BOSTON COLLEGE
Department of Economics

Econometrics Field Comprehensive Exam

May 26, 2017

ALL MUST COMPLETE PART A

PART A – ECON7772 –PROF. LEWBEL

AND THEN CHOOSE 2 PARTS OF THIS EXAM:

PART B – ECON8821 –PROF. XIAO

PART C – ECON8822 –PROF. HODERLEIN

PART D – ECON8823 –PROF. BAUM

PART E – ECON8825 –PROF. KHAN

Please use a separate bluebook for each part.

Please write your alias, part and question number(s) on each bluebook.

Please read the entire exam before writing anything.

Field Examination in Econometrics
May 2017 PART A

1. Suppose we have n iid observations of a random variable X_i. Let $\mu = E(X_i)$ and $\sigma^2 = var(X_i)$, which are both finite. Suppose that, for some constant c (which could depend on n), we have the estimator for μ given by

$$\widehat{\mu} = \left(\sum_{i=1}^{n} \frac{X_i + 1}{n} \right) c$$

a. What is the bias and variance of $\widehat{\mu}$?

b. What value of c would minimize the mean squared error of $\widehat{\mu}$?

c. Suppose c has the value given in part b. Is the resulting estimator $\widehat{\mu}$ consistent?

2. Consider the model $Y_i = a + bX_i + X_i^2 e_i$ (this is not a typo; X^2 is multiplying e). My data are a set of independently, identically distributed observations of the vector (Y_i, X_i). The unobserved errors e_i are mean zero, have finite variance, and are independent of X_j for all observations i and j. The variable X_i always lies between zero and one and has non-zero variance.

a. Discuss whether each of the following estimators is consistent for estimating a and b, and among the consistent estimators, discuss which are likely to have relatively higher or lower asymptotic variance (you do not need to give formal proofs here):

i. Ordinary least squares regression of Y on a constant and on X.

ii. Ordinary least squares regression of Y on a constant, and on X, and on X^2.

iii. Two stage least squares regression of Y on a constant and on X, using a constant and X^2 as instruments.

b. Describe an estimator that is consistent and has lower asymptotic variance than any of those listed in part a., and explain why.

3. Suppose we have n iid observations of a continuously distributed random variable X_i that can take on any real value. Let $f(x)$ denote the pdf of X at the value x. Assume $f(x)$ is finite. Consider the estimator

$$\widehat{f}_h(x) = \frac{1}{nh} \sum\nolimits_{j=1}^{n} I(x \leq X_i \leq x + h)$$

a. What does $E\left(\widehat{f}_h(x)\right)$ equal?

b. What does $E\left[\left(\widehat{f}_h(x)\right)^2\right]$ equal?

c. What assumption do we need to make about the limits of h and of the product nh as $n \to \infty$ to ensure that $f_h(x)$ converges in mean square to $f(x)$?

PART B

Time Series Econometrics, May, 2017

Question 1. (1) Is a linear combination of two covariance stationary processes still covariance stationary? Prove it if it is true. Give a counterexample if not true. (2) If $\{X_t\}$ is a covariance stationary process and $\{c_j\}$ is a sequence of real numbers that $\sum_j |c_j| < \infty$, prove that $Y_t = \sum_{j=-\infty}^{\infty} c_j X_{t-j}$ is also a covariance stationary process.

Question 2. Define the autocovariance function of a process $\{X_t\}$ as $\gamma_X(h) = \text{Cov}(X_t, X_{t+h})$. Derive the autocovariance and autocorrelation functions of an MA(∞) process:

$$X_t = \sum_{i=0}^{q} \theta_i \varepsilon_{t-i},\ (\theta_0 = 1),$$

where ε_t are iid $(0, \sigma^2)$.

Question 3. Consider OLS estimation of the following regression

$$y_t = \beta' z_t + u_t,$$

where $\{u_t\}$ is stationary ergodic with spectral density $f_{uu}(\cdot)$,

$$z_t = (1, x_t')',$$

and $\{x_t\}$ is stationary ergodic with $\text{E}(x_t x_t') = M$,

$$\widehat{\beta} = \left(\sum z_t z_t'\right)^{-1} \sum z_t y_t$$

what's the limiting distribution of $\widehat{\beta}$?

Question 4. An ADF regression (after appropriate model selection)

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + +\alpha_2 \Delta y_{t-1} + u_t,$$

is performed on Danish Bond rate time series, the results are given below

$$\Delta y_t = 0.0011 - 0.0122 y_{t-1} + 0.3916 \Delta y_{t-1},$$

and the t-ratio statistic of α_1 is -0.911, perform a unit root test based on this result.

PART C - 2 PAGES

EC 8822

Stefan Hoderlein

May, 26 2017

Instructions: this part is an one hour, closed book exam. Attempt to answer formally and avoid lengthy verbal parts. There are 40 credits in total, and the individual credits for the questions are displayed as guidance. Try to answer all questions. However, the best grade is usually awarded with less than full credit.

1. State the Kernel based estimator for the density f_X of a d-dim random vector X. Define all elements formally, including the Kernel ("a Kernel is a function K such that...."). Explain how the estimator differs from a Kernel based estimator, if X had only one dimension. (6 credits)

2. What is the bias-variance trade-off? Explain it by using the bias and variance formulas for the Kernel density estimator you have suggested in question 1, and focus in particular on the bandwidth h. Hint: you do not need derive bias and variance, just state the formulas. (4 credits)

3. based on your answers in questions 1 and 2, how would you estimate the mean regression of a random variable Y on a d-random vector X denoted $m(x)$? Hint: use the fact that in the population

$$m(x) = E[Y|X = x] = \int y f_{Y|X}(y; x) dy,$$

and derive a sample counterparts estimator. (4 credits)

4. Can you run a nonparametric regression if Y is binary? If yes, why, if not, why not? (2 credits)

5. What type of restrictive assumption necessary for (parametric) maximum likelihood based estimation of the binary choice model with d dimensional X is not required for nonparametric estimation of the binary choice

model? What is the drawback of a general kernel based estimator when compared to parametric maximum likelihood estimation (hint: focus on the rates of convergence of the two estimators)? (3 credits)

6. In which sense is a semiparametric binary choice model a good compromise between the two paradigms? (2 credits)

7. What are compliers? Why is it a disadvantage that LATE is only defined on the compliers? (4 credits)

The following two questions (i.e., question 8 and question 9) apply to the linear random coefficients model

$$Y = B_0 + X'B_1,$$

where Y is an observed random scalar, X is an observed random K-vector, $B = (B_0, B_1)'$ is an unobserved random $K+1$ vector, and X is fully independent of B. The goal is to obtain an estimator for the density f_B. To this end, define the conditional characteristic function of $Y|X = x$ as

$$\varphi_{Y|X}(t; x) = E\left[\exp(itY)|X = x\right],$$

where $t, x \in R^{K+1}$, and $i = \sqrt{-1}$.

8. Show that there is a one-to-one relation between $\varphi_{Y|X}(t; x)$ and $\varphi_B(s)$ for any value of $(t, x, s) \in R^{2K+2}$ (4 credits)

9. Why is it sufficient to show the identification of $\varphi_B(s)$, if the ultimate goal is to estimate $f_B(b)$ for any value of b. (2 credits)

10. What is a regression (random) tree? Explain its functioning at every single node. (3 credits)

11. What is pruning and why is it used? Describe an algorithm that estimates a tree. (3 credits)

12. In which sense can we think of a neural net as a semiparametric regression model (just show formula and explain briefly)? (3 credits)

PART D

PART III

ECON8823: Applied Econometrics Comprehensive, Spring 2017

Answer all questions. They are equally weighted.

1. Write an essay discussing the concept of fractional integration of a time series, how it may be detected, and how a time series exhibiting fractional integration may be modeled.

2. Write an essay discussing the appropriate techniques for modeling count data.

3. Write an essay discussing the dynamic panel data model, including the rationale for its use, the alternative estimators, and the diagnostics used to judge an estimated model's adequacy.

PART E

Spring 2017

May 26, 2017

Econometrics Field Exam Questions(Khan)

75 mins.

This exam is closed book.

1. Let X, Y be independent exponential random variables:

$$
\begin{aligned}
f(x|\lambda) &= \frac{1}{\lambda}\exp(-x/\lambda) \quad x > 0 \\
f(y|\mu) &= \frac{1}{\mu}\exp(-y/\mu) \quad y > 0
\end{aligned}
$$

We observe Z, W where

$$
\begin{aligned}
Z &= \min(X, Y) \\
W &= 1 \text{ if } Z = X \text{ and } W = 0 \text{ if } Z = Y
\end{aligned}
$$

Assume (Z_i, W_i) $i = 1, 2, ...n$ are n iid observations. Find the MLE of λ, μ, and establish its limiting distribution.

2. Consider the censored regression model in the generic form:

$$
\begin{aligned}
y^* &= x'\beta + \sigma\epsilon \\
E[\epsilon] &= 0 \\
y &= \max(0, y^*) \\
\epsilon &\perp x \quad P(\epsilon \le a) = F(a) \\
f(\epsilon) &= F'(\epsilon)
\end{aligned}
$$

(a) Evaluate δ, the conditional marginal effect:

$$\delta = \partial E[y|x]/\partial x$$

(b) Evaluate the Average Marginal Effect:

$$\bar{\delta} = E_x[\delta]$$

(c) Propose an estimator for $\bar{\delta}$ as a function of an estimator (parametric or semi-parametric, depending on your assumptions) for β_0. Establish the asymptotic properties of your estimator.