Forecasting Supply Chain Sporadic Demand Using Principal Component Analysis (PCA)

Nafi Ahmed, Shubho Roy and MD Ariful Islam
Department of Mechanical & Production Engineering
Ahsanullah University of Science & Technology
Dhaka, Bangladesh.
nafi.ipe@aust.edu, shubhoroy529@gmail.com, shaikatsameeh44@gmail.com

Abstract

Forecasting sporadic demand can be considered as one of the biggest challenges in supply chain management. It is very difficult to forecast the sporadic demand because of the irregularities present in this type of dataset. This ultimately reduces the overall supply chain performance. Traditional methods used to forecast sporadic demand include Simple Exponential Smoothing (SES), Croston's method, Multiple Linear Regression (MLR), and other parametric and non-parametric methods. However, none of these methods considers the factors behind the irregularities present in sporadic datasets. Principal component analysis (PCA), an artificial intelligence algorithm, can analyze a dataset of two or more variables and observations are explained by distinct inter-correspond variables. In this paper, a framework for forecasting sporadic demand considering multiple factors of the irregularities is presented using Principal component analysis (PCA). Furthermore, a numerical illustration is provided using automotive spare parts data to demonstrate the effectiveness of the proposed model. The main ambition of the proposed PCA model is to extract relevant information from the provided dataset and provide predictive models. The proposed forecasting model with PCA is able to reduce forecasting error and forecast the sporadic demand with a higher degree of accuracy compared to other traditional methods.

Keywords

Forecasting, Sporadic Demand, Principal Component Analysis, Supply Chain Performance, Artificial Intelligence.

1. Introduction

One of the biggest challenges for the supply chain managers is to forecast the sporadic demand as accurately as possible. In sporadic demand, it appears irregularly with a large proportion of zero values in between demand periods. The unpredictable nature of sporadic demand poses continual challenges for companies. It becomes difficult to manage sophisticated inventory systems and incurs excessive inventory or stockout costs. Moreover, in the case of spare parts, there is high devolution risk because of their particular functionalities. In today's world, the demand for spare parts is omnipresent. When a part or component stops functioning or call for replacement, the demand for spare parts emerges dramatically. So, their management is importantly maintained (Boylan et al., 2008). As demand structure is characteristically sporadic in nature, it means with different periods, numerous demand may have occurred. Sometimes it shows no demand at all. Not every demand that occurs needs to be in a single unit, a very low demand size (slow demand), or a 'constant' requirement (clumped demand). That need to say, demand sizes can be highly multivariate that leads to 'lumpy' demand (Wang & Syntetos, 2011). A convoluted determination is taken by the decision-makers while choosing the forecasting method because of the customer service. Many theoretical and practical methods are accessible for forecasting depend on the data that have been formulated for the last many years (Petropoulos et al. 2014); such as Simple Exponential Smoothing (SES), Moving Average (MA), etc. Frequently many companies practice methods proposed by Croston (1972) and Teunter et al. (2011). In the long run, it is statistically found that those methods are not highly promising in sporadic or intermittent demand cases. However, due to bias and constraints, neither of the traditional methods have shown sufficient precision in data sets of irregular demand.

Machine learning forecasts can use a large number of demand-related data and features to forecast future demand for patterns with different learning algorithms. Principal Component Analysis (PCA) is a dimensionality reduction method that is usually used to reduce a large number of input variables to a few linear combinations of data that still contains most of the information as a large dataset, and this few numbers of linear variables will correspond to Principal Components (PCs). Each PC is a linear combination of optimally weighted original variables. The first

component is required to have the largest possible variance. Using an orthogonal transformation, PCA converts a set of observations of possibly correlated variables or covariance into a set of values of linearly uncorrelated variables or PCs. PCA can be extended as Correspondence Analysis (CA) to handle qualitative variables and as Multiple Factor Analysis (MFA) to handle heterogeneous sets of variables. Principal components can be interpreted geometrically as high-dimensional data directions that absorb the maximum amount of variance and project it onto a smaller dimensional subspace while preserving most of the information. PCA provides a wide range of applications from computer vision to neuroscience, from medical data processing to psychology due to its flexibility in minimizing redundancies, and its ability to extract interpretable information. Our goal is to accurately predict the demand for automotive spare parts data using PCA that can outperform conventional methods. To summarize, this study has the following research objectives:

- Analyze the factors behind the irregularities present in the sporadic datasets
- Summarize the correlations among observed factors with a set of linear combinations
- Develop a sporadic demand forecasting model considering the factors of irregularities.
- To compare the performance of forecasting in terms of mean absolute deviation, mean absolute percent error and mean squared error with other traditional methods
- To minimize the percentage of error in forecasting sporadic demand

The remaining of this paper is organized as follows. Chapter 2 presents a brief literature review of the relevant topics. Section 3 discusses the methodology used in this study. Section 4 provides a numerical example, and Section 5 summarizes the experimental results. Lastly, Chapter 6 focuses on the final findings and recommendations for further study.

2. Background Literature

The first part of the background literature presents a brief review of the methods developed to forecast supply chain sporadic demand. It discusses the development of the methodology for forecasting the sporadic demand in the literature. The second part highlights different successful applications of Principal Component Analysis (PCA) in various research areas.

2.1 Forecasting sporadic demand

For forecasting sporadic demand, simple forecasting techniques such as Simple Exponential Smoothing (SES), Moving averages, and Naïve approaches have been used over the years. These methods are simple and easy to use (Petropoulos et al., 2013). The first specialized method for forecasting the sporadic demand was proposed by Croston (1972). He proposed a decomposition of the dataset into positive demand and arrival intervals. Rao (1973) has updated Croston's model to incorporate different terminology and applies a single exponential smoothing to non-zero demand sizes and inter demand intervals. Johnston & Boylan (1996) suggested a Poisson stream model that analyzes variations in demand with the order arrival process and describes a negative exponential distribution as inter-order arrivals. Afterward, the study measured the demand variance and add it to the demand approach that is evaluating its output by using a wide variety of operating conditions. After a few years, Syntetos & Boylan (2001) proposed a bias-correction approximation to Croston's method and successfully evaluated SBA (Syntetos & Boylan , 2005). Levén & Segerstedt (2004) suggested a procedure that used Erlang distribution with the mean and variance of the forecasted demand rate to be applied to the observed data. The system determines whether it is time to place a new order, based on probability shortages resulting from a probability distribution.

Willemain et al. (2004) has consciously generated conditions that challenge Croston's assumptions and has compared the Croston's method with Exponentially Weighted Moving Averages (EWMA) and has estimated the aggregate distribution of demand over fixed lead time by use of a new type of time series bootstrap. They apply the integral probability transformation to sporadic demand to determine consistency when forecasting an entire distribution. With the nine major industries' help, they showed that the bootstrapping method provides accurate estimates of demand distribution over a fixed lead time than exponential smoothing and Croston's method. Wallström & Segerstedt (2010) studied the moving or irregular demand for products where the forecasting time intervals are frequently zero. They proposed new tools and models to test the estimation of forecasting error measurement. Teunter et al. (2011) showed that two significant drawbacks exist in Croston's method. Firstly, the method is insufficient to deal with obsolescence issues because it is not upgraded with zero demand over numerous periods. Secondly, the method is biased, and this is true at all points in time. Boylan et al. (2008) addressed the second issue and the study proposed an unbiased model. The new method, which is unprecedented in many ways, is different from Croston's method because, instead of the demand interval, it updates the demand probability for each period. A wide-ranging modeling analysis tests the comparative performance of the current estimator. The results show their superior success and provide perspectives into the relation between demand forecasting and obsolescence. Demand is not constant for

a set of goods (or services); demand is distributed and unpredictable with a large proportion of zero values in the analyzed time series (Dombi et al., 2018). Sporadic demand has previously been researched mainly in connection with forecasting components and service parts (Vasumathi & Saradha, 2014). Several artificial intelligence-based approaches, such as Artificial Neural Networks, have been employed in forecasting sporadic demand, but they need extensive training datasets (Gutierrez et al., 2008). In recent years, Nikolopoulos et al. (2016) used Nearest Neighbors approaches to forecast the sporadic data. Ahmed & Farzana (2020) proposed Support Vector Machines (SVM) approaches for forecasting sporadic demand. Furthermore, Different bootstrapping methods have also been employed to perform the sporadic demand forecasting to obtain better results (Hasni et al., 2019).

2.2 Principal Component Analysis (PCA)

Pearson (1901) presented the first description of the PCA. He concluded that the best fitting straight line to a system of points coincides in direction with the maximum axis of the correlation. Hotelling (1933) started with the concept of factor analysis, allowing the evaluation of a smaller number of uncorrelated variables representing the original variable. Further research by Hotelling (1936) gave an accelerated version of the power method for finding Principal Components (PCs). Girshick (1936) provided some alternative derivative of PCs with the idea that sample PCs were estimates of maximum probability of underlying population PCs. Later, Gower (1966) addressed connections between PCA and other statistics and provided valuable geometric insights. PCA or PCA-related approaches such as principal component regression have also played an essential direct role in other statistical methods. Pires et al. (2008) proposed Principal Component Regression (PCR), a PCA based regression analysis that considers principal components (PCs) as independent variables instead of adopting original variables. The PCs are the linear combination of the original variables, which can be obtained by principal component analysis (PCA). The PCA transforms the original set of inter-correlated independent variables to a new set of uncorrelated variables (i.e., PCs). The use of such independent variables as PCs in multiple regression models is very useful in order to avoid the multi-collinearity problem and to classify the variables as most important in making the prediction (Abdul-Wahab et al., 2005; Çamdevýren et al., 2005; Sousa et al., 2007; Rajab et al., 2013). Multiple linear regression and PCA techniques are combined to model tropospheric ozone and to classify the significant factors influencing ozone levels (Abdul-Wahab et al., 2005). Çamdevýren et al. (2005) adopted PCs for multiple linear regression analysis in water quality research. Sousa et al. (2007) developed the PCR model using PCs as inputs for predicting ozone concentrations and compared this model with multiple linear regression and an artificial neural network feed model. Rajab et al. (2013) merged multiple regression model with PCA technique to enhance ozone level prediction. All of these studies have found that integrating PCs as independent variables in regression models has improved the model prediction and reduced the complexity of the model. Therefore, to improve the accuracy of the forecast, PCA can be implemented in sporadic demand forecasting.

3. Methodology

In this paper, Principal Component Analysis (PCA) is combined with Multiple Linear Regression (MLR) to perform Principal Component Regression (PCR) analysis. This PCR model has been applied to forecast the sporadic spare parts demand for better accuracy. A brief description of PCA, MLR, and PCR are presented in this section.

3.1 Principal Component Analysis

Principal components are new variables defined as linear combinations of the initial variables. These combinations are performed in such a way that these new variables are uncorrelated, and most information is contained in the first few components within the initial variables. The idea is that A-dimensional data will give A principal components (PCs), but PCA tries to put maximum possible information in the first few ones. It helps to concentrate the research on the first few components without paying a major penalty in terms of information loss. The number of PCs is equal to or less than the number of original variables, or the number of observations. PC1 and PC2 can be obtained by Equation (1) and (2). Similarly, the values of all the PCs can be obtained by the similar equation like equation (1) and (2). If the number of PCs and original variables are equal, the first few PCs that can be used to describe the original observations and explain much of the variation in the data set.(Abdul-Wahab et al., 2005). That helps to reduce the original data set's dimensionality.

$$PC1 = b_{11}x_1 + b_{12}x_2 + \ldots\ldots\ldots\ldots + b_{1n}x_n = \sum_{k=1}^{n} b_{1k}x_k \quad (1)$$

$$PC2 = b_{21}x_1 + b_{22}x_2 + \ldots\ldots\ldots\ldots + b_{2n}x_n = \sum_{k=1}^{n} b_{2k}x_k \quad (2)$$

where $x_1, x_2, \ldots x_k$ are the original variable in the dataset. The eigenvalues are the variance of the PCs, and b_{kk} are the eigenvectors extracted from the covariance or correlation matrix of the dataset. It is often convenient that all the variables, x_k, are standardized to zero mean and one standard deviation. The eigenvectors, b_{kk}, are calculated using covariance or correlation matrix. As the covariance or correlation matrix is symmetric positive definite, it yields an

orthogonal basis of eigenvectors, each of which has a non-negative eigenvalue. These eigenvectors, multiplied by the original inputs as equation (3), correspond to PCs and the eigenvalues are proportional to the variances explained by the PCs. There are several ways to define eigenvectors and eigenvalues; the most common approach defines an eigenvector of the matrix A as a vector b_{kk} that satisfies the equation (3).

$$|A - \lambda I| b_{kk} = 0 \tag{3}$$

where λ is a scalar called the eigenvalue associated to the eigenvector, A is the correlation/covariance matrix, and I is the identity matrix. As a consequence of the variations in the variables used in the units of necessity for automotive spare parts, a correlation matrix was used to obtain eigenvalues and eigenvectors. A $k \times k$ matrix of co-efficient, known as variable loading, is produced by the eigenvectors multiplied by the square root of eigenvalues. These loadings reflect each of the original variables to an individual PC. In addition, a new set of data values known as component scores is generated using the sum of the variable loading and the values of the original variables. In the various linear equations, these scores can be used as new variables for predicting future automotive spare parts.

3.2 Multiple Regression Analysis

Multiple regression is an extension of simple linear regression. It is used when the unknown value of a variable is calculated based on the known value of two or more variables. The predictive variable is referred to as the dependent variable. The variables used to determine the dependent variable's value are also known as the independent variables. The multiple regression allows the overall fit of the model and the relative contribution of each of its predictors to the total variance described. More precisely, multiple regression analysis helps to predict the value of Y for given values of $x_1, x_2, \ldots x_k$. In general, the multiple regression equation of Y on $x_1, x_2, \ldots x_k$ is given by equation (4).

$$Y = b_0 + b_1x_1 + b_2x_2 + \ldots\ldots\ldots\ldots + b_nx_n \tag{4}$$

where Y is the dependent variables, b_i (i=0,...., k) are the parameters generally estimated by least squares method and x_i (i = 0,... k) are the independent variables.

3.3 Principal Component Regression

In principal components regression (PCR), principal components analysis (PCA) is used to decompose the independent (x) variables into an orthogonal basis (the principal components) and select a subset of those components as the variables to predict Y. PCA has been used in linear regression to fulfill two main objectives. The first one is performed on datasets, where predictor variables are too high. This was a method of minimizing dimensionality along with the Partial Least Squares Regression. The second goal of PCR is to get rid of collinearities between variables. Because each subsequent principal component is orthogonal, PCR has been used to prevent errors caused by dependencies between assumed independent variables in regression (Hadi & Ling, 1998). The PCR analysis combines MLR and PCA for the relationship between the dependent variables and the selected input PCs (Pires et al., 2008). The principal component scores derived from the PCA are taken for PCR analysis as an independent variable within the multiple linear regression equation. The general form of PCR model is given in equation (5).

$$Y = b_1 \times PC1 + b_2 \times PC2 + \ldots\ldots\ldots\ldots + b_n \times PC_n \tag{5}$$

3.4 Model Evaluation Criteria

The MLR and PCR are compared by accepting three statistical performance basis: (i) mean absolute deviation (MAD); (ii) mean absolute percentage error (MAPE); and (iii) mean squared error (MSE); the practice of estimating forecasting methods uses the sum of elementary mistakes. Mean Absolute Deviation (MAD) measures the accuracy of the prediction by averaging the absolute value of each error. MAD is an interpretive data analysis tool, which is consisted of a certain volume of enlightening observations about the contour of a distribution and also indicates the correlation between two variables (Elamir, 2012). A data set with a shorter MAD value shows a convenient relationship with mean than a data set with larger MAD value and it can't be zero because that means the data set have no deviations, as well as all the values are same. Mean Absolute Percentage Error (MAPE) is determined by using the absolute error in specific stage, which is divided by the detect values and the values should be obvious for that period (Khair et al., 2017). MAPE is a process of measuring error where low value of MAPE gives an expected result. Mean square error is admitted to be more considerable than the usually adopted principals, the remaining sum of squares. MSE develops into zero means the expected outputs of prediction is surely harmonized by the actual data set. The values of these appraisal statistics are computed by the equations displayed in table 1 from the modeled and observed values of the dependent variable.

Table 1. Equation used to calculate the model performance indices

Performance Indices	Equation
MAD	$\frac{\sum\lvert O-P\rvert}{k}$
MAPE	$\frac{\sum\frac{\lvert O-P\rvert}{O}}{k}\times 100\%$
MSE	$\frac{1}{k}\sum(O-P)^2$

O: Observed demand, *P*: Model estimated demand, *k*: Number of observations.

4. Numerical Example

Sporadic demand is usually observed in slow-moving items such as spare parts of machinery, spare parts of motor vehicles and etc. Companies need to ensure the stock of the spare parts at the time of maintenance for operating their operations smoothly. Therefore, they need to forecast the demand for the spare parts beforehand to avoid stockouts. In this study, we have taken a real-life sales data of "Brake Pad", an important spare part in the automotive industry. We observed many zero values as well as the irregularities present in the sales data. Generally, there are some factors responsible for the variations (Schnaars, 1984). We have found five factors that might be responsible for these variations. The factors are average temperature, number of rainy days, price discounts, occasions, and inflation rate. We have taken a total of fifty-two weekly sales data from October 2017 to September 2018. Furthermore, we have collected the data of the factors corresponding to each weeks. The data of the factors are collected from the Bangladesh Bureau of Statistics. Due to space constraints, only first twenty-six sample data are shown in table 2.

Table 2. Weekly sales value with the factors of irregularities

Week Number	Sales Data (Weekly)	Average Temperature (° Celsius)	Number of Rainy Days	Price Discounts (Percentage)	Occasions (Present: 1, Absent: 0)	Inflation Rate (Percentage)
1	8	25	0	0	0	5.61
2	0	23	1	0	0	5.51
3	2	22	2	0	0	5.61
4	8	22	0	10	0	5.51
5	2	21	0	10	0	5.61
6	5	22	1	10	0	5.61
7	0	18	2	10	0	5.61
8	8	20	1	0	0	5.61
9	3	21	0	0	0	5.61
10	8	18	2	0	1	5.61
11	7	17	1	20	1	5.61
12	8	16	1	20	1	5.61
13	0	16	1	20	1	5.61
14	4	18	0	20	0	5.62
15	1	15	1	10	0	5.62
16	0	16	2	10	0	5.62
17	4	18	2	10	0	5.62
18	0	20	1	10	0	5.62
19	3	19	2	0	0	5.62
20	0	18	0	0	1	5.62
21	6	19	0	0	0	5.62
22	2	20	2	0	0	5.62
23	0	22	3	0	0	5.62
24	7	25	0	10	1	5.62
25	0	26	0	10	0	5.62
26	4	28	1	10	0	5.62

5. Experimental Results and Discussions

Principal component analysis is a method to project data in a higher-dimensional space into a lower-dimensional space by maximizing each dimension's variance. It reduces a large number of input variables to a few linear combinations of data containing most of the information as a large dataset. In order to understand the weekly sales of the Brake pad, the PCA is performed with five independent factors. In figure 1, all five factors are represented by a vector, and the direction and length of the vector indicate how each factor contributes to the three principal components. In figure 2, the scree plot shows all five components that explain total variance. The only clear break in the amount of variance accounted for by each component is between the first and third components. However, the first component by itself explains around 30% variance, and the first three components explain around 80% variance. Therefore, more components might not be needed as the first three principal components explain roughly four-fifths of the total variability in the standardized ratings.

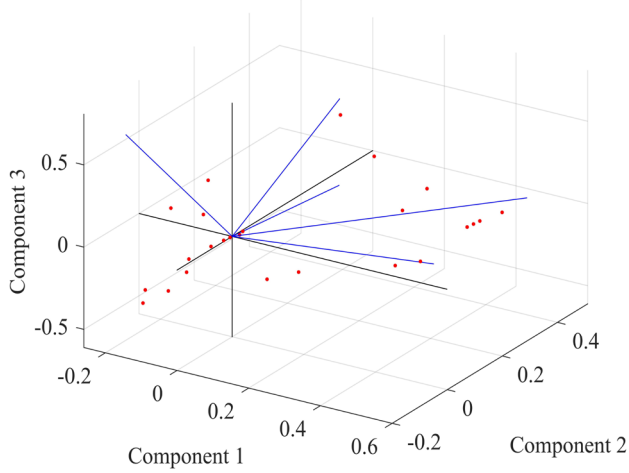

Figure 1. Principal component plots for first three components

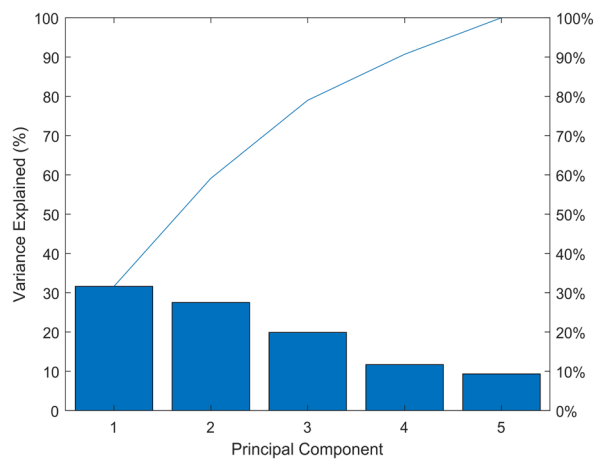

Figure 2. Scree plot of the percent variability explained by each Principal Component (PC)

Table 3 summarizes the correlation between original variables and five PCs. The value of variable loadings usually determines the output of a specific variable within a PC. The higher a variable's loading value, the more contribution the variable represents within a given PC. The bold labeled loadings in table 3 display the strong correlation between the factors and the corresponding PC. Table 3 shows that the first PC was heavily loaded with a discount, the second PC was with temperature, the third PC was with rainy days, the fourth PC was with inflation, and the fifth PC was with a discount. Component score coefficients and the values of the initial variables have been multiplied to achieve PC. These values were used in the step by step, multiple linear regression analysis as independent variables to evaluate the most appropriate PCs for the demand for spare parts. The PCR model was developed using data from October 2017 to September 2018. The model was then used to forecast the next six weeks of automotive spare parts demand.

Table 3. Component loading (correlation between original variables and five PCs)

Variables	PC1	PC2	PC3	PC4	PC5
Temp	-0.1254	**0.6804**	0.3490	0.5117	0.3709
Rainy Days	-0.3938	-0.0715	**-0.7977**	0.4154	0.1754
Discount	**0.6713**	-0.0557	-0.2299	-0.1075	**0.6940**
Occasion	0.6031	0.2901	-0.2618	0.3681	-0.5898
Inflation	-0.1206	0.6667	-0.3470	**-0.6468**	-0.0449

The results of the regression analysis are shown in table 4. The regression coefficients obtained from PCR and MLR have been used to develop the forecasting model, respectively as equation (6) and (7). The coefficients of PCR and MLR show that temperature, discount and presence of occasion has a positive impact on spare parts consumptions. It implies that the consumption of the spare parts generally increases when there are higher temperature, more price discounts and presence of occasions. In contrast, rainy days and inflation has a negative impact on sales.

Table 4. Results of regression analysis

Variables	Regression Coefficients (PCR)	Regression Coefficients (MLR)
Constant	4.0535	4.0497
Average temperature	0.0492	0.0611
Number of Rainy days	-0.3241	-0.0485
Percentage discount	0.0748	0.0765
Presence of Occasion	0.0092	0.0327
Inflation rate	-0.0013	-0.0014

The developed PCR model for forecasting sporadic demand is shown in equation (6).

$$\text{Forecast (PCR)} = 4.0535 + 0.0492\times\text{Avg. Temperature} - 0.3241\times\text{No. of Rainy Days} + 0.0748\times\text{Discount} + 0.0092\times\text{Ocassion} - 0.0013\times\text{Inflation Rate} \quad (6)$$

Similarly, the MLR model for forecasting sporadic demand is shown in equation (7).

$$\text{Forecast (MLR)} = 4.0497 + 0.0611\times\text{Avg. Temperature} - 0.0485\times\text{No. of Rainy Days} + 0.0765\times\text{Discount} + 0.0327\times\text{Ocassion} - 0.0014\times\text{Inflation Rate} \quad (7)$$

Table 5. Comparison of actual sales and forecasted sales

Week No	Actual Data	Forecasted Data for PCR (Weekly)	Forecasted Data for MLR (Weekly)
1	4	2.738	2.770
2	4	3.013	2.757
3	2	2.521	2.876
4	7	4.814	4.687
5	6	4.440	4.578
6	9	6.310	6.218

Table 5 shows the actual sales data and the forecasted data obtained from the PCR and MLR model. It can be observed that both the method perform satisfactorily compared to the actual sales data. Moreover, the models' relative performance during the forecasting period has been found to be almost identical. However, the PCR model slightly surpasses the MLR model. The PCR model is able to predict the sales with higher accuracy compared to the MLR model. For better understanding, Figure 3 illustrates the comparison of the actual data and the forecasted data obtained from the PCR model. Figure 3 shows that the actual weekly sales values for spare parts are closer to the values forecasted by the PCR model.

Table 6. Error Calculation with MAD, MSE, MAPE for PCR

Week No	Actual Data	Forecasted Data	Error	MAD	MSE	MAPE
1	4	2.738	1.262	1.262	1.593	0.315
2	4	3.013	0.987	1.124	1.284	0.281
3	2	2.521	0.521	0.923	0.946	0.274
4	7	4.814	2.186	1.239	1.904	0.283
5	6	4.440	1.559	1.303	2.010	0.279
6	9	6.310	2.689	1.534	2.880	0.282

Table 7. Error Calculation with MAD, MSE, MAPE for MLR

Week No	Actual Data	Forecasted Data	Error	MAD	MSE	MAPE
1	4	2.770	1.230	1.230	1.513	0.307
2	4	2.757	1.243	1.236	1.529	0.309
3	2	2.876	0.876	1.116	1.275	0.352
4	7	4.687	2.313	1.415	2.293	0.346
5	6	4.578	1.422	1.416	2.239	0.324
6	9	6.218	2.782	1.644	3.156	0.322

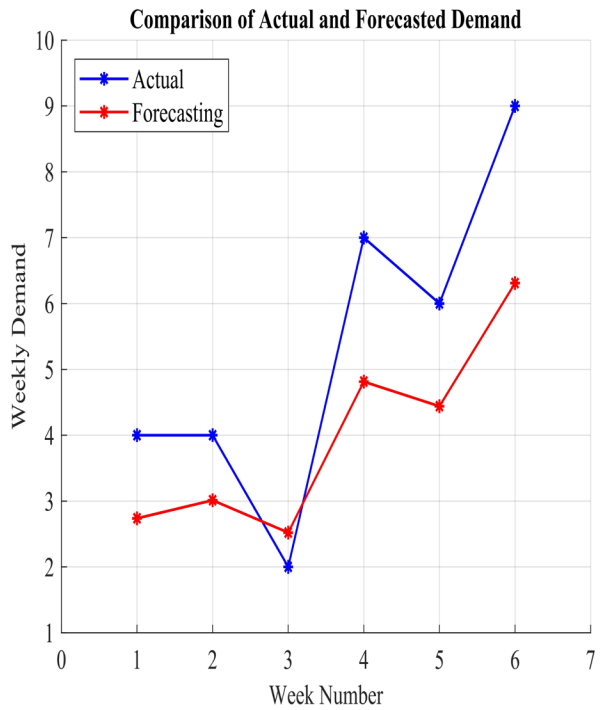

Figure 3. Comparison of actual and forecasted spare parts demand

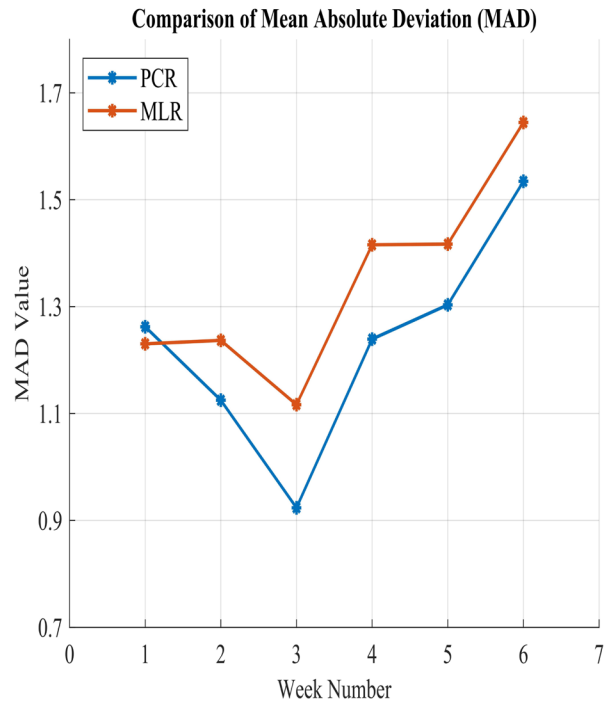

Figure 4. Comparison of Mean Absolute Deviation (MAD)

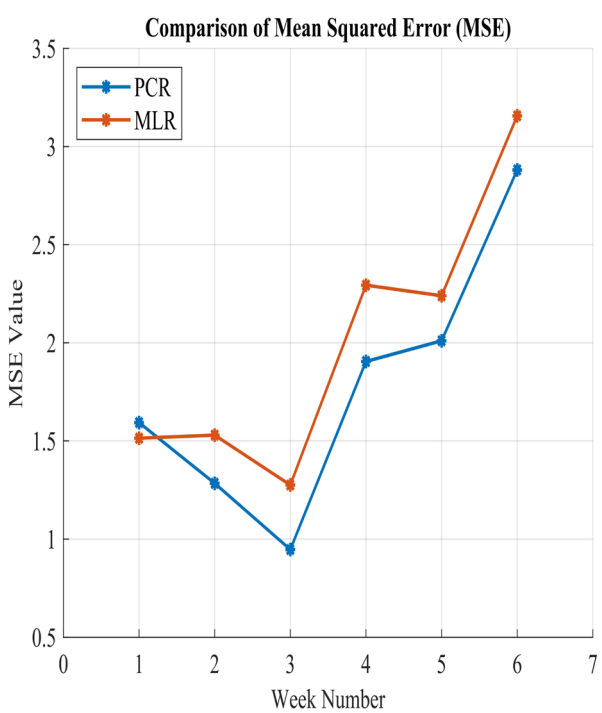

Figure 5. Comparison of Mean Squared Error (MSE)

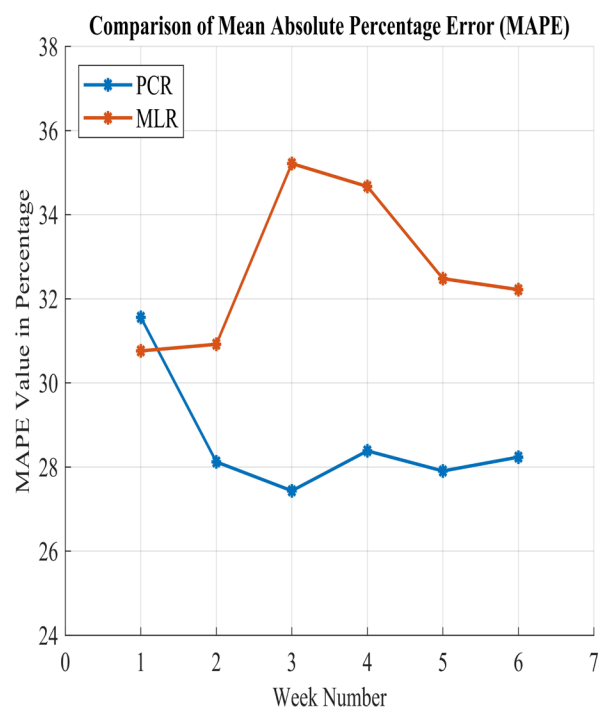

Figure 6. Comparison of Mean Absolute Percentage Error (MAPE)

5.1 Error Analysis

Error analysis is done to find the variation between the actual data and forecasted data. In table 6, the error calculations are shown for the PCR forecasting model in terms of three widely used performance indices, such as Mean Absolute Deviation (MAD), Mean Squared Error (MSE), Mean Absolute Percent Error (MAPE). The performance indices of the proposed PCR model indicates the effectiveness of the model. The MAD and MSE values are in an acceptable range. The MAPE values lie within 30%, which is a reasonable value for forecasting as suggested by Lewis, (1982). Similarly, the error calculations for the MLR model are shown in table 7. The detailed comparison of the three error indices for the PCR model and MLR model is shown graphically in figures 4, 5, and 6. Figure 4 shows that except for the first week, the PCR model outperforms the MLR model in terms of MAD. Similarly, in terms of MSE, the proposed PCR model performs better than the MLR model. Lastly, in figure 6, the comparison of MAPE is shown. It is seen that the value of MAPE for the MLR model at first increases and then decreases for subsequent weeks. However, the MAPE value for the PCR model indicates the constant better performance of the proposed model. Therefore, from the error analysis, it is observed that all of the performance statistics are in favor of the PCR model. The PCR model considers the PCs as independent variables which accounted for the contribution of all factors of irregularities and provides better forecasting.

6. Conclusion and Future research

Forecasting sporadic demand is indeed a challenging task for the practitioners because of the irregularities present and more zero values in this type of datasets. In this paper, a novel approach for forecasting sporadic demand using Principal Component Analysis is presented. The proposed model identifies some factors behind the irregularities. The insights are taken into account to perform the forecasting of the sporadic. The principal component regression (PCR) model was developed by combining multiple linear regression (MLR) and principal component analysis to identify the most important factors of irregularities behind the sporadic demand of the spare parts. Then, the developed PCR model is applied to forecast future spare parts consumptions. It is found that the first three Principal Components (PCs) are the most significant variables in the PCR model. Therefore, the factors which had significant loadings within these PCs could be considered as important predictor factors for spare parts demand forecasting. These important factors include average temperature, number of rainy days, price discounts, presence of occasions, and inflation rate. However, these factors are found to be correlated with each other. The inclusion of all these factors in the forecasting process provides better insight for identifying the patterns of the dataset. Therefore, the developed PCR model is used to forecast the future spare parts demand, which showed a high degree of forecasting accuracy. The results obtained from the developed PCR model show that some factors such as average temperature, price discounts, and presence of occasion are responsible for higher spare parts demand. On the other hand, factors such as the number of rainy days and the inflation rate negatively impact the spare parts consumption. The developed PCR model with three PCs as independent variables is also able to explain about 80 percent variation in spare parts consumption. For the forecasting period, the performance of the developed PCR model is compared to the MLR model. And the result shows that both models can predict demand closer to real demand, but the PCR model has outperformed the MLR in terms of performance indices. All the performance statistics such as Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) are found to be in favor of the PCR model. Therefore, the proposed model can be employed in sporadic demand forecasting for better forecasting accuracy.

It is recommended that future research may incorporate more factors behind the irregularities present in the sporadic demand datasets for automotive spare parts. Moreover, new factors might be identified for the forecasting sporadic demand for other products. The future works may consider other powerful regression methods such as support vector approaches for combining with the Principal Component Analysis (PCA) to obtain better accuracy in forecasting sporadic demand.

Acknowledgements

This research has been done under fully cooperation and resources of Department of Mechanical and Production Engineering, Ahsanullah University of Science & Technology (AUST). The authors express gratitude for all the efforts and cooperation to complete the research.

References

Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software*, *20*(10), 1263–1271.

Ahmed, N., & Farzana, F. (2020). Forecasting Supply Chain Sporadic Demand Using Support Vector Machine Approaches. *International Journal of Applications of Fuzzy Sets and Artificial Intelligence*, *10*, 87–102.

Boylan, J. E., Syntetos, A. A., & Karakostas, G. C. (2008). Classification for forecasting and stock control: A case study. *Journal of the Operational Research Society*, *59*(4), 473–481.

Çamdevýren, H., Demýr, N., Kanik, A., & Keskýn, S. (2005). Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling*, *181*(4), 581–589.

Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly (1970-1977)*, *23*(3), 289.

Dombi, J., Jónás, T., & Tóth, Z. E. (2018). M2odeling and long-term forecasting demand in spare parts logistics businesses. *International Journal of Production Economics*, *201*, 1–17.

Elamir, E. A. H. (2012). Mean absolute deviation about median as a tool of explanatory data analysis. *Lecture Notes in Engineering and Computer Science*, *2197*(June), 324–329.

Girshick, M. A. (1936). Principal Components. *Journal of the American Statistical Association*, *31*(195), 519–528.

Gower, J. C. (1966). Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika*, *53*(3/4), 325.

Gutierrez, R. S., Solis, A. O., & Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, *111*(2), 409–420.

Hadi, A. S., & Ling, R. F. (1998). Some Cautionary Notes on the Use of Principal Components Regression. *The American Statistician*, *52*(1), 15–19.

Hasni, M., Aguir, M. S., Babai, M. Z., & Jemai, Z. (2019). On the performance of adjusted bootstrapping methods for intermittent demand forecasting. *International Journal of Production Economics*, *216*(May 2018), 145–153.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441.

Khair, U., Fahmi, H., Hakim, S. Al, & Rahim, R. (2017). Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error. *Journal of Physics: Conference Series*, *930*(1).

Levén, E., & Segerstedt, A. (2004). Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics*, *90*(3), 361–367.

Lewis, C. D. (1982). *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. London: Butterworth Scientific.

Nikolopoulos, K. I., Babai, M. Z., & Bozos, K. (2016). Forecasting supply chain sporadic demand with nearest neighbor approaches. *International Journal of Production Economics*, *177*, 139–148.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. *Proeedings of the Royal Society*, *60*(1834), 489–498.

Petropoulos, F., Nikolopoulos, K., Spithourakis, G. P., & Assimakopoulos, V. (2013). Empirical heuristics for improving intermittent demand forecasting. *Industrial Management and Data Systems*, *113*(5), 683–696.

Pires, J. C. M., Martins, F. G., Sousa, S. I. V, Alvim-Ferraz, M. C. M., & Pereira, M. C. (2008). Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling & Software*, *23*(1), 50–55.

Rajab, J. M., MatJafri, M. Z., & Lim, H. S. (2013). Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia. *Atmospheric Environment*, *71*, 36–43.

Schnaars, S. P. (1984). Situational Factors Affecting Forecast Accuracy. *Journal of Marketing Research*, *21*(3), 290.

Sousa, S. I. V, Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, *22*(1), 97–103.

Syntetos, A.A, & Boylan, J. . (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics*, *71*(1–3), 457–466.

Syntetos, Aris A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, *21*(2), 303–314.

Teunter, R. H., Syntetos, A. A., & Zied Babai, M. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, *214*(3), 606–615.

Vasumathi, B., & Saradha, A. (2014). Forecasting of Sporadic Demand Patterns with Spare Parts. *International Journal of Innovation and Applied Studies*, *8*(1), 237–242.

Wallström, P., & Segerstedt, A. (2010). Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics*, *128*(2), 625–636.

Wang, W., & Syntetos, A. A. (2011). Spare parts demand: Linking forecasting to equipment maintenance. *Transportation Research Part E: Logistics and Transportation Review*, *47*(6), 1194–1209.

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, *20*(3), 375–387.

Biographies

Nafi Ahmed is a lecturer in the Department of Mechanical and Production Engineering at Ahsanullah University of Science and Technology (AUST), Dhaka, Bangladesh. He earned B.Sc. in Industrial and Production Engineering from Bangladesh University of Engineering and Technology (BUET) in 2017. Currently, He is pursuing M.Sc. in Industrial and Production Engineering from Bangladesh University of Engineering and Technology (BUET). He has a keen interest in research and actively involved in diverse research works. His research interests include optimization, machine learning, operation research and supply chain management. He is a member of Industrial Engineering and Operations Management (IEOM) Society, Institute for Operations Research and the Management Sciences (INFORMS) and Institution of Engineers, Bangladesh (IEB).

Shubho Roy is pursuing his B.Sc. in Industrial and Production Engineering at Ahsanullah University of Science and Technology, Bangladesh. He is a mid-level analysist in supply chain, six sigma, machine learning, and excel for business data analysis. Currently, he is involved in some research projects on machine learning and supply chain management. Besides, he has excellent skills on advance machine learning algorithm and python. His research interest include forecasting, machine learning, optimization, scheduling, advanced manufacturing and operation management.

MD Ariful Islam is pursuing his B.Sc. in Industrial and Production Engineering at Ahsanullah University of Science and Technology, Bangladesh. He is passionate about machine learning and supply chain management. Recently, he is working in a research group on machine learning and operation management. His research interests include operating systems, machine learning, operation management, artificial intelligence and lean production.