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?