Econ 220B
James Hamilton

Final Exam
Winter 2008

DIRECTIONS: No books or notes of any kind are allowed. Answer all questions on separate paper. 250 points are possible on this exam

1.) (80 points total) This question concerns the following regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is a $(T \times 1)$ vector consisting of T different observations on a variable y_t to be explained, $\mathbf{X}$ is a $(T \times k)$ matrix of observations on k explanatory variables, and $\boldsymbol{\varepsilon}$ is a $(T \times 1)$ vector of residuals. Suppose that $\mathbf{X}$ has rank k and $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$ for $\mathbf{I}_T$ the $(T \times T)$ identity matrix.

a.) (10 points) Write down the formula for the OLS estimate of $\boldsymbol{\beta}$.

b.) (20 points) Consider the special case when there are $K = 2$ explanatory variables, so that the tth row of the above vector system looks like

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t.$$

Write down the formula for the OLS F test of the hypothesis that $\beta_1 = 2\beta_2$. State (but do not derive) the distribution of this statistic under the assumptions given above.

c.) (15 points) Suppose instead that you wanted to test the hypothesis that $\beta_1 = 2\beta_2$ using a t test rather than an F test. Write down the expression for the statistic you would use and state (but do not derive) its distribution.

d.) (15 points) How would your answer to part (b) change if you were to base your test statistic on White (robust) standard errors rather than the usual OLS standard errors?

e.) (20 points) How would your answer to part (b) change if your null hypothesis were instead that $\log \beta_1 = 2 \log \beta_2$?

Econ 220B
James Hamilton

2.) (70 points) An economist is studying the effect of military service on later earnings. She has wage data on a cross section of individuals, with y_i the wage earned by individual i, and $s_i = 1$ if the individual served in the military and $s_i = 0$ otherwise. The equation of interest is

$$y_i = \alpha + \beta s_i + \varepsilon_i.$$

a.) (10 points) Explain to the economist why she might not get a consistent estimate of β from a regression of y_i on a constant and s_i.

b.) (20 points) You then offer the economist a constructive suggestion for how she might estimate β consistently using $\mathbf{x}_i$, which is the following (73×1) vector of instruments:

$$x_{1i} = \begin{cases} 1 \text{ if person } i\text{'s birthday was one of the first 5 drafted in the birthdate-based lottery} \\ 0 \text{ otherwise} \end{cases}$$

$$x_{2i} = \begin{cases} 1 \text{ if person } i\text{'s birthday was one of the second 5 drafted in the birthdate-based lottery} \\ 0 \text{ otherwise} \end{cases}$$

$$\vdots$$

$$x_{73,i} = \begin{cases} 1 \text{ if person } i\text{'s birthday was one of the last 5 drafted in the birthdate-based lottery} \\ 0 \text{ otherwise} \end{cases}.$$

(Note there are 73 different 5-day groups in a 365-day year). Give a mathematical expression for what it means for the vector $\mathbf{x}_i$ to be a relevant instrument, and an economic argument why it might hold in this case. Is there a statistic you could look at to decide if the instrument is indeed relevant?

c.) (20 points) Give a mathematical expression for what it means for the vector $\mathbf{x}_i$ to be a valid instrument, and an economic argument why it might hold in this case. Is there a statistic you could look at to decide if the instrument is indeed relevant? (Hint: are there any overidentifying restrictions that could be tested?)

d.) (20 points) If you were to implement two-stage least squares for this case by hand, describe exactly what you would do to obtain the parameter estimates in each stage. (You only need describe the calculations you'd use to arrive at your estimates of α and β, not steps you'd use to calculate standard errors for anything.)

Econ 220B
James Hamilton

3.) (100 points total) If $E[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] = \mathbf{0}$, the GMM estimate of $\boldsymbol{\theta}$ is defined as the value that minimizes

$$T[\mathbf{g}(\boldsymbol{\theta}; \mathbf{Y}_T)]'\hat{\mathbf{S}}^{-1}[\mathbf{g}(\boldsymbol{\theta}; \mathbf{Y}_T)]$$

where

$$\mathbf{g}(\boldsymbol{\theta}; \mathbf{Y}_T) = T^{-1}\sum_{t=1}^{T}\mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t)$$

and $\hat{\mathbf{S}}$ is a consistent estimate of

$$\mathbf{S} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{v=-\infty}^{\infty} E\left\{[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)]\left[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t-v})\right]'\right\}.$$

The usual asymptotic distribution of the GMM estimator is

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{V})$$

where $\mathbf{V}$ can be estimated consistently from

$$\hat{\mathbf{V}} = (\hat{\mathbf{D}}\hat{\mathbf{S}}^{-1}\hat{\mathbf{D}}')^{-1}$$

$$\hat{\mathbf{D}}' = \left.\frac{\partial \mathbf{g}(\boldsymbol{\theta}; \mathbf{Y}_T)}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Suppose that y_t is distributed i.i.d. Bernoulli, meaning that it has probability p of taking on the value unity and probability $1-p$ of taking on the value zero. Some facts about Bernoulli variables that you may find useful in answering the following questions include the following:

$$E(y_t) = p$$
$$E(y_t^2) = p$$
$$E(1-y_t)^2 = 1-p$$
$$E(y_t - p)^2 = p(1-p)$$
$$E[(y_t)(1-y_t)] = 0.$$

Note also that the probability that observation t takes on the value y_t can be written as

$$p^{y_t}(1-p)^{1-y_t}$$

and the log likelihood function is

$$\mathcal{L}(y_1, y_2, ..., y_T; p) = N\log p + (T-N)\log(1-p)$$

for $N = y_1 + y_2 + \cdots + y_T$.

a.) (10 points) Show that the maximum likelihood estimate is given by $\hat{p} = N/T$.

b.) (20 points) Use the law of large numbers and central limit theorem to calculate the plim and asymptotic distribution of $\hat{p}$ directly from your formula in (a).

c.) (10 points) If one were to view this maximum likelihood estimation as an example of GMM, what function corresponds to $\mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t)$?

d.) (10 points) Prove directly for this problem (not relying on the general results about GMM in a maximum likelihood context) that $E[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] = \mathbf{0}$.

e.) (10 points) Calculate $\mathbf{S}$ for this example.

f.) (10 points) Calculate the value of $\hat{\mathbf{D}}'$ for this example. (Hint: as a check on your answers for (e) and (f), remember that plim $\hat{\mathbf{D}}' = -\mathbf{S}$.)

g.) (10 points) Use the results from (e) and (f) to calculate the value of $\hat{\mathbf{V}}$ in the GMM formulas above. What is the relation between this value and the results you derived in part (b)?

h.) (10 points) Suppose that you fear you have mis-specified the likelihood function in that y_t is not really independent from y_s but there is instead some serial dependence, though y_t is still stationary and ergodic with $E(y_t) = p$ for all t. Is maximization of the expression $\mathcal{L}(y_1, y_2, ..., y_T; p)$ above still going to give you a useful estimate of p? Explain.

i.) (10 points) How would you modify your answers to (b) and (e) above to account for serial dependence in y_t?