Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

June 30, 2016

There are **FOUR** questions. Answer each part of each question. All questions are weighted equally. Within each question, each part will receive equal weight in grading.

I. There are two parts to this question

(a) We observe random variables $Y_1, Y_2, \ldots, Y_n$ and fixed variables $X_1, \ldots, X_n$ such that $E[Y_i|X_i] = X_i'\beta$, where β is an unknown parameter. Conditional on $X_1, \ldots, X_n$, the elements of $Y_1, Y_2, \ldots, Y_n$ are mutually independent, each with a normal distribution and variance σ^2.

(i) Find the maximum likelihood estimator for β. Compute its mean and variance.

(ii) Write down the Cramer-Rao lower bound for the variance of an unbiased estimator for β.

(iii) Find the best linear unbiased estimator for β.

(b) Consider the following model:

$$y_i = x_i'\beta + \delta T_i + u_i,$$

where y_i is log wages, x_i includes a constant, grades completed, square of grades completed and job tenure, and T_i equals 1 if individual i receives a job training, and 0 otherwise.

(i) Suppose a researcher is interested in the effectiveness of a job training program on wages. OLS estimation of the model produces a δ coefficient that is positive and highly significant ($p - value = 0.0002$). Would you conclude that the job training program is extremely effective? Why or why not? In your response, comment on which, if any, of the classical linear regression assumptions may be violated.

(ii) Now assume that $u_i = a_i + \epsilon_i$, where a_i is unobserved ability. Derive the bias in the OLS estimator of δ due to unobserved ability. State any assumptions you require for your derivation. Does it lead to an over-estimation or under-estimation of the effect of job training?

II. For $i = 1, \ldots, n$, consider the following production function $Y_i = (\alpha K_i^\rho + (1-\alpha)L_i^\rho)^{1/\rho} + \epsilon_i$, where Y is output, K is capital, and L is labor.

(a) Assume that $E[\epsilon_i|K_i, L_i] = 0$, propose a least squares, method of moments and generalized method of moments estimator for the parameter vector $\mu = (\alpha, \rho)'$.

(b) For each of the estimators in (a), give sufficient conditions for identification of μ.

(c) Is any of the estimators in (a) efficient? If yes, explain your answer. If not, propose an efficient estimator.

(d) Using the efficient estimator, propose a Wald test for the null hypothesis: $H_0 : \alpha = \rho = 1$.

(e) Using the least squares estimator, write down the score that is used to construct the score test for the same null hypothesis in (d).

III. For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, let x_{it} and β be $k \times 1$ column vectors, y_{it} and u_{it} are scalar, where $y_{it} = x_{it}'\beta + a_i + u_{it}$. Let $X_i = (x_{i1}, ..., x_{iT})$ and $u_i = (u_{i1}, ..., u_{iT})'$. Assume $E[u_{it}|X_i, a_i] = 0$ for all t.

(a) Assume $E[a_i|X_i] = 0$, $E[a_i^2|X_i] = 0$, $E[u_{it}^2|X_i, a_i] = \sigma_u^2$ and $E[u_{it}u_{i,t-\tau}|X_i] = \rho^\tau \sigma_u^2$.
Under the above scenario:

(i) Write down the conditional variance-covariance matrix of $v_i = a_i + u_i$.

(ii) Would the pooled OLS estimator yield a consistent estimator for β? To answer this question, show the probability limit of the estimator and give any conditions required for the result.

(iii) Is the pooled OLS estimator asymptotically efficient? Provide a formal argument for your answer.

(iv) If not, propose an asymptotically efficient estimator and show how you would estimate all components required.

(b) Now assume $E[a_i|X_i] \neq 0$, $E[a_i^2|X_i] > 0$, $E[u_{it}^2|X_i, a_i] = \sigma_u^2$ and $E[u_{it}u_{i,t-\tau}|X_i] = \rho^\tau \sigma_u^2$.
Under the above scenario:

(i) Show whether or not the pooled OLS estimator is consistent for β. If not, propose an estimator that would be consistent.

(ii) Is the consistent estimator (according to your answer in (i)) asymptotically efficient? Give a formal argument for your answer.

(c) If you would like to test whether scenario (a) or (b) are true in practice, how would you proceed? Describe the test statistic you would use and any specific issues that would arise in its application under the above assumptions.

IV. (a) State a Law of Large Numbers (LLN) and explain in words what it means.

(b) State a Central Limit Theorem (CLT)and explain in words what it means.

(c) Suppose that 40% of all auto accidents are partly caused by alcohol consumption and 30% of all auto accidents involve bodily injury. Further, of those accidents that involve bodily injury, 50% are partly caused by alcohol consumption.

(i) What is the probability that a randomly chosen accident is partly caused by alcohol consumption and involves bodily injury?

(ii) If a randomly chosen accident was partly caused by alcohol consumption, what is the probability that it involved bodily injury?

(iii) Are the events partly caused by alcohol consumption and involved bodily injury independent? Why or why not?

(iv) Suppose that you have data on 100 randomly sampled auto accidents, including whether the accident involved bodily injury and the amount of alcohol in the blood of each driver involved in the accident. In your sample, 20% of the accidents involve bodily injury. Form a 95% asymptotic confidence interval for the proportion of accidents involving bodily injury. If you invoke the LLN or CLT to obtain the confidence interval, state precisely why you do so.

(v) Based on your answer in (iv), do you think your data are a true random sample of the population of auto accidents? Why or why not?

(vi) Suppose that you have data on 100 randomly sampled auto accidents, including whether the accident involved bodily injury and the amount of alcohol in the blood of each driver involved in the accident. How would you test the hypothesis that the events "partly caused by alcohol consumption" and "involved bodily injury" are independent? State precisely the test statistic you would use and justify your choice.

(vii) Derive the asymptotic null distribution of your test statistic in (vi). Highlight all points in your derivation where you invoke the LLN or CLT.