

Due Wed April 5th, 2006

Dr. Bensmail

Data: C6_EPS

Many investors rely on predictions of a company's performance compiled by professional analysts. Here we examine a company's predicted earnings per share (PRE_EPS) and attempt to explain this in terms of a number of firm and industry variables. The industry variables include the expected industry changes in EPS for the next five years (PRE_IND) and the type of industry (IND). For this data set, four industries are considered: retail (IND=1), manufacturing (IND=2), life and health insurance (IND=3) and service (IND=4). The firm variable is the company change in EPS over the last five years (CL5Y).

- 1. Give the Scatterplot matrix of PRE_EPS, PRE_IND, and CL5Y. Describe the relationship between these variables. In particular, note the linear and nonlinear patterns. Fit the regression model of PRE_EPS on PRE_IND and CL5Y. Describe the adequacy of the model.
- 2. What are the values of the coefficient of determination and adjusted coefficient of determination.
- 3. From the preceding scatterplot matrix we can see that seven companies had particularly large downturns in their earnings per share over the last five years. Because this pattern may represent a nonlinear effect, an indicator variable for these seven firms was created and called "BADCL5Y". This variable was also included in the fitted regression model. Fit the new model and explain the presence of the relatively large VIF statistics.
- 4. Give the histograms of the standardized residual for the model in part (2) by level of the industry variable and discuss about the variability of each.
- 5. To assess whether there is a difference in variability by industry, the standardized residuals are raised to the 2/3 rd power. Then, one can treat the transformed residuals as the dependent variable and use regression models to see if the categorical variable industry is important. The idea here is that by going to the 2/3 rd, we are looking at the variability of the standardized residuals. The square converts the residuals to estimates of variability and 1/3 to symmetrize the distribution. Fit the regression with residuals power 2/3 as dependent variable and IND as dependent variable (One way anova analysis). Describe if the variability differs amongs industries.
- 6. By fitting the model where y is PRE_EPS and x is IND, we can see that the plot of

the fitted values of y and residual has a specific shape, this means that we encountered heteroscedasticity among industry groups. what we should do?

7. Suppose that we want to validate the results of the study:

(i) define the PRESS Statistic.

(ii) you would like to calculate PRESS statistic for two models: one using y as response and **IND** as explanatory variable and the second using \log of y as response and **IND** as explanatory variable. compare the two models.