Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

August 10, 2015

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. A man and a woman agree to meet at a certain location at 12:30pm. The man's arrival time is uniformly distributed between 12:00pm and 1:00pm, while the woman's arrival time is uniformly distributed between 12:15pm and 12:45pm. The two arrival times are independent. The person to arrive first will remain at the location for a maximum of ten minutes.

(a) Find the probability that the woman is the first to arrive.

(b) Find the probability that the man and woman meet successfully.

(c) Given that the woman arrives first, find the probability that the man and woman meet successfully.

(d) Given that the man and woman meet successfully, find the probability that the first person to arrive does not have to wait longer than five minutes for the second person to arrive.

(e) Suppose you have no idea how to solve these problems using calculus, but you know how to program in an econometrics program such as R or Stata. Explain (in words, not code) how you would write a program to estimate the probability in (d).

(f) The steps you would undertake in (e) represent the use of an estimator. What properties are you willing to claim for your estimator? Be specific.

II. **Extremum Estimation** Consider the general extremum estimation problem, where

$$\theta_0 \equiv \arg\max_{\theta \in \Theta} Q(\theta),$$

where $dim(\theta) = k$.

(a) Let $Q_n(\theta)$ be the sample analogue of $Q(\theta)$. Define $\hat{\theta}$, an estimator of θ_0, and give conditions for its consistency. Explain briefly how these conditions imply consistency of $\hat{\theta}$ for θ_0.

(b) Consider the M-estimation problem, where $Q(\theta) = E[m(y_i, x_i; \theta)]$, define $\hat{\theta}$ and give primitive conditions for the assumptions you provided for consistency in (a).

(c) Now derive an expression of the sampling error, $\sqrt{n}(\hat{\theta} - \theta_0)$, for the M-estimation problem and give conditions for its asymptotic normality. Explain briefly how the conditions you propose imply the asymptotic normality of the sampling error.

(d) Using the asymptotic distribution you provided in (c), give an estimator of the asymptotic variance, $Avar(\sqrt{n}(\hat{\theta} - \theta_0))$. Write down the Wald statistic to test the null hypothesis $H_0 : \theta_0 = 0$ and state its asymptotic null distribution.

(e) For the least squares estimation problem, where $y_i = x_i'\theta + u_i$, define θ_0 and $\hat{\theta}$. Check whether the conditions you provided for consistency in (b) apply in this problem. If not, give primitive conditions for this estimator.

III. **<u>Linear Panel Data Models</u>** For $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$, we observe a scalar y_{it} and a $k \times 1$ vector x_{it} that are related in the following way

$$y_{it} = x_{it}'\beta + a_i + u_{it}.$$

Let $X_i = (x_{i1}, ..., x_{iT})$ a $k \times T$ vector and $E[a_i|X_i] = 0$.

(a) For this part only, assume $E[u_{it}|X_i, a_i] = 0$. For each of the pooled OLS, random effects GLS, fixed-effects and first-difference estimators for β, define the estimator formally and state if it is consistent or not. Explain your reasoning briefly. (No need to provide a proof.)

(b) For this part only, assume x_{it} is scalar and let $u_{it} = w_{it}\gamma + \epsilon_{it}$, where w_{it} and γ are scalar. Give conditions that ensure that the pooled OLS estimator you gave in (a) is still consistent. Prove your result. Do these conditions imply that any of the other estimators in (a) is also consistent?

(c) Give conditions under which the pooled OLS estimator in (a) is not consistent but the fixed-effects estimator is. Prove your result. Do these conditions imply that any of the other estimators in (a) is consistent?

IV. Spending on medical care varies widely across counties in the United States, as does life expectancy. Suppose you are interested in estimating the causal effect of medical-care spending on life expectancy.

You begin by specifying the model:

$$LE_i = \beta_0 + \beta_1 M_i + \varepsilon_i \tag{1}$$

where LE_i denotes life expectancy in county i and M_i denotes spending per person on medical care in county i.

(a) Suppose some counties in your sample have healthier populations than other counties, which causes them to have higher life expectancies. Given this information, what will be the properties of the OLS estimate of β_1 in the above regression model?

(b) Continuing from (a), suppose you observe for each county a variable X that measures the health of the population. How would you incorporate X into your regression model? Justify your answer.

(c) Continuing from (a), suppose you observe an instrumental variable Z. What properties must the instrumental variable possess for it to be useful in producing a consistent estimate of β_1 with a small bias?

(d) Suppose you learn that your data do not come from the United States as we know it. Instead, they come from an imaginary country that assigns medical spending to counties randomly (i.e., half of counties are assigned high spending levels and the rest are assigned low spending). Given this information, how would you estimate β_1 ? Justify your answer. You may assume that you observe the variables Z and X from parts (b) and (c).

(e) Continuing from (d), suppose that in spite of the government assignment of spending levels, many counties spend a different amount on medical care than they are assigned. Explain how you would test whether these deviations cause bias in your estimate of β_1. Write down the test statistic you would use and state its asymptotic null distribution. You may assume that you observe the variables Z and X from parts (b) and (c).

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

Assumption (Uniform Law of Large Numbers $\{\sup_{\theta\in\Theta} |\sum_{i=1}^n f(y_i, w_i; \theta)/n - E[f(y_i, w_i; \theta)]| \xrightarrow{p} 0\}$**)**

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

(ii) *(Compactness)* Θ is compact;

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

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

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

Assumption (Consistency of Sample Average of Hessian for M-Estimators)

(i) Each element of $H(y_i, x_i; \theta)$ is bounded in absolute value by a function $b(y_i, x_i)$, where $E[b(y_i, x_i)] < \infty$;

(ii) $A_0 = -E[H(y_i, x_i; \theta_0)]$ is positive definite.

Assumption (Asymptotic Normality of Sample Average of Score for M-Estimators)

(i) $E[s(y_i, x_i; \theta_0)] = 0$;

(ii) each element in $s(y_i, x_i; \theta_0)$ has finite second moment.

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

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