Quantitative Methods Qualifying Exam

AAEC, UGA

May 17, 2018

Instructions:

- Answer all five questions.
- Begin each of the five questions with a new sheet of paper.
- Label your work with your ID number, NOT your name.
- To the extent possible, make your work easy to follow.

Good luck!!!

1. You have a simple, linear regression model, $y = X\beta + e$. You "know" that the error term is distributed according to the following process:

$$e_t = \gamma + \rho e_{t-1} + \lambda x_{3t} + u_t$$

where $u_t \sim N(0, \sigma^2)$, x_3 is the third variable in the X matrix, $t = 1, 2, ..., 50$ denotes time-ordered observations, and the regressor matrix X is full-column rank with 5 columns.

Show detailed steps that should be followed to arrive at consistent estimates of β.

2. Consider the following instrumental variable equation setup:

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

$$x_i = z_i\delta + \nu_i$$

where we have the following assumptions:

- The vector $(z_i, \nu_i, \varepsilon_i)$ is i.i.d.
- $E[z_i, \nu_i] = 0$
- $E[z_i^2] = \sigma_z^2 \in (0, \infty)$.

Define the IV estimator as:

$$b_{IV} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i y_i$$

(a) Is b_{IV} consistent? Prove your conclusion.

(b) Is b_{IV} precisely estimated? Prove your conclusion.

3. For every $i = 1, ..., n$, assume that:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Using $n = 100$ observations of $\{y_i, x_i\}_{i=1}^{n}$, a researcher (Hairy) estimates the following coefficients:

$$(b_0, b_1), (\hat{\sigma_{b_0}}, \hat{\sigma_{b_1}}), R^2, s^2$$

Suppose that another researcher (Dawg), is replicating Hairy's analysis. Dawg attempts to run the same regression that Hairy ran, but accidentally enters each observation twice. The new dataset has 200 observations, with the second $n = 100$ observations an exact repeat of the first 100 observations. Suppose that Dawg realizes his mistake, but no longer has access to a statistical package to run the regression again. How can Dawg compare his results to the results obtained by Hairy? Obtain closed form solutions for the following estimates that Dawg obtained, based on the esimates that Hairy obtained:

(a) (b_0, b_1) (i.e., the coefficient estimates),

(b) s^2 (i.e., the regression standard error),

(c) $(\hat{\sigma_{b_0}}, \hat{\sigma_{b_1}})$ (i.e., the standard errors of the coefficient estimates),

(d) R^2.

Note whether the estimates are the same or different. If they are different, state how they are different.

4. Consider the following regression model: $y_i = \beta x_{1i} + \gamma x_{2i} + \varepsilon_i$, where y, x_1, x_2, and ε are random variables (scalars) with the following properties:

- $E[x_{1i}] = 0$ and $E[x_{2i}] = 0$
- $0 < E[x_{1i}^2] < 1$ and $0 < E[x_{2i}^2] < 1$
- $E[x_{1i}x_{2i}] = 0$
- $E[\varepsilon_i | x_{1i}; x_{2i}] = 0$
- $E[\varepsilon_i^2 | x_{1i}; x_{2i}] = \sigma^2$

Suppose you have drawn an i.i.d. sample of size n from the model above.

(a) Consider $b_{OLS,full}$, the OLS estimator of β obtained from the regression of y_i on x_{1i} and x_{2i}. Prove that $b_{OLS,full}$ is a consistent estimator for β.

(b) Now consider $b_{OLS,small}$, the OLS estimator of β obtained from the regression of y_i on just x_{1i}. Prove that $b_{OLS,small}$ is a consistent estimator for β.

(c) Find the correct expressions for the variance of both $b_{OLS,full}$ and $b_{OLS,small}$.

(d) Show that $Var(b_{OLS,full}) < Var(b_{OLS,small})$.

(e) How should your result from part (d) be interpreted?

5. For decades, it had been assumed that addictions or habits are myopic in the sense that the consumer does not recognize the impact of his or her current decision on future health and preferences. The theory of rational addiction by Gary Becker and Kevin Murphy breaks from this tradition and hypothesizes that consumers can be forward-looking (i.e., rational) even when consuming addictive or habit-forming goods such as cigarettes and alcohol. This theory, if supported by data, has important implications for public policies such as tobacco control and war on drugs.

   Using cigarettes as an example, the model of rational addiction suggests the following demand relationship:

$$y_t = \alpha + \beta_1 y_{t+1} + \beta_2 p_t + \beta_3 y_{t-1} \tag{1}$$

   where y_t, y_{t+1}, and y_{t-1} are the number of cigarettes smoked at time t, $t+1$, and $t-1$, respectively; p_t is cigarette price at t. The parameters β_1 and β_3 measure the degree of forward- looking behavior and the degree of habits, respectively. There is no residual in the theoretical relationship shown in the model above because it assumes perfect foresight such that the level of y_{t+1} is known with certainty at time t. In the United States, state and federal excise taxes account for 43% of cigarette retail price. In addition, the level of state cigarette tax varies substantially across states and tax changes are known months before they are effective.

   (a) When applying equation 1 to longitudinal data on smoking, a residual term must be added, and y_{t+1} is no longer known with certainty at t. How would you estimate the parameters of equation 1?

   (b) Next, assume you estimate β_1 to be near-zero and statistically insignificant. Derive the short-run and long-run price elasticity of cigarette smoking.

   (c) How is the consistency of the OLS estimator dependent on the time-series property of the residual? Explain using math and intuition.

Econometrics Qualifying Exam
May 18, 2017

Answer all questions. Show work fully and write neatly. Good luck.

1. Answer the following questions.

   a. What does it mean for an estimator to be BLUE?

   b. Under what data generating process is OLS BLUE?

   c. Generalized Least Squares (GLS) extends the OLS estimator to be BLUE in what setting?

   d. True, Partly True, or False: GLS is to OLS as the Generalized Method of Moments is to the Method of Moments. Explain.

2. You have a simple linear demand model for cigarette at the product level:

   $$lnQ_{ijt} = \alpha_i + \beta_i lnp_{ijt} + \varepsilon_{ijt},$$

   where Q_{ijt} and p_{ijt} are the quantity and per pack price of cigarette product i (e.g., full strength Marlboro in pack, light Marlboro in carton, etc.) in state j in year t, respectively. You have sales data on 121 products from 48 contiguous states for 10 years. You also have data on the attributes of each cigarette product (e.g., tar level, whether it is mentholated, premium vs. generic, etc.). Suppose you have applied the proper econometric technique to obtain consistent estimates of the coefficients. Let $\hat{\beta}$ be the 121×1 column vector of slope coefficient estimates whose ith element is $\hat{\beta}_i$; and Ω be the 121×121 matrix of variance-covariance matrix for $\hat{\beta}$.

   With the above empirical results in hand, you are asked to investigate whether the product-level price elasticities are statistically associated with product attributes. Describe how you would proceed with your analysis? Would OLS work best in this case? If not, what is a better alternative estimator and why?

3. Assume a linear regression model that accurately represents the true data generating process:

$$y_t = \beta_o + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \varepsilon_t \qquad \text{where } \varepsilon_t \sim N(0, \sigma^2).$$

You have some information, from economic theory, that guides you toward likely values of β_2 and β_3.

a) Show how to estimate the model if the values of β_2 and β_3 are restricted to b_2 and b_3.

b) Compare the restricted estimator to the OLS estimator and describe fully when one is preferable to the other and how you propose to measure "preferable."

4. Consider two non-nested models:

(M1) $y1 = X\beta + \varepsilon$ and (M2) $y2 = Z\theta + \omega$

a) How would you go about choosing between these two non-nested models if they had the same dependent variable ($y1 = y2$)?

b) What would you do if $y1 = \ln(y2)$?

5. Consider the model

$$\begin{aligned} y_i &= y_i^* \times \mathbb{I}\{y_i^* \geq 0\} \\ y_i^* &= \alpha + \beta x_i + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2) \end{aligned}$$

where you observe i.i.d. realizations of (y_i, x_i) and $\mathbb{I}$ is an indicator function that equals 1 when the term in the {}s is true. Assume that x_i has full support.

a. Which of (α, β, σ) are identified? Explain.

b. How would the answer to part a change if we only observed $(\mathbb{I}\{y_i^* \geq 0\}, x_i)$?

c. A researcher attempts to estimate (α, β) by running an OLS regression of y_i on x_i. Explain why the estimator is biased. What is the sign of the bias?

Econometrics Qualifying Exam Retake
July 20, 2016

Answer all questions. Show work fully and write neatly. Good luck.

1. For the model $y = X\beta + \varepsilon$,
   a. list all the assumptions that make it the classic linear model.
   b. of those assumptions, explain which one if violated causes the fewest problems and how would you address the violation of that assumption (provide full details).
   c. of those assumptions, explain which one if violated causes the most or hardest problems and how would you address the violation of that assumption (provide full details).

2. You want to estimate a demand model for a consumer good with several substitutes. You have data on the quantity purchased each week for three years for the good to be modeled, along with prices for that good and four substitutes. You also have consumer income data. Economic theory tells us that the sum of the own and cross price elasticities plus the income elasticity should be zero (i.e., if all prices and incomes double, demand is unchanged). Describe in detail how to specify a demand model and test this restriction implied by economic theory.

3. An experiment was performed in the Georgia State Prison in which inmates in one cell block were randomly assigned to a vocational program or not. In another cell block inmates were allowed to voluntarily enroll in the vocational program. The inmates were followed for three years after release and those inmates in the mandatory vocational program had similar recidivism rates as those in the general prison population whereas those in the voluntary vocational program had much lower rates of recidivism. Explain.

4. Consider the estimated regression model $y=X\hat{B}+\hat{u}$ where y is (nx1), X is (nxk), $\hat{B}$ is (kx1) and $\hat{u}$ is (nx1). Given that y_o is the (scalar) value to be taken by y given x_o (1xk), $\hat{y}_o = x_o\hat{B}$ is the predicted value of y given x_o, and $\hat{u}_o = y_o - \hat{y}_o$:
   a) Show that $E[\hat{u}_o]=0$,
   b) Derive $Var[\hat{u}_o]$,
   c) Identify any necessary assumptions and then show how the above can be used to construct a $(1-\alpha)$ confidence interval for y_o,
   d) Identify the value of x_o at which that confidence interval would be the narrowest.

5. Suppose that the regression model is: $y_i = \mu + u_i$, where y_i and u_i are random variables, μ is a constant parameter, $E[u_i|x_i]=0$, $cov[u_i, u_j|x_i, x_j]=0$ for $i\neq j$, and $var[u_i|xi] = \sigma^2 x_i^2$, $x_i > 0$.
   a) Given a sample of observations on Yi and Xi, what is the most efficient estimator of μ? What is its variance?
   b) What is the OLS estimator of μ? What is its variance?
   c) Prove that the estimator in part a) is at least as efficient as the estimator in part b).
   d) Discuss and compare the asymptotic properties of these two alternative estimators of μ.

Econometrics Qualifying Exam
May 22, 2015

Please answer all questions, show work fully and write neatly. Good luck.

1. Discuss the properties of the OLS estimator for the model parameters and standard errors in the presence of stochastic regressors (i.e., explanatory variables that are not fixed in repeated sampling). Assume the stochastic regressors are uncorrelated with the error term.

2. Consider the following estimator for a linear model $y = X\beta + u$:

   $$\hat{B} = (Z'X)^{-1}Z'y$$

   where Z is a conformable matrix of exogenous variables different from but highly correlated with those in X but uncorrelated with u.

   a. Compare the properties of this estimator versus OLS.

   b. Discuss under which conditions this estimator would be preferred to OLS.

   c. What would be the impact of the level of correlation between Z and X on the estimator properties (please explain your answer)?

3. Consider the time series linear regression model $y=X\beta + u$ where y is a Tx1 vector and X is a Txk matrix of conditioning variables. The t^{th} observation is given by $y_t = x_t\beta + u_t$, where x_t is a 1xk vector. Assume that $u_t = \rho u_{t-1} + \varepsilon_t$ and that $var(u_t)=\sigma^2_u$ for all t; $var(\varepsilon_t)=\sigma^2_\varepsilon$ for all t; and $E(u_{t-s}\varepsilon_t)=0$ for all $s\geq1$. Furthermore $|\rho|<1$.

   a. Write $V(u_t)$ in terms of σ^2_ε.

   b. Show the general form of E(uu').

   c. Given your result in (a) propose a GLS (generalized least squares) estimator for this model.

   d. Show that by quasi-differencing the data by ρ (where the quasi-differenced form of z_t is given by $z_t-\rho z_{t-1}$) that the autocorrelation problem is fixed.

4. A survey of 200 households each of which had **exactly** two children in an Indian state recorded the number of boy children. 40 households had no boy child, 100 households had one boy child, and 60 households had two boy children. Let the number of boys in each category be denoted by $n_0=40$, $n_1=100$, and $n_2=60$ and assume that the numbers of boys in a two-child family are binomially distributed. The binomial probability mass function has the general form

$$P(Y=y)=\frac{m!}{y!(m-y)!}\pi^y(1-\pi)^{m-y} \quad \text{where } y=0, 1, \ldots, m \text{ and } 0<\pi<1$$

Note that in the present context, π is the probability of a boy in any given trial (birth) and that for each household in the data set the number of trials is $m=2$.

a. Write the probability of observing exactly one boy child in a household.

b. The maximum likelihood estimator of π is

$$\pi^*=\frac{0.5n_1+n_2}{n_0+n_1+n_2} \quad \text{and the observed Hessian is}$$

$$H(\pi^*)=-(1-\pi^*)^{-2}(2n_0+n_1)-(\pi^*)^{-2}(n_1+2n_2)$$

Calculate and report the maximum likelihood estimator of π and its associated standard error.

c. Formally set up and provide a Wald test of the hypothesis that $\pi=0.5$ at the $\alpha=0.05$ level of significance.

d. Note that the likelihood function for the 200 household sample can be written

$$\ell=2n_0\ln(1-\pi)+[\ln 2+\ln\pi+\ln(1-\pi)]n_1+2n_2\ln\pi$$

Formally set up and provide a likelihood ratio test of the hypothesis that $\pi=0.5$ at the $\alpha=0.05$ level of significance.

5. You estimate a linear regression model of the form $y = X\beta + \varepsilon$ using nT observations of pooled cross-section time-series data detailing household purchases of food with n=1000 households observed over T=20 years. Explain,

a. How you would estimate the model, providing specific detail on the methods with special attention to your assumptions about error structure.

b. What diagnostic test you would want to perform on your estimated models, providing specific detail on how to carry out those tests.

Quantitative Methods Preliminary Examination
Department of Agricultural and Applied Economics
July 15, 2014

Answer all parts of all questions. Show intermediate steps where appropriate and make your work easy for the graders to follow. Relax and good luck.

1. An econometrician estimates the following equation for output:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 M_t + u_t$$

where u_t is a random shock and M_t is some policy variable, say money growth. The econometrician estimates that β_2 is significantly positive.

a. What does this say about policy effectiveness?

b. Could β_2 be estimated as positive even if the policy is ineffective? Use a combination of statistical theory, mathematics and verbal explanation to fully justify your answer.

2. Consider the linear regression model $y = X\beta + \varepsilon$ where X is a (t x k) matrix which is fixed in repeated samples and $\varepsilon \sim N(0, \sigma^2\Omega)$ with Ω known. Show that $\widehat{\beta}_{GLS}$ is the MLE for β and that $\widehat{\sigma^2} = (1/T)(y - X\widehat{\beta}_{GLS})\Omega^{-1}(y - X\widehat{\beta}_{GLS})$ is the MLE of σ^2.

3. At your next job, suppose you have estimated a model:

$$y = \beta_1 + \ln(x)\beta_2 + \ln(z)\beta_3 + \varepsilon.$$

Your boss disagrees and suggest running the model without logging the regressors. Describe a formal hypothesis tests of whether your functional form is appropriate, complete with null hypothesis, test statistics, steps involved in computing the test, etc.

Quantitative Methods Preliminary Examination
Department of Agricultural and Applied Economics
July 15, 2014

4. The following regression was estimated from 16 quarterly observations (t ratios in parentheses):

$$
\begin{array}{llllll}
Y_t = & 70.7 & -0.90X_t & +0.43S_{1t} & +6.55\,S_{2t} & -2.83\,S_{3t} \\
 & (3.7) & (0.27) & (3.37) & (3.40) & (3.37)
\end{array}
$$

$$R^2 = 0.68 \qquad n = 16$$

where $S_{it} = 1$ in the i^{th} quarter and is zero otherwise. Explain the implied pattern of seasonal variation and interpret the result in a paragraph.

5. Detail all the steps involved in testing the hypothesis below for the linear regression model $y = X\beta + e$, where $X = (50 \text{ x } 6)$ for two cases.

Ho: $x_3\beta_3 + x_4\beta_4 = 0$ Ha: $x_3\beta_3 + x_4\beta_4 \neq 0$

case 1: You can assume that the errors are iid normal with zero mean.
case 2: The errors are not normally distributed.

The University of Georgia

Department of Agricultural and Applied Economics

Ph.D. Qualifying Exam Questions for Econometrics (May 2014)

Answer all questions. Show enough work for readers to clearly follow your answers.

Good luck!

1. Consider the following linear regression model:

$$y = X\beta + u.$$

X is a $N \times k$ matrix which is fixed in repeated samples with $rank(X) = k$. u is a $N \times 1$ random vector with $u \sim N(0, \sigma^2 I_N)$. Denote $\theta = (\beta', \sigma^2)'$.

(a) Show that the log likelihood function is

$$LLF(\beta, \sigma^2) = -(N/2)\ln(2\pi) - (N/2)\ln(\sigma^2) - (1/2\sigma^2)(y - X\beta)'(y - X\beta)$$

(b) You would like to test the null hypothesis that $H_0 : \sigma^2 = 1$. Show that, in this case, the restricted maximum likelihood estimator of θ becomes $(\hat{\beta}', 1)'$ where $\hat{\beta} = (X'X)^{-1}X'y$.

(c) Show that the Lagrange Multiplier (LM) test statistic for the above null hypothesis is

$$LM = N(\hat{\sigma}^2 - 1)^2/2.$$

Note that the inverse of the information matrix is

$$\mathfrak{I}_\theta^{-1} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & 2\sigma^4/N \end{bmatrix}$$

2. Consider the following linear regression model:

$$y = X\beta + u$$

where X is a $N \times 5$ matrix. You would like to test two different hypotheses about the parameters:

$$H_0^A : \beta_2 = 0 \quad \text{and} \quad \beta_3 = 0$$
$$H_0^B : \beta_2 = \beta_3 \quad \text{and} \quad \beta_3 = 0$$

(a) Express both H_0^A and H_0^B in the form of $R\beta = r$, where R contains the coefficients in a linear restriction on the coefficient vector and r contains the restricted values.

(b) Suppose that R_A and r_A are the appropriate choices for the null H_0^A, and R_B and r_B are the appropriate choices for the null H_0^B. Show that $M(R_A\beta - r_A) = R_B\beta - r_B$ where M is a nonsingular matrix. (Hint: You need to create the M matrix with real numbers.)

(c) Show that the F-statistic is invariant to nonsingular transformation of $R\beta = r$. That is, the test statistic is the same whether the null hypothesis is written as H_0^A or H_0^B.

3. The employment in a county is measured by the variable y which can be explained quite well by the model

$$y = X\beta + \epsilon$$

where β is a (10×1) vector and the ten variables in X include a constant. The government changed business tax policy ten years ago in a manner that was hoped to have a positive effect on employment, although the impact might have occurred slowly over several years. Assuming you have time series data on (X, y) for a number of counties, how would you modify the model above to test a hypothesis about the impact of the tax policy. Be specific in what you assume, what your modified model is, the hypothesis you are testing, and how you would perform the hypothesis test.

4. For a simple linear model,

$$y = X\beta + e,$$

with $X = (100 \times 6)$ and $y = (100 \times 1)$,

(a) Show that OLS estimation is inefficient if $var(e_i) = \sigma^2 + \tau x_{i3}$. That is, the model has heteroscedasticity related to the third regressor.

(b) Find the properties of the GLS estimator if you wrongly assume the heteroscedasticity follows the pattern $var(e_i) = \sigma^2 x_{i3}$.

(c) Derive the conditions under which you have made things better and those for which you have made things worse.

5. Consider four statistical tests used in econometrics: Wald test, Likelihood Ratio Test, F-test, and Lagrange Multiplier Test.

(a) For each test, describe a hypothesis for each it is used to examine.

(b) For each hypothesis in part a, show how to calculate the relevant test statistic.

(c) Indicate how each test statistic in part b is distributed and how to calculate the degrees of freedom.

Econometrics Qualifying Exam

Department of Agricultural and Applied Economics

May 31, 2013

Please Show work fully and write neatly – unreadable writing could affect your score

1. The following production function was estimated on a cross-sectional sample of firms:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(L_i) + \beta_2 \ln(K_i) + u_i$$

where Y_i is output, L_i is labor input, and K_i is capital input. Assume that all classical assumptions hold.

Explain two different methods for testing whether there are constant returns to scale. State the adequate null hypotheses and the testing procedures step by step.

2. In deciding the "best" set of explanatory variables for a regression model, some researchers follow the method of stepwise regression. In this method one proceeds either by introducing the X variables one at a time (stepwise forward regression) or by including all the possible X variables in one multiple regression and rejecting them one at a time (stepwise backward regression). The decision to add or drop a variable is usually made on the basis of the contribution of that variable to the explained sum of squares, as judged by the F test. Considering your knowledge of econometrics, would you recommend either procedure? Fully explain with attention to detail why or why not.

3. Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \qquad \text{where} \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Here $\mathbf{X}$ is an nxk matrix of rank k and its first column is the unit vector $\mathbf{i}$, $\boldsymbol{\beta}$ is a kx1 vector of unknown parameters, and $\mathbf{y}$, $\boldsymbol{\varepsilon}$, and $\mathbf{e}$ are nx1 vectors. Also define $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

a. Show that $\mathbf{e} = (\mathbf{I}-\mathbf{P})\mathbf{y}$ where $\mathbf{I}$ is an identity matrix of dimension n.
b. Show that $\mathbf{e}'\mathbf{i} = 0$.
c. Show that $\mathbf{X}'\mathbf{e} = \mathbf{0}$.
d. Show that $\mathbf{e}'\mathbf{e} = \mathbf{y}'(\mathbf{I}-\mathbf{P})\mathbf{y}$.

Second Econometrics Qualifying Exam
Department of Agricultural and Applied Economics
July 31, 2013

1. Suppose that we are interested in studying the effect of x on y. To that effect we collect data for 100,000 individuals and use OLS to estimate the following model:

$$y = \beta_1 + x\beta_2 + e$$

with the estimated model being

$$\underset{(2.0)}{y = 10} + \underset{(0.3)}{0.7x}$$

and the standard errors given in parentheses. From the estimated model, we conclude that x has an important effect on y. Provide and discussed in detail at least four reasons why this claim might be misleading.

2. In deciding the "best" set of explanatory variables for a regression model, some researchers follow the method of stepwise regression. In this method one proceeds either by introducing the X variables one at a time (stepwise forward regression) or by including all the possible X variables in one multiple regression and rejecting them one at a time (stepwise backward regression). The decision to add or drop a variable is usually made on the basis of the contribution of that variable to the explained sum of squares, as judged by the F test. Considering your knowledge of econometrics, would you recommend either procedure? Fully explain with attention to detail why or why not.

3. Consider the following simultaneous equations model:

$$y_{1t} = \beta_{12}y_{2t} + \gamma_{11}x_{1t} + u_{1t};$$
$$y_{2t} = \beta_{21}y_{1t} + \gamma_{22}x_{2t} + \gamma_{23}x_{3t} + u_{2t};$$

where y_{1t} and y_{2t} are endogenous variables, x_{1t}, x_{2t} and x_{3t} are exogenous variables, and (u_{1t}, u_{2t}) are normally distributed random disturbances with zero expected value and covariance matrix Σ.

(a) Discuss the identifiability of each equation of the system in terms of the order and rank conditions for identification.

(b) What are the Two-Stage Least Squares estimators of the coefficients in the two equations? Describe the procedure step by step.

4. Consider the following time series regression model:

$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t$ with $t = 1, 2, \ldots, T$ and $\mathbf{X}_t$ being a 1xk vector.

a. Suppose $E[u_t^2] = t^{1/2}\sigma^2$ and define the GLS estimator of $\boldsymbol{\beta}$ as $(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$. Show the exact form of the matrix $\boldsymbol{\Omega}$.

b. Continuing with the setup in a., show that an equivalent GLS estimator can be obtained by applying least squares to the model $\frac{y_t}{t^{1/4}} = \frac{\mathbf{X}_t}{t^{1/4}}\boldsymbol{\beta} + \frac{u_t}{t^{1/4}}$.

c. Now suppose instead (ignore the supposition in a.) that $u_t = \rho u_{t-1} + \varepsilon_t$. Here ε_t is a white noise disturbance uncorrelated with u_t and $|\rho| < 1$. Show how and why the model can be "quasi-differenced" (i.e. express all the data in differences rather than levels) and then estimated by OLS to correct for the autocorrelated disturbances.

5. How might you go about choosing between two model specifications such as:

(M1) $y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + e$
(M2) $y = x_1\beta_1 + \ln(x_2)\beta_2 + (x_3)^2\beta_3 + x_4\beta_4 + e$

a) Explain the steps you would take to make a decision on the best model.
b) What if the second model was $\ln(y) = x_1\beta_1 + \ln(x_2)\beta_2 + (x_3)^2\beta_3 + x_4\beta_4 + e$? How does that change your answer?

Econometrics Qualifying Exam
May 18, 2012

Answer 5 out of 6 questions
Show work fully and write neatly. Good luck.

1. An econometrician estimates the following equation for the price of beef:

$$P_t = \beta_0 + \beta_1 P_{t-1} + \beta_2 M_t + u_t$$

where u_t is a random shock and M_t is some policy variable, say money growth. The econometrician estimates that β_2 is positive and statistically significant at the p=0.07 level.

a) What does this say about policy effectiveness?

b) Could β_2 be estimated as positive even if the policy is ineffective? Use a combination of statistical theory, mathematics and verbal explanation to fully justify your answer.

2. Discuss an example of a linear regression model where it is necessary to use an instrumental variable (IV) estimator. Provide the formula for this estimator and thoroughly discuss its asymptotic properties.

3. Consider the standard linear regression model $y = \beta_0 + \beta_1 x + u$ with x being a single regressor and the model satisfying the Gauss-Markov assumptions. The usual OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased for their respective population parameters. Let $\tilde{\beta}_1$ be the estimator of β_1 obtained by assuming the intercept is zero.

   a) Find $E(\tilde{\beta}_1)$ in terms of x_i, β_0 and β_1, and identify all conditions required for $\tilde{\beta}_1$ to be unbiased for β_1.

   b) Find the variance of $\tilde{\beta}_1$ and compare it to $\hat{\beta}_1$.

   c) Discuss the trade-off one faces when choosing between $\tilde{\beta}_1$ and $\hat{\beta}_1$. Explain when you would advise an analyst to use $\tilde{\beta}_1$, and when you would advise him to use $\hat{\beta}_1$.

4. A particular heart disease has a prevalence of 1/1000 people. A test to detect this disease has a false positive rate of 5% (meaning 5% of healthy people incorrectly are tested as being ill.). Assume that the test diagnoses correctly every person who has the disease. What is the chance that a randomly selected person found to have a positive result actually has the disease?

5. For the model $y = X\beta + \varepsilon$, where $X = (100 \times 5)$ and $\varepsilon \sim N(0, \sigma^2)$,

   a) Show all the steps to efficient estimation of the model while imposing the restrictions: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$.

   b) Show all the steps to testing those joint restrictions.

   c) Show all the steps to testing the first restriction ($\beta_2 = \beta_3$) while imposing the second ($\beta_4 + \beta_5 = 1$).

6. Suppose that you need to estimate a system of two simultaneous equations, one of which is just identified and the other being over-identified. Discuss and compare the properties of the instrumental variable, two-stage and three-stage least squares estimator for the parameters of each of those two equations.

Department of Agricultural and Applied Economics
2009 PhD Econometrics Qualifying Exam

Please answer each of the following questions thoroughly:

1. For each of the following models, state which of the three classical tests (likelihood ratio test, Lagrange multiplier test, Wald test) would be easiest for testing the null hypothesis stated next to the model. Then show the steps to performing the test and the form of the test that would be used.

   a) $y = X\beta + e$ Ho: $\beta_3\beta_4 = 1$

   b) $y = (X\beta + e)^{\eta}$ Ho: $\eta = 1$

   c) $y = X\beta + Z\gamma + e$ Ho: $\gamma_1 = \gamma_2 = \gamma_3 = 0$

2. For a simple linear model, $y = X\beta + e$, with $X = (100 \times 6)$ and $y = (100 \times 1)$,

   a) Show that OLS estimation is inefficient if $var(e_i) = \sigma^2 x_{i3}$. That is, the model has heteroscedasticity related to the third regressor.

   b) Find the properties of the GLS estimator if you wrongly assume the heteroscedasticity follows the pattern $var(e) = \sigma^2 x_{i4}$.

   c) Have you made things better or worse?

3. Consider the following models:

   (i) $y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \varepsilon$

   (ii) $y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon$

   (iii) $y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + \varepsilon$

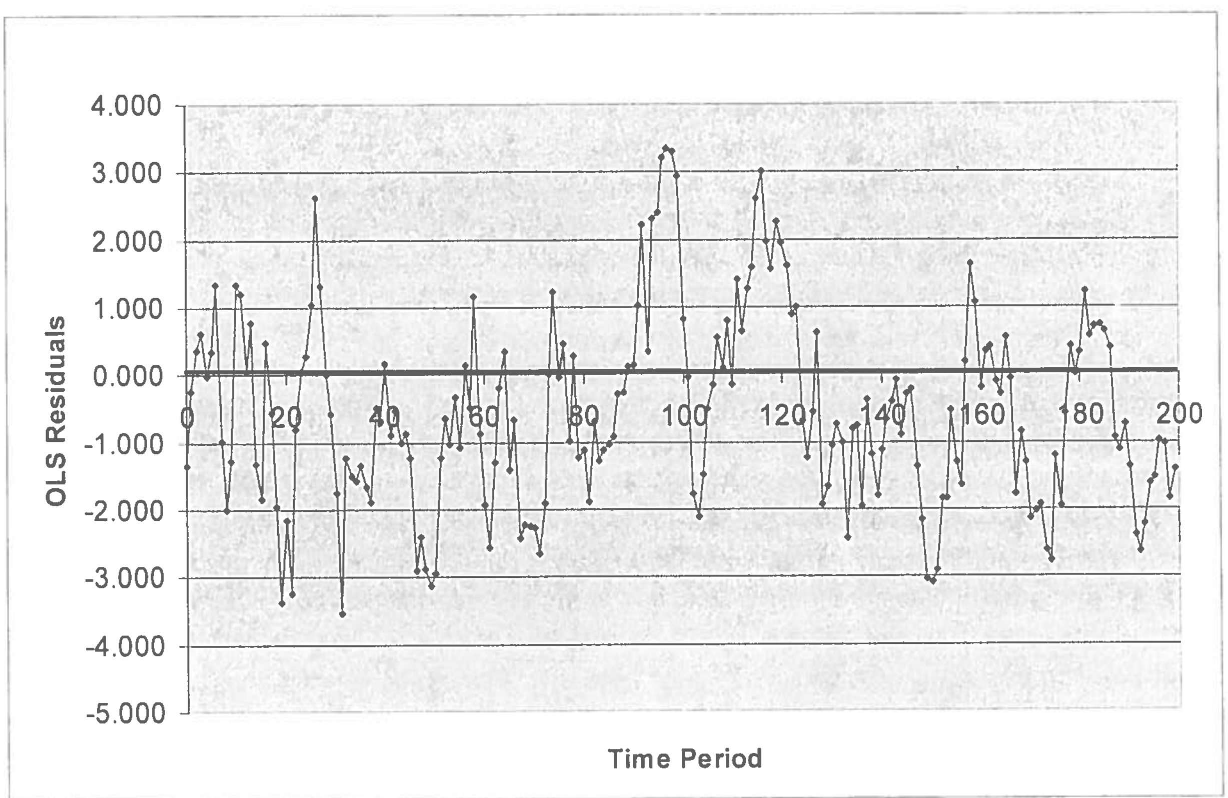
   a. If you estimate (i) but the correct model is (ii) and X_3 is correlated with X_1, what are the consequences for your estimates of the coefficients and their variances in model (i)? Would you be comfortable with the results of hypothesis tests based on model i? Why or why not?

   b. If you estimate (i) but the correct model is (ii) and X_3 is uncorrelated with X_1 and X_2, what are the consequences for your estimates of the coefficients and their variances in model (i)? Would you be comfortable with the results of hypothesis tests based on model i? Why or why not?

c. If you estimate (iii) but the correct model is (ii), what are the consequences for your estimates of the coefficients and their variances in model (iii)? Would you be comfortable with the results of hypothesis tests based on model iii? Why or why not?

4. You have estimated the model $y = \beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + \varepsilon$ using 103 observations via OLS and found that the maximum eigenvalue associated with the X'X matrix is 6000; the minimum eigenvalue is 0.006. What problem do you have? How does this affect your estimates of the coefficients, their variances, and your hypothesis tests? What can you do about this problem?

5. Below is a graph of the OLS residuals of a regression model. What major OLS assumption is likely violated in that model? Describe the details of a formal test that should allow you to ascertain whether this assumption is in fact being violated. Indicate any limitations associated with this test. Suggest possible error-term specifications that might, through GLS estimation, alleviate this problem. Generally describe (i.e. outline the basic steps of) how you would go about selecting the proper error-term specification.



Department of Agricultural and Applied Economics
2010 PhD Econometrics Qualifying Exam

1. What is multicollinearity? Explain the consequences of the presence of multicollinearity in a multiple regression model. Discuss two alternative procedures commonly used to detect multicollinearity. Outline three actions that can be undertaken to alleviate multicollinearity. Can a model that, according to the commonly used detection procedures suffers from multicollinearity, be reliably used to make statistical inferences (please explain your answer)?

2. Suppose a model suffers from a substantial heteroskedasticity problem, which has proven difficult to address through non-linear (generalized) least squares methods. What alternative course of action would you suggest that will make it possible for the model to be useful for making statistical inferences? Explain in detail the computations involved in this "correction." Also indicate any disadvantages of this approach and under which condition(s) it would not be suitable.

3. Why is OLS such an attractive estimation approach? When is it most appropriate or inappropriate? Explain using both words and math.

4. The economic health of a farm economy is measured by the variable y which can be explained quite well by the model: $y = X\beta + \varepsilon$, where β is a (5 x 1) vector and the five variables in X include a constant. The government changed farm policy ten years ago in a manner that was hoped to have a positive effect on y, although the impact might have occurred slowly over several years. Assuming you have time series data on (X, y) how would you modify the model above to test a hypothesis about the impact of the new farm policy. Be specific in what you assume, what your modified model is, the hypothesis you are testing, and how you would perform the hypothesis test.

5. Consider the following model: $y_t = \beta x_t + u_t \quad t = 1, \ldots, T$, where y_t is the dependent variable, x_t is the explanatory variable and u_t is an error term with mean 0. Assume that the explanatory variable can be described by the following equation: $x_t = \alpha z_t + \delta u_t + w_t$

   Also assume the following:

   $$\alpha \neq 0 \qquad p\lim \frac{1}{T}\sum z_t u_t = 0 \qquad p\lim \frac{1}{T}\sum u_t^2 = \sigma_u^2 \neq 0$$
   $$p\lim \frac{1}{T}\sum w_t u_t = 0 \qquad p\lim \frac{1}{T}\sum w_t^2 = \sigma_w^2 \neq 0$$
   $$p\lim \frac{1}{T}\sum z_t w_t = 0 \qquad p\lim \frac{1}{T}\sum z_t^2 = \sigma_z^2 \neq 0$$

   a. Is the OLS estimator of equation (1) consistent or inconsistent? Explain your answer.
   b. Show that z_t is a valid instrument for x_t and state the auxiliary equation.
   c. Explain how your answer to part a changes if δ=0, and explain how one can test the hypothesis H_0: δ=0.

Department of Agricultural and Applied Economics
Second 2010 PhD Econometrics Qualifying Exam

1. Discuss the finite sample properties of the OLS estimator when the error term is not normally distributed. Also when working with small samples and non-normally distributed errors, discuss the validity of the usual F-statistic to test restrictions in OLS-estimated models.

2. Consider a linear model with dependent variable Y_i (Dr. Ramirez's blood pressure on a given day), an intercept, and explanatory variables X_{1i} (1 if week day, zero otherwise), X_{2i} (number of meetings to be attended on that day), X_{3i} (number of emails responded to on that day) and X_{4i} (number of cups of coffee drank to on that day). Further assume that the model's error term is iid normal with an expected value of zero. Yesterday, Dr. Ramirez had to attend four meetings, respond to 42 emails and drank five cups of coffee. Explain how you would go about computing a 95% confidence interval for what Dr. Ramirez's blood pressure was yesterday. Also explain how this procedure would need to be adjusted to compute a confidence interval for Dr. Ramirez's average blood pressure during all weekdays when he happens to attend four meetings, responds to 42 emails and drinks five cups of coffee. Would such confidence interval be narrower or wider (please explain your answer briefly)?

3. Suppose that you have estimated the following regression:

   $\text{Ln}(salary) = \beta_0 + \beta_1 \ln(mktval) + \beta_2 \ln(sales) + \beta_3 \ln(ceoten) + u$

   Where:

   Salary = salary of the firm's Chief Executive Officer (CEO)
   Mktval = the firm's market value in 1,000s of dollars
   Sales = value of sales in 1,000s of dollars
   Ceoten = number of years the CEO has been in that job
   $u \sim N(0, \sigma^2)$

   Using 100 cross sectional observations for firms in the US, you get the following estimates for the betas (standard errors in parentheses):

   $$\underset{}{\text{Ln}(salary)} = \underset{(0.33)}{4.504} + \underset{(0.015)}{0.11} \ln(mktval) + \underset{(0.11)}{0.16} \ln(sales) + \underset{(0.08)}{0.12} \ln(ceoten)$$

   R^2=0.318, Σ exp(u-hat)=113.6, Σ u-hat^2 =113.6

   a. Detail how you would predict salaries and demonstrate whether or not your predictions would be consistent.
   b. Describe how you would measure how well the model above explains the variation in *salary* (not ln(*salary*)).

4. Statisticians often estimate models using what is called stepwise-regression. This involves considering a large set of potential regressors, then repeatedly changing the specification to drop variables which are statistically insignificant while adding back candidate regressors to see if they will now be statistically significant. What are the resulting properties of such an estimator, in finite samples and asymptotically?

5. For the model y=Xb+e where y = (100x1) and b = (6x1), state the minimum assumptions necessary to prove that:

   a. The OLS estimator for b is unbiased.
   b. The OLS estimator for b is in fact the minimum variance linear unbiased estimator for that parameter vector.

   Prove a. and b. above again stating all assumptions used in your proofs when they are needed.

4. In a simultaneous equations model with two endogenous variables, does it matter which variables are placed on the left-hand side of each equation? That is, if you estimate the two models:

(M1) $y_1 = y_2\delta + X\beta + e_1$

$y_2 = y_1\lambda + Z\gamma + e_2$

(M2) $y_1 = y_2\delta + X\beta + e_1$

$y_1 = y_2\lambda + Z\gamma + e_2$

will the estimated coefficients, standard errors, and model fit measures (like R^2) vary depending on whether you estimate the model as written in a) or in b) (in your answer make sure you address all three points (coefficients, standard errors, and model fit)? In addition, discuss from an empirical perspective whether it would be better to estimate M1 or M2 and why.

5. For a simple linear model, $y = X\beta + e$, with $X = (100 \times 6)$ and $y = (100 \times 1)$,

a) Show that OLS estimation is inefficient if $var(e_i) = \sigma^2 + \tau x_{i3}$. That is, the model has heteroscedasticity related to the third regressor.

b) Find the properties of the GLS estimator if you wrongly assume the heteroscedasticity follows the pattern $var(e) = \sigma^2 x_{i3}$.

c) Have you made things better or worse?

Econometrics Prelim Examination
Ag and Applied Economics Dept.
University of Georgia

May 20, 2008

Please answer each of the following questions. You have until 4:00 pm.
GOOD LUCK!!

1. How are the problems caused by measurement error and omitted variables different and/or the same? Which is a more severe problem? What would you do to correct each one?

2. For the model $y = X\beta + \varepsilon$ where X is (50 x 4) answer the following:
   a. If $E(\varepsilon_i^2) = \sigma_i^2 = \delta_0 + \delta_1 X_{i2}$ explain what the properties of a least squares estimator for β will be if you correct for heteroscedasticity using the wrong pattern (perhaps, $\sigma_i^2 = \lambda_1 X_{i3}$).
   b. Compare this outcome to the properties of a least squares estimator that is corrected for the presence of an AR(1) error term when autoregression is not actually present in the model.

3. When can you use an exact F-test to test a multiple hypothesis and when must you rely on asymptotic tests? Also discuss when it is advantageous to use each of the three asymptotic tests (likelihood ratio test, LaGrange multiplier test, and the Wald test).

4. In addition to causing flood damage, El Nino has created social costs by interfering with Southern Californians' enjoyment of their public beaches. Suppose the Department of Beaches has asked you to assess the welfare effects of beach closures due to storm-drain runoff from El Nino events. To do this, you need to estimate a model of local demand for public beaches, and then see how consumer's surplus from beach trips changes as this demand function shifts according to the number of days of beach closures (CLOSURES) each month. You collected survey data each month from a different random sample of Angelenos concerning the number of beach trips (TRIPS) they have made in that month as a function of the distance (DIST) they live from the beach. (Since beach access is free in most of Southern California, you will use this distance times average-travel-cost-per-mile as a rough proxy for the price of access). In the process of analyzing the effects of closures, you model demand by regressing TRIPS on DIST and CLOSURES and a set of sociodemographic characteristics such as age, income, and gender. Suppose a 1% change in DIST corresponds roughly to a 1% change in the "price" of a beach visit. Why should you be cautious about taking the results from this regression at face value (especially those concerning the price elasticity of demand for beach visits)?

5. Suppose you have been hired by a large national recreational equipment cooperative to assess individual consumer expenditures on the types of products sold by the cooperative. You are provided with some survey data on individual expenditures (EXP) by AGE, gender (FEM=1 if female) and income (INC, in thousands of dollars per year). The best-fitting model you discover is displayed below.

a.) Based on the point estimates, provide a formula that would give expected expenditures for a randomly selected female. (Two significant digits will be adequate.)

b.) Explain how you would go about testing whether expected expenditures differ by gender.

c.) What appears to be the main difference between the male and female age profiles of expenditure on recreational equipment in this sample?

d.) Does an extra $1000 of annual income have any statistically discernible effect on recreational equipment expenditures? Explain carefully.

```
|_sample 1 50
|_read(recr.dat) exp age inc fem

|_stat / pcor
NAME          N    MEAN          ST. DEV        VARIANCE       MINIMUM        MAXIMUM
EXP          50   175.70         66.647         4441.9         58.336         290.47
AGE          50   47.144         16.882         284.99         17.355         72.648
INC          50   39.511         21.061         443.57         2.3366         85.351
FEM          50  0.52000        0.50467        0.25469        0.00000         1.0000

|_genr age2=age*age
|_genr inc2=inc*inc
|_genr femage=fem*age
|_genr femage2=fem*age2
|_genr ageinc=age*inc

|_ols exp age fem inc age2 femage femage2 ageinc

 R-SQUARE =    0.9463     R-SQUARE ADJUSTED =    0.9373
VARIANCE OF THE ESTIMATE-SIGMA**2 =    278.36
STANDARD ERROR OF THE ESTIMATE-SIGMA =    16.684
SUM OF SQUARED ERRORS-SSE=    11691.
MEAN OF DEPENDENT VARIABLE =    175.70
LOG OF THE LIKELIHOOD FUNCTION = -207.311

VARIABLE   ESTIMATED  STANDARD   T-RATIO         PARTIAL STANDARDIZED ELASTICITY
  NAME    COEFFICIENT   ERROR     42 DF    P-VALUE CORR. COEFFICIENT  AT MEANS
AGE         4.9635      1.458      3.404    0.001 0.465      1.2573      1.3318
FEM        -39.113      45.82    -0.8536    0.398-0.131     -0.2962     -0.1158
INC       -0.57530     0.5049     -1.140    0.261-0.173     -0.1818     -0.1294
AGE2      -0.63225E-01 0.1844E-01 -3.429    0.001-0.468     -1.4807     -0.9003
FEMAGE     -6.0847      2.109     -2.885    0.006-0.407     -2.6707     -0.9700
FEMAGE2    0.78749E-01 0.2295E-01  3.432    0.001 0.468      2.2651      0.7275
AGEINC     0.26949E-01 0.9574E-02  2.815    0.007 0.398      0.5965      0.3174
CONSTANT   129.80       26.40      4.916    0.000 0.604      0.0000      0.7387
```