Economics 522A, Homework 10

Due Tuesday, April 24

1. Suppose that we are interested in estimating the parameters of a linear conditional mean function:
$$E[y_i|x_i] = \beta_1 + \beta_2 x_i.$$
Here, x_i is scalar.

   However, we do not observe x_i. Instead, we observe a "noisy" version
$$\tilde{x}_i = x_i + u_i.$$
Assume that u_i has mean 0 and is independent of x_i and y_i.

   (a) Will OLS of y_i on a constant and $\tilde{x}_i$ consistently estimate β_2? If not, what is the probability limit of $\hat{\beta}_2$? Make any additional assumptions necessary to justify your arguments. (E.g. assumptions on existence of moments.)

   (b) Suppose that we have a variable z_i, which is another noisy measure of x_i:
$$z_i = x_i + v_i,$$
where v_i is independent of x_i, y_i, and u_i. Is it possible to use z_i, $\tilde{x}_i$, and y_i to construct a consistent estimate of β_2? Show how to do this, and prove that your estimator is consistent.

2. Suppose that
$$(x_i, u_i)' \overset{\text{iid}}{\sim} N(0, \Sigma),$$
where Σ is a 2×2 covariance matrix. Also suppose that
$$y_i = \delta_1 + \delta_2 x_i + u_i.$$

   (a) For a given $\delta_1, \delta_2, \Sigma$, derive the distribution of y_i given x_i.

   (b) Give an example of two sets of parameter values $(\delta_1, \delta_2, \Sigma)$, and $(\tilde{\delta}_1, \tilde{\delta}_2, \tilde{\Sigma})$ which lead to the same conditional distribution of y_i given x_i. What do you conclude about the possibility to estimate δ_2 consistently, without some additional source of information?

3. Consider the following model:
$$\begin{aligned} y_i &= \beta x_i + u_i \\ x_i &= \gamma z_i + \{\gamma v_i + \lambda u_i\}. \end{aligned}$$
Suppose that u_i, v_i, and z_i are independent mean-zero normal random variables with variance σ_u^2, σ_v^2, and σ_z^2, respectively. Carry out a Monte Carlo experiment to assess the finite-sample distribution of the instrumental variables (IV) estimator b_{IV} of β, using z_i as an instrument for x_i.

   Set $\sigma_u = \sigma_v = \sigma_z = 1$, $\beta = 1$, and $\lambda = 1$. Assume i.i.d sampling with a sample size of $n = 100$.

(a) Set $\gamma = 2$ and approximate the sampling distribution of b_{IV} by Monte Carlo, using 1000 replications (so you should end up with 1000 independent draws for b_{IV}). Plot a histogram of the draws for b_{IV}. (Hint: to get greater detail using the `hist` command, use a larger number of bins than the default.)

Calculate the asymptotic approximation to the distribution of b_{IV} and compare it to the finite-sample distribution. Explicitly compare the mean, variance, and the overall shape of the two distributions.

(b) Repeat the experiment in (a) but with $\gamma = .5$. For this part and the next, when forming the histograms, it may be useful to only use the draws for b_{IV} that are within the interval $[-8, 8]$ to make the histograms easier to read.

(c) Repeat the experiment but with $\gamma = .1$. Comment on how the quality of the asymptotic approximation relates to γ.

4. This question uses the data set contained in the file `hw10.mat`. It is a Matlab binary data file, so when you download it you should save it as "binary" or "source." In matlab, type `load hw2.mat` to load the file, and `whos` to display the variables that have been loaded.

The file contains data from a study conducted at a hospital in the U.S. It has observations on 2893 patients. It records their age (`Age`), a binary variable indicating whether or not they received an influenza vaccine (`Flushot`), and a binary variable indicating whether or not they were later hospitalized for flu-related condition (`Hosp`). Note that the decision to get a flu shot was made by the patients (in conjunction with their doctors). However, before the advent of the flu season, the hospital sent out letters to a randomly chosen subset of the patients, encouraging them to get the flu shot. The variable `Letter` is equal to 1 if the patient got the letter, and 0 otherwise.

(a) Calculate the mean of `Hosp` among patients who received the vaccine (`Flushot` = 1) and those who didn't (`Flushot`=0).

(b) Regress hospitalization on a constant, the flu shot indicator, and age. Interpret the results. Provide standard errors that are "robust" against possible heteroskedasticity.

(c) Why might we be worried about interpreting the results from parts a and b as measuring the "effect" of the flu shot?

(d) Consider the equation

$$\texttt{Hosp}_i = \beta_0 + \beta_1 \texttt{Flushot}_i + \beta_2 \texttt{Age}_i + \epsilon_i.$$

where we do not assume that `Flushot` is orthogonal to ϵ_i. We do, however, assume that ϵ_i is orthogonal to `Age` and `Letter`. Use 2SLS to obtain estimates of the model parameters, and interpret the results. Report the first-stage F statistic, and standard errors under homoskedasticity, and standard errors allowing for heteroskedasticity.

(e) Comment on the exclusion restriction required for the IV analysis. Could it be violated here?