Assignment 7

Due: November 30, 2012

1. In this question, you will be considering the linear GMM estimator.

   (a) Show that when the function $\overline{m}_n(b)$ is the linear function

   $$\overline{m}_n(b) = \frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'b)$$

   then

   $$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'y.$$

   (b) Let $X \in \mathbb{R}^{n\times p}$ and $Z \in \mathbb{R}^{n\times \ell}$. Show that when $\ell = p$, $q_n(\hat{\beta}, W) = 0$ regardless of the chosen W, where $q_n(\hat{\beta}, W)$ is the GMM criterion function.

2. *The Wald estimator*: suppose you have the simple regression model

   $$y = \beta_0 + \beta_1 x + \varepsilon, \qquad \beta_1 \in \mathbb{R}$$

   and x suffers from endogeneity problems. Suppose you have an instrument z that is a binary variable (that is, it takes only values of 0 or 1).

   (a) Show that the IV estimate of β_1 is

   $$\hat{\beta}_{IV} = \frac{\overline{y}_1 - \overline{y}_0}{\overline{x}_1 - \overline{x}_0}$$

   where $\overline{y}_k$ means the mean of the subset of the data for which $z = k$, for $k = 0, 1$.

   (b) What is the interpretation of $\hat{\beta}_1$ if x is also binary?

3. {Working (1927), Hayashi}: Consider the simple supply and demand model for the market for widgets:

   $$\begin{aligned} q^d &= \alpha_0 + \alpha_1 p + u && \text{(demand equation)} \\ q^s &= \beta_0 + \beta_1 p + v && \text{(supply equation)} \\ q^d &= q^s && \text{(market equilibrium)} \end{aligned}$$

where q^d is the quantity demanded for widgets, q^s is the quantity supplied, and p is the market price. The error terms u and v represent unobserved demand and supply shifting factors. Assume $E(u) = 0$, $E(v) = 0$, and $Cov(u, v) = 0$. Assume also that p and q are distributed bivariate normal.

(a) Solve the system of equations for equilibrium price and quantity.

(b) Find the CEF $E(q|p)$. Does OLS provide a consistent estimate of α_1? β_1? Explain.

(c) Suppose you observe two more variables. The first is the price of aluminum, p_a, the primary material component in widgets. The second is the price of doohickies, p_d, a complimentary good to widgets. Assume that $Cov(p_i, j) = 0$ for $i = a, d$ and $j = u, v$. Suggest an estimation strategy.

4. {Card (1995)}: in this problem you will reproduce many of the results in Card (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". The data *card.csv* can be found on the course website as well as a document *CardReplication.pdf* describing the variables and replicating results found in Column (1) of Table 2 and Column (5) of Table 3 of Card's paper.

   (a) Estimate an OLS regression of log wage ($lwage$) on $educ$, $exper$, $expersq$, $black$, $south$, $smsa$, $reg661$ through $reg668$ and $smsa66$. Compare the results with Column (1) of the replication document.

   (b) Estimate a reduced form first stage regression of education on all the explanatory variables plus the dummy $nearc4$. Do $educ$ and $nearc4$ have (practical, statistical) significant partial correlation?

   (c) Estimate an IV regression of $lwage$ on the explanatory variables, using $nearc4$ as an instrument for $educ$. Compare confidence intervals for $\hat{\beta}_{educ}$ from this regression and the OLS regression from part (a). Also compare this to column (2) in the replication document.

   (d) Now add $nearc2$ to $nearc4$ as an instrument for education. Are the proximity variables well related to $educ$? Is there statistical evidence to suggest that $educ$ is indeed endogenous and $nearc2$ and $nearc4$ are exogenous instruments? How does the 2SLS estimate of β_{educ} compare to the OLS and simple IV estimates?

   (e) IQ scores are available for some of the men in the sample. Are iq and $nearc4$ correlated?

   (f) Regress iq on $nearc4$, $smsa66$, $reg661$, $reg662$ and $reg669$. Are iq and $nearc4$ partially correlated? What does this imply about the addition of location and region variables in the IV equation when $nearc4$ is an instrument?