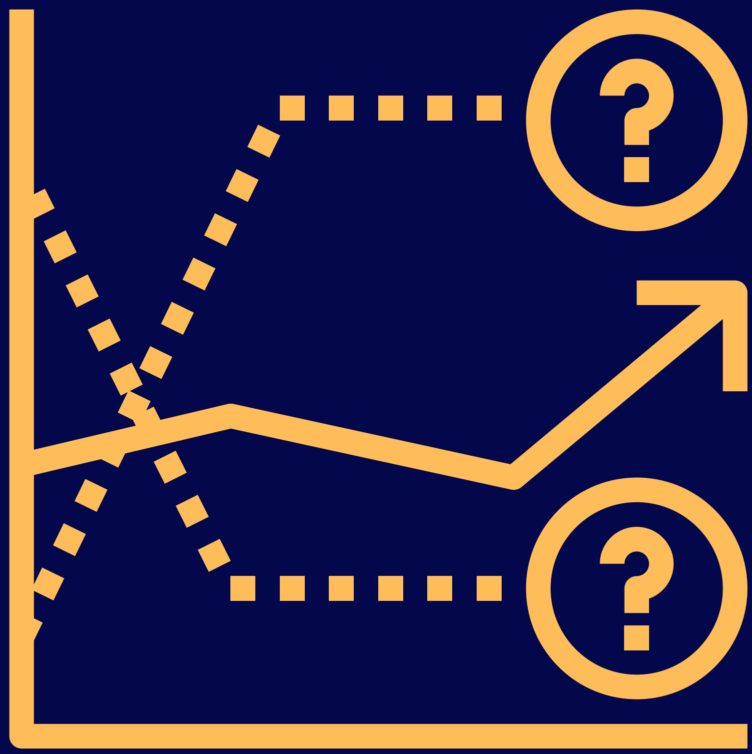


WHITEPAPER

FORECASTING MODELLING TECHNIQUES

VALIANCE

FORECASTING MODELLING TECHNIQUES

A Comparitive Study

In most cases forecasting a time-series data like sales, call volume, inventory, etc. employs the use of ARIMA modeling. Here, we discuss and compare both ARIMA and Liner Regression, an alternate approach to forecasting, in predicting sales and analyzing the impact of promotions on sales activity.

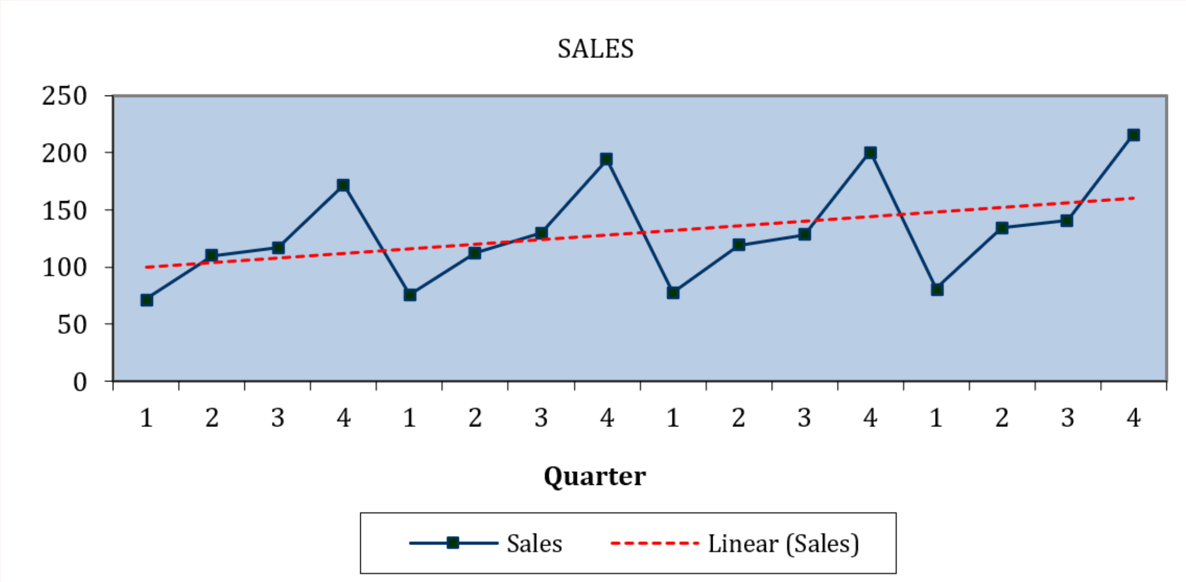
Auto-Regressive Integrated Moving Average - ARIMA

The ARIMA procedure, popularly known as Box-Jenkins methodology, uses lags of the differenced series called "auto-regressive" terms, lags of the forecast errors called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series.

A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, where:
p is the number of autoregressive terms,
d is the number of non-seasonal differences, and
q is the number of lagged forecast errors in the prediction equation.

To identify the appropriate ARIMA model for a time series, begin by identifying the order(s) of differencing needing to stationarize the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating

ARIMA technique is capable of capturing Trend, Cyclical and Seasonal patterns in the data but fails to capture the effect of other independent factors at work which are non-seasonal and non-cyclical in nature, like the impact of promotions offered by the company, change in sales-force, change in number of products being offered, change in the number of outlets, etc. Let's take an example – During winters, particularly in December, the sales of a retailer increases substantially every year due to festive season (seasonal factor). The trend for sales exhibits seasonal pattern which can be identified by graphically plotting the sales, illustrated in the graph below:



In such a scenario, ARIMA method of forecasting gives efficient results compared to any other technique. However, to measure the impact of other independent factors at work, which are non-seasonal and non-cyclical in nature, one might like to experiment with other forecasting techniques one of which is Linear Regression Analysis

Linear Regression

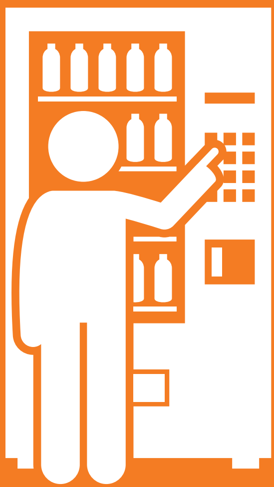
Linear Regression is also one of the techniques to forecast time series. It basically forms a linear equation by manually calculating Lag variables for the data (up to 4 lag values are recommended) and regressing the Lag variables on the dependent variable. It can also incorporate promotions data as well as other independent variables i.e. No. of outlets, Manpower, Holidays effect, etc. which may improve the forecasting accuracy of the series.

Sales(t) = A + αLag1 + βLag2 + γLag3 + δLag4 + Promo (0,1) + Other Effect Variables

We recently conducted a comparative study for one of the leading manufacturers of vending machines in the U.S. to predict sales and analyze the impact of promotions on sales at a retailer level. The next section provides food for thought, summarizing the techniques and results of this study which were derived from the two approaches discussed above.

SAMPLE CASE STUDY: ABC Ltd.

ABOUT COMPANY



A company in the U.S. manufactures vending machines which are being supplied exclusively to six retailers. The company had extended promotional offers on its products and wanted to forecast the sales incorporating the promotions which were being offered. The sales and promotions data was provided to us grouped by each retailer. We used both ARIMA and Liner Regression techniques to forecast sales and gauge the impact of promotions on the sales

BUSINESS OBJECTIVE

Does the forecasting accuracy improve significantly by incorporating promotional data in the forecasting model for all the retailers of ABC Ltd.

DATABASE USED

Sales and Promotions data for each retailer during the time period – May 2010 to March 2014

MODELLING TECHNIQUE USED

ARIMA and Linear Regression

MODELLING PROCEDURE

(Note: Modeling procedure was uniform for all retailers)

Technique 1: ARIMA (Auto-Regressive Integrated Moving Average)

- ARIMA procedure was used to optimize p, d, q values of the model.
- Further optimized p, d, q values were used to get the best forecasting model.
- Error margin was calculated between the forecasted values and the actual values.
- Model was validated using sales data for time period (April 2014 – June 2014).

Technique 2: Linear Regression

- Product identifier and Sellout unit were selected and grouped by each retailer.
- Count of missing values for Sellout unit was calculated for each product. Products with sellout unit count greater than equal to 30 were considered for forecasting.
- A final dataset was formed for the shortlisted products by including four new lagged variables (lag1, lag2, lag3 and lag4) along with promotion data. The lagged values of sellout unit data was created by using LAG function and a binary variable was created for promotional data with respect to every sellout unit data. (Note: Months having promotional activity were assigned a value '1' and months with no promotional activity were assigned a value '0').
- Linear Regression technique was used on the final dataset for each product by retailer.

RESULTS

The impact of promotions on each product by retailer is presented in the table below:

Retailer	***Total no. of Products***	***Products with Sellout>=30***	***Products with Significant p-value for Promotions***
U	50	4	0
V	83	0	0
W	79	13	1
X	91	14	1
Y	74	39	2
Z	5	3	0

In this particular case study, Linear Regression analysis proved to be a better technique than ARIMA for certain retailers and products, where we wanted to gauge the impact of an independent factor - promotions, which is non-cyclical and non-seasonal in nature, impacting the dependent variable – sales of vending machines.

In the light of the above case study, we recommend a dual approach to select a champion challenger technique using ARIMA and Linear Regression Analysis.

About Valiance

What Do We Do?

We help companies convert data into actionable intelligence using strong data science, data engineering & cloud engineering skills matched with critical business thinking.We work with:

LOB (Line of Business) teams & Digital Leaders: To build digital transformation solutions for business excellence

Technology Leaders: To create data infrastructure & platforms required to support analytical workloads

AI & ML Teams:To build next gen products powered by data intelligence

Quick Stats about us

9	50+	100+	10+
YEARS OF EXISTENCE	EMPLOYEE BASE	SUCCESSFUL PROJECTS	INDUSTRIES SERVED

DATA SCIENCE

DATA ENGINEERING

DATA CLOUD