Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

July, 2021

There are **THREE** questions. Choose and answer two of the three questions. Within each question, each part will receive equal weight in grading. You have 15 minutes to read the exam and then three hours to complete the exam.

I. Probability and Statistics

(a) Assume that all races considered in this question are choosing one winner out of two candidates, and that all voters are indifferent of the candidates.

(i) If there are only two voters, what is the probability of having a tied race (i.e., in this case, each candidate receives one vote)? What is the probability of having a tied race if there are four voters in total?

(ii) If the total number of voters is $2n$ in a specific race, what is the probability of having a tied race? If there are a total of K such races, each having 2 candidates, $2n$ voters, and all voters are indifferent of candidates, what is the probability of having *at least* one tied race out of all K such races? [Hint: if you are not sure of your answer in the first part of this question, denote your answer as A. Then continue with the second part so that your answer is a function of A.]

(b) Consider scalar random variables Y and D, where Y is continuous and D is binary taking value 0 or 1.

(i) The joint distribution of (Y, D) is $f(y, d) = y^d \cdot (1-y)^{1-d}$ if $y \in (0, 1), d \in \{0, 1\}$ and $f(y, d) = 0$ if otherwise. Calculate the marginal probability density function of Y. What is the distribution of Y called?

(ii) Following part (a), for fixed $y \in (0, 1)$, calculate the conditional probability $P[D = 1|Y = y]$ and $P[D = 0|Y = y]$. Are D and Y independent?

(iii) Now, disregard the specific joint distribution in part (a) and instead consider a general joint density $f(y, d)$ for $y \in \mathbb{R}$ and $d \in \{0, 1\}$. Let $P[D = 1] = p$. Show that $E[Y|D = 1] = E[YD]/p$ and $E[Y|D = 1] - E[Y|D = 0] = E\left[\frac{Y(D-p)}{p(1-p)}\right]$.

(c) Consider scalar continuous random variables $X_1, X_2, ..., X_n$ that are independent and identically distributed on $\mathbb{R}$.

(i) If $X_i \sim N(\mu, 1)$, derive the MLE estimator of μ and call it $\hat{\mu}$. What is the limiting distribution of $\sqrt{n}\,(\hat{\mu} - \mu)$? [Hint: the probability density function of $N(\mu, \sigma^2)$ is $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, for all $x \in \mathbb{R}$].

(ii) Following part (a), for fixed $x \in \mathbb{R}$, what is the distribution of $1(X_i \leqslant x)$ called? What is the distribution of $\sum_{i=1}^{n} 1(X_i \leqslant x)$ called? What is the mean and the variance of $\frac{1}{n} \sum_{i=1}^{n} 1(X_i \leqslant x)$ for fixed n? [Hint: If necessary, use $\Phi(.)$ to denote the cdf of $N(0, 1)$ and $\phi(.)$ to denote the pdf of $N(0, 1)$.]

(iii) Now, disregard the specific distribution in part (a). Let $F(.)$ be the marginal cdf of X_i for $i = 1, 2, ..., n$ and $F_n(.)$ be the empirical cdf such that $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1(X_i \leqslant x)$ for any $x \in \mathbb{R}$. What is the mean and the variance of $F_n(x)$ for fixed x and fixed n? What is the limiting distribution of $\sqrt{n}(F_n(x) - F(x))$ for fixed x as $n \to \infty$?

II. <u>**Linear Regression**</u>

Consider the model $y_i = \beta x_i + e_i$, where $x_i > 0$ is scalar, $E[e_i] = 0$, and $E[x_i e_i] = 0$. You have an *iid* random sample of size n. Consider the following estimators:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$\bar{\beta} = \frac{\sum_{i=1}^{n} x_i^2 y_i}{\sum_{i=1}^{n} x_i^3}$$

(a) Is $\hat{\beta}$ consistent for β? If so, prove it. If not, state additional conditions you require for consistency and prove consistency under those conditions.

(b) Is $\bar{\beta}$ unbiased for β? If so, prove it. If not, either (i) state additional conditions you require for unbiasedness and prove unbiasedness under those conditions, or (ii) explain why no such conditions exist.

(c) Is $\bar{\beta}$ consistent for β? If so, prove it. If not, state additional conditions you require for consistency and prove consistency under those conditions.

(d) Following on from (c), find the asymptotic distribution of $\sqrt{n}(\bar{\beta} - \beta)$ as $n \to \infty$. You may carry over any assumptions you made in (c) to show consistency. State any additional assumptions you require.

(e) Propose a test statistic of the null hypothesis $H_0 : E(x_i^2 e_i) = 0$. Derive the asymptotic null distribution of your test statistic. State any additional assumptions you need.

(f) Assume $E(e_i^2|x_i) = \sigma^2$. Consider the estimator $\tilde{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$. Which of $\tilde{\beta}$ and $\hat{\beta}$ is a more efficient estimator for β (asymptotically)? Justify your answer with a proof.

(g) Assume $E(e_i^2|x_i) = \sigma^2 x_i^{-0.5}$, where σ^2 is an unknown constant. Describe how you would construct a bootstrap confidence interval for β. Be specific, including about which estimator you would use and why.

III. <u>Nonlinear Estimation and Panel Data Methods</u>

(a) **Estimation and Testing in Nonlinear Models**. Suppose that $y_i = g(x_i; \beta_0) + u_i$ for $i = 1, \ldots, n$, where $dim(x_i) = \ell$, $dim(\beta_0) = k$.
In this question, you can impose the i.i.d. condition across i and all asymptotic arguments pertain to $n \to \infty$.

(i) Propose a generalized method-of-moments estimator of β_0 and provide primitive conditions for the identification of β_0 using this approach. Explain why these conditions are sufficient. Furthermore, discuss whether the moment conditions you propose just- or over-identify β_0.

(ii) Without providing sufficient conditions, state the asymptotic distribution of the sampling error of the estimator you propose in (i).

(iii) Let β_0^j denote the j^{th} element of β_0. Propose the Wald statistic to test $H_0 : \beta_0^j = \beta_0^{j'}$ for $j \neq j'$, $j = 1, \ldots, k$ and $j' = 1, \ldots, k$ (i.e. all coefficients are equal). Make sure to define all objects that the Wald statistic consists of and state its asymptotic distribution under H_0.

(b) **Measurement Error in Static Panel Models**. Assume y_{it} and x_{it}^* are scalar. Let $y_{it} = \beta_0 x_{it}^* + a_i + u_{it}$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$. Instead of observing x_{it}^*, we observe a mismeasured version of it, $x_{it} = x_{it}^* + \epsilon_{it}$. We refer to ϵ_{it} as the measurement error. In addition, we assume $E[u_{it}|x_{i1}^*, \ldots, x_{iT}^*, a_i] = 0$.
All asymptotics in this question pertain to $n \to \infty$, while holding T fixed. You can assume the i.i.d. assumption across i. However, you cannot impose the i.i.d. assumption across t unless the statement of the question allows you to.

(i) For a random variable w_{it}, let $\Delta w_{it} = w_{it} - w_{i,t-1}$. Derive the probability limit of the estimator of the slope coefficient of a regression of Δy_{it} on Δx_{it}. State all sufficient conditions required for the result invoking appropriate laws of large numbers and/or central limit theorems. However, do not impose any additional conditions on ϵ_{it} and a_i.

(ii) Now suppose there is a scalar instrument, z_{it}, which satisfies $E[u_{it}|z_{i1}, \ldots, z_{iT}, a_i] = 0$. Propose a generalized method-of-moments approach to estimate β_0. To do so, the next steps might be helpful.

1. Derive moment conditions that characterize β_0 implied by the strict exogeneity condition on z_{it}.
2. Can you estimate the moment conditions you derived? In other words, are they solely functions of observable variables? If yes, propose a generalized method of moments approach to estimate β_0. If not, impose additional assumptions such that the moment conditions are estimable. Discuss the intuition behind those additional restrictions and propose a generalized method of moments estimator of β_0.

ARE/ECN 240B Reference Sheet

Notation. θ_0, Θ, y_i, x_i, $s(y_i, x_i; \theta)$ and $H(y_i, x_i; \theta)$ pertain to the objects defined in the 240B lecture notes.

Assumption ULLN 1 $\sup_{\theta \in \Theta} |\sum_{i=1}^n f(y_i, x_i; \theta)/n - E[f(y_i, x_i; \theta)]| \xrightarrow{p} 0$, if the following conditions hold,

(i) *(i.i.d.)* $\{y_i, x_i\}_{i=1}^n$ is an i.i.d. sequence of random variables;

(ii) *(Compactness)* Θ is compact;

(iii) *(Continuity)* $f(y_i, x_i; \theta)$ is continuous in θ for all $(y_i, x_i')'$;

(iv) *(Measurability)* $f(y_i, x_i; \theta)$ is measurable in $(y_i, x_i')'$ for all $\theta \in \Theta$;

(v) *(Dominance)* There exists a dominating function $d(y_i, x_i)$ such that $|f(y_i, x_i; \theta)| \leqslant d(y_i, x_i)$ for all $\theta \in \Theta$ and $E[d(y_i, x_i)] < \infty$.

Assumption ULLN 2 $\sup_{\theta \in \Theta} |\sum_{i=1}^n f(y_i, x_i; \theta)/n - E[f(y_i, x_i; \theta)]| \xrightarrow{p} 0$, if the following conditions hold,

(i) (*Law of Large Numbers*) $\{y_i, x_i\}$ is i.i.d., and $E[f(y_i, x_i; \theta)] < \infty$ for all $\theta \in \Theta$, which implies $\sum_{i=1}^n f(y_i, x_i; \theta)/n \xrightarrow{p} E[f(y_i, x_i; \theta)]$.

(ii) (*Compactness of* Θ) Θ is in a compact subset of $\mathbb{R}^k$.

(iii) *(Measurability in $(y_i, x_i')'$)* $f(y_i, x_i; \theta)$ is measurable in $(y_i, x_i')'$ for all $\theta \in \Theta$.

(iv) (*Lipschitz Continuity*) For all $\theta, \theta' \in \Theta$, there exists $g(y_i, x_i)$, such that $|f(y_i, x_i; \theta) - f(y_i, x_i; \theta')| \leqslant g(y_i, x_i)\|\theta - \theta'\|$, for some norm $\|.\|$, and $E[g(y_i, x_i)] < \infty$.

Formula for the score statistic

$$S \equiv \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n s_i(\hat{\theta}_R)\right)' A_{nR}^{-1} C_{nR}' \left\{\widehat{Avar}\left(C_{nR} A_{nR}^{-1} \sum_{i=1}^n s_i(\hat{\theta}_R)/\sqrt{n}\right)\right\}^{-1} C_{nR} A_{nR}^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n s_i(\hat{\theta}_R)\right)$$