# Chapter 6: Further Issues in MLR

## Incorporating Nonlinearities (see Chapter 2 pp. 43–44)

So far we have focused on linear relationships between the dependent and independent variables. Sometimes this is not general enough.

One important possibility is when the dependent variable is in logarithmic form:

$$\log(y) = \beta_0 + \beta_1 X_1 + u$$

Recall the wage-education example, where we obtained a slope estimate of $\hat{\beta}_1 = 0.54$, which means that each additional year of education is predicted to increase hourly wage by 54¢. This is the same for the first year of education and the 12th year of education. (or the 100th year!). This may not be reasonable.

A better characterization ~~of~~ of how wages change maybe that each year of education increases wage by a constant percentage. For example, increasing education from 5 to 6 years, ceteris paribus, increases wages by 8% as does an increase from ~~[illegible]~~ 11 to 12 years of education.

A model that gives (approx.) a constant percentage effect is

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

In particular, if $\Delta u = 0$, then

$$\% \Delta wage \approx (100 \cdot \beta_1) \Delta educ$$

Note: we multiply $\beta_1$ by 100 to ge the percentage change in wage given one additional year of education. Since the percentage change in wage is the same for each extra year of educ this gives us an increasing return to education.

## Models with Quadratics

Quadratic functions are used quite a bit in applied econometrics to capture decreasing or increasing marginal effects.

In the simplest case

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

For example, let $y$ = wage and $x$ = exper. $\beta_1$ does not (now) measure the change in $y$ with respect to $x$, because it makes no sense to hold $x^2$ fixed while changing $x$!

Writing the estimated equation we have

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$$

then we have the approximation

$$\Delta \hat{y} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x)\Delta x \quad \text{so}$$

$$\frac{\Delta \hat{y}}{\Delta x} \approx \hat{\beta}_1 + 2\hat{\beta}_2 x$$

This means that the slope of the relationship between $x$ and $y$ depends on the value of $x$ (estimated slope is $\hat{\beta}_1 + 2\hat{\beta}_2 x$)

If we plug in $X=0$, $\hat{\beta}_1$ can be interpreted as the approx. slope going from $X=0$ to $X=1$. After that the second term $2\hat{\beta}_2 X$ must be accounted for

Ex:

$$\widehat{wage} = \underset{(.35)}{3.73} + \underset{(.041)}{.298}\, exper - \underset{(.0009)}{.0061}\, exper^2$$

$$n = 526, \quad R^2 = .093$$

This implies that exper has a diminishing effect on wage.

The first year of experience is worth roughly 30¢ per hour (\$.298)

The second is worth less [$.298 - 2(.0061)(1) \approx .286$ or 28.6¢]

From 10 to 11 years of experience wage is predicted to increase by about $.298 - 2(.0061)(10) = .176$ or 17.6¢

And so on.

~~Also~~ with $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$ the turning point is achieved at

$$x^* = \left| \frac{\hat{\beta}_1}{2\hat{\beta}_2} \right|$$

---

<u>Models with Interaction Terms</u>

Sometimes the partial effect of the dependent variable with respect to an explanatory variable may depend on another explanatory variable.

for example, consider the model

$$price = \beta_0 + \beta_1 \, sqrft + \beta_2 \, bdrms + \beta_3 \, sqrft \cdot bdrms + \beta_4 \, bthrms + u$$

the partial effect of bedrooms on price (hold all else fixed) is

$$\frac{\Delta \, price}{\Delta \, bdrms} = \beta_2 + \beta_3 \, sqrft$$

If $\beta_3 > 0$ then an additional bedroom yields a higher increase in housing price for larger houses.

## More on Goodness-of-Fit

We mentioned that adding an additional (or many additional) explanatory variables can never decrease the $R^2$. We want a measure that in some sence penalizing for adding additional variables. Then if that new measure increases when an additional variable is added we can be more certain that it is because it is helpful in explaining the variation in $y$! This is the adjusted $R^2$:

oops

↓

$$\bar{R}^2 = 1 - [SSR/(n-k-1)] / [SST/(n-1)]$$

$$= 1 - \hat{\sigma}^2 / [SST/(n-1)]$$

or in terms of $R^2$

$$\bar{R}^2 = 1 - (1-R^2)(n-1)/(n-k-1)$$

## Chapter 7 : MLR with Qualitative Information

Sometimes we want to incorporate qualitative information into our models. Consider

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + u$$

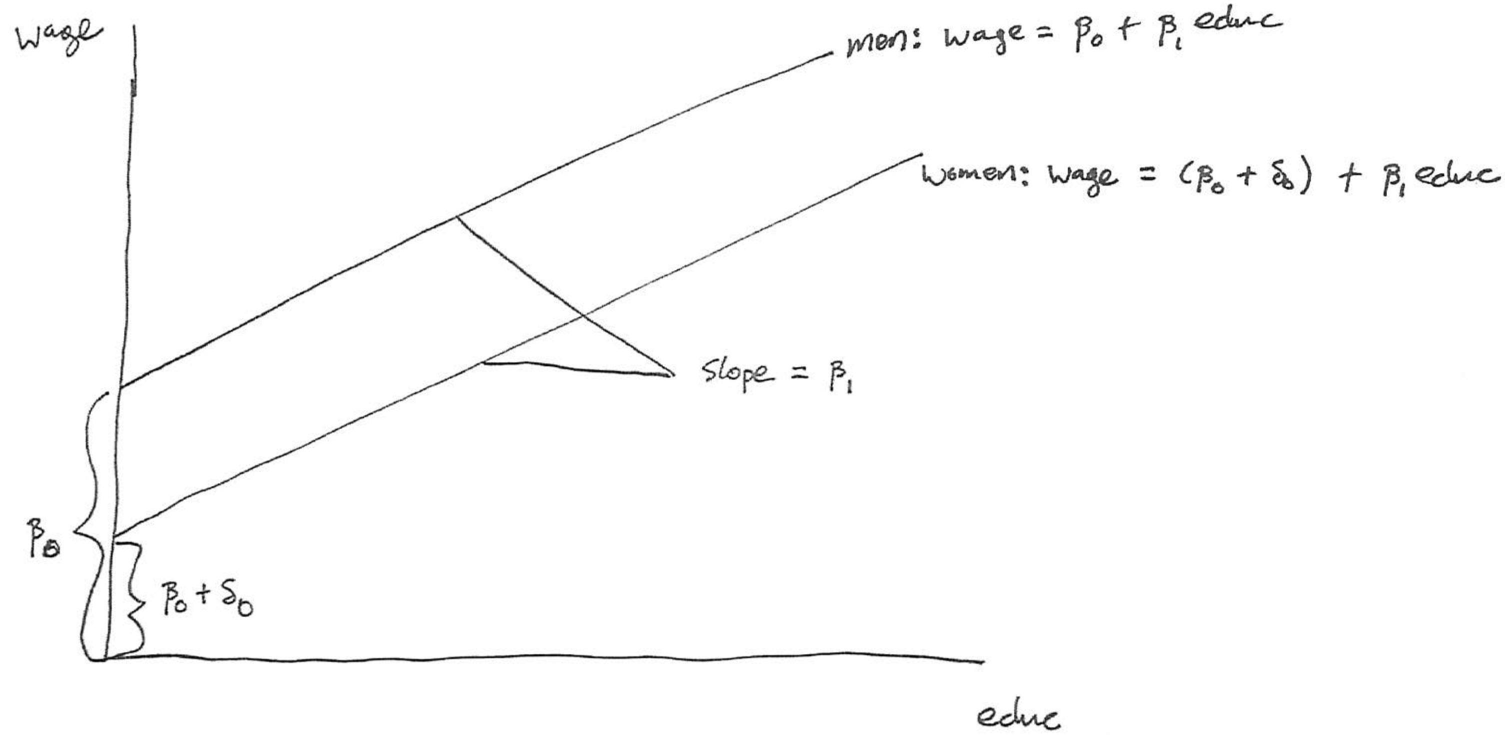
Where female is a binary variable equal to 1 if individual $i$ is a female and 0 otherwise.

The parameter $\delta_0$ has the following interpretation:

$\delta_0$ is the difference in hourly wages between males and females, given the same amount of education.

Thus $\delta_0$ determines if there is discrimination. If $\delta_0 < 0$ then, for the same level of other factors, women earn less than men on average.

Graphically



## Avoiding the Dummy Variable Trap

Including two binary variables, say $female = \begin{cases} 1 & \text{if female} \\ 0 & \text{if male} \end{cases}$

and $male = \begin{cases} 0 & \text{if female} \\ 1 & \text{if male} \end{cases}$ would introduce

perfect colinearity because $female + male = 1$, which

means the male is a perfect linear function of female.

Including the both is called the dummy variable trap.

When we in the binary variable females in wage regression we have chosen males as the <u>base group</u> or <u>bench mark group</u>. This is the group against which comparisons are made. That is why $\beta_0$ is the intercept for males and $\delta_0$ is the difference between males and females.

We could have just as easily written the model

$$wage = \alpha_0 + \gamma_0\, males + \beta_1\, educ + u$$

in which case females would be the base group, and $\alpha_0$ would be the intercept for females and $\gamma_0$ would be the difference between females and males.

We can put binary variables in the MLR as, for example

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \delta_0 female + u$$

We can test for no difference between males and females with a t test with

$$H_0: \delta_0 = 0$$

$$H_1: \delta_0 < 0$$

Rejecting $H_0$ would suggest there is evidence of discrimination.