Econ 220B Final Exam
Winter 2020

DIRECTIONS: This is an open-book, open-note exam. You are free to consult answers to old exams or any other resource. You answers should be emailed to Professor Hamilton by 11:10 a.m. 250 points are possible on this exam.

1.) (55 points total) Consider the following regression model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{y}$ is a $(T \times 1)$ vector of observations on the dependent variable, $\mathbf{X}$ is a $(T \times k)$ matrix of observations on explanatory variables, and $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of parameters we want to estimate. You can assume that $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{V})$ where $\mathbf{V}$ is a known $(T \times T)$ matrix (meaning that you don't need to estimate it from the data) and σ^2 is an unknown scalar (meaning we do need to estimate it from the data).

a.) (30 points) Suppose you were interested in testing the hypothesis that β_1 (the first element of $\boldsymbol{\beta}$) is equal to zero. Write down the formula for what you think would be the best statistic to use for purposes of testing this hypothesis. What distribution would you use to interpret this statistic? (Note: you do not need to derive or prove that this is the distribution you should use, just state it). Is the justification for using this distribution asymptotic, or is it an exact small-sample test?

b.) (15 points) Now suppose you were interested in testing the hypothesis that $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$ where $\mathbf{g} : \mathbb{R}^k \rightarrow \mathbb{R}^2$ is a known (2×1) vector-valued function whose first derivative is continuous. Write down the expression you would use for purposes of testing this hypothesis. What distribution would you use to interpret this statistic? (Note: you do not need to derive or prove that this is the distribution you should use, just state it). Is the justification for using this distribution asymptotic, or is it an exact small-sample test?

c.) (10 points) Comment on the advantages and disadvantages of the test you proposed in part (a) compared to a test based on the ratio of the OLS estimate of the first element of $\boldsymbol{\beta}$ to the White standard error (also known as the "robust" standard error or the "heteroskedasticity-consistent" standard error).

2.) (25 points) Suppose that two stationary and ergodic variables y_t and x_t are related according to $y_t = g(x_t) + u_t$ where $g(.)$ is a nonlinear function whose first derivative is continuous and $E(u_t|x_t) = 0$. The OLS coefficient from a linear regression of y_t on x_t is given by $b = \left(\sum_{t=1}^T x_t^2\right)^{-1} \left(\sum_{t=1}^T x_t y_t\right)$. Calculate the plim of b under the assumptions stated.

3.) (40 points total) Consider an OLS regression $y_t = \beta_0 + \beta_1 z_t + u_t$ where the scalar z_t is correlated with u_t. A researcher proposes to estimate the values of β_0 and β_1 by making use of an $(r \times 1)$ vector of instruments $\mathbf{x}_t$ for which it is the case that $E(\mathbf{x}_t u_t) = 0$.

a.) (10 points) What is the smallest possible value of r for which this could be feasible?

b.) (30 points) Describe a test you could use in order to assess whether the instruments $\mathbf{x}_t$ might be weak. Provide as much detail as you can on exactly how you would implement the test (though you do not need to provide any derivations of claims about the distributions of the statistics you propose to look at).

4.) (130 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

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

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

In this question you are asked to apply these results to maximum likelihood estimation of a binomially distributed random variable. Let x_t be a random variable that takes on one of the possible values $k = 0, 1, ..., n$ with probability

$$P(x_t = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$k! = \begin{cases} k(k-1)(k-2)\cdots 1 & \text{if } k = 1, 2, ..., n \\ 1 & \text{if } k = 0 \end{cases}.$$

Such a variable is said to have a binomial distribution with parameter p. This random variable has mean and variance given by

$$E(x_t) = np$$

$$E(x_t - np)^2 = np(1-p).$$

You can assume throughout this question that the value of n is known and the only thing you need to estimate is p. If we observe a sample of T observations $x_1, ..., x_T$ drawn independently from this distribution, the log likelihood is given by

$$\ell(p) = c + \textstyle\sum_{t=1}^T x_t \log p + \sum_{t=1}^T (n - x_t) \log(1-p)$$

$$c = T \log n! - \textstyle\sum_{t=1}^T \log(x_t!) - \sum_{t=1}^T \log[(n - x_t)!].$$

a.) (20 points) Show that the maximum likelihood estimate is given by $\hat{p}_{MLE} = \sum_{t=1}^T x_t/(Tn)$. Hint: note that c does not depend on p.

b.) (25 points) Use the Central Limit Theorem to find the asymptotic distribution of $\hat{p}_{MLE}$. Did you need to make any assumptions in addition to those stated above to derive this result?

c.) (25 points) The score for observation t is defined as the derivative of the log of the likelihood of observation t with respect to p, and is denoted $h(p, x_t)$. Calculate the expression for $h(p, x_t)$ for this example and show that $E[h(p, x_t)] = 0$ if $p = p_0$ and $E[h(p, x_t)] \neq 0$ if $p \neq p_0$ where p_0 denotes the true value of p.

d.) (15 points each, 60 points total) Let $\hat{p}_{GMM}$ denote the GMM estimate based on using $h(p\,, x_t)$ to construct the moment condition. Calculate the values of the following under the assumptions stated above:

i.) $\hat{p}_{GMM}$
ii.) $\hat{\mathbf{D}}'$
iii.) plim $\hat{\mathbf{D}}'$
iv.) $\hat{\mathbf{S}}$