Econ 551b Econometrics II
Problem Set 5

Prof. John Rust

Due: April 21, 1999 (Wednesday)

QUESTION 1 Derive the maximum likelihood estimator for the model

$$y_i = X_i\beta + \epsilon_i, \quad i = 1, \ldots, N$$

where ϵ_i are *IID double exponential* random variables with mean 0 and scale parameter σ^2:

$$f(\epsilon) = K \exp\left\{ - \left| \frac{\epsilon}{\sqrt{\sigma^2}} \right| \right\}.$$

1. Derive a formula for K so that f is a valid probability density (i.e. so $\int_{-\infty}^{\infty} f(\epsilon)d\epsilon = 1$).
2. Derive the maximum likelihood estimator for (β, σ^2).

Hint: Show that the MLE for β is identical to the *Least Absolute Deviations* estimator $\hat{\beta}_{\text{lad}}$ defined by:

$$\hat{\beta}_{\text{lad}} = \arg\min_{\beta} \sum_{i=1}^{N} |y_i - X_i\beta|.$$

QUESTION 2 Consider the *random utility model:*

$$\tilde{u}_d = v_d + \tilde{\epsilon}_d, \quad d = 1, \ldots, D \tag{1}$$

where $\tilde{u}_d$ is a decision-maker's payoff or utility for selecting alternative d from a set containing D possible alternatives (we assume that the individual only chooses one item). The term v_d is known as the deterministic or *strict utility* from alternative d and the error term $\tilde{\epsilon}_d$ is the random component of utility. In empirical applications v_d is often specified as

$$v_d = X_d\beta \tag{2}$$

where X_d is a vector of observed covariates and β is a vector of coefficients determining the agent's utility to be estimated. The interpretation is that X_d represents a vector of characteristics of the decision-maker and alternative d that are observable by the econometrician and ϵ_d represents characteristics of the agent and alternative d that affect the utility of choosing alternative d which are unobserved by the econometrician. Define the agent's *decision rule* $\delta(\epsilon_1, \ldots, \epsilon_D)$ by:

$$\delta(\epsilon) = argmax_{d=1,\ldots,D}\left[v_d + \tilde{\epsilon}_d\right] \tag{3}$$

i.e. $\delta(\epsilon)$ is the optimal choice for an agent whose unobserved utility components are $\epsilon = (\epsilon_1, \ldots, \epsilon_D)$. Then the agent's *choice probability* $P\{d|X\}$ is given by:

$$P\{d|X\} = \int I\{d = \delta(\epsilon)\} f(\epsilon|X) d\epsilon \tag{4}$$

where $X = (X_1, \ldots, X_D)$ is the vector of observed characteristics of the agent and the D alternatives and $f(\epsilon|X)$ is the conditional density function of the random components of utility given the values of observed components X, and $I\{\delta(\epsilon) = d\}$ is the *indicator function* given by $I\{\delta(\epsilon) = d\} = 1$ if $\delta(\epsilon) = d$ and 0 otherwise. Note that the integral above is actually a multivariate integral over the D components of $\epsilon = (\epsilon_1, \ldots, \epsilon_D)$, and simply represents the probability that the values of the vector of unobserved utilities ϵ lead the agent to choose alternative d.

Definition: The *Social Surplus Function* $U(v_1, \ldots, v_D, X)$ is given by:

$$U(v_1, \ldots, v_D, X) = E\left\{ \max_{d=1,\ldots,D} [v_d + \tilde{\epsilon}_d] \Big| X \right\} = \int_{\epsilon_1} \cdots \int_{\epsilon_D} \max_{d=1,\ldots,D} [v_d + \epsilon_d] f(\epsilon_1, \ldots, \epsilon_D | X) d\epsilon_1 \cdots d\epsilon_D \tag{5}$$

The Social Surplus function is the expected maximized utility of the agent.[1]

Problem: Prove the *Williams-Daly-Zachary Theorem:*

$$\frac{\partial U(v_1, \ldots, v_D, X)}{\partial v_d} = P\{d|X\} \tag{6}$$

and discuss its relationship to *Roy's Identity.*

Hint: Interchange the differentiation and expectation operations when computing $\partial U / \partial v_d$:

$$\begin{aligned} \frac{\partial U(v_1, \ldots, v_D, X)}{\partial v_d} &= \partial/\partial v_d \int_{\epsilon_1} \cdots \int_{\epsilon_D} \max_{d=1,\ldots,D} [v_d + \epsilon_d] f(\epsilon_1, \ldots, \epsilon_D | X) d\epsilon_1 \cdots d\epsilon_D \\ &= \int_{\epsilon_1} \cdots \int_{\epsilon_D} \partial/\partial v_d \max_{d=1,\ldots,D} [v_d + \epsilon_d] f(\epsilon_1, \ldots, \epsilon_D | X) d\epsilon_1 \cdots d\epsilon_D \end{aligned}$$

and show that

$$\partial/\partial v_d \max_{d=1,\ldots,D} [v_d + \epsilon_d] = I\{d = \delta(\epsilon)\}.$$

QUESTION 3 Consider the special case of the random utility model when $\epsilon = (\epsilon_1, \ldots, \epsilon_D)$ has a multivariate (Type I) *extreme value distribution:*

$$f(\epsilon|X) = \prod_{d=1}^{D} \exp\{-\epsilon_d\} \exp\{-\exp\{-\epsilon_d\}\}. \tag{8}$$

Show that the conditional choice probability $P\{d|X\}$ is given by the *multinomial logit formula:*

$$P\{d|X\} = \frac{\exp\{v_d\}}{\sum_{d'=1}^{D} \exp\{v_{d'}\}}. \tag{9}$$

Hint 1: Use the Williams-Daly-Zachary Theorem, showing that in the case of the extreme value distribution (8) the Social Surplus function is given by

$$U(v_1, \ldots, v_D, X) = \gamma + \log\left[\sum_{d=1}^{D} \exp\{v_d\}\right]. \tag{10}$$

[1] If we think of an economy consisting of a population of agents each with their own observed vector of utilities ϵ and $f(\epsilon|X)$ is the density function representing the distribution of these "types" in the population, then $U(v_1, \ldots, v_D, X)$ represents the indirect or maximized utility of a typical person in the population. This is the reason U is referred to as a Social Surplus Function.

where $\gamma = .577216\ldots$ is Euler's constant.
Hint 2: To derive equation (9) show that the extreme value family is *max-stable:* i.e. if $(\epsilon_1, \ldots, \epsilon_D)$ are *IID* extreme value random variables, then $\max_d\{\epsilon_d\}$ also has an extreme value distribution. Also use the fact that the expectation of a single extreme value random variable with location parameter α and scale parameter σ is given by:

$$E\{\tilde{\epsilon}\} = \int_{-\infty}^{+\infty} \epsilon \exp\{-\epsilon\} \exp\{-\exp\{-\epsilon\}\} d\epsilon = \alpha + \sigma\gamma, \tag{11}$$

and the CDF is given by

$$F(x|\alpha, \sigma) = P\{\tilde{\epsilon} \leq x|\alpha, \sigma\} = \exp\left\{-\exp\left\{\frac{-(x-\alpha)}{\sigma}\right\}\right\}. \tag{12}$$

Hint 3: Let $(\epsilon_1, \ldots, \epsilon_D)$ be *INID* (independent, non-identically distributed) extreme value random variables with location parameters $(\alpha_1, \ldots, \alpha_D)$ and common scale parameter σ. Show that this family is max-stable by proving that $\max(\epsilon_1, \ldots, \epsilon_D)$ is an extreme value random variable with scale parameter σ and location parameter

$$\alpha = \sigma \log\left[\sum_{d=1}^{D} \exp\{\alpha_d/\sigma\}\right] \tag{13}$$

QUESTION 4 Extract data in file `data3.asc` in the

`pub/John_Rust/courses/econ551/regression/`

directory on `gemini.econ.yale.edu` (either ftp to `gemini.econ.yale.edu` and login as "anonymous" and `cd pub/John_Rust/courses/econ551/regression` and `get data3.asc` or click on the hyperlink in the html version of this document). This data file contains $n = 3000$ *IID* observations (y_i, x_i) that I generated from the binary probability model:

$$y = \begin{cases} 1 & \text{with probability} \quad \Psi(x,\theta) \\ 0 & \text{with probability} \quad 1-\Psi(x,\theta) \end{cases} \tag{1}$$

where $\Psi(x, \theta)$ is some parametric model of the conditional probability of the binary variable y given x, i.e. $\Psi(x, \theta) = P\{y = 1|x, \theta\}$. Two standard models for Ψ are the *logit* and *probit* models. In the logit model we have

$$\Psi(x, \theta) = \frac{\exp\{x'\theta\}}{1 + \exp\{x'\theta\}} \tag{2}$$

and in the probit mode we have

$$\Psi(x, \theta) = \Phi(x'\theta), \tag{3}$$

where $\Phi(x)$ is the standard normal CDF, i.e.

$$\Phi(x) = \int_{-\infty}^{x} \phi(y) dy$$

where

$$\phi(y) = \frac{1}{\sqrt{2\pi}} \exp\{-y^2/2\}.$$

More generally, Ψ could take the form

$$\Psi(x, \theta) = F(x'\theta)$$

where F is an arbitrary continuous CDF.

1. Show that versions of the logit and probit models can be derived from an underlying *random utility model* where a decision maker has utility function of the form:

$$U(y, x, \epsilon, \theta) = u(y, x, \theta) + \epsilon(y), \quad y = 0, 1$$

   and takes action $y = 1$ if $u(1, x, \theta) + \epsilon(1) > u(0, x, \theta) + \epsilon(0)$ and takes action $y = 0$ if $u(1, x, \theta) + \epsilon(1) \leq u(0, x, \theta) + \epsilon(0)$. Derive the implied *choice probability* $\Psi(x, \theta) = P\{y|x, \theta\}$ in the case where $\{\epsilon(0), \epsilon(1)\}$ is a bivariate normal random vector with $E\{\epsilon(i)\} = 0$, $\text{var}\{\epsilon(i)\} = 1/2$ and $\text{cov}(\epsilon(1), \epsilon(0)\} = 0$ and $u(1, x, \theta) = x'\theta$ and $u(0, x, \theta) = 0$. What is the form of $\Psi(x, \theta)$ in the general case when $\{\epsilon(0), \epsilon(1)\}$ has an unrestricted bivariate normal distribution with mean vector μ and covariance matrix Ω? If the utility function includes a constant term, i.e. $u(1, x, \theta) = \theta_0 + x'\theta_1$ are the θ, μ and Ω parameters all separately identified if we only have access to data on (y, x) pairs?

2. Derive the form of the choice probability under the same assumptions as part 1 above but when $\{\epsilon(0), \epsilon(1)\}$ has a bivariate *Type I extreme value distribution* using the results you have obtained from QUESTION 2 and 3. By doing this you will have derived the binary logit model from first principles.

3. Using the artificially generated data in

   `pub/John_Rust/courses/econ551/regression/data3.asc`

   compute maximum likelihood estimates of the parameters $(\theta_0, \theta_1, \theta_2, \theta_3)$ of the logit and probit specifications given in equations (2) and (3) above, where $x'\theta$ is given by:

$$x'\theta = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3.$$

4. Is it possible to consistently estimate θ by doing nonlinear least squares estimation of the nonlinear regression formulation of the binary probability model

$$y = \Psi(x, \theta) + \eta \tag{4}$$

   instead of doing maximum likelihood? If so, provide a proof of the consistency of the NLLS estimator. If not, provide a counterexample showing that the NLLS estimator is inconsistent.

5. Estimate both the probit and logit specifications by nonlinear least squares as suggested in part (4). How do the parameter estimates and standard errors compare to the maximum likelihood estimates computed in part 3?

6. Is there any problem of heteroscedasticity in the nonlinear regression formulation of the problem in (4)? If so, derive the form of the heteroscedasticity and, using the estimated "first stage" parameters from part 5 above, compute second stage "feasible generalized least squares" (FGLS) estimates of θ.

7. Are the FGLS estimates of θ consistent and asymptotically normally distributed (assuming the model is correctly specified)? If so, derive the asymptotic distribution of the FGLS estimator, and if not provide a counter example showing that the FGLS estimator is inconsistent or not asymptotically normally distributed. If you conclude that the FGLS estimator is asymptotically normally distributed, is it as efficient as the maximum likelihood estimator of θ? Explain your reasoning for full credit.

8. Is it possible to determine whether the data in the file `data3.asc` are generated from a logit or probit model? In answering this question, consider whether you could estimate $\Psi(x, \theta)$ nonparametrically via non-parametric regression. Is there any way you could use the nonparametric regression estimate of Ψ to help discriminate between the logit and probit specifications?