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?