# Exam 4137 – Spring retake 2022

---

IMPORTANT:

- Read the question carefully.
- Always motivate your answers.
- Show knowledge and understanding of the concepts taught in the course.
- Your answers should be as short as possible and as long as necessary.
- Total marks of the examination are 180.
- Each sub-question carries 10 marks.

---

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?

   This is the standard formulation of the observed outcomes, where the researcher observes $Y_i$ and $D_i$ but only one of the potential outcomes. This is in the slides and the student is expected to be able to answer this question fairly easily.

   (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. $E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 0]$

   The three effects are the average treatment effect, the average treatment effect on the treated and the average treatment effect on the untreated. The last "effect" is the observed difference between treated and untreated. The latter is the only one which is always observable as it only involves realized outcomes. i-iii are causal while the last one may not be in the presence of selection.

(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?

This assumption states that treatment is uncorrelated with potential outcomes. A randomized control trial is the obvious design where such an assumption is 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?

Given the assumption that treatment is uncorrelated, all of the effects may be estimated by subtracting simple averages. However, regressing outcome on the treatment dummy would provide standard errors as well. As treatment is uncorrelated with potential outcomes, conditioning on treatment will not change the averages and all 4 differences will coincide.

(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.

This is the conditional independence assumption which is a stronger assumption than the assumption in 1c. The assumption requires us to believe that we have enough controls to believe the residual variation is random. The students may mention stratified random sampling as an empirical example though CIA is often invoked in observational studies as well.

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

The assumption implies that one needs to condition on $X$. This can be achieved by adding $X$ as a control in a regression (+interacting and computing average partial effect) or by weighting X-specific cells. The latter is not part of the curriculum.

(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?

No. If treatment is independent of potential outcomes this will continue to hold conditional on predetermined covariates.

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 $\varepsilon_i$ 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.

Schooling and crime likely reflect neighborhood and individual circumstances which may determine both crime and schooling. Without exogeneity one cannot interpret $\beta$ as the causal effect of schooling on crime even conditional on covariates, $X_i$.

(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 $\theta$ causal? What is the interpretation of $\theta$. Explain when policy-makers would be interested in this parameter?

$\theta$ is causal if the minimum schooling requirements are uncorrelated with $u$ conditional on observable neighborhood characteristics. In that case, $\theta$ is the causal effect of schooling requirements on crime. If policy-makers consider changing schooling requirements to alter crime levels, this is a policy relevant parameter.

(c) The authors also estimate the following model:

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

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

The answer to this question is basically the same as the previous question. If minimum schooling requirements are uncorrelated with $e$, then $\alpha$ can be interpreted as the effect of schooling laws on completed years of schooling. If policy-makers are interested in increasing schooling (for whatever reason) then $\alpha$ is policy-relevant.

(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 $\beta$ what are the required assumptions for the instrument to be valid?

The constant effect IV requires the existence of a first stage ($\alpha \neq 0$), conditional exogeneity of the instrument and the exclusion restriction that the effect of minimum schooling on crime may only go through schooling. Monotonicity is not required due to the assumption of constant $\beta$.Students should be aware of this destinction, and if monotonicity is mentioned as an requirement, this should count negatively.

(e) Assume that $CSL_{st}$ is correlated with $Schooling_{st}$ conditional on $X_{st}$.What is the relationship between $\beta, \theta$ and $\alpha$ ?

$\alpha$ and $\theta$ are the first stage and reduced form respectively. The assumption ensures existence of a first stage such that $\alpha \neq 0$. Thus $\beta = \frac{\theta}{\alpha}$.

(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 $\beta$? Would that change your interpretation of $\theta$? Discuss.

This example amounts to breaking the exclusion restriction as there would be a causal effect of minimum schooling laws which would not flow through the schooling channel. This effect however does not invalidate the causal interpretation of the reduced form parameter, $\theta$, as a causal effect of minimum schooling laws on crime.

(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 $\beta$? Discuss the possible advantages of including these dummies. Does this specification change which variables may be included in $X_{st}$?

This specification is an implementation of the fixed effects model which controls for all time-constant location specific effects, both observed and unobserved. Thus the remaining variation is within location over time, and differences in unobservable time-constant characteristics across locations cannot bias estimates of $\beta$. Due to the fixed effects $X_{si}$ cannot contain time-constant location specific effects. (One must include the dummies in the first stage as well, but this is left implicit in the 2SLS-specification and the students are not required to mention this.)

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 $\varepsilon_i$ is an idiosyncratic component, i.e. $\varepsilon_i$ also varies over consumers. You as a researcher do not observe $V$ or $\varepsilon_i$ 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)

The probability of buying the good equal the probability that net utility is positive

$$\begin{aligned} Pr(\text{Buys Good}) &= Pr(V - p_i + \varepsilon_i > 0) = Pr(-\varepsilon < V - p_i) \\ &= Pr(V - p_i < \varepsilon) = F(V - p_i) = \frac{e^{V-p_i}}{1+e^{V-p_i}} \end{aligned}$$

This is a very slight rewriting to the latent utility framework where the parameter on price is set to -1. This is a basic logit model. This probability is stated in equation 7.15 in the textbook with different notation.

(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 $\delta = -1$).

$V$ is given as the parameter on the constant term. The log likelihood is given as:

$$\begin{aligned} \log L(V) = & \sum_{i=1}^{N} y_i \log F(V - p_i) + (1 - y_i) \log(1 - F(V - p_i)) \\ = & \sum_{i=1}^{N} y_i \left[V - p_i - \log(1 + e^{V-p_i})\right] \\ & + (1 - y_i) \left[\log 1 - \log(1 + e^{V-p_i})\right] \\ = & \sum_{i=1}^{N} y_i [V - p_i] - \log(1 + e^{V-p_i}) \end{aligned}$$

The first-order condition:

$$\frac{\partial \log L(V)}{\partial V} = \sum_{i=1}^{N} \left[ y_i - \frac{e^{V-p_i}}{1+e^{V-p_i}} \right] = 0 \Leftrightarrow \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} \left[ \frac{e^{V-p_i}}{1+e^{V-p_i}} \right]$$

The first-order condition states that the likelihood is maximized when the predicted number of buyers is equal to the actual number of buyers . Again this formula is directly stated in the textbook (p220 eq 7:14) with the slight modification that $V$ is a constant. Thus points should be given for realizing the link between utility maximizing and the logit model given distributional assumptions.

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

The normalization of the parameter on price makes the utility interpretable in monetary terms . This is discussed in section 7.1.3 in the textbook.

(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?

For each price the students must compute $Pr(\text{Buys Good}) = \sum_q Pr(p_i = q)\frac{e^{V-q}}{1+e^{V-q}}$, which is simply the weighted mean of predicted buyers at each proce level.

$$\begin{aligned} Pr(\text{Buys Good}) &= \sum_q Pr(p_i = q)\frac{e^{V-q}}{1+e^{V-q}} \\ &= 0.16 \times 0.82 + 0.34 \times 0.63 + 0.33 \times 0.38 + 0.17 \times 0.19 \\ &\approx 0.5 \end{aligned}$$

Students will likely round at different levels, which is fine.

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 |