Lecture 16

I. Effects of Data Scaling on OLS Statistics

A. Scaling the Dependent Variable

Consider the example (6.1), p. 182

$$\widehat{bwght} = \hat{\beta}_0 + \hat{\beta}_1 \, cigs + \hat{\beta}_2 \, faminc$$

Suppose that we decide to measure birth weight in pounds, rather than in ounces. Let bwghtlbs = bwght/16 be birth weight in pounds. What happens to our OLS statistics if we use bwghtlbs as our dependent variable instead of bwght?

Consider

$$\widehat{bwght/16} = (\hat{\beta}_0/16) + (\hat{\beta}_1/16) \, cigs + (\hat{\beta}_2/16) \, faminc.$$

It follows that each new coefficient will be the corresponding old coefficient divided by 16.

What about statistical significance?

It is quite gratifying to know that the t-statistics are unaffected by the rescaling of the dependent variable. Even though the coefficients in the rescaled equation are 1/16 th the size they were before, the standard errors of the new coefficients are also 1/16 th the size they were before. Also the R^2 and overall F-statistic are unaffected by the rescaling of the dependent variable.

B. Scaling an Explanatory variable

Consider the rescaling of the cigs variable above to be packs = cigs/20. What happens to the coefficients and the other OLS statistics? Well, we can write

$$\widehat{bwght} = \hat{\beta}_0 + 20\hat{\beta}_1 (cigs/20) + \hat{\beta}_2 famine$$

$$= \hat{\beta}_0 + (20\hat{\beta}_1) packs + \hat{\beta}_2 famine$$

Thus, the intercept and slope coefficient on famine are unchanged, but the coefficient on packs is 20 times the coefficient on cigs. However, the t-statistics on the coefficients are unchanged, the R^2 and overall F-statistics are unchanged, and the Standard Error of the regression is unchanged.

What happens if we rescale the dependent variable by one constant and rescale an explanatory variable by another constant <u>at the same time</u>?

Let the original OLS regression be denoted by

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k \qquad (1)$$

Now suppose that we rescale the dependent variable y by the constant c_y resulting in a new dependent variable $y^* = c_y \cdot y$ and, at the same time, rescale the j-th explanatory variable by the constant c_j resulting in a new j-th explanatory variable $X_j^* = c_j \cdot X_j$. The new fitted equation becomes

$$\tilde{y}^* = \tilde{\beta}_0 + \tilde{\beta}_1 X_1 + \cdots + \tilde{\beta}_j X_j^* + \cdots + \tilde{\beta}_k X_k \qquad (2)$$

where we have used the tildas (~) in the second equation to distinguish the OLS coefficient estimates in the second equation from the OLS coefficient estimates in the second equation. The relationship between the OLS estimates of the second equation with those of the first are as follows:

$$\tilde{\beta}_0 = c_y \hat{\beta}_0 \;,\; \tilde{\beta}_1 = c_y \hat{\beta}_1 \;,\; \cdots \;,\; \tilde{\beta}_{j-1} = c_y \hat{\beta}_{j-1} \;,$$

$$\tilde{\beta}_j = c_y \cdot c_j \hat{\beta}_j \;,\; \tilde{\beta}_{j+1} = c_y \hat{\beta}_{j+1} \;,\; \cdots \;,\; \tilde{\beta}_k = c_y \hat{\beta}_k \;.$$

All coefficient estimates in the second equation are rescaled versions of the coefficient estimates in the first equation. For all coefficients, except the j-th, the rescaling factor is c_y whereas for the j-th coefficient the rescaling factor is $c_y \cdot c_j$. However, the t-statistics, R^2, overall F-statistic, and any subset F-statistics are left unchanged by this rescaling.

II. Beta Coefficients

As we have seen above, coefficient estimates can be rescaled to be any size we want by rescaling the dependent variable and/or ~~the~~ a given explanatory variable. Therefore, we cannot simply rely on the size of a coefficient to tell us how important the variable is in terms of its impact on the dependent variable. Of course, as we mentioned above, rescaling does not affect the statistical significances of variables.

In order to determine the "relative strength" of the explanatory variables, we can turn to the computation of regression coefficients derived from a regression of the standardized form on the dependent on the standardized forms of the explanatory variables. Consider the "Beta" equation

$$z_y = \hat{b}_1 z_1 + \hat{b}_2 z_2 + \cdots + \hat{b}_k z_k + \text{error}$$

where

$$z_y = \frac{y-\bar{y}}{\hat{\sigma}_y}, \quad z_1 = \frac{X_1-\bar{X}_1}{\hat{\sigma}_1}, \quad \ldots, \quad z_k = \frac{X_k-\bar{X}_k}{\hat{\sigma}_k}$$

and $\hat{\sigma}_y = \sqrt{\dfrac{\sum_1^N (y_i-\bar{y})^2}{N-1}}$, $\hat{\sigma}_j = \sqrt{\dfrac{\sum_1^N (X_{ij}-\bar{X}_j)^2}{N-1}}$.

The "Beta" coefficients, $\hat{b}_j$, relate to the original OLS coefficients $\hat{\beta}_j$ as in

$$\hat{b}_j = (\hat{\sigma}_j/\hat{\sigma}_y)\,\hat{\beta}_j \quad , \; j=1,2,\ldots,k.$$

Then if X_1 increases by one standard deviation then $\hat{y}$ changes by $\hat{b}_1$ standard deviations. Thus, we are measuring effects, not in terms of the original units of y on the X_j, but in standard deviation units. In a standard OLS regression it is not possible to simply look at the size of the different coefficients and conclude that the variable with the largest coefficient is "the most important." The Beta coefficients, however, allow us to make such comparisons. See the program hprice2.sas for an example of how you get PROC REG to produce beta coefficients. It should be noted that whether OLS coefficients or Beta coefficients, the t-statistics are the same in both cases, thus the statistical significance of a variable is not affected.

III. More on Logarithmic Functional Forms.

See Wooldridge's very good discussion on pp. 188-189 on why logarithms are often taken of the dependent variable when building a regression model; log is a possible solution for heteroskedastic errors and helps reduce the effects of outliers in a regression equation.

Look at p.189 for Wooldridge's discussion of the difference between "percentage change" and "percentage point change." Also see p. 682 on this topic.

Finally consider the model

$$\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 \log(X_1) + \hat{\beta}_2 X_2 .$$

we saw from previous discussion that we can interpret $\hat{\beta}_2$ in the following way:

$$\% \Delta y = \hat{\beta}_2 \cdot 100 \cdot \Delta X. \qquad (1)$$

But for large changes in X, $\Delta X >> 0$, we can use a more exact formula, namely,

$$\% \Delta y = 100 \left[\exp(\hat{\beta}_2) - 1 \right]. \qquad (2)$$

well how large does ΔX have to be before you go from using the approximation (1) to using the exact formula (2)? The answer is when the two formulas give substantially different answers!