Final Exam

January 11th, 2025, 09:00 am—January 13th, 2025, 09:00 am

Use of AI tools is permitted. You must explain how you have used the tools. When text is solely or mainly generated by an AI tool, the tool used must be quoted as a source.

Be careful not to cheat at exams! Exam cheating is for example if you:

- Copy other people's texts 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 pass grade without making use of quotation marks or source references (self-plagiarism)
- Receive help from others in contrary to the rules in the Faculty of Social Science's common part of the curriculum

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 5 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/` and report the count.

6. If you submit as a group, then you must specify on the frontpage who is responsible for each individual section of each part of the exam by their exam numbers. *All group members must contribute to each part of the exam* (but not necessarily each section/question). Do not write "Student 4 was responsible for every odd word and Student 20 every even word," or the like.

7. Name your PDF file using the exam numbers in ascending order (e.g. `4_20.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 "out-of-the-box" 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 1:** Linear Panel Data and Production Technology.

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

Part II

New Assignment

The "tiboT" Model

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

$$\begin{aligned} Y_i &= \min\{0, Y_i^*\}, \\ Y_i^* &= \alpha_o + \beta_o X_i + U_i, \quad \text{and} \\ U_i \mid X_i = x &\sim \mathrm{N}\left(\mu_o, \sigma_o^2\right), \end{aligned}$$

where

- Y_i is the observed outcome;
- Y_i^* is the latent outcome;
- X_i is a regressor, known to be continuously distributed with full support ($\mathbb{R}$);
- U_i is an unobservable; and,
- the (true) parameters $(\alpha_o, \beta_o, \mu_o) \in \mathbb{R}^3$ and $\sigma_o \in (0, \infty)$ are unknown and, for now, otherwise unrestricted.

1. For any non-random number $y^* \in \mathbb{R}$, express the cumulative distribution function (CDF) $F_{Y_i^*|X_i}(y^*|x)$ of the latent outcome Y_i^* (conditional on $X_i = x$, and evaluated at y^*) in terms of the true model parameters $(\alpha_o, \beta_o, \mu_o, \sigma_o)$. Explain the steps involved in your derivation and comment on the nature of the CDF $F_{Y_i^*|X_i}(\cdot|x)$.

2. For any non-random number $y \in \mathbb{R}$, express the CDF $F_{Y_i|X_i}(y|x)$ of the observable Y_i (conditional on $X_i = x$, and evaluated at y) in terms of $F_{Y_i^*|X_i}(\cdot|x)$. Explain the steps involved in your derivation and comment on the nature of the CDF $F_{Y_i|X_i}(\cdot|x)$.

3. Show that, in its current state, the model is *not* identifiable.

You are now told that α_o is actually zero, so that the only remaining unknown parameters are $(\beta_o, \mu_o, \sigma_o)^\top =: \boldsymbol{\theta}_o \in \mathbb{R}^2 \times (0, \infty)$.

4. Argue that the model-implied densities $f_{Y_i|X_i}(\cdot|x; \boldsymbol{\theta})$ take the form

$$f_{Y_i|X_i}(y|x; \boldsymbol{\theta}) = \left[\frac{1}{\sigma}\varphi\left(\frac{y - (\mu + \beta x)}{\sigma}\right)\right]^{\mathbf{1}(y<0)} \left[\Phi\left(\frac{\mu + \beta x}{\sigma}\right)\right]^{\mathbf{1}(y=0)},$$

where $\Phi(z) := \int_{-\infty}^{z} \varphi(t)\mathrm{d}t$ denotes the standard normal CDF, and $\varphi(z) := (2\pi)^{-1/2}\mathrm{e}^{-z^2/2}$ denotes the standard normal probability density function (PDF).
Hint: Your arguments for the tiboT may look a lot like those used for the Tobit.

5. Show that the tiboT model for the outcome Y_i — stated in terms of the non-constant regressor X_i and the unobservable U_i — is actually equivalent to the usual Tobit model stated in terms of the sign-flipped outcome $\widetilde{Y}_i := -Y_i$ for a particular choice of regressors and model unobservable.

The dataset `data.csv` contains $N = 1,000$ realizations of the pair (Y_i, X_i), generated independently of each other and in accordance with the above DGP.

6. Using the provided dataset, estimate the model parameters $\boldsymbol{\theta}_o = (\beta_o, \mu_o, \sigma_o)^\top$ and report both point estimates alongside standard errors and t ratios, gathered in a table. Explain the steps involved and justify your choices. In particular, what assumptions are needed for your standard errors to be valid?

7. Derive an expression for the partial effect(s) of changing the regressor X_i on the probability of "censoring" $\mathrm{P}(Y_i = 0|X_i = x)$.
Hint: How many partial effects are there?

8. Formally test the claim: *X_i has zero partial effect on the probability of "censoring."*

9. Show that

$$\mathrm{E}[Y_i|X_i = x] = \left[1 - \Phi\left(\frac{\mu_o + \beta_o x}{\sigma_o}\right)\right](\mu_o + \beta_o x) - \sigma_o \varphi\left(\frac{\mu_o + \beta_o x}{\sigma_o}\right).$$

Explain the steps involved.
Hint: The answer to Question 5 may come in handy.

10. Formally test the claim: *X_i has constant partial effect on the conditional mean.*

You are now told that the tiboT model as a whole is *in*correctly specified, but that one particular feature, namely the model for the conditional mean $\mathrm{E}[Y_i|X_i = x]$, *is* correctly specified.

11. Based on this new knowledge, devise a procedure for estimating the model parameters and attaching standard errors to these estimates. Provide your reasoning.
Hint: A detailed description will suffice. You do not need to report new estimation results.