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 $s\geq 1$. 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 π=0.5 at the α=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 π=0.5 at the α=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.