ECON 351* -- Introduction to NOTE 11: Multiple Linear Regression Models

Interpreting Slope Coefficients in Multiple Linear Regression Models: An Example

- Consider the following ***simple*** **linear regression model** for the birth weight of newborn babies given by the following **population regression equation (PRE)**:

$$\text{bwght}_i = \beta_0 + \beta_1 \text{cigs}_i + u_i \qquad (1)$$

where the observable variables are defined as follows:

$\text{bwght}_i \equiv$ birth weight of newborn baby born to mother i, in grams;
$\text{cigs}_i \equiv$ average number of cigarettes smoked per day during pregnancy by mother i.

Interpretation of slope coefficient β_1 on explanatory variable $cigs_i$ in simple regression model (1):

$$\frac{d\,E\left(\text{bwght}_i \mid \text{cigs}_i\right)}{d\,\text{cigs}_i} = \frac{d\,(\beta_0 + \beta_1 \text{cigs}_i)}{d\,\text{cigs}_i} = \beta_1$$

= ***unadjusted*** **marginal effect of cigs_i** on **mean birth weight** of newborn babies

= the **change in mean birth weight** of newborn babies, in grams, associated with an increase in mother's cigarette consumption during pregnancy of **one cigarette per day**

- Consider the following ***multiple*** **linear regression model** for the **birth weight of newborn babies** given by the following **population regression equation (PRE)**:

$$\text{bwght}_i = \beta_0 + \beta_1\text{cigs}_i + \beta_2\text{faminc}_i + \beta_3\text{male}_i + \beta_4\text{white}_i + u_i \qquad (2)$$

where the new explanatory variables are defined as follows:

$\text{faminc}_i \equiv$ annual family income of mother i, in thousands of 1988 dollars per year;
$\text{male}_i \equiv$ 1 if newborn baby of mother i is male, = 0 otherwise;
$\text{white}_i \equiv$ 1 if mother i is white, = 0 otherwise.

Interpretation of slope coefficient β_1 on explanatory variable *$cigs_i$* in ***multiple*** regression model (2):

Let $\underline{x}_i$ be the 1×5 row vector of regressor values for observation i: $\underline{x}_i = \begin{bmatrix} 1 & \text{cigs}_i & \text{faminc}_i & \text{male}_i & \text{white}_i \end{bmatrix}$.

$$\frac{\partial E(\text{bwght}_i \mid \underline{x}_i)}{\partial \text{cigs}_i} = \frac{\partial(\beta_0 + \beta_1\text{cigs}_i + \beta_2\text{faminc}_i + \beta_3\text{male}_i + \beta_4\text{white}_i)}{\partial \text{cigs}_i} = \beta_1$$

= ***adjusted*** **(*partial*) marginal effect of cigs_i** on ***conditional*** **mean birth weight** of newborn babies

= the change in ***conditional*** **mean birth weight** of newborn babies, in grams, associated with an increase in mother's daily cigarette consumption during pregnancy of one cigarette, **holding constant** the family income and race of the mother and the sex of the newborn child

= the change in ***conditional*** **mean birth weight** of newborn babies, in grams, associated with an **increase** in mother's daily cigarette consumption during pregnancy of one cigarette, for newborns of the **same sex** whose mothers have the **same family income** and are of the **same race**

Compare the slope coefficient β_1 in Model 1 and Model 2

Model 1 is the ***simple*** linear regression model given by population regression equation (1.1) and the corresponding population regression function (1.2):

$$\text{bwght}_i = \beta_0 + \beta_1 \text{cigs}_i + u_i \tag{1.1}$$

$$E(\text{bwght}_i \mid \text{cigs}_i) = \beta_0 + \beta_1 \text{cigs}_i \tag{1.2}$$

Model 2 is the ***multiple*** linear regression model given by population regression equation (2.1) and the corresponding population regression function (2.2):

$$\text{bwght}_i = \beta_0 + \beta_1 \text{cigs}_i + \beta_2 \text{faminc}_i + \beta_3 \text{male}_i + \beta_4 \text{white}_i + u_i \tag{2.1}$$

$$E(\text{bwght}_i \mid \text{cigs}_i, \text{faminc}_i, \text{male}_i, \text{white}_i) = \beta_0 + \beta_1 \text{cigs}_i + \beta_2 \text{faminc}_i + \beta_3 \text{male}_i + \beta_4 \text{white}_i \tag{2.2}$$

Question: How does the slope coefficient β_1 in the ***simple*** linear regression model 1 given by equations (1.1) and (1.2) differ from the slope coefficient β_1 in the ***multiple*** linear regression model 2 given by equations (2.1) and 2.2)?

Analytical Answer

- The slope coefficient β_1 in the ***simple*** linear regression model given by equations (1.1) and (1.2) is the ***unadjusted* or *total* marginal effect of cigarette consumption** on ***mean* birth weight**, because PRE (1.1) and PRF (1.2) do not account for, or control for, the effects on birth weight of any other explanatory variables apart from ***cigs$_i$***.

- Analytically, this means that the slope coefficient β_1 in the ***simple*** linear regression model (1.1)/ (1.2) corresponds to the ***total* derivative** of **mean birth weight** with respect to ***cigs$_i$***:

$$\frac{d\,E\left(\text{bwght}_i \mid \text{cigs}_i\right)}{d\,\text{cigs}_i} = \frac{d\,(\beta_0 + \beta_1 \text{cigs}_i)}{d\,\text{cigs}_i} = \beta_1 \text{ in } \textbf{Model 1}$$

= the ***unadjusted* or *total* marginal effect of *cigs$_i$*** on the **mean *birth weight* of newborns**

= the **change in *mean* birth weight**, in grams, associated with a **1-cigarette-per-day *increase* in daily cigarette consumption of the mother during pregnancy**

♦ In contrast, the slope coefficient β_1 in the ***multiple*** linear regression model 2 given by equations (2.1) and (2.2) is the ***adjusted* or *partial* marginal effect of cigarette consumption** on ***mean* birth weight**, because PRE (2.1) and PRF (2.2) account for, or control for, the effect on birth weight of other explanatory variables apart from $\boldsymbol{cigs_i}$, namely family income ($\boldsymbol{faminc_i}$), sex of the newborn ($\boldsymbol{male_i}$), and race of the mother ($\boldsymbol{white_i}$).

Analytically, this means that **the slope coefficient β_1 in the *multiple* linear regression model (2.1)/ (2.2)** corresponds to the ***partial* derivative** of **mean birth weight** with respect **to $\boldsymbol{cigs_i}$**:

$$\frac{\partial E\left(\text{bwght}_i \mid \text{cigs}_i, \text{faminc}_i, \text{male}_i, \text{white}_i, \text{mpg}_i\right)}{\partial \text{cigs}_i} = \frac{\partial(\beta_0 + \beta_1\text{cigs}_i + \beta_2\text{faminc}_i + \beta_3\text{male}_i + \beta_4\text{white}_i)}{\partial \text{cigs}_i} = \beta_1$$

in **Model 2**

= the ***adjusted* or *partial* marginal effect of $\boldsymbol{cigs_i}$** on the **mean *birth weight* of newborns**

= the **change in *conditional mean* birth weight, in grams,** associated with an increase of 1 cigarette-per-day in cigarette consumption during pregnancy, holding constant family income (***faminc***), the sex of the newborn (***male***), and the race of the mother (***white***).

= the **change in *conditional mean* birth weight, in grams,** associated with a **1-cigarette-per-day increase in cigarette consumption during pregnancy for mothers of *the same* family income and race and newborns of the same sex**.

Interpreting the slope coefficient estimate of *cigs* in Model 1

♦ Write the **OLS sample regression function (SRF) for Model 1** as

$$\tilde{bwght}_i = \tilde{\beta}_0 + \tilde{\beta}_1 cigs_i \qquad (1.3)$$

where $\tilde{\beta}_j$ denotes the OLS estimate of β_j in Model 1, and $\tilde{bwght}_i$ is the OLS estimate of mean birth weight given by the population regression function (PRF) $E(bwght_i \mid cigs_i) = \beta_0 + \beta_1 cigs_i$ for Model 1, the simple linear regression model given by PRE (10.1) and PRF (10.2).

♦ The OLS SRF (1.3) for Model 1 implies that the ***estimated* change in *mean* birth weight** associated with **a change in cigarette consumption *cigs*** of $\Delta cigs$ is

$$\Delta \tilde{bwght} = \tilde{\beta}_1 \Delta cigs \qquad (1.4)$$

♦ In Model 1, the estimated effect on ***mean* birth weight** of a **1-cigarette-per-day increase** in a mother's cigarette consumption during pregnancy can be obtained by setting $\Delta cigs = 1$ in equation (10.4):

$$\Delta \tilde{bwght} = \tilde{\beta}_1 \qquad \text{when } \Delta cigs = 1 \qquad (1.5)$$

The slope coefficient estimate $\tilde{\beta}_1$ in Model 1 is therefore **an *estimate* of the change in *mean* birth weight** associated with **a 1-cigarette-per-day increase in *cigs***, holding constant no other explanatory variables that may be related to birth weight.

Interpreting the slope coefficient estimate of *cigs* in Model 2

- Write the **OLS sample regression function (SRF) for Model 2**, obtained by OLS estimation of the multiple linear regression equation (2.1), as

$$\hat{bwght}_i = \hat{\beta}_0 + \hat{\beta}_1 cigs_i + \hat{\beta}_2 faminc_i + \hat{\beta}_3 male_i + \hat{\beta}_4 white_i \qquad (2.3)$$

where $\hat{\beta}_j$ denotes the OLS estimate of β_j in Model 2, and $\hat{bwght}_i$ is the OLS estimate of mean birth weight given by the population regression function (PRF) for Model 2, the multiple linear regression model given by PRE (2.1) and PRF (2.2).

$$E\left(bwght_i \middle| cigs_i, faminc_i, male_i, white_i\right) = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + \beta_3 male_i + \beta_4 white_i$$

- The OLS SRF (2.3) implies that the ***estimated* change in *mean* birth weight** associated with **a change in daily cigarette consumption *cigs*** of $\Delta cigs$ ***and* simultaneous changes in *family income*** of $\Delta faminc$, in ***sex* of the newborn** of $\Delta male$, and in ***race* of the mother** of $\Delta white$ is

$$\Delta\hat{bwght} = \hat{\beta}_1 \Delta cigs + \hat{\beta}_2 \Delta faminc + \hat{\beta}_3 \Delta male + \hat{\beta}_4 \Delta white \qquad (2.4)$$

- We can hold constant family income, the sex of the newborn, and the race of the mother by setting $\Delta faminc = 0$, $\Delta male = 0$ and $\Delta white = 0$ in equation (2.4); the resulting change in estimated mean birth weight is then

$$\Delta\hat{bwght} = \hat{\beta}_1 \Delta cigs \quad \text{when } \Delta faminc = 0,\ \Delta male = 0 \text{ and } \Delta white = 0 \qquad (2.5)$$

♦ In Model 2, the estimated effect on mean birth weight of a 1-cigarette-per-day increase in mother's cigarette consumption during pregnancy can be obtained by setting $\Delta \text{cigs} = 1$ in equation (2.5), or equivalently by setting $\Delta \text{cigs} = 1$, $\Delta \text{faminc} = 0$, $\Delta \text{male} = 0$ and $\Delta \text{white} = 0$ in equation (2.4):

$$\Delta \widehat{\text{bwght}} = \hat{\beta}_1 \qquad \text{when } \Delta \text{cigs} = 1 \textit{ and } \Delta \text{faminc} = 0,\ \Delta \text{male} = 0, \text{ and } \Delta \text{white} = 0 \tag{2.6}$$

The slope coefficient estimate $\hat{\beta}_1$ in Model 2 is therefore **an *estimate* of the change in *mean* birth weight** associated with a **1-cigarette-per-day increase in *cigs***, **holding constant family income (*faminc*), the sex of the newborn (*male*), and the race of the mother (*white*).**

The following *Stata* exercise is designed to illustrate the correct interpretation of the slope coefficient estimate $\hat{\beta}_1$ in the multiple linear regression model, Model 2. It also illustrates the meaning of "holding constant other variables" in multiple linear regression models. These exercises introduce you to an important post-estimation *Stata* command, the **lincom** command.

OLS Estimation of Models 1 and 2 Using *Stata*

- **OLS estimation** of the ***simple*** **linear regression model** for the birth weight of newborn babies given by **population regression equation (1)** yields the following **OLS sample regression equation**:

$$\text{bwght}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{cigs}_i + \tilde{u}_i \qquad (1^*)$$

Stata* output for Model 1 of OLS estimation command *regress:

```
. regress bwght cigs

      Source |       SS       df       MS              Number of obs =    1388
-------------+------------------------------           F(  1,  1386) =   32.24
       Model |  10496953.2     1  10496953.2           Prob > F      =  0.0000
    Residual |   451331421  1386  325635.946           R-squared     =  0.0227
-------------+------------------------------           Adj R-squared =  0.0220
       Total |   461828374  1387  332969.267           Root MSE      =  570.65

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -14.56544   2.565418    -5.68   0.000    -19.59796   -9.532918
       _cons |   3395.533   16.22586   209.27   0.000     3363.703    3427.363
------------------------------------------------------------------------------
```

- **OLS estimation** of the ***multiple*** **linear regression model** for the birth weight of newborn babies given by **population regression equation (1)** yields the following **OLS sample regression equation**:

$$\text{bwght}_i = \hat{\beta}_0 + \hat{\beta}_1\text{cigs}_i + \hat{\beta}_2\text{faminc}_i + \hat{\beta}_3\text{male}_i + \hat{\beta}_4\text{white}_i + \hat{u}_i \qquad (2^*)$$

Stata* output for Model 2 of OLS estimation command *regress:

```
. regress bwght cigs faminc male white

      Source |       SS       df       MS              Number of obs =    1388
-------------+------------------------------           F(  4,  1383) =   16.84
       Model |  21452435.7     4  5363108.93           Prob > F      =  0.0000
    Residual |   440375938  1383  318420.779           R-squared     =  0.0465
-------------+------------------------------           Adj R-squared =  0.0437
       Total |   461828374  1387  332969.267           Root MSE      =  564.29

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -13.44243   2.577508    -5.22   0.000    -18.49868   -8.386187
      faminc |   1.702555   .8631554     1.97   0.049     .0093198    3.395791
        male |   89.16755   30.35604     2.94   0.003     29.61868    148.7164
       white |   153.2959   38.70656     3.96   0.000     77.36594    229.2258
       _cons |    3177.05   40.89941    77.68   0.000     3096.818    3257.282
------------------------------------------------------------------------------
```

***Stata* Exercise: Model 2**

Question: What is the effect on the **mean birth weight of newborns** of **an *increase* in their mothers' cigarette consumption** during pregnancy **from 10 to 11 cigarettes per day (Δ*cigs* = 1)**, while **holding constant** the mother's ***family income*** at $30,000 per year (***faminc* = 30**), the ***sex of the newborn*** at 'male' (***male* = 1**), and the ***race of the mother*** at 'white' (***white* = 1**)?

Analytical Answer: Compare the expressions implied by Model 2 for **(1) the mean birth weight of newborns for whom *cigs* = 11**, *faminc* = 30, *male* = 1 and *white* = 1 and **(2) the mean birth weight of newborns for whom *cigs* = 10**, *faminc* = 30, *male* = 1 and *white* = 1. That is, compare

$$E\left(\text{bwght}_i \mid \text{cigs}_i = 11, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) = \beta_0 + \beta_1 11 + \beta_2 30 + \beta_3 + \beta_4 \qquad (3.1)$$

and

$$E\left(\text{bwght}_i \mid \text{cigs}_i = 10, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) = \beta_0 + \beta_1 10 + \beta_2 30 + \beta_3 + \beta_4 \qquad (3.2)$$

Subtract the second function from the first function: it is obvious that **this difference is simply β_1**:

$$\begin{aligned} & E\left(\text{bwght}_i \mid \text{cigs}_i = 11, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) \\ & \quad - E\left(\text{bwght}_i \mid \text{cigs}_i = 10, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) \\ & \quad = \beta_0 + \beta_1 11 + \beta_2 30 + \beta_3 + \beta_4 - \beta_0 - \beta_1 10 - \beta_2 30 - \beta_3 - \beta_4 \\ & \quad = \beta_1 (11 - 10) \\ & \quad = \beta_1 \qquad (3.3) \end{aligned}$$

- **<u>Step 1</u>:** Use a *Stata* **lincom** command to compute for Model 2 the estimated mean birth weight of a male newborn (for whom *male* = 1) who is born to a white mother (for whom *white* = 1) **who smoked 10 cigarettes per day during pregnancy** (for whom ***cigs* = 10**) and whose family income is $30,000 per year (*faminc* = 30), i.e., to compute an estimate of the conditional mean function

$$E\left(\text{bwght}_i \,\middle|\, \text{cigs}_i = 10, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) = \beta_0 + \beta_1 10 + \beta_2 30 + \beta_3 + \beta_4 .$$

Our estimate of this conditional mean function can be written as

$$\hat{E}\left(\text{bwght}_i \,\middle|\, \text{cigs}_i = 10, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) = \hat{\beta}_0 + \hat{\beta}_1 10 + \hat{\beta}_2 30 + \hat{\beta}_3 + \hat{\beta}_4 .$$

Enter the *Stata* **lincom** command:

```
. lincom _b[_cons] + _b[cigs]*10 + _b[faminc]*30 + _b[male]*1 + _b[white]*1

 ( 1)  10 cigs + 30 faminc + male + white + _cons = 0

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   3336.166   30.40569   109.72   0.000     3276.519    3395.812
------------------------------------------------------------------------------
```

- **<u>Step 2</u>:** Next, **increase *cigs* by 1 cigarette-per-day**, from 10 to 11, while holding constant family income at a value of 30 thousand dollars per year (*faminc* = 30), the sex of the newborn at 'male' (*male* = 1), and the race of the mother at 'white' (*white* = 1). Use another *Stata* **lincom** command to compute for Model 2 the **estimated mean birth weight** of a male newborn (for whom *male* = 1) who is born to a white mother (for whom *white* = 1) **who smoked 11 cigarettes per day during pregnancy** (for whom ***cigs* = 11**) and whose family income is $30,000 per year (*faminc* = 30), i.e., to compute an estimate of the conditional mean function

$$E\left(\text{bwght}_i \mid \text{cigs}_i = 11, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) = \beta_0 + \beta_1 11 + \beta_2 30 + \beta_3 + \beta_4 .$$

Our estimate of this conditional mean function can be written as

$$\hat{E}\left(\text{bwght}_i \mid \text{cigs}_i = 11, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) = \hat{\beta}_0 + \hat{\beta}_1 11 + \hat{\beta}_2 30 + \hat{\beta}_3 + \hat{\beta}_4 .$$

Enter the *Stata* **lincom** command:

```
. lincom _b[_cons] + _b[cigs]*11 + _b[faminc]*30 + _b[male]*1 + _b[white]*1

 ( 1)  11 cigs + 30 faminc + male + white + _cons = 0

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   3322.723   32.18904   103.23   0.000     3259.579    3385.868
------------------------------------------------------------------------------
```

- **<u>Step 3</u>:** Finally, use a *Stata* **lincom** command to compute **the *difference*** between (1) the estimated mean birth weight of newborns for whom *cigs* = 11, *faminc* = 30, *male* = 1 and *white* = 1 and (2) the estimated mean birth weight of newborns for whom *cigs* = 10, *faminc* = 30, *male* = 1 and *white* = 1, i.e., to compute

$$\begin{aligned}
&\hat{E}\left(\text{bwght}_i \mid \text{cigs}_i = 11, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) \\
&\quad - \hat{E}\left(\text{bwght}_i \mid \text{cigs}_i = 10, \text{faminc}_i = 30, \text{male}_i = 1, \text{white}_i = 1\right) \\
&\quad = \hat{\beta}_0 + \hat{\beta}_1 11 + \hat{\beta}_2 30 + \hat{\beta}_3 + \hat{\beta}_4 - \left(\hat{\beta}_0 + \hat{\beta}_1 10 + \hat{\beta}_2 30 + \hat{\beta}_3 + \hat{\beta}_4\right) \\
&\quad = \hat{\beta}_0 + \hat{\beta}_1 11 + \hat{\beta}_2 30 + \hat{\beta}_3 + \hat{\beta}_4 - \hat{\beta}_0 - \hat{\beta}_1 10 - \hat{\beta}_2 30 - \hat{\beta}_3 - \hat{\beta}_4 \\
&\quad = \hat{\beta}_1 (11 - 10) \\
&\quad = \hat{\beta}_1
\end{aligned}$$

Enter on one line the *Stata* **lincom** command:

```
. lincom _b[_cons] + _b[cigs]*11 + _b[faminc]*30 + _b[male]*1 + _b[white]*1 -
(_b[_cons] + _b[cigs]*10 + _b[faminc]*30 + _b[male]*1 + _b[white]*1)

 ( 1)  cigs = 0

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -13.44243   2.577508    -5.22   0.000    -18.49868   -8.386187
------------------------------------------------------------------------------
```

- **<u>Result</u>:** Compare the output of the **lincom** command in **Step 3** with the slope coefficient estimate $\hat{\beta}_1$ for the regressor ***cigs**$_i$* produced by the **regress** command used to estimate Model 2 by OLS. You will see that they are identical.

```
. lincom _b[_cons] + _b[cigs]*11 + _b[faminc]*30 + _b[male]*1 + _b[white]*1 -
(_b[_cons] + _b[cigs]*10 + _b[faminc]*30 + _b[male]*1 + _b[white]*1)

 ( 1)  cigs = 0

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -13.44243   2.577508    -5.22   0.000    -18.49868   -8.386187
------------------------------------------------------------------------------

. lincom _b[cigs]

 ( 1)  cigs = 0

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -13.44243   2.577508    -5.22   0.000    -18.49868   -8.386187
------------------------------------------------------------------------------
```

Result: The slope coefficient estimate $\hat{\beta}_1$ of *cigs* in Model 2 is an ***estimate*** of the ***change* in mean birth weight of newborns** associated with an ***increase*** of **1 cigarette per day** in the **cigarette consumption of mothers during pregnancy** ($\Delta cigs = 1$), while **<u>holding constant</u>** the **other determinants of newborns' birth weight**, namely family income (*faminc*), sex of the newborn (*male*), and race of the mother (*white*).