Class 19-20 MRM Further Issues

Pedram Jahangiry

JON M. HUNTSMAN SCHOOL OF BUSINESS
UtahStateUniversity

CHAPTER 6

Multiple Regression Analysis:

Further Issues

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

Effects of Data Scalling on OLS Statistics

If the **dependent variable** y is multiplied by a constant α , then:

- ✓ All the estimated coefficients are multiplied by α
- ✓ Residuals and so SER are multiplied by α
- ✓ t-stat, F-stat and R^2 are not affected (why?)

$$y = \beta_0 + \frac{1}{\alpha}\beta_1 \alpha x_1 + \beta_2 x_2 + u$$

If an **independent variable** x_j is multiplied by a constant α , then:

- ✓ Only the $\widehat{\beta}_j$ is multiplied by $\frac{1}{\alpha}$
- ✓ Residuals and so SER are not affected!
- ✓ **t-stat, F-stat** and R^2 are not affected (why?)

More on Functional Form

$\log(Price) = \beta_0 + \beta_1 \log(zest) + \dots$

1000

/1000

100

/100

$\log(\alpha x) = \log \alpha + \log(x)$

More on using logarithmic functional forms

- Convenient percentage/elasticity interpretation
- Slope coefficients of logged variables are invariant to rescalings
- Taking logs often eliminates/mitigates problems with outliers
- Taking logs often helps to secure normality and homoskedasticity
- Logs must not be used if variables take on zero or negative values

$\log(\alpha y) = \log \alpha + \log(y) =$

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

α

Using quadratic functional forms

Quadratic functions are used quite often in applied economics to capture decreasing or increasing marginal effects.

Example: Wage equation

Concave experience profile

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

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

```
> MRM <- lm(wage ~ exper + I(exper^2), wage1)
> stargazer(MRM, type = "text")

===============================================
                        Dependent variable:
                    ---------------------------
                               wage
-----------------------------------------------
exper                        0.298***
                             (0.041)

I(exper^2)                  -0.006***
                             (0.001)

Constant                     3.725***
                             (0.346)

-----------------------------------------------
Observations                   526
R2                            0.093
Adjusted R2                   0.089
Residual Std. Error      3.524 (df = 523)
F Statistic          26.740*** (df = 2; 523)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

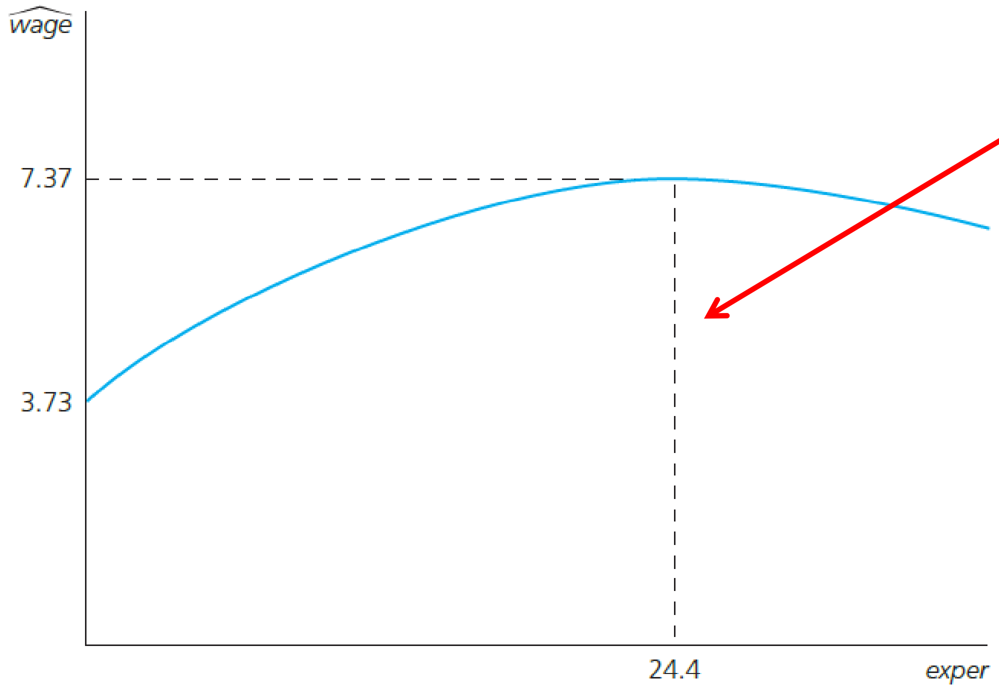
Marginal effect of experience:

$$\boxed{\frac{\Delta wage}{\Delta exper} = .298 - 2(.0061) exper}$$

The first year of experience increases the wage by some $.298, the second year by .298 - 2(.0061)(1) = $.286 etc.

Example: Wage equation (cont'd)

Wage **maximum** with respect to work experience



$$x^* = \left|\frac{\widehat{\beta}_1}{2\widehat{\beta}_2}\right| = \left|\frac{.298}{2(.0061)}\right| \approx 24.4$$

Does this mean the return to experience becomes negative after 24.4 years?

Not necessarily. It depends on how many observations in the sample lie right of the turnaround point.

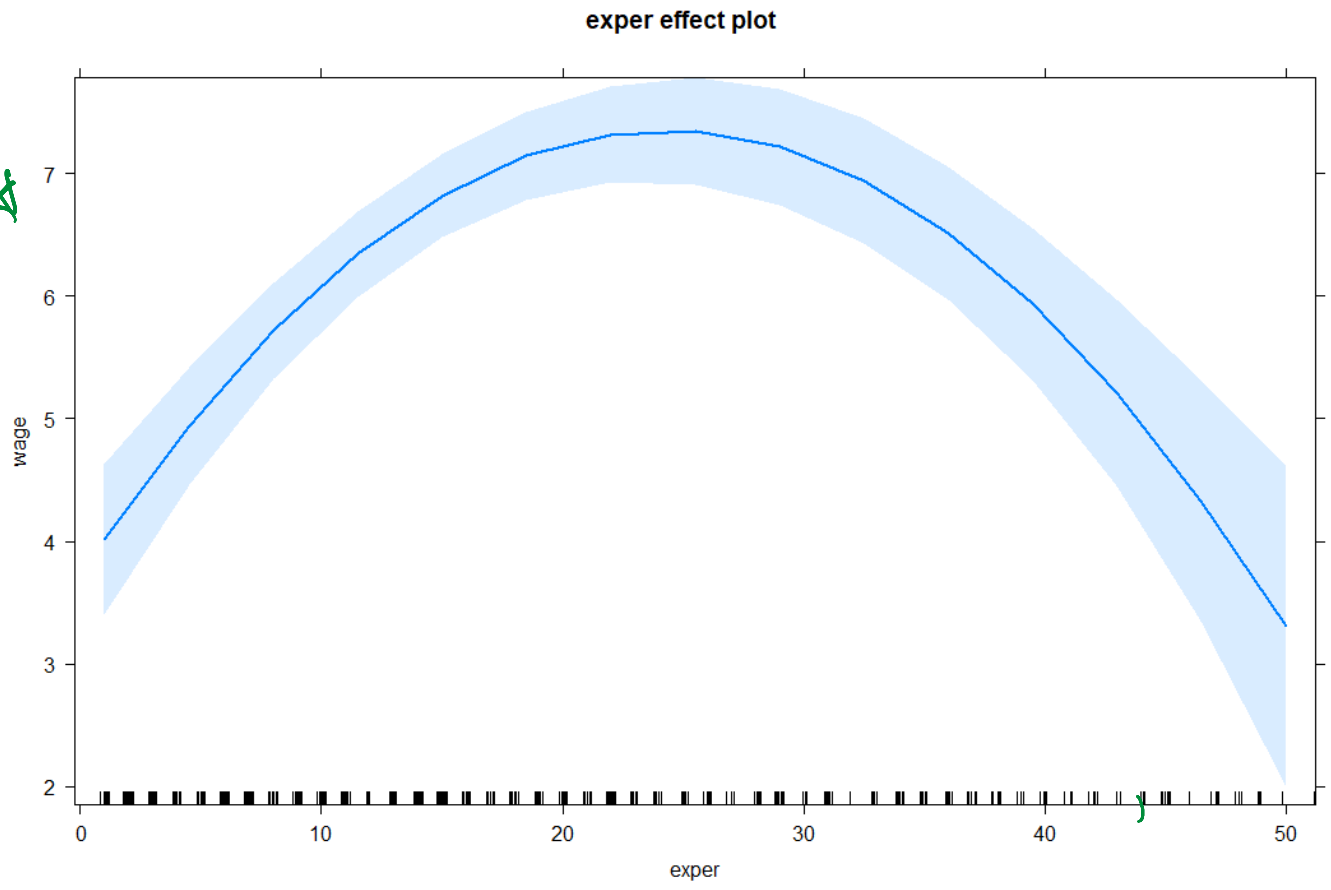
In the given example, these are about 28% of the observations. There may be a specification problem! (e.g. omitted variables).

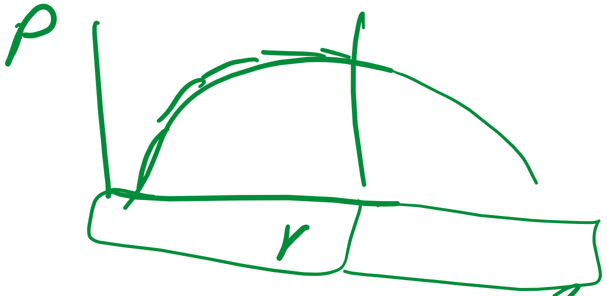
Effect plot in R (optional!)

exper : age

```
MRM <- lm(wage ~ exper + I(exper^2), wage1)

# install.packages("effects")
library(effects)
plot(effect("exper", MRM))
```





Example: Effects of pollution on housing prices

Nitrogen oxide in the air, distance from employment centers, average student/teacher ratio

$$\widehat{\log}(price) = \underset{(.57)}{13.39} - \underset{(.115)}{.902}\ \log(nox) - \underset{(.043)}{.087}\ \log(dist)$$

$$- \underset{(.165)}{.545}\ \boxed{rooms} + \boxed{\underset{(.013)}{.062}\ rooms^2} - \underset{(.006)}{.048}\ stratio$$

$$n = 506,\ R^2 = .603$$

$$\frac{d \log price}{d\ rooms} = -0.545 - 2(0.062) \times rooms$$

```
> MRM <- lm(log(price)~log(nox)+log(dist)+rooms+I(rooms^2)+stratio,data=hprice2)
> stargazer(MRM, type = "text")

===============================================
                        Dependent variable:
                    ---------------------------
                             log(price)
-----------------------------------------------
log(nox)                     -0.902***
                              (0.115)

log(dist)                    -0.087**
                              (0.043)

rooms                        -0.545***
                              (0.165)

I(rooms2)                     0.062***
                              (0.013)

stratio                      -0.048***
                              (0.006)

Constant                     13.385***
                              (0.566)

-----------------------------------------------
Observations                    506
R2                             0.603
Adjusted R2                    0.599
Residual Std. Error      0.259 (df = 500)
F Statistic          151.770*** (df = 5; 500)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Example: Effects of pollution on housing prices (cont'd)

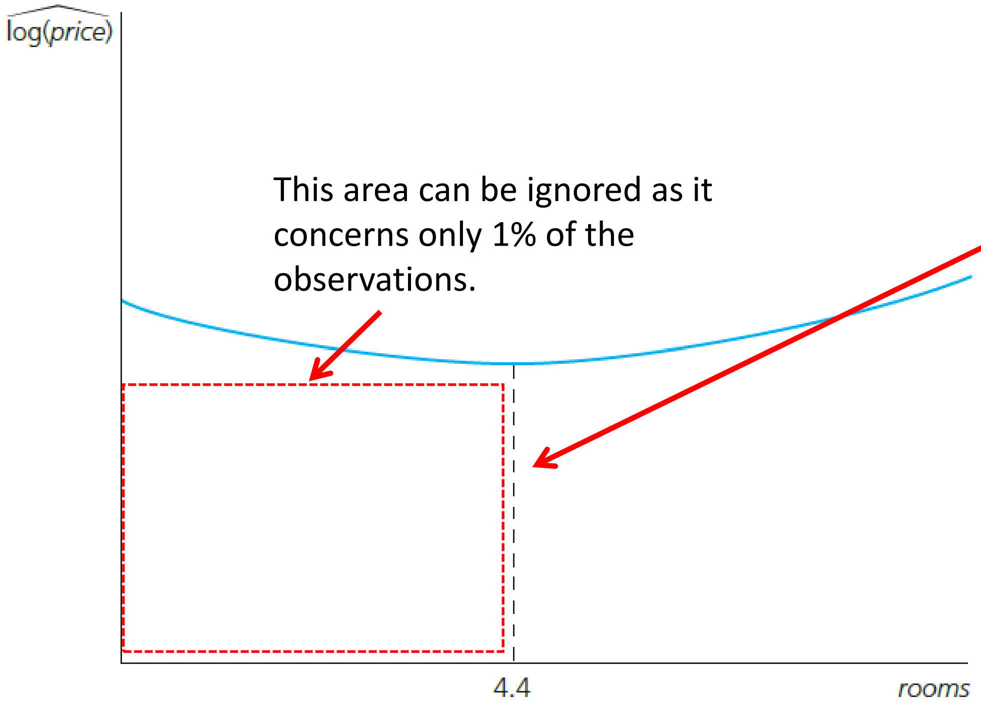
$$\Rightarrow \frac{\Delta \log(price)}{\Delta rooms} = \frac{\%\Delta price}{\Delta rooms} = -.545 + .124 rooms$$

Does this mean that, at a low number of rooms, more rooms are associated with lower prices?

$$Y = x - 2x^2$$

$$\frac{\partial y}{\partial x} = 1 - 4x = 0$$

$$x = 0.25$$



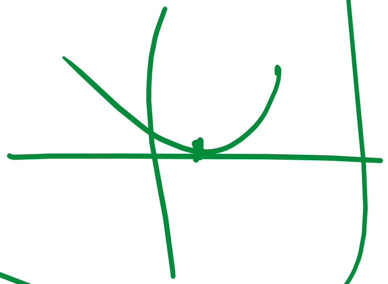
$$x^* = \left|\frac{-.545}{2(.062)}\right| \approx 4.4$$

Increase rooms from 5 to 6:

$$-.545 + .124(5) = +7.5\%\ price$$

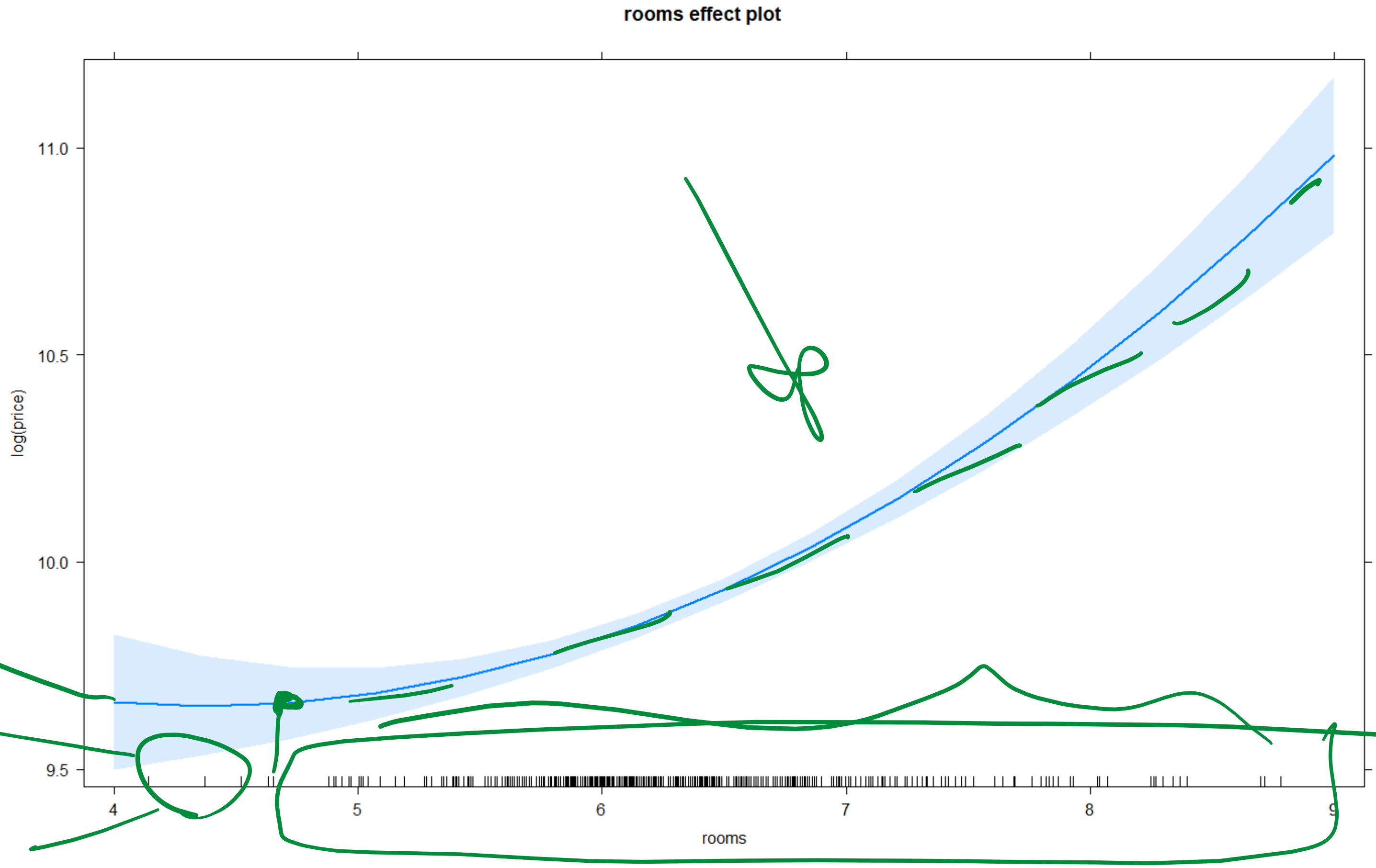
Increase rooms from 6 to 7:

$$-.545 + .124(6) = +19.9\%\ price$$



Effect plot in R (optional!)

```
MRM <- lm(log(price)~log(nox)+log(dist)+rooms+I(rooms^2)+stratio,data=hprice2)
plot(effect("rooms", MRM))
```



More examples:

Other possibilities

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 [\log(nox)]^2$$

$$+\beta_3 crime + \beta_4 rooms + \beta_5 rooms^2 + \beta_6 stratio + u$$

$$\Rightarrow \quad \frac{\Delta \log(price)}{\Delta \log(nox)} = \frac{\%\Delta price}{\%\Delta nox} = \beta_1 + 2\beta_2[\log(nox)]$$

Higher polynomials

$$cost = \beta_0 + \beta_1 quantity + \beta_2 quantity^2 + \beta_3 quantity^3 + u$$

Models with interaction terms

Interaction terms allow the partial effect of an explanatory variable, say x_1 , to depend on the level of another variable, say x_2—and vice versa.

Interaction term

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \boxed{x_1 \, x_2} + u$$

Example:

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

$$\frac{\Delta price}{\Delta bdrms} = \beta_2 + \beta_3 sqrft$$ → The effect of the number of bedrooms depends on the level of square footage

- Interaction effects complicate interpretation of parameters

$\beta_2 =$ Effect of number of bedrooms, but for a **square footage of zero!!**

Example: Effects of Attendance on Final Exam Performance

A model to explain the standardized outcome on a **final exam** in terms of **percentage of classes attended**, **prior college GPA**, and **ACT** score is:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 + \beta_5 ACT^2 + \beta_6 priGPA \cdot atndrte + u.$$

We are interested in the effects of attendance on final exam score: $\Delta stndfnl/\Delta atndrte = \beta_1 + \beta_6 priGPA$

The idea is that class attendance might have a **different** effect for students who have performed differently in the past.

$$\begin{aligned}\widehat{stndfnl} = \underset{(1.36)}{2.05} - \underset{(.0102)}{.0067}\, atndrte - \underset{(.48)}{1.63}\, priGPA - \underset{(.098)}{.128}\, ACT \\ + \underset{(.101)}{.296}\, priGPA^2 + \underset{(.0022)}{.0045}\, ACT^2 + \underset{(.0043)}{.0056}\, priGPA \cdot atndrte \\ n = 680,\ R^2 = .229,\ \bar{R}^2 = .222.\end{aligned}$$

Attendance has negative effect on final exam score? What's going on?

We must plug in interesting values of *priGPA* to obtain the partial effect. The mean value of *priGPA* in the sample is 2.59, so at the mean *priGPA*, the effect of *atndrte* on *stndfnl* is $-.0067 + .0056(2.59) \approx .0078$

What does this mean? Because *atndrte* is measured as a percentage, it means that a 10 percentage point increase in *atndrte* increases *stndfnl* by .078 from the mean final exam score.

```
MRM <-lm(stndfnl~ atndrte+ priGPA + ACT
         + I(priGPA^2) + I(ACT^2)
         + atndrte:priGPA
         , data=attend)

stargazer(MRM, type = "text")
```

Interaction term

```
==============================================
                      Dependent variable:
                  ----------------------------
                            stndfnl
----------------------------------------------
atndrte                     -0.007
                            (0.010)

priGPA                    -1.629***
                            (0.481)

ACT                         -0.128
                            (0.098)

I(priGPA2)                 0.296***
                            (0.101)

I(ACT2)                     0.005**
                            (0.002)

atndrte:priGPA               0.006
                            (0.004)

Constant                     2.050
                            (1.360)

----------------------------------------------
Observations                  680
R2                           0.229
Adjusted R2                  0.222
Residual Std. Error     0.873 (df = 673)
F Statistic         33.250*** (df = 6; 673)
==============================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

Testing if partial effect is significant?

$$H_0: \quad \beta_1 + 2.59\,\beta_6 = 0$$

```
linearHypothesis(MRM, c("atndrte+ 2.59 * atndrte:priGPA"))
```

```
Linear hypothesis test

Hypothesis:
atndrte  + 2.59 atndrte:priGPA = 0

Model 1: restricted model
Model 2: stndfnl ~ atndrte + priGPA + ACT + I(priGPA^2) + I(ACT^2) + atndrte:priGPA

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    674 519.34
2    673 512.76  1    6.5772 8.6326 0.003415 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Average partial effects (APE)

- In models with quadratics, interactions, and other nonlinear functional forms, the partial effect depend on the values of one or more explanatory variables

- Average partial effect (APE) tells us **the size of partial effect on average!**

- After computing the partial effect and plugging in the estimated parameters, average the partial effects for each unit across the sample

In the previous example: $APE_{stdndfnl} = \widehat{\beta_1} + \widehat{\beta_6}\ \ \overline{priGPA}$

More on goodness-of-fit and selection of regressors

- General remarks on R-squared
  - A high R-squared does **not** imply that there is a **causal interpretation**
  - A low R-squared does **not** preclude precise estimation of partial effects

- Adjusted R-squared
  - What is the ordinary R-squared supposed to measure?

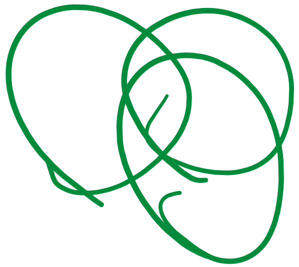
$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{(SSR/n)}{(SST/n)}$$ is an **estimate** for $$\boxed{1 - \frac{\sigma_u^2}{\sigma_y^2}}$$

Population R-squared

Adjusted R-squared (cont.)

$R^2 \uparrow$

$k \uparrow$



$R^2 \uparrow$ fact

$k \uparrow$

$\bar{R}^2$

A better estimate taking into account degrees of freedom would be

Correct degrees of freedom of numerator and denominator

$$\bar{R}^2 = 1 - \frac{(SSR/\boxed{(n-k-1)})}{(SST/\boxed{(n-1)})} = adjusted\ R^2$$

✓ The adjusted R-squared imposes a penalty for adding new regressors

✓ The adjusted R-squared increases if, and only if, the t-statistic of a newly added regressor is greater than one in absolute value

Relationship between R-squared and adjusted R-squared:

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

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

The adjusted R-squared may even get negative

$\beta_2 = \beta_3 = 0$

Using adjusted R-squared to choose between nonnested models

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_4 + u$ (1) vs $\Rightarrow \beta_2 + \beta_1 x_1$ (2)

Models are **nonnested** if neither model is a special case of the other:

F

$$rdintens = \beta_0 + \beta_1 \log(sales) + u \quad \leftarrow \quad R^2 = .061, \bar{R}^2 = .030$$

$$rdintens = \beta_0 + \beta_1 sales + \beta_2 sales^2 + u \quad \leftarrow \quad R^2 = .148, \bar{R}^2 = .090$$

- A comparison between the R-squared of both models would be unfair to the first model because the first model contains fewer parameters
- In the given example, even after adjusting for the difference in degrees of freedom, the quadratic model is preferred.

❑ **F-statistic** is used to choose between **nested** models. (Restricted and Unrestricted models)

$\bar{R}^2$

Adding regressors to reduce the error variance

- Adding regressors may excarcerbate multicollinearity problems
- On the other hand, adding regressors reduces the error variance
- Variables that are uncorrelated with other regressors should be added because they reduce error variance without increasing multicollinearity
- However, such uncorrelated variables may be hard to find!

Confidence Interval for Predictions

$\hat{y}$ new

Suppose we want a 95% CI for the future college GPA of a high school student with

colgpa

100 new students

$$sat = 1{,}200 \quad , \quad hsperc = 30 \quad , \quad hsize = 5.$$

$$\widehat{colgpa} = \underset{(0.075)}{1.493} + \underset{(.00007)}{.00149}\, sat - \underset{(.00056)}{.01386}\, hsperc - \underset{(.01650)}{.06088}\, hsize + \underset{(.00227)}{.00546}\, hsize^2$$

$$n = 4{,}137,\ R^2 = .278,\ \bar{R}^2 = .277,\ \hat{\sigma} = .560$$

```r
# confidence Intervals for predictions

MRM <- lm(colgpa~sat+hsperc+hsize+I(hsize^2),gpa2)

# Define sets of regressor variables
xvalues <- data.frame(sat=1200, hsperc=30, hsize=5)

# Point estimates and 95% prediction intervals for these
predict(MRM, xvalues, interval = "prediction", level = 0.95)
```

```
> predict(MRM, xvalues, interval = "prediction", level = 0.95)
       fit      lwr      upr
1 2.700075 1.601749 3.798402
```