Advance Journal of Pharmaceutical Research & Review
Volume 1, Issue 4, October 2024, PP: 18-28, ISSN No: 3048-491X

Demand Forecasting and Order Optimization in the Pharmaceutical Industry to Reduce Carbon Footprint.

Manjunath Ande
High School Senior Year Student, Telangana, India
Corresponding mail: manjunathande7@gmail.com

Abstract
The pharmaceutical industry plays a crucial role in global health, particularly in developing life-saving medicines. One major challenge is demanding forecasting, essential for efficient supply chain management. This paper explores the complexities of demand forecasting in the pharmaceutical sector and the risks of inaccurate predictions, such as product shortages and waste. It aims to enhance customer fulfillment and optimize inventory through reliable forecasting. Using a dataset from a third-party pharmaceutical marketing firm, predictive models, including Multiple Linear Regression and Time Series, are developed. Simulating various demand scenarios, optimal order quantities are identified to maximize profits. The models are compared against actual demand, emphasizing the need for accurate forecasting to improve operational efficiency and reduce risks. Ultimately, this research highlights the critical importance of effective demand forecasting for the industry's growth and sustainability.

Keywords: *Pharmaceutical industry, Supply chain management, predictive models, customer fulfillment rate, inventory optimization.*

I. Introduction

The global pharmaceutical market was valued at approximately $1.5 trillion in 2023 and is projected to grow at a compound annual growth rate (CAGR) of approximately 5.8% from 2024 to 2030 [1]. This substantial increase illustrates the skyrocketing demand for drugs and treatments as new diseases emerge. Discovery of new treatments and therapies remains a paramount objective in the pharmaceutical industry. In 2023, global pharmaceutical Research and Development (R&D) spending exceeded $250 billion [2]. However, one of the biggest challenges faced by manufacturing and supply companies is estimating and predicting demand. Compromises in quality and availability of reliable data, fluctuations in the market, external such as economic shifts, public health emergencies and many other factors could lead to largely erroneous predictions and hinder the ability to make informed decisions. Unreliable forecasts can have irrevocable consequences leading to both excess and inadequate demand, resulting in twin dangers: shortages of health commodities across various supply chain channels and wastage of manufactured products.

In the case of drugs, this can have serious health ramifications, such as untreated patients. For example, in 2006, the France-based pharmaceutical firm Sanofi-Aventis had to destroy nearly 500 kilograms worth of starting material that would have made up more than 10 million tablets of the antimalarial drug artesunate because orders were substantially below original estimates [3]. Shortages may lead to increased prices (if not regulated), which can adversely affect the financial stability of suppliers and the healthcare system.

Reliable demand forecasting is an efficient way to deal with stochastic demand, that is, varying demand. A literature review of 72 selectively chosen research papers underlines the exponential increase in research concentrated in the domain of health supply chain forecasting [4]. Appropriate demand forecasting plays a vital role in optimizing operational efficiency, improving customer satisfaction, and facilitating growth. By comprehensively understanding and anticipating customer requirements through a systematic approach and sophisticated tools, organizations can make data-driven decisions and capitalize on market opportunities. It indicates that the choice of forecasting technique depends on the level of technology used. Therefore, it is important to choose the most suitable approach for forecasting. Higher technology usage tends to rely more on internal data for forecasting, whereas lower technology usage may depend more on quantitative techniques, such as customer surveys.

This research elucidates several predictive models of demand forecasting, emphasizing the need to understand and interpret the results of a model optimally. The data utilized and supplied to the models is extracted from a third-party pharmaceutical marketing firm.

Information depicting the sales and purchases of their customers is taken from a third-party firm for various products. The data is then compatibly provided to the predictive models. These models provide estimations or forecast the likely demand, assisting in inventory optimization and customer fulfillment rate. In most cases, it is imperative for such firms to maintain products in their inventory as they would be needed to supply to their customers instantaneously.

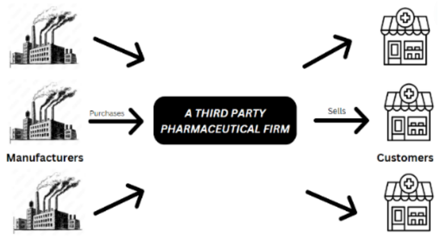

Fig. 1 Third-Party Pharmaceutical Firm

A third-party marketing firm places an order to the manufacturer with product specifications in accordance with their requirements and branding. Upon receiving the product, the firm fulfils the orders it receives from its regional clientele

II. Literature Review

Demand forecasting is an essential aspect of business management. It is an effective method to deal with varying product demand. Research on machine learning approach in demand forecasting accuracy indicates that ML techniques can significantly increase demand forecast accuracy, achieving improvements ranging from 10% to 41% [5]. A massive problem for the pharmaceutical industry's bottom line and public health is the issue of medicine expiration. Data show that every year, between 10 and 20 percent of medications are thrown out because they have expired, costing the between $30 billion and $50 billion [6]. These expiries aggravate the issue of drug shortages and jeopardize patient safety in addition to financial losses.

Demand forecasting methodologies could offer a feasible approach to address these concerns through the optimization of inventory levels and the enhancement of the synchronization between supply and demand. This study uses the sales and orders information of several products from a third-party manufacturing firm to construct predictive models using different methodologies. It aims to provide reliable model that is able to predict or forecast the upcoming demand with the least margin of error. The primary objectives of a forecasting model, in most cases, would be to increase the customer fulfillment rate and minimize unsold inventory. ‘Demand forecasting in pharmaceutical supply chains: A case study’ [7] underscored the recognition of demand forecasting as a strategic necessity for pharmaceutical companies, particularly in light of the need to enter emerging markets. Accurate forecasting is essential for effective inventory management and meeting customer demand. Overall, the study advocates for enhanced data integration and collaboration among supply chain partners to optimize demand forecasting and inventory management. It is also important to note that the use of complex models, such as symbolic regression in the study, may lead to overfitting, where the model performs well on historical data but fails to generalize to future demand scenarios. This limitation could undermine the practical applicability of the findings in real-world settings. ‘Effective Demand Forecasting in Health Supply Chains: Emerging Trend, Enablers, and Blockers’ [4] emphasizes the critical importance of demand forecasting in health supply chains, particularly in low- and middle-income countries (LMICs). A systematic literature review was conducted, analyzing 71 articles selected from an initial pool of 486, focusing on the intersection of forecasting, risks, and incentives in health supply chains. The review revealed that 44% of the articles employed qualitative methodologies, while 37% used mixed methods, and 20% utilized quantitative tools to analyze data. The findings indicate that inaccurate demand forecasting leads to an uneven distribution of risks, which hampers the ability to match supply with demand effectively. The study highlights that 15% of the total papers reviewed were relevant to demand forecasting in health supply chains, underscoring the need for more focused research in this area. The paper concludes with recommendations for stakeholders to enhance forecasting accuracy and reduce associated risks, ultimately improving health supply chain efficiency. The review was conducted on 71 articles, which may restrict the breadth of the findings. The authors indicate that exploring a larger number of articles could enhance the understanding of empirical research and the issues currently addressed in the field.

‘A dynamics approach for modeling inventory fluctuations of the pharmaceutical supply chain in covid 19 pandemic’ [8] identified 29 drivers of the bullwhip effect specific to the pharmaceutical supply chain, highlighting the critical role of information quality and lead time in managing inventory. It was found that the COVID-19 pandemic exacerbated the bullwhip effect, leading to significant demand fluctuations and inventory shortages. The research utilized systems dynamics modeling to simulate the interactions of various supply chain components, revealing that improved information sharing and safety stock levels could mitigate the bullwhip effect. Scenarios were developed to analyze the impact of these improvements on inventory gaps. The findings emphasize the necessity for better coordination and communication among supply chain players to enhance resilience. Overall, the study provides valuable insights into strategies for reducing demand variability and improving inventory management in the pharmaceutical sector.

The bullwhip effect refers to the phenomenon where small fluctuations in consumer demand at the retail level lead to increasingly larger fluctuations in demand at the wholesale, distributor, manufacturer, and supplier levels. This amplification of demand variability can result in inefficiencies, excess inventory, and increased costs throughout the supply chain. It is primarily caused by distorted information and delays in communication among supply chain participants.

'A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry' [9] highlights that the changes in technology, and shifts in healthcare provider preferences contribute to the volatility of demand forecasts. These elements complicate the forecasting process and increase the likelihood of inaccuracies. The paper emphasizes the necessity for more effective forecasting techniques to mitigate these inaccuracies. It suggests that traditional single-model approaches may not adequately address the complexities of demand forecasting in the pharmaceutical sector, thus advocating for hybrid models. The paper also discusses how improved information sharing among supply chain partners can enhance forecasting accuracy. By sharing credible demand forecasts and data, companies can better align their operations with actual market demand, thereby reducing inaccuracies. The paper also adds that industries facing complex and dynamic environments, similar to the pharmaceutical sector, can apply the insights from this research to improve their forecasting processes. Understanding how to incorporate external factors and uncertainties into forecasting models can lead to more resilient supply chains.

The key findings of the literature review emphasize the critical importance of accurate demand forecasting in the pharmaceutical supply chain to mitigate issues like inventory shortages and expired medications. It implies that the bullwhip effect, exacerbated by poor information sharing and communication delays, leads to significant demand variability. The review advocates for the adoption of hybrid forecasting models that incorporate advanced technologies and external factors to enhance accuracy. Improved collaboration among supply chain partners is essential for aligning operations with actual market demand. Overall, the study underscores the need for better data integration and strategic forecasting methodologies to optimize inventory management.
In the following study, we will use the data on purchase and sales orders to elucidate methods and construct models to forecast the demand of a medicinal product that is available in the market for customer use and we will also discuss on important associated aspects such as the Practicality, Applicability and Reliability of the models to deal with the stochastic demand in the pharmaceutical sector.

III. Data Description

The data supplied to the models in this study were obtained from the Sales and Purchases Analytical Report of a Third-Party Pharmaceutical Organization. The information is retrieved from the database that is primarily associated with billing and the generation of invoices. It maintains a record of every purchase made from the manufacturer and every unit sold to the customer. Similarly, it generates both the cumulative and individual details connected to an entity. These statistics were collected for a period of approximately three months. The data is organized from the beginning of April 2024 to approximately the end of July 2024. The compiled data highlighted situations such as the presence of an inadequate stock of medicine and the availability of surplus quality.
Several quantities in sales orders were characterized by a combination of actual and free quantities. The concept of how many units of free quantity would be attached with a particular number of order quantity was predetermined since the inception of the organization and is termed as "Scheme." These combinations varied from customer to customer.
For a comprehensive analysis and better interpretation of the data, individual sales reports were gathered to delineate how many units of the product were sold to a particular customer from the total quantity of the product in the inventory.
Overall, the data can be classified as stationary data

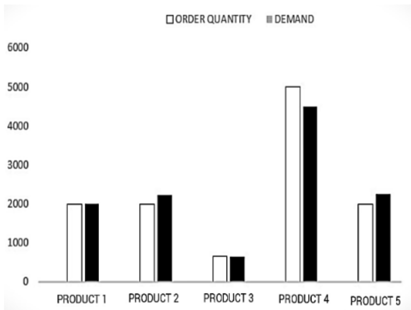

Fig. 2 Order Quantity v/s Demand

Figure 2 illustrates a clustered column comparing the quantity ordered from manufacturers and the quantity demanded from customers for five different products from the data in the time period 01 April 2024 to 19 July 2024. The vertical axis quantifies the number of orders in intervals of a thousand while the horizontal axis represents five different medicines.

The bars filled entirely with black color signify the cumulative number of demanded units of the corresponding product from the customers the third-party firm is dealing with. The bars bordered with black and filled with white color showcase the number of orders that the company placed to the manufacturer for a particular medicine.

A wide range of instances can be observed such as demand intersecting with the inventory, demand exceeding the quantity that is available demonstrating inadequate stock and demand falling short of the units available for a product delineating a scenario of excessive inventory.

IV. Methodology: Linear Regression

4.1 Linear Regression

The sales report of each client is generated from 01-04-2024 [DD-MM-YYYY] to 19-07-2024 for each of the products. Cumulatively, these sales reflect the quantity of products sold or the number of units sold of each medicine.

Each order placed by a customer for a particular product is taken along with its date. Different aspects as number of orders, total quantity, date of order placement are listed for a particular product from all the clientele.

4.2 Understanding Variance and Standard Deviation

Variance computes how much the values in a data set differ from the average of that set. Calculation of the variance involves three major steps. First, the difference between each data point and the mean of the data is found. These deviations from the mean are squared to remove negative values. Then, the mean of these squared deviations results in variance.

A higher variance indicates that the data points are more spread out from the mean, while a lower variance means that they are closer to the mean.

Standard deviation is a similar statistical measure. It indicates the typical distance between each data point and the average It is derived as the square root of variance. The variance is expressed in square units of measurement, which can make it less intuitive while analyzing the data. In contrast, standard deviation is expressed in same units of measurement which can make it more interpretable and practical.

4.3 Understanding Linear Regression

Linear regression is a mathematical method to understand the relationship between a certain number of variables. The relationship between a dependent variable and one or more independent variables is modelled as a linear function. Linear regression helps us draw a straight line through a set of data points that shows the general trend, though it may not capture each and every data point of the dataset.

Linear regression helps to find the best-fitting line that predicts the dependent variable based on independent variable.

The line has an equation like $y= mx + b$ … (1)

In Eq. (1)

‘y’ is the outcome or target variable we're trying to predict or explain.
‘m’ is the input or feature that influences the dependent variable.
‘x’ is the coefficient of the independent variable, showing the rate of change in y for each one-unit increase in ‘x’.
‘b’ is the constant term or y-intercept.

The core idea is to minimize the distance between the actual data points and the line. The closer the data points, the better the model predicts the relationship between independent variable and dependent variable.

Linear regression models are primarily categorized into two types: Simple and Multiple linear regression.

modeling the relationship between one dependent variable and one independent variable.

Multiple linear regression involves multiple independent variables to predict the dependent variable. In this context, we employ linear regression.

4.4 Constructing the predictive model

For a particular product, all the orders placed for it are listed. Orders from the same customer are listed. Parameters such as date, total quantity, client name are considered in separate columns for an optimal organization of the data.

It is also imperative to convert the month and week in our date of order placement from categorical variables to numerical variables. Different algorithms, in their basic form, understand numbers, not categories. They operate on mathematical principles that require numerical input.

By transforming categorical data into numerical form, we translate qualitative information into a language machine learning models can comprehend.

The number of months out of 12 months in a year (such as 7 for July, 2 for February) and the number of week (out of 52 weeks in a year) are calculated for each order entry. Such a numbering would further also assist us in obtaining a predicted demand for a particular month through linear regression.

Product	Month	Week	Total Quantity
G	4	14	50
G	4	14	100
G	4	14	110
G	4	14	40
G	4	14	50
G	4	14	10
G	4	14	20
G	4	14	200
G	4	15	100
G	4	15	130
G	4	16	150
G	4	17	720

Table 1: Orders throughout the month

For product G, the quantity of orders that are received in the same week are added up and cumulatively represented as the total number of orders received in that week of the particular month.

Product	Month	Week	Total Quantity
G	4	14	580
G	4	15	230
G	4	16	150
G	4	17	270

Table 2: Cumulated weekly orders in the month

With this information that is modified to meet the requirements, regression is performed in Microsoft Excel after the incorporation of Data Analysis Tool pack.

Total order quantity is considered as dependent variable and month and week numbers taken as independent variables for the analysis.

4.5 Interpreting the Results

Multiple R - 0.31579308

Multiple R is the Correlation Coefficient that tells us the strength of a linear relationship between two variables. The correlation coefficient can be any value between -1 and 1, and its absolute value indicates the relationship strength. The larger the value, the stronger the relationship is between the variables.

R^2- 0.09972527

R Square is the Coefficient of Determination, which is utilized to illustrate of the goodness of fit. It shows how many points fall on the regression line. In other words, it represents how accurately the data is synchronizing with our line of best fit. It is the sum of the squared deviations of the original datapoints from the mean.
Here the R^2 of our regression is 0.099 approximately. This implies that only a 9.9% of our values fit the regression model, showcasing a massive variation from the original values. Also, 9.9% of the variability in our dependent variables can be explained by our independent variables.

The R^2 value of our model, which is very low, is an indication that the predictive model is not a good fit for the kind of data we have.

Standard Error – 595.189398

Standard Error is an absolute measure that shows the average distance that the data points fall from the regression line.

In this case, a huge standard error says that the data points have a large separation from the regression line. These coefficients along with intercepts enable us to build a linear regression equation.

These coefficients along with intercepts enable us to build a linear regression equation.

Independent variables	Coefficient
Month	-626.7124248
Week	153.1212425
Intercept	**- 407.009018**

Table 3: Coefficients and Intercept

Y = C + m1* x1 + m2* x2 + …... (2)

In Eq. (2)
'Y' denotes the entity that we want to predict. In this case, it is the predicted demand a particular week/month.

'C' denotes the intercept from the results of running a regression.
'm' represents the coefficient independent variables in our data set. Here it refers to the coefficient of month i.e. -626.7124248.

'x' represents the quantitative representation of our variable.

[Symbol (*) represents multiplication]

Month Number	Week Number	Predicted Demand
4	14	43.8567134
4	15	196.977956
4	16	350.099198
4	17	503.220441

Table 4: Predicted demands in the weeks of April

Table 4 demonstrates the computing the predicted demand in different weeks of April and ultimately, obtaining the predicted demand for entire April.

In table 4, the predicted demand column is calculated using the Eq.3.

Predicted demand for a specific week = Intercept + Month number * Month Coefficient + Week number * Week coefficient …. (3)

Here, intercept, month and week coefficients are derived from Table 2.

Similarly, the predicted demands for the months May, June and July are also calculated.

We will now draw insights from various simulations to determine the potential profits in each scenario and assess how a firm should prepare to meet fluctuating market demands. Consider the case of 'Expected Sales of Product G in August.' Various random order quantity scenarios are analyzed, such as ordering 700 units of Product G from the manufacturer, 750 units, and so on. This approach is used to create hypothetical scenarios and evaluate the profit and loss associated with having a specific quantity of the product relative to market demand. A total of 21 plausible order quantities, ranging from 700 to 1700 units, are considered, with increments of 50 units per scenario.

Random demands are generated using the Microsoft Excel function:

NORMINV (RAND (), Mean, Standard Deviation)

This function returns the Inverse of Normal Cumulative Distribution for a specified mean and standard deviation.

The normal cumulative distribution function (CDF) provides the likelihood that a number will be less than or equal to a specific value in a bell-shaped curve. The CDF starts at 0 and goes up to 1, meaning it covers all possibilities. It helps us understand how likely something is to happen before reaching a certain value.

The inverse of a normal cumulative distribution aids in determining the value that corresponds to a given probability. The inverse CDF indicates what number delivers the chance that we are interested.

Here, the total expected demand for product G in August that was computed in previous steps is taken as the Mean in the Excel Function. The standard error that was obtained from the Regression results is taken as Standard Deviation.

Nineteen columns of Expected demands are randomly generated using the function below corresponding to the 21 order scenarios. As the demands are randomly generated by the function, it is plausible that there would be occurrences of negative demands and decimal demands.

In order to ensure that all the demands are whole numbers, the entire function is encapsulated in an integer function that rounds off the random demands in decimals to its nearest whole number.

INT (NORMINV (RAND (),1094,595))

Three hundred and ninety-nine scenarios are generated, each comprising an order quantity that the seller has and how much is demanded from the customer. This wide range of scenarios demonstrate instances of both inadequate stock and excess stock of varying degrees.

We considered a standard profit margin of Rs.25 and a salvage value of Rs.10 for product G. The profits are calculated in two ways:

1) When Demand>Order Quantity:

Then, Profit = [Order Quantity * 25]

2) When Demand<Order Quantity:

Then, Profit=
[Demand*25 – (Order Quantity – Demand) *10]

Note:
Demand refers to the quantity needed by the customer
Order quantity refers to the number of units with the seller.
[* denotes multiplication]

When the demand exceeds order quantity, we are calculating the profit by multiplying the profit margin with number of units sold.

When demand is less than the order quantity, only the demanded units are sold and this leaves some stock in the inventory. From the profit we procure by multiplying demand by 25, we subtract the salvage value of unsold units.

The negative demands that are generated randomly are adjusted to zero, as negative demand is not logically valid. The low values of 'Multiple R' and 'R^2' and the high 'standard error' indicate that the multiple linear regression model is not an efficient and optimal way to forecast the demand of such a product between variables. This means it looks for a straight line that best fits the data. However, when the data fluctuates or follows a non-linear pattern, a straight line cannot accurately represent the underlying relationship. This can lead to poor predictions and an oversimplified understanding of the data.

Additionally, if there are outliers or extreme fluctuations in the data, they can disproportionately influence the linear regression model, leading to a skewed line that doesn't represent the majority of the data points well.

This approach may not adequately capture the complexities of demand patterns, particularly in cases where data exhibits non-linear trends or significant fluctuations. In such scenarios, relying solely on linear regression can lead to inaccurate forecasts and suboptimal inventory management. Therefore, the adoption of a Time Series forecasting model, such as Holt's Trend-Adjusted Exponential Smoothing, presents a viable alternative.

V. Methodology: Trend adjusted exponential smoothing model

Time series forecasting is a method to predict future values based on historical data. Time series data consist of consecutive observations taken at regular intervals (daily, weekly, monthly, etc.), providing a useful tool for organizations to make informed decisions based on past events or trends. The objective is to model the underlying patterns or trends in the data to accurately forecast future values.

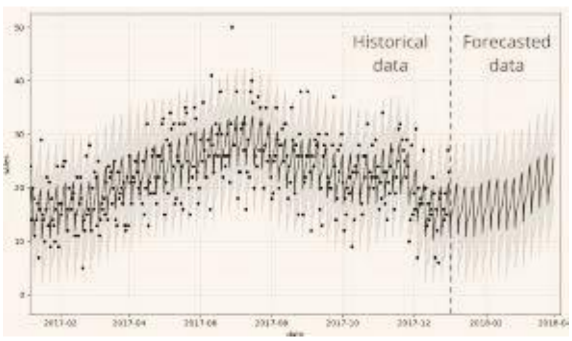

Fig. 3 Time Series Forecasting
(Shailesh https://shorturl.at/Aalyk)

Fig 3 illustrates future forecasts made by analyzing the movements in the historical data broadly categorized as trends, seasonality and variations. These are listed as the components of a time series.

5.1 Components of a Time Series:

Trend: The long-term pattern of time series is referred as trend. The trend of the data may be positive or negative reflecting the growth or decline in the values respectively. Fig 4 showcases an upward trend.

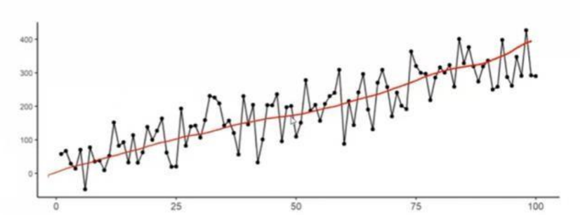

Fig. 4 Upward trend
(Shailesh https://shorturl.at/Aalyk)

Seasonality: A repeated pattern in the series that reoccurs at regular intervals (e.g., yearly, monthly). Fig.5 provides an example of such changes.

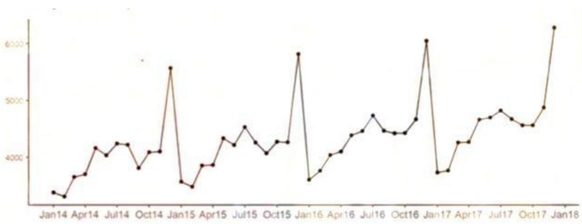

Fig.5 Considerable and periodic repetitions
(Shailesh https://shorturl.at/Aalyk)

Cyclic variations: Long term fluctuations whose periodicity of repetition is not fixed. These cycles are typically influenced by broader factors, such as economic conditions, market cycles, or business cycles, and can last for several years. Fig. 6 reflects such cyclic variations.

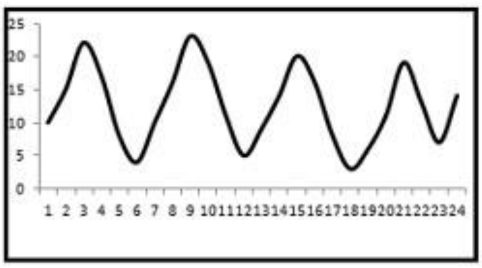

Fig.6 showing cyclic variation in the series
(Shailesh https://shorturl.at/Aalyk)

Irregular variations: They are also known as random or residual variations. They occur due to unforeseen events. Also known as unpredictable fluctuations, these variations shown by Fig.7 do not follow any pattern and are caused by random or rare occurrences.

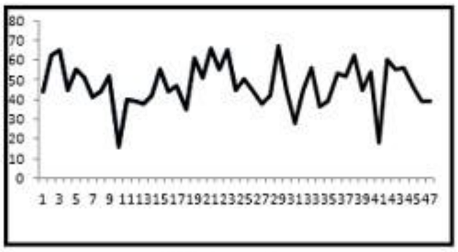

Fig.7 irregular variation in series
(Shailesh https://shorturl.at/Aalyk))

5.2 Moving average forecasting model:
It uses the average of past observations to smooth out noise and predict or forecast future values. The moving average models are primarily used for stationary data, where the demand has no observable trend or seasonality.

There are two types of moving average models:
Simple moving average and Weighted moving average

Simple Moving Average is a type of forecasting model that uses fixed number of data points for average calculation.

$$F_t = \frac{D_{t-1} + D_{t-2} + \cdots + D_{t-n}}{n} = \frac{1}{n}\sum_{i=1}^{n} D_{t-i}$$

Fig. 8 Simple Moving average

In Fig. 8,
'F_t' the forecasted demand of period 't'
'n' is the number of periods in moving average
'D_{t-i}' is the demand in the respective period

For a certain data, if we compute a 3-month and 6-month simple moving average models. The forecast based on a 6-month moving average is tend to be more accurate forecast or closer the actual demand/values in comparison to that of a 3-month moving average as we are supplying a bigger amount of historical data to our model to interpret and analyze.
Weighted moving average forecasting model is a type of moving Average model that uses a weighting scheme. Recent data points are prioritized or given more importance. This type of moving average model can be used for time-series forecasting, and can help to reduce the impact of older data points on the average.

$$F_t = W_1 D_{t-1} + W_2 D_{t-2} + \cdots + W_n A_{t-n}$$

Fig. 9 Weighted moving average

In Fig. 9,
'W_{i-1}' is the weight of i-th previous period ($0 \le W_i \le 1$)

5.3 Simple exponential smoothing model:

In the simple exponential smoothing model, the forecast value is compared with the actual value, and the weightage assigned to either the forecast or the actual value is adjusted based on the results of this comparison.

$F_{t+1} = \alpha D_t + (1- \alpha) F_t$ .... (4)

In eq 4,
$F_{(t+1)}$ is Forecast for the next period.
D_t is Actual demand for the period.
F_t is Forecasted demand for the period. α(alpha) is the weighing factor for smoothing.
The alpha (α) value is a number between 0 and 1 that determines how much weight or consideration is given to the most recent data point when forecasting.

When we give more weight or consideration to older data points, so the forecast reacts slowly to changes. we choose alpha to be close to 0 (e.g., 0.1). On the other case, we chose a value close to 1.

A relatively big **alpha (α)** value tracks and adjusts with the quick changes in the demand. This is useful if we expect rapid changes in the data.

5.4 Holt's Model

Holt's model adjusts for trends (increasing or decreasing patterns) over time. It is appropriate when demand is assumed to have a definite trend in, but has no certain seasonality.

Holt's model uses two smoothing factors: *α (alpha)* for the level and *β (beta)*for the trend. These help to weight recent observations more heavily than older ones.

Components of Holt's Model:
F_t : This denotes the forecast for the upcoming period. It predicts the demand for the future based on the level and trend in the previous period. For example, if we are trying to predict sales for *month 4,* we use information from *month 3*.
$F_t = L_{t-1} + T_{t-1}$ … (5)
To forecast for a period, we need the level and the trend of the previous period.
L_t : It is an estimate of the "current level" of demand. It reflects the overall average demand for the current period (without accounting for trends or patterns).
$L_t = \alpha D_t + (1-\alpha) (L_t-1 + T_t-1)$ ... (6)

In Eq.6

D_t is the actual observed demand for the current period. [L_{t-1}+ T_{t-1}] represents the previous estimate of demand from the previous period. 'α' is the smoothing constant for the level, a value between 0 and 1 that determines how much weight you give to the actual demand in a particular period comparison to the previous estimate or forecast we obtained for that period.

Eq.6 accounts for both the actual demand in the current period and the forecast from the previous period. It allows the forecast to incorporate new information (the actual demand) while still relying on the existing trend and level from previous periods. This balance helps make the model responsive to changes without overreacting to random fluctuations.

T_t **:** This reflects a prediction how fast the demand is increasing or decreasing over time. It captures the trend in the data. It could be compared to the slope in a linear equation.

$T_t = \beta\ (L_t - L_t\text{-}1) + (1\text{-}\beta)\ T_{t-1}$ … (7)

Eq.7 emphasizes upon updating the **trend**, which captures how the demand is changing over time gradually, such as increasing or decreasing.

In eq.7

[$\beta\ (L_t - L_t\text{-}1)$] shows how the demand is changing.

[$(1\text{-}\beta)\ T_{t-1}$] controls how much of the previous trend T_{t-1} is retained or valued. If (β beta) is small, the previous trend will have more influence on the new trend and vice versa.

5.5 Constructing the predictive model

We now again run a linear regression as the first step in this construction. But in this linear regression, we only consider "weeks which is a representative of the week number in the year in which the corresponding quantity of the product G has been sold, excluding the month parameter from the dataset for performing analysis. This can be seen in table 5.

Product	Week	Total quantity
G	14	580
G	15	230
G	16	150
G	17	720

Table 5: Rearrangement of data for Holt's Model's

Regression is performed in Microsoft Excel after the incorporation of Data Analysis Tool pack obtaining results shown in Table 6.

Parameter	Coefficient
Intercept	-96.4382353
Week	20.7529412

Table 6: Regression Analysis results

From table 6, we obtain intercept and coefficient of week to devise our model.

Intercept is taken as L_0

L_0 = -96.4382353…. (8)

The coefficient of increase or decrease is denoted by 'T'.

T_0 = 20.7529412… (9)

Initially, we compute the forecast for first period

$F_t = L_0 + T_0$

F_1 = -75.6852941…. (10)

In a similar manner, the level and the trend are calculated by also incorporating smoothing constants alpha and beta.

To generate random demands using the inverse of cumulative distribution function such as in linear regression, we need the error component.

Error component= Forecasted demand – Actual Demand
…. (11)

We square these errors and find the mean of the squared errors (MSE). Then we find the standard error (SE) which is the square root of MSE.

Now, we generate random demands using the excel function that calculates the inverse of the normal cumulative distribution for a given mean and standard deviation.

INT (NORMINV (RAND (),75.685,828.72))

The function is encapsulated with "INT" or integer function to ensure that they are no decimals for the sake of calculation.

75.685 is the forecasted demand [F_t] for the first week/period from our Holt's model that is taken as the mean in the syntax.

828.72 is the standard error (square root of MSE) that is taken as the standard deviation.

When the randomly generated demands give out a negative demand, it is manually set to zero to overcome the discrepancy as no such negative demand could exist. It also it trains the model to prepare for such scenarios when there is no demand for the product.

Holt's Trend-Adjusted Exponential Smoothing Model employs two smoothing factors: α (alpha) for the level and β (beta) for the trend, which are crucial for adapting to the dynamics of pharmaceutical demand data. The model estimates the current level of demand and the trend based on its time series, making it suitable for datasets with increasing or decreasing patterns. By integrating recent data, it effectively forecasts future demand, aligning with the stochastic nature of pharmaceutical sales

VI. Discussion

This paper explores the complexities of demand forecasting within the pharmaceutical industry, emphasizing its critical role in efficient supply chain management. The research highlights the risks associated with inaccurate predictions, such as product shortages and waste, and aims to enhance customer fulfillment and optimize inventory through reliable forecasting methods.

Two primary predictive models are employed: Multiple Linear Regression and Holt's Trend-Adjusted Exponential Smoothing Model.

Multiple Linear Regression is a widely used statistical technique that seeks to establish a linear relationship between independent variables and the dependent variable—in this case, demand. The advantages of this model include its simplicity and ease of interpretation. However, it has notable limitations, particularly when dealing with non-linear data patterns or outliers, which can skew results and lead to poor predictions. The model's reliance on historical data can also result in overfitting, where it performs well on past data but fails to generalize to future scenarios.

In contrast, Holt's Trend-Adjusted Exponential Smoothing Model offers a more nuanced approach by incorporating trends from previous periods to forecast future demand. This model is particularly effective for time series data that exhibit trends and seasonality. However, it is limited in its forecasting horizon; while Multiple Linear Regression can simulate various order scenarios to identify optimal order quantities for an entire month, Holt's model can only provide predictions for a single period—in this case, one week. This limitation arises because Holt's model focuses on the most recent data to generate forecasts, making it less suitable for long-term planning.

In summary, while both models have their strengths and weaknesses, the choice of model should align with the specific forecasting needs of the pharmaceutical sector. The research underscores the importance of accurate demand forecasting to mitigate risks and optimize inventory management, ultimately contributing to the sustainability and growth of the industry.

Acknowledgement

I would like to express my sincere gratitude to my research mentor, Ms. Nymisha Bandi, pursuing her PhD in Retail Management in McGill University in Canada, for her unwavering guidance, insightful feedback, and support throughout the course of this research. Their expertise and meticulous attention to detail were instrumental in shaping this work.
I also extend my heartfelt thanks to Mr. Kiran Kumar Ande, Chairman and Managing Director of Medipure Pharmaceuticals for generously providing the data necessary for this research. Their collaboration and willingness to share resources made this project possible.

References

[1]. Market Research Future. Canada Metal Packaging Market Material Segment Outlook 2016-2022. https://www.marketresearchfuture.com/statistical-reports/pharmaceutical-market-2232. Aug 7 2024

[2]. Statista. Total global pharmaceutical R&D spending 2014-2030. https://www.statista.com/statistics/309466/global-r-and-d-expenditure-for-pharmaceuticals. Aug 8 2024

[3]. Dolgin E. Better forecasting urged to avoid drug waste. Nature Medicine 2010; 16(1):5. https://doi.org/10.1038/nm0110-5. Aug 8 2024

[4]. Subramanian L. Effective Demand Forecasting in Health Supply Chains: Emerging Trend, Enablers, and Blockers. Logistics 2021; 5(1):12. https://doi.org/10.3390/logistics5010012. Aug 8 2024

[5]. Yani LPE, Aamer A. Demand forecasting accuracy in the pharmaceutical supply chain: a machine learning approach. International Journal of Pharmaceutical and Healthcare Marketing 2022; 17(1):1–23. https://doi.org/10.1108/ijphm-05-2021-0056Aug 8 2024

[6]. Kumar R, Singh A, Patel S. Pharmaceutical waste management and its impact on industry. Journal of Pharmaceutical Sciences 2020; 109(8):2487-2495.Aug 8 2024

[7]. Merkuryeva G, Valberga A, Smirnov A. Demand forecasting in pharmaceutical supply chains: A case study. Procedia Computer Science 2019; 149:3-10. https://doi.org/10.1016/j.procs.2019.01.100 Aug 8 2024

[8]. Tavakol P, Nahavandi B, Homayounfar M. A dynamics approach for modeling inventory fluctuations of the pharmaceutical supply chain in covid 19 pandemic. Journal of Optimization in Industrial Engineering 2023; 16(1):105–118. https://doi.org/10.22094/joie.2023.1966193.1981 Aug 8 2024

[9]. Siddiqui R, Azmat M, Ahmed S, Kummer S. A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry. Supply Chain Forum an International Journal 2021; 23(2):124–134. https://doi.org/10.1080/16258312.2021.1967081 Aug 9 2024