Problem Set 11

Both problems in this problem set use the same data used in Problem Set 4, and can be accessed by clicking on “Data for problem 4 in Problem Set 4” on the course web page. The data set has 50 rows and 6 columns. The first column is \( y \), the vector of observations on the dependent variable. Columns 2 through 6 of the data set are columns 2 through 6 of the matrix \( X \) of explanatory variables. The first column of \( X \) is a vector of ones. Thus, \( n = 50 \) and \( k = 6 \). The model is the usual

\[
y = X\beta + u
\]

where \( \beta \) is a \( k \)-vector of unknown parameters and \( u \) is the \( n \)-vector of unobserved disturbances. In Problem Set 4 you computed the OLS estimate for \( \beta \),

\[
b = \begin{bmatrix}
73.1010 \\
0.9888 \\
-1.2094 \\
0.4824 \\
-0.4833 \\
0.4021
\end{bmatrix}
\]

1. (a) Compute the bounded-influence estimate \( \hat{\beta} \) and the associated matrix \( A \) defined by equations 186 and 187 on pages 86 and 87 of the Notes, following the iterative procedure outlined on page 87. Use \( a = 1.3\sqrt{k} \). Report \( \hat{\beta} \) but not \( A \).

(b) Based on \( (\hat{\beta}, A) \) at the final iteration, compute for each observation \( (i = 1, \ldots, n) \) the residual \( e_i \), the weight \( w_i \), and the “leverage” \( \sqrt{x_i^t A^{-1} x_i} \) where \( x_i \) is the \( i \)th row of \( X \).

2. One weakness of a regression diagnostic based on OLS, such as the one in formula 101 on page 52 of the Notes, is that it is susceptible to “masking.” By this we mean that even when a single anomalous observation would show up clearly as an outlier, the presence of a second, similar anomalous observation can allow both to escape detection. We will illustrate masking in this exercise. In problem 4b of Problem Set 4
you found that observation 15 has a diagnostic of 21.4396, which is by far the largest in the sample.

(a) Augment the matrix $X$ and vector $y$ by adding a 51st observation which is exactly equal to observation 15.

(b) Re-estimate by OLS, and then compute the OLS residual and the diagnostic for each observation. What has happened to the diagnostic for observation 15?