14.385 Recitation 3

Paul Schrimpf

September 26, 2008

1 PS1 Solutions

see website.

2 Mixed Logit Models

Lecture 4 covered a handful of multinomial choice models. We talked about multinomial logit, nested logit, and multinomial probit. For more information on these and related models, see Kenneth Train's freely available book on discrete choice methods. It can be found at `http://elsa.berkeley.edu/books/choice2.html`.

One of the most popular models of discrete choice is the mixed logit, or logit with random coefficients model. In this model, the utility for person i from choice j is:

$$U_{ij} = x_{ij}\beta_i + \epsilon_{ij}$$

where ϵ_{ij} is iid extreme value and β_i has pdf $f(\beta|\theta)$ and θ are some parameters to be estimated. This type of model is very popular in IO. A typical application might look at the demand for different brands of a product. The choices indexed by j are different brands. x_{ij} are the characteristics of each brand. β_i are consumers' heterogeneous tastes for various characteristics.

A person chooses k if $U_{ik} \geq U_{ij}$ for all j. The probability of choosing k is then:

$$P(k|x) = \int \frac{e^{x_{ik}\beta}}{\sum_j e^{x_{ij}\beta}} dF(\beta;\theta)$$

This integral is typically computed through simulation.

$$\tilde{P}(k|x) = \sum_{r=1}^{R} \frac{e^{x_{ik}\beta_r(\theta)}}{\sum_j e^{x_{ij}\beta_r(\theta)}}$$

where $\{\beta_r(\theta)\}$ are R indenpendent draws from $f(\beta|\theta)$. $\tilde{P}(k|x)$ can be used to form a simulated method of moments, or simulated maximum likelihood objective function.

3 Asymptotics of Simulated Estimators

This is a quick and dirty discussion of the asymptotics of simulated extremum estimators. See Train's book and the references therein for more details and rigor.

Let $Q_n(\theta)$ denote the exact objective function. Let $\tilde{Q}_n(\theta)$ denote the simulated objective function. Assume that $\hat{\theta} = \arg\min Q_n(\theta)$ is consistent and asymptotically normal. We want to understand the behavior of $\tilde{\theta} = \arg\min \tilde{Q}_n(\theta)$. For specificity, assume that in simulating, we make R draws for each observation, and these draws are independent across observations.

3.1 Consistency

For consistency, the key condition to check is that $\tilde{Q}(\theta) = \text{plim}\,\tilde{Q}_n(\theta)$ is uniquely minimized at θ_0. Consider the first order condition:

$$\nabla\tilde{Q}_n(\theta) = \nabla Q_n(\theta) + \left(E_r(\nabla\tilde{Q}_n(\theta)) - \nabla Q_n(\theta)\right) + \left(\nabla\tilde{Q}_n(\theta) - E_r(\nabla\tilde{Q}_n(\theta))\right)$$

where E_r denotes an expectation taken over our simulated draws. If we can show that the second and third terms on the right vanish as $n \to \infty$, then we will have consistency. The third term is easy. Since we are making R independent draws for each observation, as long as R is fixed or increasing with N, $\nabla\tilde{Q}_n(\theta)$ satisfies an LLN and converges to its expectation. The second term depends on how we are simulating. If R increases with N, then it also vanishes because of an LLN. Furthermore, even if R is fixed with N, it will be zero, if our simulations result in an unbiased estimate of the gradient. In the mixed logit example above, the simulation of choice probabilities is unbiased. Therefore, NLLS, for which the first order condition is linear in $\tilde{P}$, is consistent with fixed R. However, MLE, for which the first order condition involves $\frac{1}{\tilde{P}}$, is consistent only if R increases with N. For this reason, people sometimes suggest using the method of simulated scores (MSS) instead of MSL. MSS call for simulated the score in an unbiased way and doing GMM on the simulated score.

3.2 Asymptotic Normality

As always, we start by taking an expansion of the first order condition:

$$\sqrt{n}(\tilde{\theta} - \theta_0) = (\nabla^2\tilde{Q}_n(\bar{\theta}))^{-1}(\sqrt{n}\nabla\tilde{Q}_n(\theta_0))$$

If $\tilde{\theta}$ is consistent, then $(\nabla^2\tilde{Q}_n(\bar{\theta}))^{-1} \xrightarrow{p} (\nabla^2 E\tilde{Q}(\theta_0))^{-1}$. The main thing to worry about is the behavior of the gradient. As above, it helps to break it into three pieces:

$$\sqrt{n}\nabla\tilde{Q}_n(\theta_0) = \sqrt{n}\nabla Q_n(\theta) + \sqrt{n}\left(E_r(\nabla\tilde{Q}_n(\theta)) - \nabla Q_n(\theta)\right) + \sqrt{n}\left(\nabla\tilde{Q}_n(\theta) - E_r(\nabla\tilde{Q}_n(\theta))\right)$$

Let's start with the third term. Suppose we have iid observations so that $\nabla\tilde{Q}_n = \sum_{i=1}^n \nabla\tilde{q}_{i,R}$. Let S be the variance of $\nabla\tilde{q}_{i,1}$. Then the variance of $\nabla\tilde{q}_{i,R}$ is S/R, and

$$\sqrt{n}\left(\nabla\tilde{Q}_n(\theta) - E_r(\nabla\tilde{Q}_n(\theta))\right) \xrightarrow{d} N(0, S/R)$$

Now, on to the second term. As above, it is zero if our simulations are unbiased. If our simulations are biased, then it is $O(\frac{1}{R})$. If R is fixed, then our estimator is inconsistent. If $\frac{\sqrt{n}}{R} \to 0$, then this term vanishes, and our estimator has the same asymptotic distribution as when using the exact objective function. If R grows with n, but slower that $\sqrt{n}$, then $\tilde{\theta}$ is consistent, but not asymptotically normal.

4 Selection Models

Suppose you have an outcome, y, that is a linear function of some regressors, x,

$$y = x\beta + \epsilon \tag{1}$$

but you do not observe y for the entire population, instead you only observed y if

$$z\gamma - \nu > 0 \tag{2}$$

where ν and ϵ are potentially correlated. Assume that ν and ϵ are independent of x and z. Estimating (1) by OLS using only the observations with y observed will be inconsistent because when ν and ϵ are correlated, $E[x\epsilon|z\gamma > \nu] \neq 0$ even though $E[x\epsilon = 0]$. However, if we knew $E[\epsilon|z\gamma > \nu]$ then we could do OLS on:

$$y = x\beta + E[\epsilon|z\gamma > \nu] + e \tag{3}$$

to consistently estimate β. In lecture 5 and problem set 2, we saw that if ν and ϵ are jointly normal this conditional expecation is given by the inverse mills ratio, $E[\epsilon|z\gamma > \nu] = \frac{\phi(z\gamma)}{\Phi(z\gamma)}$. What if ν and ϵ are not normal? It is always true that:

$$\begin{aligned}
E[\epsilon|z\gamma > \nu] =& \frac{1}{F_\nu(z\gamma)} \int_{-\infty}^{z\gamma} p_{\nu|z}(\nu|z) \int p_{\epsilon|\nu,z}(\epsilon|\nu, z) d\epsilon d\nu \\
& \text{independence} \\
=& \frac{1}{F_\nu(z\gamma)} \int_{-\infty}^{F_\nu^{-1} F_\nu(z\gamma)} p_\nu(\nu) \int p_{\epsilon|\nu}(\epsilon|\nu) d\epsilon d\nu \\
=& K(F_\nu(z\gamma)
\end{aligned}$$

the conditional expectation of ϵ is just some function of the probability of being included in the sample. This suggests that we can estimate the model semiparametrically by:

1. Specify a distribution for ν, estimate γ by ML
2. Run OLS of y on x and a polynomial of powers of $F_\nu(z\hat{\gamma})$

This procedure is semiparametric in the sense that it leaves the distribution of ϵ unspecified. We will learn more about semiparametric estimation later. For this model, it is possible to be even more flexible. You can also leave the distribution of ν unspecified, replace $z\gamma$ with just some unknown function, $g(z)$, and repalce $x\beta$ with some other unknown function, $\mu(x)$.

In general, the above procedure, where an unobserved disturbance is replaced by its conditional expectation, is called the control function approach. The conditional expectation is a "control function" that controls for endogeneity. As briefly mentioned in 382, 2SLS has a control function interpretation.

4.1 Extremum Estimator Computation

Problem set 2 asks some questions about how extremum estimators are computed. To answer these it helps to know a little bit about numerical optimization algorithms. Raymond Guiteras wrote some nice notes on MLE. These notes can be found on the course website. I have some slides about optimization in Matlab at `http://web.mit.edu/~paul_s/www/14.170/matlab.html`. Train's book has a nice chapter on maximization. `http://elsa.berkeley.edu/books/choice2.html`