Handout 4: Coefficient Interpretation and Statistical Significance

1 Coefficient Interpretation

1.1 Level-Level

Suppose the dependent variable and independent variable of interest are both in level form. For example, suppose we want to know the effect of an additional year of schooling on wages and from available data we fit the following equation

$$W_i = \alpha + \beta_1 Ed_i + \beta_2 Age_i + \beta_3 Exp_i + \epsilon_i \tag{1}$$

In this case, the coefficient on the variable of interest can be interpreted as the marginal effect. The **marginal effect** is how the dependent variable changes when the independent variable changes by an additional unit holding all other variables in the equation constant (i.e. partial derivative) or

$$\frac{\partial W_i}{\partial Ed_i} = \beta_1$$

Therefore, β_j can be interpreted as the change in wages from a one unit increase (or state change if dummy variable) of X_j holding all other independent variables constant.

Example: Suppose the fitted equation for (1) is

$$\hat{W}_i = 3.5 + 0.75 Ed_i + 0.25 Age_i + 0.30 Exp_i + \epsilon_i$$

Based on the data used in this regression, an additional year of education corresponds to an increase in hourly wages of \$0.75. Similarly, an additional year of experience is associated with a \$0.30 per hour wage increase.

1.2 Log-Log

Consider interpreting coefficients from a regression where the dependent and independent variable of interest are in log form. The coefficients can no longer be interpreted as marginal effects. Suppose economic theory suggests estimation of our wage equation with the dependent variable in log form and inclusion of community volunteer hours per week (Comm) also in log form. The equation of interest is now

$$\log(W_i) = \alpha + \beta_1 Ed_i + \beta_2 Age_i + \beta_3 Exp_i + \beta_4 \log(Comm_i) + \epsilon_i \tag{2}$$

We would like to interpret the coefficient on the community volunteer variable (β_4). To better understand the interpretation, consider taking the differential of (2) holding all independent variables constant except $Comm_i$.

$$d[\log(W_i)] = d[\log(Comm_i)]\beta_4$$

$$\frac{1}{W_i} dW_i = \frac{1}{Comm_i} dComm_i \beta_4$$

since $d[log(X)] = \frac{1}{X}d(X)$. This final equation can be rearranged such that,

$$\frac{\frac{100*dW_i}{W_i}}{\frac{100*dComm_i}{Comm_i}} = \beta_4$$

where the left hand side is the (partial) elasticity of W with respect to $Comm$. **Elasticity** is the ratio of the percent change in one variable to the percent change in another variable. The coefficient in a regression is a *partial* elasticity since all other variables in the equation are held constant. Therefore, β_4 can be interpreted as the percent change in hourly wages from a one percent increase in community volunteer hours per week holding education, age and experience constant.

Example   Suppose that the fitted equation for (2) is

$$\log(\hat{W_i}) = 3.26 + 0.24Ed_i + 0.08Age_i + 0.16Exp_i + 1.2\log(Comm_i)$$

Based on these regression results, a one percent increase in community volunteer hours per week is associated with a 1.2% increase in hourly wages.

1.3 Log-Level

In equation (2), education, age and experience are in level terms while the dependent variable (wage) is in log terms. We would like to interpret the coefficients on these variables. First, consider education. Take the differential holding all other independent variables constant.

$$d[\log W_i] = dEd_i\beta_1$$
$$\frac{dW_i}{W_i} = dEd_i\beta_1$$

Multiply both sides by 100 and rearrange,

$$\frac{100 * dW_i}{W_i} = 100 * dEd_i\beta_1$$
$$100 * \beta_1 = \frac{\frac{100 * dW_i}{W_i}}{dEd_i} = \frac{\%\Delta W_i}{\text{unit } \Delta Ed_i}$$

Therefore, $100 * \beta_1$ can be interpreted as the percentage change in W_i for a unit increase in Ed_i, holding all other independent variables constant. Similar derivations can derive the interpretation for the coefficients on age and experience.

Example: Consider the fitted equation for (2)

$$\log(\hat{W_i}) = 3.26 + 0.24Ed_i + 0.08Age_i + 0.16Exp_i + 1.2\log(Comm_i)$$

Therefore, holding all other independent variables constant, an additional year of schooling is associated with a 24% increase in hourly wages. Similarly, an additional year of experience is associated with a 16% increase in hourly wages.

1.4 Level-Log

Consider a regression where the dependent variable is in level terms and the independent variable of interest is in log terms. For example, consider the following equation

$$W_i = \alpha + \beta_1 Ed_i + \beta_2 Age_i + \beta_3 Exp_i + \beta_4 \log(Comm_i) + \epsilon_i \tag{3}$$

Recall from the section on level-level regressions that the coefficients on education, age and experience can be interpreted as marginal effects. We would like to interpret the coefficient on community

volunteer hours (β_4). Again, take the differential on both sides, holding all independent variables constant except community volunteer hours:

$$dW_i = d[log(Comm_i)]\beta_4$$
$$dW_i = \frac{1}{Comm_i}dComm_i\beta_4$$

Divide both sides by 100 and rearrange,

$$\frac{\beta_4}{100} = \frac{dW_i}{\frac{100*dComm_i}{Comm_i}} = \frac{\text{unit } \Delta W_i}{\%\Delta Comm_i}$$

Therefore, $\frac{\beta_4}{100}$ can be interpreted as the increase in hourly wages from a one percent increase in community volunteer hours per week.

Example: Suppose that the fitted equation for (3) is

$$\hat{W}_i = 3 + 0.67Ed_i + 0.28Age_i + 0.34Exp_i + 13.2\log(Comm_i)$$

Therefore, holding education, age and experience constant, a one percent increase in community volunteer hours per week is associated with a $0.132 increase in hourly wages.

2 Statistical Significance

Estimated parameter values are only estimates of the "true" relationship and there exists uncertainty associated with these estimates. Consider the following fitted equation:

$$\hat{W}_i = 3 + 0.50Ed_i$$
$$(1.26)(0.09)$$

The uncertainty about each estimated parameter is measured by its **standard error** or the estimated standard deviation of the coefficient. In the example above, standard errors are reported in the parentheses below the parameter estimate. The larger the standard error, the greater the uncertainty about the estimated parameter value.

Standard errors can be useful to test hypotheses about the estimated parameter values. For example, we can test the hypothesis that the slope coefficient is positive against the hypothesis that the coefficient is zero (i.e. no relationship between education and wages). A convenient way to test whether a parameter estimate is different from zero is to calculate the t-statistic. The **t-statistic** is the absolute value of the parameter estimate divided by the standard error of the estimate. The rule of thumb is if the t-statistic is greater than 2, we can reject the hypothesis that the parameter estimate is equal to zero. In other words, if the absolute value of a parameter estimate is at least twice as large as its standard error, one can be fairly confident that the true value of the coefficient is different than zero. When this holds, we say that the estimated parameter is statistically different than zero or **statistically significant**. In the example above, the t-statistic for $\hat{\beta}$ is $0.5/0.09 = 5.4 > 2$. Therefore, we are very confident that the true relationship between education and wages is positive.