Final Exam

December 18$^{\text{th}}$, 10 am—December 20$^{\text{th}}$, 10 am

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 laid down in part 4.12 of the Faculty of Social Science's common part of the curriculum on cooperation/sparring

You can read more about the rules on exam cheating on your Study Site and in part 4.12 of the Faculty of Social Science's common part of the curriculum.
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 on as well as receive and provide peer feedback on during the semester. Part II contains a new assignment. The two parts are weighted equally in the overall assessment.

Formal Requirements:

1. You are allowed and strongly encouraged work in groups of up to three students.[1] The formal requirements and the 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, the projects you have completed and submitted for peer review during the semester that were not selected are not included in the assessment.

3. Your answer to the repeat project (Part I) must satisfy the page limitations as stated in the project text (with no character limit on any submitted Python code). For the new assignment (Part II) there is no character/page limit, although 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.

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

7. Name this file according to your exam number and only that number (e.g. `12.pdf` if student `12` is working solo, or `12_34.pdf` if `12` and `34` are in a two-person group). Submit your answers using the Digital Exam portal at `eksamen.ku.dk`.

8. Along with your report, please also upload a compressed zip-folder with all the Python code you used to obtain your results for Part II. Make sure that your code runs with only minor modications (e.g., changing relevant paths).

[1] Groups of four are only allowed if you have been granted special permission to form such a group from the lecturers.

Part I

Repeat Project (50 %)

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 (50 %)

Nonlinear Regression

Consider the regression model

$$y_i = h\left(\mathbf{x}_i\boldsymbol{\beta}\right) + \varepsilon_i, \quad \varepsilon_i|\mathbf{x}_i \sim \mathrm{N}(0, \sigma_\varepsilon^2), \tag{1}$$

relating

- y_i, the scalar outcome;
- $\mathbf{x}_i := (x_{i1}, x_{i2}, x_{i3})$, a triplet of regressors with $x_{i1} \equiv 1$, x_{i2} binary (0/1), and x_{i3} uniformly distributed on $(0; 1)$;
- ε_i, a random component unobservable to the researcher ("error term").

The model parameters involve the constants $\boldsymbol{\beta} \in \mathbb{R}^3$, $\sigma_\varepsilon \in \mathbb{R}_{++}$, and the function $h : \mathbb{R} \to \mathbb{R}$. You are for now only informed that h is continuously differentiable—these quantities are otherwise complete unknowns. The dataset `cross_section.csv` contains $N = 1,000$ independent samples $\{(y_i, \mathbf{x}_i)\}_{i=1}^N$ consistent with (1).

1. Derive the log-likelihood contribution ℓ_i for the model (1) viewed as a function of $(\boldsymbol{\beta}, \sigma_\varepsilon, h)$.

2. Suppose we know that $h(z) = \gamma z$ for some $\gamma \neq 0$. Explain why the parameters $(\boldsymbol{\beta}, \gamma)$ cannot be identified. PROOF BY CONTRADICTION PLUG THEM IN THE LOG-LIKEHOOD CONTRIBUTION -> DO THEY GIVE A DIFFERENT VALUE? SUPPOSE I KNOW THE TRUE PARAMETERS. THEN PRESENT SOME ALTERNATIVES. ARE THE LOG-LIKEHOOD CONT. DIFFERENT?

   You are now informed that either $h(z) = \exp(-z)$ or $h(z) = 3\Phi(z)$ (where Φ is the standard normal cdf). BASICALLY EXPLAIN WHICH GIVES THE HIGHEST LOG-LIKELIHOOD CONTRIBUTIONS

3. Using the cross section provided, estimate the model parameters for the two specifici- ations of h and compare them. Determine which h was used to generate the data.

4. Using this h, estimate and report the average partial effects (APEs) on $E[y_i|\mathbf{x}_i = \mathbf{x}^0]$ at $\mathbf{x}^0 = (1, 1, x_3)$ for each value $x_3 = 0, .1, .2, \ldots, 1$ based on the model (1) in addition to standard errors hereof.

   Consider next the model with panel structure,

$$y_{it} = h(\mathbf{x}_{it}\boldsymbol{\beta} + c_i) + \varepsilon_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N, \tag{2}$$

   where $c_i|\mathbf{x}_i \sim \mathrm{N}(0, \sigma_c^2)$ and the $\{\varepsilon_{it}\}_{t=1}^T$ are independently distributed $\mathrm{N}(0, \sigma_\varepsilon^2)$ over time conditional on $(\mathbf{x}_i, c_i)$. Here

   - $\mathbf{y}_i := (y_{i1}, \ldots, y_{iT})$ is a vector of outcomes;
   - $\mathbf{x}_i := (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ gathers the possibly time-varying regressors $\mathbf{x}_{it}$, where again $\mathbf{x}_{it} = (1, x_{it2}, x_{it3})$ with x_{it2} binary and x_{it3} continuous;
   - c_i is an individual-specific time-invariant unobservable random component; and,
   - $\{\varepsilon_{it}\}_{t=1}^T$ are time-varying unobservable random components ("error terms").

   The model parameters now involve $(\boldsymbol{\beta}, \sigma_\varepsilon, \sigma_c, h)$, where $\sigma_c \in \mathbb{R}_{++}$ and $h : \mathbb{R} \to \mathbb{R}$ is some, as of yet, unknown function. The dataset `panel.csv` contains $N = 100$ independent samples $\{\{(y_{it}, \mathbf{x}_{it})\}_{t=1}^T\}_{i=1}^N$ consistent with (2) with $T = 10$.

5. Show that the log-likelihood contribution ℓ_i for the model (2) viewed as a function of $(\boldsymbol{\beta}, \sigma_\varepsilon, \sigma_c, h)$ equals

$$-T \log \sigma_\varepsilon + \log \left[\int_{-\infty}^{\infty} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T [y_{it} - h(\mathbf{x}_{it}\boldsymbol{\beta} + \sigma_c c)]^2 \right) \phi\left(c\right) \mathrm{d}c \right] - \frac{T}{2} \log\left(2\pi\right),$$

   and explain your derivation.

6. An alternative assumption concerning $\{\varepsilon_{it}\}_{t=1}^T$ is that $\varepsilon_{it}|\mathbf{x}_{it}, c_i \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ for each t (i.e. conditioning on $\mathbf{x}_{it}$ rather than $\mathbf{x}_i$). Briefly discuss this assumption in contrast to the one specified earlier. Would it lead to the same log-likelihood contribution? Justify your answer.

   You are informed once again, that either $h(z) = \mathrm{e}^{-z}$ or $h(z) = 3\Phi(z)$.

7. Using the panel data provided, estimate the model parameters for both specifications and compare them. Determine which h was used to generate the data.

8. Test the following two null hypotheses:

   (a) $\beta_1 = \beta_2 = \beta_3$,

   (b) $\beta_2 = \beta_3$.