Introductory Econometrics I

Multiple Regression: Further Issues

Yingjie Feng

School of Economics and Management

Tsinghua University

April 19, 2024

Outline

Units of Measurement

- Changing the units of measurement of y or some of x_j **cannot** change the interpretation of the OLS regression line.

  1. Multiply the dependent variable y by a constant $c \neq 0$:

  $$\begin{aligned} & y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \\ \rightarrow \quad & cy = c\beta_0 + c\beta_1 x_1 + \cdots + c\beta_k x_k + cu \end{aligned}$$

     - ★ All coefficients (intercept and slopes) get multiplied by c
     - ★ Standard errors of OLS estimates get multiplied by c
     - ★ Fitted values $\hat{y}_i$ and residuals $\hat{u}_i$ get multiplied by c
     - ★ R^2, t statistics (except they change sign if $c < 0$) and F statistics do not change

Units of Measurement

- Changing the units of measurement of y or some of x_j **cannot** change the interpretation of the OLS regression line.

  2. Multiply an independent variable x_j by a constant $c \neq 0$

$$
\begin{aligned}
& y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_k x_k + u \\
\rightarrow \quad & y = \beta_0 + \beta_1 x_1 + \cdots + c^{-1}\beta_j (c x_j) + \cdots + \beta_k x_k + u
\end{aligned}
$$

  - ★ The slope on x_j gets divided by c (others do not change)
  - ★ The standard error of $\hat{\beta}_j$ gets divided by c (others do not change)
  - ★ Fitted values $\hat{y}_i$ and residuals $\hat{u}_i$ do not change
  - ★ R^2, t statistics (except t stat. on the new x_j changes sign if $c < 0$) and F statistics do not change

Units of Measurement

- Changing the units of measurement of y or some of x_j **cannot** change the interpretation of the OLS regression line.

  3. For dependent variable $y > 0$, we transform $\log(y)$ into $\log(cy)$ for $c > 0$

$$\log(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

$$\rightarrow \quad \log(cy) = \beta_0 + \log(c) + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

     - ⋆ The intercept in the regression increases by $\log(c)$
     - ⋆ Slopes (and their standard errors) and residuals do not change
     - ⋆ Each fitted values $\hat{y}_i$ increases by $\log(c)$
     - ⋆ R-squared, t and F statistics (except the intercept) do not change

Units of Measurement

- Changing the units of measurement of y or some of x_j **cannot** change the interpretation of the OLS regression line.

  4. For an independent variable $x_j > 0$, we transform $\log(x_j)$ into $\log(cx_j)$

$$
\begin{aligned}
& y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j \log(x_j) + \cdots + \beta_k x_k + u \\
\rightarrow \quad & y = \beta_0 - \beta_j \log(c) + \beta_1 x_1 + \cdots + \beta_j \log(cx_j) \cdots + \beta_k x_k + u
\end{aligned}
$$

  - ⋆ The intercept (and its s.e.) changes but the slopes (and their s.e.) do not
  - ⋆ Fitted values and residuals do not change
  - ⋆ R^2 and test statistics (except for those relating to the intercept) do not change

Units of Measurement

- All previous claims can be shown *algebraically*
  - Check them with data: BWGHT.DTA
    - y: infant birth weight
    - x: number of cigarettes smoked by mother; family income
  - $bwghtlbs = bwght/16$, $packs = cigs/20$, and $famindol = 1,000 \cdot faminc$.
    - `reg bwght cigs faminc`
    - `reg bwghtlbs cigs faminc`
    - `reg bwght packs faminc`
    - `gen famincdol = 1000*faminc`
    - `gen lfamincdol = log(famincdol)`
    - `reg lbwght cigs lfaminc`
    - `reg lbwght cigs lfamincdol`

Units of Measurement

- All previous claims can be shown *algebraically*
- The bottom line: nothing unexpected happens
  - We **cannot** change the importance of an effect, goodness of fit, or statistical inference by changing units of measurement of variables.
- Changing units and then taking logs only changes the intercept.
  - Recall: a change in logs approximates the **relative** change (free of units of measurement)
  - In particular, elasticities are free of units of measurement (units of measurement of x and y when regressing $\log y$ on $\log x$ are irrelevant)

Beta Coefficients

- A **beta coefficient** can be useful (only for interpreting results) when some of the x_j or y have units that are not easily understood
  - Example: how important is a one-point increase in a test score?
- Useful to ask:
  - "How many standard deviations will y change when x_j increases by one standard deviation?"
  - This allows us to see how important an effect is **relative to** the population
- We call such estimates the beta coefficients
  - This is done by standardizing y and each x_j:

$$y_i' = \frac{y_i - \bar{y}}{sd(y)}, \quad x_{ij}' = \frac{x_{ij} - \bar{x}_j}{sd(x_j)}$$

  - Do it by hand or have Stata compute the **beta coefficients**.

Beta Coefficients: Example

- Use ATTEND.DTA:
  - $\hat{\beta}_{priGPA} \approx 5\hat{\beta}_{ACT}$. Does it mean $priGPA$ has a more important effect?
  - 1 sd increase in $priGPA$ increases $\widehat{final}$ by about .222 sds
  - 1 sd increase in ACT increases $\widehat{final}$ by about .297 sds (a larger movement in the distribution of final exam score)
  - Nothing changes in terms of fit or testing

```
reg final skipped priGPA ACT, beta
```

Source	SS	df	MS
Model	3032.09408	3	1010.69803
Residual	12029.853	676	17.7956405
Total	15061.9471	679	22.1825435

Number of obs	=	680
F(3, 676)	=	56.79
Prob > F	=	0.0000
R-squared	=	0.2013
Adj R-squared	=	0.1978
Root MSE	=	4.2185

final	Coefficient	Std. err.	t	P>|t|	Beta
skipped	-.0793386	.0352349	-2.25	0.025	-.0918918
priGPA	1.915294	.372614	5.14	0.000	.2215126
ACT	.4010639	.0532268	7.54	0.000	.2972548
_cons	12.37304	1.171961	10.56	0.000	.

Beta Coefficients: Example

- **Note**: no "beta coefficient" for the intercept (the intercept is zero when all variables have zero sample averages)
- We could also do the calculation by hand.

```
sum final skipped priGPA ACT
```

Variable	Obs	Mean	Std. dev.	Min	Max
final	680	25.89118	4.709835	10	39
skipped	680	5.852941	5.455037	0	30
priGPA	680	2.586775	.5447141	.857	3.93
ACT	680	22.51029	3.490768	13	32

- For example, holding other factors fixed, if $\Delta priGPA = .545$ (one sd),

$$\widehat{\Delta final} = 1.915(.545) = 1.0437$$

- This is equivalent $1.0437/4.7098 \approx .222$ sd of $final$.

Outline

Outline

More on Logarithms

$$\log(y) = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2 + u$$

- Recall
  - β_1: the elasticity of y with respect to x_1
  - $100\beta_2$: approximate percentage change in y when $\Delta x_2 = 1$
- But the approximation may be bad, especially for larger changes.

$$\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 \log(x_1) + \hat{\beta}_2 x_2,$$

- The more precise calculation is

$$\Delta\widehat{\log(y)} = \hat{\beta}_2 \Delta x_2 \quad \Rightarrow \quad \log \frac{\hat{y} + \Delta\hat{y}}{\hat{y}} = \hat{\beta}_2 \Delta x_2$$

$$\%\Delta\hat{y} = 100 \cdot [\exp(\hat{\beta}_2 \Delta x_2) - 1]$$

$$\%\Delta\hat{y} = 100 \cdot [\exp(\hat{\beta}_2) - 1] \qquad \text{if } \Delta x_2 = 1$$

- If $\hat{\beta}_2$ is not "too large", $100[\exp(\hat{\beta}_2) - 1] \approx 100 \cdot \hat{\beta}_2$.

More on Logarithms

- HPRICE2.DTA:

  (log) median house price, pollution and median number of rooms

```
. reg lprice lnox rooms

      Source |       SS           df       MS      Number of obs   =       506
-------------+----------------------------------   F(2, 503)       =    265.69
       Model |  43.4513652         2  21.7256826   Prob > F        =    0.0000
    Residual |  41.1308598       503  .081771093   R-squared       =    0.5137
-------------+----------------------------------   Adj R-squared   =    0.5118
       Total |   84.582225       505  .167489554   Root MSE        =    .28596

------------------------------------------------------------------------------
      lprice | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnox |  -.7176736   .0663397   -10.82   0.000    -.8480106   -.5873366
       rooms |   .3059183   .0190174    16.09   0.000      .268555    .3432816
       _cons |   9.233738   .1877406    49.18   0.000     8.864885     9.60259
------------------------------------------------------------------------------

. * increase
. di 100*(exp(.306)-1)
35.798231

. * The same as
. di 100*(exp(_b[rooms])-1)
35.787137

. * decrease
. di 100*(exp(-.306)-1)
-26.361338
```

More on Logarithms

- HPRICE2.DTA:

  (log) median house price, pollution and median number of rooms

  - More precise percentage change of median house price due to adding one more room is 35.8%
  - $100\hat{\beta}_2\% = 30.6\%$: imprecise approximate, but it is between the two estimates (for increase and decrease)

More on Logarithms

- **Reasons for Using the Natural Log**

  1. The coefficients have percentage change interpretations (units of measurement of these variables are irrelevant)
  2. When $y > 0$, models with $\log(y)$ as the dependent variable often more closely satisfy the classical linear model assumptions such as normality
  3. In most cases, taking the log greatly reduces the variability of a variable, making OLS estimates less sensitive to outlier (or extreme) values

More on Logarithms

- **Limitations of Using the Natural Log**

    1. If $y \geq 0$ but $y = 0$ is possible, we cannot use $\log(y)$. Sometimes $\log(1 + y)$ is used, but interpreting the coefficients is difficult, and $\log(1 + y) \geq 0$ if $y \geq 0$.

    2. It is harder to predict y when we have estimated a model for $\log(y)$.

    3. In cases where y is a fraction and close to zero for many observations, $\log(y)$ can have *more* variability than y.

More on Logarithms

- **Some (Not-so-Hard) Rules on Using Logarithms**

  1. Logs are often used for dollar amounts that are always positive, as well as for variables such as population, especially when there is a lot of variation.
  2. Logs are used less often for variables measured in years, such as schooling, age, and experience.
  3. Logs are used less infrequently for variables that are already percents or proportions (e.g., unemployment rate)
     - Careful: percentage point change (use y) and percentage change (use $\log y$)
     - An increase from 8 to 9 is 1 percentage point change, but 12.5% percentage change ($\log 9 - \log 8 \approx .118$)
  4. Do not compare R^2 from regressing y and regressing $\log y$

Outline

Models with Quadratics

- Reg y on $\log x$: x has a diminishing effect (log function is concave).
- But sometimes it is not flexible enough.
- Models with **quadratics**
  - deliver increasing or decreasing effects
  - contain the constant effect as a special case, which can be easily tested
  - allow for a turning point, which may be of interest.
    - For example, what is the optimal number of students at a high school for high school performance?

Models with Quadratics

- Consider the model

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

  - One single explanatory variable x
  - But two **regressors**, $x_1 = x$ and $x_2 = x^2$

- The slope of y with respect to x depends on β_1 and β_2, and the value of x:

$$\frac{dy}{dx} = \beta_1 + 2\beta_2 x \qquad \text{(holding } u \text{ fixed)}$$

- Estimation is straightforward: just define a new variable x^2, and include it along with x as a regressor.

Models with Quadratics

- Given the estimated coefficients,

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

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

1. $\hat{\beta}_2 < 0$: the slope is initially positive but decreases as x increases. The function has a hump shape
2. $\hat{\beta}_2 > 0$: the slope is initially negative but increases as x increases. The function has U-shaped

- The turning point is

$$x^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

- OLS calculation doesn't change; just be careful about the interpretation

Models with Quadratics

- **EXAMPLE:** A log($wage$) equation with $exper^2$ (WAGE1.DTA)

$$\begin{aligned} \widehat{lwage} &= \underset{(.106)}{0.128} + \underset{(.007)}{.090}\, educ + \underset{(.005)}{.041}\, exper - \underset{(.0001)}{.0007}\, exper^2 \\ n &= 526,\ R^2 = .300 \end{aligned}$$

  - Estimated return to education $\approx 9.0\%$ (the model **assumes** this is the same for all years of experience and education)
  - Each year of experience is worth less than the preceding year

    Partial effect of $exper$ (taking derivatives):

$$\frac{\Delta \widehat{lwage}}{\Delta exper} \approx .041 - 2(.0007)exper = .041 - .0014\ exper$$

  - 4.1% is approximately the return to the 1st year of experience;

    The return from 10 to 11 is about

$$.041 - .0014 \times (10) = .027 \qquad \Leftrightarrow \qquad 2.7\%.$$

Models with Quadratics

- **EXAMPLE:** A log($wage$) equation with $exper^2$ (WAGE1.DTA)

$$\begin{aligned}\widehat{lwage} &= \underset{(.106)}{0.128} + \underset{(.007)}{.090}\, educ + \underset{(.005)}{.041}\, exper - \underset{(.0001)}{.0007}\, exper^2 \\ n &= 526,\ R^2 = .300\end{aligned}$$

  - Partial effect of $exper$: to be more precise, not use a calculus approximation

    Return from 10 to 11

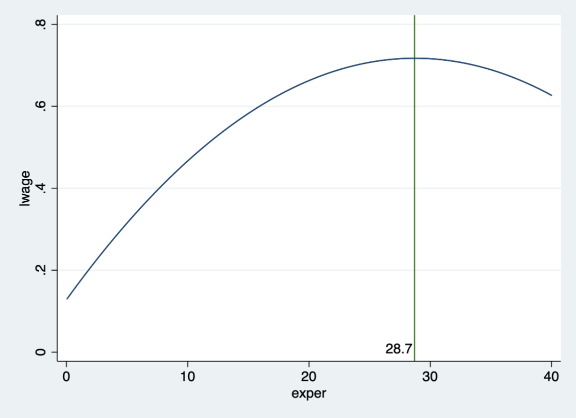
$$[.041(11) - .0007(11)^2] - [.041(10) - .0007(10)^2] \approx .026 \quad \Leftrightarrow \quad 2.6\%$$

  - Do the exact calculation for larger changes in $exper$.

Models with Quadratics

```
. gen exper2=exper^2

. reg lwage educ exper exper2

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(3, 522)       =     74.67
       Model |  44.5393713         3  14.8464571   Prob > F        =    0.0000
    Residual |   103.79038       522  .198832146   R-squared       =    0.3003
-------------+----------------------------------   Adj R-squared   =    0.2963
       Total |  148.329751       525   .28253286   Root MSE        =    .44591

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0903658    .007468    12.10   0.000     .0756948    .1050368
       exper |   .0410089   .0051965     7.89   0.000     .0308002    .0512175
      exper2 |  -.0007136   .0001158    -6.16   0.000     -.000941   -.0004861
       _cons |   .1279975   .1059323     1.21   0.227    -.0801085    .3361035
------------------------------------------------------------------------------

. di _b[exper]+2*_b[exper2]*10
.02673771

. di (_b[exper]*11+_b[exper2]*11^2)-(_b[exper]*10+_b[exper2]*10^2)
.02602415
```

Models with Quadratics



Models with Quadratics

- The curve turns at about $exper^* = .041/[2 \cdot (.000714)] \approx 28.7$.
  - About 23% of the observations have $exper > 29$
- Quadratic model is more complicated to interpret
  - We need good statistical evidence for keeping x^2 (e.g., $exper^2$)
  - Use a t-test

$$H_0 : \beta_{exper^2} = 0 \qquad vs. \qquad H_1 : \beta_{exper^2} \neq 0$$

  - t-ratio $= -6.16$: reject H_0

Models with Quadratics

- The curve turns at about $exper^* = .041/[2 \cdot (.000714)] \approx 28.7$.
  - About 23% of the observations have $exper > 29$
- We already know $exper$ affects $lwage$. But if we did want to test

$$\begin{aligned} H_0 &: \quad exper \text{ has no effect on } lwage \\ H_1 &: \quad exper \text{ does have an effect on } lwage \end{aligned}$$

  - This would be

$$H_0\text{: } \beta_{exper} = 0,\ \beta_{exper^2} = 0.$$

  - Use an F test. But usually, if the quadratic term is insignificant, we go back to a linear model.

Outline

Models with Interaction Terms

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- β_1: partial effect of x_1 on y; β_2: partial effect of x_2 on y
- **Important restriction**: effect of x_1 never depends on x_2 (vice versa)
- Sometimes we expect the partial effect of one variable (e.g., education) depends on another variable (e.g., intelligence)
- Solution: add an **interaction term**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

$$\frac{\Delta y}{\Delta x_1} = \beta_1 + \beta_3 x_2, \quad \frac{\Delta y}{\Delta x_2} = \beta_2 + \beta_3 x_1$$

  - $H_0 : \beta_3 = 0$ means the partial effects are constant. It should be tested.

Models with Interaction Terms

$$\frac{\Delta y}{\Delta x_1} = \beta_1 + \beta_3 x_2$$

- β_1: the partial effect (PE) of x_1 on y when $x_2 = 0$.
  - But $x_2 = 0$ may be far from a legitimate, or interesting part of population
- Two interesting parameters: PEs evaluated at mean of the other variable

$$\begin{aligned} \delta_1 &= \beta_1 + \beta_3 \mu_2 \qquad && (\mu_2 = \mathbb{E}[x_2]) \\ \delta_2 &= \beta_2 + \beta_3 \mu_1 \qquad && (\mu_1 = \mathbb{E}[x_1]) \end{aligned}$$

  - Estimates: $\hat{\delta}_1 = \hat{\beta}_1 + \hat{\beta}_3 \bar{x}_2$, $\hat{\delta}_2 = \hat{\beta}_2 + \hat{\beta}_3 \bar{x}_1$
  - Alternative implementation: rewriting the model

$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u$$

    - Reg y_i on 1, x_{i1}, x_{i2}, $(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)$, and s.e. are obtained as well

Models with Interaction Terms

- **EXAMPLE:** Does the effect of attending classes depend on priGPA?

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u$$

  - $stndfnl$: standardized final score; $atndrte$: percentage of classes attended

```
. reg stndfnl atndrte priGPA ACT

      Source |       SS           df       MS      Number of obs   =       680
-------------+----------------------------------   F(3, 676)       =     56.79
       Model |  133.822385         3  44.6074616   Prob > F        =    0.0000
    Residual |  530.941183       676  .785415951   R-squared       =    0.2013
-------------+----------------------------------   Adj R-squared   =    0.1978
       Total |  664.763568       679  .979033237   Root MSE        =    .88624

------------------------------------------------------------------------------
     stndfnl | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     atndrte |   .0053337   .0023687     2.25   0.025     .0006827    .0099846
      priGPA |   .4023727   .0782803     5.14   0.000      .248671    .5560744
         ACT |   .0842571   .0111821     7.54   0.000     .0623013    .1062129
       _cons |  -3.343655   .2990985   -11.18   0.000    -3.930929   -2.756381
------------------------------------------------------------------------------
```

Models with Interaction Terms

- **EXAMPLE:** Does the effect of attending classes depend on priGPA?

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA \cdot atndrte + u$$

  - $stndfnl$: standardized final score; $atndrte$: percentage of classes attended

```
. gen atnpriGPA=atndrte*priGPA

. reg stndfnl atndrte priGPA ACT atnpriGPA
```

Source	SS	df	MS
Model	140.819497	4	35.2048742
Residual	523.944071	675	.776213439
Total	664.763568	679	.979033237

Number of obs	=	680
F(4, 675)	=	45.35
Prob > F	=	0.0000
R-squared	=	0.2118
Adj R-squared	=	0.2072
Root MSE	=	.88103

stndfnl	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
atndrte	-.0208926	.0090469	-2.31	0.021	-.0386561	-.0031291
priGPA	-.5544979	.3280652	-1.69	0.091	-1.198649	.0896531
ACT	.0816979	.011149	7.33	0.000	.059807	.1035889
atnpriGPA	.0114617	.0038175	3.00	0.003	.0039661	.0189573
_cons	-1.135889	.7931751	-1.43	0.153	-2.693276	.421498

Models with Interaction Terms

- **EXAMPLE:** Does the effect of attending classes depend on priGPA?

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 atnpriGPA0 + u$$

  - $stndfnl$: standardized final score; $atndrte$: percentage of classes attended

```
reg stndfnl atndrte priGPA ACT atnpriGPA0
```

Source	SS	df	MS
Model	140.819498	4	35.2048744
Residual	523.94407	675	.776213437
Total	664.763568	679	.979033237

Number of obs	=	680
F(4, 675)	=	45.35
Prob > F	=	0.0000
R-squared	=	0.2118
Adj R-squared	=	0.2072
Root MSE	=	.88103

stndfnl	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
atndrte	.0087588	.0026166	3.35	0.001	.0036212	.0138964
priGPA	.3819215	.0781179	4.89	0.000	.2285382	.5353047
ACT	.0816979	.011149	7.33	0.000	.059807	.1035889
atnpriGPA0	.0114617	.0038175	3.00	0.003	.0039661	.0189573
_cons	-3.558406	.3058231	-11.64	0.000	-4.158885	-2.957927

$$atnpriGPA0 = (atndrte - 81.7)(priGPA - 2.587)$$

Models with Interaction Terms

- The interaction term is statistically significant (p-value $= .003$).
  - β_1 is hard to interpret: "1 percentage point increase in attendance for someone with $priGPA = 0$ decreases the score by 0.02 sd."
- But if we use $(priGPA - \overline{priGPA})(atndrte - \overline{atndrte})$
  - β_1 is the partial effect of attendance for those with an **average** $priGPA$
  - 10 percentage points increase in attendance increases final score by 0.088 sd
- Positive coefficient on the interaction: return to attending classes is higher for people with higher prior GPA.

$$\frac{\Delta \widehat{stndfnl}}{\Delta \widehat{atndrte}} \approx .0088 + .011(priGPA - 2.587)$$

  - If $priGPA$ is 1.0 above its mean value, the return to attendance is $.0088 + .011 = .0198$

Outline

Outline

Adjusted R-Squared

- **Nested model**:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad\qquad\qquad + u$$

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \beta_{k+1} x_{k+1} + \cdots + u$$

  - One is a special case of the other
  - Recall: the usual R^2 never decreases (usually increases), when one or more variables are added to a regression (if no observations are lost)
  - Use t test to decide if we want to include a single new variable
  - Use F test to decide if we want to add a group of new variables

- But sometimes we want to compare **nonnested** models (neither is a special case of the other). For example,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1} + \beta_{k+1} x_{k+1} + u$$

Adjusted R-Squared

- Goal: have a goodness-of-fit measure that penalizes adding additional explanatory variables. (The usual R^2 has no penalty!)
- Recall:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{(SSR/n)}{(SST/n)},$$

  - SSR/n: estimate $\sigma_u^2 = \mathbb{V}[u]$; SST/n: estimate $\sigma_y^2 = \mathbb{V}[y]$ (consistent but not unbiased estimators)
  - **Population R-squared**: the amount of population variation in y explained by $x_1, ..., x_k$.

$$\rho^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}$$

- **Adjusted *R*-squared** (also called "R-bar-squared"):
  - use $SSR/(n-k-1)$ and $SST/(n-1)$ as the unbiased estimators

Adjusted R-Squared

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

- With more regressors: SSR falls, but so does $df = n - k - 1$.
  - $\bar{R}^2$ can increase or decrease
  - For $k \geq 1$, $\bar{R}^2 < R^2$ unless $SSR = 0$
  - For small n and large k, $\bar{R}^2$ can be much smaller than R^2
  - It is possible that $\bar{R}^2 < 0$, especially if df is small. (But $R^2 \geq 0$ always)
  - Important: the R-squared form of the F statistic uses the usual R-squared, not the adjusted R-squared

Adjusted R-Squared: Final Remarks

- We *do not* emphasize goodness of fit because focusing on making R^2 or $\bar{R}^2$ as large as possible can lead to silly mistakes
- In most economic applications we care more about the causal interpretation
- Better to ask
  - Does adding a particular variable reduce the bias, or in other words, does omitting it cause bias?

Outline

Controlling for Too Many Factors in Regression

- Remember the ceteris paribus interpretation of regression.
  - Sometimes it **does not** make sense to hold other factors fixed when studying the effect of a particular variable.
- **Example:** effect of spending per student on math pass rate (MEAP93.DTA)

$$\Delta\widehat{math10} = (6.23/100)\%\Delta spend \approx .06(\%\Delta spend)$$

  - If spending increases by 10%, pass rate increases by 0.6 percentage point
  - Now add the teacher-student ratio, $staff$, and log of the average teacher salary, $lsalary$
  - The coefficient on $lspend$ is negative (but not very statistically different from zero). Does spending no longer matter?

Controlling for Too Many Factors in Regression

```
reg math10 lexpend lnchprg
```

Source	SS	df	MS
Model	8063.82429	2	4031.91215
Residual	36753.3562	405	90.7490276
Total	44817.1805	407	110.115923

Number of obs	=	408
F(2, 405)	=	44.43
Prob > F	=	0.0000
R-squared	=	0.1799
Adj R-squared	=	0.1759
Root MSE	=	9.5262

math10	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lexpend	6.22969	2.972634	2.10	0.037	.3859705	12.07341
lnchprg	-.3045853	.0353574	-8.61	0.000	-.3740923	-.2350783
_cons	-20.36075	25.07288	-0.81	0.417	-69.64998	28.92848

```
reg math10 lexpend lnchprg lsalary staff
```

Source	SS	df	MS
Model	8559.25797	4	2139.81449
Residual	36257.9225	403	89.9700311
Total	44817.1805	407	110.115923

Number of obs	=	408
F(4, 403)	=	23.78
Prob > F	=	0.0000
R-squared	=	0.1910
Adj R-squared	=	0.1830
Root MSE	=	9.4853

math10	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lexpend	-20.05237	11.69585	-1.71	0.087	-43.04487	2.940139
lnchprg	-.2801704	.0384148	-7.29	0.000	-.3556888	-.204652
lsalary	26.32343	11.22514	2.35	0.020	4.25628	48.39058
staff	.2510484	.1140019	2.20	0.028	.0269357	.475161
_cons	-98.81947	44.03569	-2.24	0.025	-185.3878	-12.25113

Controlling for Too Many Factors in Regression

- It is tempting to **over control** because often R^2 or $\bar{R}^2$ increases.
- In the previous example,
  - Why should we control for the teacher-student ratio and teacher's salary?
  - We want to allow spending to increase these variables
  - Once we hold those fixed, the role of spending is limited
    - Spending other than to affect these two variables has no effect on performance. But this does *not* mean total spending has no effect.
  - Do not include these factors unless we want to recover some effect of spending that works not through teacher-student ratio and salary
- Different models for different purposes: focus on the ceteris paribus interpretation! (more on this when we discuss program evaluation)