Econometrics Preliminary Exam

Agricultural and Resource Economics, UC Davis

August 8, 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.

1. Given the following linear regression model

$$Y = X\beta + u$$
$$E[u|X] = 0$$
$$E[uu'|X] = \sigma^2\Omega = \Sigma$$

where $Y = (y_1, \ldots, y_T)'$ and $u = (u_1, \ldots, u_T)'$ are $T \times 1$ vectors, $X = (x_1, \ldots, x_T)'$ is an $T \times k$ matrix and β is a $k \times 1$ vector.

   (a) Show that the OLS estimator of β is unbiased.

   (b) Find the variance-covariance matrix of the OLS estimator of β.

   (c) Show that the OLS estimator of β is consistent.

   (d) Is the OLS estimator efficient? If not, propose an efficient estimator of β.

   (e) Now assuming that there is no heteroskedasticity. Suppose that the disturbance term in the above model is an AR(1) process, i.e., $u_t = \rho u_{t-1} + \epsilon_t$.

      i. Consider the case where $|\rho| < 1$. Is the OLS estimator consistent? Is yes, give sufficient conditions for its consistency.

      ii. Now what happens to your answer in (i) if $\rho = 1$?

2. Consider the following objective function, $Q(\theta) = E[m(y_i, x_i; \theta)]$, where θ and x_i are $k \times 1$ vectors and y_i is a scalar.

   (a) Given n observations of $\{y_i, x_i\}$, propose an estimator of $\theta_0 = \arg\max_{\theta\in\Theta} Q(\theta)$ and give conditions for its consistency.

   (b) Derive an expression for the sampling error, $\sqrt{n}(\hat{\theta} - \theta_0)$, and give conditions for the asymptotic normality of the estimator in (a). Make sure to give the expression for the asymptotic variance, $Avar(\sqrt{n}(\hat{\theta} - \theta_0))$. Explain briefly how the conditions you propose imply the asymptotic normality of the sampling error.

(c) Now consider the following generalizations of ordinary least squares (OLS) estimation, ridge regression ($\hat{\theta}_R$) and LASSO ($\hat{\theta}_L$). Let θ^j denote the j^{th} element of θ,

$$\hat{\theta}_R = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n}(y_i - x_i'\theta) + \lambda \sum_{j=1}^{k}(\theta^j)^2$$

$$\hat{\theta}_L = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n}(y_i - x_i'\theta) + \lambda \sum_{j=1}^{k}|\theta^j|$$

where λ is specified by the empirical researcher.

i. Show that the OLS estimator is a special case of $\hat{\theta}_R$ and $\hat{\theta}_L$.

ii. Define the population objective function for $\hat{\theta}_R$ and $\hat{\theta}_L$.

iii. Now consider the conditions you specified in (a) and (b). Do $\hat{\theta}_R$ and $\hat{\theta}_L$ fulfill these conditions? If not, which conditions do they violate? Be precise and clear in your answer.

3. Consider the following estimation problems.

(a) Let W_1 and W_2 be dependent random variables, where $W_2 = \alpha W_1 + \eta$, and η and W_1 are i.i.d. N(0,1) random variables. Suppose you observe an i.i.d. random sample of $\{W_{1i}, W_{2i}\}_{i=1}^n$, propose an estimator of α and give conditions for its consistency and asymptotic normality. Is the estimator you proposed efficient?

(b) Now consider the two-period dynamic panel data model, where

$$Y_{i2} = \rho Y_{i1} + X_{i2}'\beta + \alpha_i + u_{it},$$

Given an i.i.d. random sample of $\{Y_{i1}, Y_{i2}, X_{i1}, X_{i2}\}_{i=1}^n$, can you use insights from the estimation strategy you proposed in (a) to estimate β? State any additional assumptions you require. Define the estimator you propose and its objective function.

4. This question is based on a 1996 paper by Steve Levitt in the Quarterly Journal of Economics, which addresses the effect of imprisonment on violent crime. Levitt's data are measured annually at the state level, i.e., one observation for each U.S. state in each year from 1980-1993. Consider the equation:

$$gcriv = \beta_0 + \beta_1 gpris + \beta_2 gincpc + \varepsilon$$

where *gcriv* denotes the annual growth rate in violent crime, *gpris* denotes the annual growth rate in the number of prison inmates per resident, and *gincpc* denotes per capita income.

The parameter of interest is β_1, which measures the marginal effect of an increase in imprisonment on the crime rate.

(a) Would ordinary least squares (OLS) produce a consistent estimate of β_1? Justify your answer clearly in words.

(b) Write the model in the notation from question 1. Clearly define each term and show which property of the model determines whether OLS is consistent for β_1.

(c) Levitt observes two additional variables: (i) *final1* is a dummy variable denoting a final decision in the current year on legislation to reduce prison overcrowding, and (ii) *final2* is a dummy variable denoting a final decision in the last two years on legislation to reduce prison overcrowding. Levitt uses these variables to instrument for $gpris$. What would need to be true for this instrumental variables (IV) estimator to produce a consistent estimate of β_1? Justify your answer clearly in words.

Using Levitt's data, we estimated the model by OLS and IV. The STATA output is shown on the last page of the exam.

(d) Both the OLS and IV estimation uses the robust command to correct the standard errors for heteroscedasticity. How might the results differ if this correction were not done?

(e) The IV estimate of β_1 is a larger negative number than the OLS estimate. Explain in words whether this result makes sense.

(f) Describe how you would check the strength of *final1* and *final2* as instruments for imprisonment. Explain the implications for the results if the instruments are weak.

```
. regress gcriv gpris gincpc, robust

Linear regression                                 Number of obs =     714
                                                  F(  2,   711) =   15.62
                                                  Prob > F      =  0.0000
                                                  R-squared     =  0.0461
                                                  Root MSE      =  .08661
------------------------------------------------------------------
             |               Robust
       gcriv |     Coef.  Std. Err.    t   P>|t|    [95% Conf. Interval]
-------------+----------------------------------------------------
       gpris |-.1954158  .0531582  -3.68  0.000   -.2997817   -.09105
      gincpc | .4667109  .1468838   3.18  0.002    .1783331    .7550887
       _cons | .0043254  .0106637   0.41  0.685   -.0166106    .0252615
------------------------------------------------------------------


. ivregress gcriv (gpris=final1 final2) gincpc, robust

Instrumental variables (2SLS) regression          Number of obs =     714
                                                  F(  2,   711) =    8.82
                                                  Prob > F      =  0.0002
                                                  R-squared     =       .
                                                  Root MSE      =  .10484
------------------------------------------------------------------
             |               Robust
       gcriv |     Coef.  Std. Err.    t   P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------
       gpris |-1.082207  .3154422  -3.43  0.001   -1.701517   -.4628977
      gincpc | .3798519  .2007459   1.89  0.059   -.0142738    .7739776
       _cons | .0684567  .0255884   2.68  0.008    .0182188    .1186946
------------------------------------------------------------------
Instrumented:  gpris
Instruments:   gincpc final1 final2
------------------------------------------------------------------
```