Econ 8825 May 21, 2019

Spring 2019

1. (30) This question pertains to the multinomial choice model studied in class.

   Assume the dependent variable takes one of $J+1$ mutually exclusive and collectively exhaustive alternatives (numbered from $j=0$ to $j=J$).

   Recall the model was based on utility maximization. Specifically, for individual i, alternative j is assumed to have an unobservable indirect utility y^*_{ij} and the alternative with the highest indirect utility is assumed chosen.

   Thus the observed variable y_{ij} has the form

   $$y_{ij} = I[y^*_{ij} > y^*_{ik} \text{ for } k = 0, ..., J]$$

   An assumption of joint continuity of the indirect utilities rules out ties (with probability one);

   In the model for this question, the indirect utilities are further restricted to have the linear form

   $$y^*_{ij} = x'_{ij}\beta_0 + \epsilon_{ij}$$

   for $j = 0, ..., J$, where the $J+1$ dimensional vector ϵ_i (with elements ϵ_{ij}) of unobserved error terms is assumed to be jointly continuously distributed and independent of the $(J+1) \times k$-dimensional matrix of regressors X_i (whose j^{th} row is x_{ij}).

   The purpose of this question is to focus on identification and estimation of β_0, from a random sample of observations of the vector (y_i, x_i), where $y_i \equiv (y_{i0}, ...y_{iJ}), x_i \equiv (x_{i0}, ...x_{iJ})$.

   We will first take a parametric approach.

   (a) Assume the vector $\epsilon_{ij}, j = 0, 1, ...J$ is multivariate normal and distributed independently of the regressors x_{ij}. List all scale and location normalizations needed to identify β_0.

   (b) Write down the likelihood function for this model.

   (c) Propose an algorithm to construct a simulated MLE (SMLE). What motivates an SMLE over an MLE?

2. (30) Consider a variation of the standard *censored* selection model, where we get to observe both hours worked and wages of individuals. Hours worked are censored at 0 and we only observe wages for those who are working, and hence their hours worked is positive.

$$
\begin{aligned}
y_{1i} &= \max(w_i'\delta_0 + \eta_i, 0) \\
d_i &= I[y_{1i} > 0] \\
y_i &= d_i(x_i'\beta_0 + \epsilon_i)
\end{aligned}
$$

Observed variables: $(y_{1i}, d_i, w_i, y_i, x_i)$ and we assume $(\eta_i, \epsilon_i) \sim N(\mathbf{0}, \mathbf{\Sigma})$, where the (1,1) element of $\mathbf{\Sigma}$ is σ_1^2, the other diagonal element is σ_2^2, and the off diagonal terms are γ_0. We wish to estimate $(\delta_0, \beta_0, \gamma_0, \sigma_1^2, \sigma_2^2)$.

(a) Write the likelihood function for this model.

(b) Propose a NLLS estimator for the parameters in this model.

(c) Propose a computationally friendly Heckman type two-step estimator where the first step involves a convex optimization procedure and the second step is closed form.

3. (15) Frechet(1951) was interested in knowing whether knowing the marginal distributions of continuous random variables X and Y tells us anything about their joint distribution.

He showed that, given knowledge of the distribution functions $F(\cdot)$ and $G(\cdot)$ of X and Y, respectively, their joint distribution $K(\cdot,\cdot)$is such that for all a, b:

$$\max(F(a) + G(b) - 1, 0) \leq K(a, b) \leq \min(F(a), G(b))$$

Prove this result.