Final Exam

January 13th, 2024, 9:00 am—January 15th, 2024, 9:00 am

Be careful not to cheat at exams!

Exam cheating is for example if you:

- Copy text written by other people without making use of quotation marks and source referencing, so that it may appear to be your own text
- Use the ideas or thoughts of others without making use of source referencing, so it may appear to be your own idea or your thoughts
- Reuse parts of a written paper that you have previously submitted and for which you have received a passing grade without making use of quotation marks or source references (self-plagiarism)
- Receive help from others in contrary to the rules laid down in the Faculty of Social Sciences' common part of the curriculum on cooperation/sparring

You can read more about the rules on exam cheating on your Study Site and in the Faculty of Social Science's common part of the curriculum (https://socialsciences.ku.dk/education/studentservices/regulations/curricula/). **Exam cheating is always sanctioned by a written warning and expulsion from the exam in question. In most cases, the student will also be expelled from the University for one semester.**

(This exam consists of 4 pages in total.)

Introduction and Formal Requirements

The exam consists of two parts. Part I involves (re)submitting one of the projects that you have had the opportunity to work and receive peer feedback on during the semester. Part II contains a new assignment. These parts are weighted approximately equally in the overall assessment.

Formal Requirements:

1. You are allowed and strongly encouraged to work in groups of at most three students. Formal requirements and assessment criteria are the same for individuals and groups.

2. You are only assessed based on what you hand in for the final exam. Hence, projects not selected for the final are not included in the assessment.

3. Your answer to Part I must satisfy the formal requirements (pages, font, spacing, margins, etc.) as stated in the project text (with no character limit on code). For Part II there is no character limit, but it will be to your advantage to be brief and to the point.

4. Formulate your answers to the exam questions in English only and compile them into a single pdf file. Include a frontpage as well as your answers to all parts of the exam.

5. On the frontpage, you must provide a count of the characters (including spaces, math, and formulas, but excluding tables and figures) for each of the exam parts. There is no globally accepted way of counting mathematical characters. Copy-paste the text including math and formulas into `https://charcounter.com/` or `https://charactercounttool.com` and report that number.

6. If you submit as a group, you must specify on the frontpage who is responsible for each individual section of each part of the exam by their name. *All group members must contribute to each part of the exam* (but not necessarily each section/question). Do not write "Munk-Nielsen was responsible for every odd word and Sørensen every even word," or the like.

7. Name your pdf file using the group members last names sorted alphabetically (e.g. `munk-nielsen_sorensen.pdf`).

8. Along with your exam answers, upload a compressed zip-folder with all the Python code you used to obtain your results for each part. Your code must be transparent and run with only minor modifications (e.g. changing relevant paths).

9. Submit your pdf and zip files using the Digital Exam portal at eksamen.ku.dk before the end of the exam. Late submissions by email or other means will not be accepted.

Part I

Repeat Project

The project to be (re)submitted is

Project 2: High-dimensional Linear Models and Convergence in Economic Growth.

See the course website on Absalon for data and project description.

Part II

New Assignment

Heckmania

Consider the following data-generating process (DGP), composed by the equations

$$s_i = \mathbf{1}\left(\delta_o x_i + v_i \in (a, b)\right), \qquad \text{(“selection”)}$$
$$y_i = \beta_o x_i + u_i, \qquad \text{(“outcome”)}$$

where the unobservables u_i and v_i satisfy

$$\begin{aligned} \mathrm{E}\left[u_i \mid v_i = v\right] &= \gamma_o v, \\ v_i \mid x_i = x &\sim \mathrm{N}\left(0, \sigma^2\left(x, \alpha_o, \eta_o\right)\right), \end{aligned}$$

and u_i and x_i are *mean* independent conditional on v_i, i.e. $\mathrm{E}[u_i|x_i = x, v_i = v]$ does not depend on x. The remaining information structure goes as follows:

- a and b are known constants satisfying $-\infty \leqslant a < b \leqslant +\infty$, which are specified below.
- x_i is a scalar random variable, which is continuously distributed with full support ($\mathbb{R}$).
- The functional form of $\sigma(x, \alpha, \eta)$ is known to be $\mathrm{e}^{\alpha+\eta x}$ [such that $\sigma^2(x, \alpha, \eta) = \mathrm{e}^{2(\alpha+\eta x)}$].
- The (true) model parameters $\alpha_o, \beta_o, \gamma_o, \delta_o$ and η_o are unknown elements of $\mathbb{R}$.
- The pair of s_i and x_i is (always) observable.
- We only observe y_i when $s_i = 1$.

Suppose for now that $a = -\infty$ and $b = +\infty$.

1. Show that β_o is identified. Explain your steps. *Hint: What is observable and what is not?*

2. Granted a random sample of observables of size N, suggest a consistent estimator of β_o, and argue its consistency as $N \to \infty$.

3. Is your choice of estimator also asymptotically efficient? Provide your reasoning.

Keeping $b = +\infty$, suppose for the remaining tasks that $a = 0$.

4. Is β_o still identified based on the approach suggested earlier? Provide your reasoning.

5. Show that selection probability function $\mathrm{P}(s_i = 1 \mid x_i = x)$ in terms of model parameters is given by

$$\mathrm{P}(s_i = 1 \mid x_i = x) = \Phi\left(\frac{\delta_o x}{\mathrm{e}^{\alpha_o + \eta_o x}}\right).$$

   Explain your steps.

6. Show that the triplet $(\alpha_o, \delta_o, \eta_o)$ is not identifiable based on the joint distribution of s_i and x_i.

In what follows you are informed that α_o is actually zero, such that the only unknown parameters are $(\beta_o, \gamma_o, \delta_o, \eta_o)$. The accompanying dataset `data.csv` contains $N = 1,000$ realizations of independent triplets (y_i, x_i, s_i) in agreement with the above DGP and observational rules (for the current a and b).

7. Estimate the model parameters associated with the *selection* equation. Report the point estimates alongside standard errors and t ratios. Explain the steps involved. What assumptions are needed for your standard errors to be valid?

8. Derive an expression for the partial effect(s) of changing the regressor x_i on the probability of selection. *Hint: How many partial effects are there?*

9. Formally test the claim: "x_i has zero partial effect on selection."

10. Work out an expression for the conditional mean function $\mathrm{E}[y_i \mid x_i = x, s_i = 1]$ in terms of model parameters. Explain the steps involved.

11. Estimate the model parameters associated with the *outcome* equation. Report the point estimates alongside standard errors and t ratios. Explain the steps involved and to what extent your standard errors are valid.