Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

July 8, 2019

There are **FOUR** questions. Answer each part of each question. All questions are weighted equally. Within each question, each part will receive equal weight in grading. You have 20 minutes to read the exam and then four hours to complete the exam.

I. Consider (X, Y) with joint p.d.f. $f_{X,Y}(x,y) = \begin{cases} (2x^3 + y) & 0 \leqslant x \leqslant 1, \quad 0 \leqslant y \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$

(i) Obtain $f_X(x)$, the marginal density of X.

(ii) Obtain $f_Y(y)$, the marginal density of Y.

(iii) Obtain $E[Y]$.

(iv) Obtain $\Pr[Y \leqslant 0.5]$.

(v) Are X and Y independent? Explain.

(b) This part has various unrelated questions.

(i) Given $P(A \cap B) = P(A|B) \times P(B)$ prove Bayes rule.

(ii) Prove that $Var[X] = E[X^2] - \mu^2$ where $\mu = E[X]$.

(iii) Prove that if a continuous random variable X has moment generating function $M(t)$ then $M'(0) = E[X]$.

(c) Suppose we have a random sample $x_1, ..., x_n$ of size n from a distribution with c.d.f. $F(x;\theta) = 1 - \exp(-x^2/\theta^2)$, density $f(x;\theta) = (2x/\theta^2)\exp(-x^2/\theta^2)$, $x \geqslant 0$, $\theta > 0$.

The r^{th} moment of X is given by $E[X^r] = \theta^r \Gamma(1 + \frac{r}{2})$ where $\Gamma(\cdot)$ is the gamma function and $\Gamma(1.5) \simeq 0.8862$, $\Gamma(2) = 1$, $\Gamma(2.5) \simeq 1.3293$, and $\Gamma(3) = 2$.

(i) Obtain the first-order conditions for the MLE of θ.

(ii) Is there an explicit solution for $\widehat{\theta}$? If so, give it.

(iii) Give the limit distribution of $\sqrt{n}(\widehat{\theta} - \theta)$ as $n \to \infty$.

(iv) Suppose $\widehat{\theta} = 3.2$ and $n = 100$. Do you reject $H_0 : \theta = 3$ against $H_a : \theta \neq 3$ at level 0.05?

(d) Continue with the same setup as part (c).

(i) Provide a precise algorithm to generate by a computer a random sample of X from the distribution given in part (c) for the case $\theta = 2$.

(ii) How would you use a computer to determine whether the MLE in part (c) is unbiased for θ?

(iii) How would you use a computer to determine whether the MLE in part (c) is most likely consistent for θ?

(iv) Consider the alternative estimator $\widetilde{\theta} = \bar{X}/\Gamma(1.5)$.

Obtain the mean and variance of this estimator.

(v) Given your answers in part(c)(iii) and part(d)(iv) which estimator is more efficient asymptotically: $\widehat{\theta}$ or $\widetilde{\theta}$?

II. **Linear Regression**

Consider the model $y_i = \beta_0 + \beta_1 x_i + e_i$, where x_i is scalar, $E[x_i] = 1$, $E[e_i|x_i] = 0$ and $E[e_i^2|x_i] > 0$. You have an *iid* random sample of size n. Define a dummy variable d_i that equals one if $x_i > 1$ and zero otherwise. Consider the following estimators:

$$\tilde{\beta}_1 = \frac{m^{-1}\sum_{i=1}^{m} y_i - m^{-1}\sum_{i=m+1}^{n} y_i}{m^{-1}\sum_{i=1}^{m} x_i - m^{-1}\sum_{i=m+1}^{n} x_i}$$

$$\bar{\beta}_1 = \frac{c^{-1}\sum_{i=1}^{n} d_i y_i - (n-c)^{-1}\sum_{i=1}^{n}(1-d_i)y_i}{c^{-1}\sum_{i=1}^{n} d_i x_i - (n-c)^{-1}\sum_{i=1}^{n}(1-d_i)x_i}$$

where $m = n/2$ and $c = \sum_{i=1}^{n} d_i$. For simplicity, assume n is an even number.

(a) Is $\tilde{\beta}_1$ unbiased for β_1? If so, prove it. If not, state additional conditions you require for unbiasedness and prove unbiasedness under those conditions.

(b) Is $\tilde{\beta}_1$ consistent for β_1? If so, prove it. If not, either (i) state additional conditions you require for consistency and prove consistency under those conditions, or (ii) explain why no such conditions exist.

(c) Is $\bar{\beta}_1$ consistent for β_1? If so, prove it. If not, either (i) state additional conditions you require for consistency and prove consistency under those conditions, or (ii) explain why no such conditions exist.

(d) Following on from (c), find the asymptotic distribution of $\sqrt{n}(\bar{\beta}_1 - \beta_1)$ as $n \to \infty$. State any additional assumptions you require.

(e) Derive the finite-sample variance of the two estimators conditional on $\{x_1, x_2, ..., x_n\}$. State any additional assumptions you require.

(f) Write down a 95% confidence interval for β_1. State any additional assumptions you require.

(g) Is $\tilde{\beta}_1$ or $\bar{\beta}_1$ a more efficient estimator for β_1? Explain in words.

(h) Now, suppose we observe another variable z_i and that $E[y_i|x_i, z_i] = \gamma_0 + \gamma_1 x_i + \gamma_2 z_i$. Is $\bar{\beta}_1$ consistent for γ_1? If so, prove it. If not, either (i) state additional conditions you require for consistency and prove consistency under those conditions, or (ii) explain why no such conditions exist.

III. **Variance Estimation in Linear Models.**

(a) Consider the linear regression model $y = X\beta_0 + e$, where y and e are $n \times 1$ vectors, β_0 a $K \times 1$ vector and X an $n \times K$ matrix (n denotes the sample size). Assume $e|X \sim N(0, \sigma_0^2 I_n)$, $\sigma_0^2 > 0$. The OLS residuals are $\hat{e} = y - X\hat{\beta}$.

  (i) Show that $E(s^2) = \sigma_0^2$, where $s^2 = \hat{e}'\hat{e}/(n-K)$.

  (ii) Show that s^2 is independent of $\hat{\beta}$. Explain in words the implications of this result for inference about β_0.

  (iii) Derive the distribution of s. You may use either asymptotic or finite-sample theory. State any additional assumptions you require.

(b) For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, $y_{it} \overset{i.i.d.}{\sim} N(\mu_i, \sigma_0^2)$, where the i.i.d. assumption holds across t and the cross-sectional independence assumption is maintained (across i). Note that $\{\mu_i\}_{i=1}^n$ and σ_0^2 are unknown.

  (i) Derive the maximum likelihood estimator of σ_0^2.

  (ii) Under the assumptions of this problem, derive the bias of this estimator.

  (iii) Propose an unbiased estimator of σ_0^2.

  Note: The density of a random variable $X \sim N(\mu, \sigma^2)$ evaluated at x is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{ -\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2} \right\}.$$

IV. **Estimation of Binary Outcome Models under Exogeneity and Endogeneity**.

Binary outcomes, such as employment status or product choice, are widely studied in economics. Consider the general class of binary outcome models, where for $i = 1, \ldots, n$, $y_i|x_i \overset{i.i.d.}{\sim} Bernoulli(F(x_i'\theta_0))$, where $F(.)$ takes values between 0 and 1, $dim(x_i) = dim(\theta_0) = k$. All asymptotics in this question pertain to $n \to \infty$.
Notation: For a vector v, v^j refers to its j^{th} element.

(a) Assume that the proposed model is correctly specified.

(i) Propose a consistent and asymptotically efficient estimator of θ_0.

(ii) Provide conditions for its consistency and asymptotic normality. Make sure to write down its asymptotic distribution.

(iii) Construct the Wald, score and likelihood ratio statistics of the null hypothesis $H_0 : \theta_0^1 = 0$. Make sure to write down the asymptotic distribution of the test statistics under the null hypothesis.

(iv) Now suppose the model is misspecified, describe exactly what aspects of your answer in (i) and (ii) would change.

(b) Now let $y_i = h(x_i'\theta_0) + u_i$. Suppose $E[x_i^1 u_i] \neq 0$, whereas $E[u_i | x_i^2, x_i^3, \ldots, x_i^k] = 0$. There are two instruments however, z_i^1 and z_i^2, which satisfy $E[z_i^1 u_i] = 0$, $E[z_i^2 u_i] = 0$.

(i) Propose a consistent and asymptotically efficient estimator of θ_0.

(ii) Derive conditions for its consistency and asymptotic normality. Make sure to write down the asymptotic distribution.

(iii) Propose a consistent estimator of the asymptotic variance you derived in the previous question.

(iv) Does your model have over-identifying restrictions? If yes, propose a test of these restrictions and write down its asymptotic distribution under the null hypothesis. If not, explain why.