Exam 4137 – Spring 2026

IMPORTANT: Always motivate your answers. Show knowledge and understanding of the concepts taught in the course. Make sure not only to repeat material from lectures and curriculum, but engage with the empirical examples and show that you understand what it means in the particular setting in addition to using math. Your answers should be as short as possible and as long as necessary. Each lettered subtask carries 10 points, for a total of 150.

---

1. (20 points) Short questions.

   (a) TRUE or FALSE - No points without an explanation!

      i. In order for OLS regression coefficients to be unbiased, we must assume that the error term is homoskedastic

      ii. The common trends assumption can be tested by comparing pre-trends

   (b) Interpret the following assumptions. $Y(d,z)$ and Y_{it}^{d} are potential outcomes under treatment state d and instrument value$z$, D_i is actual treatment and X_{it} are covariates for observation i at time t.

      i. $Y_i(d,z) = Y_i(d)$

      ii. $\mathbb{E}(Y_{i1}^{0} - Y_{i0}^{0} \mid D_i, X_{it}) = \mathbb{E}(Y_{i1}^{0} - Y_{i0}^{0} \mid X_{it})$

2. (80 points) You work with the application and admissions system for high schools in the Oslo Education Agency. You are tasked with analyzing the causal effect of attending different schools because your employer is worried about differences in school quality. You have a dataset on an entire cohort of Oslo students at academic track high schools that contains information on which high school $D \in (1,2,...,15)$ they attended, and annual earnings at age 30, Y_i. You are interested in well-defined average causal effects on earnings of attending one school j relative to another j'.

   (a) Using potential outcomes notation, define and interpret the potential and realized outcomes.

   (b) What do you need to assume about potential outcomes for a simple difference in means between students attending school j and j' to be causal? What causal effect is that?

   (c) In the reasonable case that independence does not hold, will you over- or underestimate the true effect? Use potential outcomes notation to express the bias and explain why.

   (d) The dataset also contains GPA_i, the grade point average of each student from compulsory school. Explain how you could use this to weaken the independence assumption from b)

   (e) What regression would you run to exploit the weaker assumption from d)? What average treatment effect would you estimate?

(f) Still unconvinced about the assumption you need to make, you dig up data on the admissions thresholds c_j to each of the 15 schools. Let schools now be indexed by their cutoff so that school 1 is the highest-threshold school, $c_1 \geq c_2 \geq ... \geq c_{15}$.You can also construct $Z_{ij} = GPA_i \geq c_j$, a dummy for whether student i's grades are good enough to be accepted to school j. Focus on the causal affect of attending the highest-ranked school ($D_i = 1$), among students with GPA better than the cutoff of the second-highest ranked school ($GPA_i > c_2$), who applied to school 1 above school 2, using the instrument for threshold crossing to the highest-ranked school ($Z_{i1} = GPA_i > c_1$). Explain what assumptions you need to make to utilize this in a regression discontinuity framework.

(g) You estimate the following system using 2SLS

$$\begin{aligned} Y_i &= \alpha + \beta 1(D_i = 1) + \gamma(GPA_i - c_1) + \theta(GPA_i - c_1)Z_{ij} + \varepsilon_i \\ 1(D_i = 1) &= \zeta + \pi Z_{i1} + \delta(GPA_i - c_1) + \eta(GPA_i - c_1)Z_{i1} + \mu_i \end{aligned}$$

Explain the roles of the $\gamma, \theta, \delta$and η terms. How do you interpret the coefficients π and β?

(h) Explain the logic behind at least two different robustness checks you could perform to increase your trust in the assumptions behind this estimate.

3. (50 points) In order to fully understand how the application system to high school shapes outcomes, you also need to understand students' (and parents) preferences for schools. Next, you are tasked with surveying the literature on school preferences, and you find the influential paper *"What Parents Want: School Preferences and School Choice*" by Simon Burgess, Ellen Greaves, Anna Vignoles, Deborah Wilson, published in the Economic Journal in 2015. In this paper, the authors use high school application data from England to estimate a conditional logit model of preferences for school characteristics. Their model is the classic conditional/multinomial logit model, where

$$P_{is} = \frac{e^{X_{is}\beta + W_i X_{iz}\gamma}}{\sum_{l=1}^{n} e^{X_{il}\beta + W_i X_{il}\gamma}}$$

governs the probability that student i chooses school s. The specification includes school characteristics such as school quality and distance (in X_{is}), and indicators for family socioeconomic status (in W), which allows different SES-group families to have different preferences.

(a) Derive an expression for the log likelihood from this probability, and explain how you would use maximum likelihood to estimate the parameters based on this expression.

(b) A simplified excerpt of their results is found in Table 1. Focusing on the first column of the table and one or two coefficients of choice, what can you learn about the preferences of parents in the highest SES group from signs and

Table 1: School choice preference estimates from Burgess et al., 2015

	Reference group (highest SES)	Interaction: lowest SES
Academic quality: KS2 mean points score over 3 years (SD units)	0.270 (0.087)	-0.363 (0.108)
Share free school meals (SD units)	-0.539 (0.161)	0.939 (0.185)
Share English as additional language (SD units)	0.171 (0.109)	-0.044 (0.152)
Share white (SD units)	0.233 (0.115)	-0.272 (0.154)
Any special educational need (SD units)	-0.230 (0.075)	0.165 (0.087)
Faith school (binary indicator)	-0.817 (0.111)	-0.071 (0.144)
Distance rank	-0.644 (0.091)	0.261 (0.108)
N	42,047	
No. households	7,905	

relative magnitudes? How are preferences of the lowest SES group different?

(c) How much longer is a high-SES parent willing to send their kids to get a 1 SD increase in academic quality?

(d) An elasticity of demand measures how much relative demand for a school increase when a characteristic x_k increase by 1%, e.g. $\frac{\partial P_{is}}{\partial x_{isk}} \frac{x_{isk}}{P_{is}} = \beta_k x_{isk}(1-P_{is})$. The authors report elasticity estimates based on Partial Effects at the Average (PEA) , i.e. evaluating the elasticity for the average school and individual. An excerpt of the authors' elasticity estimates is found in Table 2, which shows that the elasticity of demand with respect to school quality is 4 for high-SES students in row 1. Use the numbers in the Tables to compute this elasticity and interpret it.

(e) Do you think that this partial elasticity at the average is a reasonable way to summarize the heterogeneous effects from this model? What is the alternative, and what is the difference between the two?

Table 2: Demand elasticity estimates from Burgess et al. (2015)

	Mean school characteristic X	Elasticity (Lowest SES)	Elasticity (Highest SES)
Academic quality (KS2, SD units)	17.7	-1.38	4.00
		(0.631)	(0.80)
Free school meals (SD units)	1.2	0.40	-0.54
		(0.057)	(0.096)
English as additional language (SD units)	0.9	0.10	0.13
		(0.052)	(0.070)
White (SD units)	3.9	-0.13	0.76
		(0.259)	(0.308)
Any special educational need (SD units)	1.9	-0.10	-0.37
		(0.062)	(0.084)
Distance rank	4.0	-0.26	-0.55
		(0.063)	(0.061)

The mean number of schools in the choice set is 6.1. We therefore fix $\bar{P} = 1/6.1 = 0.164$.