# Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

August 19, 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. **(a)** Consider $(X, Y)$ with joint p.d.f. $f_{X,Y}(x,y) = \begin{cases} 6xy^2 & 0 < x < 1, \quad 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$

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

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

**(iii)** Obtain $E[Y]$ and $Var[Y]$.

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

**(v)** Are $X$ and $Y$ independent? Explain.

**(b)** These are unrelated questions.

**(i)** Suppose random variable $X$ has density $f_X(x) = \frac{1}{3}$, $2 \leqslant x \leqslant 5$.

Find the moment generating function of $X$.

**(ii)** Suppose $X$ has mean 1 and standard deviation 2.

Provide a range $[a, b]$ for which $\Pr[a \leqslant X \leqslant b\ ] \geqslant 0.75$.

**(iii)** On average there is a 1% chance of having breast cancer. 90% of mammograms detect breast cancer when it is there. 5% of mammograms detect breast cancer when it is **not** there.

What is the probability that a test on a randomly chosen person detects cancer?

**(c)** Suppose we have a random sample $x_1, ..., x_n$ of size $n$ from a distribution with density $f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp(-x^2/2\theta)$, $-\infty < x < \infty$, $\theta > 0$. $X$ has first four moments $E[X] = 0$, $E[X^2] = \theta$, $E[X^3] = 0$, $E[X^4] = 3\theta^2$.

**(i)** Obtain the first-order conditions for the MLE of $\theta$.

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

**(iii)** Using standard results for the MLE, give the limit distribution of $\sqrt{n}(\widehat{\theta} - \theta)$.

**(iv)** Suppose $\widehat{\theta} = 2.7$ and $n = 200$. Do you reject $H_0 : \theta = 3$ against $H_a : \theta \neq 3$ at level 0.05?

**(d)** These are unrelated questions.

**(i)** Obtain the density of $Y = X^3$ when $X$ has density $f(x) = 2x\exp(-x^2)$, $x \geqslant 0$.

**(ii)** If $\widehat{\theta} \overset{a}{\sim} N(\theta, 4)$ find the asymptotic distribution of $\widehat{\gamma} = \widehat{\theta}^2$.

**(iii)** Suppose a random variable $X$ has mean $\mu$. Provide an example of an estimator of $\mu$ that is biased for $\mu$ but is consistent for $\mu$.

**(iv)** Suppose a random variable $X$ has mean $\mu$. Provide an example of an estimator of $\mu$ that is unbiased for $\mu$ but is inconsistent for $\mu$.

## II. Linear Regression

Consider the linear regression model $y = X\beta + e, E(X'e) = 0$. Consider the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$.

(a) Is $\hat{\beta}$ an unbiased estimator for $\beta$? If yes, provide a proof. If no, state additional sufficient condition(s) for unbiasedness and prove unbiasedness under these assumptions.

(b) Show that $\hat{\beta}$ is consistent for $\beta$. State any assumptions you make.

(c) Find the asymptotic distribution of $\hat{\beta}$. State any assumptions you make.

(d) Suppose you wish to test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$, where $\beta_1$ is the first element of $\beta$. Write down the t statistic for this hypothesis and derive its asymptotic null distribution. State any assumptions you make.

(e) Suppose $\beta_1 = 0.1$, and consider the hypothesis test in part (d). What is the power of the test in the limit as $n \to \infty$.

(f) Suppose $\beta_1 = 0.1$, and consider the hypothesis test in part (d). Using asymptotic theory, what sample size would give approximately the same power as if $\beta_1 = 0.9$ and n=100?

(g) Suppose you observe the variable $Z$ instead of $X$, where $Z = X + u$. Under what conditions is $\tilde{\beta} = (Z'Z)^{-1}Z'y$ consistent for $\beta$? Prove consistency and describe a realistic empirical setting under which these conditions would apply. If there are no such settings, explain what is unrealistic about the conditions.

III. **Generalized Least Squares**

(a) Consider the linear regression model $y = X\beta + e, E(e|X) = 0$. Assume the observations $(y_1, x_1), (y_2, x_2), ..., (y_n, x_n)$ are independent and identically distributed. Consider two estimators: $\hat{\beta} = (X'X)^{-1}X'y$ and $\tilde{\beta} = (X'WX)^{-1}X'Wy$, where W is a positive definite matrix.

(i) Suppose $E(ee'|X) = \sigma^2 I_n$. Which variance is larger, $var(\tilde{\beta}|X)$ or $var(\hat{\beta}|X)$? Prove it.

(ii) Suppose $E(ee'|X) = W^{-1}$. Which variance is larger, $var(\tilde{\beta}|X)$ or $var(\hat{\beta}|X)$? Prove it.

(b) Let $y_{it} = x'_{it}\beta + a_i + u_{it}$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$. All asymptotics in this question pertain to $n \to \infty$ while holding $T$ fixed. Let $X_i = (x_{i1}, \ldots, x_{iT})$ and $u_i = (u_{i1}, \ldots, u_{iT})'$. Assume $E[a_i|X_i] = 0$, $E[u_{it}|X_i, a_i] = 0$, $E[a_i^2|X_i] = \sigma_a^2$, $E[u_{it}u_{i,t-\tau}|X_i, a_i] = \rho^\tau$ for any integer $\tau$.

Under the above scenario:

(i) Write down the conditional variance covariance matrix of $v_i = a_i\mathbf{1}_T + u_i$, $\Omega = E[v_i v_i'|X_i]$, where $\mathbf{1}_T$ is a $T \times 1$ vector of ones.

(ii) Would the pooled OLS estimator yield a consistent estimator for $\beta$? Write down the estimator and justify your answer by showing the probability limit of the pooled OLS estimator. Make any additional assumptions you require to show the probability limit of the pooled OLS estimator

(iii) Is the pooled OLS estimator asymptotically efficient? If yes, explain formally. If not, propose an asymptotically efficient estimator.

IV. (a) Suppose that for $i = 1, \ldots, n$, $y_i^* = x_i'\beta_0 + u_i$, but the researcher can *only* observe $y_i = y_i^* 1\{y_i^* \geqslant 0\}$ and $x_i$, where $1\{A\}$ equals 1 when the event $A$ holds and zero otherwise. Assume that $E[u_i|x_i] = 0$, where $dim(x_i) = dim(\beta_0) = k$, and maintain the i.i.d. assumption across $i$.

(i) What is the conditional expectation of $y_i^*$ given $x_i$? What is the conditional expectation of $y_i$ given $x_i$? Given your answer, can $\beta_0$ be consistently estimated using OLS regression of $y_i$ on $x_i$? Explain formally, but without deriving the probability limit of the OLS estimator.

(ii) Now suppose that $u_i|x_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$, propose a consistent estimator of $\beta_0$. Provide sufficient conditions for its consistency.

(iii) Derive the asymptotic distribution for the estimator you propose in (ii). Provide any conditions required for the result. Is the estimator efficient? If yes, state clearly the conditions under which it is efficient. If not, propose an efficient estimator.

(iv) Let $\beta_0^j$ denote the $j^{th}$ element of $\beta_0$. Propose the Wald statistic to test $H_0 : \sum_{j=1}^k \beta_0^j = 1$. Derive the asymptotic distribution of the statistic under the null hypothesis. State any conditions required for the result.

(b) In the following question, let $i = 1, \ldots, n$, $t = 1, \ldots, T$, all asymptotic arguments pertain to $n \to \infty$ while holding $T$ fixed. You can maintain the i.i.d. assumption across $i$, but make no assumptions along the time-series dimension. For a time-varying variable $w_{it}$, $w_i \equiv (w_{i1}, \ldots, w_{iT})$.

Note: For each of the following questions, make sure to solve for the closed-form solution of each estimator if it exists.

(i) Let $y_{it} = \rho_0 y_{i,t-1} + \beta_0 x_{it} + \alpha_i + \epsilon_{it}$, where $y_{i,t-1}$ is sequentially exogenous and $x_{it}$ is strictly exogenous, $dim(\beta_0) = dim(x_{it}) = 1$. Define the exogeneity condition for this problem clearly. Propose two estimators of $(\rho_0, \beta_0)'$, one based on a just-identifying set of moment condtions and the other on an over-identifying set.

(ii) Provide sufficient conditions for the consistency of the estimator using the over-identifying moment conditions in (i).

(iii) Now suppose that $y_{i,t-1}$ and $x_{it}$ are both *sequentially* exogenous. Propose a just-identified and an over-identified estimator of $(\rho_0, \beta_0)'$. Give sufficient conditions for the consistency of the over-identified estimator.

(iv) Suppose that in (iii), instead of sequential exogeneity, the lagged dependent variable only satisfied contemporaneous exogeneity, i.e. $E[\epsilon_{it}|y_{i,t-1}, a_i] = 0$. Would the just-identified estimator you proposed in (iii) be consistent? If not, derive its probability limit. Give any additional conditions required for the result.