Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

July 6, 2018

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 examand then four hours to complete the exam.

I. Probability and Statistics

(a) Consider (X, Y) with joint p.d.f. $f_{X,Y}(x, y) = \begin{cases} 3xy^2/16 & 0 \leqslant x \leqslant 2, \quad 0 \leqslant y \leqslant 2. \\ 0 & \text{otherwise.} \end{cases}$

- (i) Obtain $f_X(x)$, the marginal density of X.
- (ii) Obtain $f_Y(y)$, the marginal density of Y.
- (iii) Obtain $E[X]$ and $Var[X]$.
- (iv) Obtain $\Pr[X \leqslant 1]$.
- (v) Obtain $E[X|Y = 1]$.

(b) Suppose we have a random sample $x_1, ..., x_n$ of size n from a distribution with density $f(x; \theta) = \theta^3 \exp(-\theta^3 x)$, $x > 0$, $\theta > 0$. X has mean $1/\theta^3$ and variance $1/\theta^6$.

- (i) Obtain the first-order conditions for the MLE of θ. (Note: for the MLE of θ and not of θ^3).
- (ii) Is there an explicit solution for $\widehat{\theta}$? If so, give it.
- (iii) Using standard results for the MLE, give the limit distibution of $\sqrt{n}(\widehat{\theta} - \theta)$.
- (iv) Suppose $\widehat{\theta} = 2.7$ and $n = 100$. Do you reject $H_0 : \theta = 3$ against $H_a : \theta \neq 3$ at level 0.05?

(c)
- (i) Suppose the continuous random variable X on $(-\infty, \infty)$ has moment generating function $M_X(t) = e^{2t+32t^2}$. Obtain $E[X]$.
- (ii) State a law of large numbers and conditions on X_i, $i = 1, ...n$, that guarantee that a law of large numbers is satisfied.
- (iii) Suppose that on average there is a 3% chance of a voter fraudulently voting (e.g. through voting twice or voting when not a citizen) in the 2016 U.S. Presidential Election (if all fraudulent votes were for Hillary Clinton this would reduce her vote by 3.75 million and give President Trump a victory of 0.75 million votes). The Voter Integrity Commission is looking into fraudulent voting. Suppose they develop a test that correctly detects 95% of the cases of fraudulent voting, but also incorrectly detects fraudulent voting in 10% of the cases when it is **not** there.
  What is the probability of fraudulent voting if the test detects fraudulent voting?

II. **Linear Regression**

Consider the model

$$y_i = x_i'\beta + e_i$$
$$z_i = y_i + u_i,$$

where x_i is $k \times 1$, $E(x_i e_i) = 0$, $E(y_i u_i) = 0$, and $E(x_i u_i) = 0$. You have a sample of size n. You do not observe y_i, so you estimate β by regressing z_i on x_i. Call your estimator $\tilde{\beta}$.

Credit will be given for answers that avoid imposing unnecessarily strong assumptions.

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

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

(c) Following on from (b), find the asymptotic distribution of $\sqrt{n}(\tilde{\beta} - \beta)$ as $n \to \infty$. State any additional assumptions you need.

(d) Propose a test statistic of the null hypothesis $H_0 : \beta = \beta_0$. Derive the asymptotic null distribution of your test statistic. State any additional assumptions you need.

(e) Suppose $E(x_i u_i) = \eta \neq 0$ and that you observe a $k \times 1$ variable w_i such that $E(x_i w_i') = \Pi$, where Π is full rank. Otherwise, the model is as above. Propose a consistent estimator for β and prove that it is consistent. State any additional assumptions you need.

(f) Following on from part (e), propose a test statistic of the null hypothesis $H_0 : \eta = 0$. State the asymptotic null distribution of the test statistic you propose and any assumptions required for this asymptotic distribution to be valid.

III. **Heterogeneity in Linear Models.**

Parts (a) and (b) are independent.

(a) Consider the model $y_i = x_i'(\beta + u_i)$, where x_i is $k \times 1$ and the $k \times 1$ vector $u_i \sim iid(0, \Omega)$. Given a sample of size n, consider the OLS estimator $\hat{\beta} = \left(\sum_{i=1}^n x_i x_i'\right)^{-1}\left(\sum_{i=1}^n x_i y_i\right)$. The vector x_i includes a constant.

  (i) Is $\hat{\beta}$ consistent for β? If so, prove it. If not, state additional conditions that are sufficient for consistency and prove consistency under those conditions.

  (ii) Following on from (a.i), find the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta)$ as $n \to \infty$. State any additional assumptions you use.

  (iii) Propose a test of the null hypothesis that the coefficients on x_i are constant across i. Derive the asymptotic null distribution of your test statistic. State any additional assumptions you need.

(b) Suppose you have a random sample of individuals $i = 1, \ldots, n_g$ in clusters $g = 1, \ldots, G$, and consider the model $y_{gi} = x_{gi}'\beta + a_g + u_{gi}$.

Note: You cannot assume random effects in any part of this problem.

  (i) Propose an estimator that is consistent for β. Give sufficient conditions for its consistency as $n_g \to \infty$ for $g = 1, \ldots, G < \infty$. Can you estimate a_g consistently under these asymptotics?

  (ii) Now suppose that the slope coefficient is also heterogeneous across clusters, specifically $y_{gi} = x_{gi}'\delta_g + \alpha_g + \epsilon_{gi}$. What is the probability limit of the estimator in (b.i) as $n_g \to \infty$ for $g = 1, \ldots, G < \infty$? Derive this probability limit formally and state any additional conditions you require for the result.

  (iii) Define the population average of δ_g. Is the probability limit of the estimator in (b.ii) equal to the population average of δ_g? If not, propose an estimator that would be consistent for the population average as $n_g \to \infty$ for $g = 1, \ldots, G < \infty$. Can you estimate the cluster-specific slope coefficients δ_g consistently for all $g = 1, \ldots, G$ under these asymptotics?

IV. **Estimation of Nonlinear Models with Panel Data**.

Panel data are used extensively to estimate weather impacts on outcomes, such as mortality rates, labor productivity, agricultural yield, etc. Let y denote an outcome and w denote a $k \times 1$ vector of weather variables, e.g. temperature and precipitation. For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, the following model describes the relationship between y_{it}, w_{it} and the unobservables,

$$y_{it} = h(w_{it}; \gamma_0) + a_i + u_{it}.$$

where $dim(\gamma) = \ell$ and $h(w; \gamma)$ is known given w and γ. Assume that weather (w_{it}) is strictly exogenous.

Note: You cannot assume random effects or impose any restrictions on time series dependence or heteroskedasticity in any part of this problem. However, you can assume the i.i.d. assumption across cross-sectional units $i = 1, \ldots, n$.

(a) Propose a consistent estimator of γ as $n \to \infty$ assuming T is fixed. Provide conditions for its consistency. Give primitive conditions wherever possible.

(b) Consider the following special cases of the above model:

(i) $h(w_{it}; \gamma) = w_{it}'\gamma$;

(ii) $h(w_{it}; \gamma) = t(w_{it}'\gamma)$, where $t(.)$ is a transformation function.

For each of the above models, comment on which assumptions in (a) would be immediately fulfilled and give primitive conditions whenever required.

(c) Now suppose that the researcher would like to understand the dynamic relationship between y and w and hence adds the lag of y to the model, specifically,

$$y_{it} = g(x_{it}; \beta_0) + \alpha_i + \epsilon_{it}$$

where $x_{it} = (y_{i,t-1}, w_{it}')'$ is a $(k+1)$ vector, $dim(\beta) = m$ and $g(x; \beta)$ is known given x and β. Suppose you apply the same estimation procedure you proposed in (a) to estimate β_0. Is the population analogue of this estimator equal to β_0, the true parameter value? Provide a formal answer.

(d) Propose an efficient generalized-method-of-moments (GMM) estimator of β as $n \to \infty$ assuming T is fixed. Provide conditions for its consistency for β_0. Give primitive conditions wherever possible. Be sure to specify the number of time periods required for the model to be just-identified and/or over-identified.

(e) Provide conditions for the asymptotic normality of the estimator you proposed in (d) assuming it is over-identified. Give primitive conditions wherever possible.

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

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