**Qualifying Exam: CAS MA 575.**

Boston University, Fall 2006

1. Suppose that we are interested in the coefficients $\beta$ of a linear model $E(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\beta$, where $\mathbf{Y}$ is $n \times 1$ and $\mathbf{X}$ is $n \times p'$. Furthermore, suppose that it is of interest to partition that model in the form $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, for $n \times p_i$ matrices $X_i$, $i = 1, 2$. Finally, suppose that an investigator creates a partially orthogonal design, in which $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ has the property that $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$.

   (a) Show that the least squares estimate of $\beta$ takes the form $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$, where
   $$\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} \quad \text{and} \quad \hat{\beta}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{Y} \ .$$

   (b) Show that the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are uncorrelated. That is, show that the covariance matrix $\mathrm{Var}(\hat{\beta})$ has a block-diagonal form
   $$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$
   for some matrices $A_1, A_2$.

   (c) Specify an appropriate $F$-statistic for the testing problem
   $$H_0 : E(y|\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2) = \mathbf{x}_1'\beta_1$$
   versus
   $$H_1 : E(y|\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2) = \mathbf{x}_1'\beta_1 + \mathbf{x}_2'\beta_2 \ .$$
   Be sure to define all terms in your statistic, and to state the corresponding degrees of freedom.

2. (a) Let $\mathbf{X}$ be an $n \times p'$ matrix, with linearly independent columns (i.e., of full column-rank). By subscript "$(i)$" we will mean "without the $i$th case". So $\mathbf{X}_{(i)}$ is the $(n-1) \times p'$ matrix created by removing the $i$th row of $\mathbf{X}$ i.e., $\mathbf{x}_i$.
   Let $\mathbf{Y}$ be an $n \times 1$ response vector, and consider the linear regression model $E(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\beta$. Define $\hat{\beta}_{(i)}$ to be the estimate of $\beta$ without the $i$th case, and $\hat{\beta}$, the estimate with all the data.
   Show that
   $$\hat{\beta}_{(i)} = \hat{\beta} - \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\,\hat{e}_i}{1 - h_{ii}} \ ,$$
   using the formula
   $$(\mathbf{X}_{(i)}'\mathbf{X}_{(i)})^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}}{1 - h_{ii}} \ .$$

(b) Consider the figure below. The solid line shows the regression of $y$ on $x$ based on only the scatter of data points in the left-hand portion of the plot. The dotted line shows the same regression as the first, but with the point in the *upper* right-hand corner included too. Similarly, the dashed line shows the same regression as the first, but with the point in the *lower* right-hand corner included too.

Comment on the degree of (i) outlying-ness, (ii) leverage, and (iii) influence of the point in the upper right-hand corner. Do the same for the point in the lower right-hand corner. Justify your answer through appropriate description of the likely values of the statistics $t_i, h_{ii}$, and $D_i$. (That is, the outlier $t$-test value, the hat-matrix entry, and Cook's distance.)