Erasmus School of Economics

MOOC Econometrics

Test Exercise 2

Notes:

- See website for how to submit your answers and how feedback is organized.
- For parts (e) and (f), you need regression results discussed in Lectures 2.1 and 2.5.

Goals and skills being used:

- Use matrix methods in the econometric analysis of multiple regression.
- Employ matrices and statical methods in multiple regression analysis.
- Give numerical verification of mathematical results.

Questions

This test exercise is of a theoretical nature. In our discussion of the F-test, the total set of explanatory factors was split in two parts. The factors in X_1 are always included in the model, whereas those in X_2 are possibly removed. In questions (a), (b), and (c) you derive relations between the two OLS estimates of the effects of X_1 on y, one in the large model and the other in the small model. In parts (d), (e), and (f), you check the relation of question (c) numerically for the wage data of our lectures.

We use the notation of Lecture 2.4.2 and assume that the standard regression assumptions A1-A6 are satisfied for the unrestricted model. The restricted model is obtained by deleting the set of g explanatory factors collected in the last g columns X_2 of X. We wrote the model with $X = (X_1\ X_2)$ and corresponding partitioning of the OLS estimator b in b_1 and b_2 as $y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X_1 b_1 + X_2 b_2 + e$. We denote by b_R the OLS estimator of β_1 obtained by regressing y on X_1, so that $b_R = (X_1'X_1)^{-1}X_1'y$. Further, let $P = (X_1'X_1)^{-1}X_1'X_2$.

(a) Prove that $E(b_R) = \beta_1 + P\beta_2$.

(b) Prove that $\text{var}(b_R) = \sigma^2(X_1'X_1)^{-1}$.

(c) Prove that $b_R = b_1 + Pb_2$.

Now consider the wage data of Lectures 2.1 and 2.5. Let y be log-wage (500×1 vector), and let X_1 be the (500×2) matrix for the constant term and the variable 'Female'. Further let X_2 be the (500×3) matrix with observations of the variables 'Age', 'Educ' and 'Parttime'. The values of b_R were given in Lecture 2.1, and those of b in Lecture 2.5.

(d) Argue that the columns of the (2×3) matrix P are obtained by regressing each of the variables 'Age', 'Educ', and 'Parttime' on a constant term and the variable 'Female'.

(e) Determine the values of P from the results in Lecture 2.1.

(f) Check the numerical validity of the result in part (c). Note: This equation will not hold exactly because the coefficients have been rounded to two or three decimals; preciser results would have been obtained for higher precision coefficients.