IIA-3 Econometrics: Additional Notes

Emre Usenmez

Easter Break 2025

Notes are based on the following:

Hardy, M A (1993), *Regression with Dummy Variables*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-093, Newbury Park, CA: Sage

Halvorsen, R, and Palmquist, R (1980) *Interpretation of dummy variables in semilogarithmic equations*, American Economic Review 70:474-475.

Interpretation of Regression with Dummy Variables

We will use income difference as an example.

Let's begin the discussion with descriptive statistics.

Descriptive Statistics

Given that dummy variables provide qualitative information, the mena value of a dummy variable indicate the proportion, or relative frequency, of cases in the category coded as 1. This is because the mean value is calculated by adding up all the cases and dividing it by the number of cases. Since all the cases are either 1 or 0, this ends up being the same as the relative frequency. Therefore, for dummy variables, the formulas for the mean and for a proportion are equivalent.

Similarly, the formula for the variance of a dummy variable can be related to the more general variance formula for continous measures. When X_i is continuous:

$$Var(X_i) = \frac{\sum X_i^2}{N} - \Big(\frac{\sum X_i}{N}\Big)^2 \tag{1}$$

and when X_i is a dummy variable, $\sum X_i^2$ reduces to n_j, the number of cases coded as 1. Thus, if we denote the proportion of cases coded as 1 as $p_j = \frac{n_j}{N}$ then the variance for a dummy variable becomes:

$$Var(X_i) = p_j - p_j^2 = p_j(1 - p_j). \tag{2}$$

That is, the variance for a dummy variable is the product of the proportion of cases coded as 1 and its complement. This means, the maximum variability in a dummy variable is obtained when the number of cases are evenly split between 1 and 0.

Dummy Explanatory Variables

Suppose we are interested in looking at potential income differences between left and right handed people.

Consider the model

$$\text{Model 1: } Y_i = f(\text{handedness}) = \beta_0 + \beta_1 \text{ left}_i + u_i.$$

When the independent variable is continuous, the distribution of predicted values of the dependent variable are also continuous and so the regression coefficient indicates a slope. On the other hand when it is a dummy variable, the predicted value of the dependent variable changes by the estimated coefficient each time membership in a specified category is switched on or off.

Using some data we obtained from a survey of 3,211 respodents (2,290 of which are right handed), suppose we estimate the following relationship:

$$\begin{aligned} Y_i &= 78,210.9 - 32,020.9 \text{ left} \\ se &: (910.9) \qquad (1,710.6) \\ R_1^2 &: 0.09792 \\ F_1 &: 348.3 \end{aligned}$$

This means predicted income for respondents who are left handed is £46,190.00 which is £32,020.90 less than predicted income of £78,210.9 for right handed people. These are also the same as the mean values of each group. That is, if we average the income for all left handed people in the survey, we would get £46,190.00, and averaging all right handed people would yield £78,210.90.

The coefficient of the independent variable measures the effect of being left handed on income. Since that is the case, the standard error of this coefficient is then the standard error of the difference between expected income for right handed people and expected income for left handed people.

When testing against a null hypothesis that there are no difference in expected income between the two groups, i.e. zero effect, the t test reduced to the ratio of the coefficient of the standard error. Since this model has only one independent variable, the F test for the model is a test of the same null hypothesis of zero effect. The value of F is the square of t value. This means, the null hypothesis that $\beta_1 = 0$ is equivalent to the null hypothesis $\mathbb{H}_0 : \mu_{\text{left}} - \mu_{\text{right}} = 0$. Both t test for $\hat{\beta}_1$ and the F test for the model itself are essentially difference of means tests.

Suppose now we instead want to estimate income as a function not of handedness but occupational categories: upper white-collar jobs (`upwc`), lower white-collar (`lowwc`), skilled craftsmen (`skill`), operatives such as welders, stitchers in manufacturing etc. (`oper`), service workers such as barbers, janitors (`serv`), and laborers (`labor`):

$$\text{Model 2: } Y_i = f(\text{handedness, occupation}) = \beta_0 + \beta_1 \text{ lowwc}_i + \beta_2 \text{ skill}_i + \beta_3 \text{ oper}_i + \beta_4 \text{ serv}_i + \beta_5 \text{ labor}_i + u_i.$$

Notice that we are using upper white-collar workers as the reference group and regress without it to avoid perfect multicollinearity. Suppose again that we use data from a survey with 3211 sample points and we now estimate the following relationship:

$$\begin{aligned} Y_i &= 107,020.1 - 30,210.2 \text{ lowwc} - 37,570.1 \text{ skill} - 51,480.2 \text{ oper} - 62,670.7 \text{ serv} - 66,120.1 \text{ labor} \\ se &: (1,600.8) \quad (2,740.4) \quad (2,150.5) \quad (2,160.8) \quad (2,890.7) \quad (2,720.3) \\ R_2^2 &: 0.224 \\ F_2 &: 185.0 \end{aligned}$$

The constant reports the expected income for the reference group, upper white-collar workers, to be £107,020.1. The remaining coefficients report the effect of being in a particular occupational category compared with the reference category. So the coefficient $\hat{\beta}_1$ indicates that on average lower white-collar workers earn £30,210.20

less than upper white-collar workers, i.e. they earn on average £76,800.90. Similarly, laborers earn £66,120.10 less than upper white-collar workers on average, or £40,900.00.

In terms of testing for the effect of occupational category on income, notice that distinctions among occupational groups are captured by the entire set of dummy variables and not by any single variable. Therefore, F test would be appropriate here with the null hypothesis that $\beta_1 = \cdots = \beta_5 = 0$, meaning that F test is a test of the hypothesis that the expected value of income (Y_i) for all occupational groups is the same.

Also notice that F test can be a test of significance of R^2 here. This is because the F-test can be expressed as the ratio of R^2/k to $(1 - R^2)/(N - k - 1)$. Therefore, if null hypothesis is rejected then a nonzero amount of variation in income is explained by the respondent's occupational category. In this example,

$$F_{5,3205} = \frac{0.224/5}{(1 - 0.224)/(3211 - 5 - 1)} = 185$$

which is significant at better than .001 level. Thus occupation is significant.

Since occupation is significant, we can now test if the expected income for each occupational category is significantly different from that of the reference group using t test. That is, with t test on the coefficients we are checking if the effect of being in the designated category rather than in the reference group is significant.

One other thing to check here is whether the occupational categories are different from each other. That is, we need to check if, for example, the expected income for laborers ($\hat{\beta}_6$) is different, or indeed smaller, than the expected income for operatives ($\hat{\beta}_4$) for example. To test this, notice that $\beta_j = \mathbb{E}(Y_i \mid DUMMY_j = 1) - \mathbb{E}(Y_i \mid REFERENCE)$. Because of this, the difference in expected income for included categories is equal to the difference between their coefficients (i.e. $\beta_j - \beta_k$). So, for example to test for a difference in the effects of laborer and operatives, we'd use a t test for the difference in regression coefficients:

$$t = \frac{\beta_j - \beta_k}{\sqrt{Var(\beta_j) + Var(\beta_k) + 2Cov(\beta_j \beta_k)}}.$$

We can do similar operations for other comparisons.

We can put together these two qualitative measures, handedness and occupation, to see if handedness differences in income persists when we control for income differences in occupation. The model then becomes:

$$\begin{aligned} \text{Model 3: } Y_i &= f(\text{handedness, occupation}) \\ &= \beta_0 + \beta_1 \text{ left}_i + \beta_2 \text{ lowwc}_i + \beta_3 \text{ skill}_i + \beta_4 \text{ oper}_i + \beta_5 \text{ serv}_i + \beta_6 \text{ labor}_i + u_i. \end{aligned}$$

where the upper white-collar is still the reference occupation category.

Once more suppose that we use data from a survey with 3211 respondents. Our estimation now becomes:

$$\begin{aligned} Y_i &= 108,110.4 - 16,760.0 \text{ left} - 28,420.1 \text{ lowwc} - 35,660.4 \text{ skill} - 46,040.5 \text{ oper} \\ se &: (1,580.9) \quad (1,720.4) \quad (2,710.1) \quad (2,130.3) \quad (2,200.9) \\ &\quad - 55,120.7 \text{ serv} - 56,470.8 \text{ labor} \\ &\quad (2,950.9) \quad (2,860.2) \\ R_3^2 &: 0.24624 \\ F_3 &: 174.4 \end{aligned}$$

The constant indicates an expected income when all independent variables are set to zero, so it tells us that the expected income for right handed upper white-collar workers is £108,110.40.

The coefficient for `left` indicates that once the variation in income linked to occupational category is taken into account along with the fact that handedness is not uniformly distributed across all occupational categories, left handed people still average £16,760.00 less income than right handed people. This is a reduction from the difference in earnings of £32,020.90 when handedness was the only explanatory variable. The reduction in

the magnitude of the coefficient of `left` suggests that one reason left handed workers average lower incomes is because they are concentrated in occupations that in general commands lower salaries or earnings.

Similarly, the partial regression coefficients associated with occupation dummy variables estimate the effect on expected income of membership in each of the designated categories rather than the reference upper white-collar group, controlling for handedness differences in both income and the distribution of respondents across occupational categories.

To decide whether the partial effects of handedness or the partial effects of occupation, while controlling for other variables, are statistically significant, an F test would be appropriate. However, we would not use the F test for the model as a whole but instead use the incremental F test. Suppose we want to assess whether occupational categories contribute to this model. For this we can compare the R^2s of this third model, and the first model where `left` was the only explanatory variable. The null hypothesis is that once handedness differences are controlled for in both income and occupational category, the expected value of earned income is the same across occupational categories; i.e. $\beta_2 = \cdots = \beta_6 = 0$. The test for the significance of occupation controlling for handedness is:

$$F_{5,3204} = \frac{\dfrac{R_3^2 - R_1^2}{k_3 - k_1}}{\dfrac{1 - R_3^2}{N - k_3 - 1}} = \frac{\dfrac{0.24624 - 0.09792}{6 - 1}}{\dfrac{1 - 0.24624}{3211 - 6 - 1}} = 126.1$$

Here, the numerator calculates the increments to R^2 that results from specifying the effects of occupational category relative to the difference in the number of independent variables between the two models.

The denominator is the proportion of variance left unexplained when both handedness and occupation are included divided by the degrees of freedom.

Notice that this last model estimates 12 different values for the predicted income. This is because handedness has 2 categories and occupation has 6 categories, together they generate 12 distinct subgroups. With one mean per each handedness-by-occupation subgroups, these predicted values correspond to 12 subgroup means. However we are making a simplifying assumption that the estimated income difference between left handed people and right handed ones (i.e. the effect of `left`) is the same across all occupational groups. In other words, we assume that the income differences across occupational groups are the same for left handed and right handed people. In our model so far, the difference between left handed and right handed workers is always $\hat{\beta}_1 = 16,760$ regardless of occupation. So for example a left handed skilled worker is expected to earn $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 =$ £55,690.00 while a right handed skilled worker is expected to earn $\hat{\beta}_0 + \hat{\beta}_3 =$ £72,450.00 on average. The difference is $\hat{\beta}_1 =$ £16,760.

We will tackle this later on but first lets see what happens when we add some quantitative variables to our model.

Dummy and Quantitative Explanatory Variables

Suppose now income is a function of not only handedness and occupation but also education and employment duration which are measured in years:

$$\begin{aligned}\text{Model 4: } Y_i &= f(\text{handedness, occupation, education, tenure})\\ &= \beta_0 + \beta_1\ \text{left}_i + \beta_2\ \text{lowwc}_i + \beta_3\ \text{skill}_i + \beta_4\ \text{oper}_i + \beta_5\ \text{serv}_i + \beta_6\ \text{labor}_i + \beta_7\ \text{educ}_i + \beta_8\ \text{dur}_i + u_i.\end{aligned}$$

Suppose again that after estimating using data from a survey with 3211 respondents, our estimation now becomes:

$$\begin{aligned}Y_i = 57,610.1 &\ - 11,880.1\ \text{left} \ - 23,160.1\ \text{lowwc} - 23,430.7\ \text{skill} - 31,660.6\ \text{oper}\\ se: (3,590.0) &\quad (1,690.4) \qquad (2,610.8) \qquad (2,230.7) \qquad (2,370.5)\\ &- 39,180.5\ \text{serv} - 36,060.8\ \text{labor} \ + 2,820.0\ \text{educ} \ + 840.7\ \text{dur}\end{aligned}$$

$$(2,990.9) \quad (3,060.4) \quad (230.1) \quad (60.6)$$

$R_4^2 : 0.31459$

$F_4 : 183.7$

The intercept is now the expected income for right handed upper white-collar workers who have 0 years of schooling and 0 years of working in their occupation. Once the variation in income due to occupation, education, and duration of employment is partialled out, the expected income for left handed and right handed people differ by £11,880.10. Coefficients for each occupational dummy variables estimate the net difference in expected income for each occupational group relative to the reference group. So, for example, skilled workers earn £23,430.70 less than upper white-collar workers on average. Similarly, with handedness, occupation, and education held constant, each additional year on the job translates into another £840.70 in earnings. Under similar conditions, an additional year of education is associated with an increase of £282.00 in expected income.

Assessing Group Differences

The models thus far share an assumption that the effect of any single explanatory variable is the same across the range of other explanatory variables. That is, as highlighted earlier, there is an assumption that the effect of `left` is the same across all occupational groups. To test the validity of this assumption we can introduce an interaction term. In order to test for interaction effects, we add five interaction terms to the model that multiplies `left` with each of the occupational dummy variables:

$$\begin{aligned}\text{Model 5: } Y_i &= f(\text{handedness, occupation, education, tenure}) \\ &= \beta_0 + \beta_1\ \text{left}_i + \beta_2\ \text{lowwc}_i + \beta_3\ \text{skill}_i + \beta_4\ \text{oper}_i + \beta_5\ \text{serv}_i + \beta_6\ \text{labor}_i + \beta_7\ \text{educ}_i + \beta_8\ \text{dur}_i \\ &\quad + \beta_9\ \text{leftlow}_i + \beta_{10}\ \text{leftskill}_i + \beta_{11}\ \text{leftop}_i + \beta_{12}\ \text{leftserv}_i + \beta_{13}\ \text{leftlab}_i + u_i.\end{aligned}$$

Here, the interaction term `leftop` for example, would be coded as 1 if the respondent to the survey is both left handed and an operative such as a welder, stitcher in manufacturing, etc.

Now suppose our estimation becomes:

$$\begin{aligned}Y_i = 57,940.8 &- 37,930.3\ \text{left} &&- 22,740.9\ \text{lowwc} &&- 24,180.4\ \text{skill} &&- 34,270.2\ \text{oper} \\ se : (3,580.7) &\quad (6,100.1) &&\quad (2,800.2) &&\quad (2,320.7) &&\quad (2,560.3) \\ &- 45,130.4\ \text{serv} &&- 42,020.8\ \text{labor} &&+ 2,920.9\ \text{educ} &&+ 840.0\ \text{dur} \\ &\quad (3,720.5) &&\quad (3,990.0) &&\quad (230.1) &&\quad (60.6) \\ &+ 15,010.2\ \text{leftlow} &&+ 23,260.2\ \text{leftskill} &&+ 29,840.8\ \text{leftop} &&+ 35,280.0\ \text{leftserv} \\ &\quad (8,230.0) &&\quad (7,050.0) &&\quad (6,720.5) &&\quad (7,610.0) \\ &+ 33,830.9\ \text{leftlab} \\ &\quad (7,470.3)\end{aligned}$$

$R_5^2 : 0.32138$

$F_5 : 116.46$

First, we would like to check if our approach of allowing for differential effects of handedness and occupation resulted in statistically significant improvement in model's fit. Just as we did earlier when we assessed whether the partial effects of occupation improved the model, we use incremental F-test where we compare the R^2s of this model and the model without the interaction terms. The null hypothesis is that once all

the other variables are controlled for, the expected value of earned income is the same across all interaction terms; i.e. $\beta_9 = \cdots = \beta_{13} = 0$. The test therefore is:

$$F_{5,3197} = \frac{\frac{R_5^2 - R_4^2}{k_5 - k_4}}{\frac{1 - R_5^2}{N - k_5 - 1}} = \frac{\frac{0.32138 - 0.31459}{13 - 8}}{\frac{1 - 0.32138}{3211 - 13 - 1}} = 6.4$$

which is statistically significant at better than 0.001 level. Although the increment to explanatory power is far from overwhelming, the F-test suggests that the large sample size has enabled us to estimate the differential effects with reasonable accuracy.

Interpretation:

- Intercept still has similar interpretation as Model 4 whereby the expected income for right handed upper white-collar workers who have 0 years of education and 0 years of working in their occupation is £57,940.80.
- In order to have a better understanding of the coefficients of the occupational groups - and to simplify things a bit - suppose the coefficients for 'educ' and 'dur' are 0. Let's map out the 12 handedness-by-occupation subgroups as follows:

	Right	Left
Upper white-collar	$\hat{\beta}_0$	$\hat{\beta}_0 + \hat{\beta}_1$
Lower white-collar	$\hat{\beta}_0 + \hat{\beta}_2$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_9$
Skilled	$\hat{\beta}_0 + \hat{\beta}_3$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 + \hat{\beta}_{10}$
Operative	$\hat{\beta}_0 + \hat{\beta}_3$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_4 + \hat{\beta}_{11}$
Service	$\hat{\beta}_0 + \hat{\beta}_5$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_5 + \hat{\beta}_{12}$
Laborer	$\hat{\beta}_0 + \hat{\beta}_6$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_6 + \hat{\beta}_{13}$

  Notice that the coefficient for 'left' no longer provides an estimate of the average effect of left handedness across all occupational groups as it did in Model 4. Recall in that model the expected income for left and right handed people differed by £11,880.10 once the variation in income due to occupation, education, and duration of employment are partialled out. Here in Model 5 $\hat{\beta}_1$ instead estimates the difference in income between left and right handed upper white-collar workers. The t-test for this coefficient is then a test of the null hypothesis that 'left' has no significant effect on expected income - net of other variables - for upper white-collar workers. That is, the null hypothesis is that expected income for left handed upper white-collar workers is equal to the expected income for right handed upper white-collar workers, after controlling for variation in income due to education and duration of employment. If this null hypothesis is rejected, then we learn that left handed workers among upper white-collar workers average lower incomes than the right handed workers among upper white-collar workers, controlling for all other factors specified in the model.
- Similarly, the coefficients for the occupation dummy variables no longer provide an estimate of the average effect of being in a particular occupational category versus the reference group as it did in Model 4. Recall that in that model the expected incomes for skilled workers were £23,430.70 less than the upper white-collar workers on average, irrespective of handedness (ie left and right handed workers together in each category). Here in Model 5,

  $\hat{\beta}_2$ instead estimates the difference in expected earnings between lower white-collar and upper white-collar workers who are right handed. Right handed lower white-collar workers average £22,740.90 less in income than right handed upper white-collar workers.

  $\hat{\beta}_3 = -24,180.4$ means, right handed skilled laborers earn on overage £24,180.40 less income than right handed upper white-collar workers.

  That is, once interaction terms are specified, the coefficients for the original set of variables (in this

example, 'left', 'lowwc', 'skill', 'oper', 'serv', and 'labor') refer to comparisons involving the reference categories:

$\hat{\beta}_1$ measures the effect of being left handed for the reference category for occupation, i.e. upper white-collar workers;

$\hat{\beta}_2$ through $\hat{\beta}_6$ measure the effects of being in an occupational category other than upper white-collar for the right handed workers, i.e. the reference category for handedness.

The t tests associated with the regression coefficients $\hat{\beta}_2$ through $\hat{\beta}_6$ are therefore tests for significant differences among occupational groups for right handed workers that can be generalized to the population.

- The coefficients for the interaction terms estimate the differential effect of occupation by handedness. Alternatively, these coefficients can also be seen as estimates of the differential effect of being left handed by occupational category.

  This is because the difference in predicted income between, say, lower white-collar workers and upper white-collar workers is captured by $\hat{\beta}_2 = -22,740.9$ for right handed workers and by $\hat{\beta}_2 + \hat{\beta}_9 = -22,740.9 + 15,010.2 = -7,730.7$ for left handed workers. Therefore, $\hat{\beta}_9 = 15,010.2$ estimates the difference in the effect of being lower white-collar worker for left handed workers relative to right handed ones. So, the earnings gap between lower white-collar workers and upper white-collar workers is £15,010.20 narrower for left handed workers than right handed ones, or -£7,730.70 rather than -£22,740.90.

- The difference in expected income between left and right handed workers who are upper white-collar workers is $\hat{\beta}_1 = -37,930.3$. On the other hand, the left-right difference among lower white-collar workers is $\hat{\beta}_1 + \hat{\beta}_9 = -37,930.3 + 15,010.2 = -22,920.1$.

  Here $\hat{\beta}_9$ estimates the difference in the effect of being left handed for lower white-collar workers relative to upper white-collar workers. So the difference in expected income between left and right handed workers who are lower white-collar workers is -£22,920.10.

  This illustrates that the differences in expected income by occupation for left handed workers are captured by the sum of two coefficients: the coefficient of an occupation dummy variable, β_j plus the coefficient of the relevant interaction term, β_{jk}. The connection between β_j and β_{jk} can be defined as follows:

$$\beta_j = \mathbb{E}(Y_i \mid \text{right, occ}_j) - \mathbb{E}(Y_i \mid \text{right, occ}_{\text{ref}})$$

  which is the difference between the expected income given a right handed worker in occupation category j and the expected income given a right handed worker in the reference occupation category. Similarly,

$$\begin{aligned}\beta_{jk} &= \big[\mathbb{E}(Y_i \mid \text{left, occ}_j) - \mathbb{E}(Y_i \mid \text{left, occ}_{\text{ref}})\big] - \big[\mathbb{E}(Y_i \mid \text{right, occ}_j) - \mathbb{E}(Y_i \mid \text{right, occ}_{\text{ref}})\big] \\ &= \big[\mathbb{E}(Y_i \mid \text{left, occ}_j) - \mathbb{E}(Y_i \mid \text{left, occ}_{\text{ref}})\big] - \beta_j\end{aligned}$$

  Rearranging this then gives:

$$\beta_j + \beta_{jk} = \mathbb{E}(Y_i \mid \text{left, occ}_j) - \mathbb{E}(Y_i \mid \text{left, occ}_{\text{ref}}). \tag{3}$$

- The t tests for the coefficients of the interaction terms are therefore testing whether the net income differential between specific occupational groups and the reference group is the same for left and right handed workers.

- If the coefficients for the interaction terms had been negative, we would have had evidence that the earnings differences between upper white collar-workers and remaining occupational groups were larger for left handed workers than right handed workers.

  That is, the occupational differences in earning had been identified for right handed workers through the negative coefficients for the occupation dummy variables. These would have been even larger for left handed workers because of the extra negative effect captured by coefficients for the interaction terms.

- But these coefficients of interaction terms are positive. Therefore, it appears that the differences in earnings across occupational groups are more pronounced for right handed workers, and more compact for left handed workers. In fact, there may very well be no significant occupational differences in income among left handed workers.

  It also seems that the left/right difference in expected income becomes narrower as we move down the occupational scale.

Should the partial effects of education and duration of employment be the same for all subgroups?

This question effectively means testing the hypotheses that $\beta_{\text{educ(left)}} = \beta_{\text{educ(right)}}$ and $\beta_{\text{dur(left)}} = \beta_{\text{dur(right)}}$. Since it is testing the variability of relationships, we would add two new interaction terms to Model 5 that interacts education with left handedness, and duration with left handedness:

$$\begin{aligned}\text{Model 6: } Y_i &= f(\text{handedness, occupation, education, tenure}) \\ &= \beta_0 + \beta_1\ \text{left}_i + \beta_2\ \text{lowwc}_i + \beta_3\ \text{skill}_i + \beta_4\ \text{oper}_i + \beta_5\ \text{serv}_i + \beta_6\ \text{labor}_i + \beta_7\ \text{educ}_i + \beta_8\ \text{dur}_i \\ &\quad + \beta_9\ \text{leftlow}_i + \beta_{10}\ \text{leftskill}_i + \beta_{11}\ \text{leftop}_i + \beta_{12}\ \text{leftserv}_i + \beta_{13}\ \text{leftlab}_i + \beta_{14}\ \text{lefteduc}_i \\ &\quad + \beta_{15}\ \text{leftdur}_i + u_i.\end{aligned}$$

Interpretation

The interpretation of the coefficients for the variables that are the same as those in Model 5 is similar, though now we assess these effects controlling for the differential impact of education and duration of employment by handedness in addition to other independent variables.

Suppose our estimation gives us:

$$\begin{aligned}Y_i = 49,620.5 &- 16,670.3\ \text{left} &&- 21,550.4\text{lowwc} &&- 21,670.9\ \text{skill} &&- 31,320.1\ \text{oper} \\ se: (4,350.9) &\quad (9,010.3) &&\quad (2,810.7) &&\quad (2,420.5) &&\quad (2,680.8) \\ &- 42,810.2\ \text{serv} &&- 38,510.3\ \text{labor} &&+ 3,590.1\ \text{educ} &&+ 800.3\ \text{dur} \\ &\quad (3,780.9) &&\quad (4,110.0) &&\quad (290.4) &&\quad (70.6) \\ &+ 10,860.5\ \text{leftlow} &&+ 14,490.5\ \text{leftskill} &&+ 19,870.6\ \text{leftop} &&+ 26,760.1\ \text{leftserv} \\ &\quad (8,290.0) &&\quad (7,440.4) &&\quad (6,720.5) &&\quad (7,990.5) \\ &+ 22,390.7\ \text{leftlab} &&- 1,720.7\ \text{lefteduc} &&+ 140.2\ \text{leftdur} \\ &\quad (8,150.2) &&\quad (470.5) &&\quad (150.2)\end{aligned}$$

$R_6^2 : 0.32434$
$F_6 : 102.25$

For most, but not all, of the variables, the coefficient estimates and significant tests remain largely unchanged compared to Model 5.

The coefficient for `leftdur` is not significant which suggests that working with the same employer for additional duration pays left and right handed workers about equally. Since we now have a product term, the interpretation of the coefficient for `dur` is also modified. Now, this coefficient estimates the effect of additional years with the same employer for right handed workers at £800.30 per year. The coefficient for `leftdur` estimates the difference between the net effect of job duration for left and right handed workers to be £140.20, making each additional year worth £940.50 for left handed workers. However, the size of the standard error for `leftdur` indicates that the evidence of this difference is weak. Therefore, we are led to the conclusion that duration operates in roughly the same way for left and right handed workers.

The situation with education is different. The coefficient for `educ` tells us that each additional year of education is associated with £3,590.10 in additional income for the right handed workers, of course when controlling for the effects of other variables. The coefficient for `lefteduc` indicates that for left handed workers, each additional year of education pays £1,720.70 less than that, or £1,860.40. From the standard errors, we can see that this difference is statistically significant. This means, in the population, an additional year of education is associated with a smaller average increment to income for left handed workers than for right, net of other variables in the equation. Since there is a difference between left and right handed workers in this regard, this in turn means that the estimated effect of education in Model 5 - i.e. the average effect of education for left and right handed workers - underestimates the return on additional years of schooling for the right handed workers, and overestimates the effect for left handed ones.

Therefore, we now know that the net effect of duration for left handed workers is not significantly different from right handed workers, but that the net effect of education is. However, we do not yet know whether education significantly affects the expected level of income for left handed workers. To answer this question we must return to equation (3) on page 3 and test the sum of the coefficients estimating the effect of education for left handed workers:

$$\beta_j + \beta_{jk} = \beta_7 + \beta_{14} = 3,590.1 + (-1,720.7) = 1,860.4$$

If following t-test we reject the null hypothesis that this coefficient is 0, we can conclude that education does affect expected income for both left and right handed workers but left handed workers average a lower rate of return on education than do right handed workers.

Notice that when we in Model 5 we constrained the net effect of education on income to be the same for left and right handed workers. In that model, the coefficient for `left` indicated that left handed upper white-collar workers on average earned almost £40,000 less (£37,930.30) per year than their right handed counterparts, controlling for other variables. However, in Model 6 we estimate handedness-specific effects for education which gives us the interpretation that among upper white-collar workers, the net difference in expected income between left and right handed workers is nonsignificant at the standard .05 level for a two-tailed test.

Although the standard error for the coefficient for `left` has increased from Model 5 to Model 6, the more important change is in the point estimate for the coefficient itself where the coefficient of 16,670.3 is less than half that of 37,930.3 in Model 5. This reduction in the difference in earnings is reflected at 0 years of education and 0 years of job duration. This gap nevertheless increases as the years of education increases. That is, the relative earnings advantage of right handed upper white-collar workers over their left handed counterparts increases with the level of education. So, for example, what was a difference of £16,670 at 0 years of education and job duration becomes an average difference of £44,300 for upper white-collar workers with 16 years of education: $16,670.3 + 16 \times 1,720.7 \approx 44,300$.

Now consider $\hat{\beta}_{10}$, the coefficient for `leftskill`. It is not only smaller in Model 6 than in Model 5 but it is also no longer significant at .05 level. Accordingly, we can no longer reject the null hypothesis that the net effect of being a skilled worker rather than an upper white-collar worker is the same for left and right handed workers, once we control for the other factors in general and handedness-specific effect of education in particular.

In comparing the results of Models 5 and 6, one thing that comes across is that the net difference in expected income, Y_i, between left-handed and right-handed upper white-collar workers is partially attributable to the fact that attaining a higher level of education does not produce the same income dividend for left handed workers as it did for right handed ones. One this difference is allowed in determining income - i.e. once we allow the increment to expected income associated with additional schooling to be smaller for left handed workers than right-handed ones - there emerges one possible explanation as to why upper white-collar workers who are left handed had lower expected income than their counterparts who are right-handed. It appears that their additional education was valued at a lower rate of return.

Notice that this does not necessarily invalidate the initial observation that left-handed upper white-collar workers were at an income disadvantage relative to right-handed upper white-collar workers. Instead Model 6 shows one mechanism involved in producing that disadvantage. In addition, we also learn that among

skilled workers what initially appeared as a handedness difference in the effect of being skilled versus upper white-collar is, in part, a function of the differential returns to education that accrue to left and right handed workers.

Recap

In the absence of specified product terms (i.e. interaction effects), the coefficients for independent variables reports the "average" effects, or when other independent variables are included in the specification, "average" partial effects. When we expand the specification to include product terms - thereby removing the constraint of equivalent effects across all groups), we can compare the R^2 vlaues from the two specifications and determine whether relaxing the constraint of equal subgroup effects results in a significant improvement in the fit of the model.

Once we estimated Model 6, we'd use t-tests on $\hat{\beta}_2$ through $\hat{\beta}_6$ to test the net occupation effects for the left handed workers. Similarly, we'd use a t-test on β_1 to test the net effect of being left handed on expected income for upper white-collar workers.

In order to test whether the effect of an independent variable is significant for the left handed workers, we'd construct a t test for the sum of two coefficients.

In order to test whether the effects of two independent variables are significantly different from each other - e.g. to test whether skilled workers are different from lower white-collar workers - we'd construct a t test for the difference between two coefficients.

Finally, by examining the t test for the product terms, we can determine whether the effects of explanatory variables differ by handedness.

Interpretation of Dummy Variables in Semilogarithmic Equations

In this example, we use the log transformation of the dependent variable, earnings, and leave the independent variables in their original metric:

$$\begin{aligned}\text{Model 7: } \ln(Y_i) &= f(\text{handedness, occupation, education, tenure}) \\ &= \beta_0 + \beta_1\ \text{left}_i + \beta_2\ \text{lowwc}_i + \beta_3\ \text{skill}_i + \beta_4\ \text{oper}_i + \beta_5\ \text{serv}_i + \beta_6\ \text{labor}_i + \beta_7\ \text{educ}_i + \beta_8\ \text{dur}_i \\ &\quad + \beta_9\ \text{leftlow}_i + \beta_{10}\ \text{leftskill}_i + \beta_{11}\ \text{leftop}_i + \beta_{12}\ \text{leftserv}_i + \beta_{13}\ \text{leftlab}_i + \beta_{14}\ \text{lefteduc}_i \\ &\quad + \beta_{15}\ \text{leftdur}_i + u_i.\end{aligned}$$

The interpretation of the coefficients would differ depending on whether the independent variable is continuous or a dummy variable. When X_{ki} is a continuous measure, the associated coefficient is interpreted as the relative change in Y for a given absolute change in X - e.g. the proportional change in income for a one-year change in job duration.

Halvorsen and Palmquist (1980) have shown that this is not a correct way to interpret when X_{ki} is a dummy dependent variable. Since the coefficient for the dummy variable captures the difference in subgroup means between the designated and reference groups in units of the dependent variable, when $Y^* = \ln Y$, the coefficient $\hat{\beta}_k$ already expresses relative change in units of $\ln Y$. In this case, the coefficient of a dummy variabke in a semilogarithmic regression actually equals to

$$\hat{\beta}_k = \ln\left(\frac{1 + \hat{Y}_{designated} - \hat{Y}_{reference}}{\hat{Y}_{reference}}\right)$$

where $\hat{Y}_{designated}$ is the predicted value of $\hat{Y}$ for the group coded 1. In order to find the percentage effect of the dummy variable on Y - measured in the original units of Y rather than relative to the log-transformed distribution - it is necessary to use the inverse of the logarithmic function, or the antilog function. The

percentage difference associated with being in the group coded 1 rather than in the reference group is then equal to

$$100(e^{\hat{\beta}_k} - 1).$$

So, suppose our Model 7 gives the following estimates:

$$\begin{array}{rllll}
Y_i = 8.353 & -0.632 \text{ left} & -0.244\text{lowwc} & -0.174 \text{ skill} & -0.328 \text{ oper} \\
se : (0.056) & (0.115) & (0.036) & (0.031) & (0.035) \\
& -0.585 \text{ serv} & -0.510 \text{ labor} & +0.049 \text{ educ} & +0.014 \text{ dur} \\
& (0.049) & (0.053) & (0.004) & (0.001) \\
& +0.215 \text{ leftlow} & +0.230 \text{ leftskill} & +0.272 \text{ leftop} & +0.480 \text{ leftserv} \\
& (0.106) & (0.095) & (0.093) & (0.102) \\
& +0.301 \text{ leftlab} & -0.006 \text{ lefteduc} & +0.013 \text{ leftdur} & \\
& (0.104) & (0.006) & (0.002) &
\end{array}$$

$$R_7^2 : 0.42489$$
$$MeanRSS : 0.23654$$

Then, if we want to find the expected income differential between left-handed and right handed workers, then

$$100\big(e^{-0.632} - 1\big) = -0.468.$$

Therefore, the expected value of Y for the designated group - left handed upper white-collar workers - is 46.8% lower than the value for the reference group - right handed upper white-collar workers. Since the coefficient is statistically significant, this effect is significant.

The effect of being left handed for lower white-collar workers equal to $\hat{\beta}_1 + \hat{\beta}_9$; for skilled workers it is $\hat{\beta}_1 + \hat{\beta}_{10}$; for operatives, $\hat{\beta}_1 + \hat{\beta}_{11}$; for service workers $\hat{\beta}_1 + \hat{\beta}_{12}$; and for laborers $\hat{\beta}_1 + \hat{\beta}_{13}$. To see if there are significant handedness differences net of other specified effects, we'd use the t test as described under the "Assessing Group Differences" section.

So overall, then, we can say that the net effect of being left-handed in income differs by occupation: Left handed upper white-collar workers are at a significant income disadvantage relative to their right handed counterparts and this disadvantage is maintained across all but one - low white-collar - workers (since $\hat{\beta}_9$, the coefficient for `leftlow`, is not significant.)