14.385b Problem Set 1

Fall 2007

1. Another way to think about GMM is as setting a linear combination of moment conditions equal to zero. Consider notation as in class, where $g_i(\beta)$ is an $m \times 1$ vector of functions of the data and a $p \times 1$ vector of parameters β, with $E[g_i(\beta_0)] = 0$, and $\hat{g}(\beta) = \sum_i g_i(\beta)/n$. Let $\hat{L}$ be $p \times m$ matrix that can depend on the data. Consider an estimator β solving

$$\hat{L}\hat{g}(\beta) = 0$$

a) Computationally, would you prefer this version of GMM or one that minimizes $\hat{g}(\beta)'\hat{A}\hat{g}(\beta)$ for a positive semi-definite matrix $\hat{A}$? What if $m = p$? What if $m > p$?

b) Assume that $\hat{L} \xrightarrow{p} L$ for a matrix L with LG nonsingular, that $\sqrt{n}\hat{g}(\beta_0) \xrightarrow{d} N(0, \Omega)$, that $\sqrt{n}\partial\hat{g}(\bar{\beta})/\partial\beta \xrightarrow{p} G$ for any $\bar{\beta} \xrightarrow{p} \beta_0$. Derive the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$.

c) What is the optimal, asymptotic variance minimizing choice of L? Prove your answer. (Hint: Consider the same linear model analogy as in the notes and a linear estimator of the form $(LG)^{-1}LY$.)

2. One way to form tests of subsets of moment conditions is to add parameters to the model, one for each moment condition. Consider the same notation as in the notes, where $g_i(\beta) = (g_i^1(\beta)', g_i^2(\beta)')'$, Ω is partitioned conformably, and the null hypothesis is $H_0 : E[g_i^1(\beta_0)] = 0$, where $g_i^1(\beta)$ is $m_1 \times 1$. Consider augmenting the moment conditions by adding an $m_1 \times 1$ vector of parameters γ, with new moment functions

$$h_i(\beta, \gamma) = \begin{pmatrix} g_i^1(\beta) - \gamma \\ g_i^2(\beta) \end{pmatrix}.$$

a) Consider the null hypothesis $\gamma_0 = 0$. Show that this is the same as the null hypothesis $E[g_i^1(\beta_0)] = 0$.

b) Describe the optimal two-step GMM estimator for β for the new moment restrictions under the null hypothesis $H_0 : \gamma_0 = 0$ and under the alternative $\gamma_0 \neq 0$.

c) Derive formulas for the likelihood ratio, Wald, and LM tests of this null hypothesis on the parameter γ in the augmented moment condition model.

d) What is the relationship of these tests to the tests of subsets of moment conditions described in the notes?

3. Consider a model where the conditional mean and variance are linked as follows:

$$E[y_i|X_i] = X_i'\beta_0 + \alpha_0(e^{X_i'\gamma_0})^{\lambda_0}, Var(y_i|X_i) = e^{X_i'\gamma_0}.$$

a. Assuming that y_i is normally distributed conditional on X_i, derive the conditional likelihood for y_i given X_i.

b. Derive the first order conditions for the MLE. If y_i is not normal, will the MLE estimator still be consistent? Will it be efficient?

c. Consider a conditional moment restrictions model based on this set up. Derive a formula for the optimal instruments.

d. How would you construct a GMM estimator that is approximately efficient?

4. This is an empirical exercise using the Angrist and Krueger (1991) data from Josh Angrist's website. The model is linear model with schooling on the right hand side using quarter of birth as an instrument.

a. Calculate and report the 2LSLS, LIML, FULLER, and two-step optimal GMM estimator and heteroskedasticity consistent standard errors for each. How do the heteroskedasticity consistent standard errors compare with the usual ones?

b. Using the instrument selection criteria from Donald and Newey (2001), select the 2SLS estimator among using each quarter of birth and using all quarters of birth as instruments.

c. Calculate the JLIML estimator and heteroskedasticity and many instrument robust standard errors using all of the instruments. How does this compare with LIML and the usual standard errors?

5. This is a Monte Carlo exercise. Consider a linear simultaneous equations model of the form

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_i + \varepsilon_i, \\ x_i &= \pi_1 + \pi_2 z_i + v_i \end{aligned}$$

where ε_i and vi are joint normal, each with variance 1, and with covariance $\rho = .4$. Let the z_i be $N(0,1)$ and the sample size be 200.

a. What is the value of the concentration parameter in this model?

b. For values of the concentration parameter ranging from 8 to 24 in steps of 4, compute the rejection frequencies for nominal 5 and 10 percent tests concerning $H_0 : \beta_2 = 0$ using the usual standard errors, for 5000 replications, along with standard errors for those rejection frequencies.

c. In your view, is weak instruments a problem across this range of concentration parameters?