Problem Set 5

G63.2707 Fall 2002
Modern Statistical Inference And Econometrics Due October 21

Typographical errors in the Notes: About two-thirds of the way down on page 49, the phrase “unbiased estimator $s^2$ for $\sigma$” should be replaced by “unbiased estimator $\sigma^2$ for $\sigma^2$.” Also, in formula 95 on page 50, $\hat{\sigma}$ should be replaced by $s$. Finally, in the bottom line on page 70, $\hat{\sigma}$ should be replaced by $s$.

The problems concern the classical linear regression model

$$y = X\beta + u$$

where $y$ is the $n$-vector of observations on the dependent variable, $X$ is the $n \times k$ matrix of observations on the explanatory variables, $\beta$ is a $k$-vector of unknown parameters, and $u$ is the $n$-vector of unobserved disturbances. We assume that conditional on $X$, $u$ has mean 0 and covariance matrix $\sigma^2 I$. The ordinary least squares (OLS) estimate of $\beta$ is $b = (X^t X)^{-1} X^t y$ and the vector of residuals is $e = y - Xb$. An unbiased estimate of $\sigma^2$ is $s^2 = e^t e / (n - k)$. The symmetric idempotent matrix $H = X(X^t X)^{-1} X^t$ is called the “hat matrix.” Its $i$th diagonal element is denoted $h_{ii}$. Let $y_i$, $u_i$ and $e_i$ denote the $i$th elements of $y$, $u$ and $e$ respectively, and let $x_{ij}$ denote the element of $X$ in the $i$th row and $j$th column.

1. Suppose that the columns of the matrix $X$ of explanatory variables are orthogonal, meaning

$$\sum_{i=1}^{n} x_{ij} x_{im} = 0 \quad \text{if } j \neq m.$$

Find simple expressions (not involving matrices) for the components of the OLS coefficient vector $b$ in this case.

2. In this problem you will show that $\max_i h_{ii} \to 0$ as $n \to \infty$ in a simple regression ($k = 2$) with a linear time trend. The model is

$$y_i = \beta_1 + \beta_2 x_{i2} + u_i \quad i = 1, \ldots, n$$
where \( x_{i2} = i \) so that the \( n \times 2 \) matrix of explanatory variables is

\[
\begin{bmatrix}
1 & 1 \\
1 & 2 \\
1 & 3 \\
\vdots & \vdots \\
1 & n \\
\end{bmatrix}
\]

For computational simplicity, assume \( n \) is odd.

(a) By subtracting \( (n + 1)/2 \) times the first column from the second column, the matrix of explanatory variables becomes

\[
X = \begin{bmatrix}
1 & -(n - 1)/2 \\
\vdots & \vdots \\
1 & -1 \\
1 & 0 \\
1 & 1 \\
\vdots & \vdots \\
1 & (n - 1)/2 \\
\end{bmatrix}
\]

Explain why this transformation leaves the hat matrix \( H \) unchanged.

(b) Find \( X^tX \) and \( (X^tX)^{-1} \). [Note: The fact that

\[
1^2 + 2^2 + \ldots + \left( \frac{n - 1}{2} \right)^2 = \frac{n(n - 1)(n + 1)}{24}
\]

is helpful.]

(c) Show that \( \max_i h_{ii} \to 0 \) as \( n \to \infty \).

3. Access the data for this problem by clicking on “Data for problem 3 in Problem Set 5” on the course web page. The data set has 40 rows and 3 columns. The first column is \( y \). Columns 2 and 3 of the data set are columns 2 and 3 of \( X \). The first column of \( X \) is a vector of ones. Thus, \( n = 40 \) and \( k = 3 \).

The purpose of this computational exercise is to familiarize you with symptoms of non-linearities.

(a) Find \( b \), and compute the vector \( e \) of residuals.
(b) Examine a scatter-plot of the residuals versus the third column of $X$. [Note: In MATLAB, the command `plotmatrix` works well.] Does the expectation of the residual appear to be related to the level of this explanatory variable? If so, what is the apparent relationship?

(c) Expand the model to

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i3}^2 + u_i \quad i = 1, \ldots, n$$

by adding a fourth column to the $X$ matrix, with $x_{i4} = x_{i3}^2$ for all $i$. For this model, find $b$ (which is now a 4-vector) and $s$. Also compute the estimated standard errors of the components of $b$ (the square roots of the diagonal elements of $s^2(X^tX)^{-1}$). Examine plots of residuals versus the explanatory variables. Is there still clear visual evidence of misspecification?