First Year Quantitative Comp Exam
Spring, 2011

Part I - 203A

Instruction: Answer every question.

Suppose that random vectors $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and (x_1, x_2, x_3) are distributed independently. The random vector $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ is normally distributed with mean (c_1, c_2, c_3) and variance

$$\begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix},$$

while the random vector (x_1, x_2) is normally distributed with mean (d_1, d_2, d_3) and variance

$$\begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}.$$

The values of σ_{ii} $(i = 1, 2, 3)$ and of ω_{ij} $(i, j = 1, 2)$ are unknown. Let

$$y_1 = \alpha + \beta\ x_1 + \gamma\ \varepsilon_1$$

where α, β, and $\gamma \neq 0$ are parameters of unknown value. Let

$$y_2 = s(x_2) + \varepsilon_2$$

where s is an unknown continuous function. Let

$$y_3 = \varepsilon_3$$

and let y be defined by

$$y = y_1 + y_2 + y_3$$

1. Derive expressions for the density of y_1 conditional on x_1 and for the (unconditional) density of y, both in terms of the unknown function and parameters.

2. Derive an expression for the moment generating function of y, in terms of unknown function and parameters.

3. Can you show that, under the above specifications, y_1 and y_2 are independently distributed? If your answer is YES, prove that they are independently distributed. If your answer is NO, provide a counter example.

4. Suppose that $s(x_2) = x_2$. If it is known that the value of y lies in the interval $(0, 1)$, derive an expression for the probability that y_1, y_2 and y_3, all lie in the interval $(0, 1/3)$ and derive an expression for the probability that at least one y_i lies in the interval $(0, 1/3)$.

5. Suppose again that $s(x_2) = x_2$. Derive an expression for the expected value of y given that $y_3 > 0$.

6. Determine what parameters (or combination of parameters) and functions, if any, can be identified when the function s is unknown and the only observable variables are y, x_1, and x_2. Answer the same question for the case when only y and x_2 are observable. Explain in detail.

7. Suppose now that y_1, y_2, y_3, x_1, and x_2 are observable, that $s(0) = 0$, and that the density of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ is an everywhere positive unknown function, $f_{\varepsilon_1,\varepsilon_2,\varepsilon_3}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. Determine what parameters, or combination of parameters, and what functions or features of functions are identified. Answer the same question when y_1, y_2, y_3, x_1, and x_2 are observable, $s(0) = 0$, and it is known that
$$f_{\varepsilon_1,\varepsilon_2,\varepsilon_3}(\varepsilon_1, \varepsilon_2, \varepsilon_3) = g_{\varepsilon_1}(\varepsilon_1)\, g_{\varepsilon_2}(\varepsilon_2)\, g_{\varepsilon_3}(\varepsilon_3)$$
for everywhere positive, unknown functions $g_{\varepsilon_1}(\cdot)$, $g_{\varepsilon_2}(\cdot)$, and $g_{\varepsilon_3}(\cdot)$.

Part II - 203B

INSTRUCTION: ANSWER EVERY QUESTION.

1 Question 1

You are given a linear regression model

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + \varepsilon_i$$

such that (i) $(x_{i1}, x_{i2}, x_{i3}, x_{i4}, \varepsilon_i)$ is iid; (ii) $(x_{i1}, x_{i2}, x_{i3}, x_{i4})$ is independent of ε_i; and (iii) $(x_{i1}, x_{i2}, x_{i3}, x_{i4})' \sim N(0, I_4)$ and $\varepsilon_i \sim N(0,1)$. You would like to test

$$H_0 : \beta_1\beta_2 - \beta_3\beta_4 = 0$$
$$H_1 : \beta_1\beta_2 - \beta_3\beta_4 \neq 0$$

Propose a test statistic, and critical value. Provide a rigorous justification. Clarify whether your justification is based on asymptotic or exact finite sample properties.

2 Question 2

You are given a linear model

$$y_i = x_i\beta + \varepsilon_i$$

such that x_i and ε_i are independent of each other, and $\varepsilon_i \sim N(0,1)$. You are interested in $\theta = \exp(\beta)$.

1. Conclude that

$$\Pr\left[\widehat{\beta} - \frac{1.96}{\sqrt{\sum_{i=1}^n x_i^2}} \leq \beta \leq \widehat{\beta} + \frac{1.96}{\sqrt{\sum_{i=1}^n x_i^2}}\right] = 95\% \qquad (*)$$

   Hint: The question is not looking for an approximate justification. You should establish exact equality. It can be done by first proving

$$\sqrt{\sum_{i=1}^n x_i^2}\,\left(\widehat{\beta} - \beta\right) \sim N(0,1).$$

2. Because $\exp(t)$ is monotonically increasing in t, say, we can convince ourselves that equation (*) implies

$$\Pr\left[\exp\left(\widehat{\beta} - \frac{1.96}{\sqrt{\sum_{i=1}^n x_i^2}}\right) \leq \exp(\beta) \leq \exp\left(\widehat{\beta} + \frac{1.96}{\sqrt{\sum_{i=1}^n x_i^2}}\right)\right] = 95\% \qquad (**)$$

3. Using

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \frac{1}{E\left[x_i^2\right]}\right)$$

and delta method, prove that

$$\sqrt{n}\left(\exp\left(\widehat{\beta}\right)-\exp\left(\beta\right)\right)\xrightarrow{d}N\left(0,\frac{\left(\exp\left(\beta\right)\right)^{2}}{E\left[x_i^2\right]}\right)$$

Also prove that

$$\frac{\sqrt{n}\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\left(\exp\left(\widehat{\beta}\right)-\exp\left(\beta\right)\right)}{\exp\left(\widehat{\beta}\right)}\xrightarrow{d}N\left(0,1\right)$$

based on which prove that

$$\Pr\left[\exp\left(\widehat{\beta}\right)-1.96\frac{\exp\left(\widehat{\beta}\right)}{\sqrt{\sum_{i=1}^{n}x_i^2}}\leq\exp\left(\beta\right)\leq\exp\left(\widehat{\beta}\right)+1.96\frac{\exp\left(\widehat{\beta}\right)}{\sqrt{\sum_{i=1}^{n}x_i^2}}\right]\rightarrow 95\% \qquad (***)$$

4. Equation (**) implies that a valid 95% confidence interval for θ is

$$\left(\exp\left(\widehat{\beta}-\frac{1.96}{\sqrt{\sum_{i=1}^{n}x_i^2}}\right),\ \exp\left(\widehat{\beta}+\frac{1.96}{\sqrt{\sum_{i=1}^{n}x_i^2}}\right)\right)$$

Approximate equality (***) implies that a valid 95% asymptotic confidence interval for θ is

$$\left(\exp\left(\widehat{\beta}\right)-1.96\frac{\exp\left(\widehat{\beta}\right)}{\sqrt{\sum_{i=1}^{n}x_i^2}},\ \exp\left(\widehat{\beta}\right)+1.96\frac{\exp\left(\widehat{\beta}\right)}{\sqrt{\sum_{i=1}^{n}x_i^2}}\right)$$

Comparison of these two confidence intervals suggests that we may have

$$\exp\left(\widehat{\beta}+\frac{1.96}{\sqrt{\sum_{i=1}^{n}x_i^2}}\right)\approx\exp\left(\widehat{\beta}\right)+1.96\frac{\exp\left(\widehat{\beta}\right)}{\sqrt{\sum_{i=1}^{n}x_i^2}}$$

$$\exp\left(\widehat{\beta}\right)-1.96\frac{\exp\left(\widehat{\beta}\right)}{\sqrt{\sum_{i=1}^{n}x_i^2}}\approx\exp\left(\widehat{\beta}\right)-1.96\frac{\exp\left(\widehat{\beta}\right)}{\sqrt{\sum_{i=1}^{n}x_i^2}}$$

Provide a justification of such conjecture. Hint: This is an open-ended question.

3 Question 3

Consider the following two-equation model

$$\begin{aligned} y_{i1} &= x_{i1}'\beta + y_{i2}\theta + \varepsilon_{i1} \\ y_{i2} &= x_{i2}'\beta + y_{i1}\theta + \varepsilon_{i2} \end{aligned}$$

This is a model of social interaction, where observations are drawn from pairs of individuals. The model is such that the outcome y_{i1} of the first agent depends not only on his own characteristic x_{i1} but also on the outcome y_{i2} of the second agent. (Your GPA not only depends on your effort and intelligence, but also on your roommate's GPA.) Using the intuition on identification of the linear simultaneous equations models, discuss how the degree of social interaction θ can be identified. Discuss your identifying assumptions in detail.

Part III - 203C

Instruction: Answer every question.

3.1 Question 1 (20 pts.)

You are stranded on a deserted island. Suppose in your spare time you want to run a linear regression using T observations and K variables. The T observations may represent the number of coconuts on a palm tree, the height of the tree, and other factors; being hungry, you think a regression will help you locate bountiful trees. You lack a computer to run the cross-sectional regression but find the Kalman Filter algorithm implemented on a device that takes all the usual inputs for the Kalman Filter.

1. How can you perform your regression using the Kalman Filter?
2. You notice that trees have significant variability with more mature trees (older, taller) bearing more coconuts. How would you modify your Kalman Filter to perform generalized least squares?

3.2 Question 2 (10 pts)

The Fibonacci sequence is 1, 1, 2, 3, 5, 8, 13, ... What is the ratio of adjacent terms in the limit? Prove your answer.

3.3 Question 3 (20 pts)

Consider a stationary process Y_t given by $Y_t = \alpha Y_{t-1} + \epsilon_t$. Suppose a moving average of n terms is formed $V_{t,n} = \frac{1}{n}\sum_{k=1}^{n} Y_{t-k}$.

1. What is $E(V_{t,n})$?
2. Show that
$$Var(V_{t,n}) = \frac{1}{n}\left[R_0 + 2(1-\frac{1}{n})R_1 + 2(1-\frac{2}{n})R_2 + ... + \frac{2}{n}R_{n-1}\right],$$
where R_j is autocovariance at lag j. Hint: Do this for $n = 1, 2, 3$ and find the pattern.
3. Contrast the solutions for $Var(V_{t,n})$ when α is near zero or α near one? How do the $Var(V_{t,n})$ differ? Explain.

3.4 Question 4 (10 pts)

Let X_t and Y_t be two random sequances. Show that $S_{X+Y}(w) = S_X(w) + S_Y(w) + S_{XY}(w) + S_{YX}(w)$ where $S_{XY}(w)$ is the cross-spectrum. Does this contradict the fact that the Fourier transform of a sum is the sum of Fourier transforms?

3.5 Question 5 (20 pts)

Consider a real data series $x_0, x_1, ..., x_{n-1}$ where n is even. How does the Fourier transform $d_x(f)$ relate to the Fourier transform $d_y(f)$ if $\{y_t\} = x_0, x_1, ..., x_{n-1}, x_0, x_1, ..., x_{n-1}$, i.e. a "doubled" dataset? Does this change if $\{y_t\} = x_0, x_0, x_1, x_1, ..., x_{n-1}, x_{n-1}$?