Workshop2. Functional Form

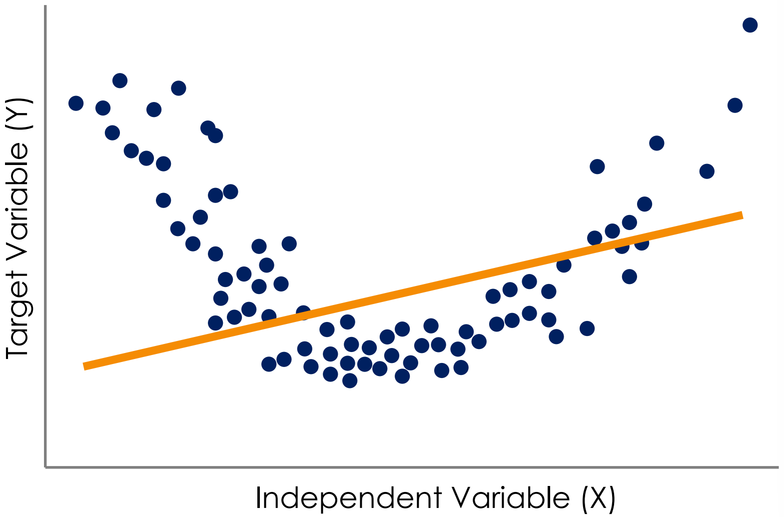
UNAM – FE   Econometrics I

Esp. Humberto Acevedo Assistant. Emilio Sandoval

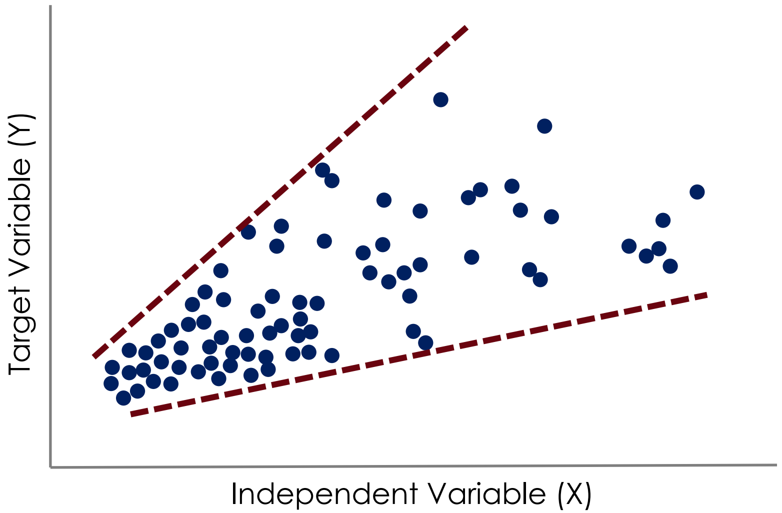
Aug 22, 2022

OLS Assumptions

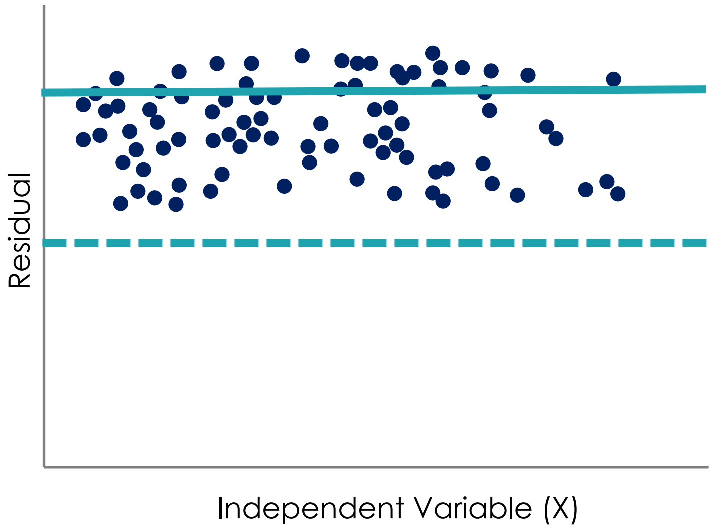
Linearity



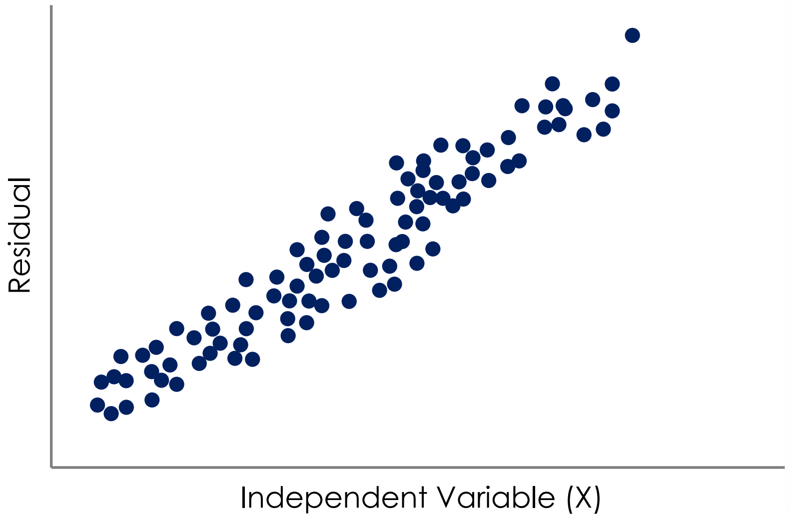
Homoscedasticity



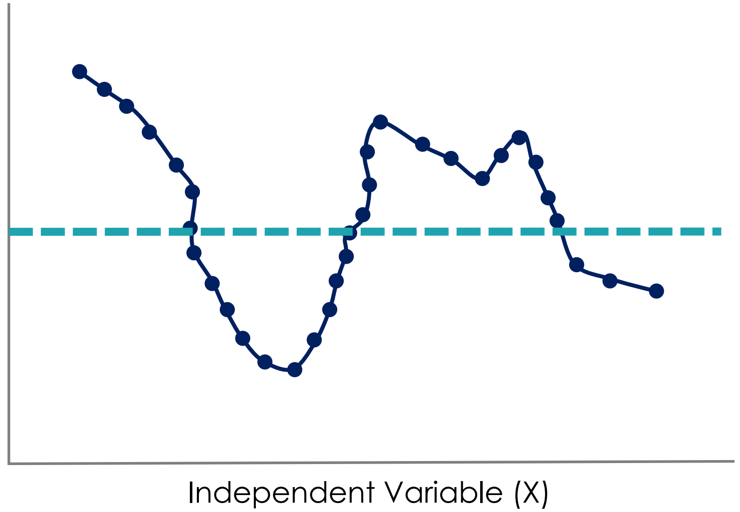
Zero-Mean Errors



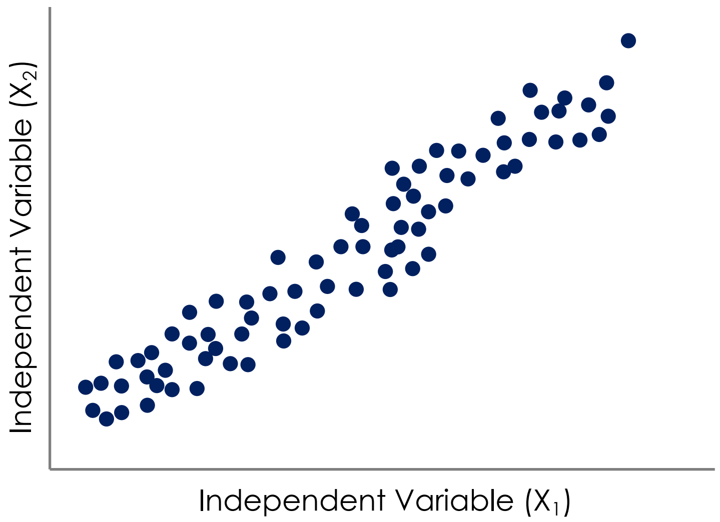
Endogeneity



Autocorrelation of Errors



Multicollinearity

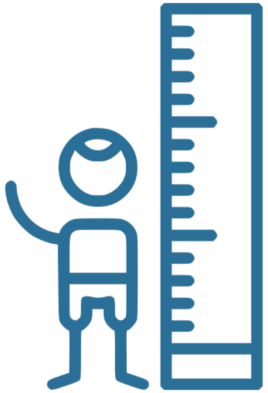


Properties of **estimators**

Unbiasedness

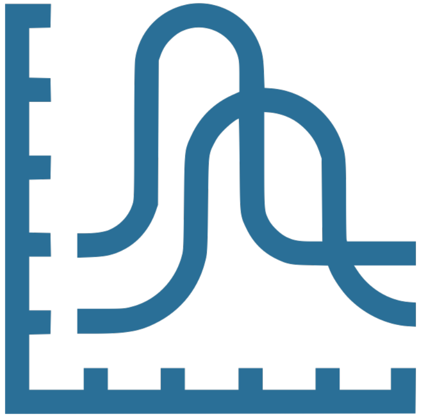
If expected vale is identical to population parameter

Consistency



Estimator approaches parameter as sample size increases

Efficiency



If two estimator are unbiased, we choose the one with smaller variance (it is efficient)

Sufficiency

It is sufficient if information in it about the parameter is enough

Okay but…what is the importance of functional form?

Think about the complex world which is not always linear…how can we show a relationship that is not linear in a linear regression?

Importance of **functional form**

"linearity in parameters but not necessarily in variables"

$$Y_1 = \ b_0 + b_0 X_{1i} + b_2 X_{2i}^2 + u_i$$

For example:

$$Wage_1 = \ \underbrace{b_0}_{\text{Linear}} + \underbrace{b_1}_{\text{Linear}} exp_i + \underbrace{b_2}_{\text{Linear}} \overbrace{exp_i}^{\text{Not Linear}} + \underbrace{b_3}_{\text{Linear}} \overbrace{exp_i^2}^{\text{Not Linear}} + u_i$$

Let's check several functional forms used in econometrics and other related fields

Linear

$$Y = \beta_1 + \beta_2 X$$

Log-Linear (log-log)

$$\ln Y = \beta_1 + \beta_2 \ln X$$

Log-lin

$$\ln Y = \beta_1 + \beta_2 X$$

Lin-log

$$Y = \beta_1 + \beta_2 \ln X$$

Reciprocal

$$Y = \beta_1 + \beta_2 \left(\frac{1}{x}\right)$$

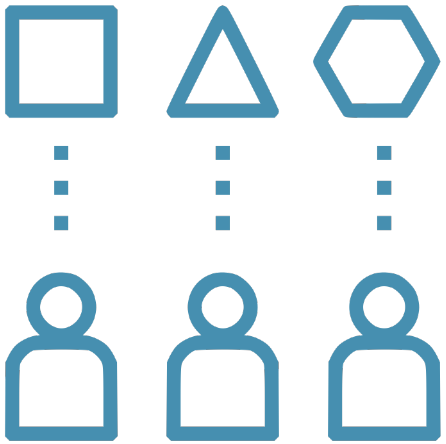
Log-Reciprocal

$$\ln Y = \beta_1 - \beta_2 \left(\frac{1}{x}\right)$$

Models

You have to ask yourself about

Why this model?



Characteristics

Appropiate cases

Log-Linear **model**

Economists look for the rate of growth in some economic variables such as population, GDP, monetary supply, employment…

$$\overbrace{\ln Y_1}^{\text{Log side}} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^2 + u_i$$

or

$$Y_1 = \underbrace{\beta_0 + \beta_1\, lnX_{1i} + \beta_2 \ln X}_{\text{Log side}} + u_i$$

Notice that just one side of the equation is logarithmic

The difference is on the assigned variable

$$\overbrace{\ln Y_1}^{\text{Log side}} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^2 + u_i$$

log - lin

or

$$Y_1 = \underbrace{\beta_0 + \beta_1 lnX_{1i} + \beta_2 \ln X}_{\text{Log side}} + u_i$$

lin - log

Log-Linear **model**

$$\overbrace{\ln Y_1}^{\text{Log side}} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^2 + u_i$$

Estimated percentage change in your dependent variable for a unit change in your independent variable

$$Y_1 = \underbrace{\beta_0 + \beta_1 \, lnX_{1i} + \beta_2 \ln X}_{\text{Log side}} + u_i$$

Estimated unit change in your dependent variable for a percentage change in your independent variable

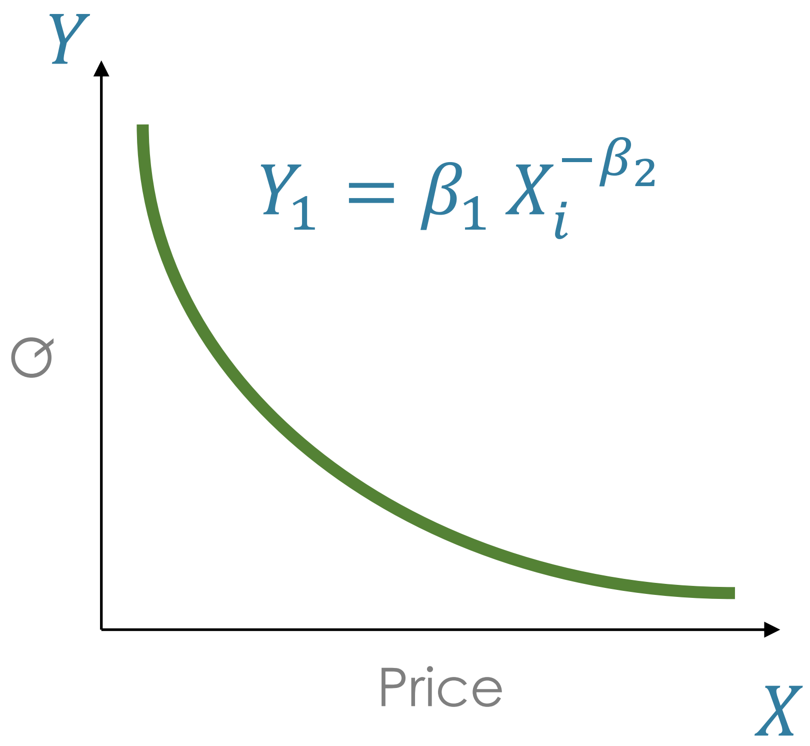
On the other hand, if we want to use elasticities we adopt a log-log model

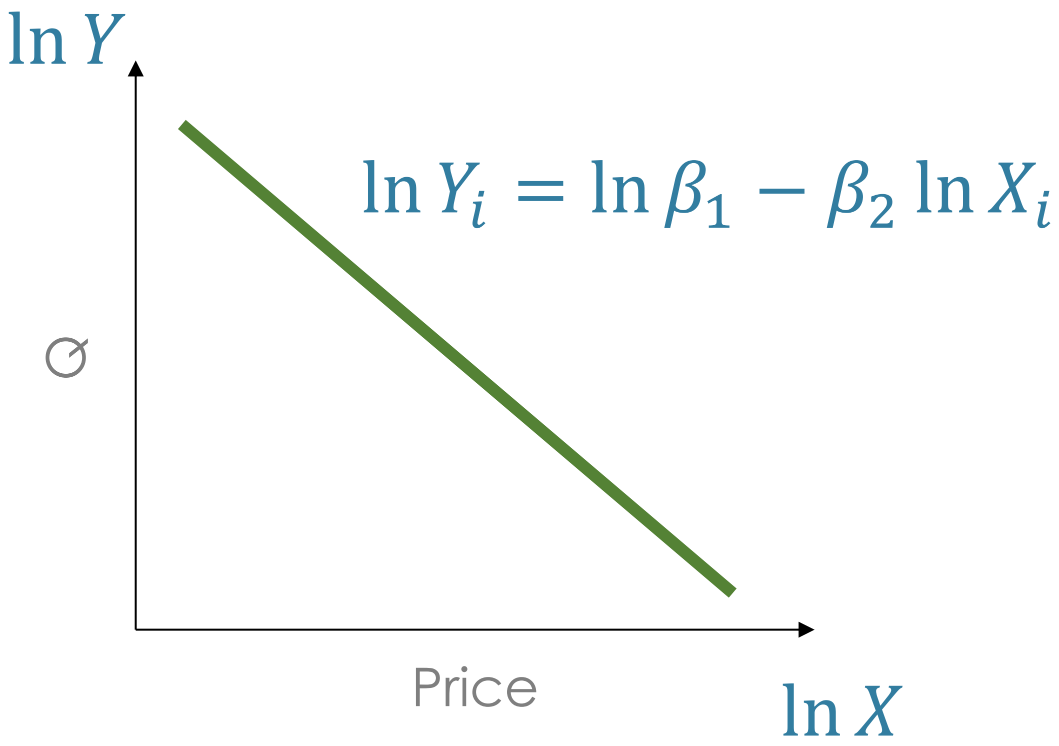
$$Y_1 = \beta_1 X_i^{\beta_2} + e^{ui}$$

or

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i$$

Log-Log **model**





What we may focus on is the slope β_2 which measures the elasticity of Y respect to X

or

the percentage change in Y for a small percentage change in X

NOTE: there is a difference between percentage change and percentage points

For example, the rate of unemployment is stated as percentage

Imagine that Mexico unemployment rate for 2019 is 4%, but in 2020 it grows to 8%

It is said that the percentage point change in unemployment rate is 2

but

The percentage change is $\frac{8-6}{6}$ or $\approx 33\%$

Log-Log **model**

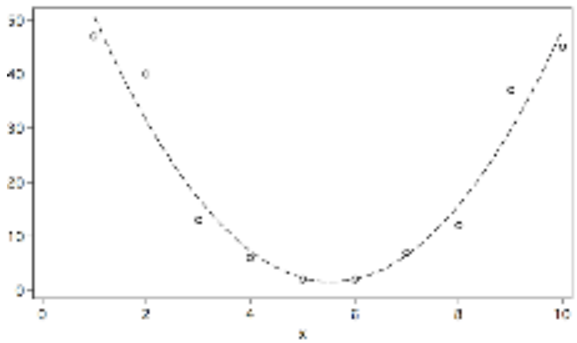
Suppose we have the following equation

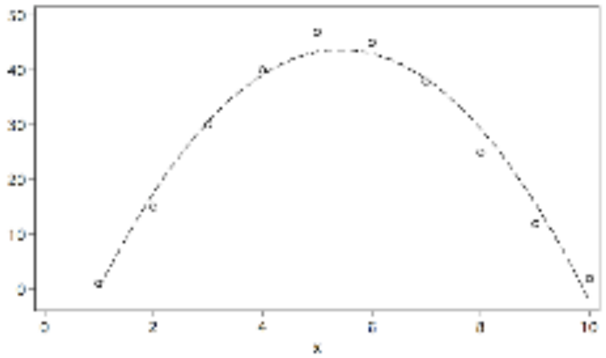
$$y = \alpha + \beta_1 X_1 + \beta_2 X_2^2$$

Stata will output these coefficients

Model A	Coef.
x	-.1839751
x2	1.016747
constant	.2076584

Model D	Coef.
x	.1839751
x2	-1.016747
constant	99.79234





Practice (**Quadratic**)

The data set in MUS08 contains data panel information about wages and salaries in the USA from 1976-1982.

Then we plot the relationship between salary and experience.

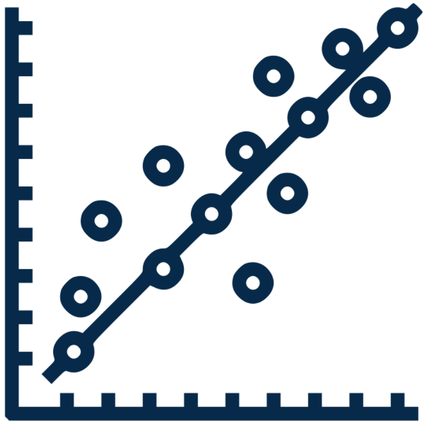
Finally, generate linear regression with `lwage` as dependent and `exp`, `exp2`, `wks` and `ed` as independent.

STATA COMMANDS

1. `graph twoway (scatter lwage exp) (qfit lwage exp) (lowess lwage exp)`
2. `regress lwage exp exp2 wks ed`

Practice (**Quadratic**)

We can read the output as:



$R^2 = 0.28$

Salaries increase in 0.6% for each additional work week

Salaries increase in 0.6% for each additional year of education

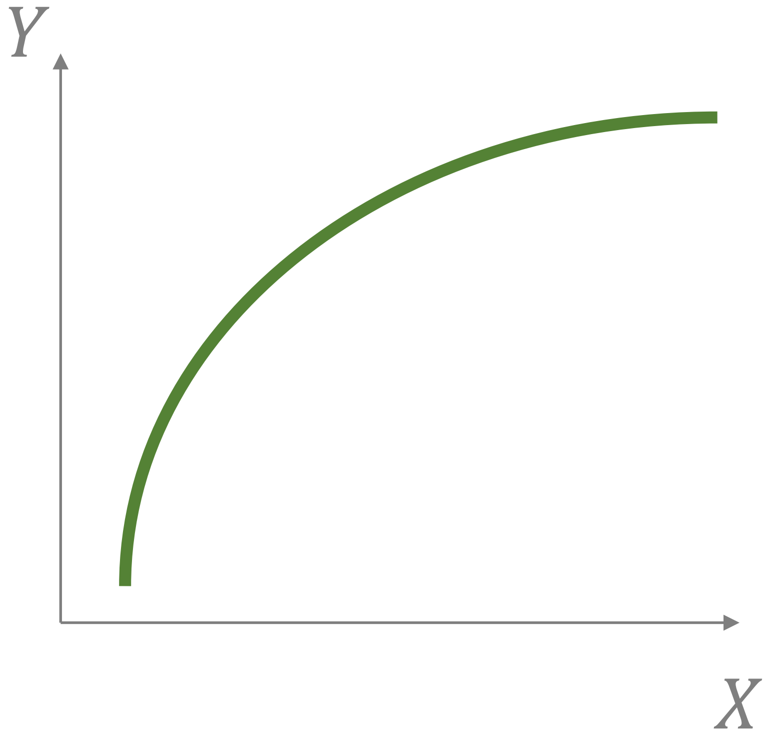
How to find the slope of the curve?

Hint: look into x_2^2

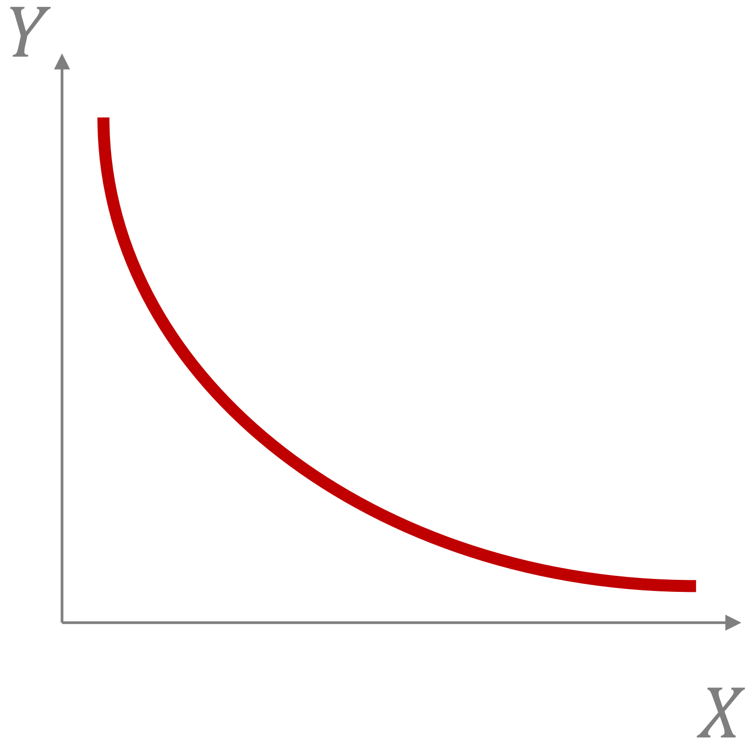
$$y = \beta_0 + \beta_1 \boxed{X_1} + \beta_2 \boxed{X_2^2}$$

Practice (**Tipping point**)

$$y = \beta_0 + \beta_1 X_1 \boxed{+} \beta_2 \boxed{X_2^2}$$



$$y = \beta_0 + \beta_1 X_1 \boxed{-} \beta_2 \boxed{X_2^2}$$



How to find the tipping point?

By solving for it

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2$$

Then calculate the derivative with respect to X

$$y' = \beta_1 + 2\beta_2 X_2^2$$

Finally, equal to zero and solve for finding tipping points

Practice (**Tipping point**)

1. State the output in econometrics terms

$$\hat{y} = 4.097 + 0.44675\,\text{exp}\;0.00072\,exp^2 + 0.0058wks + 0.7604ed + \varepsilon$$

2. If we take first derivative respect to exp we have

$$y' = 0.4467 - (2\times(0.00072)exp)$$

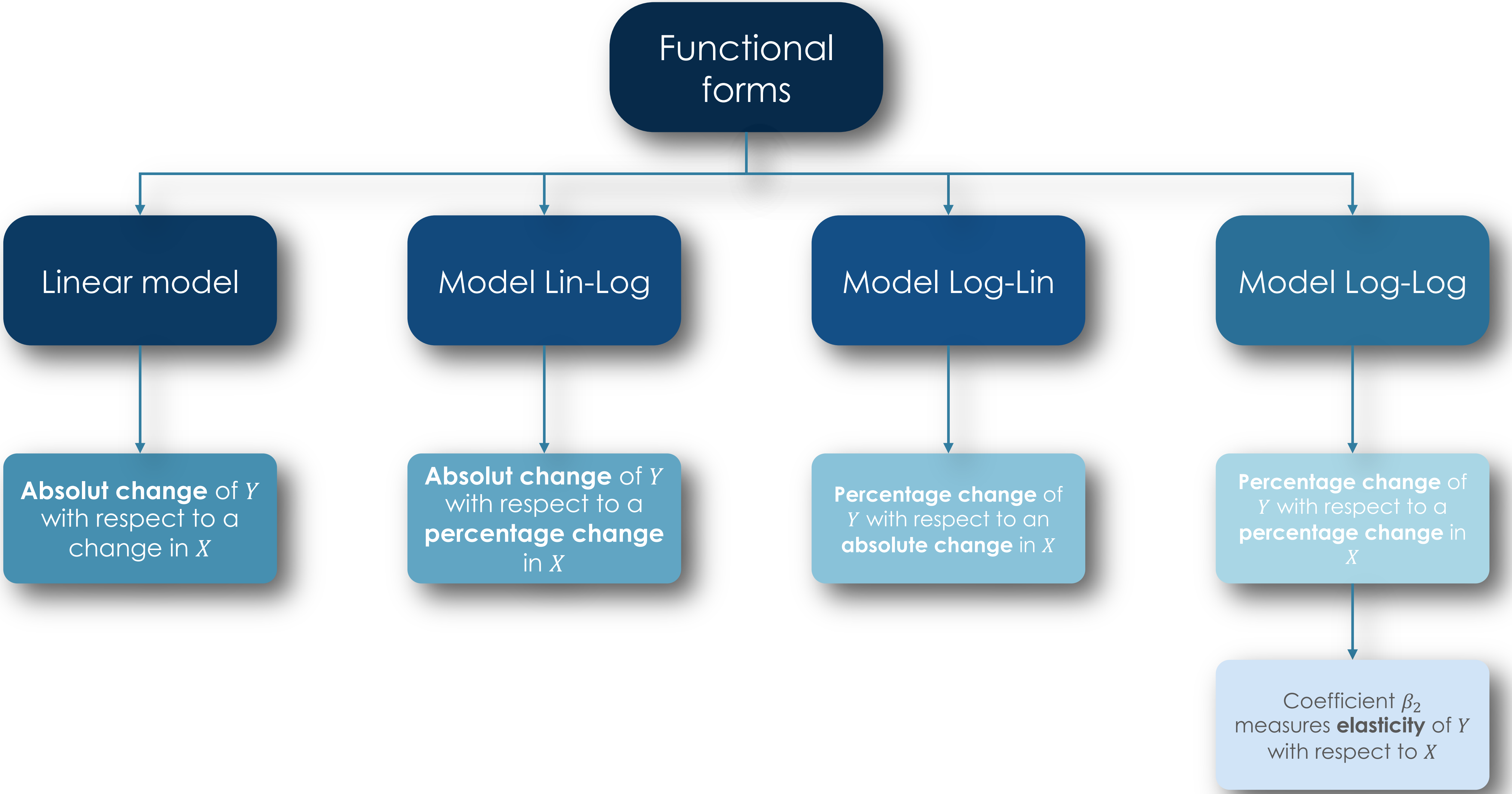
3. If we solve for exp we have:

$$exp = \frac{0.4467}{2\times 0.00072} = 31$$

Practice (**Tipping point**)

We can read the output as:

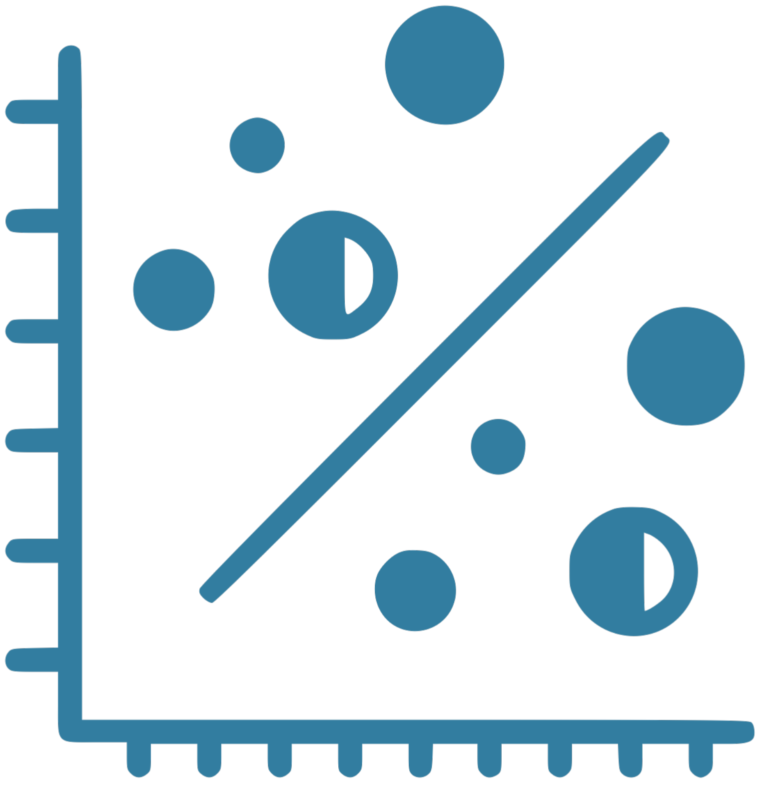
Salaries increase as experience grows up to 31 years $\frac{0.0447}{(2\times 0.00072)}$ and from there it decreases


Functional forms
Linear model
Model Lin-Log
Model Log-Lin
Model Log-Log
Absolut change of Y with respect to a change in X
Absolut change of Y with respect to a **percentage change** in X
Percentage change of Y with respect to an **absolute change** in X
Percentage change of Y with respect to a **percentage change** in X
Coefficient β_2 measures **elasticity** of Y with respect to X

Ramsey Regression Equation Specification Error Test (RESET) is a general specification test for linear regression models

It tries to prove if non-linear combinations of adjusted values have any explanatory force over dependent variable

If any of this non-linear combinations have explanatory influence over Y, then model is specified incorrectly



We use auto.dat dataset. We will try to explain car prices depending on mileage, weight, engine capacity, and whether car was produced in USA or not.

Run a linear regression.

Standardize data (mean 0 and variance 1).

Then we generate powers over adjusted values (y^2, y^3, y^4).

Finally we run the regression using powers.

STATA COMMANDS

1. `reg price mpg weight foreign`
2. `predict y`
3. `sum y`
4. `replace y = (y-r(mean))/r(sd))`
5. `gen y2 = y^2`
6. `gen y3 = y^3`
7. `gen y4 = y^4`
8. `reg price mpg weight foreign y2 y3 y4`

We will prove coefficient significance in order to inspect powers using F test (apply test over linear constraint)

Due to output, we cannot reject H_0, our model is well specified.

Then, apply Ramsey test over well specified form.

What is the conclusion?

STATA COMMANDS

```
9.  test y2 y3 y4
10. reg price mpg weight foreign
11. ovtest
```

References

- **Salvatore, D., & Sarmiento, J. C.** (1983). *Econometría* (No. HB141 S39). McGraw-Hill.
- **Kumari K., J. Pract Cardiovasc, Wooldridge, J.** (2020). *Introductory econometrics : a modern approach*. Boston, MA: Cengage. Gujarati, D. & Porter, D. (2009). *Basic econometrics*. Boston: McGraw-Hill Irwin.
- **Gujarati, D. N.** (2009). Basic econometrics. Tata McGraw-Hill Education.