Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

July 6, 2017

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.

I. Applied Probability

1. State a Law of Large Numbers (LLN). Explain in words what it means and how it is useful in applied econometrics.
2. State a Central Limit Theorem (CLT). Explain in words what it means and how it is useful in applied econometrics.
3. The United States Government finances 50% of all child births through the Medicaid program. Most non-Medicaid child births are financed by private medical insurance. Five out of every six infants who die before the age of one have Medicaid-financed births. The overall infant mortality rate in the United States is 6. (The infant mortality rate is the average number of deaths of infants under one year old per 1,000 live births.)
   (a) What is the probability that a randomly chosen infant's live birth was financed by Medicaid and this infant will die before the age of one?
   (b) What is the infant mortality rate for Medicaid-financed births?
   (c) Are the events "birth financed by Medicaid" and "die before the age of one" independent? Explain.
   (d) Based only on the information given, can you infer whether Medicaid causes higher infant mortality? Explain your reasoning.
   (e) Suppose that you have data on 900 live births. In your sample, 2% of the infants died before reaching the age of one. Form a 95% asymptotic confidence interval for the infant mortality rate. If you invoke the LLN or CLT to obtain the confidence interval, state precisely why you do so.
   (f) Based on your answer in (e), do you think your data are a true random sample of the population of child births? Why or why not?

II. **Linear Regression** Consider the regression model: $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, $E[\varepsilon_i | X_{1i}, X_{2i}] = 0$.

(a) Show that ordinary least squares (OLS) provides an unbiased estimate of $\beta = [\beta_0\ \beta_1\ \beta_2]'$. State any assumptions you make.

(b) Show that OLS provides a consistent estimate of $\beta = [\beta_0\ \beta_1\ \beta_2]'$. State any assumptions you make.

(c) Explain the difference between unbiasedness and consistency.

(d) Suppose you were to omit X_{2i} from the regression and estimate the coefficient on X_{1i} by OLS. Derive the probability limit of your estimator and state the conditions under which it is consistent for β_1.

(e) Which estimate of β_1 is more efficient, the one for a regression that includes X_{2i} or the one that excludes X_{2i}? Explain your reasoning. If your answer depends on any properties of the data or model, then state those properties. (No mathematics is necessary; an answer in words is fine.)

(f) Suppose that $E[\varepsilon_i^2 | X_{1i}, X_{2i}] = \sigma_i^2$, which varies across i. Derive the variance of the OLS estimator for the full model.

III. **Estimation with Censored Outcome Variables**. Suppose that $y_i^* = x_i'\beta + u_i$, but we observe $y_i = y_i^* 1\{y_i^* > 0\}$.

Note: See the reference sheet for relevant formulas for this problem.

(a) Assume that $u_i | x_i \sim N(0, 1)$. Propose an efficient estimator of β.

(b) Give sufficient conditions for consistency and asymptotic normality of the estimator you proposed in (a).

(c) Propose Wald and Lagrange Multiplier tests for the null hypothesis $H_0 : \beta = 0$.

(d) What is the expectation of y_i (not y_i^*) conditional on x_i?

(e) Consider the setting where x_i is not exogenous, specifically $E[u_i | x_i] \neq 0$. Propose a method to estimate β under this assumption that takes into account the censoring. Give sufficient conditions for consistency.

IV. **What Fixed Effects Can and Cannot Fix**

1. Fixed effects estimation was introduced to control for time-invariant unobservables in the linear model. Define the linear fixed effects (FE) estimator and give conditions for its consistency and explain in words to an empirical researcher the specific type of unobservable heterogeneity it can control for.

2. Consider a dataset that contains individuals' earnings and their years of schooling over a short period of time. Four different researchers are interested in the effect of schooling on earnings, and they each have different concerns:

   (a) Researcher A is concerned about some individuals responding differently to increases in their schooling than would other individuals;

   (b) Researcher B believes it is very important to control for lagged earnings, since earnings tend to be serially correlated;

   (c) Researcher C is worried that individuals with different unobservable ability may choose different years of schooling, e.g. higher-ability individuals are more likely to go to graduate school;

   (d) Researcher D believes that the variability in income shocks experienced by different individuals in the sample is different, e.g. lesser-educated individuals tend to experience greater variability in their income shocks on average than higher-educated individuals.

   You are expected to address each of the empirical researcher's concerns separately. For each one:

   (i) Write down the model that addresses the potential problem raised by the researcher.

   (ii) Derive the probability limit of the FE estimator if applied to data generated by the model in (i). Give sufficient conditions for the result.

   (iii) If FE is inconsistent, explain formally which assumption in (1) is violated and explain to the empirical researcher the meaning of the probability limit and the reasons behind the inconsistency of FE.

ARE 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 $E[m(y_i, x_i; \theta)] < \infty$ for all $\theta \in \Theta$, which implies $\sum_{i=1}^n m(y_i, x_i; \theta)/n \xrightarrow{p} E[f(y_i, x_i; \theta)]$.

(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

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

GMM Expression for the Sampling Error

$$
\sqrt{n}(\hat{\theta} - \theta_0) = -\left(\mathcal{H}_0' W \frac{1}{n}\sum_{i=1}^n \left.\frac{\partial h(y_i, w_i; \theta)}{\partial \theta}\right|_{\theta=\theta*}\right)^{-1} \mathcal{H}_0' W \frac{1}{\sqrt{n}}\sum_{i=1}^n h(y_i, w_i; \theta_0) + o_p(1)
$$

where

$$
\mathcal{H}_0 = E\left[\left.\frac{\partial h(y_i, w_i; \theta)}{\partial \theta'}\right|_{\theta=\theta_0}\right]
$$

Normal Distribution

Let $Z \sim N(0,1)$, then its density is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{1}{2}z^2\right\} \tag{1}$$

Its cdf is denoted by $\Phi(z)$.

Let $X \sim N(\mu, \sigma)$, then its density is given by

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

and its cdf $F(x) = \Phi((x-\mu)/\sigma)$.