Written Exam at the Department of Economics
Winter 2019–20

Advanced Microeconometrics

Retake Exam

February 14th, 2020 (9:00–12:00)
(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 discrete choice model for a sample of individuals $i = 1, \dots, N$:

$$y_i = \underset{j\in\{1,\dots,J\}}{\arg\max}\{u_{ij}\}, \qquad u_{ij} = v_{ij} + \varepsilon_{ij}, \qquad v_{ij} = x_{ij}'\beta + w_i'\gamma_j \tag{1}$$

where $y_i \in \{1, \dots, J\}$ is the alternative chosen by individual i and u_{ij} denotes the utility derived from choosing alternative j for individual i. Utility, u_{ij}, is composed of an observed and deterministic part of utility, v_{ij}, and unobserved and random component of utility, ε_{ij}.

The observed part of utility $v_{ij} = x_{ij}'\beta + w_i'\gamma_j$ depends on a vector of choice specific observed explanatory variables, x_{ij}, that vary with both individuals i and choice alternative j, and a vector of observed characteristics, w_i, only specific to the decision maker i.

Assume further that ε_{ij} is iid extreme value type 1 distributed with location parameter $\mu = 0$ and scale parameter σ, such that conditional choice probabilities has the logit form

$$\Pr(y_i = j \mid v_{i1}, \dots, v_{iJ}; \sigma) = \frac{\exp(v_{ij}/\sigma)}{\sum_{k=1}^{J} \exp(v_{ik}/\sigma)}$$

Question 1.1: Discuss the identification of this model and explain if the scale and level of utility are identified from the choice data. In particular, discuss whether parameters β, γ_j and σ^2 are identified given the observable variables $\{y_i, x_{i1}, \dots, x_{iJ}, w_i\}$. If some parameters are unidentified, discuss which normalizations that are necessary to achieve identification.

Question 1.2: Given the model outlined above, derive the corresponding log-likelihood function for a random sample of N observations $\{y_i, x_{i1}, \dots, x_{iJ}, w_i\}_{i=1}^{N}$, and describe how to obtain Maximum Likelihood estimates of the parameters of interest.

Question 1.3: Let $p_{ij} \equiv \Pr(y_i = j \mid w_i, x_{i1}, \dots, x_{iJ})$ and show that the partial effects of the observed outcome with respect to marginal changes in covariates x_{ik} and w_i can be written as

$$\frac{\partial p_{ij}}{\partial x_{ik}} = p_{ij}\left(\mathbb{1}\{k=j\} - p_{ik}\right)\beta/\sigma$$

and

$$\frac{\partial p_{ij}}{\partial w_i} = p_{ij}\left(\gamma_j/\sigma - \sum_{l=1}^{J} p_{il}\gamma_l/\sigma\right)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function.

Discuss also whether the partial effects are identified if σ is unidentified?

Question 1.4: Derive the odds ratio p_{ik}/p_{il} and show that this ratio of choice probabilities between alternatives k and l does not depend on x_{ij} for any alternative j other than k and l. Discuss the implications of this.

Question 1.5: You work with a co-author on a residential choice model where you try to model in what region $y_i \in \{1, \ldots, J\}$ that households choose to locate. You are interested in predicting the effect on location choices from a counterfactual change in attributes of a particular region (such as house prices, school quality, crime rates, pollution, etc.). Your co-author is concerned that the substitution patterns imposed by the logit model are too restrictive and suggests that you instead work with the probit model where ε_{ij} is multivariate normal.

Why is your co-author concerned, and how could probit potentially help to address this issue?

Problem 2 (30%)

We now consider the Probit model, which has the same structure as above except that ε_{ij} in Eq. (1) now follows a multivariate normal distribution. Specifically, ε_{ij} is an element in the J dimensional vector $\varepsilon_i = (\varepsilon_{i1}, .., \varepsilon_{iJ})'$, where $\varepsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

Question 2.1: To estimate the parameters of the probit model, your co-author suggests that you construct an estimator based on the following optimization problem:

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

where u_{iM} is a sample of M random draws $u_{iM} = \{u_i^{(1)}, \ldots, u_i^{(M)}\}$ from the standard normal distribution, for each $i = 1, \ldots, N$. Here we have defined $z_i = \{x_{i1}, \ldots, x_{iJ}, w_i\}$ to simplify notation.

Describe the principle of the estimation method your co-author is referring to. As part of your answer, you are expected to provide and justify a possible expression of $\widehat{f}(y_i \mid z_i, u_{iM}; \theta)$, and to outline the steps of the corresponding estimation approach.

Hint (for the very detailed answer): You can always obtain draws from the multivariate normal by rescaling a J-vector $u_i^{(m)}$ of independent draws from the standard normal distribution. In particular, $\varepsilon_i^{(m)} = u_i^{(m)}\mathbf{L} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{L}$ is the lower triangular Cholesky matrix such that $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}'$.

Question 2.2: How do you recommend to choose the number of random draws M in Question 2.1? In particular, explain how this number affects the bias of the estimator. Is the estimator consistent for a fixed number of draws, e.q. $M = 1$. *(no derivations expected).*

Question 2.3: How would you modify the optimization problem in Eq. (2) to implement an estimator $\widehat{\theta}$ with the remarkable property that it is consistent for $M = 1$? Describe *briefly* the corresponding estimation approach.

Problem 3 (20%)

Consider the following `MATLAB` functions:

```
function [p] = f1(V)
    p = exp(V)./sum(exp(V));
end

function [p, y] = f2(V, M)
    J=numel(V);
    U = V - evrnd(0,1,M,J);
    [Vmax, y] = max(U,[],2);
    for j=1:J
        p(j)=mean(y==j);
    end
end
```

and the following piece of code:

```
rng(123);
V=[0,1,2,3];
M=10000;

fprintf('f1 = ');
fprintf('%10.4f ', f1(V));
fprintf('\n');

fprintf('f2 = ');
fprintf('%10.4f ', f2(V, M));
fprintf('\n');
```

which produces the following output:

```
f1 =     0.0321     0.0871     0.2369     0.6439
f2 =     0.0306     0.0897     0.2306     0.6491
```

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 `evrnd(mu, sigma, m, n)` produces a $m \times n$ matrix of random draws from the type 1 extreme value distribution with location parameter `mu` and scale parameter `sigma`. `MATLAB` returns the version suitable for modeling minima rather than maxima. We need the mirror image of this distribution which is why we take the negative value]

Question 3.2: Explain why the output of `f1()` and `f2()` look similar.