



UNIVERSITÉ DU LUXEMBOURG
Mathematics Research Unit (RMATH)

Bachelor en Mathématiques

Experimental Mathematics

Summer semester 2019-2020

The Linear Regression Model Applied to Finance

Authors:

Anthony PANTELIS

Julien PEZZI

Supervisor:

Dr. George KERCHEV

Contents

1	Linear Regression Model	3
1.1	What Is Linear Regression ?	3
1.2	Relationship and Slope	3
1.3	Example	3
2	How to trace a regression line ?	4
2.1	Axes and Means	4
2.2	Slope and y -intercept	4
2.3	The Regression Line and its Representation	5
3	Correlation and Standard Error of the Estimate	5
3.1	The Coefficient of Correlation - R	5
3.2	The Coefficient of Determination - R^2	6
3.3	The Standard Error of the Estimate	7
4	An Exponential Model	8
4.1	How to Find the Exponential Model	8
4.2	Regression vs Exponential Model	9
5	Representing Stock Data	10
5.1	Some Beginner's Notions...	10
5.2	Choosing a Stock, Representing Its Data	10
6	The Algorithm	13
6.1	The code	13
6.2	Its Functioning	16
7	Conclusion	17

1 Linear Regression Model

1.1 What Is Linear Regression ?

Linear regression is a statistic model used to find a linear relationship between a dependent and an independent variable. The independent variable is on the x axis and the dependent variable is on the y axis. What we try to understand is what happens to the dependent variable, as the independent variable changes.

1.2 Relationship and Slope

1. If both the independent and the dependent variables increase, we say that there is a **positive relationship**.
2. If on the other hand, as the independent variable increases, the dependent variable decreases, we say there is a **negative relationship**.

We take several observations and try to find a line that will “fit” through all these different points. The line we obtain is called a **regression line**. In reality, “fit” means that we have minimized the difference between the estimated value (the point in the regression line) and the actual value (the point representing the original observation.) In order to find the regression line, we use a method called the "least squares method." The regression line is of the form $y = b_0 \pm b_1x$, where b_0 represents the y -intercept and b_1 is its slope. The + or – that precedes b_1 depends upon the kind of relationship there is between the two variables.

1. For a **positive relationship** between the two variables, we obtain a **positive slope**.
2. For a **negative relationship**, we obtain a **negative slope**.

1.3 Example

Before we begin, we need to choose of an example, which we will try to model with a regression line.

The following table shows the evolution of the number of members of a rugby club from 2001 to 2006 :

Year	2001	2002	2003	2004	2005	2006
Rank x_i	1	2	3	4	5	6
Number of members y_i	70	90	115	140	170	220

2 How to trace a regression line ?

2.1 Axes and Means

It has been proved that the regression line has to pass by intersection of the axes of the means of its two variables. In order to find out where we need to trace the two axes, we calculate the mean for each column x and y , where $(x; y)$ represent some given observations. For the example that precedes, we have :

$$\text{mean}(x) = \bar{x} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5$$

and

$$\text{mean}(y) = \bar{y} = \frac{70 + 90 + 115 + 140 + 170 + 220}{6} = \frac{805}{6} = 134.2$$

The first thing we need to do is trace the vertical axis, $\bar{x} = 3.5$ and the horizontal axis, $\bar{y} = 134.2$. We have to keep in mind that the intersection will be one of the two points through which our regression line will pass. After that, we also have to calculate the values, shown in the following table :

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
1	70	-2.5	-64.2	6.3	160.5
2	90	-1.5	-44.2	2.3	66.3
3	115	-0.5	-19.2	0.3	9.6
4	140	0.5	5.8	0.3	2.9
5	170	1.5	35.8	2.3	53.7
6	220	2.5	85.8	6.3	214.5

2.2 Slope and y -intercept

To determine the slope b_1 of our regression line we add up the sixth and fifth column separately and then divide the former by the latter. Basically,

$$b_1 = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}.$$

In our example,

$$b_1 = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} = \frac{160.5 + 66.3 + 9.6 + 2.9 + 53.7 + 214.5}{6.3 + 2.3 + 0.3 + 0.3 + 2.3 + 6.3} = \frac{507.5}{17.8} = 28.5$$

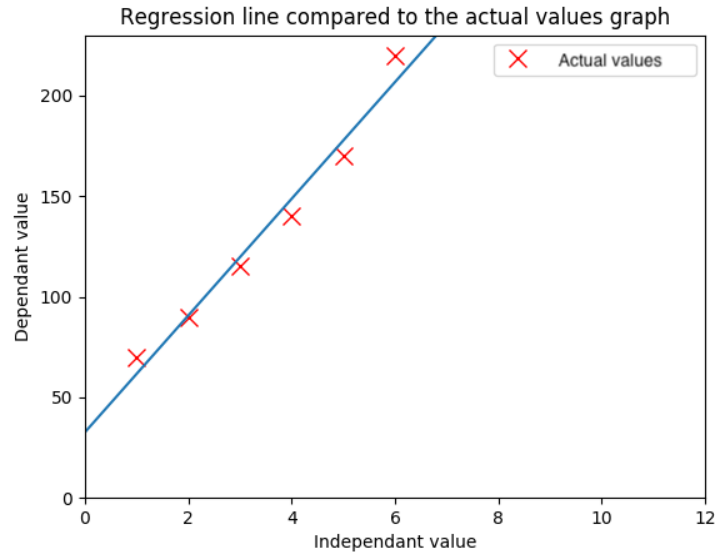
For now, we have that : $\hat{y} = b_0 + 28.5x$.

Finally, to determine the y -intercept b_0 we need to remember of something we have already mentioned. The regression line has to pass from the intersection of the axes \bar{x} and \bar{y} . Hence, we have :

$$134.2 = b_0 + 3.5b_1 \iff b_0 = 134.2 - 3.5b_1 \iff b_0 = 134.2 - 3.5 \cdot 28.5 = 34.5$$

2.3 The Regression Line and its Representation

So, our regression line is : $\hat{y} = 34.5 + 28.5x$.



3 Correlation and Standard Error of the Estimate

3.1 The Coefficient of Correlation - R

In statistics, we have defined a coefficient R (denoted sometimes as r) as the correlation coefficient. From the correlation coefficient, we can derive two basic pieces of information concerning our two values x, y : the strength and the direction of their linear relationship. The value of the correlation coefficient R has one fundamental property to its interpretation : It's always between -1 and 1. Its interpretation is fairly easy. When the value of R approaches either -1 or 1 then we say that we have a strong linear relationship. Otherwise, if R 's value is closer to 0 then the relationship is weak. In other words, the more $|R|$ gets closer to 1, the more the relationship between the two variables becomes stronger. For a more detailed explanation look at the following examples :

If r is :

- **Exactly -1.** A perfect downhill (negative) linear relationship;
- **-0.7.** A strong downhill (negative) linear relationship;
- **-0.5.** A moderate downhill (negative) linear relationship;

- **-0.3** A weak downhill (negative) linear relationship;
- **0** No linear relationship
- **0.3** A weak uphill (positive) linear relationship;
- **0.5** A moderate uphill (positive) linear relationship;
- **0.7** A strong uphill (positive) linear relationship;
- **Exactly +1.** A perfect uphill (positive) linear relationship

Depending on the sign of R , one can derive, as mentioned before, the direction of the linear relationship of the two variables.

In this paper, we will not be particularly interested in calculating the correlation coefficient R , as much as the determination coefficient R^2 . There are several reasons why, but they all break down to the fact that R^2 is easier to interpret.

3.2 The Coefficient of Determination - R^2

The coefficient of determination, denoted R^2 (or r^2), is a measure of a quality of a prediction of a regression line. In simple terms, R^2 (pronounced "R-squared") tells us how accurately a regression line estimates actual values.

In reference to the previous example, in order to calculate R^2 , we need to, first of all, calculate the following table. As you can see we have already calculated the third column (the distance between the actual values and the mean \bar{y} .) The fifth column represents the estimated values. We will denote the estimated values by \hat{y} . To calculate the estimated values we take the expression of our regression line $\hat{y} = 34.5 + 28.5x$ and calculate the different \hat{y} values, $\forall x$. Finally, the last column is the distance between these estimated values and the mean \bar{y} .

x	y	$y - \bar{y}$	$(y - \bar{y})^2$	\hat{y}	$\hat{y} - \bar{y}$	$(\hat{y} - \bar{y})^2$
1	70	-64.2	4121.6	63	-71.2	5069.4
2	90	-44.2	1953.6	91.5	-42.7	1823.3
3	115	-19.2	368.6	120	-14.2	201.6
4	140	5.8	33.6	148.5	14.3	204.5
5	170	35.8	1281.6	177	42.8	1831.8
6	220	85.8	7361.6	205.5	71.3	5083.7

The coefficient of determination R^2 is calculated with the help of the following formula :

$$R^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2}$$

In our case we have :

$$R^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2} = \frac{5069.4 + 1823.3 + 201.6 + 204.5 + 1831.8 + 5083.7}{4121.6 + 1953.6 + 368.6 + 33.6 + 1281.6 + 7361.6} = \frac{14214.3}{15120.6} = 0.94$$

$R^2 = 0.94$ is a very good fit. If the actual values and estimated values were not as close, then we would probably get something like $R^2 = 0.6$ instead. In fact, if we obtain $R^2 = 1$, then we say we have a perfect fit between actual and estimated values. In other words, as the value of R^2 approaches 1, we say there's a better fit and we mean that we have a better estimation. Another way to measure goodness of fit is the "Standard Error of the Estimate". In that case we are interested in the distance between the estimated and the actual values.

3.3 The Standard Error of the Estimate

The standard error of the estimate is a measure of the accuracy of predictions. You can interpret the Standard Error of the Estimate (denoted S_e) as a standard deviation in the sense that, if you have a normal distribution for the prediction errors, then you will expect about two-thirds of the data points to fall within a distance S_e either above or below the regression line. Also, about 95% of the data values should fall within $2S_e$, and so forth.

To calculate the standard error of the estimate, we have to compare estimated values, calculated before, and the actual values. The distance between the estimated values and their respective actual values is called the error. In the following table, we've placed the estimated values denoted \hat{y} , the difference between \hat{y} and y , and finally the square of this difference.

x	y	\hat{y}	$\hat{y} - y$	$(\hat{y} - y)^2$
1	70	63	-7	49
2	90	91.5	1.5	2.3
3	115	120	5	25
4	140	148.5	8.5	72.3
5	170	177	7	49
6	220	205.5	-14.5	210.3

Now, we do the sum of the terms of the last column which is :

$$49 + 2.3 + 25 + 72.3 + 49 + 210.3 = 407.9.$$

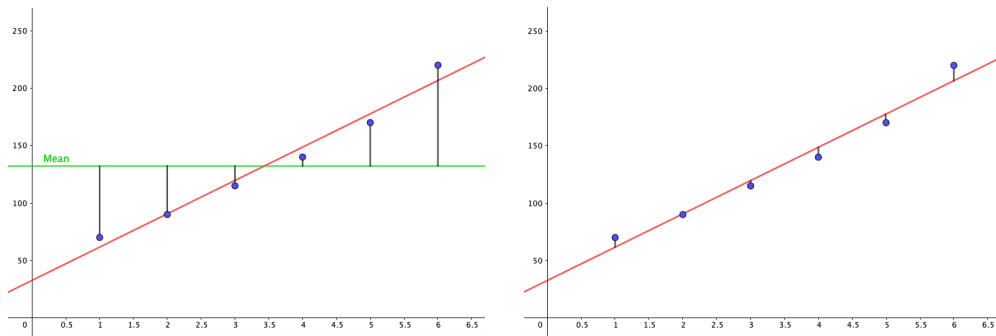
Finally, to find the standard error we have to apply the following formula :

$$S_e = \sqrt{\frac{\sum(\hat{y} - y)^2}{n - 2}}$$

with n , the number of values. In our case, we obtain :

$S_e = \sqrt{\frac{407.9}{4}} = \sqrt{102} = 10.1$. Therefore, the standard error of the estimate is 10.1.

The coefficient of determination R^2 compares the distance between actual values to the mean, whereas the standard error of the estimate compares the actual values with the estimate values.



Difference between the coefficient of determination and the standard error of the estimate

4 An Exponential Model

4.1 How to Find the Exponential Model

We will use the same example as before. For our exponential model we will have to set $z_i = \ln(y_i)$. The process to determine an exponential equation that well interprets our observations is going to be very similar to the one we already completed to determine our regression line. As we did before we have to make a table for our observations :

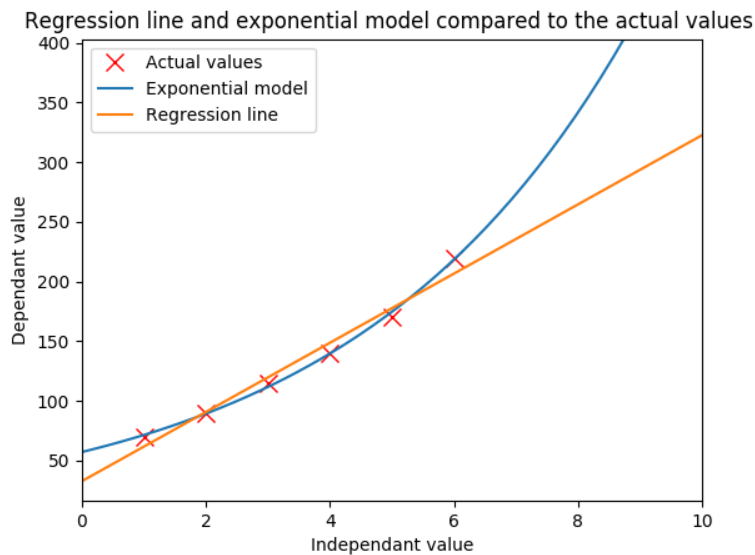
Year	2001	2002	2003	2004	2005	2006
Rank x_i	1	2	3	4	5	6
Number of members y_i	70	90	115	140	170	220
z_i	$\ln(70) = 4.2$	$\ln(90) = 4.5$	4.7	4.9	5.1	5.4

We will then quickly calculate, a regression line formula for x and z , following the aforementioned process. The regression line we obtain, is of the form $z = 4.1 + 0.2x$. After having determined the regression line we only

have to solve $z = \ln(y)$.

$$\begin{aligned} z &= \ln(y) \\ \iff \ln(y) &= 4.1 + 0.2x \\ \iff e^{\ln(y)} &= e^{4.1+0.2x} \\ \iff y &= e^{4.1+0.2x} \\ \iff y &= e^{4.1} \cdot e^{0.2x} \\ \iff y &= 60.3e^{0.2x}. \end{aligned}$$

The exponential model's equation is $y_e = 60.3 \cdot e^{0.2x}$.



4.2 Regression vs Exponential Model

Since we don't really have a quality indicator for the exponential model, as we do for the linear regression one (R , R^2 , Standard error of estimate, etc.), we have to find another way to decide which of the two models is best suitable for our example. A solid way to do that is by observing the two graphs we were able to obtain by our python project. Even by just watching the scatter plot, we can make out that the points follow a more curve-like trail (and not a linear one). As a result, we can understand with absolute certainty that the exponential model is more adequate for this example than the linear one.

5 Representing Stock Data

5.1 Some Beginner's Notions...

A (capital) stock of a corporation, is the total of the shares into which the ownership of the corporation is divided. Yahoo Finance is an easy tool, with which one can obtain historical data for a grand variety of stocks of different corporations.

In Yahoo Finance we can also find a chart where the historical data are graphically represented. As we will do later on, while the x -axis represents the days of the year, the y -axis represents the adjusted close of the stock. The close of the stock is the cash value of the last transacted price before the market closes for the day. The adjusted close is nothing else than the closing price, but having factored in anything that might affect the stock price after the market closes. The adjusted closing price might be a little different from the closing price, but most of the times the two prices are equal.

5.2 Choosing a Stock, Representing Its Data

In this part, we will choose a stock of a certain corporation and use the two models (a regression line and the exponential model) to represent its data. We have chosen to work with the Tesla, Inc. stock (TSLA).

The idea is that we are going to choose a certain period of time, let's say January 14th to February 26th 2020. We will make a table containing all the data we will have gathered, as we did before. As we mentioned before, the x -axis will represent the days of the year between January 14th and February 26th, while the y -axis will represent the adjusted close of the stock.

Finally, when we will have determined a regression line formula and an exponential expression, we will calculate -using both formulas- two predictions for a future date. We will choose for example April 14th. That day the adjusted close of the stock was equal to USD 709.89, according to the historical data of Yahoo Finance. Whichever of the two predictions is closer to the real adjusted closing price will determine, which of the aforementioned models is the most precise and reliable.

For the chosen time period, we have the following observations ¹ :

¹Every observation is rounded to the second decimal place.

Date	x : Rank of The Day	y : Adjusted Close
2020-01-14	1	537.92
2020-01-15	2	518.50
2020-01-16	3	513.49
2020-01-17	4	510.50
2020-01-21	8	547.20
2020-01-22	9	569.56
2020-01-23	10	572.20
2020-01-24	11	564.82
2020-01-27	14	558.02
2020-01-28	15	566.90
2020-01-29	16	580.99
2020-01-30	17	640.81
2020-01-31	18	650.57
2020-02-03	21	780.00
2020-02-04	22	887.06
2020-02-05	25	734.70
2020-02-06	26	748.96
2020-02-07	27	748.07
2020-02-10	30	771.28
2020-02-11	31	774.38
2020-02-12	32	767.29
2020-02-13	33	804
2020-02-14	34	800.03
2020-02-18	38	858.4
2020-02-19	39	917.42
2020-02-20	40	899.41
2020-02-21	41	901
2020-02-24	44	833.79
2020-02-25	45	799.91
2020-02-26	46	778.78

In order to avoid having to enter one value at a time in the Python program we have created the following lists. Tesla's stock from 14/01 to 26/02

Days :1 2 3 4 8 9 10 11 14 15 16 17 18 21 22 25 26 27 30 31 32 33 34 38 39 40 41 44 45 46

Corresponding values : 537.92 518.5 513.49 510.5 547.2 569.56 572.2 564.82 558.02 566.9 580.99 640.81 650.57 780 887.06 734.7 748.96 748.07 771.28 774.38 767.29 804 800.03 858.4 917.42 899.41 901 833.79 799.81 778.8

All we have to do is to copy-paste the lists in the program.

We will now use our python program ² to rapidly calculate the formulas for the two models.

²You can find its algorithm in the next section of this paper.

```

Python 3.7.4 (v3.7.4:e09359112e, Jul 8 2019, 14:54:52)
[Clang 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license()" for more information.
>>>
RESTART: /Users/julien/Desktop/Maths expérimentales/Regression line and exponen
tial model.py
Please enter your list of independent values separated by spaces.1 2 3 4 8 9 10
11 14 15 16 17 18 21 22 25 26 27 30 31 32 33 34 38 39 40 41 44 45 46
Please enter your list of dependent values separated by spaces.537.92 518.5 513.
49 510.5 547.2 569.56 572.2 564.82 558.02 566.9 580.99 640.81 650.57 780 887.06
734.7 748.96 748.07 771.28 774.38 767.29 804 800.03 858.4 917.42 899.41 901 833.
79 799.91 778.8
Please enter the greatest value you want to have on the X axis.80
The coefficient of determination R^2 is 0.8998844866556768 .
The regression line formula is 8.71989836250706 x+ 500.48704498400133 .
The exponential model's equation is exp( 0.01282372248954822 x)*exp( 6.238959285
254934 .

```

So our regression line is :

$$y_l = 8.71989836250706x + 500.48704498400133$$

and our exponential model is

$$\begin{aligned}
y_e &= e^{6.238959285254934} \cdot e^{0.01282372248954822x} \\
&= 512.3250492 \cdot e^{0.01282372248954822x} .
\end{aligned}$$

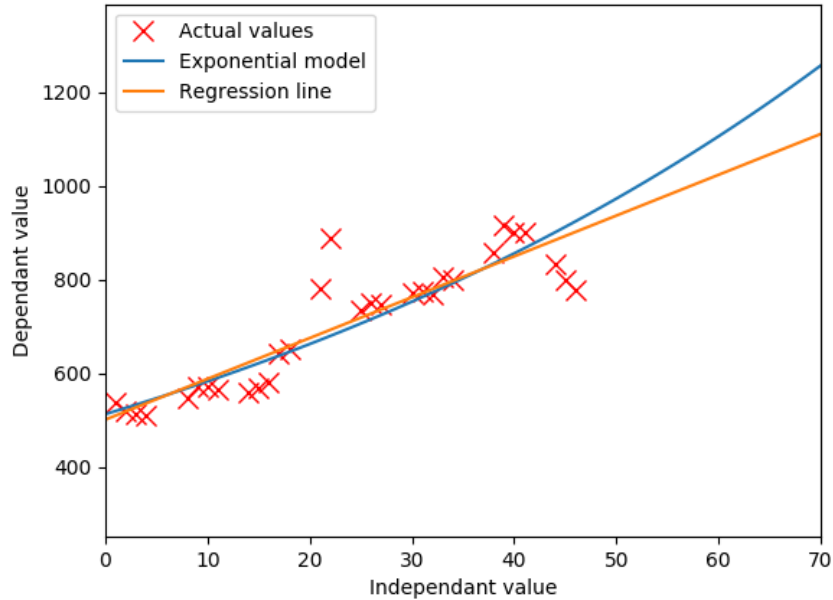
As we said before, we will use each model to make a prediction for April 14th. The date corresponds to an x equal to 61. And so we have :

$$\begin{aligned}
512.3250492 \cdot e^{0.01282372248954822 \cdot 61} &= y_e(61) \\
8.71989836250706 \cdot 61 + 500.48704498400133 &= y_l(61) \\
512.3250492 \cdot e^{0.7822470719} = y_e(61) & \quad 531.9138001 + 500.48704498400133 = y_l(61) \\
512.3250492 \cdot 2.186379702 = y_e(61) & \quad 531.9138001 + 500.48704498400133 = y_l(61) \\
1120.14 = y_e(61) & \quad 1032.40 = y_l(61)
\end{aligned}$$

As we can see the values that we have obtained are not really close to the actual value of the stock. Nevertheless, this is quite normal. Finding a model that accurately predicts a stock's future is really difficult and a much more complicated work than those two models. The reason for that, is that only one parameter (the evolution of the stock's value) is taken under consideration, whereas the stock's value depends on many factors such as risks, customer request, geopolitical situation and so on. Even though the models are not that close to reality, we can see that $y_e(61) > y_l(61) > 709.89$, so the linear regression's prediction is closer to reality than the exponential one. A stock's evolution doesn't follow a strict mathematical model but is highly unpredictable.³

³That does not, of course, mean that this kind of data cannot be modeled

Regression line and exponential model compared to the actual values



6 The Algorithm

6.1 The code

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 import math
4 plt.clf() #to clear the graph of any functions we could have
           #traced before
5
6 I_v= input('Please enter your list of independent values
           #separated by spaces.')
7 Dv= input('Please enter your list of dependent values separated
           #by spaces.')
8 absc= int(input('Please enter the greatest value you want to
           #have on the X axis.)) #
           #horizontal scale
9
10 # vertical scale
11
12 Ivd=[] # It's the same list as below but we need to keep the
           #values we entered.
13 Iv=[] # It's the list corresponding to the independent values.
14 Dvy=[] # It's the list corresponding to the dependent values y.
15 Dvz=[] # It's the list corresponding to the dependent values z
           #(y-values to which we apply the
           #ln-map).
```

```

16
17
18 for e in I_v.split():
19     f = float(e)
20     Iv.append(f)
21     Ivd.append(f)
22
23 for g in Dv.split():
24     h = float(g)
25     Dvy.append(h)
26     Dvz.append(h)
27
28 for i in range(len(Dvz)):
29     Dvz[i]=math.log(Dvz[i])
30
31 def scatterplot(x,y,xtitle,ytitle,title):
32     """
33     Function generating the scatter plot y depending on x
34     Inputs : coordonates of x, coordonnates of y, abscissis
35             title, Y title, graph title
36     Output : scatter plot y depending on x
37     """
38     plt.plot(x, y, linestyle="none" ,marker="x" ,color="red",
39             markersize="10", label="
40             actual values")
41
42     plt.xlabel(xtitle) #legend on the absciss axis
43     plt.ylabel(ytitle) # legend on the ordinate axis
44     plt.title(title) # display a title
45
46 scatterplot(Iv, Dvy, "Independant value", "Dependant value", "
47             Actual values")
48
49 # We calculate x-x_bar :
50 c=sum(Iv)
51 mx=c/len(Iv) # It's the mean of x.
52 for i in range(len(Iv)):
53     Iv[i]=Iv[i]-mx # We subtract the mean to each element of
54                   # the list containing the x-
55                   # values.
56
57 # We calculate y-y_bar :
58 ay=sum(Dvy)
59 my=ay/len(Dvy) # It's the mean of y.
60 for i in range(len(Dvy)):
61     Dvy[i]=Dvy[i]-my # We subtract the mean to each element of
62                     # the list containing the y-
63                     # values.
64
65 # We calculate z-z_bar :
66 az=sum(Dvz)
67 mz=az/len(Dvz) # It's the mean of z.
68 for i in range(len(Dvz)):
69     Dvz[i]=Dvz[i]-mz # We subtract the mean to each element of
70                     # the list containing the z-

```

```

values.
61
62 # We calculate the product between (x-x_bar) and (y-y_bar) :
63 pIv_Dvy= [Iv[i]*Dvy[i] for i in range(min(len(Iv),len(Iv)))]+
max(Iv,Dvy,key=len)[min(len(Iv)
, len(Dvy)):]
64 # We calculate the square of (x-x_bar) :
65 pIv_Iv= [Iv[i]*Iv[i] for i in range(min(len(Iv),len(Iv)))]+max(
Iv,Iv,key=len)[min(len(Iv),len(
Iv)):]
66 # We calculate the product between (x-x_bar) and (z-z_bar) :
67 pIv_Dvz= [Iv[i]*Dvz[i] for i in range(min(len(Iv),len(Iv)))]+
max(Iv,Dvz,key=len)[min(len(Iv)
, len(Dvz)):]
68 # We calculate the square of (y-y_bar) :
69 pDvy_Dvy= [Dvy[i]*Dvy[i] for i in range(min(len(Dvy),len(Dvy))
]+max(Dvy,Dvy,key=len)[min(len(
Dvy),len(Dvy)):]
70
71 s_xx=sum(pIv_Iv) # It's the sum of the elements of the square
of (x-x_bar).
72 s_yy=sum(pDvy_Dvy) # It's the sum of the elements of the square
of (y-y_bar).
73
74 s_xz=sum(pIv_Dvz) # It's the sum of the elements of the product
between (x-x_bar) & (z-z_bar).
75 c_xz= s_xz/s_xx # It corresponds to the coefficient of the
regression line slope for x and
z.
76 y_iz= mz- c_xz*mx # It corresponds to the y-intercept of the
regression line for x and z.
77
78 s_xy=sum(pIv_Dvy) # It's the sum of the elements of the product
between (x-x_bar) & (y-y_bar).
79 c_xy= s_xy/s_xx # It corresponds to the coefficient of the
regression line slope for x and
y.
80 y_iy= my- c_xy*mx # It corresponds to the y-intercept of the
regression line for x and y.
81
82 y_hat=[c_xy*l+y_iy for l in Ivd] # It's the list containing the
y--hat value for all the
independant values.
83 for i in range(len(y_hat)):
84     y_hat[i]=y_hat[i]-my # We subtract the mean to each
element of the list
containing the y_hat-values
.
85
86 # We calculate the square of (y_hat-y_bar) :
87 py_hat= [y_hat[i]*y_hat[i] for i in range(min(len(y_hat),len(
y_hat)))]+max(y_hat,y_hat,key=
len)[min(len(y_hat),len(y_hat))
:]

```

```

88 s_py_hat= sum(py_hat) # It's the sum of the elements of the
                        # square of (y_hat-y_bar).
89 rr= math.sqrt(s_py_hat/((s_yy))) # It corresponds to the
                        # coefficient of determination R^
                        # 2.
90
91 print("The coefficient of determination R^2 is", rr, ".")
92 print("The regression line formula is",c_xy,"x+",y_iy, ".")
93 print("The exponential model's equation is exp(",c_xz,"x)*exp("
                        # ,y_iz, ".")
94
95 plt.title("Regression line and exponential model compared to
            # the actual values ")
96 x= np.linspace(0,absc)
97 y=[c_xy*t+y_iy for t in x] # regression line
98 z=[math.exp((c_xz)*t)*math.exp(y_iz) for t in x] # exponential
                        # model
99
100 plt.xlim(0,absc)
101 plt.ylim(int(min(y)-min(y)/2),(max(y)*1.25))
102 plt.plot(x,z,label="Exponential model")
103 plt.plot(x,y,label="Regression line")
104 plt.legend(["Actual values","Exponential model","Regression
            # line"])
105 plt.show()

```

6.2 Its Functioning

First of all, the program is going to ask you to enter two lists of values :

1. the independent values (X-axis values)
2. the dependent values corresponding to the independent values (Y-axis values).

The only condition of the algorithm is that we must enter two lists of the same length containing only floats. For example, if we enter a first list (independent values) with six values and a second one (dependent values) with five values, it will return a dimension error.

If all conditions are satisfied, the program will return a graph of both the linear regression and the exponential model for the values we asked for. Moreover, we will find the expressions of the two models, as well as the value of R^2 of the linear regression.

7 Conclusion

In this paper, we have worked closely with the linear regression model and tried to see how and if it can be applied to financial matters and, more specifically, to the modeling of stocks. We then introduced some quality indicators, as well as a whole new and more advanced model (the exponential model). After some mathematical "experiments" though, we arrived to the following conclusion.

Even though we can construct complex formulas to show us the quality of our model, the reality of things is that it is always dependent to the kind of data we are trying to represent. In our first example, we saw a quite predictable example where we obtained high values for R^2 for both models. If we were to make a future prediction it would probably be very close to reality, if not equal to it.

Due to the R^2 value being close to 1, in the first example, the regression line was precise and our predictions were accurate. However, in the second one, this was not the case for the predictions, even though once again the regression line was precise. Each problem has its own factors and risks that have to be taken under consideration. While the two models we introduced in our paper seem to be very much trustworthy for modeling simple data, when trying to model stock data or other more complex systems, we need to put the work and create even more complex models. We can also use the multiple linear regression (denoted MLR), which can take several factors. For example, the following picture is a linear regression with two parameters.

