DATA ANALYTICS

Unit 3: Forecasting with Exponential, Croston’s and Regression

Jyothi R.
Department of Computer Science and Engineering

TRIPLE EXPONENTIAL SMOOTHING (HOLT-WINTER MODEL)

- Moving averaging and single and double exponential smoothing techniques can handle data as long as the data do not have any seasonal component associated with it.

- When there is seasonality in the time-series data, techniques such as moving average, exponential smoothing, and double exponential smoothing are no longer appropriate.

- In most cases, the fitted error values (actual demand minus forecast) associated with simple exponential smoothing and Holt's method will indicate systematic error patterns that reflect the existence of seasonality.

- For example, presence of seasonality may result in all positive errors, except for negative values that occur at fixed intervals.

- Such pattern in error would imply existence of seasonality.

- Such time series data require the use of a seasonal method to eliminate the systematic patterns in error.

TRIPLE EXPONENTIAL SMOOTHING (HOLT-WINTER MODEL)

Multiplicative Model

Level (or Intercept) equation:

$$L_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)\left[L_{t-1} + T_{t-1}\right]$$

Trend equation:

$$T_t = \beta \times (L_t - L_{t-1}) + (1 - \beta)\, T_{t-1}$$

Seasonal equation:

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma) S_{t-c}$$

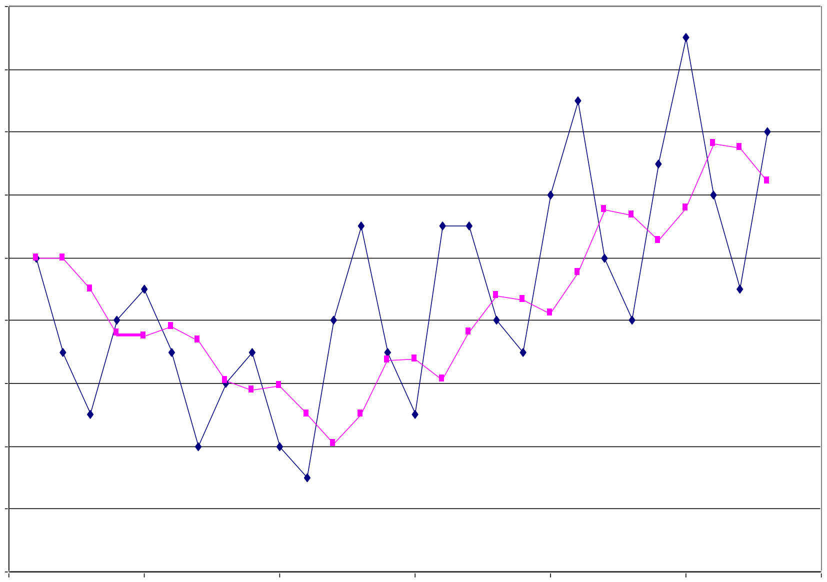
The forecast F_{t+1} using triple exponential smoothing is given by

$$F_{t+1} = \left[L_t + T_t\right] \times S_{t+1-c}$$

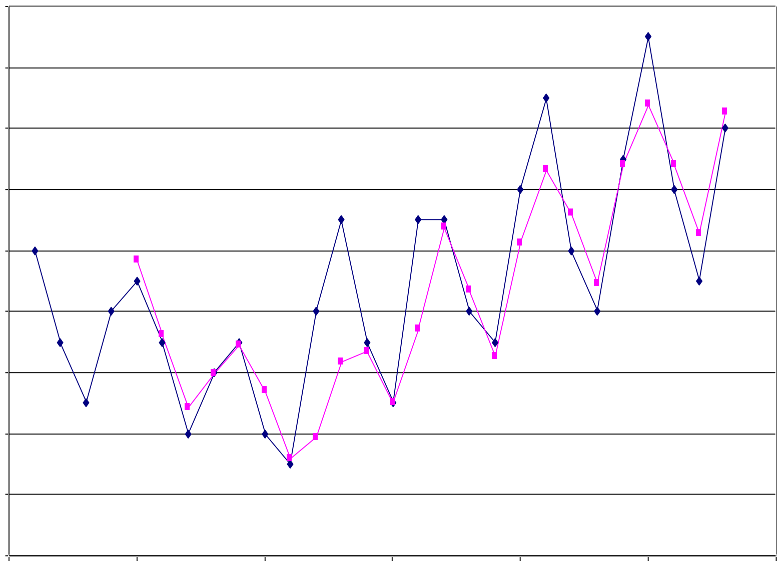
DATA ANALYTICS

Exponential Models

Quarterly Saw Sales Forecast: Holt's Method



Quarterly Saw Sales Forecast: Winter's Method (Multiplicative seasonality)



- RMSE for this application is: $\alpha = .3$ and $\beta = .1$

  RMSE = 155.5

- The plot also showed the possibility of seasonal variation that needs to be investigated.

- RMSE for this application is:

  $\alpha = 0.4$, $\beta = 0.1$, $\gamma = 0.3$ and

  RMSE = 83.36

- Note the decrease in RMSE.

Holt-Winter's Model

Additive Model

- The seasonal component in Holt-Winters' method.
- The basic equations for Holt's Winters' additive method are:

$$L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-s}$$

$$F_{t+m} = L + mT_{t-1} + S_{t+m-s}$$

- The initial values for L_s and T_s are identical to those for the multiplicative method.
- To initialize the seasonal indices we use

$$S_1 = y_1 - L_s, \quad S_2 = y_2 - L_s, \cdots, S_s = Y_s - L_s$$

Computing the Seasonality Index

- Several techniques exist to calculate the initial seasonality index (Winters, 1961; Makridakis *et al.*, 1998;Taylor 2011).

- The initial seasonality index can be calculated using a technique called method of simple.

- Several variations to the procedure discussed in next section exist, such as ratio-to-moving average.

Predicting Seasonality Index Using Method of Averages

- The following steps are used for predicting the seasonality index using method of averages:
- **STEP 1**
- Calculate the average of value of *Y* for each season that is, if the data is monthly data, then we need to calculate the average for each month based on the training data.
- Let these averages be $\bar{Y}_{1}, \bar{Y}_{2}, \bar{Y}_{3} \ldots\ldots \bar{Y}_{c}$

- **STEP 2**
- Calculate the average of the seasons' averages calculated in step 1 (say $\bar{Y}$).

- **STEP 3**
- The seasonality index for season *k* is given by the ratio $\bar{Y}_k / \bar{Y}$.
- to the procedure explained above is first divide the value of Y_t with its yearly average and calculate the seasonal average
- We will use first 3 years of data to calculate the seasonality index for various months.

Predicting Seasonality Index Using Method of Averages

- The seasonality index based on first 3 years of data using method of averages.

- Seasonality index can be interpreted as percentage change from the trend line.
- For example, the seasonality index for January is approximately 1.088 or 108.8% (textbook example).

- This implies that in January, the demand will be approximately 8.8% more from the trend line. The seasonality index for March is 0.8885 or 88.85% (textbook example).

Predicting Seasonality Index Using Method of Averages

- TABLE 1: Seasonality index using method of averages

Month	Sale Quantity (2012)	Sale Quantity (2013)	Sale Quantity (2014)	Monthly Average $\bar{Y}_k$	Seasonality Index $\bar{Y}_k / \bar{\bar{Y}}$
January	3002666	4447581	4634047	4028098.00	1.087932
February	4401553	3675305	3772879	3949912.33	1.066815
March	3205279	3477156	3187110	3289848.33	0.888541
April	4245349	3720794	3093683	3686608.67	0.9957
May	3001940	3834086	4557363	3797796.33	1.02573
June	4377766	3888913	3816956	4027878.33	1.087872
July	2798343	3871342	4410887	3693524.00	0.997568
August	4303668	3679862	3694713	3892747.67	1.051375
September	2958185	3358242	3822669	3379698.67	0.912808
October	3623386	3361486	3689286	3558053.33	0.960979
November	3279115	3670362	3728654	3559377.00	0.961337
December	2843766	3123966	4732677	3566803.00	0.963342
Average of monthly averages				3702528.22	

Predicting Seasonality Index Using Method of Averages

- This implies that the demand in March will be 11.15% less from the trend line.
- Note that, multiplicative model is used in this example.
- To start the triple exponential smoothing, we need to set the starting values of level and trend.

  $L_{36} = Y_{36}/S_{36} = 4732677/0.9633 = 4912983.494$
- The initial value of trend (T_{36}) can be calculated based on second and third year by using

$$T_{36} = \frac{1}{12}\left[\frac{Y_{36}-Y_{24}}{12} + \frac{Y_{35}-Y_{23}}{12} + \frac{Y_{34}-Y_{22}}{12} + \cdots + \frac{Y_{25}-Y_{13}}{12}\right]$$

$$T_{36} = \frac{1}{12}\left[\frac{4732677-3123966}{12} + \frac{3728654-3670362}{12} + \ldots + \frac{4634047-4447581}{12}\right] = 21054.35$$

Predicting Seasonality Index Using Method of Averages

- The forecast for period 37 using triple exponential smoothing is given by

$$F_{37} = [L_{36} + T_{36}] \times S_{37-12} = [L_{36} + T_{36}] \times S_{25}$$

- The seasonal index S_{25} (seasonality index for January) is 1.088.
- Substituting the values of L_{36}, T_{36} and S_{25}, we get

$$F_{37} = [4912983.494 + 21054.35] \times 1.088 = 5368233.2$$

Predicting Seasonality Index Using Method of Averages

- TABLE 3: Forecasting using triple exponential smoothing (values differ for different round off values of parameters)

Month t	Actual Demand	L_{t-1}	T_{t-1}	S_t	F_t	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
37	3216482	4912983.49	21054.35	1.09	5367895.97	4.62858E+12	0.668872
38	3453239	4301229.28	−295349.93	1.07	4273531.48	6.7288E+11	0.237543
39	5431651	3759825.78	−418376.71	0.89	2969014.38	6.06458E+12	0.453386
40	4241851	4228345.39	25071.45	1.00	4235127.90	45200134.5	0.001585
41	3909887	4255577.53	26151.79	1.03	4391900.21	2.32337E+11	0.123281
42	3216437	4131354.31	−49035.71	1.09	4441041.44	1.49966E+12	0.380733
43	4222004	3722098.55	−229145.74	1.00	3484457.63	5.43975E+11	0.174691
44	3621034	3729543.06	−110850.61	1.05	3804603.81	33697874118	0.050695
45	5162201	3562820.55	−138786.56	0.91	3125486.18	4.14821E+12	0.394544
46	4627176	4138038.05	218215.47	0.96	4186269.09	1.94399E+11	0.095286
47	4623945	4503072.73	291625.07	0.96	4609319.06	213918263.4	0.003163
48	4599368	4799566.34	294059.34	0.96	4906905.01	94579010434	0.066865

Predicting Seasonality Index Using Method of Averages

- The forecast for the period 37 to 48 for the data in Table 2 is given in Table 3 .
- Note that the values such as seasonality index are rounded to two decimals, the forecast values will be different if the actual seasonality index values are used.
- The RMSE and MAPE using triple exponential smoothing are 1228588.29 and 0.2208 (22.08%), respectively.
- The values of $a = 0.32$, $b = 0.5$, and $\gamma = 1$ are used for calculating the level, trend, and seasonal components.

- It is important to note that the exponential smoothing techniques are very sensitive to initial values of level, trend, and seasonal index

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- Products such as spare parts may have intermittent demands.
- Exponential smoothing models discussed so far in the chapter will produce biased estimate when used for intermittent demand.
- Croston (1972) developed a model that uses two separate exponential smoothing equations for predicting mean time between demands and the magnitude of demand whenever the demand occurs.

- That is, Croston's method has two components:

  (a) Predicting time between demand and

  (b) magnitude of the demand.

- The primary objective of Croston's method is to forecast mean demand per period.

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- Let
- Y_t = Demand at time t (Y_t may take value 0)
- F_t = Forecasted demand
- TD_t = Time between the latest and the previous non-zero demand in period t
- FTD_t = Forecasted time between demand at period t

The following steps are used for forecasting demand:

- If $Y_t = 0$ then $F_{t+1} = F_t$ and $FTD_{t+1} = FTD_t$ **Eqn 8**

Eqn 9 If $Y_t \neq 0$ the $F_{t+1} = \alpha \times Y_t + (1-\alpha)F_t$ and $FTD_{t+1} = \beta \times TD_t + (1-\beta) \times FTD_t$

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- *α and β* are smoothing constants for forecasted demand and forecasted time between demands, respectively
- Once the forecasted demand and time between demands are known, then the mean demand per period D_{t+1}, is given by

Eqn 10:

$$D_{t+1} = \frac{F_{t+1}}{FDT_{t+1}}$$

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- Quarterly demand for spare parts of avionics system of an aircraft
- Use the demand during the quarters 1 to 4 as training data to forecast demand for periods 5 to 16 using Croston's method.

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- Quarterly demand for spare parts of avionics system of an aircraft is shown in Table 4:
- Use the demand during the quarters 1 to 4 as training data to forecast demand for periods 5 to 16 using Croston's method.

Quarter	1	2	3	4	5	6	7	8
Demand	20	12	0	18	16	0	20	22
Quarter	9	10	11	12	13	14	15	16
Demand	0	28	0	0	30	26	0	34

TABLE 13.8 Quarterly demand for avionic system spares

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- Procedure used for starting values of F_t and FTD_t is shown in the table here:
- TD_4 = 2 since the elapsed time from the previous demand and current demand period is 2 (4 – 2).
- The forecasted time between demand is the average TD_t values up to t = 4.
- So, FTD_4 = (1+2)/2 = 1.5.
- The forecasted demand F_4 for t = 4 is (20 + 12 + 18)/3 = 16.67.
- Note that the total value is divided by 3 (not 4) since only 3 quarters had non-zero demand.
- So, the starting values for Croston's method are.

Quarter	Demand	TD_t	FTD_t	F_t
1	20			
2	12	1		
3	0			
4	18	2	1.5	16.67

$TD_4 = 2$, $FTD_4 = 1.5$, and $F_4 = 16.67$

Let $\alpha = \beta = 0.2$. Then

$$F_5 = 0.2 \times 18 + (1 - 0.2) * 16.67 = 16.936$$

$$FTD_5 = 0.2 \times 2 + (1 - 0.2) \times 1.5 = 1.6$$

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- Forecasted demand for periods 5 to 16 using Croston's method.

Quarter	Demand	TD_t	FTD_t	F_t	$D_t = (F_t/FTD_t)$
1	20				
2	12	1			
3	0				
4	18	2	1.5000	16.67	11.11333
5	16	1	1.6000	16.936	10.585
6	0		1.4800	16.7488	11.31676
7	20	2	1.4800	16.7488	11.31676
8	22	1	1.5840	17.39904	10.98424
9	0		1.4672	18.31923	12.48585
10	28	2	1.4672	18.31923	12.48585
11	0		1.5738	20.25539	12.8707
12	0		1.5738	20.25539	12.8707
13	30	3	1.5738	20.25539	12.8707
14	26	1	1.8590	22.20431	11.94417
15	0		1.6872	22.96345	13.61034
16	34	2	1.6872	22.96345	13.61034

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND

- **Example: lubricant sales**
- Several years ago, an oil company requested forecasts of monthly lubricant sales
- One of the time series is shown in the table below.
- The data contain small counts, with many months registering no sales at all, and only small numbers of items sold in other months.

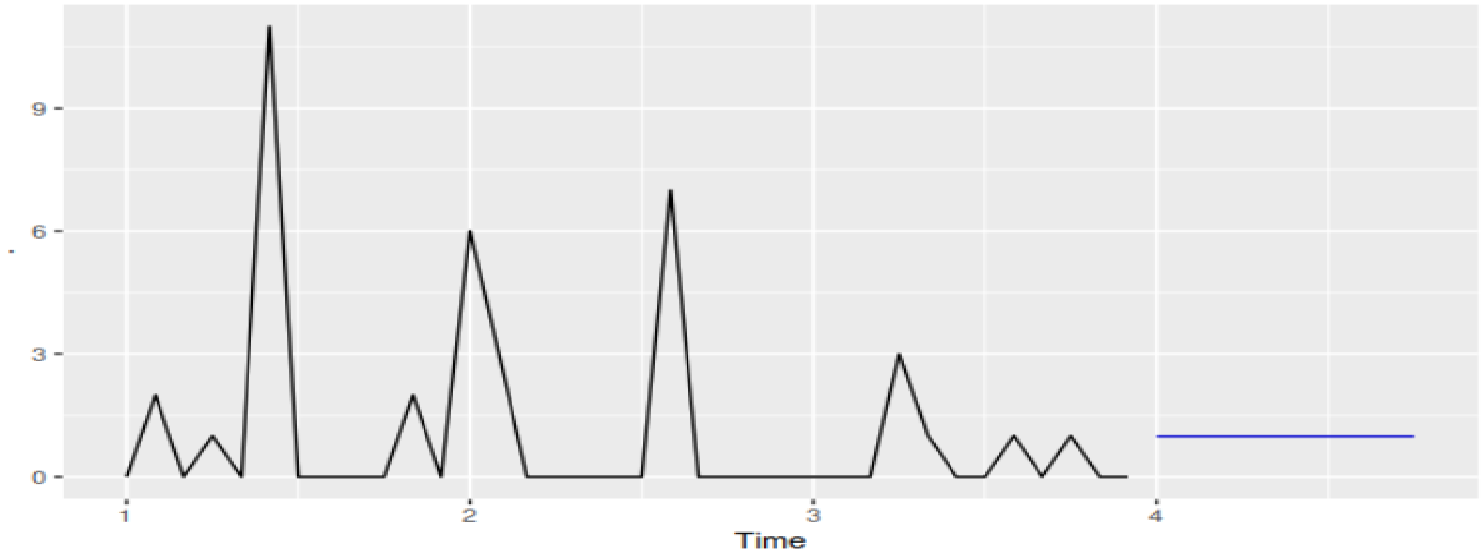
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0	2	0	1	0	11	0	0	0	0	2	0
2	6	3	0	0	0	0	0	7	0	0	0	0
3	0	0	0	3	1	0	0	1	0	1	0	0

- There are 11 non-zero demand values in the series, denoted by q.
- The corresponding arrival series a is also shown in the following table.

i	1	2	3	4	5	6	7	8	9	10	11
q	2	1	11	2	6	3	7	3	1	1	1
a	2	2	2	5	2	1	6	8	1	3	2

- Applying Croston's method gives the demand forecast 2.750 and the arrival forecast 2.793.
- So the forecast of the original series is $\hat{y}_{T+h|T} = 2.750/2.793 = 0.985$

CROSTON'S FORECASTING METHOD FOR INTERMITTENT DEMAND



- An implementation of Croston's method with more facilities (including parameter estimation) is available in the **tsinterminant package** for R.
- Forecasting models that deal more directly with the count nature of the data are described in Christou & Fokianos (2015).

Forecasting Stories 1: The Power of a Seasonality Index

- Read the second entry in a series on time series analysis and seasonality, and see how, through 2 simple use cases, the power of a seasonality index is uncovered.
- The strange fact was we had **performed poorly on all weeks**. Following are the weekly attainment figures: Week 1: 75% Week 2: 77% Week 3: 79% Week 4: 81%. What was amiss? What is the real-life forecasting story?

- **CASE 1 : Forecast over-indexed in April**

April Forecast and Actuals Comparison

	FY19 Forecast	FY19 Actuals	FY18 Actuals	FY17 Actuals
April Average(Mn)	16.5	14.3	15.5	15.6
Year Average(Mn)	18.6	18.4	19.1	19.5
April Seasonality	89%	78%	81%	80%

Forecasting Stories 1: The Power of a Seasonality Index

- Here is what happened: As we can see from the image,
- **April forecast** seasonality was **over indexed by 11%**, i.e. at 89% of yearly average while actuals were trending towards 78%.
- What does a seasonality index mean?

Seasonality Index Calculation

	FY19 Forecast	FY19 Actuals	FY18 Actuals	FY17 Actuals
Week 1(Mn)	18.2	17.1	18.5	17.9
Week 2(Mn)	18.4	17.0	17.6	18.7
Week 3(Mn)	19.6	16.5	17.5	18.6
Year Average(Mn)	18.6	18.4	19.1	19.5

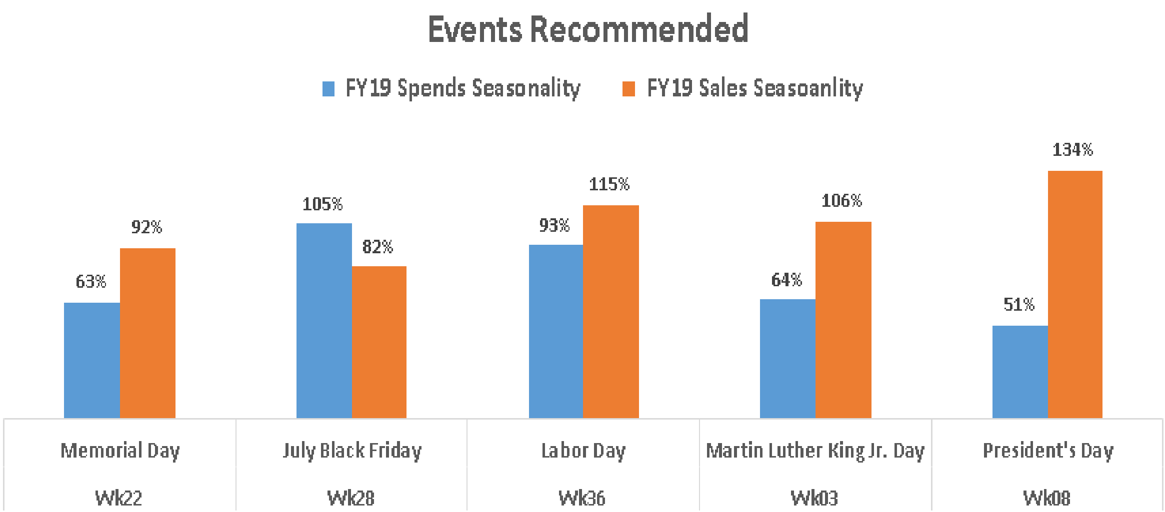
Week 1	98%	93%	97%	92%
Week 2	99%	=F73/F$75	92%	96%
Week 3	105%	90%	91%	95%

Forecasting Stories 1: The Power of a Seasonality Index

- It simply refers to all variables **normalized to a range close to 1**, so that all variables are comparable. As explained in the image, we divide each number by their yearly average to calculate the index. This way, the average of all values in the entire variable column is always 1.
- Hence interpreting the April seasonality, April being holiday is low performing month for this product.
- The forecast does partially take this into account, with 89% target compared to average of the year.
- However, actual performance across years can be seen ~80%.
- Hence the targets or f**orecast need to be even lower to be realistic**.

Forecasting Stories 2: The Power of a Seasonality Index

- Have a look at the spends and sales seasonality indices:



- What conclusions can you draw?
  - The first observation was that we are spending a much higher proportion of marketing budget on July Black Friday, whereas President's day week results in higher sales.
  - If we **reallocate the spends** from July Black Friday day to President's week, we would end up with a higher ROI without adding a single penny to the budget.

• Forecasting Stories 2: The Power of a Seasonality Index

Forecast over-indexed in April

- For President's day, the sales seasonality index is 134% while your marketing spends index is 51%!
- And hence we should **reallocate the spends** from July Black Friday day to President's week.

Comparing 2 seasonality indexes can give some power-packed insights

- Numbers after all, are good or bad only when they are compared against another.
- The rest of the story is based on events, specially holidays in the US Calendar.
- Not all of the following are holidays, and different events impact sales in different ways.
- Also, some events are more important than others.

Regression for Forecasting

- Parker and Segura (1971) claimed regression can predict more accurately than exponential smoothing

- Regression is particularly useful when there is one or more explanatory variable in addition to the dependent variable Y_t

The forecast value at time *t* can be written as

$$F_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_n X_{nt} + \varepsilon_t$$

Here F_t is the forecasted value of Y_t, and X_{1t}, X_{2t}, etc. are the predictor variables measured at time *t*.

References

Text Book:

“Business Analytics, The Science of Data-Driven Making”, U. Dinesh Kumar, Wiley 2017 Chapter-13.7-13.9

Image and Case Study

https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics

https://otexts.com/fpp2/stationarity.html

THANK YOU

Jyothi R

Assistant Professor, Department of Computer Science

jyothir@pes.edu