Computational Econometrics with R

Further Issues in Multiple Regression Analysis: Wage Equation

Labor economists and policy makers are interested in the returns to education and work experience. We use data on monthly earnings, education, experience, several demographic variables and IQ scores for 935 men in the US in 1980 to approach the topic with OLS. The following variables are available in `wages2.csv`:

Variable	Explanation
WAGE	monthly earnings, in dollar
HOURS	average weekly hours of work
IQ	IQ score
SCORES	scores achieved in a test of work-related abilities
EDUC	years of education
EXPER	years of work experience
TENURE	years with current employer
AGE	age in years
MARRIED	1 if married, 0 otherwise
BLACK	1 if black, 0 otherwise
SOUTH	1 if living in south, 0 otherwise
URBAN	1 if living in a city, 0 otherwise

1. Preliminaries

1.1 Import the file `wages2.csv` to R and label the variables appropriately.

1.2 Plot a histogram of the variable WAGE and provide meaningful axis labels and titles.

1.3 Estimate the regression model

$$\text{WAGE}_i = \beta_0 + \beta_1 \text{EDUC}_i + \beta_2 \text{EXPER}_i + \varepsilon_i \tag{1}$$

by OLS and interpret your results.

2. Effects of Data Scaling on OLS Statistics

2.1 Generate a series WAGETH which measures the variable WAGE in thousands of dollars rather than in dollars.

2.2 Estimate the regression model

$$\text{WAGETH}_i = \beta_0 + \beta_1 \text{EDUC}_i + \beta_2 \text{EXPER}_i + \varepsilon_i \tag{2}$$

by OLS. What happens to the estimates of β_0 through β_2 in comparison to those of equation (1)? Write down a general rule for what happens to the parameter estimates when you change the units of measurement of the dependent variable.

2.3 Generate the series EDUCM which measures the variable EDUC in month rather than in years, i.e. EDUCM = EDUC $\cdot$ 12.

2.4 Estimate the regression model

$$\mathrm{WAGE}_i = \beta_0 + \beta_1 \mathrm{EDUCM}_i + \beta_2 \mathrm{EXPER}_i + \varepsilon_i \tag{3}$$

by OLS. What happens to the estimates of β_0 through β_2 in comparison to those of equation (1)? Write down a general rule for what happens to the parameter estimates when you change the units of measurement of an independent variable.

2.5 To figure out which of the independent variables has the greatest effect on the dependent variable, it is useful to obtain regression results when all variables involved have been standardized to z-scores (by subtracting off the mean and by dividing by the standard deviation of the respective variable) as in

$$\mathrm{WAGE\_Z}_i = \beta_1 \mathrm{EDUC\_Z}_i + \beta_2 \mathrm{EXPER\_Z}_i + \varepsilon_i. \tag{4}$$

Estimate equation (4) by OLS and interpret the results.

3. Functional Form and Choice of Additional Regressors

3.1 So far, we have focused on linear relationships between WAGE and the independent variables EDUC and EXPER. An alternative characterization of how the wage changes with education and experience is that each year of education or experience increases the wage by a constant percentage as in

$$\ln(\mathrm{WAGE})_i = \beta_0 + \beta_1 \mathrm{EDUC}_i + \beta_2 \mathrm{EXPER}_i + \varepsilon_i. \tag{5}$$

Estimate equation (5) by OLS and interpret your results.

3.2 Estimate a sixth regression model

$$\ln(\mathrm{WAGE})_i = \beta_0 + \beta_1 \mathrm{EDUC}_i + \beta_2 \mathrm{EXPER}_i + \beta_3 \mathrm{TENURE}_i + \beta_4 \mathrm{IQ} + \varepsilon_i \tag{6}$$

by OLS. Interpret your results.

3.3 Test the two hypothesis (seperately) that the additional regressors TENURE and IQ have no effect on ln(WAGE), i.e. $\mathrm{H}_0 : \beta_3 = 0$ and $\mathrm{H}_0 : \beta_4 = 0$, by using the t-test and interpret your result.

3.4 Compare the appropriate coefficients of determination (R-squared) of model specifications (5) and (6). Which model performs better in terms of the goodness of fit?