Qualifying Exam: CAS MA 575.

Boston University, Fall 2007

1. Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \ ,$$

for $i = 1, \ldots, n$, where we assume $E[e_i|X] = 0$, $\mathrm{Var}(e_i|X) = \sigma^2$, and the e_i are uncorrelated.

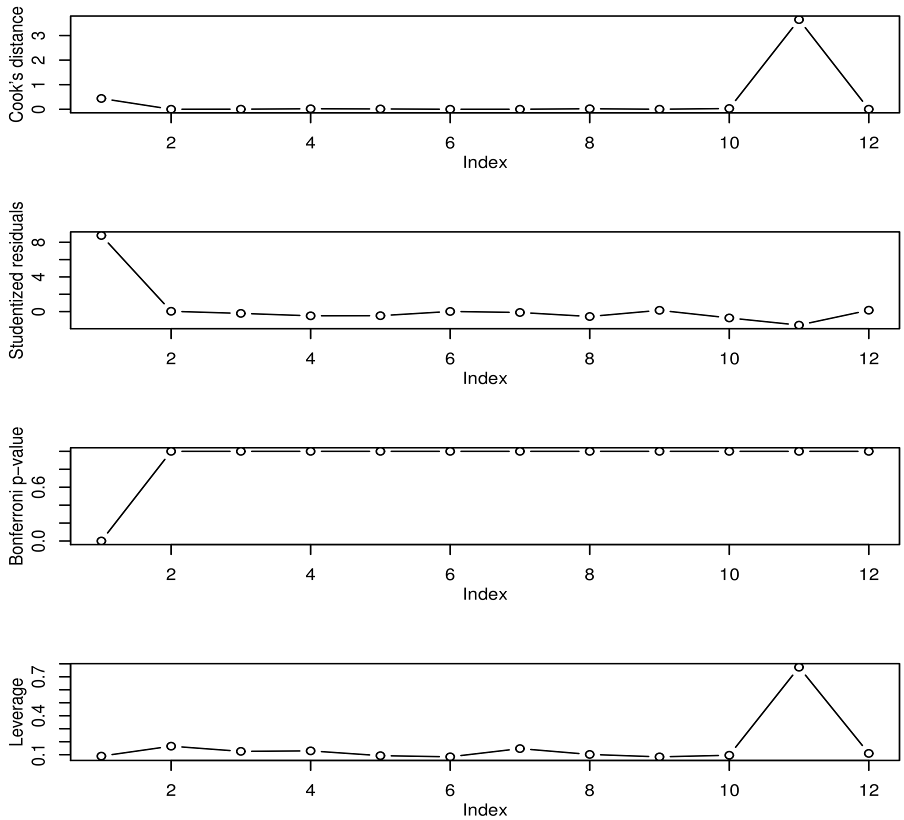
a. Let $\hat{\beta}_1$ be the usual OLS estimate of the slope β_1. Denote by $\hat{\beta}_1^{(i)}$ the OLS estimate of β_1 when the response of the i-th observation is changed from y_i to $y_i + \delta$, and all other observations are left unchanged. Show that

$$\hat{\beta}_1^{(i)} - \hat{\beta}_1 = \delta \, \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad .$$

b. Utilizing the expression in part (a), provide an illustration of the case in which the i-th observation is an outlier but not influential. Provide an illustration of the case in which the i-th observation is influential but not an outlier.

c. The figure on the next page shows diagnostic plots for a set of 12 simulated data points (x_i, y_i), after fitting a simple linear regression model. The plots show, from top to bottom, Cook's distance D_i, the standardized residuals r_i, Bonferroni corrected p-values for those residuals, and the leverage values h_{ii}. Based on these plots, provide a schematic diagram (i.e., a sketch) of what the scatterplot for the data might look like.

d. Let $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ be the OLS estimate of the intercept β_0 with the original data, and let $\hat{\beta}_0^{(i)}$ be the estimate when the i-th response is perturbed in the manner described above in part (a). Find a concise expression for the change in the estimate of β_0 i.e., for $\hat{\beta}_0^{(i)} - \hat{\beta}_0$.


Cook's distance
Studentized residuals
Bonferroni p−value
Leverage
Index

2. Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \ ,$$

for $i = 1, \ldots, n$, where we assume $E[e_i|X] = 0$, $\mathrm{Var}(e_i|X) = \sigma^2$, and the e_i are correlated so that $\mathrm{Cov}(y_i, y_{i+1}) = \rho\sigma^2$, for $i = 1, \ldots, n-1$.

a. Show that the ordinary least squares estimate $\hat{\beta}_1$ of the slope β_1 is unbiased, despite the presence of correlation in the errors. That is, show that

$$E[\hat{\beta}_1|X] = \beta_1 \ .$$

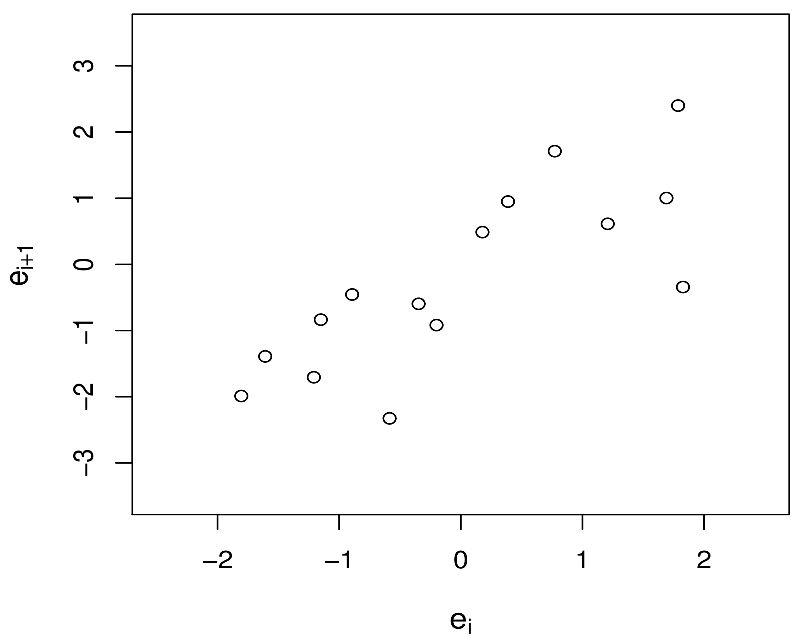
b. Show that the variance of the ordinary least squares estimate $\hat{\beta}_1$ of the slope β_1 can be expressed as

$$\mathrm{Var}(\hat{\beta}_1|X) = \left(\frac{\sigma}{SXX}\right)^2 \left[\sum_{i=1}^{n}(x_i - \bar{x})^2 + 2\rho\sum_{i=1}^{n-1}(x_i - \bar{x})(x_{i+1} - \bar{x})\right] \ .$$

(Hint: Begin by writing $\hat{\beta}_1$ as a linear combination of the y_i's i.e., in the form $\sum_{i=1}^{n} c_i y_i$, for appropriate choice of constants c_i.)

c. The correlation structure in this model is sometimes referred to as a 'lag-one serial correlation', in that the adjacent error terms e_i and e_{i+1} each have correlation ρ. In the presence of positive lag-one serial correlation, a plot of the residual pairs $(\hat{e}_i, \hat{e}_{i+1})$ may look something like that shown in the figure of the next page.

Design a test for lag-one serial correlation in a regression model. That is, assume a model like that described above, and design a test for whether or not ρ is equal to zero. Provide as much detail as you are able, regarding the definition of your test, its computation, and the determination of significance.



3
2
1
0
−1
−2
−3
e_{i+1}
−2
−1
0
1
2
e_i