ECNS561, Fall 2011

Take-home Graded Practice Opportunity

Due date: October 21, 2011, 5:00 p.m.

1. Prove that $\boldsymbol{X'X}$ is a positive definite matrix if $\boldsymbol{X}$ is full rank. That is, show that $\mathbf{c}'\boldsymbol{X'Xc} > 0$, where $\boldsymbol{c}$ is a scalar such that $\boldsymbol{c} \neq 0$. *Hint*: if $\boldsymbol{X}$ is full rank, then $\boldsymbol{Xc} \neq 0$ for any $c \neq 0$.

2. Prove that the projection and residual maker matrices are (a) symmetric and (b) idempotent.

3. Consider a linear model: $y = \boldsymbol{X\beta}+\varepsilon$. Suppose that $E[\boldsymbol{X}'\varepsilon] = 0$ and $Var[\varepsilon|\boldsymbol{X}] = \sigma^2$. However, $E[\varepsilon|\boldsymbol{X}] \neq 0$.

   (a) Does $E[\varepsilon^2|\boldsymbol{X}] = \sigma^2$?

   (b) What does $E[\varepsilon|\boldsymbol{X}] \neq 0$ imply about the OLS estimators?

4. The following problem will help you understand the effects of scaling on linear OLS regression analysis.

   (a) Let $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ be the estimated parameters from the regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Furthermore, assume that c_1 and c_2 are constants, with $c_2 \neq 0$. Now, assume that $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\boldsymbol{\beta}}_1$ are the estimated parameters from the regression of $c_1\boldsymbol{Y}$ on $c_2\boldsymbol{X}$.
   Show that $\tilde{\boldsymbol{\beta}}_1 = (c_1/c_2)\hat{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_0 = c_1\hat{\boldsymbol{\beta}}_0$.
   What does this imply about scaling the dependent and independent variables? How does scaling affect the estimated parameters?

   (b) Let $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ be OLS estimates from the regression of $\log(y)$ on $\boldsymbol{X}$. For a constant $c > 0$, let $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\boldsymbol{\beta}}_1$ be estimates from a regression of $\log(cy)$ on $\boldsymbol{X}$. Show how you can write $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\boldsymbol{\beta}}_1$ as a function of $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$. How does the constant scale the regression coefficients?

5. Consider the following data set describing the number of days until (negative) and after (positive) harvest and the premium (in cents per bushel) for having a 1% increase in wheat protein levels.

Days	Premium
-2	200
-1	180
0	20
1	22
2	20

(a) Explain the economic intuition about what happened at harvest time to the quality level of wheat. How is that reflected in the premiums? What kind of quality was available before the harvest and what quality is available after the harvest?

(b) Using matrix algebra (by hand), perform the following:

i. Compute estimated parameters for the model $Premium = f(Days) + \varepsilon$.

ii. Determine the variance-covariance matrix.

iii. Construct a 95% confidence interval around the population parameter associated with the days.

iv. Test a 90% confidence level hypothesis that the average premium before the harvest was different than the average premium after the harvest. Interpret.

6. Applied Practice

Use the data set "Heart" available under the "Homeworks" category on the class website. Complete the following:

(a) Explore the data using a software package. Provide some basic intuition about the relationships between individuals' weight and their diastolic and systolic blood pressure measures. No regressions.

(b) Read only the columns "Weight" and "Diastolic" from this data set into IML (or another matrix language). Read each column into a separate vector.

(c) Using matrix algebra, compute the parameter estimates of the model: $Diastolic = \beta_0 + \beta_1 Weight + \varepsilon$.

(d) Compute the variance-covariance matrix and the standard errors vector. What are the standard errors for each estimated parameter?

(e) For each estimated parameter, test the hypothesis (95% confidence level) that the true population value of the parameter is zero. At the 95% confidence level, what is the marginal effect of weight on the diastolic blood pressure?

(f) Using the result in (e), what are the potential economic implications of a national health plan? Think about what is happening with the number of individuals in the United States who are overweight. Who might be better off? Who might be worse off?

(g) Hypertension is a condition describing high blood pressure. What are some adverse effects of hypertension? At what level does hypertension become critical to ones health? How much weight, on average, would people in the sample need to get to a dangerous level of hypertension?