Problem Set 1

Directions: Answer all questions. Clearly label all answers. Show all of your code. Compute standard errors for all parameter estimates. Turn in the following to me via your dropbox (in a folder labeled 'MatlabPS2.1') in Sakai by 11:59 p.m. on Thursday, July 26, 2012:

- m-file(s)
- a log file (from off the cluster)
- matsub.oXXXXX file
- pdf version of your writeup with its LaTeXsource code

Put the names of all group members at the top of your writeup (each student must turn in his/her own materials).

1. Multinomial Logit. Using the dataset `nlsy97.mat`, consider the following model

$$\begin{aligned} d_{ijt} &= \beta_{1j} + \beta_{2j} male_i + \beta_{3j} AFQT_i + \beta_{4j} Mhgc_i + \beta_{5j} hgc_{it} + \beta_{6j} exper_{it} + \\ &\quad \beta_{7j} Diploma_{it} + \beta_{8j} AA_{it} + \beta_{9j} BA_{it} + \varepsilon_{ijt} \end{aligned}$$

   where $\varepsilon_{ijt} \overset{iid}{\sim}$Type I Extreme Value, and d_{ijt} is an indicator that $choice_{it} = j$ for individual i in time t. $choice_{it}$ takes on values 1, 2, or 3, corresponding to school, work, and other, respectively. The variables are in $N \times T$ form, with $T = 5$ corresponding to ages 20-24 (inclusive) of the individual, and are summarized as follows:

`ID`	:	unique individual identifier
`age`	:	age (in years)
`male`	:	1 if male, 0 if female
`AFQT`	:	ASVAB test score normalized to have mean-zero, s.d. 1
`Mhgc`	:	mother's highest grade completed
`hgc`	:	highest grade completed
`exper`	:	labor market experience (in years)
`Diploma`	:	1 if has HS diploma (or GED), 0 otherwise
`AA`	:	1 if has associate's degree, 0 otherwise
`BA`	:	1 if has bachelor's degree, 0 otherwise
`log_wage`	:	log (hourly wage)
`activity`	:	1 if in school, 2 if working (not in school), 3 if other

   (a) Reshape the data from "wide" panel to "long" panel

(b) Using `fminunc`, estimate a multinomial logit model, normalizing the utility of "other" to zero. Report the log likelihood value and the sample size.

(c) Double check that your `fminunc` results are correct by doing the same estimation using Matlab's `mnrfit` function.

(d) What is the interpretation of the parameter β_{5j}?

2. Unobserved heterogeneity (Fixed and Random Effects). Using `nlsy97.mat`, consider the following model

$$\begin{aligned}\ln(wage_{it}) &= \gamma_1 + \gamma_2 male_i + \gamma_3 AFQT_i + \gamma_4 Mhgc_i + \gamma_5 hgc_{it} + \gamma_6 exper_{it} + \\ &\quad \gamma_7 Diploma_{it} + \gamma_8 AA_{it} + \gamma_9 BA_{it} + \varepsilon_{it}\end{aligned}$$

(a) Estimate γ under the assumption that $\gamma_3 = \gamma_4 = 0$ and ε_{it} is iid across time and people, and independent of the regressors.

(b) Repeate (a), but allow γ_3 and γ_4 to be freely estimated. How do your estimates change? How does this relate to unobserved heterogeneity?

(c) Now estimate $\tilde{\gamma} = \begin{bmatrix} \gamma_5 & \gamma_6 & \gamma_7 & \gamma_8 & \gamma_9 \end{bmatrix}'$ assuming $\varepsilon_{it} = \alpha_i + u_{it}$, and that the usual fixed effects assumptions also hold. The formula for this is

$$\hat{\gamma}_{FE} = \left(\sum_i \sum_t \ddot{x}'_{it} \ddot{x}_{it} \right)^{-1} \sum_i \sum_t \ddot{x}'_{it} \ddot{y}_{it}$$

where $\ddot{x}_{it} = x_{it} - \bar{x}_i$ and $\ddot{y}_{it} = y_{it} - \bar{y}_i$. (Note: be careful of missing wage observations and the fact that this is an unbalanced panel.)

Compare your estimates of $\tilde{\gamma}$ with your estimates of γ. What other potentially important predictors of wages were in the error term in (a) and (b) but are differenced out in (c)?

(d) Estimate γ assuming $\varepsilon_{it} = \alpha_i + u_{it}$, and that α_i is independent of the regressors (i.e. the usual random effects assumptions also hold—see lecture slides for details). Be sure to recover the estimated variance parameters and the standard errors for all parameter estimates [except the error variances]. Compare your results from (b), (c), and (d).