Questions and Exercises - Tutorial #2

Ettore Gallo

April 23, 2018

Measures Of Fit

1. Given that $y_i = \hat{y}_i + \hat{u}_i$, provide an intuitive explanation of what it is measured by the total sum of squares (SST), the explained sum of squares (SSE) and the residual sum of squares (SSR).

$$SST \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{1}$$

$$SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{2}$$

$$SSR \equiv \sum_{i=1}^{n} {\hat{u}_i}^2 \tag{3}$$

2. Demonstrate that $SST = SSE + SSR$ (*Hint:* start with the definition of SST and sum/subtract $\hat{y}_i$ in the summation operator.)

3. The R^2 measures "how well the explanatory variable or independent variable x explains the dependent variable y" (Wooldridge, section 2.3) and it can be interpreted "as the fraction of the sample variation in y that is explained by x" (*ibid.*). On the basis of equations (1), (2) and (3), how would you intuitively define R^2?

4. Sketch a hypothetical scatterplot of data for an estimated regression with $R^2 = 0.9$. Sketch a hypothetical scatterplot of data for an estimated regression with $R^2 = 0.5$. *(Exercise 4.3 from Stock and Watson)*

5. How does $\bar{R}^2$ differ from R^2? Why is R^2 useful in a regression model with multiple regressors?

6. Demonstrate that under MLR.1 through MLR.5, the standard error of the regression ($\hat{\sigma}^2$ or SER) is an unbiased estimator of the error variance $Var(u)$, i.e. demonstrate that $E(\hat{\sigma}^2) = \sigma^2$.

Biases in Multiple Regression Analysis

1. Consider the multiple regression model containing three independent variables, under Assumptions MLR.1 through MLR.4:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \tag{4}$$

2. Which of the following can cause OLS estimators to be biased?

   - Heteroskedasticity.
   - Omitting an important variable.
   - A sample correlation coefficient of .95 between two independent variables both included in the model.

3. A researcher is interested in the effect on test scores of computer usage. Using school district data, she regresses district average test scores on the number of computers per student. Will β_1 be an unbiased estimator of the effect on test scores of increasing the number of computers per student? Why or why not? If you think β_1 is biased, is it biased up or down? Why? *(Exercise 6.1, Review the Concepts, Stock and Watson)*

4. Explain why two perfectly multicollinear regressors cannot be included in a linear multiple regression. Give two examples of a pair of perfectly multicollinear regressors. *(Exercise 6.3, Review the Concepts, Stock and Watson)*