Written Exam at the Department of Economics
Winter 2019–20

Advanced Microeconometrics

Final Exam

January 4th, 2020 (13:00–16:00)
(3-hour closed book exam)

Answers only in English.

This exam question consists of 6 pages in total.

Falling ill during the exam
If you fall ill during an examination at Peter Bangs Vej, you must:

- Contact an invigilator who will show you how to register and submit a blank exam paper.
- Leave the examination.
- Contact your GP and submit a medical report to the Faculty of Social Sciences no later than five (5) days from the date of the exam.

Be careful not to cheat at exams!
You cheat at an exam, if during the exam, you:

- Make use of exam aids that are not allowed.
- Communicate with or otherwise receive help from other people.
- 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.
- Or if you otherwise violate the rules that apply to the exam.

Note: The percentage weights indicated on each problem should only be regarded as indicative. The final grade will ultimately be based on an overall assessment of the quality of the answers to the exam questions in their totality.

Problem 1 (50%)

Consider the following censored regression model, for a sample of individuals $i = 1, \ldots, N$:

$$y_i = \max\{0, y_i^\star\}, \qquad y_i^\star = x_i'\beta + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right), \tag{1}$$

where the explanatory variables are contained in the $K \times 1$ vector x_i, and are related to the latent variable $y_i^\star$ through the vector of regression coefficients β.

Question 1.1: Discuss *briefly* the identification of the model (without any derivations). In particular, explain if σ^2 is identified, and why.

Question 1.2: Show that the conditional expectation of the observed outcome is

$$\mathrm{E}[y_i \mid x_i] = x_i'\beta\, \Phi\left(\frac{x_i'\beta}{\sigma}\right) + \sigma\, \phi\left(\frac{x_i'\beta}{\sigma}\right), \tag{2}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote, respectively, the cumulative distribution function(CDF) and the probability density function (PDF) of the standard normal distribution $\mathcal{N}(0, 1)$.

Hint 1: To get started, remember that $\mathrm{E}[y_i \mid x_i] =$
$\mathrm{E}[y_i \mid x_i, y_i = 0] \Pr(y_i = 0 \mid x_i) + \mathrm{E}[y_i \mid x_i, y_i > 0] \Pr(y_i > 0 \mid x_i)$.

Hint 2: If $z \sim \mathcal{N}(0, 1)$, then for any constant $\alpha \in \mathbb{R}$ it holds that $\mathrm{E}[z \mid z > \alpha] = \phi(\alpha)/[1 - \Phi(\alpha)]$.

Question 1.3: One of your colleagues suggests you construct an estimator of $\theta = (\beta', \sigma^2)'$ based on the following optimization problem:

$$\widehat{\theta} = \arg\min_{\theta} \left[\frac{1}{N} \sum_{i=1}^{N} \widehat{m}(y_i, x_i; \theta, u_{iM})\right]^2 \tag{3}$$

where $\widehat{m}(y_i, x_i; \theta, u_{iM})$ is simulated using a sample of M random draws $u_{iM} = \{u_i^{(1)}, \ldots, u_i^{(M)}\}$ from the standard normal distribution, for each $i = 1, \ldots, N$.

Describe the principle of the estimation method your colleague is referring to. As part of your answer, you are expected to provide and justify a possible expression of $\widehat{m}(y_i, x_i; \theta, u_{iM})$ *[hint: you may or may not use the result in Eq.* (2) *to do this]*, and to outline the steps of the corresponding estimation approach.

Question 1.4: How do you recommend to choose the number of random draws M in Question 1.3? In particular, explain how this number affects the bias of the estimator *(no derivations required)*.

Question 1.5: How would you modify the optimization problem in Eq. (3) to improve the efficiency of the estimator $\widehat{\theta}$? Describe *briefly* the corresponding approach.

Problem 2 (30%)

Consider the following two-parameter model

$$y \sim \mathcal{N}(\theta_1 + \theta_2, 1)\,, \tag{4}$$

with prior distributions $\theta_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\theta_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Question 2.1: Derive the conditional distributions $p(\theta_1 \mid \theta_2, y)$ and $p(\theta_2 \mid \theta_1, y)$.

Hint: Given the symmetry of the problem, you need to do the derivations only once.

Question 2.2: Outline the different steps of a Gibbs sampler that can be designed to produce random draws from the posterior distribution of θ_1 and θ_2. Be as precise as possible.

Question 2.3: Assuming we observe $y = 4$ and we set $\mu_1 = \mu_2 = 50$ and $\sigma_1^2 = \sigma_2^2 = 100$, we run the Gibbs sampler derived in Question 2.2 for 1,000 iterations. The corresponding trace plots of the two parameters θ_1 and θ_2, as well as the trace of their sum $\theta_1 + \theta_2$, are shown in Fig. 2.1.

Does the algorithm converge in any sense? Comment on the trace plots and explain the results, both intuitively and formally.

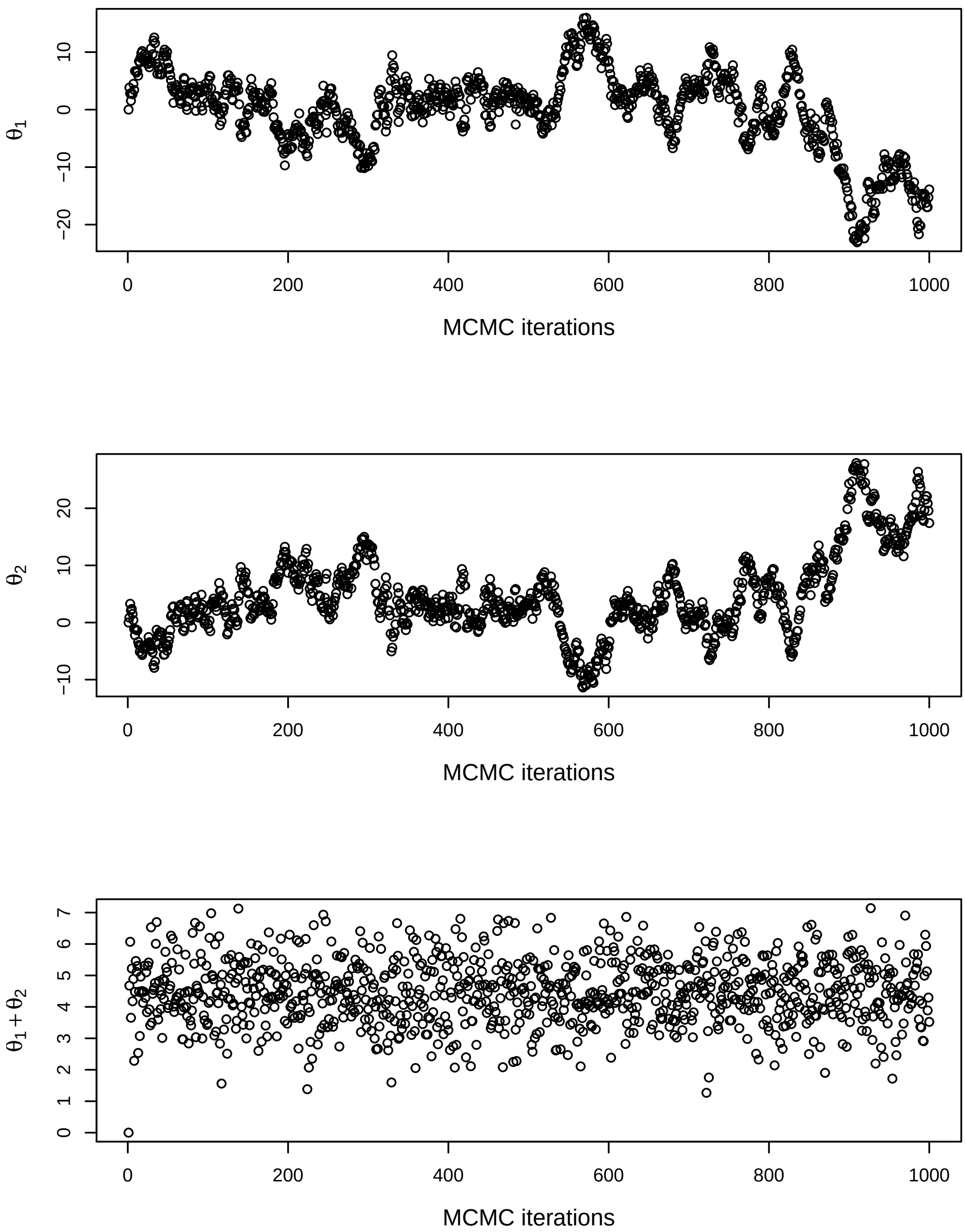

Figure 2.1: Trace plots of the Gibbs sampler for the parameters θ_1, θ_2, and their sum $\theta_1 + \theta_2$.

Problem 3 (20%)

Consider the following MATLAB functions:

```matlab
function [x] = simul1(n, fun, a, b)
    z = betarnd(a, b, n, 1);
    x = mean(fun(z));
end

function [x] = simul2(n, fun, a, b)
    z = rand(n, 1);
    x = mean(fun(z) .* betapdf(z, a, b));
end
```

and the following piece of code:

```matlab
rng(123);
h = @(x) (x - 3).^2;
n = 10000;
fprintf('simul1 output = %6.4f\n', simul1(n, h, 2, 3));
fprintf('simul2 output = %6.4f\n', simul2(n, h, 2, 3));
```

which produces the following output:

```
simul1 output = 6.7954
simul2 output = 6.7502
```

Question 3.1: Express in mathematical terms what these two functions do. You should just provide a few equations to answer this question. Be explicit about the notation.

[Note: The `MATLAB` *function* `betarnd(a, b, m, n)` *produces a $m \times n$ matrix of random draws from the Beta distribution with shape parameters a and b, while the function* `betapdf(z, a, b)` *returns the probability density function of the corresponding Beta distribution evaluated at each entry of z.]*

Question 3.2: Explain precisely the two approches implemented by the functions `simul1()` and `simul2()`, and why the corresponding results look similar.