ECON 351* -- Addendum to NOTE 13

Goodness-of-Fit in the Multiple Linear Regression Model

3.4 Limitations of R^2

The R^2 can be used to compare the goodness-of-fit of alternative sample regression equations only if the regression models satisfy **two conditions**.

(1) The models must have the ***same*** **regressand**, or ***same*** **dependent variable**.

Reason: TSS, ESS, and RSS depend on the units in which the regressand Y_i is measured.

(2) The models must have the ***same*** **number of regressors and regression coefficients** -- i.e., the ***same*** **value of k**.

Reason: Adding additional regressors to a regression equation -- i.e., increasing the value of k -- always increases the value of R^2.

- **ESS** is an ***increasing*** **function** of the number of regressors k.
- **RSS** is a ***decreasing*** **function** of the number of regressors k.

- Therefore, $\mathbf{R^2}$ is an ***increasing*** **function of the number of regressors k**.

- Consider three alternative regression models of North American car prices

$price_i$ = price of car i, in US dollars
wgt_i = weight of car i, in pounds
mpg_i = miles per gallon of car i
$foreign_i$ = 1 if car i is foreign, = 0 if car i is domestic
$length_i$ = length of car i, in inches
$turn_i$ = turning radius of car i, in inches

Model 1:

PRE: $price_i = \beta_1 + \beta_2 wgt_i + \beta_3 mpg_i + \beta_4 foreign_i + \beta_5 length_i + \beta_6 turn_i + u_i$

Model 2: Drops regressor mpg_i from Model 1;
sets $\beta_3 = 0$ in Model 1.

PRE: $price_i = \beta_1 + \beta_2 wgt_i + \beta_4 foreign_i + \beta_5 length_i + \beta_6 turn_i + u_i$

Model 3: Drops regressors mpg_i and $turn_i$ from Model 1;
sets $\beta_3 = 0$ and $\beta_6 = 0$ in Model 1.

PRE: $price_i = \beta_1 + \beta_2 wgt_i + \beta_4 foreign_i + \beta_5 length_i + u_i$

- **Values of R^2 for Models 1, 2 and 3**

For the OLS sample regression equations of Models 1, 2 and 3:

R^2 for Model 1 > R^2 for Model 2 > R^2 for Model 3

This would be true even if $\beta_3 = 0$ and $\beta_6 = 0$; i.e., even if the true population values of the slope coefficients β_3 and β_6 were zero.

OLS SRE for Model 1 (k = 6)

```
. regress price wgt mpg foreign length turn

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  5,    68) =   17.29
       Model |  355491949     5  71098389.9           Prob > F      =  0.0000
    Residual |  279573447    68  4111374.22           R-squared     =  0.5598
-------------+------------------------------           Adj R-squared =  0.5274
       Total |  635065396    73  8699525.97           Root MSE      =  2027.7

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   5.918179   1.023498     5.78   0.000      3.87582    7.960538
         mpg |   -22.7625   72.13908    -0.32   0.753    -166.7138    121.1888
     foreign |   3273.408   687.0594     4.76   0.000     1902.402    4644.413
      length |  -78.78337   35.09348    -2.24   0.028    -148.8113   -8.755416
        turn |  -149.7014   116.5489    -1.28   0.203    -382.2711    82.86829
       _cons |   8548.568   5726.899     1.49   0.140     -2879.28    19976.42
------------------------------------------------------------------------------
```

OLS SRE for Model 2 (k = 5): omits $\boldsymbol{mpg_i}$ from Model 1

```
. regress price wgt foreign length turn

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  4,    69) =   21.88
       Model |  355082608     4  88770652.0           Prob > F      =  0.0000
    Residual |  279982788    69  4057721.57           R-squared     =  0.5591
-------------+------------------------------           Adj R-squared =  0.5336
       Total |  635065396    73  8699525.97           Root MSE      =  2014.4

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   6.011528   .9733945     6.18   0.000     4.069659    7.953397
     foreign |   3318.753   667.4639     4.97   0.000     1987.198    4650.307
      length |  -77.25884   34.53175    -2.24   0.028    -146.1478   -8.369885
        turn |  -145.9884   115.1943    -1.27   0.209    -375.7946    83.81788
       _cons |   7334.719   4214.683     1.74   0.086    -1073.344    15742.78
------------------------------------------------------------------------------
```

OLS SRE for Model 3 (k = 4): omits mpg_i and $turn_i$ from Model 1

```
. regress price wgt foreign length

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  3,    70) =   28.39
       Model |   348565467     3   116188489           Prob > F      =  0.0000
    Residual |   286499930    70  4092856.14           R-squared     =  0.5489
-------------+------------------------------           Adj R-squared =  0.5295
       Total |   635065396    73  8699525.97           Root MSE      =  2023.1

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   5.774712   .9594168      6.02    0.000     3.861215    7.688208
     foreign |   3573.092    639.328      5.59    0.000     2297.992    4848.191
      length |  -91.37083   32.82833     -2.78    0.007    -156.8449    -25.8968
       _cons |   4838.021    3742.01      1.29    0.200    -2625.183    12301.22
------------------------------------------------------------------------------
```

4. The Adjusted R^2

4.1 Definition of Adjusted R^2

$$\bar{R}^2 \equiv 1 - \frac{RSS/(N-k)}{TSS/(N-1)} = 1 - \frac{\hat{\sigma}^2}{s_Y^2}$$

where

$$\hat{\sigma}^2 = \frac{RSS}{N-k}$$ = the unbiased estimator of the error variance σ^2;

$$s_Y^2 = \frac{TSS}{N-1} = \frac{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}{N-1}$$ = the sample variance of the Y_i values.

4.2 Relationship Between R^2 and Adjusted R^2

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{N-1}{N-k}.$$

(1) For values of $k > 1$, $\bar{R}^2 < R^2$.

(2) $\bar{R}^2$ can be negative, even though R^2 is non-negative.

4.3 Guidelines for Using Adjusted R^2

1. $\bar{R}^2$ can be used to compare the goodness-of-fit of two regression models **only if the models have the *same* regressand**.

2. $\bar{R}^2$ **should never be the sole criterion** for choosing between two or more sample regression equations.

Example of Using Adjusted R^2

Model 1:

PRE: $\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{mpg}_i + \beta_4 \text{foreign}_i + \beta_5 \text{length}_i + \beta_6 \text{turn}_i + u_i$

Model 2: Sets $\beta_3 = 0$ in Model 1.

PRE: $\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_4 \text{foreign}_i + \beta_5 \text{length}_i + \beta_6 \text{turn}_i + u_i$

Model 3: Sets $\beta_3 = 0$ and $\beta_6 = 0$ in Model 1.

PRE: $\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_4 \text{foreign}_i + \beta_5 \text{length}_i + u_i$

- Compare ***unadjusted*** $\mathbf{R^2}$ values for Models 1, 2 and 3

  R^2 for Model 1 (k = 6) = **0.5598**
  R^2 for Model 2 (k = 5) = **0.5591**
  R^2 for Model 3 (k = 4) = **0.5489**

  Note: R^2 for Model 1 > R^2 for Model 2 > R^2 for Model 3

- Compare ***adjusted*** $\mathbf{R^2}$ values for Models 1, 2 and 3

  Adjusted R^2 for Model 1 (k = 6) = **0.5274** worst fit to sample data
  Adjusted R^2 for Model 2 (k = 5) = **0.5336** best fit to sample data
  Adjusted R^2 for Model 3 (k = 4) = **0.5295**