PROBLEM SET

14.171 – Software Engineering for Economists

September 5th, 2008

Instructor: Matt Notowidigdo; noto [at] mit (dot) edu

Introduction:

This problem set contains four Stata exercises that are somewhat longer and more involved than the in-class exercises. In many cases, the problems require being able to implement specific estimators. If the estimators are unfamiliar, there are links to relevant papers and/or pages from Greene ("Econometric Analysis" textbook) that you can consult.

Problem 1: Correcting for measurement error of a known (estimated) form

URL: http://web.mit.edu/econ-gea/14.171/problem1/

In this problem you will write an estimator that will correct for attenuation bias. There is a Stata DO file that you can use to get started. I have also posted the Card-Lemieux paper that describes this estimator in more detail.

We are interested in estimating the returns to ability, but we have a noisy measure of ability. The true model is the following:

$$w_i = \alpha + \beta a_i^* + \varepsilon_i$$

where i indexes one of the K individuals, w is the observed wage, and a measures true ability. Instead of observing true ability directly, we observe a_i which is a noisy measure of a_i^*. Thus when we regress observed wages on our noisy measure of ability, we will estimate a biased and attenuated estimate of β.

In this setting, assume that our (noisy) measure of a_i is actually an average of a random sample of noisy measurements of true ability, and that for some reason we only have (1) the number of observations used in computing the average ability and (2) the sample variance of the observations that were used. For some reason we don't have the individual measurements, but instead for each individual we have only the average ability, the sample variance of the individual measurements, and the number of observations used to calculate the average ability. Note that if we had multiple noisy measurements rather than just the average and sample variance, then we could use 2SLS to get an unbiased estimate of β. Instead for each individual we have N_i and s_i^2, which we can use to get an unbiased estimate of β under the assumption that the measurement error of each noisy measurement of ability is i.i.d.

A measurement-error corrected estimator can be written as follows:

$$\beta_{ME-CORR} = \left(\frac{X'X}{K} - \hat{\Sigma} \right)^{-1} \left(\frac{X'y}{K} \right)$$

HINT: If the first column of X is the constant term and the second column is the noisy measure of ability a, then $\hat{\Sigma}$ is the following:

$$\hat{\Sigma} = \frac{1}{K}\sum_{i=1}^{K}\hat{V}_i$$

$$\hat{V}_i = \begin{bmatrix} 0 & 0 \\ 0 & \frac{s_i^2}{N_i} \end{bmatrix}$$

You should also compute the asymptotic standard errors (see end of A.1.1 of Card-Lemieux paper). You may implement this problem in either Stata (using "matrix accum" and "matrix vecaccum") or in Mata (or do it both ways!).

Problem 2: Bootstrapped standard errors for Heckman two-step estimator

URL: http://web.mit.edu/econ-gea/14.171/problem2/

In this problem you will implement the Heckman two-step (sample selection) estimator. Use the DO file for this problem (problem2.do) posted at above URL. The relevant pages from Greene are also posted above which go over the sample selection model; see the middle of page 784.

The last part of this problem is to construct bootstrapped standard errors. Instead of computing the formula for analytic standard errors, write a loop to calculate the bootstrapped standard errors (type "help bsample"). For 1,000 replications, the bootstrapped standard errors should be very close to the standard errors reported by "heckman …, twostep." Compare the simplicity of your resulting code to the formula at the bottom of page 785. In more complicated two-step estimators, computing the analytic standard errors might be very difficult, and bootstrapping the standard errors can be a helpful alternative.

Problem 3: Conditional logit

URL: http://web.mit.edu/econ-gea/14.171/problem3/

In this problem you will implement the conditional logit maximum likelihood estimator (MLE) using Stata's ML language. The log-likelihood function is in the middle of page 698 of Greene (relevant pages are posted at above URL).

There are two ways to implement this estimator (you should try to implement both if you have time):

(1) Enumerate all possible sequences of 1's and 0's that have same sum for each panel.

(2) Use the clever recursion proposed by Krailo and Pike (1984); the paper is posted at the URL above. Their notation is not very clear, so I have also posted a note on how to implement their recursive formula, reusing the notation from Greene for clarity.

If you implement the estimator both ways, you will see that the implementation in (1) runs considerably slower (in fact, for T > 10, the implementation of (1) is almost computationally infeasible). But using the recursive formula, T > 100 can be handled with ease. Make sure to always check for better algorithms; they can save you time!

Problem 4: Non-linear IV using GMM in Mata

URL: http://web.mit.edu/econ-gea/14.171/problem4/

In this problem you will implement a non-linear instrumental variables (IV) estimator in Mata using GMM. The moment condition is the following:

$$E[Z'\varepsilon] = E[Z'(y - e^{X'\beta})] = 0$$

The minimization problem (using the empirical analogs of the moments above) is the following:

$$\min_{\beta}(m'Wm)$$

$$m = \left(\frac{1}{N} Z'(y - e^{X'\beta})\right)$$

$$W = \frac{(Z'Z)^{-1}}{N}$$

The solution to this minimization problem is the non-linear IV estimator.

BONUS: Calculate the standard error of this estimate.