ECON7772 Econometric Methods

Professor Arthur Lewbel

SAMPLE MIDTERM EXAM QUESTIONS

INSTRUCTIONS:
Write your name on the exam book. Write your answers in the exam book.
You have one hour and fifteen minutes to complete this exam.
You may bring notes on a single sheet of paper to the exam.
Answer all the problems clearly, completely, and concisely. Show all your work.

1. Suppose we flip a fair penny many times and we flip a fair quarter once. Let the random variable X_n be a zero or one, depending on if the $n'th$ flip of the penny is a tail or a head. Also let the random variable Y be a zero or one, depending on if the flip of the quarter is a tail or a head.
a. Does $X_n \to^d Y$? Explain your answer.
b. Does $X_n - Y \to^p 0$? Explain your answer.

2. Let the Gauss-Markov assumptions hold for $Y = X\beta + \varepsilon$. Let $\widehat{\beta}$ be the Ordinary Least Squares (OLS) estimator of the k vector β and let $\widehat{\varepsilon}$ be the n vector of OLS residuals.
a. Write $\widehat{\varepsilon}$ as a function of just X and ε.
b. Using your answer to part a, find the simplest expression you can for $cov(\widehat{\beta}, \widehat{\varepsilon})$, which is a k by n matrix.

3. Suppose $Y_i = X_i'B + e_i$ for a K vector B. Assume all of the Gauss-Markov assumptions hold, each element of each X_i lies between 1 and 5, and that each e_i is normal. Consider the following estimator: For any sample size $n > 10$, discard all but the last 10 observations, and let $\widehat{B}$ be the ordinary least squares estimate of B using just these last 10 observations. Answer each of the following questions, and show why your answer is correct:
a. Is $\widehat{B}$ unbiased?
b. Is $\widehat{B}$ most efficient among linear estimators?
c. Is $\widehat{B}$ consistent?
d. Is $\widehat{B}$ asymptotically normal?
e. Suppose that instead of the last 10 observations, we now used the last $n/2$ observations to estimate $\widehat{B}$. Which of your answers above would change and which would stay the same?

4. For a random variable Y_i that can only equal zero or one, the log of the probability of an observation Y_i can be written as $Y_i \ln\left[prob(Y_i = 1)\right] + (1 - Y_i)\ln\left[prob(Y_i = 0)\right]$. Assume that $prob(Y_i = 1) = a$ and that $0 < a < 1$. Suppose our data consists of n independent observations of Y_i.

a. What is the log likelihood function for this model and data?

b. What is the score function and the information matrix for this model and data?

c. What is the maximum likelihood estimator for a in this model?

d. Is the maximum likelihood estimator for a unbiased in this model? Prove your answer.

5. Assume that (ε_i, X_i) for $i = 1, ..., n$ are bounded iid random vectors, $X_i > 0$, and $E\left(\varepsilon_i \mid X_i\right) = 0$. Assume $Y_i = e^{\theta_0 X_i} + \varepsilon_i$ with $|\theta_0| \leq 1$. Consider the estimator $\widehat{\theta} = \arg\min_{|\theta| \leq 1} \sum_{i=1}^{n} \left(Y_i - e^{\theta X_i}\right)^2$. Either prove that $plim\left(\widehat{\theta}\right) = \theta_0$, or show that some assumption we would use to prove this condition may be violated.

ECON7772 Econometric Methods

SAMPLE MIDTERM EXAM - ANSWERS

1. a. Yes, $X_n \to^d Y$ does hold. Let $F_n(X_n)$ and $G(Y)$ be the cdf's of X_n and of Y. They both have the same binomial distribution function (1/2 chance of equaling zero and 1/2 chance of equaling one) for every n. That is, $F_n(0) = G(0) = 1/2$ and $F_n(1) = G(1) = 1$. Therefore for every possible value of x (that is, for $x = 0$ and for $x = 1$) we have $|F_n(x) - G(x)| = 0$ for all n, so $\lim_{n\to\infty} |F_n(x) - G(x)| = 0$, which satisfies the definition of convergence in distribution.

1. b. No $X_n - Y \to^p 0$ does not hold. There is a 50% chance that $|X_n - Y|$ $= 1$ (this happens when $X_n = 0$ and $Y = 1$ or vice versa, that is, when on the n'th toss, the penny comes up heads and the quarter comes up tails, or vice versa). Applying the definition of convergence in probability, having $X_n - Y \to^p 0$, requires that $\lim_{n\to\infty} prob(|X_n - Y| \geq \varepsilon) = 0$ for any small ε, while in our case we have $\lim_{n\to\infty} prob(|X_n - Y| \geq \varepsilon) = 1/2 \neq 0$ for any small (less than one) value of ε.

2. a. $\widehat{\varepsilon} = Y - X\widehat{\beta} = Y - X(X'X)^{-1}X'Y = (X\beta + \varepsilon) - X(X'X)^{-1}X'(X\beta + \varepsilon)$ $= (X\beta + \varepsilon) - (X\beta + X(X'X)^{-1}X'\varepsilon) = \varepsilon - X(X'X)^{-1}X'\varepsilon = [I_n - X(X'X)^{-1}X']\varepsilon$

2. b. We know for ordinary least squares that $E(\widehat{\beta}) = \beta$ and $\widehat{\beta} - E(\widehat{\beta}) =$ $(X'X)^{-1}X'\varepsilon$. Also, $E(\widehat{\varepsilon}) = [I_n - X(X'X)^{-1}X']E(\varepsilon) = 0$ so $cov(\widehat{\beta}, \widehat{\varepsilon}) = E\left[\left(\widehat{\beta} - E(\widehat{\beta})\right)\widehat{\varepsilon}'\right]$ $= E\left[\left(\widehat{\beta} - E(\widehat{\beta})\right)([I_n - X(X'X)^{-1}X']\varepsilon)'\right] = E\left[(X'X)^{-1}X'\varepsilon\varepsilon'[I_n - X(X'X)^{-1}X']'\right] =$ $(X'X)^{-1}X'E(\varepsilon\varepsilon')[I_n - X(X'X)^{-1}X'] = \sigma^2(X'X)^{-1}X'I_n[I_n - X(X'X)^{-1}X'] = \sigma^2\left[(X'X)^{-1}X' - (X'X)^{-}\right.$ $= \sigma^2\left[(X'X)^{-1}X' - (X'X)^{-1}X'\right] = 0$, a k by n matrix of zeroes. This shows that the estimated residuals $\widehat{\varepsilon}$ are uncorrelated with the estimated coefficients $\widehat{\beta}$.

3 a. Yes, $\widehat{B}$ is unbiased. Let $\widetilde{X}$ be the matrix containing the last 10 observations of X and let $\widetilde{Y}$ and $\widetilde{e}$ be the vectors of the last 10 observations of Y and e. Then $E\left(\widehat{B}\right) = E\left[\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\widetilde{Y}\right] = E\left[B + \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\widetilde{e}\right] = B + \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'E\left(\widetilde{e}\right) =$ B.

3 b. No, $\widehat{B}$ is not most efficient among linear estimators. Intuitively, throwing away information from discarding observations loses efficiency, and OLS with all the

data is BLUE. Formally, $\sum_{i=n-10}^{n} X_i X_i' < \sum_{i=1}^{n} X_i X_i'$ because of the bounds on each X_i, and therefore
$Var\left(\widehat{B}\right) = \sigma^2 \left(\widetilde{X}'\widetilde{X}\right)^{-1} = \sigma^2 \left(\sum_{i=n-10}^{n} X_i X_i'\right)^{-1} > \sigma^2 \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} = Var(\text{OLS}$ estimate of $B)$.

3 c. No, $\widehat{B}$ is not consistent. Intuitively, the number of observations used to estimate $\widehat{B}$ stays fixed at 10, and consistency requires the number of observations to go to infinity. Formally: $plim\left(\widehat{B}\right) = plim\left[B + \left(\sum_{i=n-10}^{n} X_i X_i'\right)^{-1}\left(\sum_{i=n-10}^{n} X_i e_i\right)\right]$
$= B + \left(lim \sum_{i=n-10}^{n} X_i X_i'/10\right)^{-1}\left(plim \sum_{i=n-10}^{n} X_i e_i/10\right)$, and this last probability limit does not converge to zero, or to any constant. The limit of $\sum_{i=n-10}^{n} X_i X_i'/10$ may also not be well defined, but is positive and finite because of the bounds on X_i.

3 d. Yes, $\widehat{B}$ is asymptotically normal. Not only is it asymptotically normal, it is exactly normal for all sample sizes because $\widehat{B} = B + \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\widetilde{e}$ is linear in the normal vector $\widetilde{e}$, since B and $\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'$ are not random.

3 e. The only answer that would change is c. One gets the same results as above in parts a., b., and d. when replacing 10 with $n/2$, but in part c. we get $lim \sum_{i=1}^{n/2} X_i X_i'/(n/2)$ is finite and positive because of the bounds on X_i and $plim \sum_{i=n/2}^{n} X_i e_i/(n/2) = E(X_i e_i) = 0$ by the law of large numbers (which applies because $X_i e_i$ are i.n.i.d with finite moments higher than two) so now $plim\left(\widehat{B}\right) = B$.

4 a.The Log likelihood function is $\ln L = \sum_{i=1}^{n} \ln(Prob(Y_i))$
$= \sum_{i=1}^{n} Y_i \ln[prob(Y_i = 1)] + (1 - Y_i)\ln[prob(Y_i = 0)] = \sum_{i=1}^{n} Y_i \ln(a) + (1 - Y_i)\ln(1 - a)$.

4 b. The score function is $s(Y_i \mid a) = \partial[Y_i \ln(a) + (1 - Y_i)\ln(1 - a)]/\partial a = \frac{Y_i}{a} - \frac{1-Y_i}{1-a} = \frac{Y_i - a}{(1-a)a}$. The matrix $J_0 = var[s(Y_i \mid a_0)] = E\left[\left(\frac{Y_i - a}{(1-a)a}\right)^2\right] = E\left(\frac{Y_i^2 - 2aY_i + a^2}{(1-a)^2 a^2}\right) = \frac{a - 2a^2 + a^2}{(1-a)^2 a^2} = \frac{1}{(1-a)a}$ and so the information matrix is $I_n = nJ_0 = \frac{n}{(1-a)a}$. Note that the matrices J_0 and I_n are both scalars here, because the parameter vector a is a scalar.

4 c. The first order conditions for the ML estimator $\widehat{a}$ are

$0 = \frac{1}{n}\sum_{i=1}^{n} s\left(Y_i \mid \widehat{a}\right)$, so $0 = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i-\widehat{a}}{(1-\widehat{a})\widehat{a}} = \frac{\overline{Y}-\widehat{a}}{(1-\widehat{a})\widehat{a}}$, which when solved for $\widehat{a}$ gives $\widehat{a} = \overline{Y}$, the sample mean of the Y_i data. (One can also check that the second order conditions for a maximum, $\partial^2 \ln L/\partial a^2 < 0$, are satisfied).

4. d. $E(\widehat{a}) = E(\overline{Y}) = E(Y) = 1 * [prob(Y = 1)] + 0 * [prob(Y = 0)] = prob(Y = 1) = a$, so while ML estimators are not unbiased in general, in this particular application the ML estimator $\widehat{a}$ is unbiased.

5. Let $Q_n(\theta) = \frac{-1}{n}\sum_{i=1}^{n}\left(Y_i - e^{\theta X_i}\right)^2$. We multiply by $\frac{-1}{n}$ to make it a max and to so it has a plim:

By law of large numbers (using Y and X bounded): $Q_0(\theta) = plim\frac{-1}{n}\sum_{i=1}^{n}\left(Y_i - e^{\theta X_i}\right)^2 = -E\left(\left(Y - e^{\theta X}\right)^2\right)$ is finite. Checking the remaining conditions for consistency of an extremum estimator:

For identification, $0 = \frac{\partial Q_0(\theta)}{\partial\theta} = -2E\left(\left(Y - e^{\theta X}\right)\left(-Xe^{\theta X}\right)\right)$ so $0 = E\left(\left(e^{\theta_0 X} - e^{\theta X}\right)\left(-Xe^{\theta X}\right)\right)$ which, since X is strictly positive, holds only at $\theta = \theta_0$.

For compactness, $|\theta_0| \leq 1$ so $-1 \leq \theta_0 \leq 1$. This is a closed interval, which is a compact set.

For smoothness, $Q_n(\theta)$ is differentiable in θ.

For stochastic equicontinuity, again using the law of large numbers, $plim\frac{\partial Q_n(\theta)}{\partial\theta} = -2E\left(\left(Y - e^{\theta X}\right)\left(-Xe^{\theta X}\right)\right)$. Call this function $c(\theta)$, which is bounded by boundedness of Y, X, and θ. It follows that $|\frac{\partial Q_n(\theta)}{\partial\theta}| = |c(\theta)| + o_p(1) = O_p(1)$.

So all the conditions to show that $\widehat{\theta}$ is consistent are satisfied.