Econ 388E - Fall 2018
Assignment 2: Discrete Choice Models
Due Friday, Nov 9

Partial Answers:
1a) $\widehat{\theta_0}$ = -0.708, SE($\widehat{\theta_0}$) = 0.386
c) $\widehat{\theta_0}$ = -1.132, SE($\widehat{\theta_0}$) = 0.625
2d) $\widehat{\theta_0}$ = -0.391, SE($\widehat{\theta_0}$) = 0.660 (note that this maximization problem appears to be sensitive to initial values, so try multiple ones)
3d) $\widehat{\theta_0}$ = -0.798, SE($\widehat{\theta_0}$) = 0.194 (note: this was with 10000 sim draws - your estimates could be different if you used a lot less)
f) $\widehat{\theta_0}$ = -0.948, SE($\widehat{\theta_0}$) = 0.152

1) The file "ps2.dat" contains data on labor force participation of married women (originally from Mroz (1987)). The dataset has 4 variables (columns) in the following order:

y_i- Dummy Variable = 1 if woman participated in labor force
$x_{1i}-$ age of woman
$x_{2i}-$ woman's education level (in years)
z_i- mean education level (in years) of parents of woman

a) Use ML to estimate a binary probit model on this data, with labor force participation as the endogenous variable, and x_{1i} and x_{2i} (but not z_i) as explanatory variables (plus a constant term). You can think of this as the following model:

$$y_i = I(\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \epsilon_i > 0)$$

where y_i is equal to 1 if the woman participated in the labor force and ϵ_i is distributed $N(0,1)$. Compute standard errors of your estimates.

b) Given your estimated parameters, what is the effect of an additional year of eduction on the probability of an *average* (in terms of explanatory variables and unobservables) woman working?

c) Use ML to estimate a binary logit model (assuming ϵ_i is distributed logistic, i.e. as the difference in two i.i.d. extreme value deviates) on the data. Again compute standard errors. Why are there big differences in the estimated coefficients between the logit and probit models? What about the difference in the ratio of estimated coefficients (i.e. θ_1/θ_2) between the two models? Explain why the former differences are bigger than the latter.

2) Considering the same dataset:

a) Give me an economic story (i.e. an economic model) where x_{2i} is correlated with ϵ_i? Given this model, in what direction would you expect your estimate of θ_2 to be biased?

b) Let's now allow for possible endogeneity of x_{2i}. Since this is a non-invertible model, we generally will need to specify how the variable x_{2i} is determined. Assume that a woman's educational level depends on her age (e.g. for older women, there were likely more significant barriers to education at the time of their youth), her parent's average educational level (z_i), and unobservables. This results in the two equation system:

$$\begin{aligned} y_i &= I(\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \epsilon_i > 0) \\ x_{2i} &= \theta_3 + \theta_4 x_{1i} + \theta_5 z_i + \eta_i \end{aligned}$$

or the "reduced form"

$$\widetilde{y}_i = \begin{pmatrix} y_i \\ x_{2i} \end{pmatrix} = f(\widetilde{x}_i, \widetilde{\epsilon}_i, \theta) = \begin{pmatrix} I((\theta_0 + \theta_2\theta_3) + (\theta_1 + \theta_2\theta_4)\, x_{1i} + \theta_2\theta_5 z_i + \theta_2\eta_i + \epsilon_i > 0) \\ \theta_3 + \theta_4 x_{1i} + \theta_5 z_i + \eta_i \end{pmatrix}$$

c) Assume that

$$\begin{pmatrix} \epsilon_i \\ \eta_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & \sigma_\eta^2 \end{bmatrix}\right.$$

independently of x_{1i} and z_i. How does ρ relate to the possible "endogeneity" of x_{2i}? Would you expect ρ to be positive, negative, or zero? What is the importance of our assumption that z_i is does not directly determine y_i?

d) Estimate these two equations jointly using Maximum Likelihood and compute standard errors. Note that there are a total of 8 parameters to estimate. Note that while the "Second Stage" equation is not invertible in the unobservable(s), the "First Stage" equation is invertible. As a result, it may be easiest to factor the likelihood $p(\widetilde{y}_i|\widetilde{x}_i, \theta) = p(y_i, x_{2i}|x_{1i}, z_i, \theta)$ into the product of a marginal and a conditional likelihood, i.e.

$$\begin{aligned} p(y_i, x_{2i}|x_{1i}, z_i, \theta) &= p(x_{2i}|x_{1i}, z_i, \theta)p(y_i|x_{2i}, x_{1i}, z_i, \theta) \\ &= p(x_{2i}|x_{1i}, z_i, \theta)p(y_i|x_{2i}, x_{1i}, z_i, \eta_i, \theta) \end{aligned}$$

Note that the second equality follows because the first stage equation is invertible - this implies that conditioning on $(x_{2i}, x_{1i}, z_i, \theta)$ is equivalent to conditioning on $(x_{2i}, x_{1i}, z_i, \eta_i, \theta)$. The likelihood component $p(x_{2i}|x_{1i}, z_i, \theta)$ is a simple change of variable formula. The component $p(y_i|x_{2i}, x_{1i}, z_i, \eta_i, \theta)$ can be calculated without simulating (i.e. using the normal CDF Φ), but you will need to use results on distributions of conditional normals to obtain the distribution $p(\epsilon_i|\eta_i, \theta)$ to use in that integration problem.

e) Test the hypothesis that x_{2i} is correlated with ϵ_i. You should be able to do this with a simple t-test on your estimated parameters.

f) Consider the alternative model

$$\begin{aligned} y_i &= I(\theta_0 + \theta_1 x_{1i} + (\theta_2 + \sigma_\tau \tau_i)x_{2i} + \epsilon_i > 0) \\ x_{2i} &= \theta_3 + \theta_4 x_{1i} + \theta_5 z_i + \eta_i \end{aligned}$$

Everything is as before, except the additional unobservable τ_i. Suppose that this new unobservable τ_i is independent of $(x_{1i}, x_{2i}, z_i, \eta_i, \epsilon_i)$ and has a lognormal distribution, i.e.

$$\ln(\tau_i)|x_{1i}, x_{2i}, z_i, \eta_i, \epsilon_i \;\tilde{}\; N(0, 1)$$

Descibe what you think is an economic interpretation of τ_i.

g) Write down (**do not program or estimate!!!**) a simulated likelihood function for this new model.

h) Write down (**do not program or estimate!!!**) a set of simulated moments that can be used to

estimate this new model. Be sure to include at least as many moment conditions as parameters (now there are 9 parameters). Discuss informally how the various moments would provide information on the parameters (I'm particularly interested in which of your moments you think might provide information on the variance related parameters σ_τ^2, σ_η^2, and ρ)

i) Discuss the tradeoffs between the two possible estimators proposed in g) and h) - what are advantages of the SML estimator, and what are advantages of the MSM estimator?

j) Are the simulated moments in part h) continuous (smooth) in the parameter vector θ? Can you write down an alternative set of simulated moments that are smooth in θ? (One way to do this is to use the importance sampling approach discussed in class, but there are other ways, e.g. simulators where you simulate some of the unobservables, but analytically itegrate over other unobservables).

k) Given that we are allowing ϵ_i to be correlated with η_i, does it seem reasonable to assume that τ_i is independent of η_i? Discuss the economic content behind these assumptions.

l) Look up STATA's "ivprobit" command and read how it works. This command can estimate the model in b) and c) - what are the assumptions required for "ivprobit" to produce consistent estimates? Why can "ivprobit" not be used to estimate the model in part f)?

3) The file "sim3.dat" contains simulated data on a panel of consumer decisions whether or not to purchase a grocery item in a given week. The dataset has four columns:

i - consumer number ($i = 1, ..., N$)
t - week number ($t = 1, ..., T$)
y_{it} - purchase decision ($= 1$ if purchased the product in week t)
p_{it} - price of the product faced by consumer i in week t.

Consider the following discrete choice model of behavior. Consumer i purchases the product in week t (i.e. $y_{it} = 1$) iff:

$$U_{it} > 0$$

where U_{it}, the normalized utility from purchasing the product, is given by:

$$U_{it} = \theta_0 + \theta_1 p_{it} + \theta_2 y_{i,t-1} + \sigma_\alpha \alpha_i + \epsilon_{it}$$

Note that this utility function depends not only on price, but on the consumer's choice last period (i.e. we allow for state dependence). In addition, note that there are two unobservables entering U_{it}, α_i, a random effect which is constant across time, and ϵ_{it}, which varies across time. Assume that:

$$\begin{aligned} \epsilon_{it} &\sim \text{i.i.d. difference between 2 Type 1 Extreme Value deviates (i.e. "logit" error)} \\ \alpha_i &\sim N(0,1) \text{ independently across } i \end{aligned}$$

and that both unobservables are distributed independently of p_{it} (i.e. price is exogenous in this model). The fixed variance of the ϵ_{it} represents the multiplicative normalization of the discrete choice model. σ_α measures the variance (and thus the relevance) of the random effect (relative to the fixed variance of ϵ_{it}). Also assume that

$$y_{i0} = 0 \text{ for all } i$$

This simple "initial condition" assumption is needed for deriving choice probabilities at $t = 1$.

In thinking about maximum likelihood estimation of the above model, note that there is correlation in each individual's decisions over time. Thus, one needs to consider the *joint* likelihood of consumer i's decisions in all the time periods. In other words, the likelihood of the data for consumer i is:

$$L_i = \Pr(y_{i1}, ..., y_{iT} \mid p_{i1}, ..., p_{iT}, y_{i0}; \theta) \tag{1}$$

Integrating out the unobservable α_i results in:

$$L_i = \int \Pr(y_{i1}, ..., y_{iT} \mid p_{i1}, ..., p_{iT}, y_{i0}, \alpha_i; \theta) p(\alpha_i)$$

a) Show that factoring the inner joint probability into the product of conditional probabilities results in:

$$L_i = \int \left[\prod_{t=1}^{T} \Pr(y_{it} \mid p_{it}, y_{it-1}, \alpha_i; \theta) \right] p(\alpha_i) \tag{2}$$

(the key here is to argue why y_{it-k} for $k > 1$ disappear from the above equation)

b) Would the same be true if this was done *without* integrating out α_i? In other words does the following equality hold? Why or why not?

$$L_i = \Pr(y_{i1}, ..., y_{iT} \mid p_{i1}, ..., p_{iT}, y_{i0}; \theta) = \prod_{t=1}^{T} \Pr(y_{it} \mid p_{it}, y_{it-1}; \theta)$$

c) Argue that each of the elements in the product of (2) is a simply binary logit probability, i.e. that the likelihood can be written as:

$$L_i = \int \left[\prod_{t=1}^{T} \left[y_{it} \frac{e^{\theta_0+\theta_1 p_{it}+\theta_2 y_{i,t-1}+\sigma_\alpha \alpha_i}}{1+e^{\theta_0+\theta_1 p_{it}+\theta_2 y_{i,t-1}+\sigma_\alpha \alpha_i}} + (1-y_{it}) \frac{1}{1+e^{\theta_0+\theta_1 p_{it}+\theta_2 y_{i,t-1}+\sigma_\alpha \alpha_i}} \right] \right] p(\alpha_i) \tag{3}$$

d) Estimate this model using simulated maximum likelihood. The parameters you are estimating are $(\theta_0, \theta_1, \theta_2, \sigma_\alpha)$. Note that since the "logit" probabilities are analytically computable, the only integral that needs to be simulated here is the integral over α_i. In other words, for consumer i, use the following simulated likelihood:

$$\widehat{L_i} = \frac{1}{S} \sum_s \left[\prod_{t=1}^{T} \left[y_{it} \frac{e^{\theta_0+\theta_1 p_{it}+\theta_2 y_{i,t-1}+\sigma_\alpha \alpha_{i,s}}}{1+e^{\theta_0+\theta_1 p_{it}+\theta_2 y_{i,t-1}+\sigma_\alpha \alpha_{i,s}}} + (1-y_{it}) \frac{1}{1+e^{\theta_0+\theta_1 p_{it}+\theta_2 y_{i,t-1}+\sigma_\alpha \alpha_{i,s}}} \right] \right] \tag{4}$$

where the $\alpha_{i,s}$ are $S = 1000$ (or as many as you like, e.g. if takes too much time) draws from the $N(0,1)$ distribution. Use *different* draws for simulating the integral for every consumer i. Compute standard errors of your estimates (assume that $\sqrt{N}/S \to 0$ as $N \to \infty$, so you can ignore simulation error in computing these standard errors)

e) What do your estimates say about state dependence versus heterogeneity in explaining correlation in y_{it} across time?

f) Re-estimate the model fixing $\sigma_\alpha = 0$ (i.e. assuming there is no time invariant unobserved heterogeneity). Note that now there is no need to simulate. What happens to your estimate of θ_2? Why do you think this happens?

g) Using only linear regressions (e.g. linear probability models), can you think of a way to "crudely" test the null hypothesis that there is no state dependence in your data (i.e. whether $\theta_2 = 0$)? What is the key "exclusion" restriction behind this test?