Econometrics

Geometric Interpretation of OLS, Non-linearities and Omitted Variable Bias

April 4, 2017

Today's plan

- Introduction to non-linearities
- Omitted Variable Bias (OVB) and Multiple Regression

Non-linearities. Motivation

Non-linearities. Motivation

- Linear estimation is easy and handy, but not always the best option.

Non-linearities. Motivation

- Linear estimation is easy and handy, but not always the best option.
- Suppose that you want to estimate the effect of the tax rate on tax revenue.

Non-linearities. Motivation

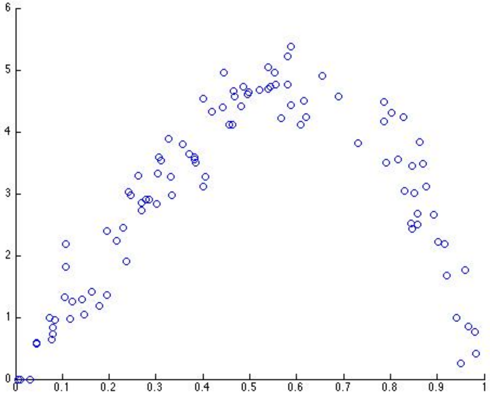
- Linear estimation is easy and handy, but not always the best option.
- Suppose that you want to estimate the effect of the tax rate on tax revenue.
- Is a linear estimation as the one following a good starting point? why, why not?

$$REV_c = \beta_0 + \beta_1 TAXRATE_c + u_c$$

note: subscript c stands for country

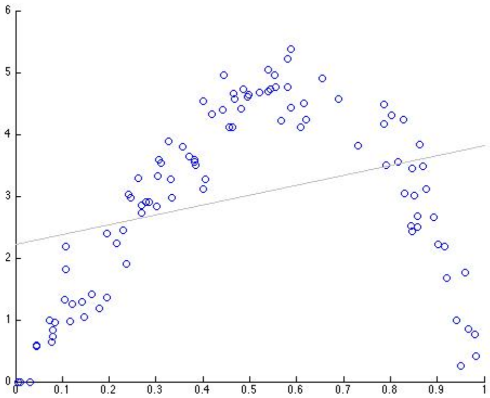
Laffer Curve (note: simulated data)

x-axis: tax rate per country. y-axis: Invented Revenue in billions of USD

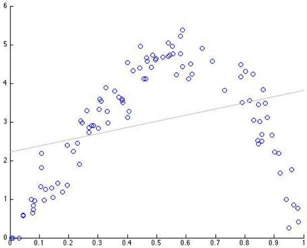


Laffer Curve (note: simulated data)

x-axis: tax rate per country. y-axis: Invented Revenue in billions of USD

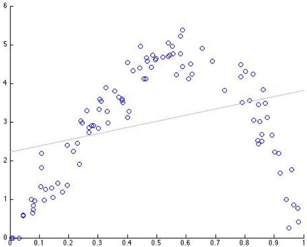


Laffer Curve (note: simulated data)



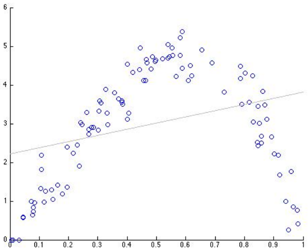
$\hat{\beta}_1 = 1.59$
95% confidence interval = [0.62,2.55]

Laffer Curve (note: simulated data)



$\hat{\beta}_1 = 1.59$
95% confidence interval = [0.62,2.55]
Interpretation?

Laffer Curve (note: simulated data)



$\hat{\beta}_1 = 1.59$

95% confidence interval = [0.62,2.55]

Interpretation? Increasing tax rate increases revenue. At *all* levels.

- Common sense (and lots of data) tells us that cannot be true.
- True model is

$$REV_c = \beta_0 + f(TAXRATE_c) + u_c$$

- where the shape of f is to be estimated
- How can we estimate f?

Semi-parametric estimation

- One option: semiparametric estimation. Very roughly speaking, divide the sample into N blocks, and run OLS for each of those N blocks.

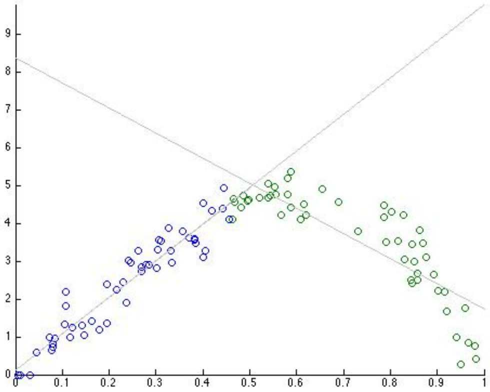
Semi-parametric estimation

- One option: semiparametric estimation. Very roughly speaking, divide the sample into N blocks, and run OLS for each of those N blocks.
- Then, put together all the 'lines' resulting from all different N estimations.

Semi-parametric estimation

- One option: semiparametric estimation. Very roughly speaking, divide the sample into N blocks, and run OLS for each of those N blocks.
- Then, put together all the 'lines' resulting from all different N estimations.
- Next slide shows the case for 2 subsamples.
- Problem? By dividing sample you lose power. Parameters estimation becomes less precise (as $N \to \infty$).

Laffer Curve again



Semi-parametric estimation

- The previous regression does not look bad.
- Problem: The kink in the neighbourhood of 0.5 may lead to suboptimal policy recommendations around that point
- Increasing the number of subsamples yields a smoother plot.
- But as we see we lose power.

Semi-parametric estimation

- In reality, more complicated than that.
- Instead of splitting sample, one assigns different weights to different values

Semi-parametric estimation

- In reality, more complicated than that.
- Instead of splitting sample, one assigns different weights to different values
- (More on weights later on at the end of this course)
- Let's see what other alternative possibilities we have

Non-linear estimation

- Polynomials
- Logarithmic transformations

Non-linear estimation

- Polynomials
- Logarithmic transformations

Let's take one step back before we get into them

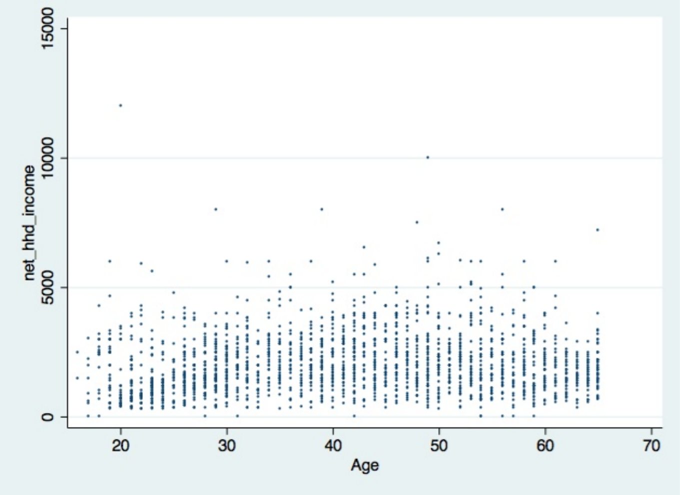
Multiple Regression

- Let's take a look at the effect of Age on Wage
- Data: Germany, 2009, GLES (German Longitudinal Election Study)
- For simplicity, we will use wage as a dependent variable (not log(wage) as we see it is normally done)
- N=1,847

Multiple Regression

$$\omega_i = \beta_0 + \beta_1 \times age_i + u_i$$

the Wage Regression. Germany, 2009. N=1,847



the Wage Regression. Germany, 2009. N=1,847

OLS results

the Wage Regression. Germany, 2009. N=1,847

OLS results

$$\omega_i = \beta_0 + \beta_1 \times age_i + u_i$$
$$\hat{\beta}_0 = 1,778 \quad \hat{\beta}_1 = 6.48$$

Can you interpret these?

the Wage Regression. Germany, 2009. N=1,847

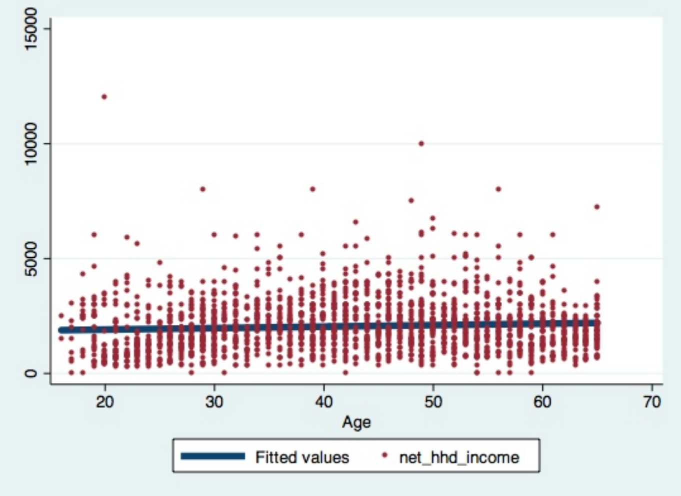
OLS results

$$\omega_i = \beta_0 + \beta_1 \times age_i + u_i$$
$$\hat{\beta}_0 = 1,778 \quad \hat{\beta}_1 = 6.48$$

Can you interpret these?
95% Confidence intervals:

$$\hat{\beta}_0 \in [1601, 1955] \quad \hat{\beta}_1 \in [2.49, 10.47]$$

the Wage Regression. Germany, 2009. N=1,847



the Wage Regression. Germany, 2009. N=1,847

Before proceeding, it's always a good idea to check that OLS is roughly a good estimation of our model.
A good visual and quick way is to plot residuals vs. fitted values.

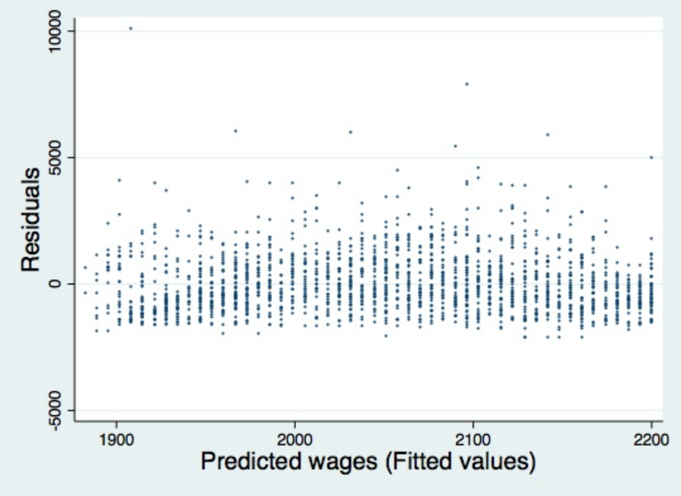
the Wage Regression. Germany, 2009. N=1,847

Before proceeding, it's always a good idea to check that OLS is roughly a good estimation of our model.
A good visual and quick way is to plot residuals vs. fitted values.

Fitted values: $\hat{\omega}_i = \hat{\beta}_0 + \hat{\beta}_1 age_i$
Residuals: $\hat{u}_i = \omega_i - \hat{\omega}_i$

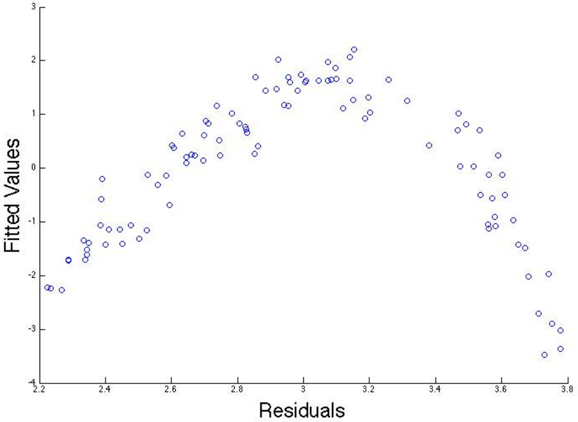
Recall from previous lessons: this must look as the *most boring* scatterplot ever

the Wage Regression. Germany, 2009. N=1,847



the Wage Regression. Germany, 2009. N=1,847

Compare this to the residuals of the Laffer Curve seen above



the Wage Regression. Germany, 2009. N=1,847

- All in all, not too bad.
- However, can we explain wages only based on age?

the Wage Regression. Germany, 2009. N=1,847

```
. reg   net_hhd_income age if age<66

      Source |       SS       df       MS              Number of obs =    1862
-------------+------------------------------           F(  1,  1860) =   10.13
       Model |  14034424.5     1  14034424.5           Prob > F      =  0.0015
    Residual |  2.5761e+09  1860  1384973.12           R-squared     =  0.0054
-------------+------------------------------           Adj R-squared =  0.0049
       Total |  2.5901e+09  1861  1391770.24           Root MSE      =  1176.8

------------------------------------------------------------------------------
net_hhd_in~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   6.482308   2.036353     3.18   0.001     2.488531    10.47609
       _cons |   1778.883   90.25982    19.71   0.000     1601.862    1955.905
------------------------------------------------------------------------------
```

the Wage Regression. Germany, 2009. N=1,847

. reg net_hhd_income age if age<66

Source	SS	df	MS
Model	14034424.5	1	14034424.5
Residual	2.5761e+09	1860	1384973.12
Total	2.5901e+09	1861	1391770.24

Number of obs	=	1862
F(1, 1860)	=	10.13
Prob > F	=	0.0015
R-squared	=	0.0054
Adj R-squared	=	0.0049
Root MSE	=	1176.8

net_hhd_in~e	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	6.482308	2.036353	3.18	0.001	2.488531	10.47609
_cons	1778.883	90.25982	19.71	0.000	1601.862	1955.905

the Wage Regression. Germany, 2009. N=1,847

```
. reg   net_hhd_income age if age<66

      Source |       SS       df       MS              Number of obs =    1862
-------------+------------------------------           F(  1,  1860) =   10.13
       Model |  14034424.5     1  14034424.5           Prob > F      =  0.0015
    Residual |  2.5761e+09  1860  1384973.12           R-squared     =  0.0054
-------------+------------------------------           Adj R-squa[illegible] 0.0049
       Total |  2.5901e+09  1861  1391770.24           Root MSE      =  1176.8

------------------------------------------------------------------------------
net_hhd_in~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   6.482308   2.036353     3.18   [illegible]  2.488531    10.47609
       _cons |   1778.883   90.25982    19.7[illegible] [illegible].000  1601.862    1955.905
------------------------------------------------------------------------------
```

the Wage Regression. Germany, 2009. N=1,847

- All in all, not too bad.
- However, can we explain wages only based on age?

the Wage Regression. Germany, 2009. N=1,847

- All in all, not too bad.
- However, can we explain wages only based on age?
- We cannot even explain 1% of the variability!

the Wage Regression. Germany, 2009. N=1,847

- All in all, not too bad.
- However, can we explain wages only based on age?
- We cannot even explain 1% of the variability!
- What else may be explaining wage?

the Wage Regression. Germany, 2009. N=1,847

- All in all, not too bad.
- However, can we explain wages only based on age?
- We cannot even explain 1% of the variability!
- What else may be explaining wage?

the Wage Regression. Germany, 2009. N=1,847

- All in all, not too bad.
- However, can we explain wages only based on age?
- We cannot even explain 1% of the variability!
- What else may be explaining wage?

the Wage Regression. Germany, 2009. N=1,847

- All in all, not too bad.
- However, can we explain wages only based on age?
- We cannot even explain 1% of the variability!
- What else may be explaining wage?
- Ability, experience, years of studying, connections, parents' education, discount factor, height, physical appearance, school attended,...
- (these have all been proven in one way or another to affect wages. You saw most of the papers in PoE)

the Wage Regression. Germany, 2009. N=1,847

$$\omega_i = \beta_0 + \beta_1 \times age_i + \beta_2 \times educ_i + u_i$$

where $educ_i = \{1, 2, 3, 4, 5\}$, 1=lowest degree and 5=highest degree

For simplicity, you can think of *educ* as number of years in education.

the Wage Regression. Germany, 2009. N=1,847

OLS results

the Wage Regression. Germany, 2009. N=1,847

OLS results

$$\omega_i = \beta_0 + \beta_1 \times age_i + u_i$$

$$\hat{\beta}_0 = 1,778 \quad \hat{\beta}_1 = 6.48$$

95% Confidence intervals: $\hat{\beta}_0 \in [1601, 1955] \quad \hat{\beta}_1 \in [2.49, 10.47]$

$$\omega_i = \beta_0 + \beta_1 \times age_i + \beta_2 \times educ_i + u_i$$

$$\hat{\beta}_0 = 728.03 \quad \hat{\beta}_1 = 11.53 \quad \hat{\beta}_2 = 280.12$$

95% Confidence intervals:

$$\hat{\beta}_0 \in [467, 990] \quad \hat{\beta}_1 \in [7.51, 15.67] \quad \hat{\beta}_2 \in [228.2, 332]$$

$$R^2 = 0.07$$

What's going on?

the Wage Regression. Germany, 2009. N=1,847

```
.
. reg    net_hhd_income age education if age<66

      Source |       SS       df       MS              Number of obs =    1847
-------------+------------------------------           F(  2,  1844) =   61.38
       Model |   160278172     2  80139085.8           Prob > F      =  0.0000
    Residual |  2.4077e+09  1844  1305675.43           R-squared     =  0.0624
-------------+------------------------------           Adj R-squared =  0.0614
       Total |  2.5679e+09  1846  1391085.41           Root MSE      =  1142.7

------------------------------------------------------------------------------
net_hhd_in~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   11.53778   2.055819     5.61   0.000     7.505799    15.56976
   education |   280.1277   26.45242    10.59   0.000     228.2479    332.0076
       _cons |   728.0315   133.2504     5.46   0.000     466.6941     989.369
------------------------------------------------------------------------------
```

the Wage Regression. Germany, 2009. N=1,847

- Education greatly improves the fit of our model
- Education *changes* the value of the coefficient for Age.
- The latter is what we refer to as OVB - Omitted Variable Bias

the Wage Regression. Germany, 2009. N=1,847

- Education greatly improves the fit of our model
- Education *changes* the value of the coefficient for Age.
- The latter is what we refer to as OVB - Omitted Variable Bias
- Once we control for Education, the impact of age on wage doubles.

the Wage Regression. Germany, 2009. N=1,847

- corr(educ,age) = -0.19
- Education $\Rightarrow$ positive effect on wage

the Wage Regression. Germany, 2009. N=1,847

- corr(educ,age) = -0.19
- Education ⇒ positive effect on wage
- Young people *on average* more educated
- Age captures that: if we forget education level, then the difference of average wage between old and young people is 'corrected' / 'smoothed' / 'softened' by the fact that young people are more educated

the Wage Regression. Germany, 2009. N=1,847

- corr(educ,age) = -0.19
- Education ⇒ positive effect on wage
- Young people *on average* more educated
- Age captures that: if we forget education level, then the difference of average wage between old and young people is ‘corrected’ / ‘smoothed’ / ‘softened’ by the fact that young people are more educated
- Wage difference *on average* between old and young people is small, not because age does not matter, but because younger have more education
- Need to compare equally educated young and old people ⇒ that’s what controlling for other variables

Table of average monthly wages in euros per household. Sample: 18 - 60 years old.
"Young" $=$ 18-39 years old
"Old" $=$ 40-59 years old

Table of average monthly wages in euros per household. Sample: 18 - 60 years old.
"Young" = 18-39 years old
"Old" = 40-59 years old

	YOUNG	OLD
Whole sample	1,900 $N = 794$	2,250 $N = 844$

Table of average monthly wages in euros per household. Sample: 18 - 60 years old.
"Young" = 18-39 years old
"Old" = 40-59 years old

Table of average monthly wages in euros per household. Sample: 18 - 60 years old.
"Young" = 18-39 years old
"Old" = 40-59 years old

	YOUNG	OLD
Whole sample	1,900 $N = 794$	2,250 $N = 844$
High Educ.	1,996 $N = 199$	3,470 $N = 88$
Mid. Educ.	2,087 $N = 31$	2,928 $N = 33$
Low Educ.	1,861 $N = 564$	2,073 $N = 723$

Omitted Variable Bias

- Once we control for education, we see that age really matters
- It was significant before, but with a downward bias.

Omitted Variable Bias

- If we forget to include a variable that is correlated to a variable we are including, we will get bias
- See algebraic proof (below)
- See geometric interpretation - alt. set of slides

Omitted Variable Bias. Need to Show

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2}$$

and then use this to show that

$$\hat{\beta}_1 \longrightarrow_p \beta_1 + \rho_{xu} \frac{\sigma_u}{\sigma_x}$$

R^2

- It gives a measure of how well we can explain the data using our model
- $\frac{\text{`what we can explain'}}{\text{`what we can explain'} + \text{`what we cannot'}}$

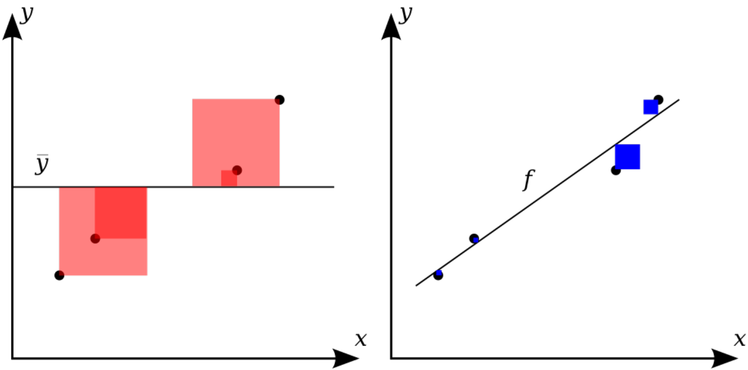
R^2

- It gives a measure of how well we can explain the data using our model
- $\frac{\text{'what we can explain'}}{\text{'what we can explain'} + \text{'what we cannot'}}$
- $R^2 = 1 - \frac{RSS}{TSS}$

R^2

- It gives a measure of how well we can explain the data using our model
- $\frac{\text{'what we can explain'}}{\text{'what we can explain'} + \text{'what we cannot'}}$
- $R^2 = 1 - \frac{RSS}{TSS}$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} = 1 - \frac{RSS}{TSS}$$



"Coefficient of Determination" by Orzetto - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons, wiki page R^2

R^2

- Why can it only increase when we include regressors?
- Algebraic and Geometric Interpretation
- Adjusted-R^2
- To think about:
  - Is a high R^2 necessarily good or imperative?
  - How is it possible that we *may* get a negative value?
  - How can you guarantee that R^2 is equal to 1?

R^2

- Why can it only increase when we include regressors?

R^2

- Why can it only increase when we include regressors?
- $(i)\ Min \sum_i (y_i - \alpha - \beta_1 x_{1i})^2$
- $(ii)\ Min \sum_i (y_i - \alpha - \beta_1 x_{1i} - \beta_2 x_{2i})^2$

R^2

- Why can it only increase when we include regressors?
- (i) $Min \sum_i (y_i - \alpha - \beta_1 x_{1i})^2$
- (ii) $Min \sum_i (y_i - \alpha - \beta_1 x_{1i} - \beta_2 x_{2i})^2$
- By construction, $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$ are such that (ii) is minimized

R^2

- Why can it only increase when we include regressors?
- (i) $Min \sum_i (y_i - \alpha - \beta_1 x_{1i})^2$
- (ii) $Min \sum_i (y_i - \alpha - \beta_1 x_{1i} - \beta_2 x_{2i})^2$
- By construction, $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$ are such that (ii) is minimized
- If x_2 cannot help reduce the Sum of Squared Errors, then $\beta_2 \to 0$

R^2

- Why can it only increase when we include regressors?
- $(i)\ Min \sum_i (y_i - \alpha - \beta_1 x_{1i})^2$
- $(ii)\ Min \sum_i (y_i - \alpha - \beta_1 x_{1i} - \beta_2 x_{2i})^2$
- By construction, $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$ are such that (ii) is minimized
- If x_2 cannot help reduce the Sum of Squared Errors, then $\beta_2 \rightarrow 0$
- Geometric interpretation: you are increasing the projection space by one dimension. At worst, you can stay exactly where you are. And in most cases, you get actually closer to the N-dimensional vector y

Adjusted R^2

- Imposes a ‘penalty’ each time we include a new regressor
- In other words, prevents R^2 from increasing mechanically by including *any* variable.

Adjusted R^2

- Imposes a 'penalty' each time we include a new regressor
- In other words, prevents R^2 from increasing mechanically by including *any* variable.

$$R^2 = 1 - \frac{RSS}{TSS} \qquad R^2_{adj} = 1 - \frac{n-1}{n-1-k}\frac{RSS}{TSS}$$

k is number of regressors other than the constant.

Adjusted R^2

- Imposes a ‘penalty’ each time we include a new regressor
- In other words, prevents R^2 from increasing mechanically by including *any* variable.

$$R^2 = 1 - \frac{RSS}{TSS} \qquad R^2_{adj} = 1 - \frac{n-1}{n-1-k}\frac{RSS}{TSS}$$

k is number of regressors other than the constant. Note: *when everything else is constant*

Adjusted R^2

- Imposes a 'penalty' each time we include a new regressor
- In other words, prevents R^2 from increasing mechanically by including *any* variable.

$$R^2 = 1 - \frac{RSS}{TSS} \qquad R^2_{adj} = 1 - \frac{n-1}{n-1-k}\frac{RSS}{TSS}$$

k is number of regressors other than the constant. Note: *when everything else is constant*

$$\frac{\partial R^2_{adj}}{\partial k} = -\frac{RSS}{TSS}(n-1)(-1)(n-1-k)^{-2}(-1) = -\frac{RSS}{TSS}(n-1)\frac{1}{(n-1-k)^2} < 0$$

R^2

- Is a high R^2 necessarily good or imperative?
- How is it possible that we *may* get a negative value?
- How can you guarantee that R^2 is equal to 1?

R^2

- Is a high R^2 necessarily good or imperative?
- How is it possible that we *may* get a negative value?
- How can you guarantee that R^2 is equal to 1?

In cases where negative values arise, the mean of the data provides a better fit to the outcomes than do the fitted function values.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} = 1 - \frac{RSS}{TSS}$$

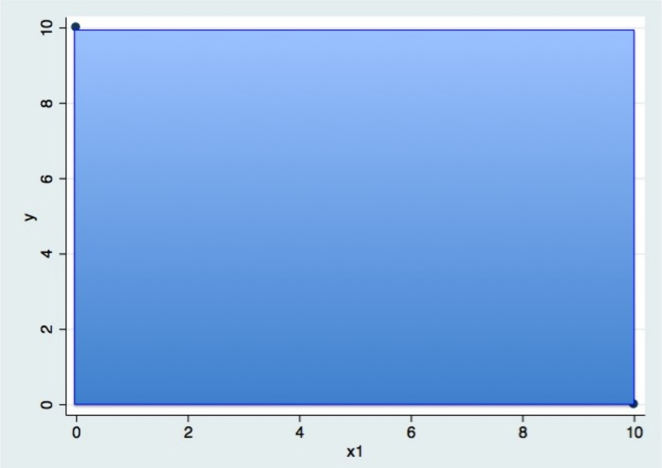
$$y = \begin{pmatrix} 10 \\ 0 \end{pmatrix} x_0 = \begin{pmatrix} 0 \\ 10 \end{pmatrix}$$

```
reg y x1, noc

      Source |       SS       df       MS              Number of obs =       2
-------------+------------------------------           F(  1,      1) =    0.00
       Model |          0     1           0           Prob > F      =  1.0000
    Residual |        100     1         100           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -1.0000
       Total |        100     2          50           Root MSE      =      10

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |          0          1     0.00   1.000    -12.7062     12.7062
------------------------------------------------------------------------------
```

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} = 1 - \frac{RSS}{TSS}$$



$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} = 1 - \frac{RSS}{TSS}$$