameter on price to -1, i.e. $Utility = V + \delta p_i$,where $\delta = -1$).

(c) In the model, the parameter on price is set to -1. How does this affect the interpretation of the estimated value for V?

(d) You are given the output from a logit model estimated by maximum likelihood with the parameter on price constrained to -1. The model is estimated on a sample of consumers facing 4 different prices, $p_i \in (1,2,3,4)$. Estimated parameters are shown in Table 1. The shares of consumers facing each price is displayed in Table 2. What is the predicted share of consumers who choose to buy the good?

Table 1: Logit output

	Coefficient	**Standard error**	**z**
Constant	2.52	0.022	114.42
Price	-1	-	-

Table 2: Frequencies of prices faced by consumers

Price	**Share of sample**
1	0.16
2	0.34
3	0.33
4	0.17