Econometrics Problem Sets
William M. Boal

Problem Set 8
"Two-Variable Regression: Algebraic Properties"

(8.1) [Regression without an intercept] Consider the two-variable model *without* an intercept, that is $y = \beta x + \varepsilon$. Notice that there is only one β parameter for this model—the slope. There is no intercept because the true population regression function is here assumed to pass through the origin. The least-squares estimator for this model minimizes the following function: $f(\beta) = \sum (y_i - \beta x_i)^2$ where the sum runs from i=1,…,n.

a. Show that the first-order necessary condition (FONC) for a minimum implies $0 = \sum x_i (y_i - \beta x_i)$. Justify each step of your proof.

b. Show that this FONC implies $\hat{\beta} = \dfrac{\sum x_i y_i}{\sum x_i^2}$. Justify each step of your proof.

c. Also show that this FONC implies $\sum \hat{\varepsilon}_i \, x_i = 0$, where the least-squares residuals are defined as $\hat{\varepsilon}_i = y_i - \hat{\beta} x_i$. Justify each step of your proof.

(8.2) [Regression without a slope] Consider the two-variable model *without* a slope, that is $y = \beta + \varepsilon$. Notice that there is only one β parameter for this model—the intercept. There is no slope because the true population regression function is here assumed to be horizontal. The least-squares estimator for this model minimizes the following function: $f(\beta) = \sum (y_i - \beta)^2$ where the sum runs from i=1,…,n.

a. Show that the first-order necessary condition (FONC) for a minimum implies $0 = \sum (y_i - \beta)$. Justify each step of your proof.

b. Show that this FONC implies $\hat{\beta} = \dfrac{1}{n} \sum y_i$. Justify each step of your proof.

c. Also show that this FONC implies $\sum \hat{\varepsilon}_i = 0$, where the least-squares residuals are defined as $\hat{\varepsilon}_i = y_i - \hat{\beta}$. Justify each step of your proof.

(8.3) [Algebraic properties of least-squares] Suppose we estimate the two-variable regression model $y = \beta_1 + \beta_2 x + \varepsilon$, using ordinary least-squares (LS). Denote the LS fitted values $\hat{y}_i$ and denote the LS residuals $\hat{\varepsilon}_i$ for i = 1,…, n. Which of the following expressions must necessarily be zero? Which of the following are generally nonzero?

a. $\sum x_i$
b. $\sum (x_i - \bar{x})$
c. $\sum (x_i - \bar{x})^2$
d. $\sum (\hat{y}_i - \bar{y})$
e. $\sum \hat{\varepsilon}_i$
f. $\sum \hat{\varepsilon}_i^2$
g. $\sum x_i \hat{\varepsilon}_i$
h. $\sum \hat{y}_i \hat{\varepsilon}_i$
i. $\sum x_i \hat{y}_i$

(8.4) [Algebraic properties of least squares] A certain researcher has analyzed state-level data on college attendance and the price of cheese. This researcher has fitted a two-variable regression model $y = \beta_1 + \beta_2 x + \varepsilon$, where y denotes the fraction of persons in the state between the ages of 18 and 22 who are attending college, and x denotes the average price of American cheese in that state. "I find that the sum of the least-squares residuals is zero," claims the researcher excitedly, "and the sample correlation between my residuals and the price of cheese is zero! This proves that the relationship is strong and that my model is a good one!" Do you agree or disagree? Explain your reasoning.

(8.5) [Algebraic properties of least squares] Suppose a model has been estimated by two-variable least-squares. As usual, the least-squares residuals are defined as $\hat{\varepsilon}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$. Let a bar ($\bar{}$) denote the sample mean. Then the sample correlation between the least-squares residuals and the regressor x can be defined as

$$\textit{sample corr} = \frac{\frac{1}{n}\sum(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})(x_i - \bar{x})}{\frac{1}{n}\sum(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2 \ \frac{1}{n}\sum(x_i - \bar{x})^2}$$

a. What is the value of $\bar{\hat{\varepsilon}}$? Why?
b. Use the algebraic properties of least squares to prove that *sample corr* must necessarily equal zero. Justify each step of your proof. [Hint: First, eliminate $\bar{\hat{\varepsilon}}$ from numerator. Then split the numerator into two separate sums. Use the algebraic properties to show that each of these sums equals zero.]

(8.6) [Sum-of-squares decomposition and r-square] A model has been estimated by two-variable least-squares, but not all the information is readable due to a printing error. We can read only the sum of squared residuals and the total sum of squares:

$\sum \hat{\varepsilon}_i^2 = 78$ and $\sum (y_i - \bar{y})^2 = 120$.

a. Compute the "explained sum of squares" $\sum (\hat{y}_i - \bar{y})^2$.
b. Compute the r^2 value.

(8.7) [Sum-of-squares decomposition and r-square] A model has been estimated by two-variable least-squares, but not all the information is readable due to a printing error. We can read only the total sum of squares and the explained sum of squares:

$$\sum (y_i - \bar{y})^2 = 75 \qquad \text{and} \qquad \sum (\hat{y}_i - \bar{y})^2 = 24 .$$

a. Compute the "sum of squared residuals" $\sum \hat{\varepsilon}_i^2$.
b. Compute the r^2 value.

(8.8) [Least-squares principle] Suppose the line $y_i = \beta_1 + \beta_2 y_i$ were estimated by least squares by mistake. Note that y_i appears on both sides of the equation.

a. What would be the least-squares estimates of β_1 and β_2 ? Why?
b. What would be the sum of squared residuals? Why?
c. What would be the value of r^2 ? Why?

[end of problem set]