Econ 240B Spring 2009
Problem Set 4

This problem set is due in class on Monday May 4th, 2009

1. Derive the Multinomial Logit Model. Define the latent utility of choosing alternative m, $m = 1, \ldots, M$, to be:

$$y_{im}^* = x_i' \beta_m + \epsilon_{im}$$

where conditional on $x_{im}, m = 1, \ldots, M$, $\epsilon_{im}, m = 1, \ldots, M$ are independently and identically distributed with type I extreme value distribution function: $F(\epsilon) = \exp(-\exp(-\epsilon))$. Individual i chooses alternative m if and only if it yields the highest latent utility, i.e.

$$y_{im} = \begin{cases} 1 & \text{if} \quad y_{im}^* \geq y_{im'}, \forall m' \neq m, m' = 1, \ldots, M \\ 0 & \text{otherwise} \end{cases}$$

Show that

$$P(y_{im} = 1 | x_i) = \frac{\exp(x_i' \beta_m)}{\sum_{m'=1}^{M} \exp(x_i' \beta_{m'})}$$

How do you interpret the coefficients β_m?

You may take $M = 3$ if you find it notationally cubersome to work with general M.

2. Answer the following questions as true, false or uncertain and explain your answers.

   (a) consider the limited depedent variable model

$$y(t) = \begin{cases} x(t)\beta + u(t) & \text{if} \quad x(t)\beta + u(t) > 0 \\ 0 & \text{otherwise} \end{cases}$$

   Regressing $y(t)$ on $x(t)$ using only those observations for which $y(t) \neq 0$ yields estimates that overstates the true value of β.

   (b) Consider the limited dependent variable model

$$y(t) = \begin{cases} x(t)\beta + u(t) & \text{if} \quad x(t)\gamma + e(t) > 0 \\ 0 & \text{otherwise} \end{cases}$$

   where $(u(t), e(t)) \sim N(0, \Sigma)$. Applying nonlinear least square to the regression equation of $y(t)$ on $x(t)$ for the sample of observation with $y(t) \neq 0$:

$$y(t) = x(t)\beta + \rho \frac{\phi(x(t)\gamma)}{\Phi(x(t)\gamma)} + \eta(t)$$

   will yield consistent estimate of the parameter β. ϕ and Φ are the density and the c.d.f. of a standardized normal distribution, and ρ is a parameter.

   (c) The answer to question (2.b) does not change if merely $E(u(t) | e(t)) = \rho e(t)$ where $e(t) \sim N(0, 1)$.

(d) Consider the probit model $\delta(t) = 1$ when $y(t) = x(t)\beta + e(t) > 0$ with $e(t) \sim iid\ N(0,1)$, and $\delta(t) = 0$ otherwise. Applying weighted nonlinear LS to the regression equation $\delta(t) = 1 - \Phi\left(-x(t)'\beta\right) + u(t)$(where Φ is the standard normal cdf) yields a consistent estimator for β that is as efficient as the maximum likelihood estimator for β.

3. This is an empirical exercise about multinomial logit models.

   Use the data set bell_female.dat. Consider a three-state classification of a woman's hours of work: she doesn't work at all(designate by setting the discrete variable b=1); she works part of the year which implies that $0 <$ Weeks < 20(designated by b=2); and she works most of the year with Weeks ≥ 20(b=3). Estimate the probability $Pr(b = j|x)$ using a multinomial logit model and test whether marriage influences the likelihood that a woman works part of the year instead of most of the year.