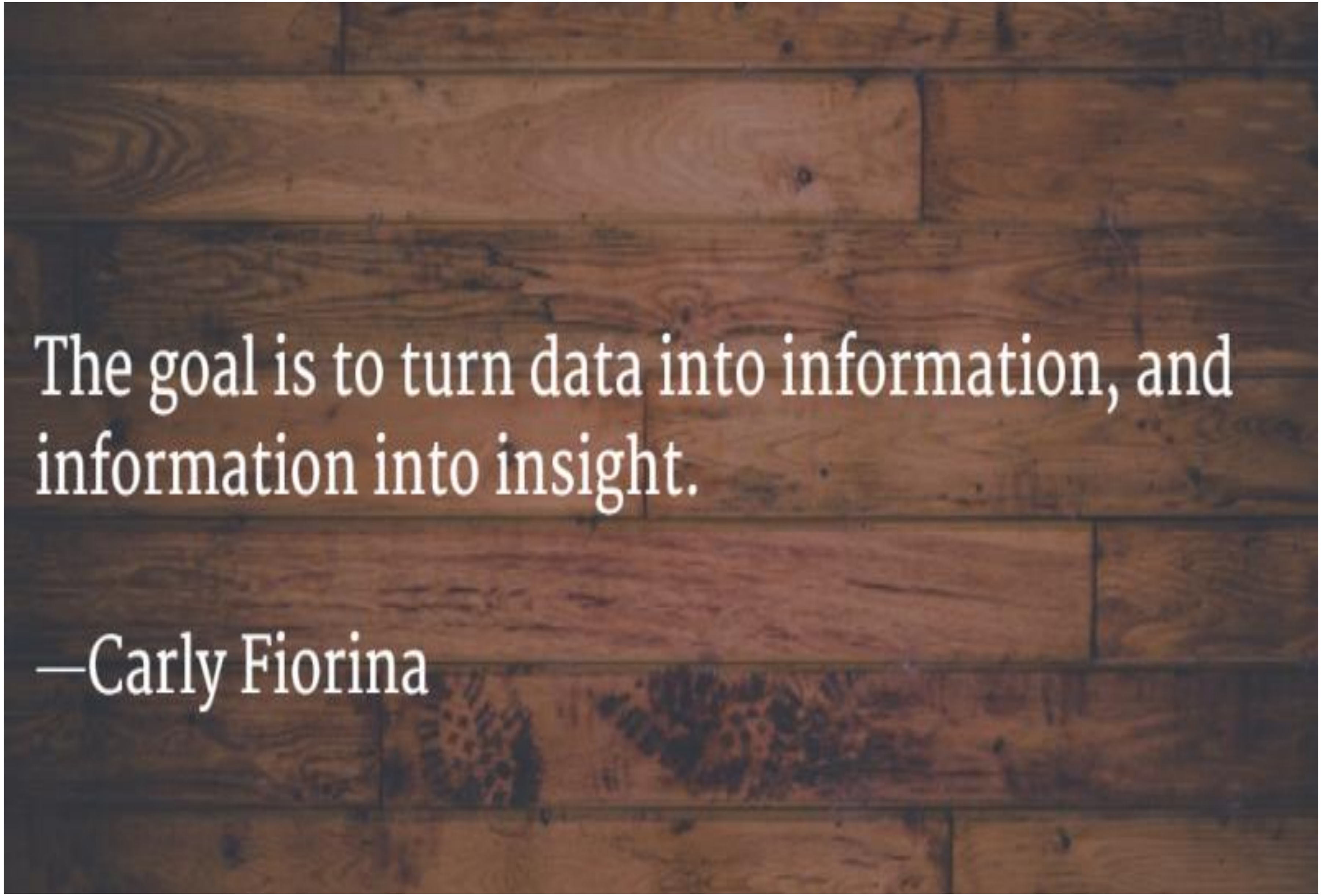
विद्या परं दैवतम्
IIM
VISAKHAPATNAM

Quantitative Analytics tools for financial decisions

Time Series Analytics Part 1

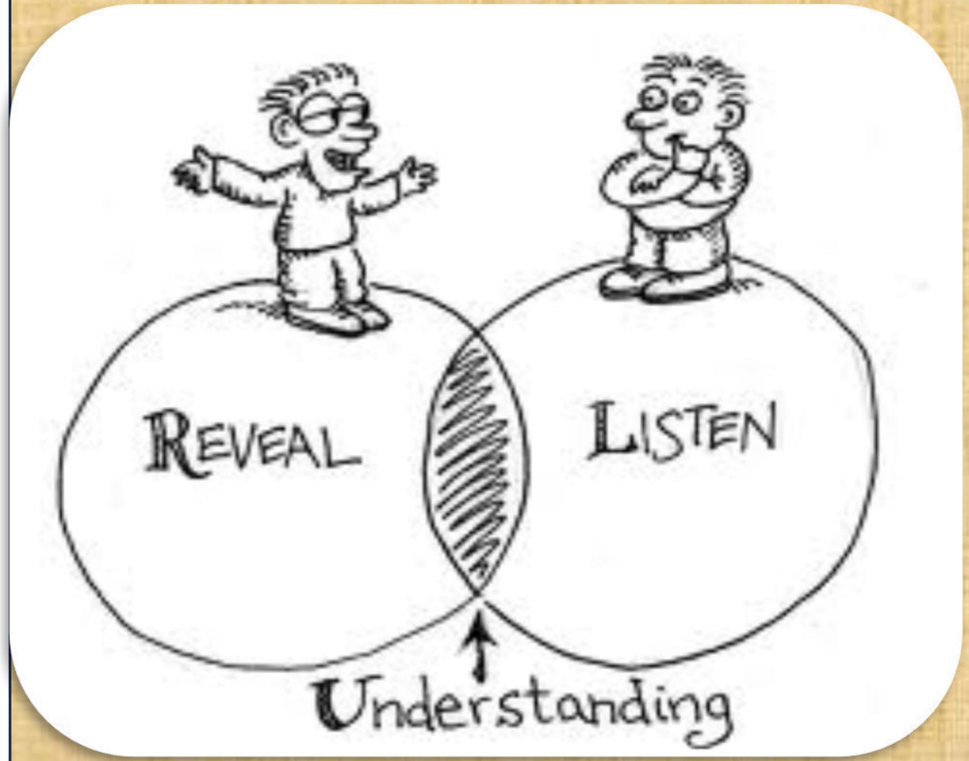
Module 2 Session 7 & 8



The goal is to turn data into information, and information into insight.

—Carly Fiorina

Getting Deeper Understanding…



Deeper understanding of Finance using Time Series Analytics!

Today’s Agenda

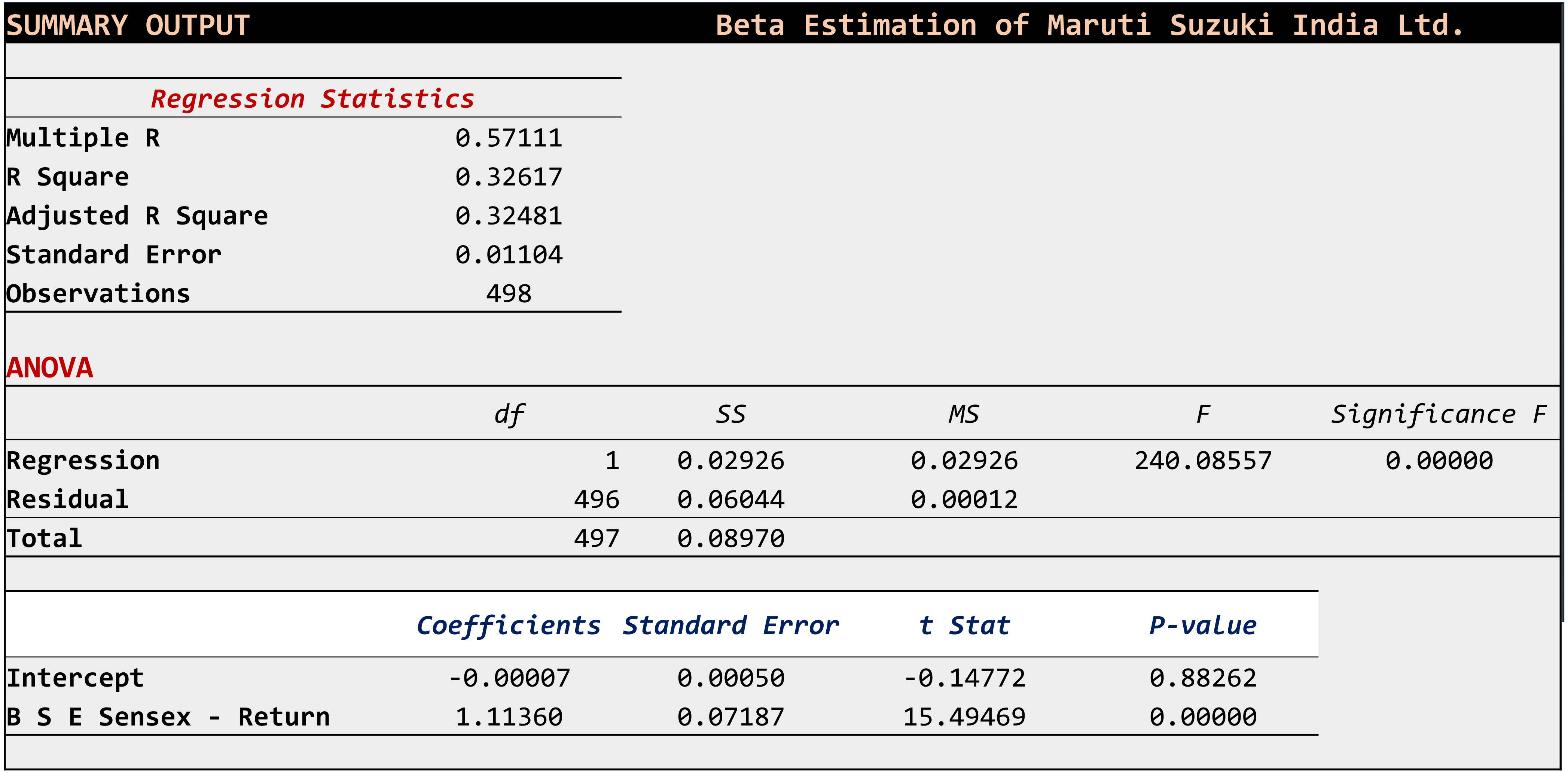
Time Series Analytics

Single Index Model……

$$R = \alpha + \beta Rm + \varepsilon$$

- …gives us the estimation of β;
- …gives us the ideas and the estimation of Systematic Risk and Unsystematic Risk

Getting more meanings out of the following!!!



SUMMARY OUTPUT Beta Estimation of Maruti Suzuki India Ltd.

Regression Statistics	
Multiple R	0.57111
R Square	0.32617
Adjusted R Square	0.32481
Standard Error	0.01104
Observations	498

ANOVA

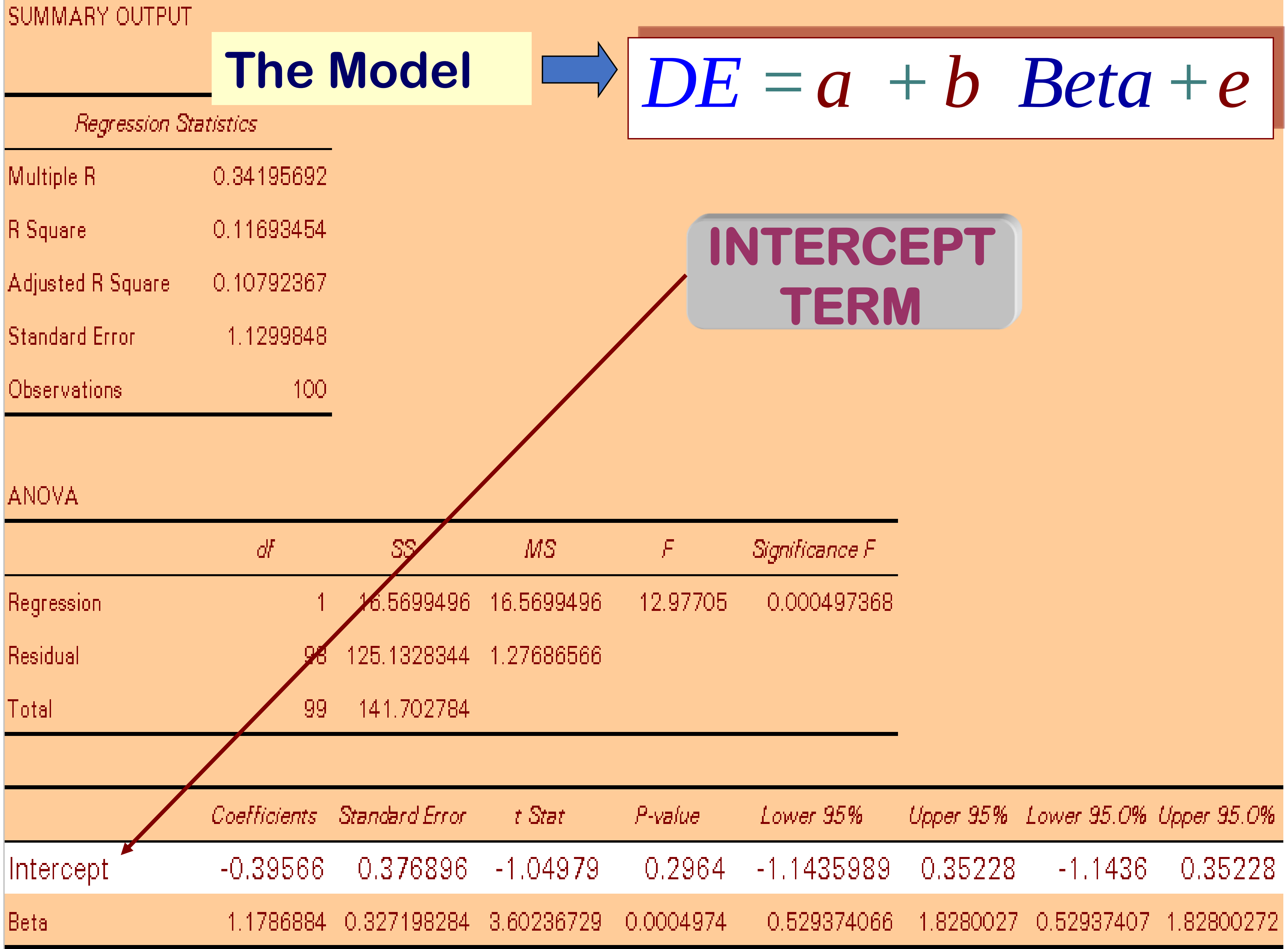
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.02926	0.02926	240.08557	0.00000
Residual	496	0.06044	0.00012		
Total	497	0.08970			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.00007	0.00050	-0.14772	0.88262
B S E Sensex - Return	1.11360	0.07187	15.49469	0.00000

The Model

$$DE = a + b \ Beta + e$$

INTERCEPT TERM



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.34195692
R Square	0.11693454
Adjusted R Square	0.10792367
Standard Error	1.1299848
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	16.5699496	16.5699496	12.97705	0.000497368
Residual	98	125.1328344	1.27686566		
Total	99	141.702784			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.39566	0.376896	-1.04979	0.2964	-1.1435989	0.35228	-1.1436	0.35228
Beta	1.1786884	0.327198284	3.60236729	0.0004974	0.529374066	1.8280027	0.52937407	1.82800272

Remember-

Life is full of challenges!!!

Now… is the time to take on challenges that come in way of Financial Modeling!!!!

Interpret the parameters of the following two models…

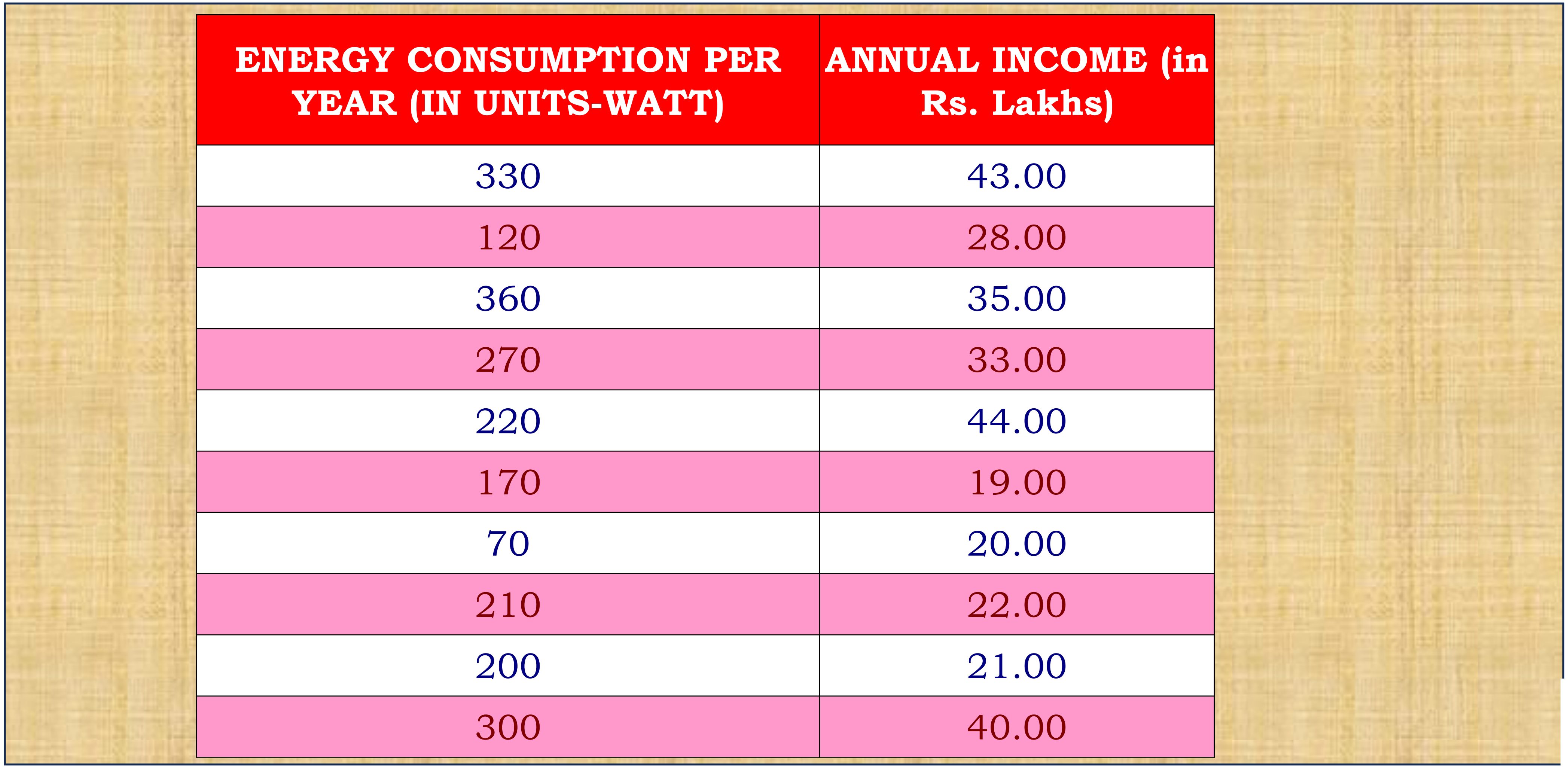
$$\text{Operating\_Profit} = \alpha + \beta\ \text{Rev\_size} + \varepsilon$$

$$\ln(\text{Operating\_Profit}) = \alpha + \beta\ \ln(\text{Rev\_size}) + \varepsilon$$

What is the insight we get?

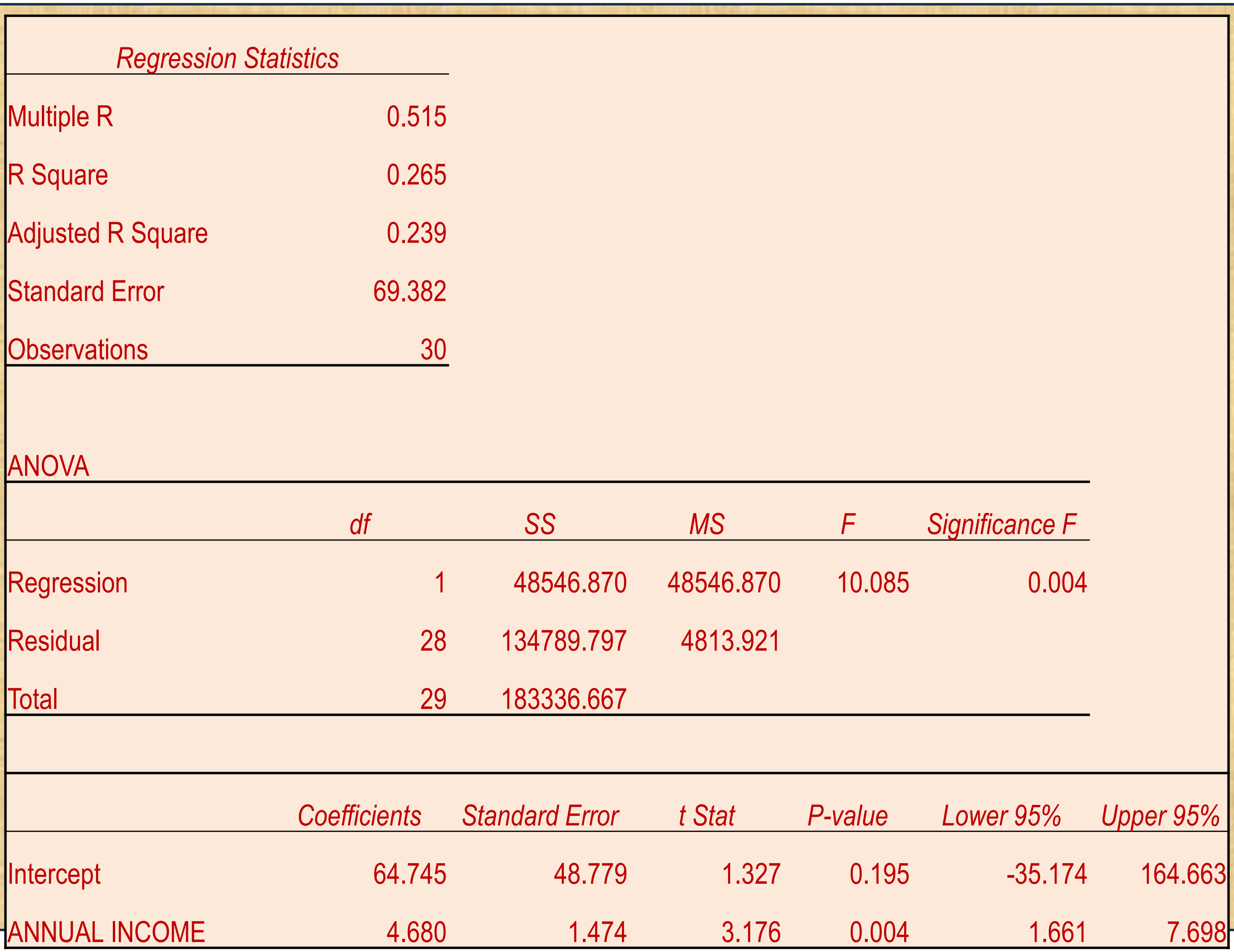
She is working on a research project related to energy consumption and income of households…she wants to determine the elasticity of energy consumption …

She got the data as follows…



ENERGY CONSUMPTION PER YEAR (IN UNITS-WATT)	ANNUAL INCOME (in Rs. Lakhs)
330	43.00
120	28.00
360	35.00
270	33.00
220	44.00
170	19.00
70	20.00
210	22.00
200	21.00
300	40.00

She run simple Regression on it and get the following result…



Regression Statistics	
Multiple R	0.515
R Square	0.265
Adjusted R Square	0.239
Standard Error	69.382
Observations	30

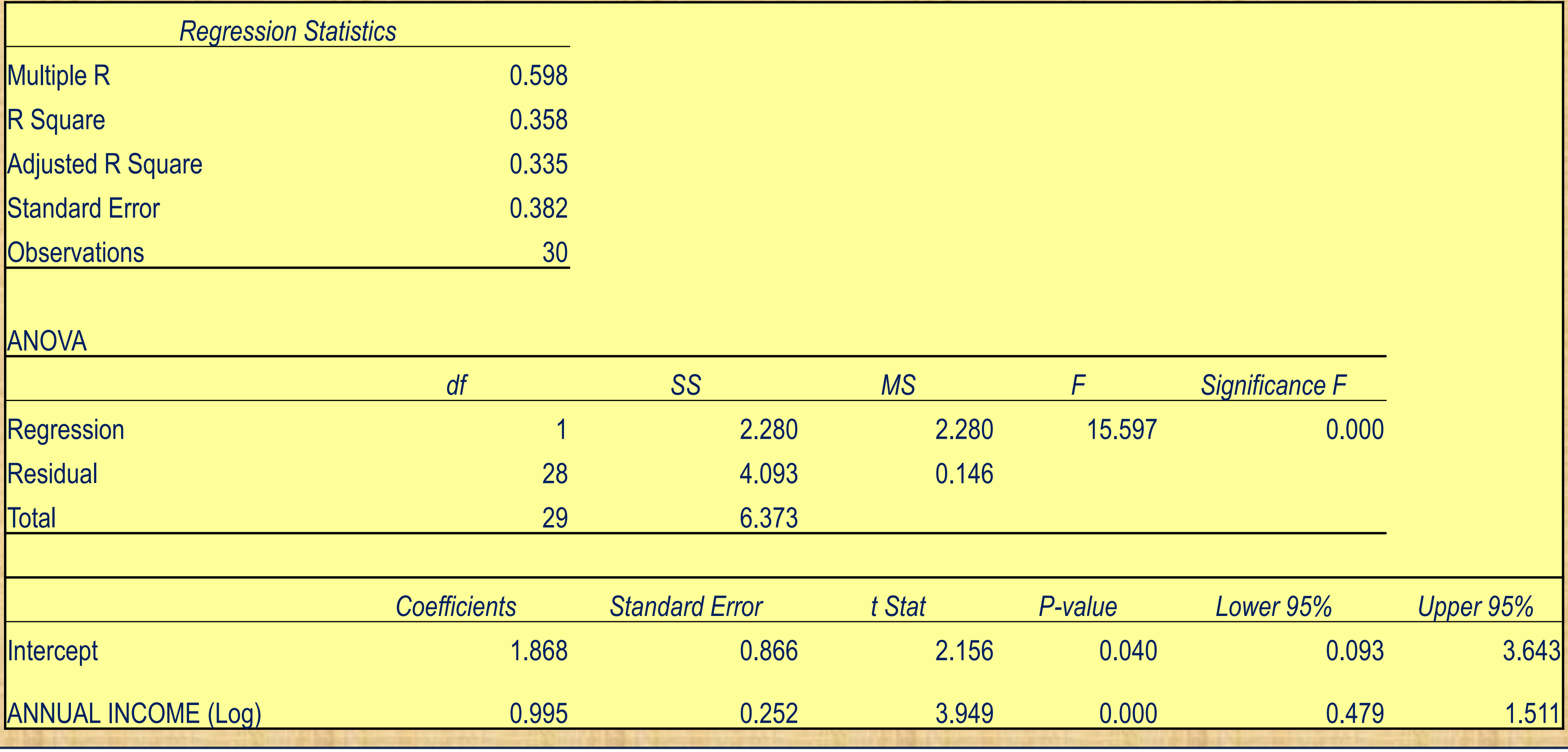
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	48546.870	48546.870	10.085	0.004
Residual	28	134789.797	4813.921		
Total	29	183336.667			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	64.745	48.779	1.327	0.195	-35.174	164.663
ANNUAL INCOME	4.680	1.474	3.176	0.004	1.661	7.698

Getting more meanings out of the following!!!

Then, somebody suggested her to use Log-Log Regression and the output of the same is given below…



Regression Statistics	
Multiple R	0.598
R Square	0.358
Adjusted R Square	0.335
Standard Error	0.382
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2.280	2.280	15.597	0.000
Residual	28	4.093	0.146		
Total	29	6.373			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.868	0.866	2.156	0.040	0.093	3.643
ANNUAL INCOME (Log)	0.995	0.252	3.949	0.000	0.479	1.511

SUMMARY OUTPUT **Oil & Natural Gas Corpn. Ltd.**

Regression Statistics	
Multiple R	0.41386
R Square	0.17128
Adjusted R Square	0.17016
Standard Error	0.01537
Observations	743

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.03617	0.03617	153.15291	0.00000
Residual	741	0.17502	0.00024		
Total	742	0.21119			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.00111	0.00057	-1.96354	0.04996
S & P B S E Sensex	**0.92942**	**0.07510**	**12.37550**	**0.00000**

SUMMARY OUTPUT **Reliance Industries Ltd.**

Regression Statistics	
Multiple R	0.53501
R Square	0.28623
Adjusted R Square	0.28527
Standard Error	0.01372
Observations	743

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.05596	0.05596	297.15646	0.00000
Residual	741	0.13954	0.00019		
Total	742	0.19550			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.00073	0.00050	1.43585	0.15147
S & P B S E Sensex	**1.15599**	**0.06706**	**17.23823**	**0.00000**

MULTIPLE REGRESSION MODEL

- **The General form of the Multiple Regression Model is -**

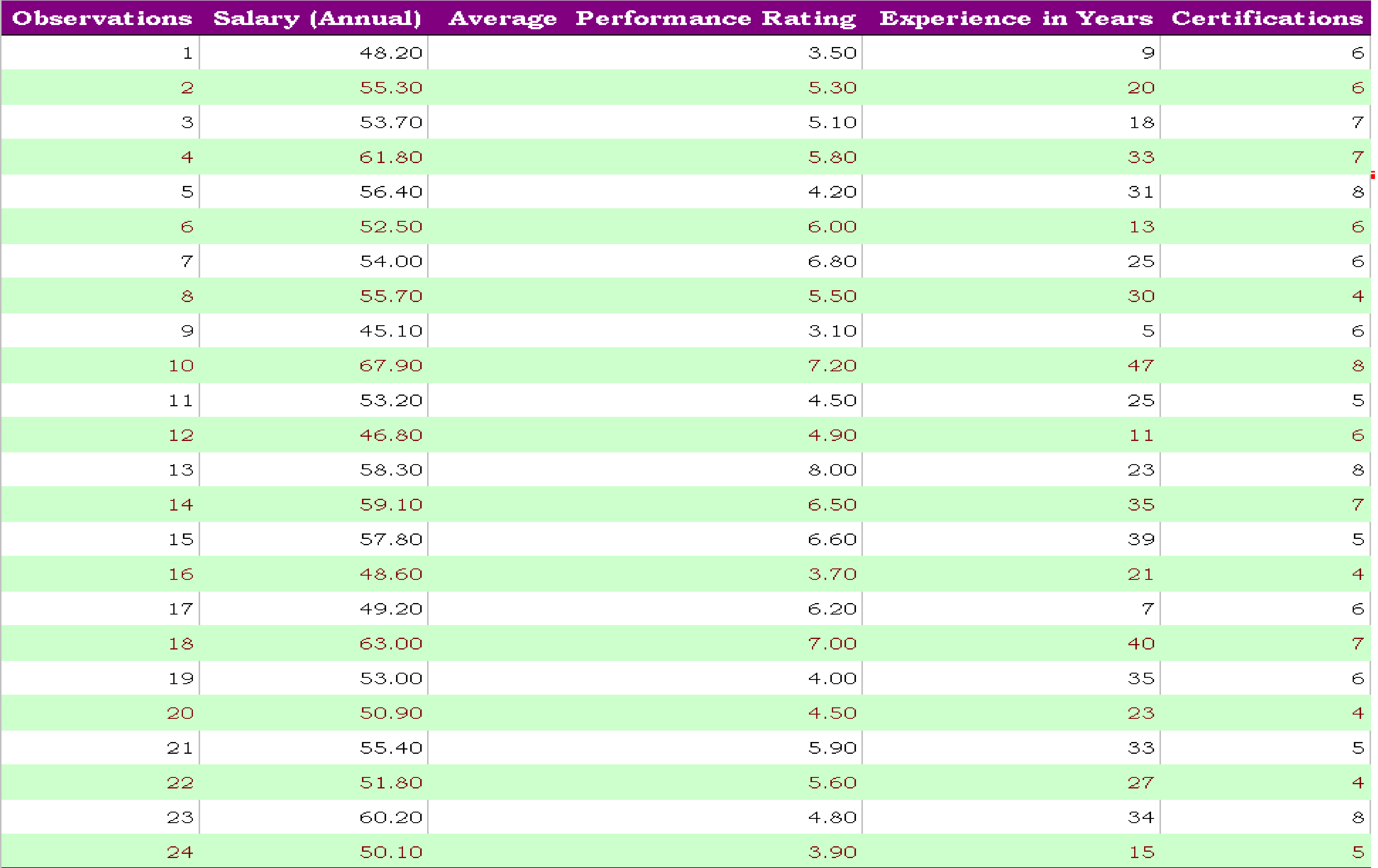
$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_\kappa X_{ki} + \varepsilon_i \;;\; i = 1, \ldots, n$$

- **Where-**
  - **α is called the Intercept Parameter.**
  - **β_is are called the Slope Parameters in the relation between variables *Y and X's.***
  - **ε_i is called Error Term or Disturbance Term or Noise Term.**

Example…(Multiple Regression)

Mr. A. S. Mittal, Director (HR) of a mid-size company has got the data describing the Salary earned during a year (Rs. in thousands) by a machinist in the factory along with the average performance rating over the past 3 years, the years of service, and the number of different machines each employee is certified to operate.

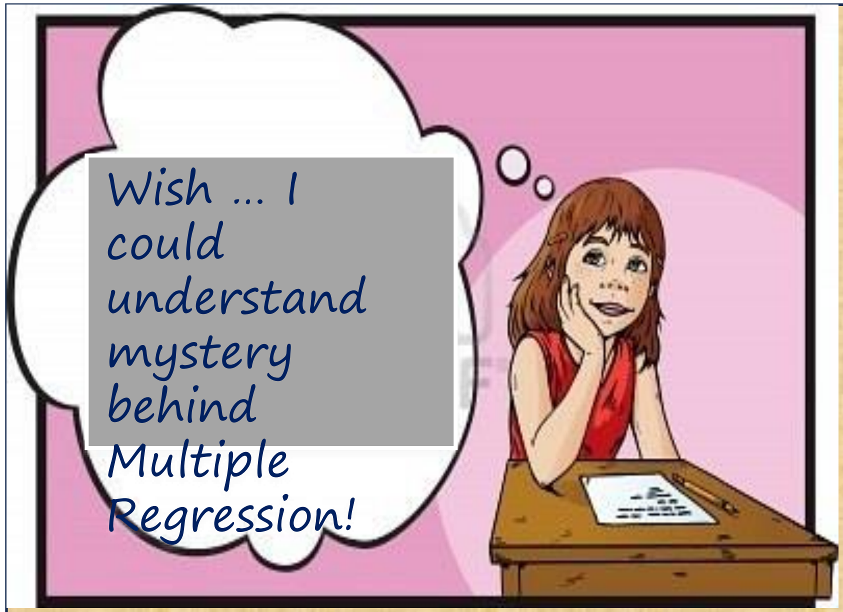
Mr. Mittal wants to build a model so that the average salary to be earned by an employee during a year can be estimated given his or her performance rating in a year, years of service, and certifications.



Observations	Salary (Annual)	Average Performance Rating	Experience in Years	Certifications
1	48.20	3.50	9	6
2	55.30	5.30	20	6
3	53.70	5.10	18	7
4	61.80	5.80	33	7
5	56.40	4.20	31	8
6	52.50	6.00	13	6
7	54.00	6.80	25	6
8	55.70	5.50	30	4
9	45.10	3.10	5	6
10	67.90	7.20	47	8
11	53.20	4.50	25	5
12	46.80	4.90	11	6
13	58.30	8.00	23	8
14	59.10	6.50	35	7
15	57.80	6.60	39	5
16	48.60	3.70	21	4
17	49.20	6.20	7	6
18	63.00	7.00	40	7
19	53.00	4.00	35	6
20	50.90	4.50	23	4
21	55.40	5.90	33	5
22	51.80	5.60	27	4
23	60.20	4.80	34	8
24	50.10	3.90	15	5

विद्या परं दैवतम्
IIM
VISAKHAPATNAM

Getting more meanings out of the following!!!



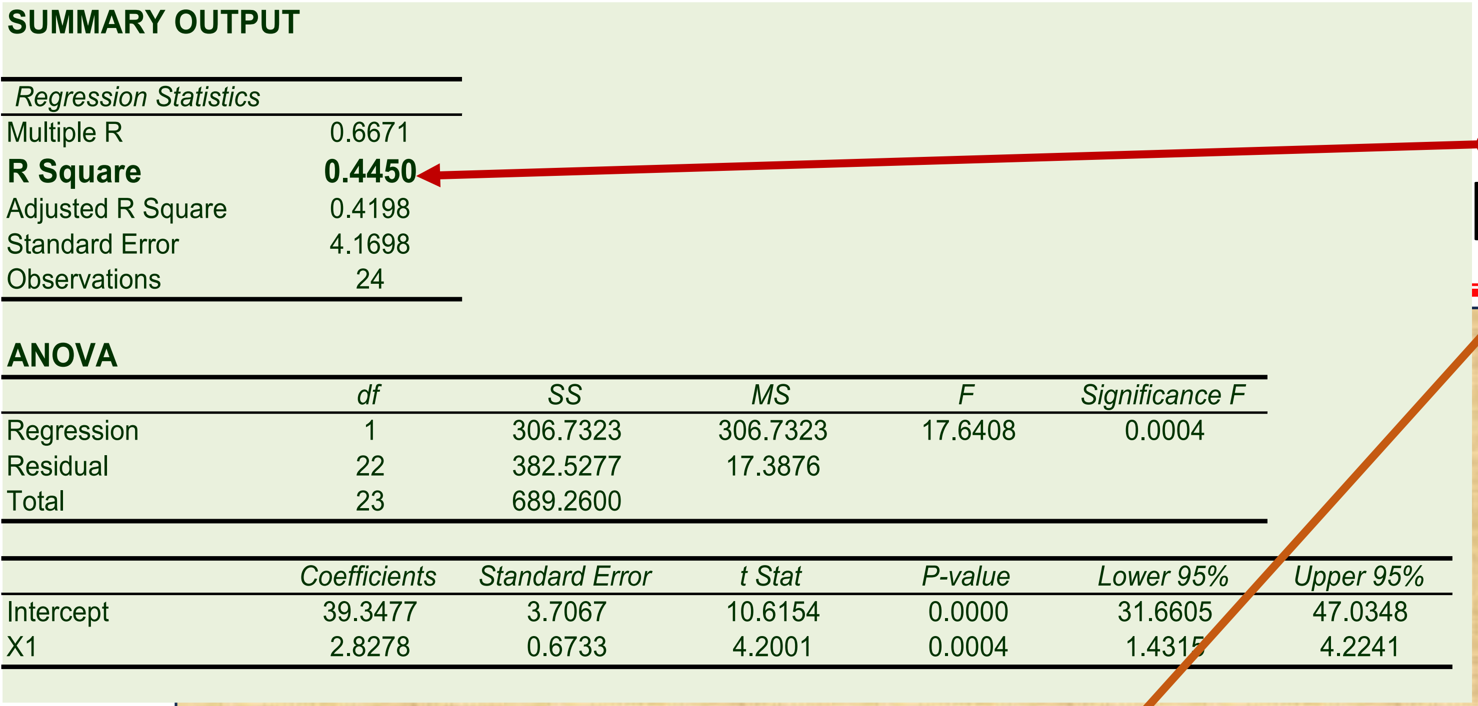
Assume that you are working with Simple Regression Model…

$$Y = \alpha + \beta X_1 + \varepsilon$$

- What would happen to our results if we introduce another variable X_2 into our model?
- If we run a model only with X_1 and only with X_2 **separately**, can we say that total explained variation will be **the sum of their R^2**?

lowing!!

Look at the Values of R^2!



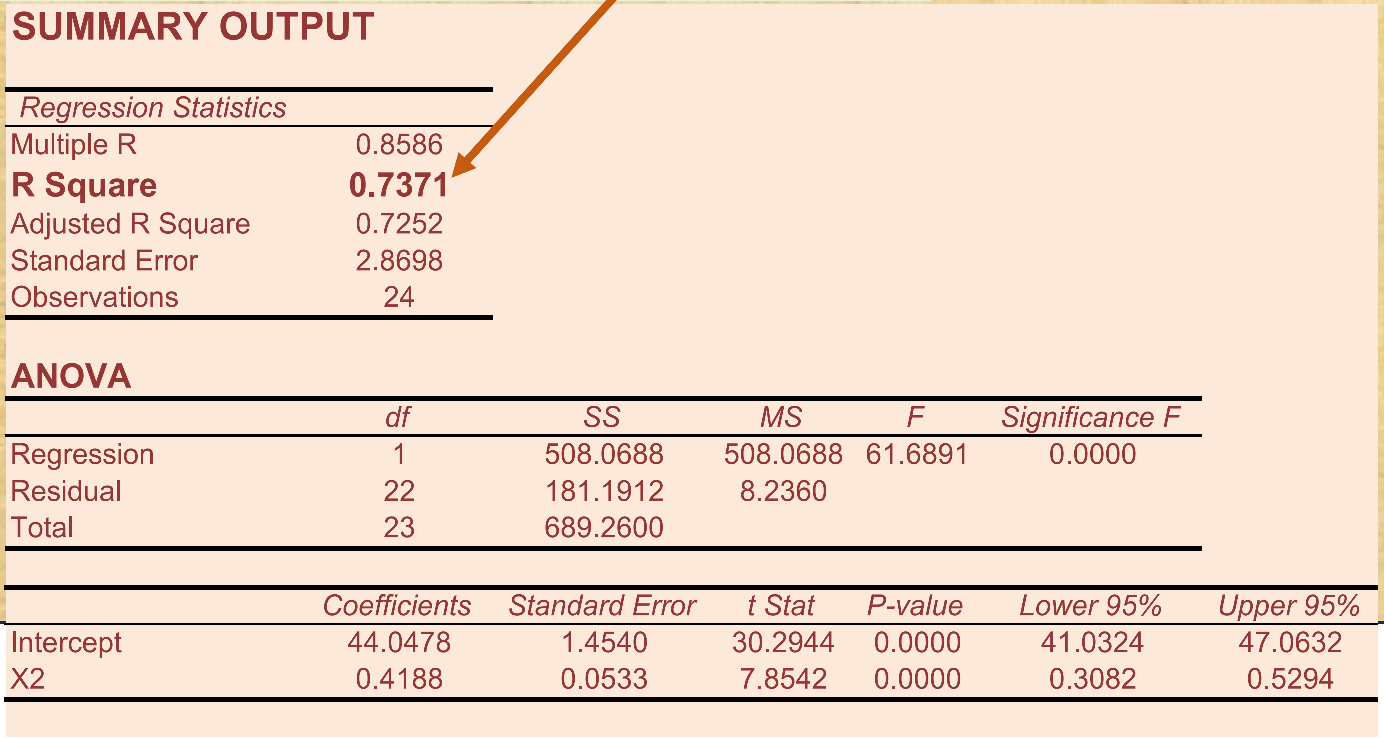
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6671
R Square	**0.4450**
Adjusted R Square	0.4198
Standard Error	4.1698
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	306.7323	306.7323	17.6408	0.0004
Residual	22	382.5277	17.3876		
Total	23	689.2600			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	39.3477	3.7067	10.6154	0.0000	31.6605	47.0348
X1	2.8278	0.6733	4.2001	0.0004	1.4315	4.2241



SUMMARY OUTPUT

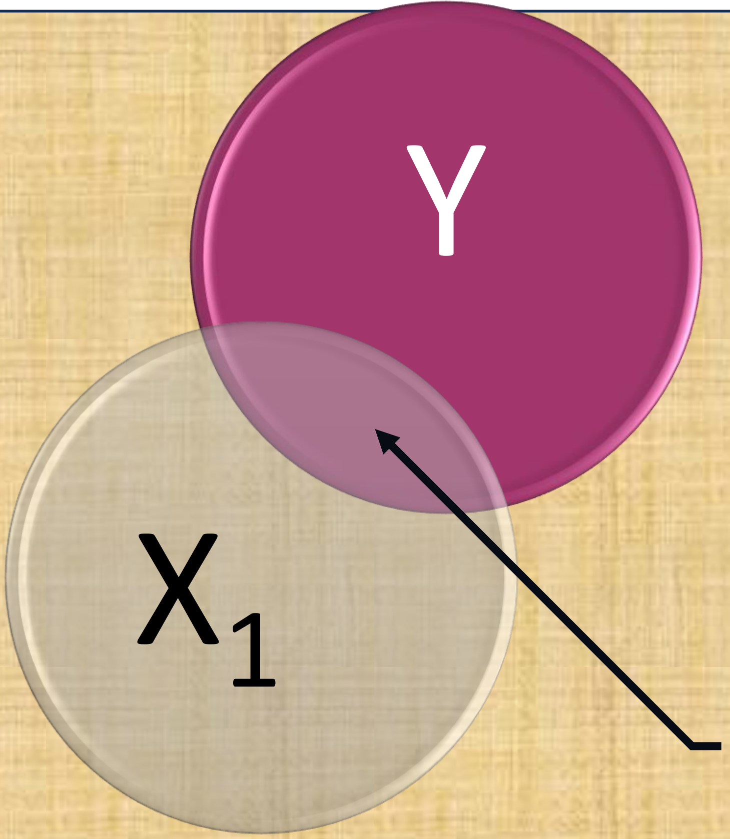
Regression Statistics	
Multiple R	0.8586
R Square	**0.7371**
Adjusted R Square	0.7252
Standard Error	2.8698
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	508.0688	508.0688	61.6891	0.0000
Residual	22	181.1912	8.2360		
Total	23	689.2600			

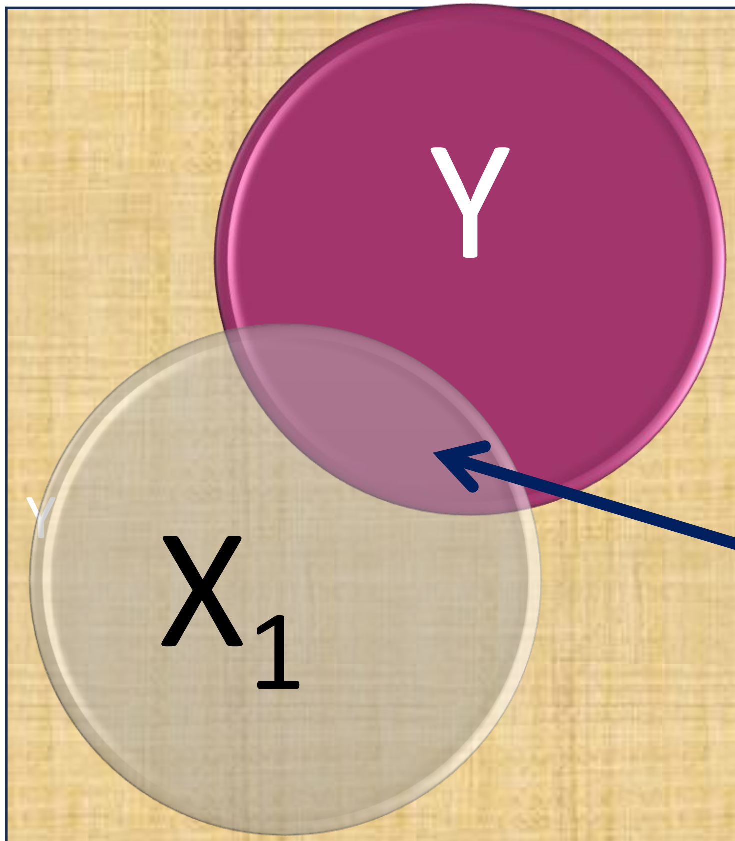
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	44.0478	1.4540	30.2944	0.0000	41.0324	47.0632
X2	0.4188	0.0533	7.8542	0.0000	0.3082	0.5294

Let’s understand Multiple Regression through Venn Diagrams …



This area of Y is EXPLAINED by X_1.

This Venn Diagram suggests the following regression model …



SUMMARY OUTPUT

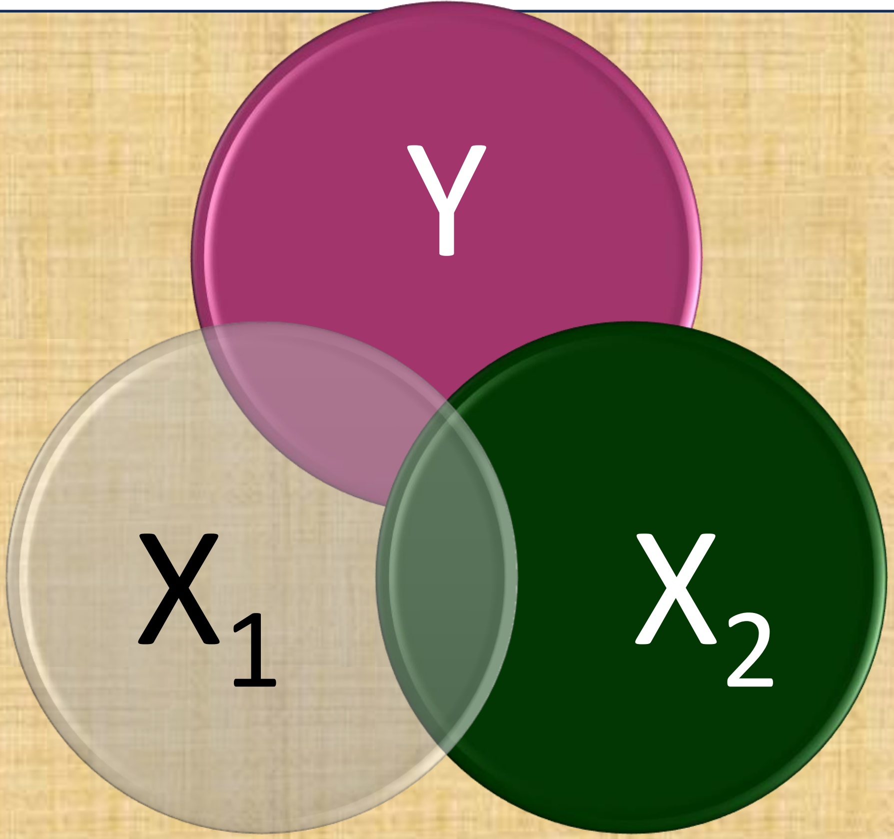
Regression Statistics	
Multiple R	0.6671
R Square	**0.4450**
Adjusted R Square	0.4198
Standard Error	4.1698
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	306.7323	306.7323	17.6408	0.0004
Residual	22	382.5277	17.3876		
Total	23	689.2600			

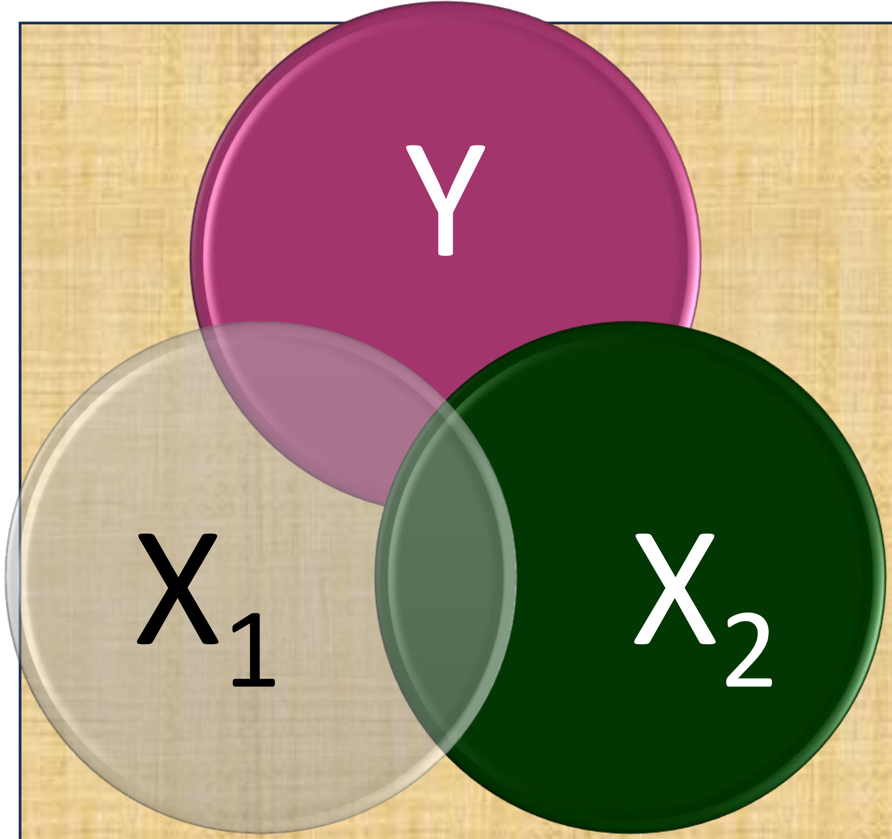
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	39.3477	3.7067	10.6154	0.0000	31.6605	47.0348
X1	2.8278	0.6733	4.2001	0.0004	1.4315	4.2241

What X_2 is supposed to do in the regression model?



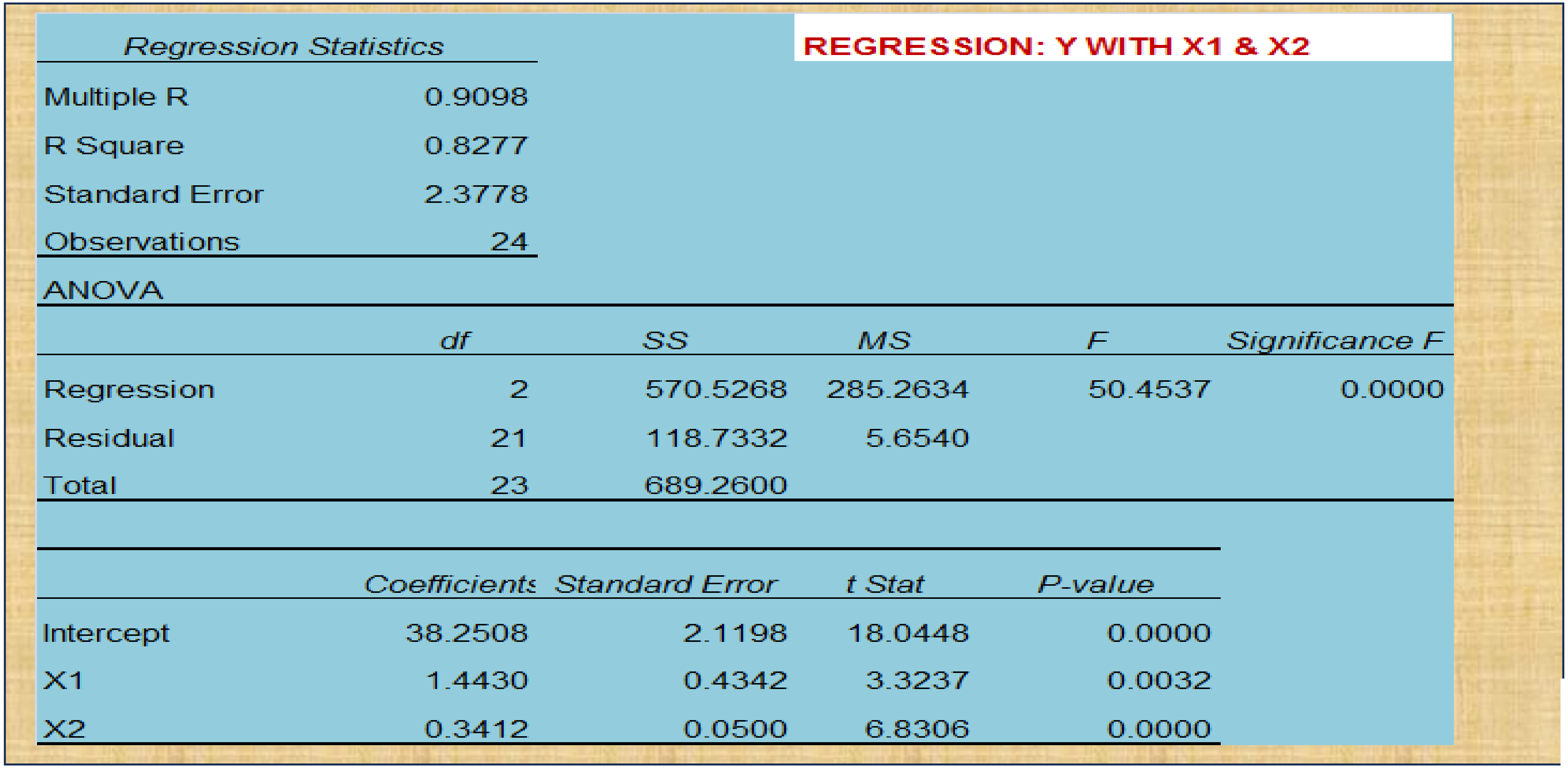
X_2 tries to explain that part of Y which is not explained by X_1 and through that which is NOT there already in X_1.

It means that X_2 will explain the residuals of Y which are left after regressing with X_1.



Will the whole of X_2 explain the residuals of Y ? NO.

Only that part of X_2 which is not contained in X_1!!!



REGRESSION: Y WITH X1 & X2

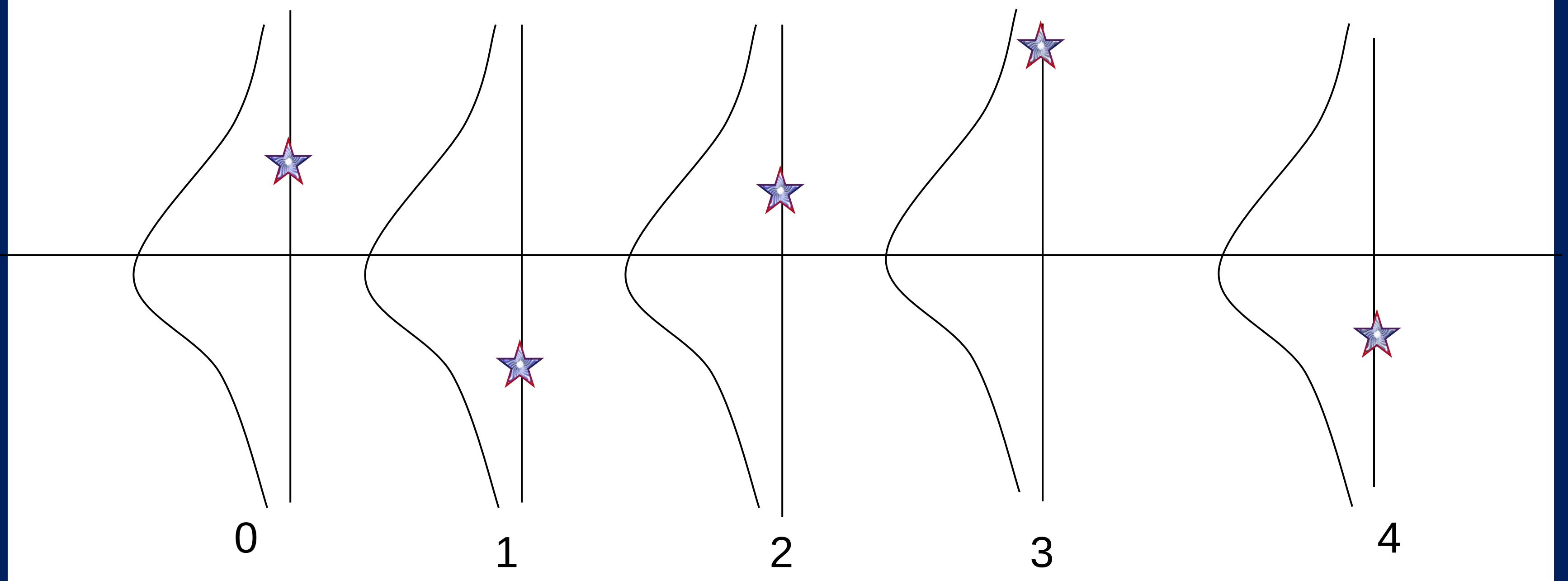
Regression Statistics	
Multiple R	0.9098
R Square	0.8277
Standard Error	2.3778
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	570.5268	285.2634	50.4537	0.0000
Residual	21	118.7332	5.6540		
Total	23	689.2600			

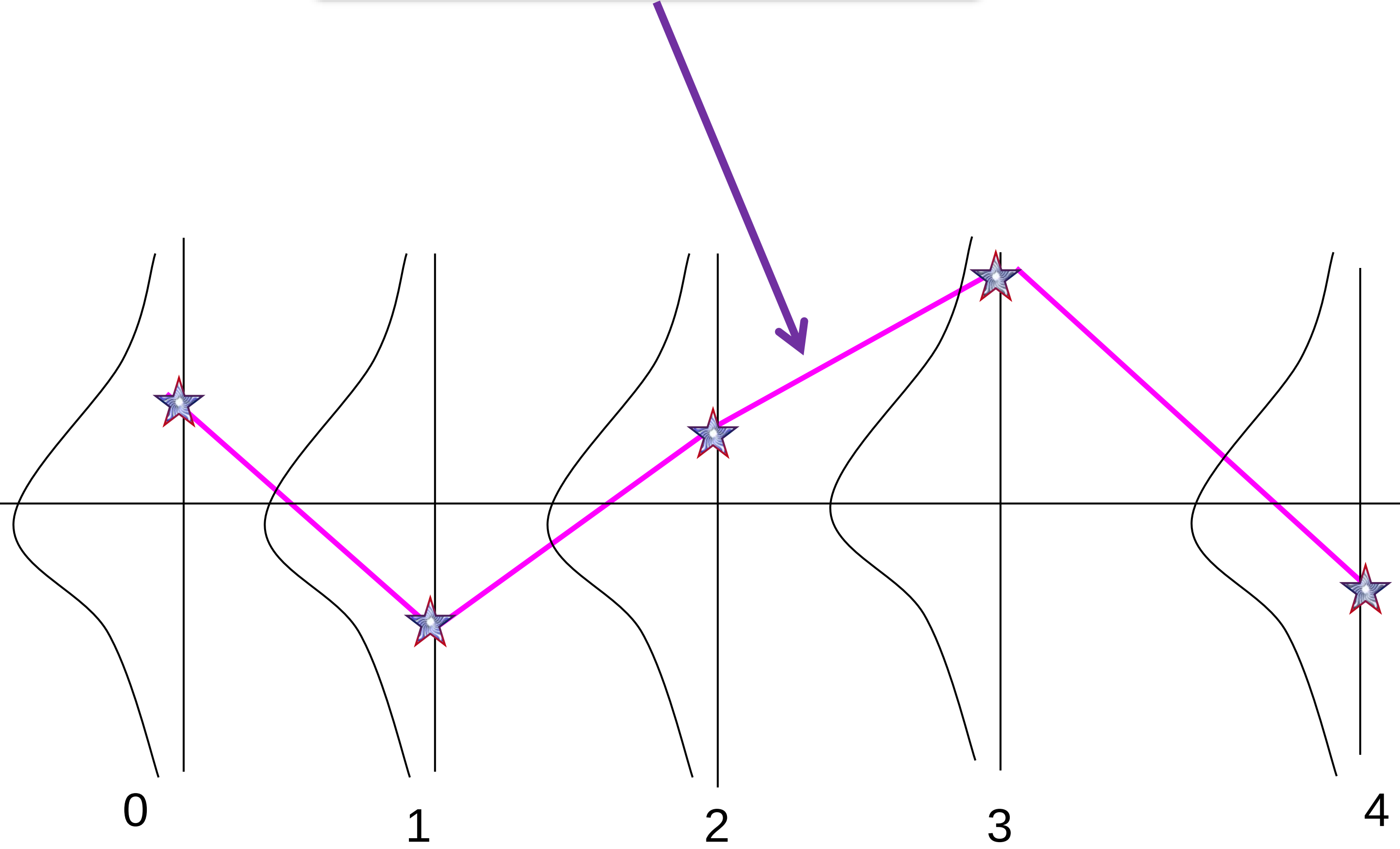
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	38.2508	2.1198	18.0448	0.0000
X1	1.4430	0.4342	3.3237	0.0032
X2	0.3412	0.0500	6.8306	0.0000

How many possible values can happen on day zero?

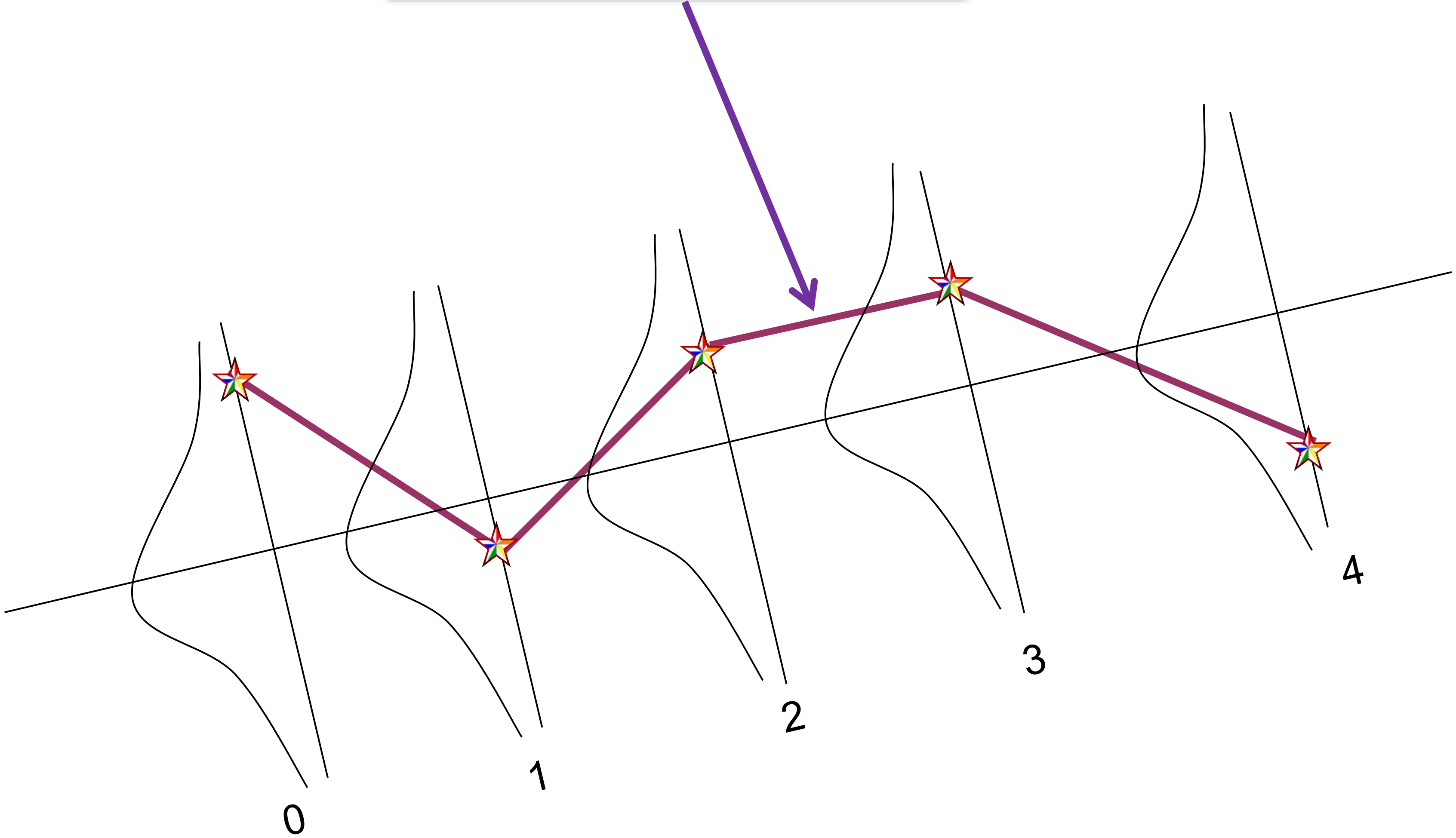


But, we realized only ONE!!! Similarly, for other days.

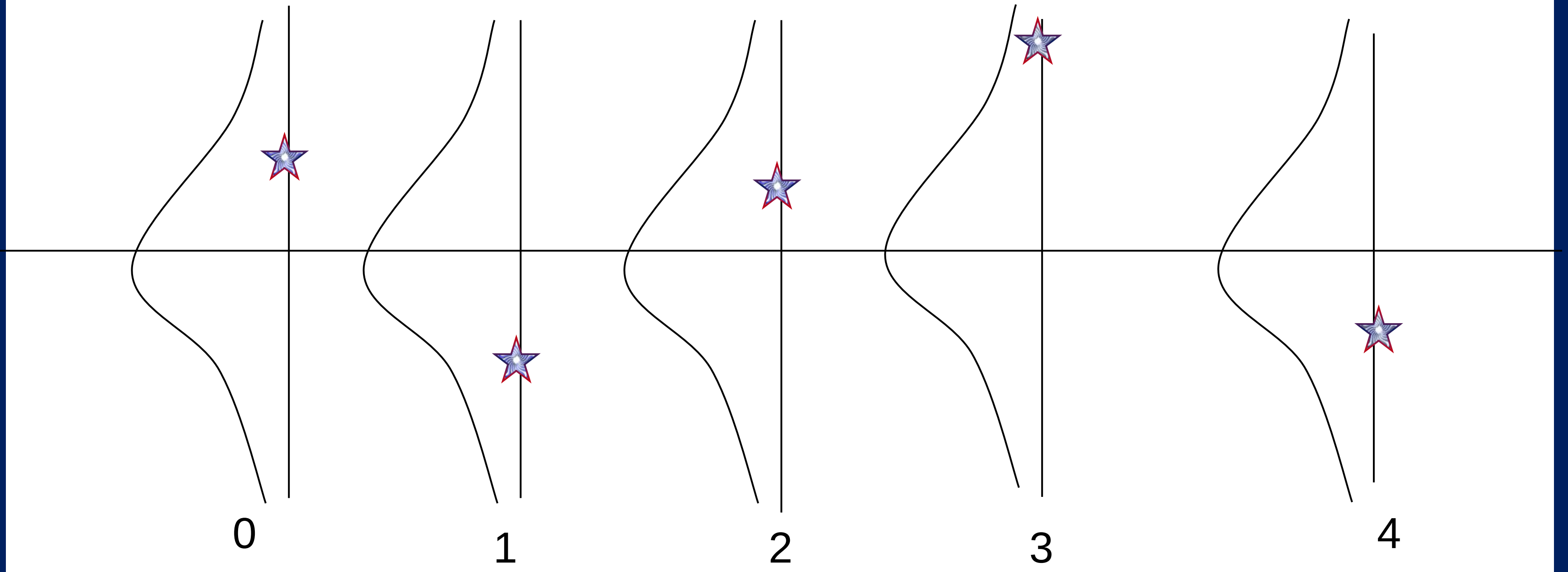
This is time series!!!



This is time series!!!

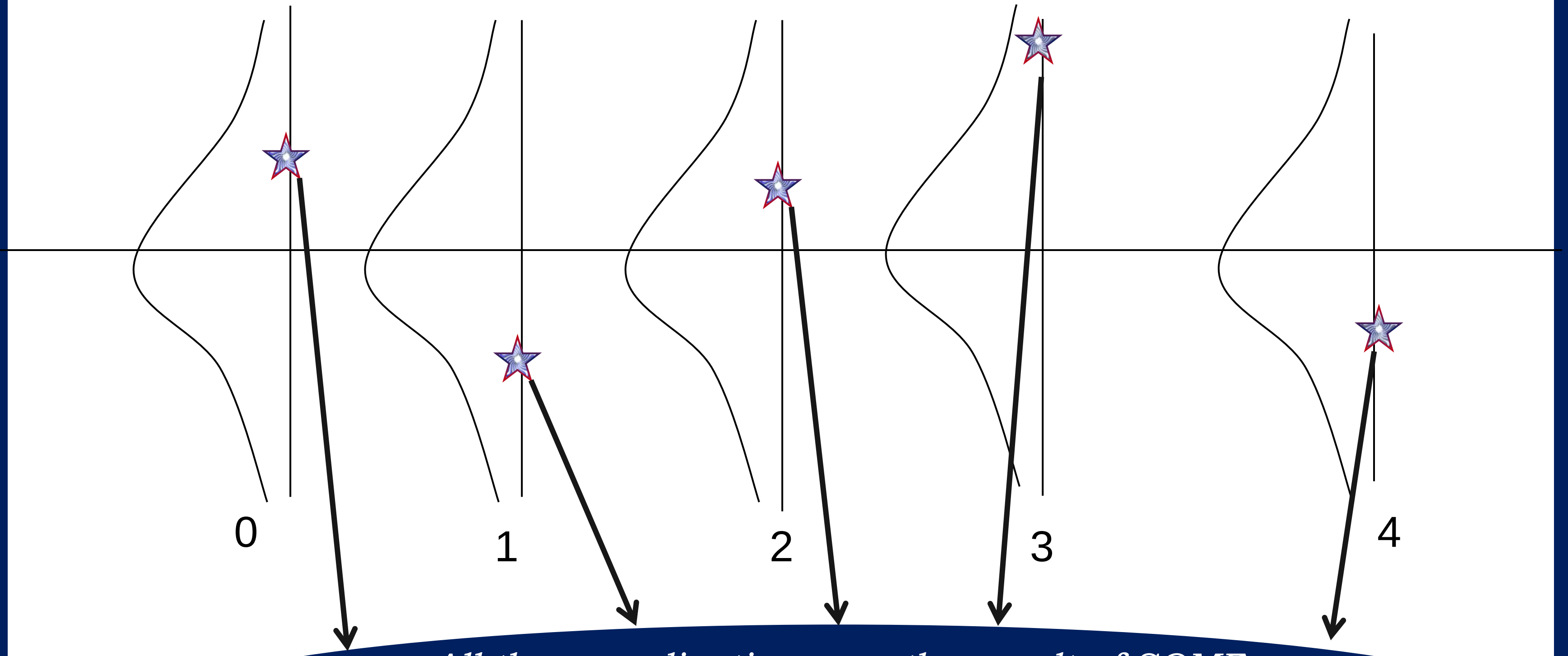


Important thing to note is … that …



Time Series deals with ONE realization out of infinite possibilities.

And, this one realization is the result of some underlying stochastic process!



All these realizations are the result of SOME underlying stochastic process whose behaviour is modeled through Normal Probability Distribution.

Now, we define Time Series…

- Realization of a particular stochastic process is called a **TIME SERIES.**

- Every ***realized*** observation of a stochastic process is understood as a **TIME SERIES.**

Now, let’s move to the building block of a stochastic process in finance.

Let’s take the equation…

- $P_t = P_0 e^{rt+\sigma dW} \Leftrightarrow \log P_t = \log(P_0) + rt + \ \sigma W_t$

- Taking derivative of the equation, we get –

$$\frac{dP}{P} = r\, dt + \ \sigma dW_t$$

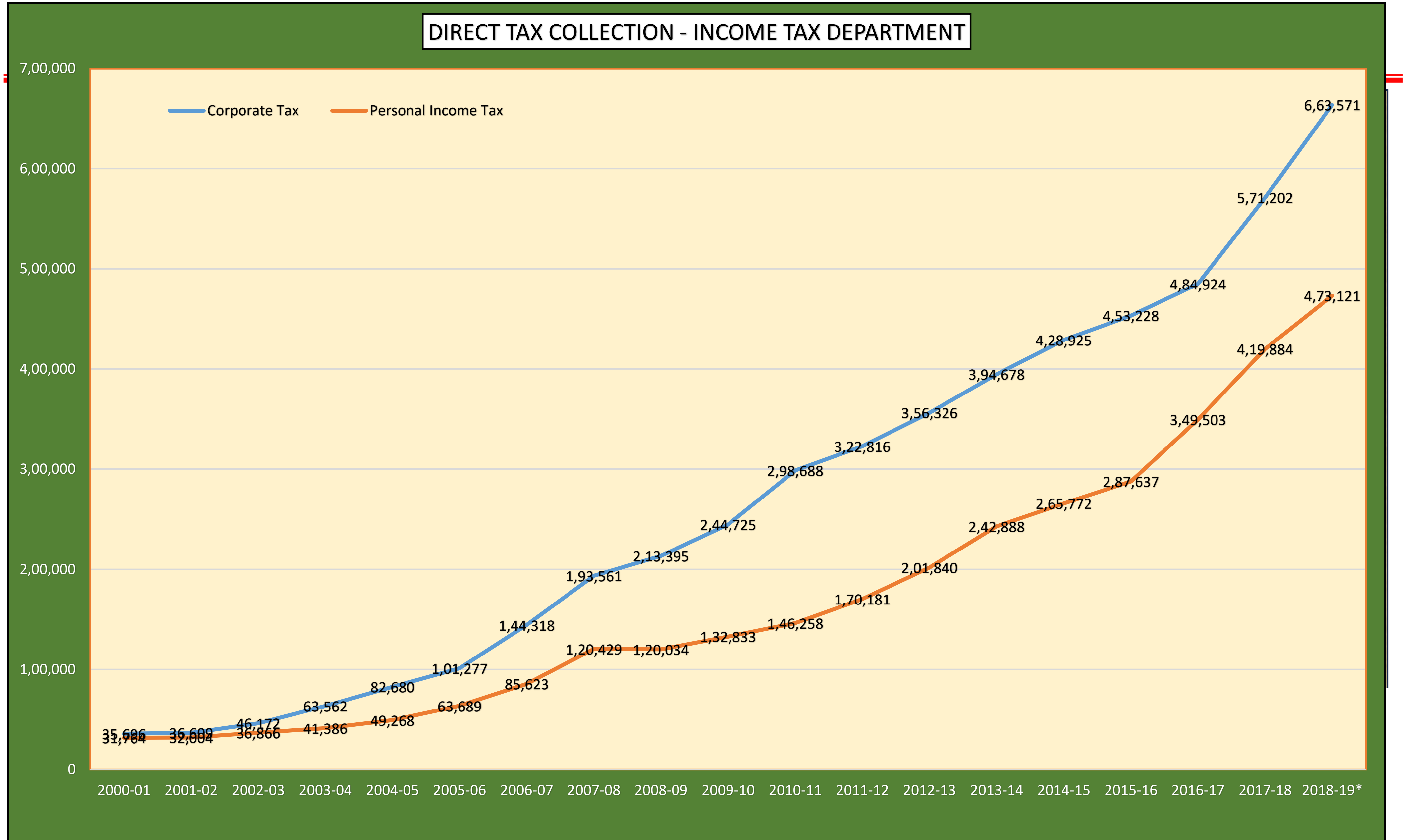
$$dP = r\ P\ dt + \ \sigma\ P\ dW_t$$

What is this equation?

- Also, try to understand what this equation represents

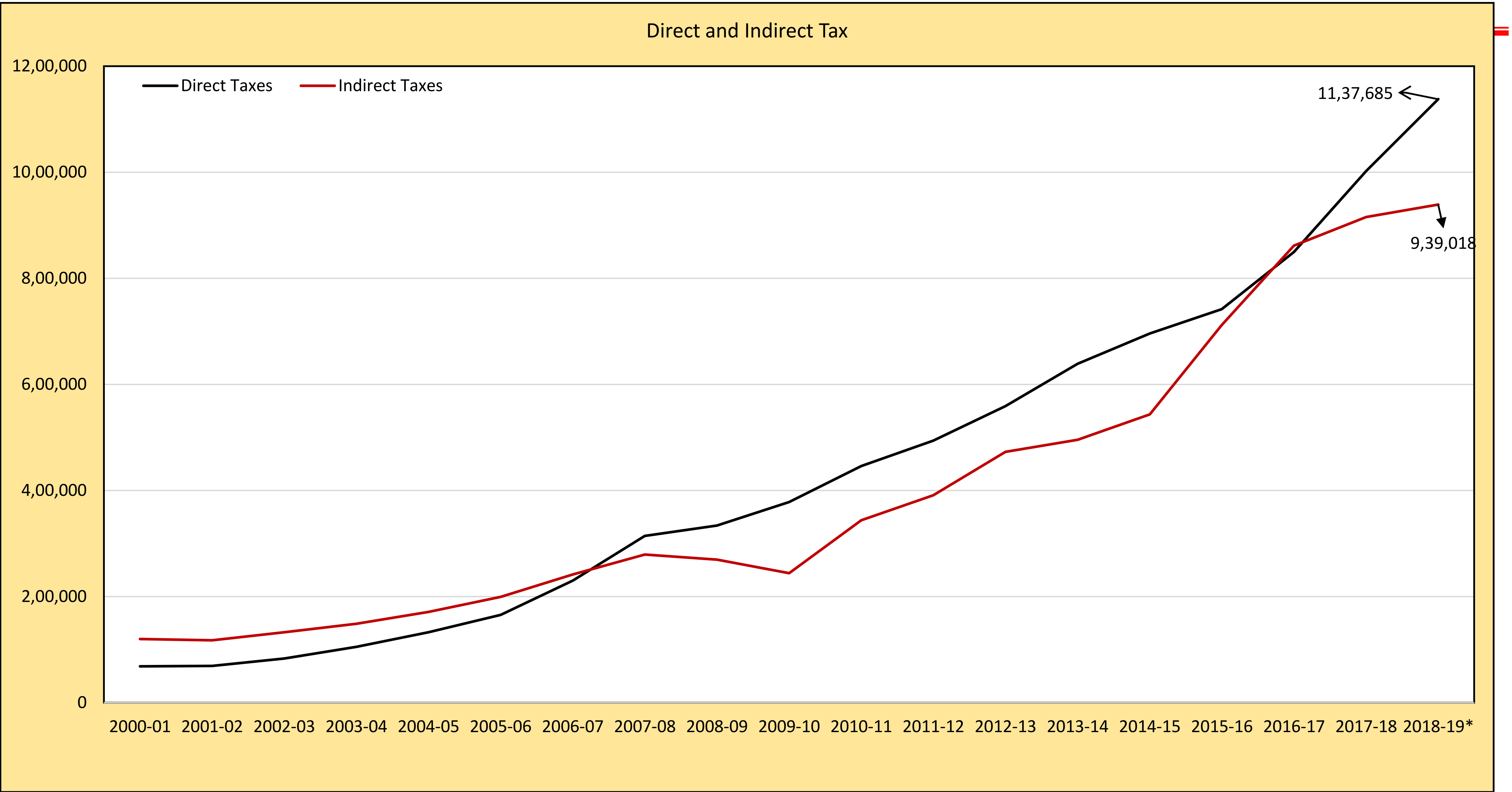
$$\log P_t - \log(P_0) = \ rt + \ \sigma W_t$$

Let’s have a look at the tax data of the Income Tax Department of GOI…

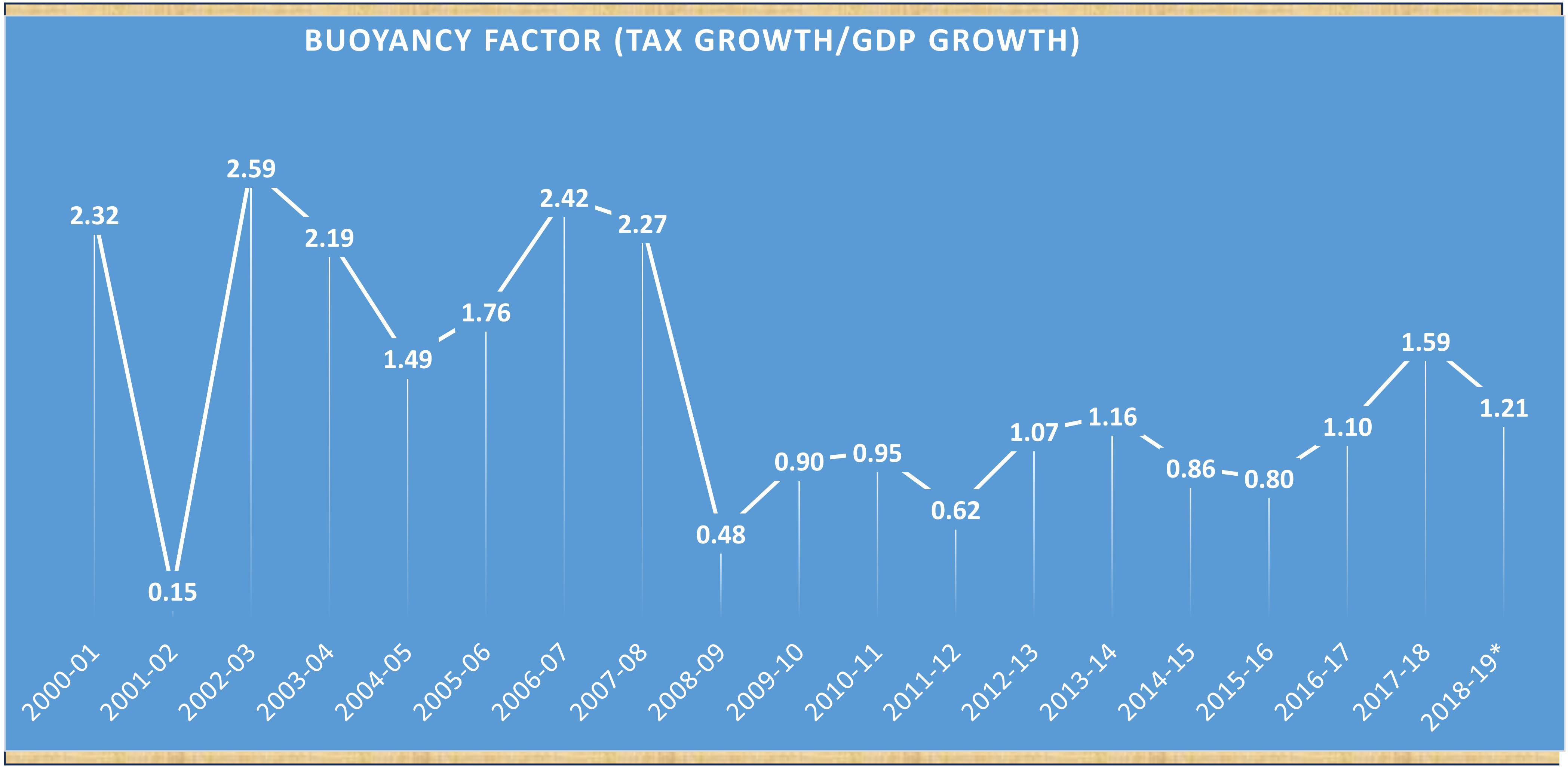


DIRECT TAX COLLECTION - INCOME TAX DEPARTMENT

What comes to your mind when you see the following data?…

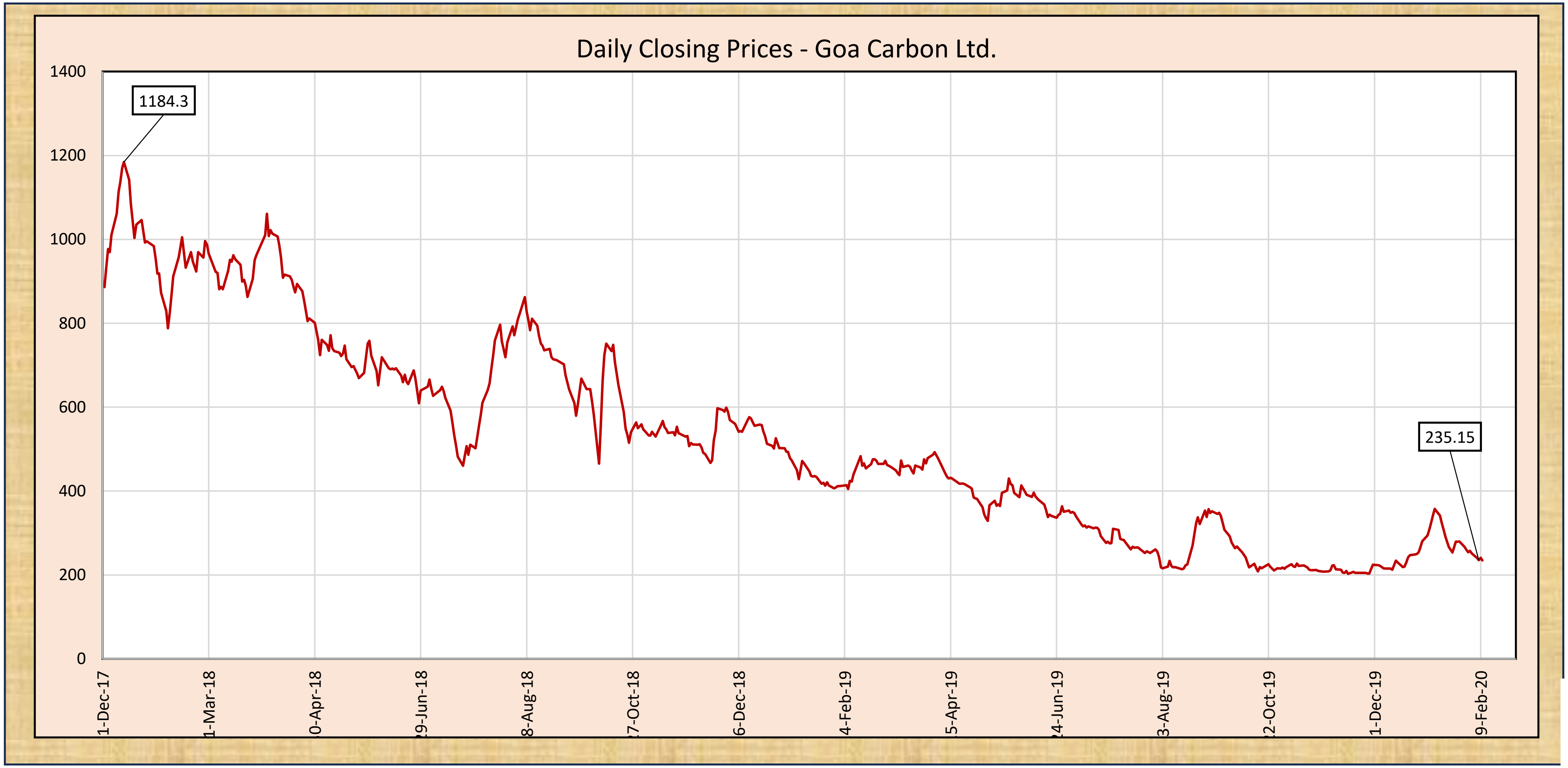


Do you see the difference in the behaviour of earlier charts and this chart? Will you use same tools for analysis?



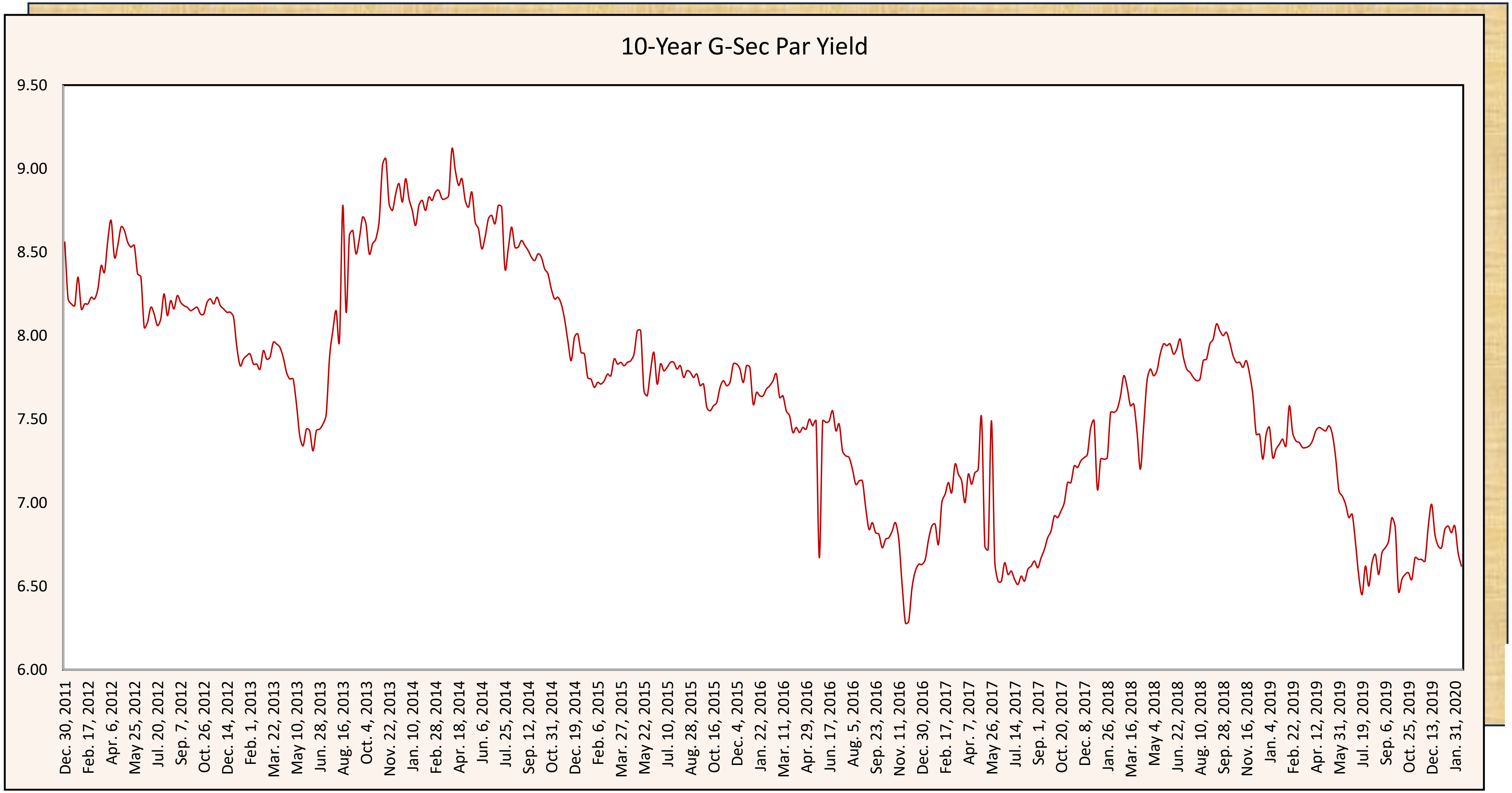
What difference you find in the share price data?

Look at the share prices of Goa Carbon Ltd.



विद्या परं दैवतम्
IIM
VISAKHAPATNAM

Look at the behaviour of 10-Year GOI bonds



Think…

What kind of data you are seeing?

Do we have some challenges in analyzing the data?

Can we use OLS for the analysis of such data for meaningful inferences?

Let's dare to enter the world of TIME SERIES!!!

Issues related to Time Series

Issues:

- **FORECASTING**
- **ESTIMATION**
- **ESTABLISHING THE RELATION**

Specific issues in Finance related to Time Series

Issues:

- **Forecasting** – Share Prices, Interest Rates.
- **Estimating** – Mean (Return) and Variance (Risk)
- **Establishing the relation** – Relation between the spot rates and forward rates

Features of Time Series Data

Time Series data may be related to their own past

- (such as autoregressive process, moving average process)

They may be subject to deterministic trends, seasonality, cycles etc.

- (Such trend in sales, growth rate of profit)

In multi-variate context, they may have cross-correlations.

- (For example; advertisement expenses for the month of January will affect the sale not only in the month of January but may affects the sale of February and coming months.)

Revisit the purposes for which time series are used …

- Forecasting
- Estimation
- Establishing relation

For all these, we need ***INFORMATION****!*

But, where is the ***INFORMATION*** *hidden in a time series?*

Where is the information hidden in a time series?

Is it in **TIME**?

Is it in the **OWN PAST** of a time series?

Is it in **OWN RESIDUAL**?

Is it in **SOME OTHER TIME SERIES**?

The whole fight in a time series modeling is …

Identify the source of information in a time series which can be used for forecasting, estimation and establishing relation.

Once the source of information is identified, how to use the same to estimate the parameters of the model to get the **BEST estimators**!

Take a pause, meditate on what we have discussed and learned so far!

When we run from SILENCE, we run from OURSELVES; please come back to yourself!

I trust that NOW we are capable of taking challenges of time series!

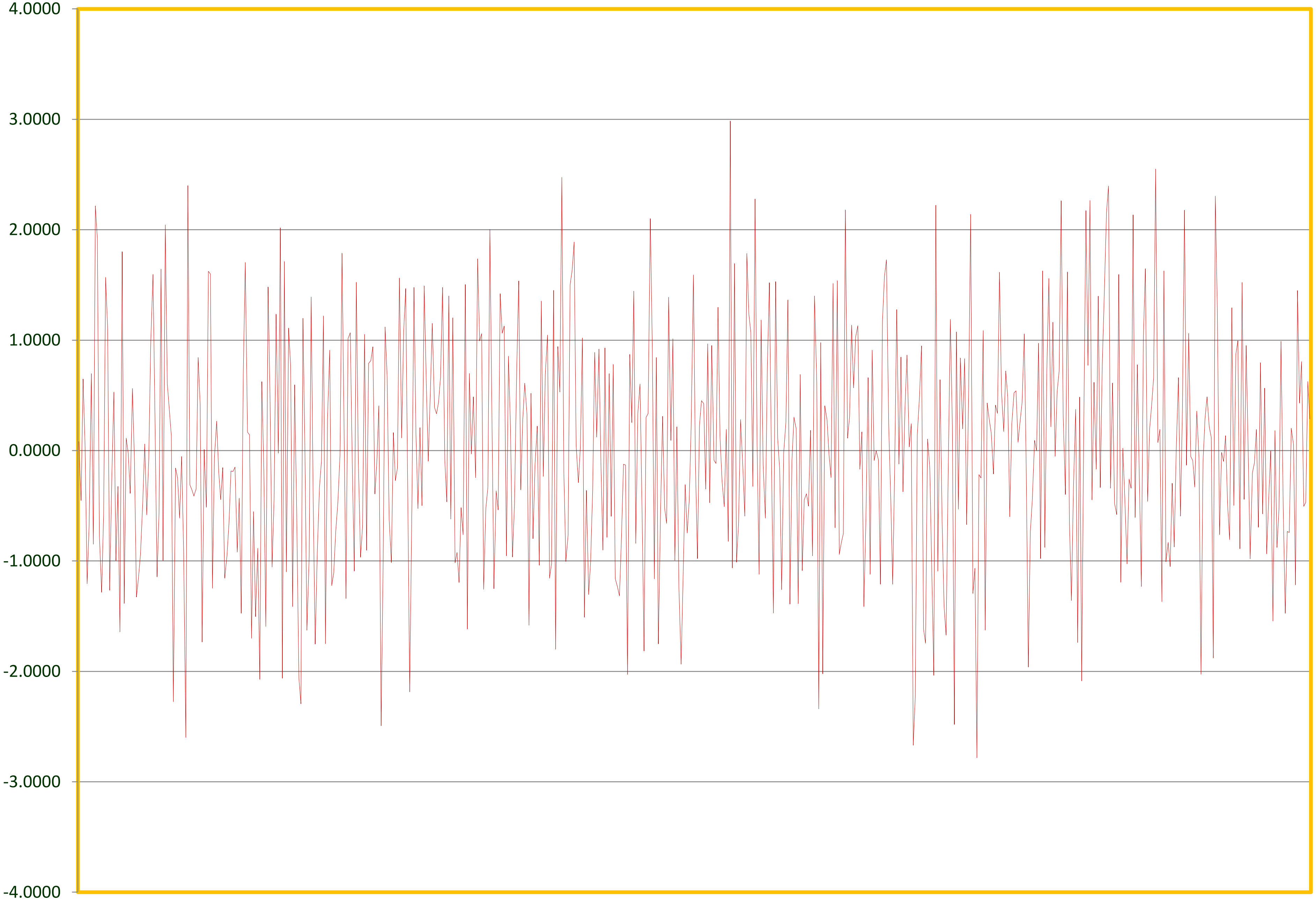
Tools to Identify a underlying behavior of Time Series

Time Series Plot

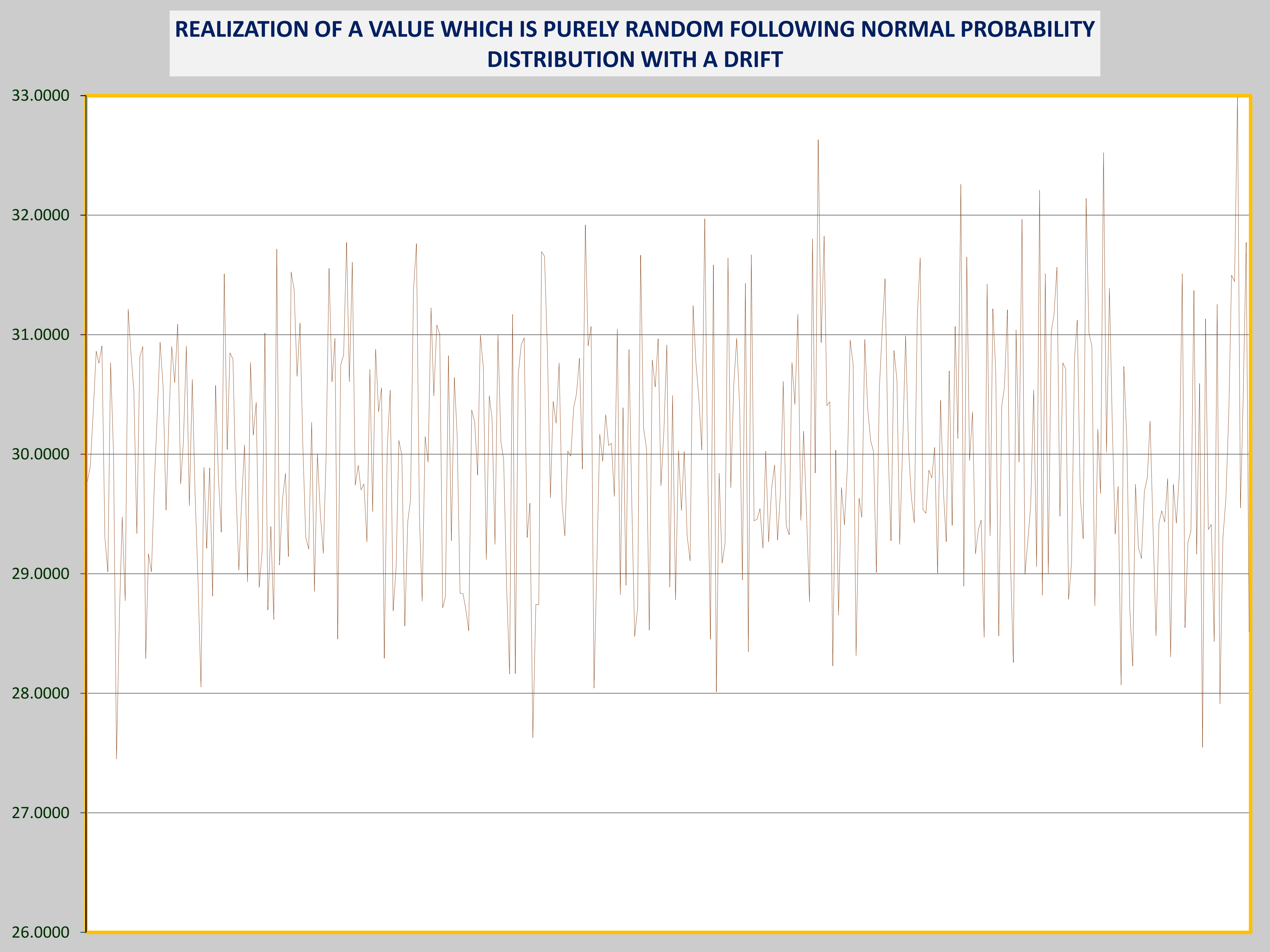
Correlogram

REALIZATION OF A VALUE WHICH IS PURELY RANDOM FOLLOWING NORMAL PROBABILITY DISTRIBUTION

A S



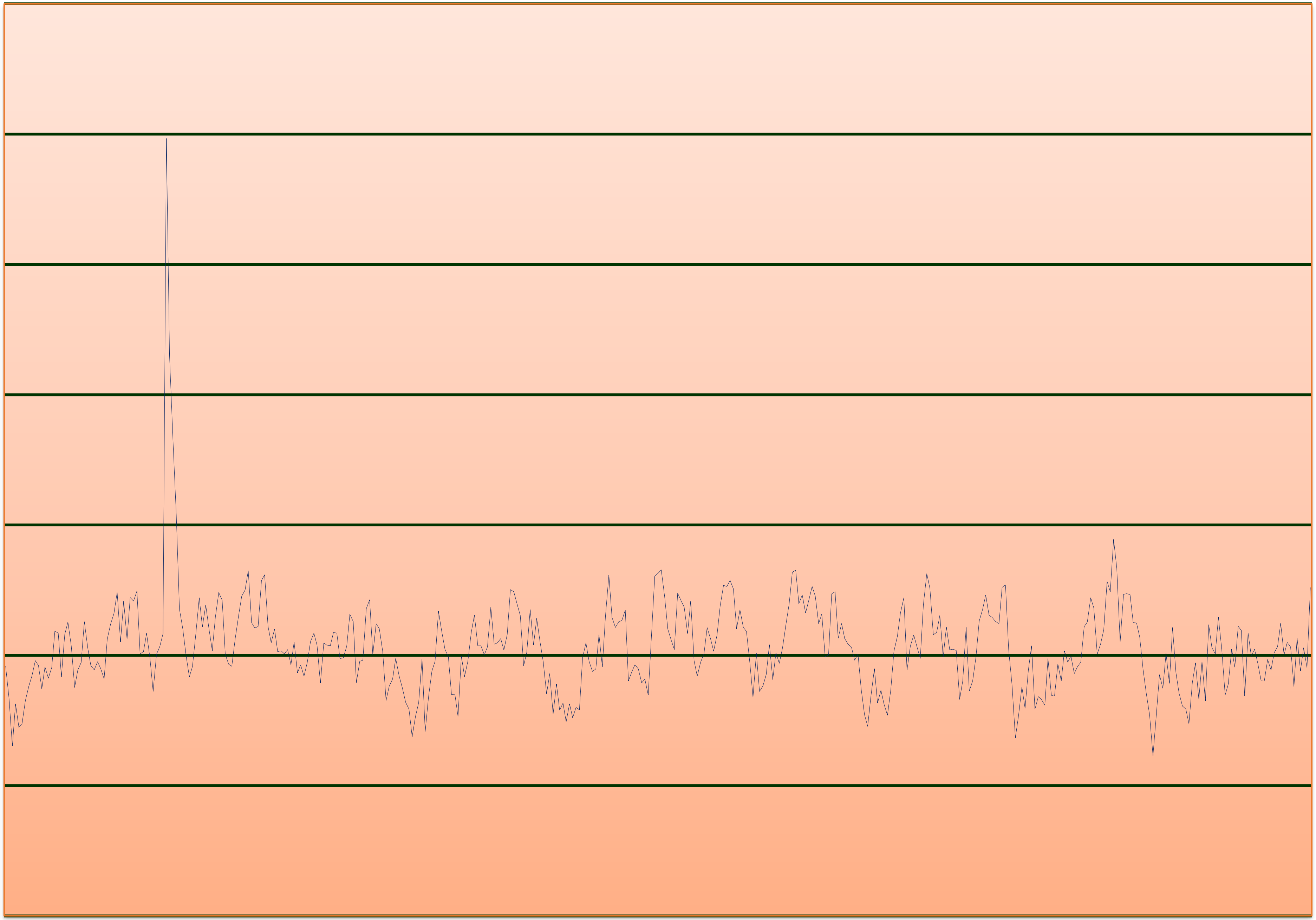
An


REALIZATION OF A VALUE WHICH IS PURELY RANDOM FOLLOWING NORMAL PROBABILITY DISTRIBUTION WITH A DRIFT
33.0000
32.0000
31.0000
30.0000
29.0000
28.0000
27.0000
26.0000

Another Time Series…..is its behaviour same as those of the earlier ones?

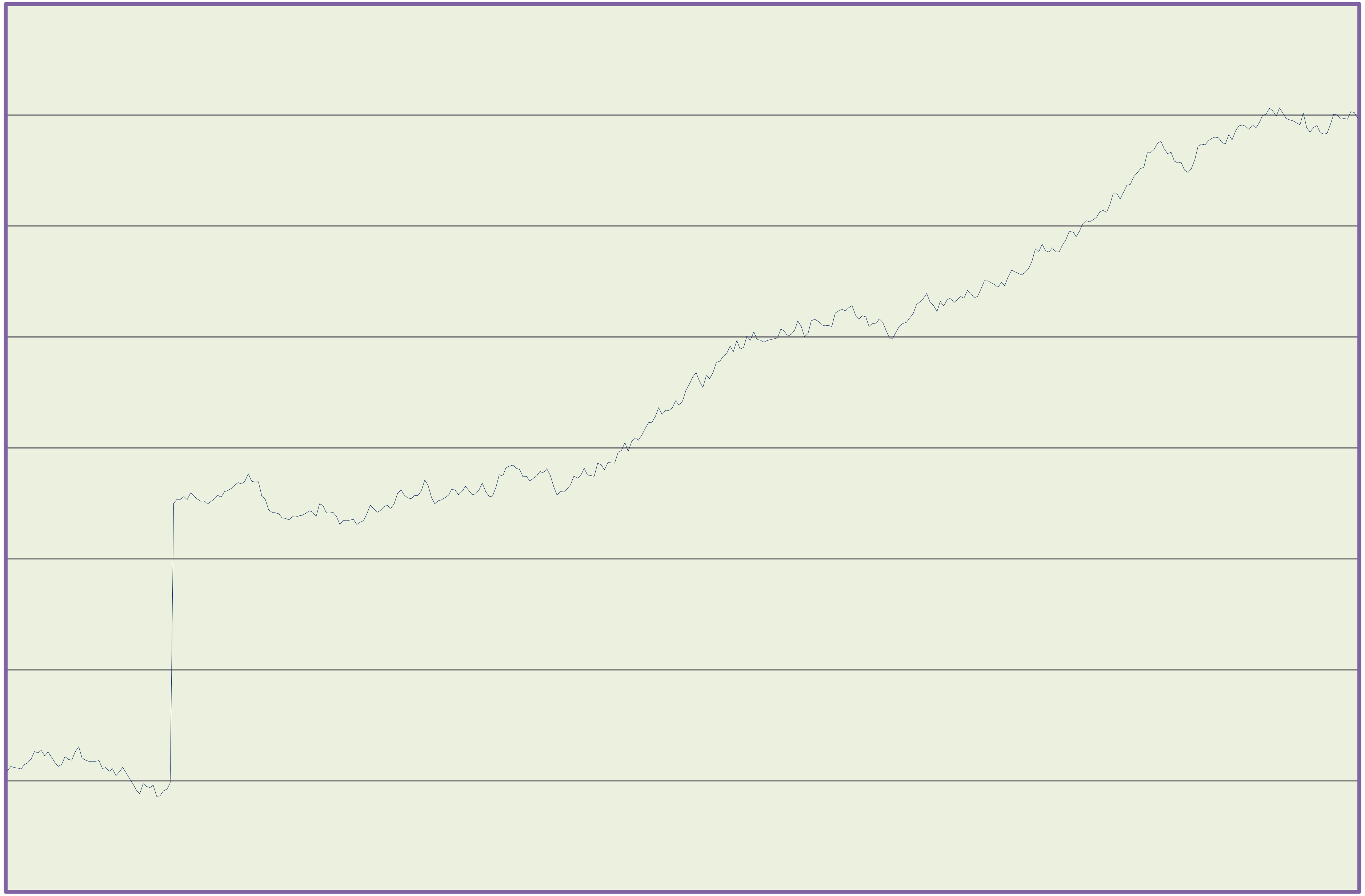
Lo

REALIZATION OF A VALUE WHICH IS HAVING SHOCKS

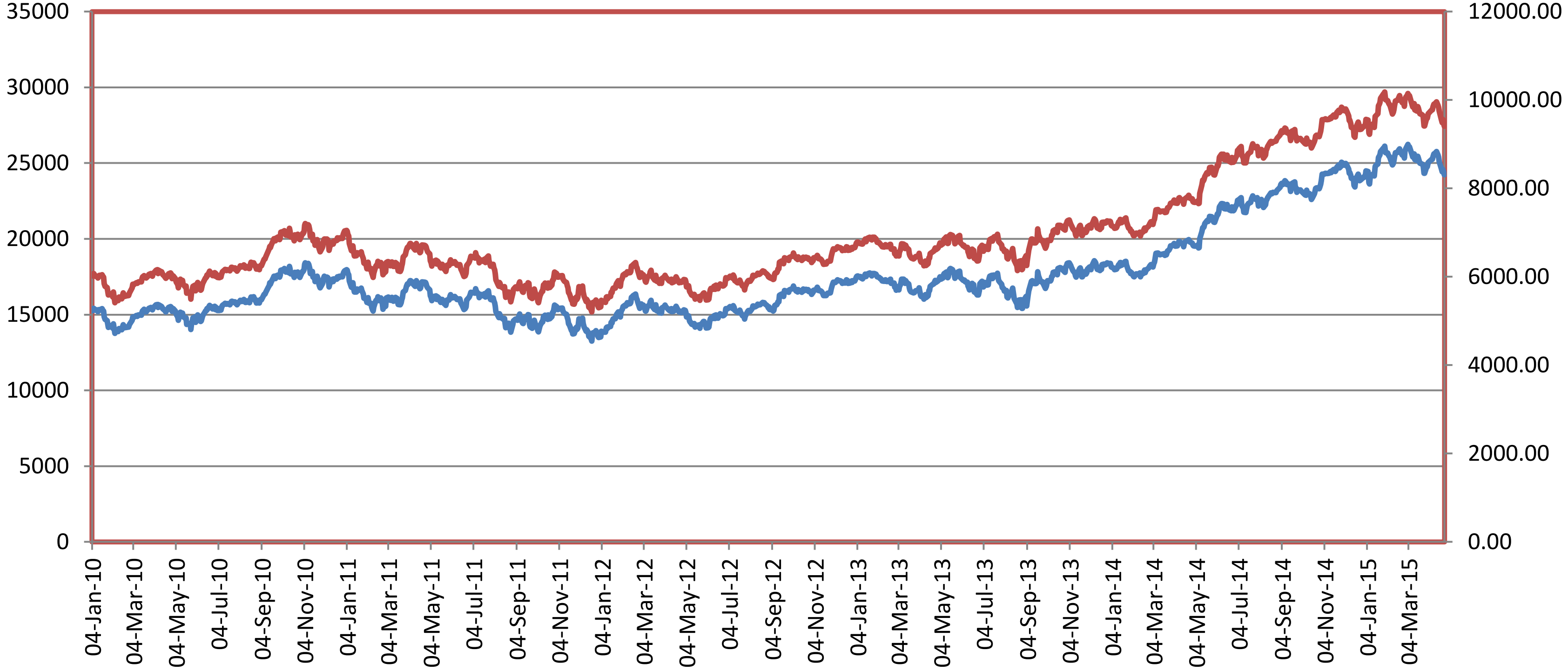


Look at this graph – a different one?

REALIZATION OF A VALUE WHICH IS HAVING SHOCKS



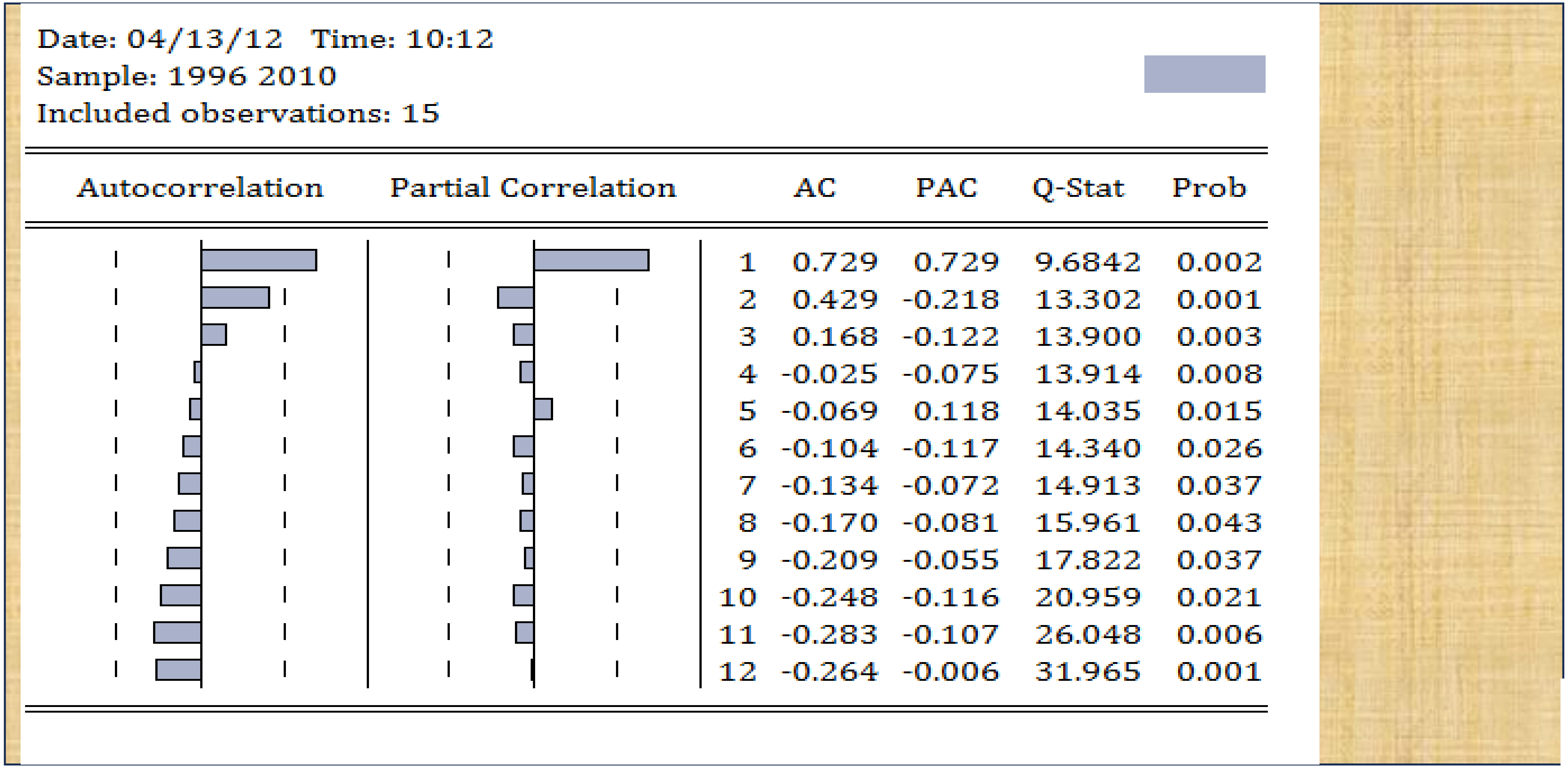
Look at the following … tell us what you would like to do with them?



Another tool to understand the behaviour of a time series …

Correlogram …

Correlogram



```
Date: 04/13/12   Time: 10:12
Sample: 1996 2010
Included observations: 15

   Autocorrelation     Partial Correlation        AC     PAC    Q-Stat   Prob

                                             1   0.729   0.729  9.6842   0.002
                                             2   0.429  -0.218  13.302   0.001
                                             3   0.168  -0.122  13.900   0.003
                                             4  -0.025  -0.075  13.914   0.008
                                             5  -0.069   0.118  14.035   0.015
                                             6  -0.104  -0.117  14.340   0.026
                                             7  -0.134  -0.072  14.913   0.037
                                             8  -0.170  -0.081  15.961   0.043
                                             9  -0.209  -0.055  17.822   0.037
                                            10  -0.248  -0.116  20.959   0.021
                                            11  -0.283  -0.107  26.048   0.006
                                            12  -0.264  -0.006  31.965   0.001
```

ACF and PACF

- Autocorrelation function (ACF)

$$\rho(p_{or}q)=\frac{\mathrm{cov}(y_t,y_{t-p_{or}q})}{\mathrm{var}(y_t)}$$

- Partial ACF (PACF)

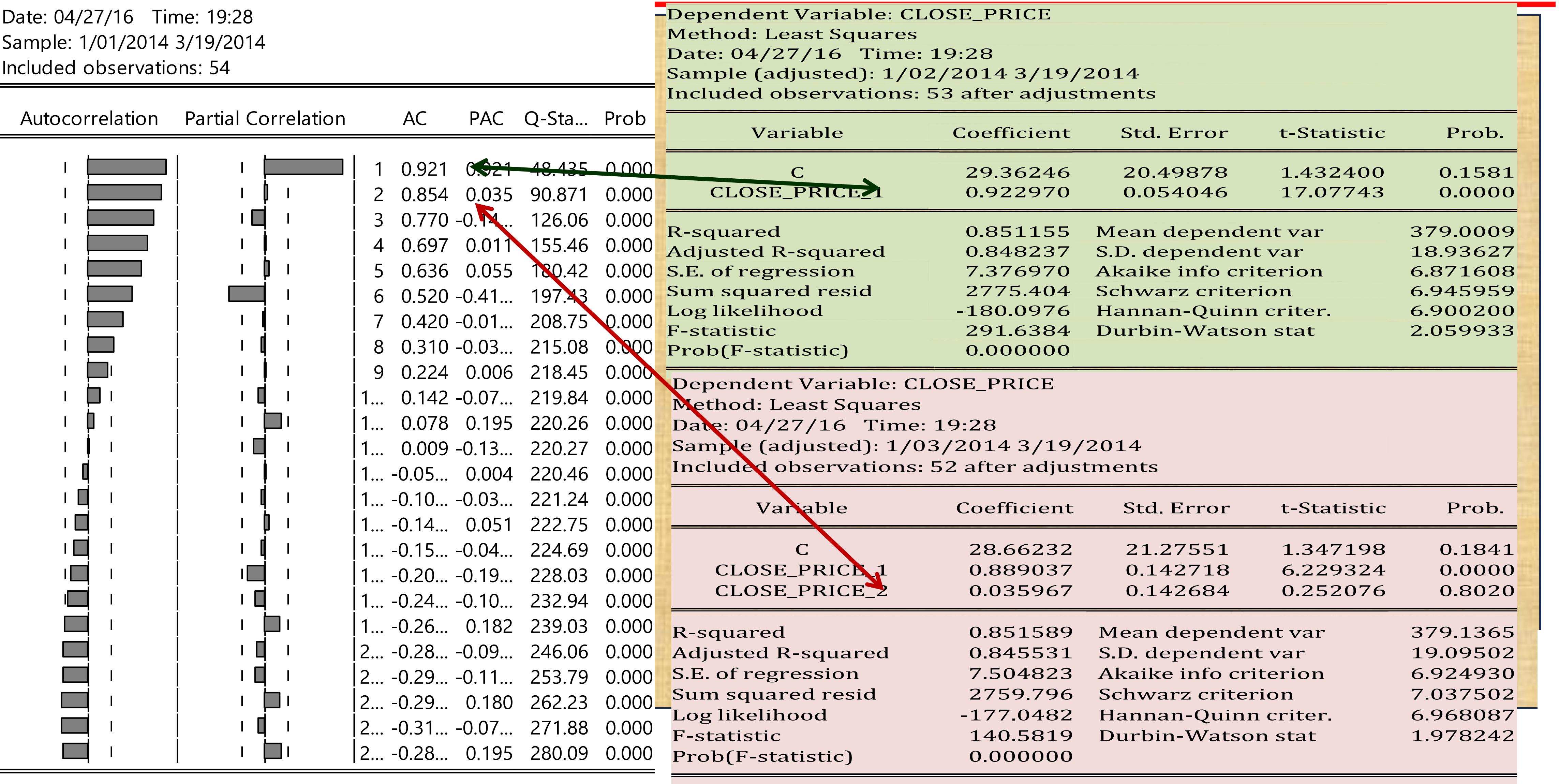
$$\phi(p)=\frac{\rho_p-\sum_{j=1}^{p-1}(\phi_{p-2,j}-\phi_{pp}\phi_{p-2,p-j})\rho_{p-j}}{1-\sum_{j=1}^{p-1}(\phi_{p-2,j}-\phi_{pp}\phi_{p-2,p-j})\rho_j},\quad p\geq 3$$

Autocorrelation and Partial Autocorrelation…

Autocorrelation is simple correlation between X_t and, say, X_{t+h}, it is a correlation between a series but with a lag.

While Partial Autocorrelation is a correlation between observations X_t and X_{t+h} after removing the linear relationship of all observations that fall between X_t and X_{t+h}.

Look…Think…Infer



Date: 04/27/16 Time: 19:28
Sample: 1/01/2014 3/19/2014
Included observations: 54

Autocorrelation	Partial Correlation		AC	PAC	Q-Sta...	Prob
		1	0.921	0.921	48.435	0.000
		2	0.854	0.035	90.871	0.000
		3	0.770	-0.14...	126.06	0.000
		4	0.697	0.011	155.46	0.000
		5	0.636	0.055	180.42	0.000
		6	0.520	-0.41...	197.43	0.000
		7	0.420	-0.01...	208.75	0.000
		8	0.310	-0.03...	215.08	0.000
		9	0.224	0.006	218.45	0.000
		1...	0.142	-0.07...	219.84	0.000
		1...	0.078	0.195	220.26	0.000
		1...	0.009	-0.13...	220.27	0.000
		1...	-0.05...	0.004	220.46	0.000
		1...	-0.10...	-0.03...	221.24	0.000
		1...	-0.14...	0.051	222.75	0.000
		1...	-0.15...	-0.04...	224.69	0.000
		1...	-0.20...	-0.19...	228.03	0.000
		1...	-0.24...	-0.10...	232.94	0.000
		1...	-0.26...	0.182	239.03	0.000
		2...	-0.28...	-0.09...	246.06	0.000
		2...	-0.29...	-0.11...	253.79	0.000
		2...	-0.29...	0.180	262.23	0.000
		2...	-0.31...	-0.07...	271.88	0.000
		2...	-0.28...	0.195	280.09	0.000

```
Dependent Variable: CLOSE_PRICE
Method: Least Squares
Date: 04/27/16   Time: 19:28
Sample (adjusted): 1/02/2014 3/19/2014
Included observations: 53 after adjustments

        Variable          Coefficient    Std. Error    t-Statistic      Prob.

           C               29.36246      20.49878      1.432400        0.1581
     CLOSE_PRICE_1         0.922970      0.054046      17.07743        0.0000

R-squared                  0.851155   Mean dependent var          379.0009
Adjusted R-squared         0.848237   S.D. dependent var          18.93627
S.E. of regression         7.376970   Akaike info criterion       6.871608
Sum squared resid          2775.404   Schwarz criterion           6.945959
Log likelihood            -180.0976   Hannan-Quinn criter.        6.900200
F-statistic                291.6384   Durbin-Watson stat          2.059933
Prob(F-statistic)          0.000000
```

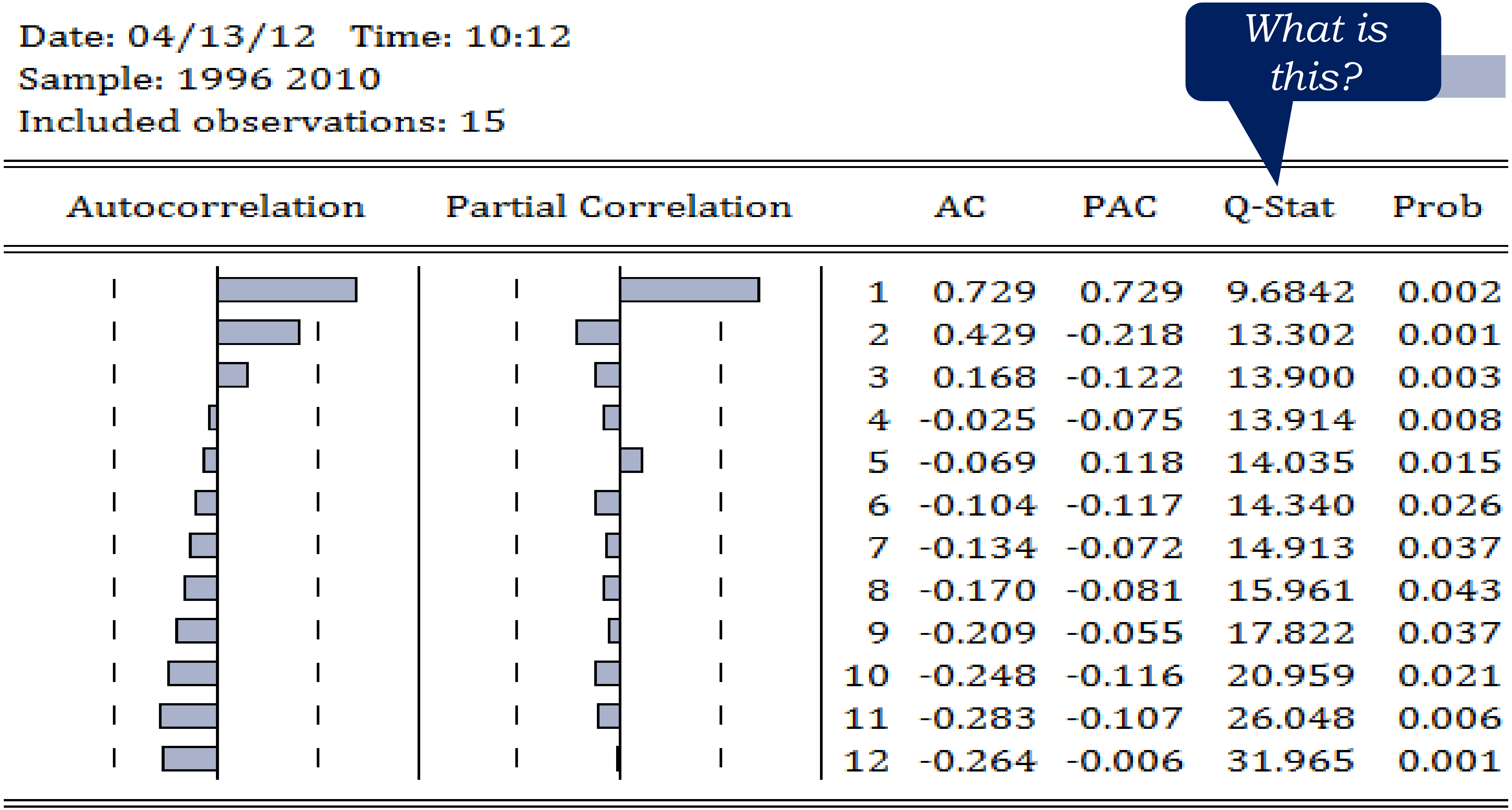
```
Dependent Variable: CLOSE_PRICE
Method: Least Squares
Date: 04/27/16   Time: 19:28
Sample (adjusted): 1/03/2014 3/19/2014
Included observations: 52 after adjustments

        Variable          Coefficient    Std. Error    t-Statistic      Prob.

           C               28.66232      21.27551      1.347198        0.1841
     CLOSE_PRICE_1         0.889037      0.142718      6.229324        0.0000
     CLOSE_PRICE_2         0.035967      0.142684      0.252076        0.8020

R-squared                  0.851589   Mean dependent var          379.1365
Adjusted R-squared         0.845531   S.D. dependent var          19.09502
S.E. of regression         7.504823   Akaike info criterion       6.924930
Sum squared resid          2759.796   Schwarz criterion           7.037502
Log likelihood            -177.0482   Hannan-Quinn criter.        6.968087
F-statistic                140.5819   Durbin-Watson stat          1.978242
Prob(F-statistic)          0.000000
```

Correlogram…



Date: 04/13/12 Time: 10:12
Sample: 1996 2010
Included observations: 15

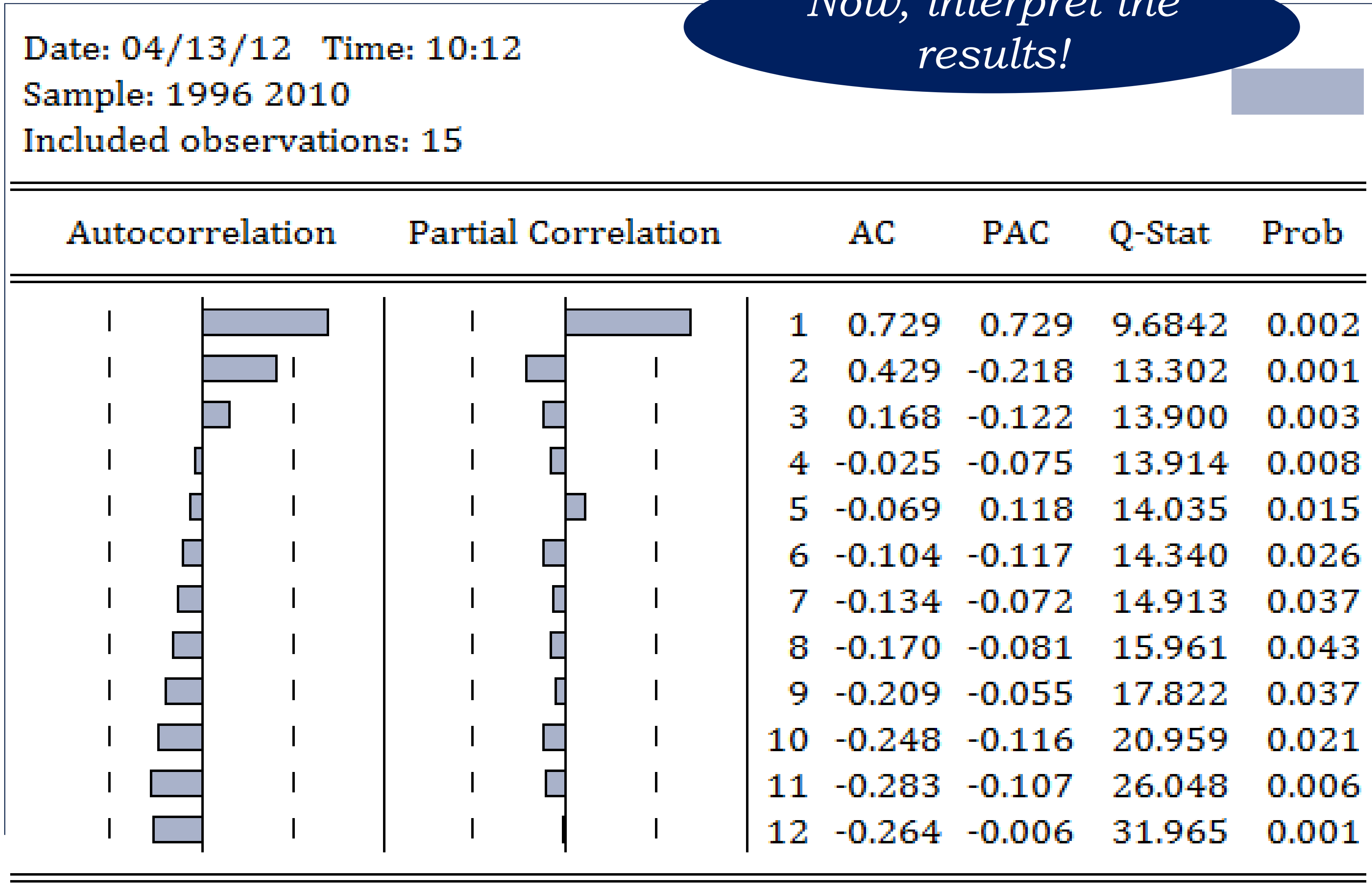
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.729	0.729	9.6842	0.002
		2	0.429	-0.218	13.302	0.001
		3	0.168	-0.122	13.900	0.003
		4	-0.025	-0.075	13.914	0.008
		5	-0.069	0.118	14.035	0.015
		6	-0.104	-0.117	14.340	0.026
		7	-0.134	-0.072	14.913	0.037
		8	-0.170	-0.081	15.961	0.043
		9	-0.209	-0.055	17.822	0.037
		10	-0.248	-0.116	20.959	0.021
		11	-0.283	-0.107	26.048	0.006
		12	-0.264	-0.006	31.965	0.001

What is this?

A discussion on Ljung-Box Q statistic

- Ljung-Box Q statistic: The Ljung-Box Q-test is a quantitative way to test for autocorrelation at multiple lags jointly.
- The null hypothesis for this test is that the first *m* autocorrelations are jointly zero, $H_0: \rho_1 = \rho_2 = \ldots = \rho_m = 0$.
- Alternatively, at least one of ρ_i is not zero.
- Note that it is a test related to **AUTOCORRELATION**.
- Ljung-Box Q statistic is having asymptotic χ^2 distribution

Correlogram…



Now, interpret the results!

Date: 04/13/12 Time: 10:12
Sample: 1996 2010
Included observations: 15

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.729	0.729	9.6842	0.002
		2	0.429	-0.218	13.302	0.001
		3	0.168	-0.122	13.900	0.003
		4	-0.025	-0.075	13.914	0.008
		5	-0.069	0.118	14.035	0.015
		6	-0.104	-0.117	14.340	0.026
		7	-0.134	-0.072	14.913	0.037
		8	-0.170	-0.081	15.961	0.043
		9	-0.209	-0.055	17.822	0.037
		10	-0.248	-0.116	20.959	0.021
		11	-0.283	-0.107	26.048	0.006
		12	-0.264	-0.006	31.965	0.001

Let’s work with these ways of forecasting…

We have data about…

- Sales Data of ONGC
- Price Data of Maruti
- Price Data of Maruti and Sensex