Notes

Friday, August 19, 2022 9:55 AM

Myroslava Sánchez Andrade A01730712 | 19/08/2022

The simple Linear Regression

The simple regression model measures the linear relationship between 2 random variables (X and Y), where X explains the movements of Y, so Y depends on the movement of Y. This model estimates a linear equation (regression line) to represent how much Y (on average) moves with movements of X, and what is the expected value of Y when X = 0.

- Introduction

Used to understand the linear relationship between two variables (assuming that one variable can be used as a predictor of the other variable). The simple regression considers only one independent variable (the multiple regression can include more) and **one dependent variable**.

Use:

- Understanding the relationship between a dependent variables and one or more independent variables.
- Predicting or estimating the expected value of the dependent variable.

A linear regression model does not capture non-linear relationships (unless with specific mathematical transformations of the variables).

Regression model with the Market Regression Model:

The Market Model states that the expected return of a stock is given by its alpha coefficient plus its market beta coefficient times the market return.

$$E[R_i] = \alpha + \beta(R_M)$$

We can express the same equation using β_0 as alpha, and β_1 as market beta:

$$E[R_i] = \beta_0 + \beta_1(R_M)$$

We can estimate beta 0 and beta 1 coefficients by running a simple linear regression model, specifying that the market return is the independent variable and the stock return is the dependent variable. The market regression model can be expressed as:

$r_{(i,t)} = b_0 + b_1 * r_{(M,t)} + \varepsilon_t$

Where:

ε_t is the error at time t.

(Which behaves as a normal distributed random variable)

$r_{(i,t)}$ is the return of the stock i at time t.

$r_{(M,t)}$ is the market return at time t

b_0 and b_1 are called regression coefficients

- Types of data structures

The market model is a time-series regression model, we looked at the relationship between 2 variables representing one feature or attribute (returns) of two "subjects" over time: a stock and a market index. We try to understand how changes in the return over time is related to changes in the stock return (independent variable => "time-series").

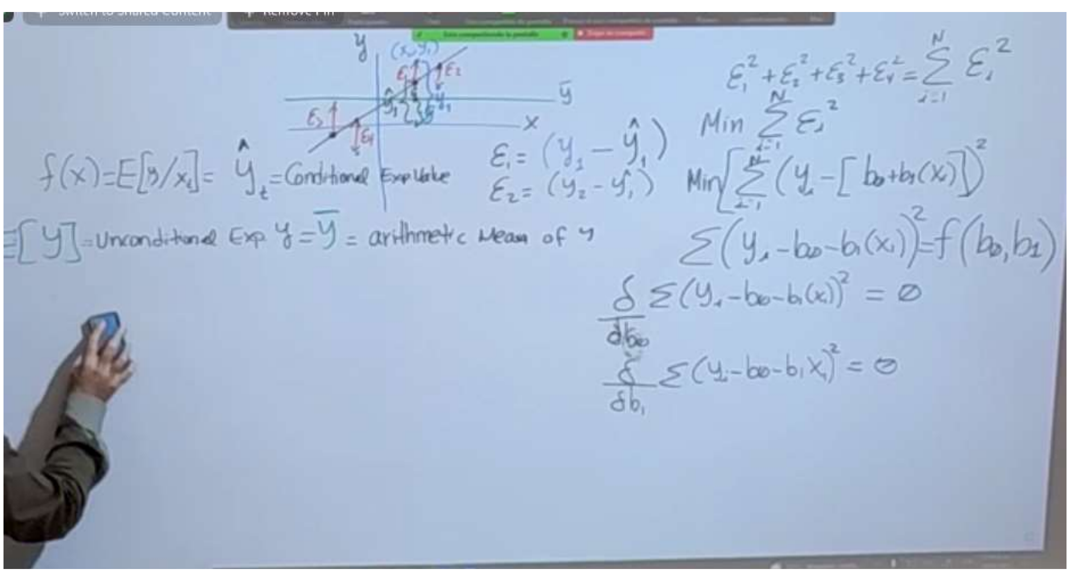
Data structures used in regression models:

- Times-series

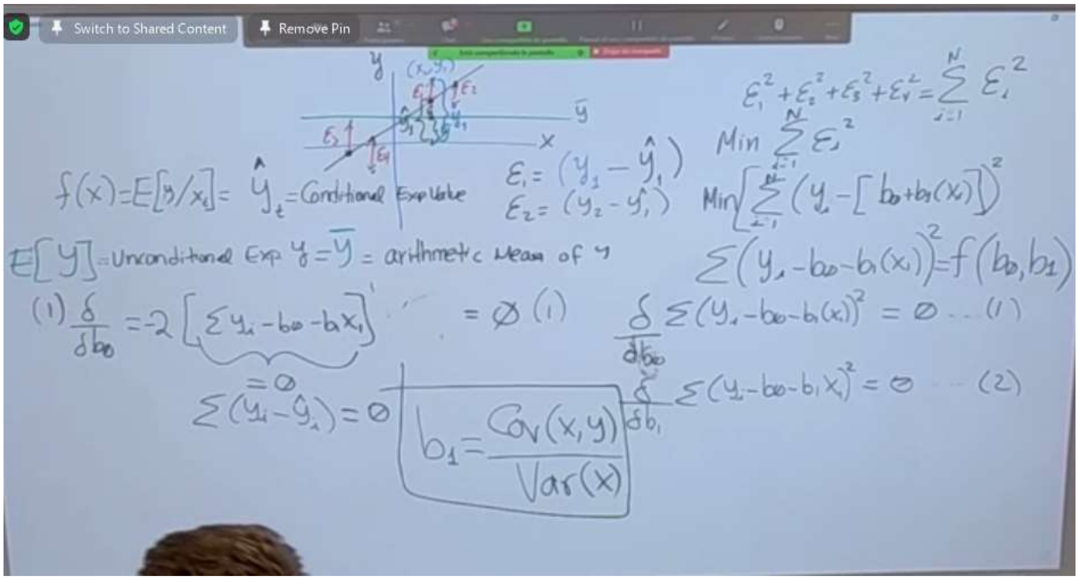
- Cross-sectional
  Model where the independent variables(s) represent(s) a specific characteristic of many subjects in only one period of time. The dependent variable would be a feature or characteristic of the subjects in one point in time.
  Example: analyze how annual earnings per share (when announced at the end of the year) of a set of 100 firms is related to the stock price change (return) of these firms right at the end of the year; the subjects are the 100 public firms; the dependent variable is the stock return of each of the 100 firms; and the independent variable would be the earnings per share of these 100 firms disclosed to investors at the end of the year.

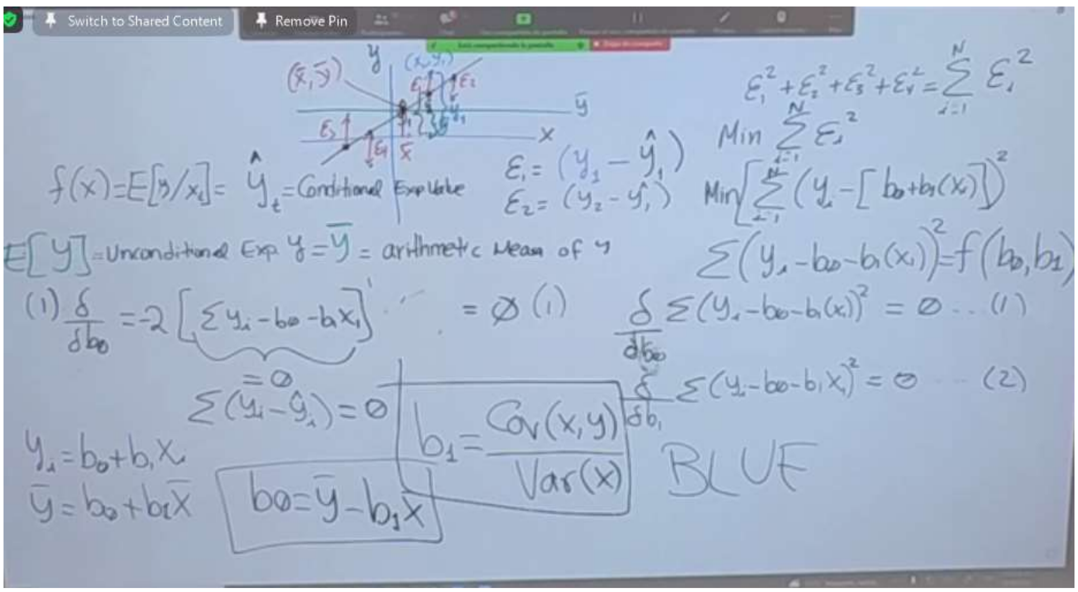
- Panel data
  Combines time-series with cross-sectional structures. It has two different regression models: **pulled time series regression** => time series regression of more than one subject; **panel data regression** => this model accounts for changes in both subjects and periods.
  Example: taking the above example, if we add 10 years of data for each of the 100 firms we can design either *a) a pulled time-series regression*, or *b) a panel-data regression.* For the **b)**, we would need to estimate a more sophisticated model since there are basically two sources or types of relations: a) how changes in the dependent variable of subjects are related to changes in the independent variables of the same subjects, and b) how average (taking the average over time) changes in the dependent variable among different subjects are related to average changes in the independent variable. The first type of relation is called within-subjects differences, while the second one is called between-subjects differences.

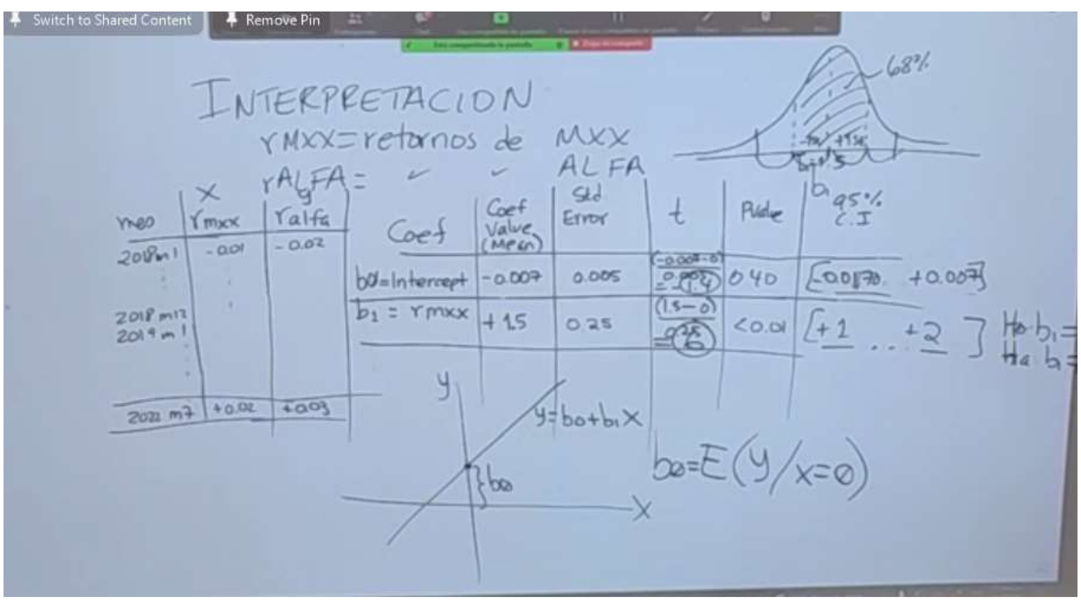
The regression model can also be classified based on the number of independent variables:
1 independent variable => simple regression model
2+ independent variables => multiple regression model



Unconditional expected value = arithmetic mean
Conditional expected value = conditional mean = regression function


Switch to Shared Content
Remove Pin
f(x)=E[y/x]= ŷ = Conditional Exp Value
E[Y] = Unconditional Exp y = ȳ = arithmetic Mean of y
Min Σ ε²
b₁ = Cov(x,y) / Var(x)


Switch to Shared Content
Remove Pin
f(x)=E[y/x]= ŷ = Conditional Exp Value
E[Y] = Unconditional Exp y = ȳ = arithmetic Mean of y
b₁ = Cov(x,y) / Var(x)
BLUE
b0 = ȳ - b₁x̄


Switch to Shared Content
Remove Pin
INTERPRETACION
rMXX = retornos de MXX
rALFA = ALFA
68%
95% C.I
Coef
Coef Value (Mean)
Std Error
t
P-value
b0 = Intercept
-0.007
0.005
0.40
b1 = rmxx
+1.5
0.25
<0.01
y = b0 + b1X
b0 = E(y/x=0)