Field Examination: Econometrics

Department of Economics
University of California, Berkeley

August 2016

Instructions: You have 180 minutes to answer THREE out of the following four questions. Please make your answers elegant, that is, clear, concise, and, above all, correct. Good luck!

1. Suppose $\Theta \subseteq \mathbb{R}^m$ is open, $(\mathcal{X}, \mathcal{X}, P_X)$ is a probability space, and $\psi : \mathcal{X} \times \Theta \to \mathbb{R}^m$ is some function. Assume $X_i \sim IID - P_X$ for $i = 1, ..., n$ and suppose $(\hat{\theta}_n)_n$ is such that

$$n^{-1/2} \sum_{i=1}^n \psi(X_i, \hat{\theta}_n) = o_{P_X}(1).$$

Moreover, suppose

I. There exists a $\theta_0 \in \Theta$ such that $\lambda(\theta_0) = 0$, where $\lambda(\theta) \equiv E[\psi(X, \theta)]$.

II. There exists a $d_0 > 0$ such that $\sup_{|\tau - \theta_0| \leq d_0} Z_n(\tau, \theta_0) = o_{P_X}(1)$, where

$$Z_n(\tau, \theta) \equiv \frac{\{\psi(X_i, \tau) - \psi(X_i, \theta) - \{\lambda(\tau) - \lambda(\theta)\}\}}{\sqrt{n} + n|\lambda(\tau)|}.$$ 

III. $E[|\psi(X, \theta_0)|^2]$ is non-zero and finite.

(a) Suppose $P_X(|\hat{\theta}_n - \theta_0| \leq d_0) \to 1$. Show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i, \theta_0) + \sqrt{n}\lambda(\hat{\theta}_n) = o_{P_X}(1).$$

(b) Suppose $\theta \mapsto \psi(x, \theta)$ is not differentiable, but $\theta \mapsto \lambda(\theta)$ is. Could you still show that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically Gaussian? Be precise about the assumptions you need over $\lambda$ and its derivatives.

(c) Suppose $\psi(x, \theta) \equiv d \log f(x, \theta)/d\theta$, where $f(\cdot, \theta)$ is the pdf (with respect to the Lebesgue measure) indexed by $\theta$. Show that in this case the asymptotic variance coincides with the Fisher information matrix

$$\int \psi(x, \theta)\psi(x, \theta)' f(x, \theta) dx.$$ 

(d) Suppose the following hold:

1. There exists a $a > 0$ such that $|\lambda(\theta)| \geq a|\theta - \theta_0|$ for $|\theta - \theta_0| \leq d_0$.

2. There exists a $b > 0$ such that $E[\sup_{|\tau - \theta| \leq b} |\psi(x, \tau) - \psi(x, \theta)|] \leq bd$.

3. There exists a $c > 0$ such that $E[\sup_{|\tau - \theta| \leq b} |\psi(x, \tau) - \psi(x, \theta)|^2] \leq cd$ for $|\theta - \theta_0| + d \leq d_0$.

Show that these assumptions and assumptions I and II imply assumption III.

[This part is hard. You might be able to show the results with a different set of assumptions; these assumptions are only intended to be sufficient.]
2. [65 points] In each of periods $t = 0, \ldots, T$ an agent makes the binary choice $Y_t \in \{0, 1\}$. The econometrician observes the $T + 1$ choice sequence, she does not observe the vector of unobserved agent attributes $A$. Refer to $A$ as an agent’s (unobserved) type. Conditional on $A$ choice follows the stationary first order Markov chain:

$$
\Pr(Y_t = y | Y_{t-1}^t, A) = \Pr(Y_t = y | Y_{t-1}, A),
$$

where $Y_{t-1}^t = (Y_{t-1}, Y_{t-2}, \ldots, Y_0)'$ is the $t \times 1$ vector of past choices.

(a) [5 points] Let

$$
\pi_0(A) = \Pr(Y_t = 1 | Y_{t-1} = 0, A) \\
\pi_1(A) = \Pr(Y_t = 1 | Y_{t-1} = 1, A)
$$

denote the transition probabilities as a function of agent type. Let $p(A)$ denote the steady-state probability of being in state $Y_t = 1$. Solve for $p(A)$.

(b) [10 points] Let $Y_t = 1$ denote employment in period $t$ and $Y_t = 0$ non-employment. Interpret the estimand

$$
\Lambda(y_0, s) = \mathbb{E}_A[\Pr(Y_s = Y_{s-1} = \cdots = Y_1 = 1 | Y_0 = y_0, A)]
$$

and explain why, in general, it would not coincide with

$$
\Pr(Y_s = Y_{s-1} = \cdots = Y_1 = 1 | Y_0 = y_0).
$$

Present, and interpret, a sufficient condition for (1) and (2) to equal one another. Is this condition plausible when $Y_t$ measures employment? [4 to 6 sentences]

(c) [5 points] Assume that the unobserved attribute vector may take one of $K$ configurations:

$$
A \in \mathcal{A} = \{a_1, \ldots, a_K\}.
$$

Let $\rho = (\rho_1, \ldots, \rho_K)'$ denote the population frequency of each type of agent. Let

$$
\Pr(Y_0 = 1 | A = a_k) = \gamma_k
$$

for $a_k \in \mathcal{A}$ and $k = 1, \ldots, K$ parameterize the initial condition of the process for each type of agent. Similarly let $\pi_0(a_k) = \pi_{0,k}$ for $k = 1, \ldots, K$ be the probability of choice $Y_t = 1$ given that $Y_{t-1} = 0$ for each type of agent. Define $\pi_{1,k}$ similarly. Note that $\sum_{k=1}^K \rho_k = 1$ for $k = 1, \ldots, K$. Explain why $2^{T+1} \geq 4K - 1$ is a necessary condition for identification. [2 to 3 sentences]

(d) [5 points] Write $\Lambda(y_0, s)$ in terms of the parameters introduced in part (c) above.
(e) [5 Points] The econometrician observes choice in periods $t = 0, \ldots, T$ for each of $i = 1, \ldots, N$ randomly sampled agents. Let $\theta = (\pi', \gamma', \rho')'$ be the full “common” parameter. Assume that $A_i$ is observed. Write down the $i^{th}$ agent’s contribution to the complete data likelihood.

(f) [10 Points] Write down agent $i$’s contribution to the integrated likelihood, which marginalizes over the distribution of $A_i$. Say $\theta$ was known. What is the posterior probability that agent $i$ is of type $k$ after observing her choice sequence, that is:

$$\tilde{p}_{ki}(\theta) = \Pr(A = a_k| Y^T_0 = y_0^T; \theta).$$

HINT: Use Bayes’ Law and the $i^{th}$ agent’s contribution to the complete data likelihood.

(g) [15 Points] Describe, in detail, how the EM algorithm can be used to maximize the integrated likelihood as a function of $\theta$.

(h) [10 Points] Describe a joint fixed effects maximum-likelihood estimator for $\theta$ and the vector of incidental parameters $\{A\}_{i=1}^N$. Is the resulting estimate of $\theta$ consistent as $N \to \infty$? Will it be consistent under sequences where both $N$ and $T$ grow large? How fast does $T$ need to grow relative to $N$? A narrative answer is okay, just try to be clear about the main issues involved, and assumptions needed, for positive or negative answers to the questions. [5 to 15 sentences]

3. Suppose $\{y_t : 1 \leq t \leq T\}$ is an observed time series generated by the model

$$y_t = \delta t + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T,$$

where $u_0 = 0$ and $\varepsilon_t \sim i.i.d. \mathcal{N}(0,1)$, while $\delta$ is a parameter of interest and $\rho \in (-1,1)$ is a (possibly) unknown nuisance parameter.

(a) Find the log likelihood function $L(\delta, \rho)$ and, for $r \in (-1,1)$, derive $\hat{\delta}(r) = \arg \max_{\delta} L(\delta, r)$, the maximum likelihood estimator of $\delta$ when $\rho$ is assumed to equal $r$.

Suppose $\rho = 1$.

(b) Find the limiting distribution (after appropriate centering and rescaling) of $\hat{\delta}(1)$, the “oracle” estimator of $\delta$.

Suppose also that $\hat{\rho} - 1 = O_p(1/T)$.

(c) Is $\hat{\delta}(\hat{\rho})$ a consistent estimator of $\delta$?

(d) Is $\hat{\delta}(\hat{\rho})$ asymptotically equivalent to $\hat{\delta}(1)$?
4. Suppose \( \{(y_t, x_t) : 1 \leq t \leq T\} \) is an observed time series generated by the cointegrated system

\[
 y_t = \theta_0 x_t + u_t,
\]

where

\[
\begin{pmatrix}
 u_t \\
 \Delta x_t
\end{pmatrix} \sim i.i.d. \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)
\]

with initial condition \( x_0 = 0 \).

It can be shown that

\[
\frac{1}{T} \sum_{t=1}^{T} x_t u_t \xrightarrow{d} \begin{pmatrix} B_x(r) \\ \int_0^1 B_x(r) dB_y(r) \end{pmatrix},
\]

where \( B_x \) and \( B_y \) are independent Wiener processes.

Let \( z_t = (y_t, x_t)' \) and define the function

\[
h_T(z_t, \theta) = \left( \frac{T^{-1/2} x_t \sqrt{T}}{T^{-3/2} x_t^3} \right) (y_t - \theta x_t).
\]

(a) Show that \( \Theta_T = \{\theta_0\} \), where \( \Theta_T = \{\theta : \sum_{t=1}^{T} E[h(z_t, \theta)] = 0\} \).

Let

\[
\hat{\theta}_W = \arg \min_{\theta} g_T(\theta)' W g_T(\theta), \quad g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} h_T(z_t, \theta),
\]

where \( W \) is a symmetric, positive definite \( 2 \times 2 \) matrix.

(b) It can be shown that

\[
T(\hat{\theta}_W - \theta_0) \xrightarrow{d} \int_0^1 B_W(r) dB_y(r),
\]

where \( B_W \) is some functional of \( B_x \) and \( W \). Verify this claim and express \( B_W \) in terms of \( B_x \) and \( W \).

Let \( \omega^2_W = \int_0^1 B_W(r)^2 dr \).

(c) Find \( W^* \), a value of \( W \) for which \( \omega^2_W \) is minimal, and express \( \omega^2_W \) in terms of \( B_x \).

(d) Propose a feasible estimator \( \hat{\theta} \) satisfying

\[
T(\hat{\theta} - \theta_0) \xrightarrow{d} \int_0^1 B_{W^*}(r) dB_y(r).
\]