

DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF BUSINESS ADMINISTRATION (ENGLISH)
BUSINESS ADMINISTRATION (ENGLISH) PROGRAM
MASTER’S THESIS

**COMBINED FORECASTS FOR STOCK KEEPING
UNITS WITH INTERMITTENT DEMAND PATTERN**

Gizem HALİL

Supervisor

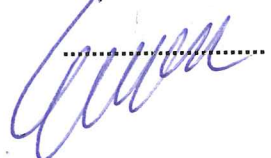
Associate Professor Aysun KAPUCUGİL İKİZ

İZMİR – 2019

MASTER THESIS/PROJECT
APPROVAL PAGE

University : Dokuz Eylül University
Graduate School : Graduate School of Social Sciences
Name and Surname : Gizem HALİL
Title of Thesis : Combined Forecasts for Stock Keeping Units with Intermittent Demand Pattern
Defence Date : 25/07/2019
Supervisor : Assoc.Prof.Dr.Aysun KAPUCUGİL İKİZ

EXAMINING COMMITTEE MEMBERS

<u>Title, Name and Surname</u>	<u>University</u>	<u>Signature</u>
Assoc.Prof.Dr.Aysun KAPUCUGİL İKİZ	- Dokuz Eylül University	
Prof.Dr.Sabri ERDEM	- Dokuz Eylül University	
Prof.Dr.Mehmet Emre GÜLER	- İzmir Katip Çelebi Üniversitesi	
Unanimity ☒		
Majority of votes ☐		

The thesis titled as "**Combined Forecasts for Stock Keeping Units with Intermittent Demand Pattern**" prepared and presented by Gizem HALİL is accepted and approved.

Prof. Dr. Metin ARIKAN
Director

DECLARATION

I hereby declare that this master's thesis titled as "**Combined Forecasts for Stock Keeping Units with Intermittent Demand**" has been written by myself in accordance with the academic rules and ethical conduct. I also declare that all materials benefited in this thesis consist of the mentioned resources in the reference list. I verify all these with my honour.

05.03/2019

Gizem HALİL



ABSTRACT
Master's Thesis
Combined Forecasts for Stock Keeping Units with Intermittent Demand
Gizem HALİL

Dokuz Eylül University
Graduate School of Social Sciences
Department of Business Administration (English)
Business Administration (English) Program

Effective inventory management is vital for the success of companies as it provides a competitive advantage. Many institutions hold different types of stock keeping units (SKU) on hand. To give an example, SKUs can have a normal pattern like finished goods; have intermittent demand pattern which has no demand for numerous periods and infrequent demand arrivals with highly variable sizes like spare parts. Accurate forecasting is crucial for inventory management. However, forecasting is a challenging task for the SKUs with intermittent demand. The irregular pattern causes traditional forecasting methods like moving averages or single exponential smoothing (SES) to perform poorly. To overcome this problem, several forecasting methods like the seminal method, “Croston’s method (CR),” modifications on CR – such as Syntetos-Boylan Approximation (SBA), and new methods like Artificial Neural Networks and Bootstrapping are proposed. Judgmental forecasting methods, which are based on expert opinions, are frequently used in practice as well. These methods are generally used when the historical data is absent and when there are significant changes in the environment and in the time series. Also, it is possible to combine statistical and judgmental forecasts. The studies state that combining forecasts help to eliminate biases and lead to increased accuracy.

This study proposes a framework for combining statistical and judgmental methods to generate forecasts with higher accuracy, which at the same time is easy to implement and interpret by the practitioners. First, The SKUs are categorized according to their demand patterns (intermittent, lumpy,

erratic, and smooth) by using Syntetos-Boylan-Croston (SBC) categorization scheme. Then SES, CR, and SBA, as one of the variants of CR, are evaluated to select the best method and parameters for each SKU. In parallel, judgmental forecasts should be obtained by using structured methods that reflect the experience and knowledge about the SKUs. Lastly, the generated statistical and judgmental forecasts are combined based on the weighted averaging procedure.

The proposed framework is demonstrated on two real-world datasets for spare parts from the automotive and small home appliances industries. The results showed that SES can perform well for SKUs with intermittent demand. Combined forecasts generate higher accuracy than the actual judgmental forecasts of the companies. For most SKUs in the first dataset (automotive), statistical forecasts are the best performing methods. For the second dataset (small home appliances), combination forecasts outperformed both judgmental and statistical forecasts.

The outcomes of this study can help institutions to manage their inventory successfully by generating more accurate forecasts. The costs that occur due to stock-outs, overstocks, and risk of obsolescence will decrease, while the customer satisfaction level increases. Also, some spare parts which are strategic for the companies' operations are costly and have a high stock value. Improving the management of such parts will end up with significant savings.

Keywords: Statistical Forecasting, Judgmental Forecasting, Combining Forecasts, Intermittent Demand, Stock Keeping Units, Spare Parts.

ÖZET
Yüksek Lisans Tezi
Kesintili Talep Yapısına Sahip Stok Tutma Birimleri için Birleştirilmiş
Tahminler
Gizem HALİL

Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü
İşletme Anabilim Dalı
İngilizce İşletme Programı

Günümüzün rekabetçi koşullarında etkin envanter yönetimi, rekabet avantajı sağladığından şirketlerin başarısı için hayati öneme sahiptir. Birçok kurum farklı talep yapılarına sahip stok tutma birimlerini (STB) elinde bulundurmaktadır. STB'ler nihai ürünler gibi normal bir yapıda olabilecekleri gibi, yedek parçalar gibi çok uzun süre talep görmeyen, gerçekleşen taleplerin ise değişken miktarlarda olduğu kesintili talep yapısına da sahip olabilirler

Doğru tahmin, envanter yönetimi için çok önemlidir. Bununla birlikte, tahminleme kesintili talep yapısına sahip STB'ler için zorlu bir süreçtir. Düzensiz talep yapısı, hareketli ortalamalar veya üstel düzeltme (ÜD) gibi geleneksel tahminleme yöntemlerinin düşük performans göstermesine neden olur. Bu sorunun üstesinden gelmek için literatürde, “Croston yöntemi (CR)”, Syntetos-Boylan Yaklaşımı (SBY) gibi CR yönteminin varyasyonları ve Yapay Sinir Ağları ve bootstrapping gibi yeni yöntemler önerilmiştir. Uygulamada uzman görüşlerine dayanan yargısal tahminleme yöntemleri de sıklıkla kullanılmaktadır. Bu yöntemler genellikle geçmiş veriler bulunmadığında ve çevrede ve zaman serilerinde önemli değişiklikler olduğunda kullanılır. Ayrıca, istatistiksel ve yargısal tahminleri birleştirmek mümkündür. Çalışmalar, tahminlerin birleştirilmesinin sapmaları ortadan kaldırmaya yardımcı olduğunu ve tahminlerin doğruluğunu artırdığını belirtmektedir.

Bu çalışma, kesintili talep yapısına sahip STB'ler için daha yüksek doğrulukta tahminler üretmek amacıyla istatistiksel ve yargısal yöntemleri

birleştiren, aynı zamanda uygulayıcılar tarafından uygulanması ve yorumlanması kolay, bütünleşik bir tahminleme modeli önermektedir. Önerilen metodolojik çerçevede ilk olarak, STB'ler Syntetos-Boylan-Croston (SBC) sınıflandırma şemasını kullanarak talep modellerine göre sınıflandırılır. Ardından ÜD, CR ve SBY yöntemleri ile oluşturulan tahminler her bir STB için en iyi yöntemi ve parametreleri seçmek üzere değerlendirilir. Paralel olarak, yapısal yöntemler kullanılarak uzmanların söz konusu STB'ler hakkındaki deneyim ve bilgisini yansıtan yargısal tahminler elde edilmelidir. Son olarak, oluşturulan istatistiksel ve yargısal tahminlerin ağırlıklı ortalama prosedürüne göre birleştirilmesiyle oluşturulan tahminlerin performansı her STB için hesaplanır ve her STB için en iyi performans gösteren yöntem seçilir.

Önerilen metodolojik çerçeve, otomotiv ve küçük ev aletleri sektörlerinde faaliyet gösteren iki firmadan alınan satış sonrası yedek parça taleplerini içeren gerçek veri setlerinde denenmiştir. Sonuçlar, literatürde belirtildiğinin aksine, ÜD'nin kesintili talebe sahip STB'ler için iyi bir performans gösterdiğini göstermiştir. Her iki veri seti için de bütünleşik tahminler şirketlerin kullanmakta olduğu yargısal tahminlerden daha yüksek doğruluk sağlamaktadır. İlk veri setindeki (otomotiv) çoğu STB için, istatistiksel tahminler en yüksek performansı gösterir. İkinci veri seti için (küçük ev aletleri), bütünleşik tahminler hem yargısal hem de istatistiksel tahminlerden daha yüksek doğruluğa sahip tahminler üretmiştir.

Bu çalışmanın sonuçları, kurumların en doğru tahminleri üreterek envanterlerini başarıyla yönetmelerine yardımcı olabilir. Yetersiz stoktan dolayı talebin karşılanamaması, fazla stok tutma ve stoktaki STB'lerin değer yitirmesinden dolayı ortaya çıkan maliyetler düşerken, taleplerin zamanında karşılanabilmesiyle müşteri memnuniyeti artar. Şirketlerin operasyonları için stratejik olan bazı yedek parçalar çok pahalıdır ve stok değeri yüksektir. Bu parçaların envanter yönetiminin daha doğru tahminler ile iyileştirilmesi önemli tasarruflarla sonuçlanacaktır.

Anahtar Kelimeler: İstatistiksel Tahminleme, Yargısal Tahminleme, Bütünleşik Tahminler, Kesintili Talep, Yedek Parçalar, Stok Tutma Birimi.

COMBINED FORECASTS FOR STOCK KEEPING UNITS WITH INTERMITTENT DEMAND PATTERN

CONTENTS

THESIS APPROVAL PAGE	ii
DECLARATION	iii
ABSTRACT	iv
ÖZET	iv
ABBREVIATIONS	xi
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF APPENDICES	xv
INTRODUCTION	1

CHAPTER 1

INTERMITTENT DEMAND FORECASTING

1.1. DEMAND DATA CATEGORIZATION	7
1.1.1. Specification of Non-Normal Demand Patterns	8
1.2. DEMAND CATEGORIZATION SCHEMES	9
1.2.1. Williams Categorization Scheme	9
1.2.2. Johnston and Boylan Categorization Scheme	11
1.2.3. Eaves Categorization Scheme	11
1.2.4. Syntetos – Boylan – Croston (SBC) Categorization Scheme	12
1.2.5. Kostenko – Hyndman (KH) Categorization Scheme	14
1.3. QUANTITATIVE FORECASTING METHODS	15
1.3.1. Simple Moving Average (SMA)	16
1.3.2. Weighted Moving Average (WMA)	17
1.3.3. Single Exponential Smoothing	17

1.3.4. Croston's Method (CR)	18
1.3.5. Syntetos and Boylan Approximation (SBA)	20
1.3.6. Teunter-Syntetos-Boylan (TSB) Method	22
1.3.7. Bootstrapping	23
1.3.8. Artificial Neural Networks (ANN)	24
1.4. QUALITATIVE FORECASTING METHODS	25
1.4.1. The Jury of Executive Opinion	27
1.4.2. Delphi Method	27
1.4.3. Forecasting by Analogies	29
1.4.4. Scenario Forecasting	30

CHAPTER 2

COMBINING QUANTITATIVE AND QUALITATIVE FORECASTING

2.1. USING JUDGMENT IN MODEL BUILDING PROCESS	32
2.2. COMBINING FORECASTS	34
2.3. JUDGMENTAL ADJUSTMENT OF STATISTICAL FORECASTS	36
2.4. STATISTICAL CORRECTION OF JUDGMENTAL FORECASTS	39

CHAPTER 3

METHODOLOGICAL FRAMEWORK ON COMBINING FORECASTS

3.1. DATA PREPARATION	43
3.2. CATEGORIZATION OF THE DEMAND DATA	43
3.3. BUILDING A STATISTICAL FORECAST MODEL	44
3.4. BUILDING JUDGMENTAL FORECAST MODEL	48
3.5. COMBINING STATISTICAL AND JUDGMENTAL FORECASTS	49
3.6. EVALUATION	51

CHAPTER 4
DEMONSTRATION OF THE PROPOSED FRAMEWORK

4.1. DATA PREPARATION	61
4.2. CATEGORIZATION OF THE DEMAND DATA	63
4.3. BUILDING A STATISTICAL FORECAST MODEL	67
4.4. BUILDING A JUDGMENTAL FORECASTING MODEL	72
4.5. COMBINING STATISTICAL AND JUDGMENTAL FORECASTS	73
4.6. RESULTS/FINDINGS	76
CONCLUSION	92
REFERENCES	98

ABBREVIATIONS

ANN	Artificial Neural Networks
CE	Cumulative Error
CR	Croston's Method
CV2	Square of Coefficient of Variation
EWMA	Exponential Weighted Moving Average
KH	Kostenko-Hyndman
LS	Leven and Segerstedt
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAR	Mean Absolute Rate
MASE	Mean Absolute Scaled Error
MdAPE	Median Absolute Percentage Error
ME	Mean Error
MSE	Mean Square Error
MSR	Mean Square Rate
p	Average Inter-Demand Interval
PB	Percentage Better
PBt	Percentage Best
PIS	Periods in Stock
RGRMSE	Relative Geometric Root Mean Square Error
RMSE	Root Mean Square Error
sAE	Scaled Absolute Error
SBA	Syntetos-Boylan Approximation
SBC	Syntetos-Boylan-Croston
sCE	Scaled Cumulative Error
SES	Single Exponential Smoothing
SKU	Stock Keeping Unit
SMA	Simple Moving Average
sMAE	Scaled Mean Absolute Error

S-MAPE	Symmetric Mean Absolute Percentage Error
sMSE	Scaled Mean Square Error
sPIS	Scaled Periods in Stock
sSE	Scaled Square Error
TSB	Teunter-Syntetos-Babai Method
VZ	Viswanathan and Zhou
WMA	Weighted Moving Average
WSS	Willemain-Smart-Schwarz



LIST OF TABLES

Table 1: Suggested Forecasting Methods for Each Demand Category	13
Table 2: Croston Method's Algorithm	19
Table 3: TSB Forecasting Model	22
Table 4: Performance Measures	53
Table 5: Descriptive Statistics for Dataset A	61
Table 6: Descriptive Statistics for Dataset B	63
Table 7: Demand Patterns for Dataset A	64
Table 8: Demand Patterns for Dataset B	66
Table 9: Partial Accuracy Table for Dataset A	72
Table 10: An Example of Combining Forecasts for Dataset B	75
Table 11: SBC Categorization for Datasets A and B	77
Table 12: Evaluation: Initial Values for Demand Patterns (Dataset A)	78
Table 13: Evaluation: Statistical Methods for Demand Patterns (Dataset A)	79
Table 14: Evaluation: Alpha Values for Demand Patterns (Dataset A)	80
Table 15: Evaluation: Initial Values for Demand Patterns (Dataset B)	81
Table 16: Evaluation: Statistical Methods for Demand Patterns (Dataset B)	83
Table 17: Evaluation: Alpha Values for Demand Patterns (Dataset B)	84
Table 18: Partial Accuracy Table: All Forecasts (Dataset A)	86
Table 19: Performance of Combined Forecasts Based on p and CV^2 (Dataset A)	87
Table 20: Partial Accuracy Table: All Forecasts (Dataset B)	89
Table 21: Performance of Combined Forecasts Based on p and CV^2 (Dataset B)	90

LIST OF FIGURES

Figure 1: Demand Patterns	9
Figure 2: Williams' Categorization Scheme	10
Figure 3: Eaves' Categorization	12
Figure 4: Eaves' Categorization Scheme	12
Figure 5: Syntetos-Boylan-Croston (SBC) Categorization Scheme	13
Figure 6: Disputed Parameter Space of Kostenko - Hyndman Categorization	14
Figure 7: Kostenko-Hyndman Categorization Scheme	15
Figure 8: Statistical Forecasting Methods for Intermittent Demand	16
Figure 9: Model Building Process	33
Figure 10: Combination of Objective and Subjective Forecasts	35
Figure 11: Judgmental Adjustment of Objective Forecasts	37
Figure 12: Quantitative Correction of Judgmental Forecasts	39
Figure 13: Stages of the Proposed Methodological Framework	42
Figure 14: Forecasting Process of Croston's Method	45
Figure 15: SBC Classification for Dataset A	64
Figure 16: Visual Representation of Demand Patterns Selected from Dataset A	65
Figure 17: SBC Classification for Dataset B	65
Figure 18: Visual Representation of Demand Patterns Selected from Dataset B	66
Figure 19: RStudio – tsintermittent SBA Codes for Product 11 (From Dataset A)	68
Figure 20: RStudio – Output of SBA Forecasts for Product 11 (From Dataset A)	69

LIST OF APPENDICES

Appendix 1: R Output of SBC Categorization Scheme for Dataset A	app. p.1
Appendix 2: R Formulation of <i>sexsm</i> Function (for SES)	app. p.2
Appendix 3: R Formulation of <i>crost</i> Function (for CR and SBA)	app. p.3
Appendix 4: R Codes for Forecasting the Demand	app. p.4
Appendix 5: Regression Tables for Dataset A – p & CV ²	app. p.5
Appendix 6: Regression Tables for Dataset B – p & CV ²	app. p.9



INTRODUCTION

For many companies, inventory is one of the most expensive assets and it may represent up to 50% of the total investment (Heizer and Render, 2014: 512). To survive in the competitive world, management of this asset is vital for the organizations. Inventory management is redressing the balance between the investment made in inventory and the customer service level. With successful management, companies can fulfill customer orders on time and increase customer satisfaction. They can have better inventory planning, which will help them to save money and time. Also, they can increase the efficiency and productivity in their institutions. It provides a competitive advantage over competitors; thus, it is directly related to the success of a company.

Companies hold different SKUs such as finished goods and work-in-process items with regular demand, which shows a smooth pattern or service parts which are used for after-sales services or for repairing and maintenance purposes that have an irregular and intermittent demand pattern. Characteristics of the SKUs play an important role in managing the inventory since they require different management methods. In the intermittent demand pattern, there are many periods with zero demand, and the demand arrivals are infrequent. Also, when demand occurs, the size is highly variable changing from thousands of units per month to a few per year. Spare parts (automotive, maritime, aviation), heavy machinery, and electronics are some examples that represent an intermittent demand pattern (Ghobbar and Friend, 2003; Willemain, 2004; Syntetos and Boylan, 2005). Also, for products that are at the “*end of their lifecycle*”, this pattern typically occurs (Nagaria, 2017).

Forecasting, which is the process of predicting the by using historical and present data and trend analysis, is the primary aspect of inventory management. Though companies cannot see the future of their business, forecasting helps them to reduce the risks, understand the future conditions, and make better decisions. Good forecasting results in increased customer satisfaction since it lets companies provide the demanded SKU on time, decreased stock outs, better scheduling of production, lowered safety stock levels, and reduced obsolescence (Ahmet, 2016). However, forecasting is not always an easy task for all SKUs. Because of the infrequent and

varying demand expectations in intermittent demand pattern, companies hold large and unnecessary amount of stock on hand which may result in risk of obsolescence or they face with stock-outs (Ghobbar and Friend, 2003: 2098). In order to overcome such inventory management problems that intermittent demand cause, accurate forecasting is vital. However, due to the complex structure of these SKUs, forecasting with traditional methods is a challenging task and they require specialized methods to generate more accurate predictions. Thus, in the literature several methods are proposed.

Forecasting methods can be classified into two categories as quantitative and qualitative forecasting. Quantitative forecasting methods are objective and use statistical techniques to predict the future. These methods use historical numerical data. Oppositely, qualitative forecasting techniques are subjective, and they are based on expert judgment.

In the literature, there are several quantitative forecasting methods used for forecasting the smooth demand like exponential smoothing, moving averages, autoregressive integrated moving average and regression. However, because of the specific characteristics of the SKUs with intermittent demand pattern, forecasting is a challenging task for such products, and traditional forecasting methods do not perform well. To overcome the problems caused by the intermittent demand and generate more accurate forecasts, several forecasting methods have been proposed in the literature. A pioneering study of Croston (1972) showed that using traditional methods such as single exponential smoothing is not appropriate for intermittent demand forecasting. He proposed a technique called "*Croston's method (CR)*". After CR, many modifications on it such as "*Syntetos-Boylan Approximation (SBA)*", "*Teunter-Syntetos-Babai (TSB)*", "*Levén and Segerstedt*" modification and new methods such as Artificial Neural Networks and Bootstrapping are proposed.

Another way for forecasting the demand is by using qualitative methods. Especially when the historical data is absent, there are significant changes in the environment, or the time series are highly variable, using qualitative or judgmental forecasting can be used. These methods are generally based on the opinions of the experts. The study of Lawrence (2006) showed that when the forecaster has information about the forecasting environment gained by experience and up-to-date

information, the accuracy of judgmental forecasts increases. The jury of executive opinion, Delphi method, forecasting by analogies and scenario forecasting are some of the judgmental forecasting methods, and they are explained in detail in later sections. In practice, companies frequently use judgment in forecasting. However, this can introduce bias in the forecasts. Judgmental forecasting can be used solely to generate forecasts by adjusting statistical forecasts by using judgment or by combining independently generated statistical and judgmental forecasts.

Instead of using only one method for forecasting the demand, one can combine statistical and judgmental forecasts to achieve a higher level of accuracy. Choosing the single best method prevents the contribution of other forecasting methods. Combining different methods bring different information sources together. In the literature, it is suggested that a linear combination of two or more forecasts generated by different forecasting methods provide more accurate final forecasts than the combined forecasts individually since it averages the forecast errors. Also combining avoids the biases that may occur when adjusting statistical forecasts judgmentally or using judgmental forecast solely.

As mentioned earlier, forecasting the demand for SKUs which have non-smooth demand pattern is a challenging task. Companies may use inappropriate methods for forecasting intermittent demand. Another issue is that; statistical forecasting methods are not capable of capturing the dynamics of the business context like promotions, changes in the environment, or changes in production processes. This kind of information is included in the forecasting process by using judgmental forecasting methods. However, using only judgmental forecasting may cause bias, which will decrease the forecast accuracy. To avoid these problems and increase the accuracy level of the forecast, statistical and judgmental forecasts can be linearly combined. Even though there are some studies on judgmental adjustments of the statistical forecasts, a combination of the forecasts in intermittent demand forecasting has not truly investigated in the literature.

The main goal of this study is building a forecasting framework that generates more accurate forecasts than the constituent forecasts based on different forecast accuracy measures by combining the statistical forecasting methods with the judgmental forecasts. The study also aims to provide a framework which can easily be

applied and interpreted by the practitioners in the companies who do not have a strong statistical background. The objectives of the study are:

- to investigate demand patterns for SKUs observed when managing inventories for categorization;
- to explain the forecasting methods proposed in the literature both for smooth and non-smooth demand patterns;
- to explore judgmental forecasting methods used in the practice and combination strategies of statistical and judgmental forecasting;
- to build a framework for combining statistical and judgmental forecasts;
- to assess the performance of the proposed framework by using accuracy measures

To achieve the goal of this study, a methodological framework is proposed. This framework is implemented on two real-world datasets from automotive and small-home appliances, which include historical demand information for after-sales spare parts. Spare parts are typical examples for SKUs having intermittent demand pattern. The inventory of such SKUs should be managed effectively since they are strategic for the operations of the companies. Thus, accurate forecasting is crucial.

The first stage of the implementation process is preparing the data for forecasting by eliminating the inappropriate data from the dataset. In the second stage, the demand data is categorized into subgroups by using the categorization schemes which are proposed for the SKUs that have intermittent demand data. Categorization process groups the SKUs that have similar characteristics together so provides convenience for forecasting. The third stage is building a statistical forecast model by using various statistical forecasting methods that are proposed in the literature for SKUs with intermittent demand. Parallel to this stage, judgmental forecast model is also constructed. For generating judgmental forecasts, there are several structured methods in the literature. However, in practice, these forecasts may rely solely on the opinions of the experts in a company. After generating both statistical and judgmental forecasts, the next stage is the combination of them. In the final stage, the accuracy of statistical, judgmental, and the combination forecasts are evaluated by using suitable accuracy measures for intermittent demand pattern.

In the literature, several studies are focusing on statistical forecasting of the SKUs by using different statistical methods. Also, judgmental forecasting is widely used in practice. Judgmentally adjusting the statistical forecasts is discussed both for the SKUs with smooth and intermittent demand pattern. However, combining these forecasts for generating more accurate forecasts in the intermittent demand case is not tackled before. This study aims to contribute to the existing literature in this context.

The proposed framework for combining the statistical and judgmental forecasts is applied on two real-world datasets. Thus, the results may differ for other datasets from other companies. The statistical forecasting methods used in the application process are selected since it is stated in the literature that they perform well on the SKUs with intermittent demand pattern and they are easier to use and interpret for the practitioners. However, more sophisticated methods like Artificial Neural Networks (ANN) may perform better. One another issue is that the judgmental forecasts are not generated by using any structured methods in both companies, which may result in lower forecast accuracy. Therefore, the results that are observed in this study are the lower limits of the potential increases in forecast accuracy for the SKUs with intermittent demand.

In this study is expected that the SKUs in the datasets are independent of each other since the companies did not provide any information about the SKUs due to the anonymity. Also, it is expected that the historical data is recorded appropriately and accurately. It is assumed that the judgmental forecasts are generated by the experts, and they have knowledge about the factors that affect the demand for SKUs that they are responsible for.

Chapter 1 of this study includes the literature review for the given problem, the forecasting methods, the combination processes, and the performance measures for the generated forecasts. In Chapter 2, the methodological framework for combined forecasts is explained. Chapter 3 includes a demonstration of the proposed framework on two real-world datasets. Finally, Chapter 4 gives the findings and the discussion of the results of the implementation process, and the limitations and recommendations for future studies.

CHAPTER 1

INTERMITTENT DEMAND FORECASTING

Effective management of the inventory is vital for the success of companies from all industries as it provides a competitive advantage over the competitors. Many institutions hold different types of stock keeping units (SKU) on hand. Most of these SKUs are finished goods or work-in-process items with a normal demand pattern. However, some SKUs may have an intermittent demand pattern, which is not as smooth as the finished goods. The SKUs with intermittent pattern have several periods with no demand. When the demand occurs, it occurs randomly, and the demand size is highly variable. These SKUs may be spare parts that are used for engineering or after-sales and retailing or other items from all kinds of organizations at any level of the supply chain (Syntetos, 2001:1). These items may establish a considerable amount of stock value in the organizations. Thus, improving the inventory management of them may provide significant savings for the companies. Nevertheless, managing the inventory of such SKUs is difficult due to the intricate pattern they have. One of the most important aspects of inventory management is accurate forecasting of the demand. There are several methods proposed in the literature that are used for forecasting the demand, such as moving averages, exponential smoothing, ARIMA, etc. However, these methods perform poorly for the SKUs that have intermittent demand pattern. In order to overcome the problem in accurately forecasting the demand, various forecasting methods specialized for non-smooth demand pattern are proposed. Boylan et al. (2008: 473) mentioned that “*The selection of the appropriate forecasting methods is mostly based on the values of parameters that define the current state of the SKU*”. Considering this, some demand categorization schemes are recommended in the literature. These schemes show how to categorize the goods and which forecasting method should be used for which demand pattern.

In the following sections, further information about the categorization of demand data, several quantitative (statistical) forecasting methods both for smooth and non-smooth demand patterns, qualitative (judgmental) forecasting methods, combination of statistical and judgmental forecasts and the forecast performance

measurements that are used for measuring the accuracy of the forecasts will be explained, respectively.

1.1. DEMAND DATA CATEGORIZATION

In order to select the best performing forecasting method and inventory control policy, it is essential to categorize the demand patterns into groups or categories. Using the appropriate forecasting method may increase the availability of the SKUs, and there may be a decrease in the inventory costs. Organizations deal with a high number of SKUs. Thus, they need to evaluate these SKUs as groups instead of an individual basis. Categorization also helps SKUs to be grouped into categories with similar characteristics.

“Infrequent demand occurrences and variable demand sizes when demand occurs mean demand to be non-normal since demand per unit period or lead time demand cannot be represented by the normal distribution” (Syntetos, 2001: 365). There are differences between these non-normal patterns. To find out the characteristics of demand data, two parameters are used, which are “*the average inter-demand interval*” (p) and the “*square of the coefficient of variation*” (CV^2). p shows the regularity of demand by measuring the average number of periods between two non-zero demands and it is calculated as follows:

$$p = \frac{\sum \text{Intervals between nonzero demand periods}}{\text{Number of nonzero demand periods}} \quad (1)$$

CV measures the variation in the demand size. CV is calculated as follows:

$$CV = \frac{\text{Standard deviation of demand values}}{\text{Average demand over periods}} \quad (2)$$

In the following sections, different non-normal demand patterns will be explained, and various demand categorization schemes will be discussed.

1.1.1. Specification of Non-Normal Demand Patterns

To find out the best performing forecasting method which provides the highest forecasting accuracy level, it is essential to specify and categorize the demand patterns according to their similarities. In the following, four demand patterns that are introduced in the literature are defined briefly.

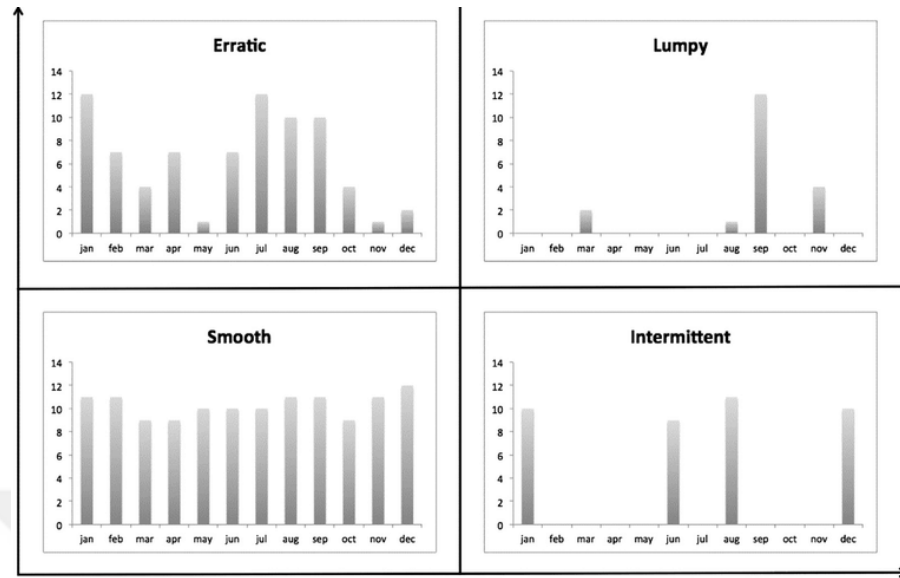
SKUs that have *smooth demand* pattern show a regular demand over time, and the non-zero demand size shows a little variation. So, these items are easy to forecast with low levels of forecast error. Majority of the finished goods have a smooth pattern.

Intermittent demand is often referred to as sporadic demand as well. Silver et al. (1998) defined this kind of demand as “*infrequent in the sense that the average time between consecutive transactions is considerably larger than the unit period, the latter being the interval of forecast updating.*” Infrequent demand occurrences may be defined as intermittent demand. In other words, intermittence refers to the occurrence of the demands but not to the demand sizes of the occurring demands. Such demand patterns can be seen in SKUs that are used as service parts in industries such as military goods, automotive and aviation. Both after-sales and maintenance spare parts form all industries also have an intermittent demand pattern since they are demanded rarely.

The irregularity of demand is not related to how often the demand occurs but with the high variability of the size of occurring demand. Silver et al. (1998) states that demand is *erratic* if the variability is large relative to the mean. In other words, the demand for SKUs that have erratic demand is varied and unstable.

In *lumpy demand*, there are many periods without any demand occurring. When the demand occurs, it is variable in size. Since both demand size and demand occurrences are highly variable, it is challenging to forecast items with lumpy demand (Syntetos, 2001: 31). In order to identify an SKU as lumpy both p and CV^2 should to be considered since the degree of lumpiness is related both with the degree of intermittence and irregularity. Forecasting the spare parts demand and inventory control of such SKUs, lumpy demand is the biggest challenge (Bartezzaghi and Kalchschmidt, 2011: 90). Figure 1 represents the four demand patterns.

Figure 1: Demand Patterns



Source: Constantino et al., 2018: 59.

To classify the demand patterns into the given groups, different categorization schemes are proposed. Following, these schemes are explained briefly.

1.2.DEMAND CATEGORIZATION SCHEMES

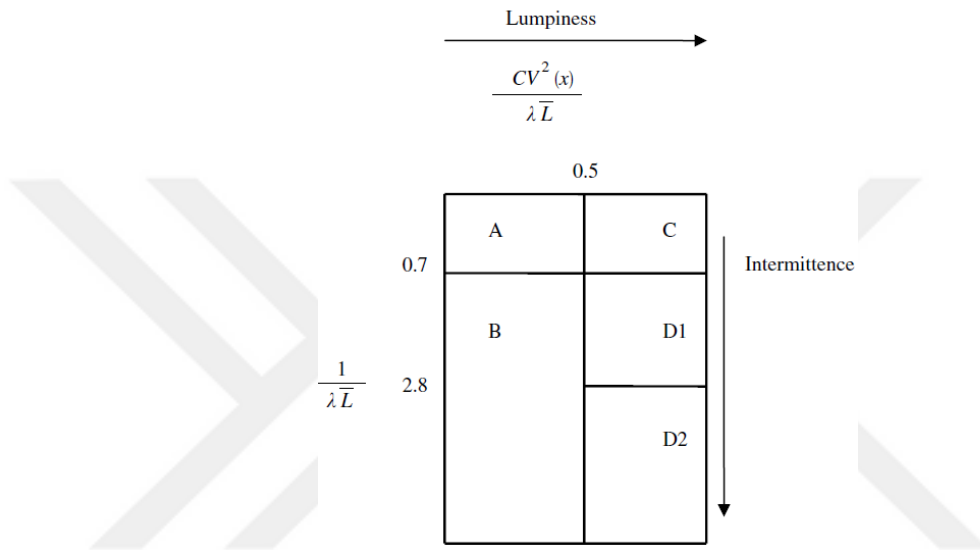
To identify the best performing forecasting methods and stock control policies for each type of demand pattern that are given in *Section 1.1.1*, Williams (1984: 941) proposed to categorize the patterns to identify the most proper forecasting methods for pattern type. The idea of variance partition is used in the study. After this study, several authors proposed different categorization schemes which are generated with different rules to determine best performing forecasting methods and inventory control parameters. The followings explain studies that are presented in the demand categorization literature.

1.2.1. Williams Categorization Scheme

This method is proposed to determine the best forecasting and inventory control methods for the resulting demand patterns based on variance partition. The

variance of the demand during lead time is divided into the “*variance of the demand sizes*”, “*transaction variability*”, and “*variance of the lead times*”. The categorization scheme classifies the demand of SKUs into smooth, slow-moving, and sporadic. Figure 2 shows the Williams’ Categorization Scheme.

Figure 2: Williams’ Categorization Scheme



Source: Syntetos et al., 2005:496.

In this matrix λ refers to the “*mean demand arrival rate*”, $\bar{L}$ to the “*mean lead time duration*” and $CV^2(x)$ to “*squared coefficient of variation of demand sizes*”. $\frac{CV^2(x)}{\lambda \bar{L}}$ is the CV^2 of demand during the lead time and, it shows the lumpiness of data. If the ratio is higher, it means that demand is lumpier. $\frac{1}{\lambda \bar{L}}$ is the number of lead times between consecutive demands. It shows the frequency of demand occurrence or the degree of intermittence. The categorization scheme divides the intermittent demand into five categories as A, B, C, D₁ and D₂. Category B is slow moving, D₁ is sporadic (lumpy), D₂ is highly sporadic (lumpy), and others are smooth. Cut-off values given in the study are chosen arbitrarily and they are based on the dataset that the researchers used. Thus, these values make sense only for the situation analyzed in the research.

1.2.2. Johnston and Boylan Categorization Scheme

Authors compared CR, which is a pioneer and specifically designed for intermittent demand, with Exponentially Weighted Moving Average (EWMA) on a theoretically generated demand data. They introduced average inter-demand interval (p). In their study, they considered many different possible conditions by using different p , smoothing constants, lead times, and demand size distributions. Results of the simulation study showed that Croston's method outperforms EWMA for p -values greater than 1.25 review periods. Johnston and Boylan (1996: 116) defined a decision rule for intermittent demand and mentioned that these SKUs have intermittent demand pattern. Authors were the first ones who formally confirmed the importance of p value as a classification parameter (van Kampen et al., 2012: 863).

1.2.3. Eaves Categorization Scheme

Eaves (2002: 59) analyzed the data from the Royal Air Force (RAF). The analysis showed that Williams' classification did not appropriately describe the demand structure. The author modified the model proposed by Williams (1984) and defined a categorization scheme based on three measures: "*transaction variability* ($\frac{c_n^2}{L}$)", "*demand size variability* ($\frac{c_z^2}{nL}$)", "*lead time variability* (c_L^2)". c_z refers to the CV of the demand size, c_n to the CV for the transaction per unit time and c_L to the CV for the lead time.

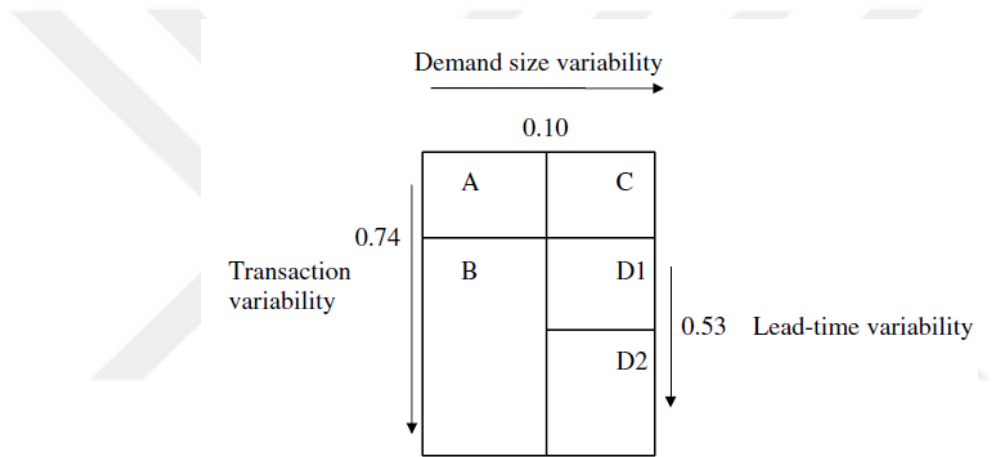
Eaves (2002: 127) also reclassified the SKUs in the dataset into five categories, "*smooth*", "*irregular*", "*slow-moving*", "*mildly intermittent*", and "*highly intermittent*". Figure 3 shows the table of categorization of demand proposed by Eaves (2002). The cut-off values are determined according to the characteristics of the dataset and sufficient subsample size considerations. Therefore, it does not have broad applicability. Eaves' categorization scheme is given in Figure 4.

Figure 3: Eaves' Categorization

<i>Lead-time demand component</i>			<i>Demand pattern classification</i>
<i>Transaction variability</i>	<i>Demand size variability</i>	<i>Lead-time variability</i>	
Low	Low	Low High	Smooth
Low	High		Irregular
High	Low		Slow moving
High	High		Mildly intermittent
High	High		Highly intermittent

Source: Eaves and Kingsman, 2004: 432.

Figure 4: Eaves' Categorization Scheme



Source: Syntetos et al., 2005: 496.

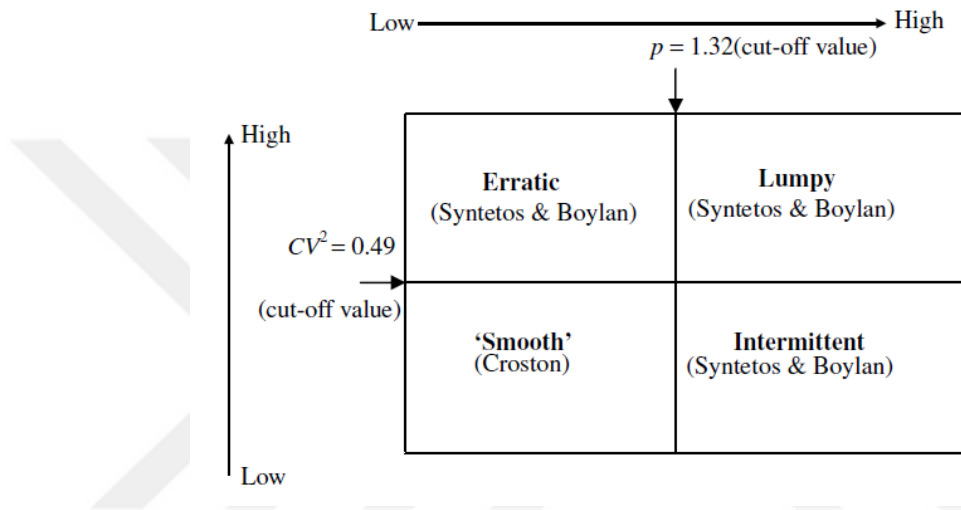
The cut-off values are determined as transaction variability is 0.74, demand size variability is 0.10 and lead time variability is 0.53 where category A is smooth, B is slow moving, C is irregular, D1 is mildly intermittent, and D2 is highly intermittent.

1.2.4. Syntetos – Boylan – Croston (SBC) Categorization Scheme

The study conducted by Syntetos et al. (2005: 499) compared EWMA, Croston's method (CR), and Syntetos-Boylan Approximation (SBA) method based on the theoretical analysis of the mean square error (MSE). The aim of this study was to determine the regions of superior performance and define the demand patterns accordingly. They developed a model of four demand categories: "erratic", "lumpy", "smooth", and "intermittent". The categorization scheme proposed by the authors is

based on p and the CV^2 of demand sizes. Comparing the theoretical MSE values of EWMA, CR and SBA, the cut-off values are determined as $p=1,32$ and $CV^2=0,49$. SBC scheme was empirically tested by using 3.000 SKUs from a company operating in automotive industry and the validity is confirmed. Figure 5 visualizes the categorization scheme recommended by the authors.

Figure 5: Syntetos-Boylan-Croston (SBC) Categorization Scheme



Source: Boylan et al., 2008:476.

Table 1 represents the cut-off values; p , and CV^2 of each demand pattern and the suggested forecasting methods for each of them. Syntetos and Boylan categorization suggest that forecasting the smooth demand with CR gives the highest accuracy while erratic, lumpy, and intermittent demands are forecasted best by using SBA. This study contributes the identification of the CV^2 as a new categorization parameter for demand forecasting purposes.

Table 1: Suggested Forecasting Methods for Each Demand Category

Condition	Demand Pattern	Method
$p \leq 1.32; CV^2 \leq 0.49$	Smooth	Croston
$p > 1.32; CV^2 \leq 0.49$	Intermittent	SBA
$p \leq 1.32; CV^2 > 0.49$	Erratic	SBA
$p > 1.32; CV^2 > 0.49$	Lumpy	SBA

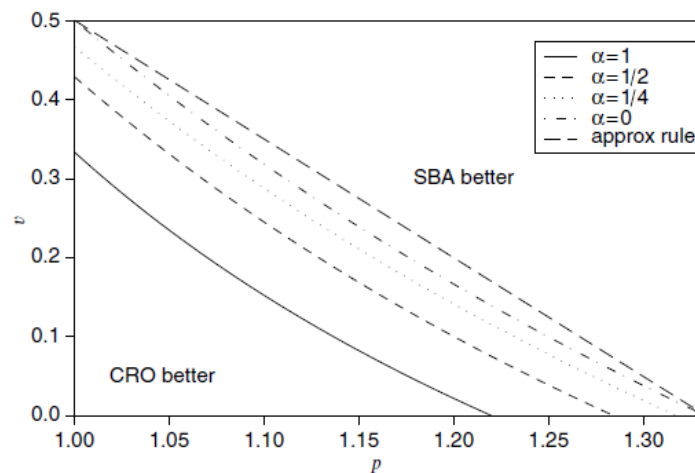
1.2.5. Kostenko – Hyndman (KH) Categorization Scheme

KH scheme is developed as an extension of the SBC scheme by Kostenko and Hyndman (2006: 1257). Authors introduced a more accurate and simpler rule than SBC and suggested using SBA method in smooth pattern whenever:

$$CV^2 > 2 - (3/2)p \quad (3)$$

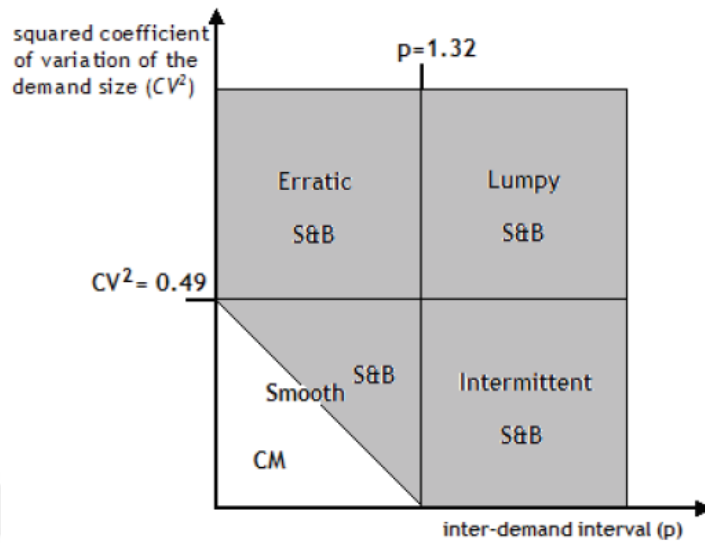
Heinecke et al., (2013: 458) mentioned that “*The bottom-left quadrant of the SBC categorization scheme is the parameter space area where the application of the proposed approaches potentially results in SKUs being assigned to different categories*”. According to KH categorization scheme, for SKUs with CV^2 value of 0.4 and p value of 1.25 SBA is used while the SBC categorization scheme suggests using CR for the same SKUs. Figure 6 shows the disputed parameter space where the vertical axis is the CV^2 and the horizontal axis shows the p , and Figure 7 shows the categorization of demand pattern based on KH categorization.

Figure 6: Disputed Parameter Space of Kostenko - Hyndman Categorization



Source: Kostenko and Hyndman, 2006: 1257.

Figure 7: Kostenko-Hyndman Categorization Scheme



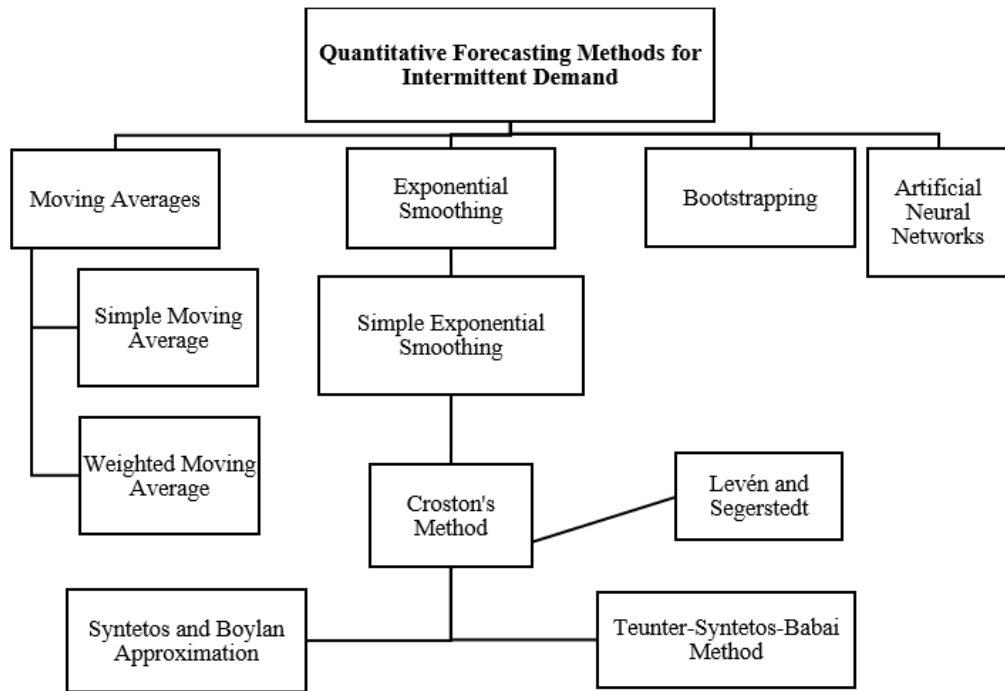
Source: Babiloni et al., 2010:118.

1.3. QUANTITATIVE FORECASTING METHODS

Forecasting is the process of predicting a future activity, event, or occurrence based on historical information. This process can be classified into two groups as qualitative and quantitative forecasting. Qualitative methods are subjective, and they are based on the judgment of experts, and it is appropriate to use such methods when there is no numerical data available and the forecasting environment or the time series are highly variable. Quantitative (statistical) methods are based on mathematical models. It is appropriate to use such methods when numerical past data is available. These methods also assume that the patterns in the past data will continue into the future. Quantitative forecasting methods are divided into two: “*time-series models*” which includes the historical data into the prediction of the future and “*causal models (regression)*” that develops the statistical relationship between variables.

This section of the study covers the widely used statistical time-series models for forecasting the demand for SKUs with intermittent demand pattern. Figure 8 represents the classification of the statistical forecasting methods explained in the following sections that are used for forecasting the SKUs that have intermittent demand pattern.

Figure 8: Statistical Forecasting Methods for Intermittent Demand



1.3.1. Simple Moving Average (SMA)

In the SMA method, by averaging the last selected number of periods the forecasts are calculated. After the occurrence of a new demand, the newest value is added to the computation of new mean while the oldest observation is dropped. “A *small number of periods makes forecasting sensitive while a large number of periods results with poorer response to a step change*” (Hanke and Wichern, 2014: 65). In moving average, equal weights are assigned to each observation. This method is not appropriate if there is a trend or seasonality.

For a moving average of order k , $MA(k)$, the SMA can be calculated as follows:

$$Y_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k} \quad (4)$$

where

Y_{t+1} = the forecast value for the next period

Y_t = the actual value at period t

k = the number of terms in the moving average

SMA is one of the most frequently used methods for forecasting intermittent demand. Sani and Kingsman (1997: 711) compared SMA, single exponential smoothing, and Croston's method by using cost and service level as performance measurements. Their results showed that SMA outperforms other methods.

1.3.2. Weighted Moving Average (WMA)

SMA method weights all terms equally. If there is a trend presented, it is appropriate to use weights to give higher importance on recent values. Thus, forecasting techniques to be more responsive to changes. WMA weights the last actual demand higher. Thus, it recovers some of the disadvantages caused by the SMA method. The formulation for WMA is given in Equation 5.

$$Y_{t+1} = \frac{\sum((Weight\ for\ period\ n)(Demand\ in\ period\ n))}{\sum Weights} \quad (5)$$

When choosing the weights, there is no formula. Therefore, some experience is required for determining these weights. Heizer and Render (2014: 146) states that *“if the last period is weighted too heavily, the forecast may reflect a large unusual change in demand pattern too quickly.”* If all terms have the same weight, WMA gives the same forecasts with SMA.

1.3.3. Single Exponential Smoothing

In practice Simple exponential smoothing (SES) is a commonly used due to its straightforwardness and robustness. It is a parameter-based method. It provides an EWMA of all observed values. This method aims to estimate the current level to use

when forecasting the future values. Using SES is suitable for data with no trend or seasonal pattern.

SES revises an estimate by using more recent experiences. The calculations require a predetermined parameter; smoothing constant “ α .” Most recent observation receives the highest weight and the older observations receive less weight. Smoothing constant is a value between 0 and 1, and it is determined judgmentally and based on the data’s characteristics. Silver et al. (1998) suggest to choose a smoothing constant value between 0.1 and 0.3 if forecasting is done monthly.

The general formulation of exponential smoothing is:

$$Y_{t+1} = \alpha X_t + (1 - \alpha)Y_t \quad (6)$$

where

Y_{t+1} = the new smoothed value or the forecast value for the next period

α = the smoothing constant ($0 < \alpha < 1$)

X_t = the new observation or the actual value of the series in period t

Y_t = the old smoothed value of the forecast for period t

1.3.4. Croston’s Method (CR)

While forecasting the demand, SES is a widely used method by the practitioners. However, for forecasting the intermittent demand, Croston (1972: 298) proved that using SES is inadequate. In order to overcome this problem, Croston (1972) proposed a new forecasting method, which separately estimates the size of non-zero demand and inter-arrival time between non-zero demands by using SES. The algorithm for CR estimates is given in Table 2.

Croston (1972: 294) assumes that demand occurs as a Bernoulli process; intervals between demands are independent and identically distributed. Demand sizes are also assumed to be independent and normally distributed. Rao (1973: 639) made some modifications on several expressions in Croston’s paper. The conclusions or the forecasting procedure did not affect by these corrections, and forecasting performance did not change.

Table 2: Croston Method's Algorithm

If $A_t = 0$	If $A_t \neq 0$
$Z_t = Z_{t-1}$	$Z_t = \alpha A_t + (1 - \alpha) Z_{t-1}$
$P_t = P_{t-1}$	$P_t = \alpha Q_t + (1 - \alpha) P_{t-1}$
$Q = Q + 1$	$Q = 1$

According to the CR, estimates of size and interval provides an estimate of the mean demand per period;

$$Y_t = \frac{Z_t}{P_t} \quad (7)$$

where

Z_t = estimate of the mean size of a nonzero demand

P_t = estimate of the mean interval between nonzero demands

A_t = the observed value of the demand

Y_t = estimate of mean demand per period

Q = time interval since last demand.

α = smoothing constant between 0 and 1.

A study conducted by Willemain et al. (1994: 537) comparing CR with SES showed that CR is superior to SES, although real data analysis in some cases presents a more modest benefit. A later study by Johnston and Boylan (1996: 117) obtained similar results who stated that the CR method always performs better than SES when p is greater than 1.25 periods. Ghobbar and Friend (2003: 2101) compared 13 different methods, including the ones used by aviation companies. These methods are additive winter, multiplicative winter, seasonal regression, component service life, weighted calculation of demand rates, weighted regression demand forecasters, CR, SES, EWMA, trend-adjusted exponential smoothing, weighted moving average, double exponential smoothing (Holt), and adaptive-response-rate single exponential smoothing. The results showed the superiority of the weighted moving average, Holt, and CR for intermittent demand. For the manufacturers dealing with intermittent demand, Croston's concept has been claimed to be a great value, and it is available in ERP type solution (Teunter et al., 2011: 606).

To generate a method which is appropriate to use both for slow- and fast-moving items, Levén, and Segerstedt (2004: 363) proposed a modification to CR. Levén and Segerstedt (LS_t) model is given as follows:

$$LS_t = \alpha \frac{Z_t}{P_t} + (1 - \alpha)LS_{t-1} \quad (8)$$

where

Z_t = estimate of the mean size of a nonzero demand

P_t = estimate of the mean interval between nonzero demands

α = smoothing constant between 0 and 1

However, though the authors mentioned that this method is unbiased, it is found to be biased by Boylan and Syntetos (2007: 513). Also, LS has a larger mean squared error value than CR.

1.3.5. Syntetos and Boylan Approximation (SBA)

Syntetos and Boylan (2001: 460) found out that there is an error in mathematical derivation of the expected estimate of demand per period which causes bias and low forecast accuracy and reported that the original method of Croston is biased due to an error in the mathematical derivation process of CR. Thus, they proposed a revision, Syntetos-Boylan Approximation (SBA), which approximately corrects the error in CR's demand estimates.

CR estimates the average demand size ($E(Z_t)=\mu$) and the mean interval between non-zero demands ($E(P_t)=p$) after demand occurs. The forecast for the next period according to CR is given in Equation 9.

$$E(Y_t) = E\left(\frac{Z_t}{P_t}\right) = \frac{E(Z_t)}{E(P_t)} = \frac{\mu}{p} \quad (9)$$

The bias in CR arises occurs since it is assumed that demand size and demand interval estimators are independent

$$E\left(\frac{Z_t}{P_t}\right) = E(Z_t) \left(\frac{1}{P_t}\right) \quad (10)$$

but

$$E\left(\frac{1}{P_t}\right) \neq \frac{1}{E(P_t)} \quad (11)$$

Syntetos and Boylan showed that the expected demand per period is not μ/p but

$$E\left(\frac{Z_t}{P_t}\right) \approx \frac{\mu}{P} \cdot \left(1 + \frac{\alpha}{2 - \alpha} \cdot \frac{(p - 1)}{p}\right) \quad (12)$$

For $\alpha=1$:

$$E\left(\frac{Z_t}{P_t}\right) = \mu \left[-\frac{1}{p - 1} \ln\left(\frac{1}{p}\right) \right] \quad (13)$$

Ignoring the term $(p-1)/p$ in Equation 12, the authors proposed a new estimator that deals with the bias of CR. Where α is the smoothing constant value used for updating the inter-demand intervals, a new estimator of mean demand is given as:

$$Y_t = \left(1 - \frac{\alpha}{2}\right) \frac{Z_t}{P_t} \quad (14)$$

Eaves and Kingsman (2004: 436) assessed that SBA is the best forecasting method for spare parts inventory within SBA, SES, CR, and moving average. They also showed that this method allows the lowest stock-holdings across all demand patterns, including smooth, irregular, slow-moving, and intermittent. To confirm the performance of SBA method, Syntetos and Boylan (2005: 312) compared the

forecasting methods, including CR and SBA in their study. The simulation exercise covering 3.000 fast intermittent SKUs from the automotive industry showed that SBA is the most accurate forecasting method within SMA, SES, CR, and SBA. Another study held by Syntetos et al. (2005: 500) that proposed a demand categorization scheme to guide the selection of forecasting methods suggests that some demand categories (smooth) give better results with CR while others (erratic, lumpy and intermittent) give better results with SBA. Gutierrez et al. (2008: 415) also stated that SBA performs better than the CR and SES.

1.3.6. Teunter-Syntetos-Boylan (TSB) Method

Teunter et al. (2011: 606) mentioned that since Z_t and P_t are updated only in periods that a demand occurs, CR is not suitable to deal with obsolescence. To overcome this issue, they proposed a new modification of CR. Different from CR and SBA, Teunter-Syntetos-Boylan (TSB) method estimates the probability of non-zero demand rather than interval size. Instead of updating the estimates only when demand occurs, TSB updates the estimates every period. Since the estimates are updated in every period, TSB is unbiased.

The TSB model is given in Table 3.

Table 3: TSB Forecasting Model

If $A_t = 0$	If $A_t \neq 0$
$P_t = (1-\beta) P_{t-1}$	$P_t = \beta + (1-\beta) P_{t-1}$
$Z_t = Z_{t-1}$	$Z_t = \alpha X_t + (1-\alpha) Z_{t-1}$
$Y_t = P_t Z_t$	$Y_t = P_t Z_t$

where

A_t = demand for an item at time t ,

P_t = estimate of the probability of a demand occurrence in period t ,

Z_t = estimate of mean demand size at the in period t ,

X_t = actual demand size in period t

Y_t = estimate of mean demand per period at the end of period t for period $t + 1$

α, β smoothing constants ($0 \leq \alpha, \beta \leq 1$)

In TSB, the probability of demand is updated more often than the demand size, so different smoothing constant values are used. Therefore, the TSB method achieves high flexibility. The essential advantage of this method is that updating the forecasts at each period (whether demand occurs or not) allows the estimate P_t to approach zero if there is a long run of periods without a demand. TSB can be used to estimate the risk of obsolescence, and it allows decisions about whether to continue to stock an item or not. A threshold for P may be set. If this level is exceeded then management can decide that the product is obsolete (Waller, 2015: 2).

1.3.7. Bootstrapping

When spare parts have an intermittent demand, forecasting is a challenging task because of the particular nature of their demand pattern. There are two forecasting approaches used in the literature to deal with such demand patterns. These are parametric and non-parametric (bootstrapping) approaches. Parametric methods are based on the assumption that the historical demand of a SKU follows a particular probability distribution. Then, point forecasts are estimated by using past data. Some researchers support the idea that instead of point forecast estimations, one can forecast the entire distribution of all possible lead-time demand values, often reporting average demand per period (Hasni et al., 2017: 1). To do so, they use a non-parametric approach, which is firstly proposed by Efron (1979). Non-parametric approaches to forecasting permit a reconstruction of the empirical distribution of the data. Thus, they are making distributional assumptions redundant (Syntetos et al., 2015: 1748). It takes random samples from the original data. These samples are used to build up a histogram of the lead-time demand distribution.

An initial study of Efron (1979) was followed by numerous studies developed in the context of forecasting intermittent demand. Alternative bootstrapping methods may be found in (Snyder, 2002: 11; Willemain et al., 2004: 380 (hereafter WSS); Porras and Dekker, 2008; Teunter and Duncan, 2009: 324; and Viswanathan and Zhou, 2008 (hereafter VZ))

Hua et al. (2007: 53) suggests that when historical data of spare part demand are limited, using bootstrapping is a powerful tool for estimating the lead-time demand

distribution for spare parts. Also, as the data become more erratic, the actual demand size distribution may not conform to any standard theoretical distribution. In such a situation, it may be beneficial to use bootstrap approaches (Syntetos et al., 2015: 1747). The most known bootstrapping method, WSS, for SKUs with intermittent demand pattern, is proposed by Willemain et al. (2004). This method is tested by the authors, comparing its performance with SES and CR. The results showed that non-parametric bootstrapping method outperforms the traditional parametric methods. However, Gardner and Koehler (2005: 1) stated that the authors did not use the correct lead time demand distribution for SES and CR. Also, the authors did not consider any of the proposed modifications of CR. To overcome these drawbacks Syntetos et al. (2015: 1750) published a study that compares WSS, SES, CR, and SBA by using 7.000 SKUs from jewelry data which have moderately irregular series with short lead times and electronics data which is more intermittent and have longer lead times. Results showed that, in cases of less regular demand series and shorter lead times, the WSS method outperforms the parametric methods. However, if the irregularity of the demand increases and the lead times get longer, SBA performs the best while WSS is worst among WSS, SES, CR, and SBA. Syntetos et al. (2015: 1751) also stated that although bootstrapping does have some advantages, whether or not the WSS method is worth the considerable added complexity is questionable. The performance of the VZ method is measured by comparing VZ with SBA and WSS in the study of Zhou and Viswanathan (2011: 484). Results showed that if the level of intermittency is moderate, VZ outperforms SBA and WSS for long lead times and long histories. If the level of intermittency is higher, SBA outperforms VZ for short histories and short lead times.

1.3.8. Artificial Neural Networks (ANN)

Recently neural networks gained the attention of researchers due to its good performance in accurately forecasting the demand and its capability to solve decision-making problems. ANN are computing models inspired by the human brain process information, which are made of simple and highly interconnected processing elements (neurons) and used for information processing and pattern identification. Due to their restrictive functional forms, traditional statistical time-series methods may fit poorly

to the underlying data generating process while flexible ANN models learn the process of data generating from the data, without needing human intervention (Kourentzes, 2013: 200). Since ANN can learn a non-linear function from samples without needing any distribution assumption, they have attracted practitioners and researchers recently (Lolli, et al., 2017:116)

ANN have some advantages over traditional forecasting methods. They can approximate any continuous function to the desired accuracy. Thus, they are more flexible than traditional methods. They can catch both linear and nonlinear patterns in the data while traditional time series are based on the assumption that time series are generated from a linear process. Also, ANNs are data-driven self-adaptive methods with an assumption-free nature (Zhang et al., 1998: 35). Lastly, ANN allow for an interaction between the demand size and the demand intervals between non-zero demands. Identifying this interaction from the data would provide useful information (Kourentzes, 2013: 199).

Gutierrez et al. (2008: 418) compared the NN method for time series with lumpy demand to SES, CR, and SBA. On a set of 24-time series, they found out that the NN method outperforms the traditional methods with different smoothing parameters. Pour et al. (2008: 1033) compared the performance of the ANN model they proposed, hybrid method (HPM), with different ANN models like MLP, generalized regression neural network (GRNN) and Elman recurrent neural network (RNN) and SBA on 30-time series. Results of this study showed that the method proposed by the authors outperformed the others. Kourentzes (2013: 202), proposed two ANNs, which are NN-Dual and NN-Rate, namely. In his study, he compared the proposed ANN methods with CR and its variants by using both forecasting accuracy and inventory metrics. The results of forecasting accuracy and inventory metrics are conflicting; ANNs achieved high inventory cost savings while their forecast accuracies were poor.

1.4. QUALITATIVE FORECASTING METHODS

In real world applications, judgmental forecasting is commonly used. Especially when the historical data is absent, there are significant changes in

environment or time-series are highly variable, it may play a crucial role in the forecasting process. If the judgment of management is not included in forecasting process, serious problems may occur. For example, Nike's experiment with forecasting software which cost \$400 million was unsuccessful and caused massive inventory write-offs due to the system's inaccuracy and lack of management input (Worthen, 2003).

Qualitative or judgmental forecasting methods are mostly based on the knowledge of experts. Research of Lawrence et al. (2006: 494) showed that when the forecaster has valuable contextual knowledge, knowledge of forecaster gained by experience about the forecasting environment, and up-to-date information, the accuracy of the judgmental forecasts increases. Study of Sanders and Ritzman (1992: 50) also shows that contextual knowledge has a vital role in judgmental forecasting while technical knowledge (knowledge about data analysis and formal forecasting procedures) has a little point. Another statement of the authors indicates that as the variability of the data, which is measured by the CV of the time-series, increases the judgmental forecasts generated by the practitioners become preferable.

Using judgmental forecasting techniques provides both advantages and disadvantages. One of the significant advantages of the judgmental forecasting methods is that they can identify the changes more quickly and interpret better their possible effects on the future. They do not require any mathematical background of the forecaster. The subjective methods are capable of capturing the forces that statistical methods cannot deal with the expertise of the forecasters. Besides the advantages, judgmental methods have disadvantages as well. The most critical disadvantage of judgmental forecasts is since they are subjective, they are almost always biased. They may not be consistently accurate over time. Also, it may take years of experience to learn how to convert judgment into a good forecast (Wilson and Keating, 2008: 20).

There are three settings that judgmental forecasts are used. First, sometimes, forecasters cannot use statistical forecasts since there is no available data. If this is the case, judgmental forecasting is the only available approach. In the second situation, there is available data, and using this data; forecaster generates statistical forecasts. Later these forecasts are adjusted judgmentally by using contextual knowledge.

Moreover, the last setting is the combination of judgmental and statistical forecasts, which are generated independently. This section of the study focuses on the judgmental forecasting methods.

1.4.1. The Jury of Executive Opinion

The jury of the executive opinion is one of the most common methods used in judgmental forecasting. This qualitative forecasting model relies on the opinions and expertise of high-level managers or experts. The opinions of related managers and executives within an organization who have the best insights about the firm's future situation are collected. The final group forecast becomes the forecast for a product as a blend of opinions of these executives from different functional areas such as sales, marketing, and production. Attendance of the executives from different areas provides a breadth of opinion to the forecasting process. The opinions of the executives can either be collected by using personal interviews or group meetings. Though in group meeting there is a chance of discussion of various viewpoints; there may also occur the domination of an expert with a strong personality which may affect the final forecast (Wilson and Keating, 2008:18). In the jury of executive opinion method, companies can also use statistical models to help with the analysis (Wright, 2013).

1.4.2. Delphi Method

The Delphi method is similar to the jury of executive opinion. This method is developed in the 1950s at the Rand Corporation by Olaf Helmer and Norman Dalkey. Assuming that a group's forecast is more accurate than an individual's forecast, the Delphi method aims to construct consensus forecasts from a group of experts in a structured, iterative manner (Hyndman and Athanasopoulos, 2018: 111). There is a facilitator to implement and manage the process. The steps of the Delphi method are as following:

1. Assemble a panel of experts. The number of experts contributing is between 5 and 20 with diverse expertise.
2. Set the forecasting tasks and deliver them to the experts.

3. Experts return their initial forecasts and justifications. Summarize the initial forecasts to provide feedback to the experts.
4. Deliver the feedback to forecasters so they can review their forecasts based on the views of others. This process continues through two-three rounds until the experts reach a satisfactory level.
5. By aggregating the forecasts of experts, construct the final forecasts.

Any Delphi procedure has four key features, which are anonymity, iteration, controlled feedback, and aggregation to provide a group response.

In the Delphi method, anonymity is one of the primary advantages. During the processes, participating experts remain anonymous at all times. Anonymity can reduce the effects of dominant individuals, which is generally a problem while using group-based processes used to collect and synthesize the information (Dalkey, 1969:16). They express their opinions privately without having social and political pressure. Geographic dispersion of the subjects and use of electronic communication facilitates the confidentiality and minimize the problems related to group dynamics such as manipulation or duress to accept a viewpoint (Hsu and Sandford, 2007: 2).

The process of experts submitting forecast, receiving feedback, and reviewing their forecasts according to the feedback is iterated until a satisfactory level of consensus between the experts is reached. After each iteration, the opinions of each expert are summarized, and each group member is informed about other anonymous members of the group. With the iteration, experts have the opportunity to change their opinions without losing face in the eyes of the remaining group members (Hanke and Wichern, 2014: 439). Usually, two or three iterations are sufficient, since the experts may drop out as the number of iterations increase (Hyndman and Athanasopoulos, 2018: 113).

Controlled feedback in the Delphi method is designed to reduce the effect of noise. Noise refers to the communication that occurs in a group process which both alters the data and deals with other interests instead of focusing on problem-solving. With controlled feedback, an organized summary of the prior iteration is distributed to the experts. Using the feedback, experts have an opportunity to generate additional insights. After multiple iterations, experts are expected to be more problem-solving

oriented, to offer their opinions more insightfully, and to minimize the effect of noise (Hsu and Sandford, 2007: 2).

Group response is usually generated by giving equal weight to all of the experts' forecasts.

1.4.3. Forecasting by Analogies

In forecasting, it is natural to use an analogy. It contains information about in the past how people behaved in a similar situation. Forecaster identifies a pattern that happened in the past and applies the same pattern to a new issue. It is expected analogies to be useful in forecasting decisions in conflict situations since they provide essential information for situations that are difficult to forecast (Green and Armstrong, 2007: 365).

An example of using analogies is about the task of designing a new judgmental decision-making curriculum for high schools in Israel. In the group, each member was asked to forecast how long would it take to complete the curriculum. Their responses were between 18-30 months. One of the members of the group, who was an expert in curriculum design, was asked to consider analogous curricula developments around the world. The expert mentioned that 40% of the analogous groups he considered were not able to finish the task while for the rest it took between 7 to 10 years to complete. The project in Israel was completed in 8 years (Kahneman and Lovallo, 1993: 24).

As mentioned before, people frequently prefer to use analogies when they forecast. However, they do it in an unstructured way. Armstrong (1985:124) shows that using structured judgmental forecasting methods provide higher accuracy than using unstructured ones. Green and Armstrong (2007: 367) proposed a structured approach for forecasting with analogies, which may encourage experts to consider more information on analogies and to process it effectively. First, an administrator is appointed. Structured analogies have five steps:

1. The administrator describes the target situation briefly and accurately
2. The administrator chooses experts who have knowledge about similar situations to the target situation. Having at least five experts is suggested.

3. Experts identify and describe as many analogies as they can. Based on each of these analogies, forecasts are generated.
4. Experts list similarities and differences between their analogies and target situation. Later they rate the similarity of each analogy to the target situation on a scale.
5. From experts' analogies, administrator drives the forecasts by using a set rule. This rule can be weighted average where the weights can be guided by the ranking scores of each analogy by the experts (Hyndman and Athanasopoulos, 2018: 116).

Green and Armstrong (2007: 372) stated that experienced forecasters on analogies who have more than two analogies generate the most accurate forecasts.

1.4.4. Scenario Forecasting

Scenario forecasting approach aims to generate forecasts based on scenarios that designed to be feasible but not necessarily most likely. These scenarios either portray a snap-shot picture of a future state or a plausible evolution from the present onwards (Bunn and Salo, 1993: 291). When generating the scenarios, all possible factors, their impacts, interactions between them, and forecasting targets are considered. Working with the scenarios help managers to understand the role of uncertainties better. Also, some extremes like the best, middle, and the worst-case scenarios can be identified when building forecasts based on analogies. Documenting these extremes can lead to early contingency planning (Hyndman and Athanasopoulos, 2018: 117). In capital intensive industries like oil companies, vehicle manufacturers, and electric suppliers which have long planning horizon, scenario techniques found to be more popular.

CHAPTER 2

COMBINING QUANTITATIVE AND QUALITATIVE FORECASTING

An example that briefly explains the importance of combining forecasts is given by Armstrong (2001: 417) as following: *“Imagine that there is a murder that has to solved, but there is limited budget. Would it be better to spend all our money on doing one task well (for example DNA test) or should we spread money over many small tasks like ballistic tests, the murder weapon and checking alibis? Standard procedure in matters of life and death is combining evidence from different approaches”*. With this example, it can be understood better why one should not try to choose single best method and find out other methods as well, which may have something to contribute to improving accuracy. Combination method appose different information sources, so the information upon which the forecast is based (Lawrence et al., 2006: 508)

In some situations, more than one forecasting technique is applied to the same event. Mostly, forecasters choose the technique which provides the lowest forecast error. However, if the forecasting techniques use different information or different methods, discarding the other techniques by choosing the one that has the best performance may cause the loss of some valuable information. Combination of forecasts obtained from more than one forecasting technique into a single better forecast prevents valuable information loss (Wilson and Keating, 2008: 404). After many studies on combining forecasts, literature suggests combining two or more forecasts linearly to generate final forecasts that are much better than the individual forecasts that are being combined.

Judgmental adjustments are affected by the judgment’s negative effects. They can introduce bias. To avoid judgment’s adverse effects, forecaster generally prefers to combine judgmental and statistical forecasts, instead of using judgmental adjustments and forecasts. Clemen (1989: 559) states that combining the forecasts helps to eliminate biases and lead to increased accuracy since it averages the forecasting errors.

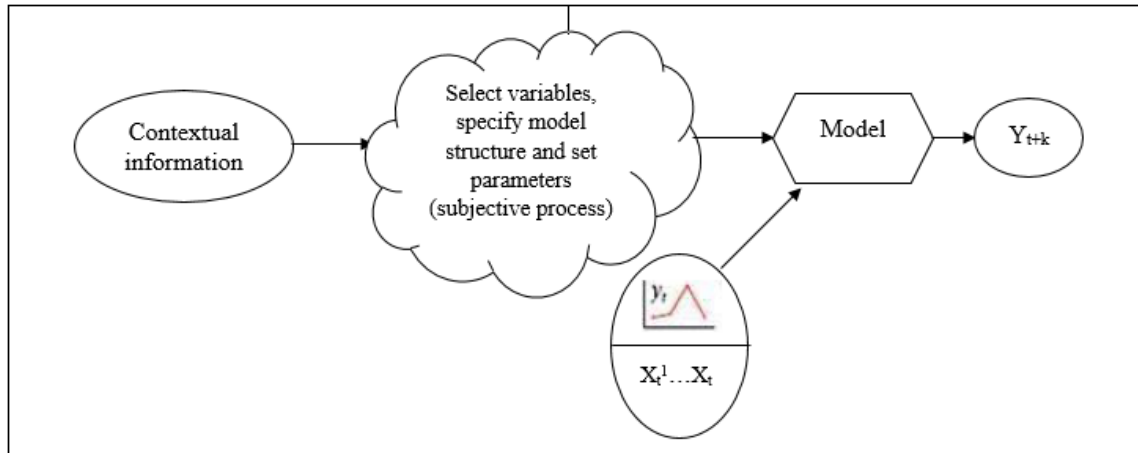
The accuracy of the overall forecast improves as the component forecasts include useful and independent information. As the data and methods differ, the overall accuracy improves. The combined forecasts should be absent of biases or forecast errors should be negatively correlated so that they can cancel each other. In practice, this is rare. Because of this reason, the practitioner may combine the forecasts that have uncorrelated forecast errors. However, in real life, forecast errors are almost always positively correlated (Armstrong, 2001: 419).

In some cases, the forecaster may not be able to decide which forecasting method is the most accurate. In such a state, combining forecasts are expected to be useful. Each method in the combination contributes some information that helps the accuracy to improve. This makes combining better than using one single best method. Combining forecasts is also useful if there is uncertainty in forecasting situation. In these situations, combining reduces the error. Armstrong (2001: 426) defined the uncertain conditions as *“Conditions can be judged as uncertain based on unexplained variations in the past or on expected volatile changes in the future”*.

2.1. USING JUDGMENT IN MODEL BUILDING PROCESS

While building a statistical model, using the input of judgment is required (Bunn and Wright, 1991: 509). In model building, judgment is used for selecting and developing the quantitative forecast, selecting the variables, setting the parameters, and specifying the structure of the model. Judgment here provides information that statistics cannot determine. This method is considered as the least subject to the harmful effects of judgmental biases, most effective and most objective (Sanders and Ritzman, 2004: 521). Figure 9 represents the model building process.

Figure 9: Model Building Process



Source: Webby and O'Connor, 1996: 100

Bunn and Wright (1991: 509) suggested while building statistical models, judgmental inputs are required in four general areas which are “*the selection of the variables*”, “*model specification*”, “*parameter estimation*”, and “*data analysis*”.

Variable selection and data analysis. Many variables can be related to the forecast series. It is possible to determine important causal variables by using statistical diagnosis. Techniques like multiple regression or bootstrapping make it possible to automate the tasks of model specification and parameter estimation, but still, judgment is required for identifying the variables (Webby and O'Connor, 1996: 100). They stated that “Indeed, even in the statistical and econometric literature, conventional wisdom is that variable selection should be essentially judgmental and not automated.” Though it is important to use statistical diagnostics identification of key causal variables, judgment should be used for selecting the key variables. For example, the bootstrapping method is vulnerable when unaccounted broken-leg cues are affecting the data although this method is successful when the variables have linear relationship and variables are well designed. The term broken-leg is defined as “the sudden and major events that cannot be captured in a formal model” (Meehl, 1957: 270). Using judgment considering these unexpected events makes the modeling process more effective.

Model specification. Is the process of determining the essential predictor variables and also involving these variables into regression function in the right way (Hanke and Wichern, 2014: 303). Turner (1990) states that “in practice, judgmental

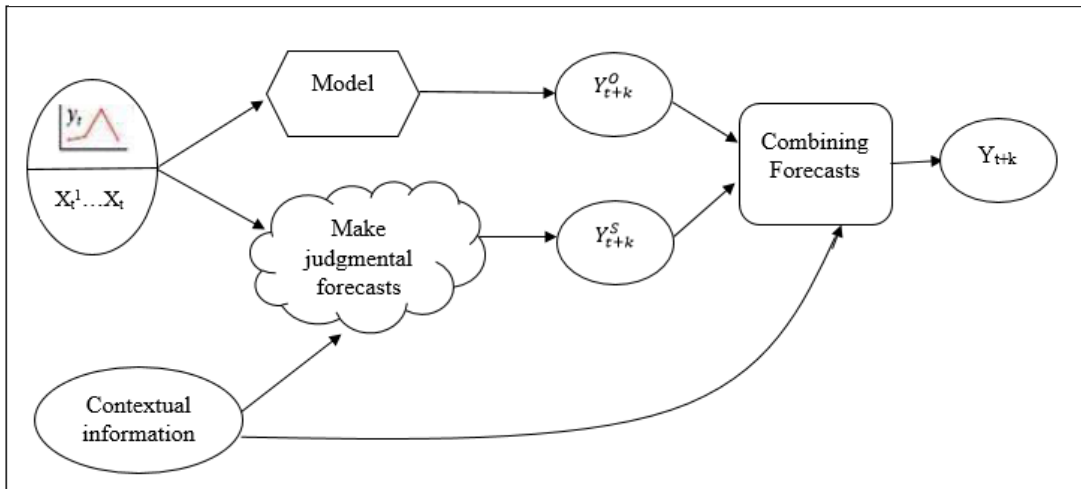
adjustments are frequently made on econometric models to have more information about the model”.

Parameter estimation. Harvey and Durbin (1986) demonstrated how to set the parameters which are related to a specific event, so the effect of that event diminished over time. Setting the parameters is not an easy task cognitively since in parameter estimation, it is crucial to understand the underlying model as well as the influence of the special events. (Webby and O'Connor, 1996: 101). By using judgments, special events can be identified, and the effects of them will be automatically assessed on the time series.

2.2. COMBINING FORECASTS

Because of the market complexity, using the results gained from a single forecasting technique are not sufficient for decision making (Mancuso and Werner, 2013: 249). Constituent forecasts have different information that could be included in the forecasting process. Thus, the overall accuracy of the forecasts increases. Combination procedure can range from mechanical methods such as taking a simple or weighted average of the constituent forecast, to using judgment to determine how to forecasts should be combined (Lawrence, et al., 2006: 508). Using an equal arithmetic average of the individual methods is the simplest approach for combining. Equal weighing provides improvements in accuracy for many forecasts (Clemen, 1989: 559). More complex weighting schemes may also be successful. For instance, if the practitioner believes that there are some issues that may cause one constituent forecasting method to be perform better than the other, s/he can weigh this method heavier. However, this complexity and the time waste it causes would make combining forecasts less desirable for an organization (Sanders and Ritzman, 2004: 521). The combination process of objective and subjective forecasts is shown in Figure 10.

Figure 10: Combination of Objective and Subjective Forecasts



Source: Webby and O'Connor, 1996: 100.

Based on historical information, a statistical model is generated at the same time as judgmental forecasts are executed. Contextual information is used for yielding statistical and objective forecasts. Subjective and objective forecasts are combined, and the final forecast is generated based on the contextual information. Contextual information refers to any kind of information which are “non-time series information” in the form of knowledge which is related to the event of forecast. If there is contextual information, judgmental adjustment of statistical forecasts may be beneficial since the judge could provide information to the forecasts that the statistical model could not (Webby and O'Connor, 1996: 103). This information includes up to date knowledge about the current changes, experience, and events which occur in forecasting in the that can affect the forecasts. For instance, promotions, holidays, advertising campaigns, changes in technology, economic fluctuations, etc.

. Caniato et al. (2009) conducted a case study which combines the forecasts in a cement company. The overall accuracy level o the forecast has increased. The study of Song et. al (2013: 307) aiming to forecast tourism demand showed that combined forecasts produce more accurate results. Graefe et al. (2014: 50) conducted a research on forecasting the elections by combining the forecasts. The results showed that the combination reduces the errors that are generated by constiuent forecasts. Also, authors stated that averaging the forecasts by using equal weights generate better results without adding complexity. In intermittent demand context, there are some

studies that focuses on adjusting the statistical forecast by using judgmental adjustments (Syntetos et al., 2009; Davydenko and Fildes, 2013). In both studies, adjustment process increased the accuracy. However, adjustments are subject to biases which may harm the forecast accuracy (Eroğlu and Croxton, 2010: 117). In order to avoid such biases, mechanically integrating statistical and judgmental methods should be preferred (Sanders and Ritzman, 2001: 405). Due to the nature of intermittent demand pattern which includes highly variations, judgmental methods may improve the accuracy of the forecasts. Nevertheless, mechanical combination of the statistical and judgmental forecasts is not discussed for the SKUs that have intermittent demand.

2.3. JUDGMENTAL ADJUSTMENT OF STATISTICAL FORECASTS

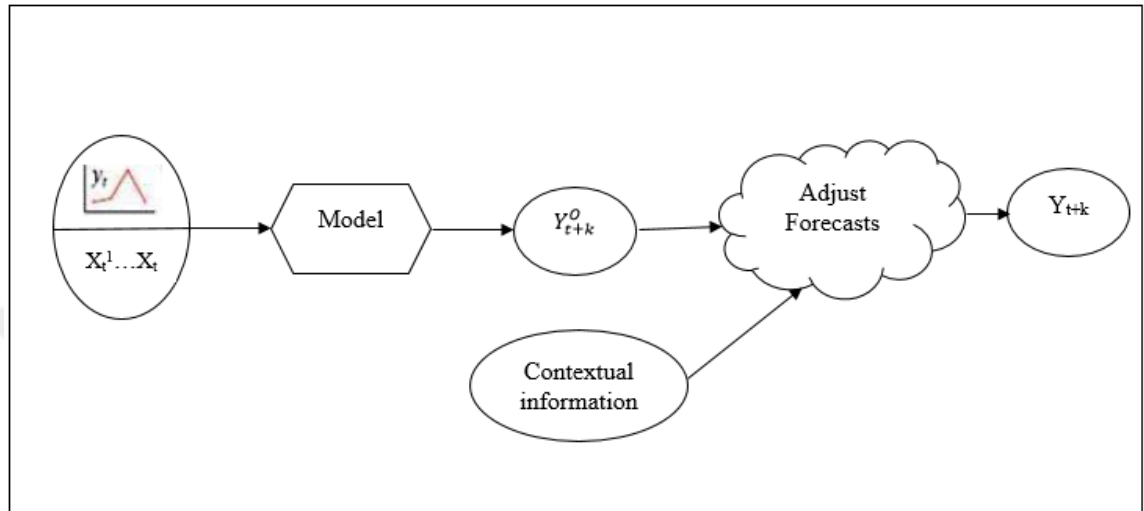
As mentioned previously, combining judgmental and statistical forecasts are shown to generate forecasts that have higher accuracy than using the forecasts individually. This integration can be done either by mechanical approaches like “*simple averaging*” or an “*adjustment of statistical forecasts*” by forecasters (Mahmoud, 1989: 599). Studies have shown that human judgment plays an important role and highly preferred in forecasting. It is a common practice in business for practitioners to adjust statistical forecasts based on contextual factors to get more realistic results (Turner, 1990: 332). A survey held by Sanders and Manrodt (1994: 98) at 96 US corporations showed that about 45% of the forecasters always made judgmental adjustments while 9% forecasters responded that they never did.

Judgmental adjustment process has two steps. First, statistical forecasts are generated to be used as a baseline for forecaster’s adjustments. Next, judgmental adjustments based on contextual information are applied on statistically generated forecasts to have final forecasts. Process of judgmental adjustment is demonstrated in Figure 11.

Judgmental adjustment approach also has the advantage of time since it allows the latest information to be incorporated rapidly. Sanders and Ritzman (1992: 40) proposed several reasons that forecasters prefer to use judgmental adjustments. First, people generally feel more comfortable with their judgment. Second, in real-life product forecasting, additional information like information about the competitors,

inventory, or possible production problems are important, and this additional information often increases the forecast accuracy.

Figure 11: Judgmental Adjustment of Objective Forecasts



Source: Webby and O'Connor, 1996: 100.

In their later study, Sanders and Ritzman (2004: 518) stated that the primary reason for the judgmental adjustment is to integrate the latest and important information about the environment, product and experiences into the forecasts.

Sanders and Ritzman, (2001: 405) proposed three principles concerning *when* to have judgmental adjustments on statistical forecasts. First, “*statistical forecasts should be adjusted based on valuable domain knowledge*”. Domain knowledge is the knowledge that a forecaster gain by experience. Second principle is that the “*adjustment of statistical forecasts if there is high uncertainty*”. For adjustments to be successful, judgment should provide additional information which statistical forecasts are not able to capture like discontinuities and pattern changes in the data. The last principle is the “*adjustment of the statistical forecasts in the light of known changes in the environment*”. For example, the forecaster can adjust the forecast with the information of changes in production level for a specific product.

Sanders and Ritzman (2001: 405) also proposes some principles about *how* to adjust statistical forecasts. First, they suggest that the adjustment process should be structured. Other suggestion of the researchers was documenting all of the adjustments

and relate them periodically to forecast accuracy. Lastly, it is suggested to use the mechanical combination of judgmental and statistical forecasts instead of judgmental adjustments. Even though judgmental adjustments improve accuracy, it can introduce bias. Combining the forecasts mechanically provides the advantages of both statistical and judgmental forecasting while reducing bias.

If there is no contextual information, practitioners should content with whatever cues they can gather from time-series data and the provided forecasts (Gönül, 2013: 78). In such cases, there are two sources of information that forecasters can use to form their adjustments. The first source is the “*characteristics of the underlying time series*”. The second source is the “*accuracy and reliability of the statistical forecasts*”.

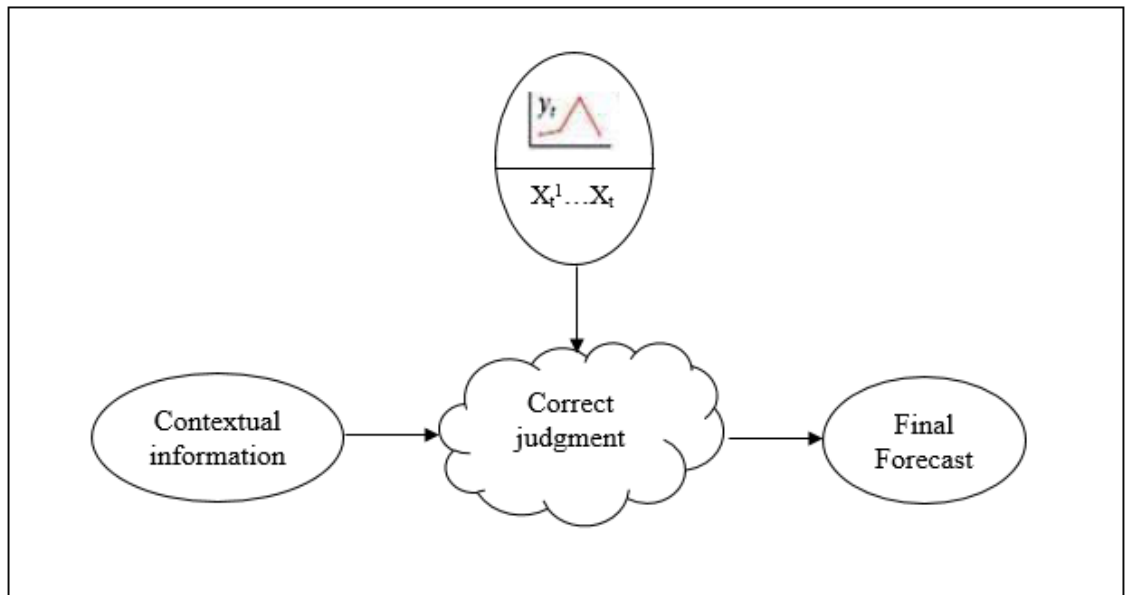
Bunn and Wright (1991: 508) have criticized judgmental adjustment of statistical forecasts since it is highly subject to judgmental biases due to its informal and ad hoc nature. To overcome this problem, application of judgmental in structured, methodological, and rule-based manner was proposed (Gönül, 2013: 86). There are some studies of applied methods for generating judgmental adjustments. Reinmuth and Geurts (1972: 294) used Bayesian techniques. They integrated the information they got from a customer survey into the adjustment made by marketing staff. To provide structure to forecast adjustment and acceptance process, some researchers like Cook et al. (1984: 203), Wolfe and Flores (1990: 391), Yüksel (2007: 1065) have tried to integrate analytical hierarchy process (AHP). Also, neural networks were examined. These methods provided some improvement in the process. However, implementation of these methods was still a problem. They were generally hypothetical and case-based, so it became complicated for organizations to implement (Gönül, 2013: 86)

Despite the overall advantages of this method, there may occur some biases as well. Eroğlu and Croxton (2010: 117) have named three biases, which are “optimism bias”, “anchoring bias”, and “overreaction bias”. Optimism bias refers to forecasters, generally projecting high values, which results in positive forecast errors. Anchoring bias occurs when the forecaster makes adjustments in the right direction, however adjusted forecasts stay close to the statistical ones. Lastly, overreaction bias occurs when a forecast adjustment generates a more massive error in the right direction. Eroğlu and Croxton (2010: 118) have also stated that personality characteristics and motivational orientation impacts the adjustment process and biases may occur.

2.4. STATISTICAL CORRECTION OF JUDGMENTAL FORECASTS

This methodology relies on forecasting judgmentally and reducing the judgment's error by statistically correcting the bias of judgmental forecasts directly (Sanders & Ritzman, 2004: 520). Though this method can lead to large improvements in accuracy, it has not received much attention. Figure 12 represents the statistical correction of judgmental forecasts process.

Figure 12: Quantitative Correction of Judgmental Forecasts



Source: Sanders and Ritzman (2004: 519)

The study of Elgers et al. (1995: 1395) showed that correcting the systematic bias reduces the MSEs (mean square error). Another study held by Goodwin (2000: 271) showed that correcting the judgmental forecasts of the managers reduce MdAPE (median absolute percentage error) values to 11.8%. In the same study combining provided MdAPE value of 14.9%. Theil (1971) developed an approach to perform quantitative correction of the bias and showed that MSE of a set of forecasts could be decomposed into “mean” and “regression bias”. He stated that these biases could be removed from past forecasts by regressing the actuals onto the forecast. Where A and F represents actuals and judgmental forecasts respectively, S_A and S_F are the standard

deviations of actuals and forecasts and r is the correlation coefficient between the actuals and forecast, forecast error MSE is represented mathematically as follows:

$$MSE = (A - F)^2 + (S_F - rS_A) + (1 - r^2)S_A^2 \quad (15)$$

The first term of the equation represents the “mean bias”; the tendency of the model to either over forecast or under forecast. The second term is “regression bias”, which is a consistent deficiency of the forecast to track the pattern in the actuals. The last term of the equation represents the “random error”; the unexplained variations in the actuals. After regressing the actuals on the forecasts, the form becomes:

$$\hat{A}_t = a + bF_t \quad (16)$$

where $\hat{A}_t$ is the estimated actual at time t and F_t represents the forecast at time t . When $\hat{A}_t$ is used as the corrected forecast, both mean and regression biases are removed from the past forecasts. Then assuming that the biases continue into the future, the correction can remove the systematic bias from the future forecasts.

The biases may change in time due to the changes in the market or environment. Sanders and Ritzman (2004: 520) stated that the rapid or sporadically changes in biases between periods, using quantitative correction of judgmental forecasts method can be problematic. Also, they mentioned that using this method is best suitable when there is no quantitative information or in environments where the corrected judgment can be combined with quantitative forecasts.

CHAPTER 3

METHODOLOGICAL FRAMEWORK ON COMBINING FORECASTS

In practice, both statistical and judgmental methods are used to forecast demand. These methods have their unique advantages and disadvantages. Statistical methods use numerical data to predict the future. The predictions are objective. However, they are only based on past data. These methods are not capable of incorporating the effects of the environmental or any additional internal or external factor. Passing these effects up causes the forecast accuracy to decrease. Contrarily of statistical methods, judgmental methods rely on the experience of the experts. This managerial judgment can incorporate the information that the statistical methods are not capable of catching if there is a change in the environment. In the literature, it is stated that using judgmental methods are useful if there is no historical data, the time series and the forecasting environment are highly variable (see Sanders and Ritzman, (2004)). However, the judgment of the experts can be biased, which may lower the forecast accuracy.

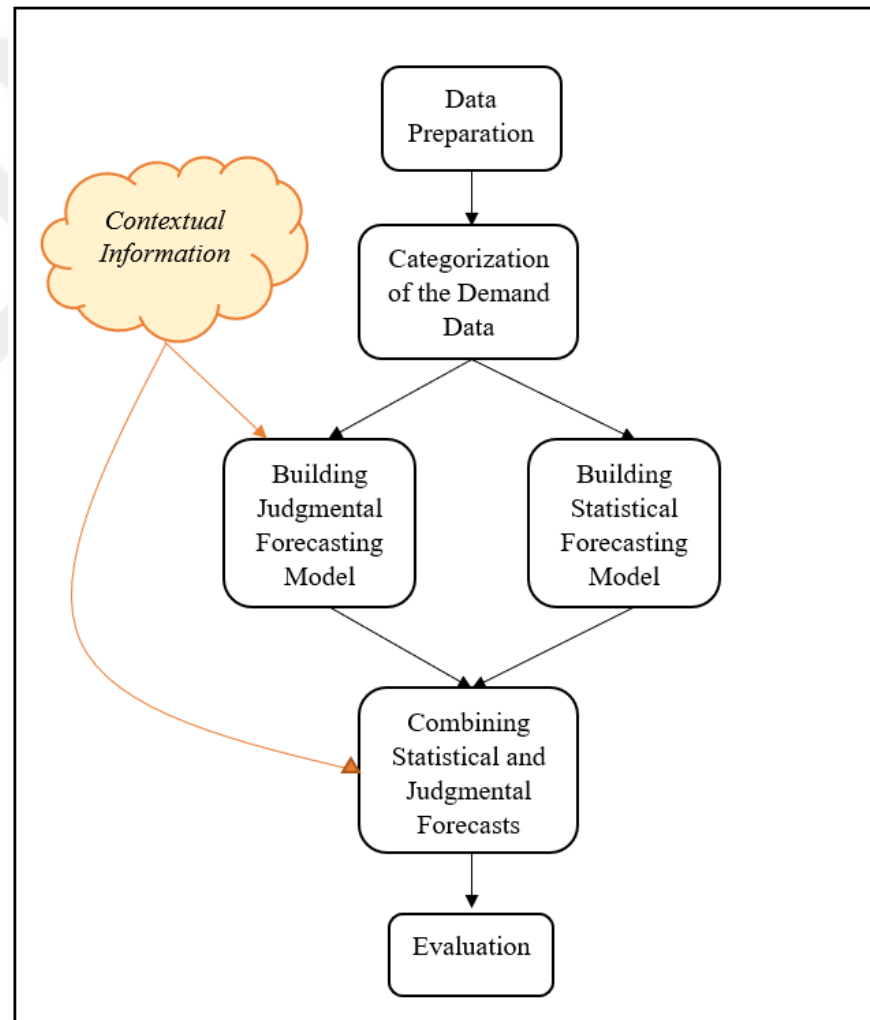
Complex and highly variable structure of the intermittent demand pattern turns forecasting into a challenging task for researchers and practitioners. There are several statistical forecasting methods that are proposed to generate accurate forecasts for the SKUs that have such a pattern. Nevertheless, these methods are not capable of capturing the variations that occur in SKUs with intermittent demand by the nature of this pattern. To come up with these variations, judgmental forecasting methods can be used. Combining statistical and judgmental forecasts is an option to increase the accuracy of the forecasts and to avoid the disadvantages of using statistical or judgmental forecasting methods solely (see Clemen (1989)).

This study aims to build a forecasting model that combines statistical and judgmental forecasts in order to get more accurate results for the SKUs that have intermittent demand pattern and also which is easy to implement and interpret by the practitioners who do not have a statistical background. In the literature, there are several studies about combining forecasts showed that this procedure increases the forecast accuracy ((Franses and Legerstee, 2011), (Wang and Petropoulos, 2016)). The studies in the literature applied combination for SKUs that have a smooth demand

pattern. However, combining statistical and judgmental forecasts for the SKUs with intermittent demand is not covered in earlier studies. In this context, this study will be the first to compare the forecast performances of statistical, judgmental, and combination forecasts.

To achieve the purpose of this study, some structured steps for the forecasting process are defined. The proposed stages of the methodological framework for combining statistical and judgmental forecasts are demonstrated in Figure 13.

Figure 13: Stages of the Proposed Methodological Framework



In the first stage, the data is prepared for the forecasting process. The second step is the categorization of the demand data based on demand patterns (intermittent, smooth, lumpy and erratic) by using the categorization schemes that are proposed for

the SKUs with intermittent demand pattern (see *Section 1.2.2*). In the third stage, statistical forecasting model is built by using the appropriate methods for intermittent demand, which are explained in *Section 1.3*. Parallel to statistical forecast model building, the judgmental forecasting model is also built by using the methods which are explained in *Section 1.4*. The best performing forecasting method for each SKU is selected by using accuracy measures appropriate for intermittent demand data. In the fourth stage, the statistical forecasts that the best performing method generated are combined with the judgmental forecasts by using combining procedures. Finally, the accuracies for statistical, judgmental, and combination forecasts are calculated, and the best performing methods are selected for each SKU.

In the following sections of this chapter, each stage of the methodological framework for combining forecasts for SKUs with intermittent demand is explained briefly.

3.1. DATA PREPARATION

The first stage of the proposed framework covers preparing SKUs data having intermittent demand pattern for building a forecast model. The data usually gathered from companies may not be suitable for the forecasting process. For instance, a historical data for an SKU may have missing values, consist only of zeros or a high number of consecutive zeros especially in the beginning or ending periods which may refer to that SKU is a new product, or it is not in use anymore. Such SKUs should be eliminated from the dataset to transform the remaining data into an appropriate form.

3.2. CATEGORIZATION OF THE DEMAND DATA

The second stage is the categorization of the demand data. Categorizing the data according to the SKUs' demand patterns plays an essential role in selecting the best performing forecasting method and inventory control parameters. Companies hold high numbers of SKUs, so it is suggested that, instead of evaluating them on an individual basis, it is more effective to assess them as groups with similar characteristics. In the literature, as mentioned in *Section 1.2.2*, there are several

categorization schemes proposed by various researchers (see Williams (1984), Johnston and Boylan (1996), Syntetos (2001), Ghobbar and Friend (2002), Eaves (2002), Syntetos et al. (2005), Kostenko and Hyndman (2006) and Boylan et al. (2008)).

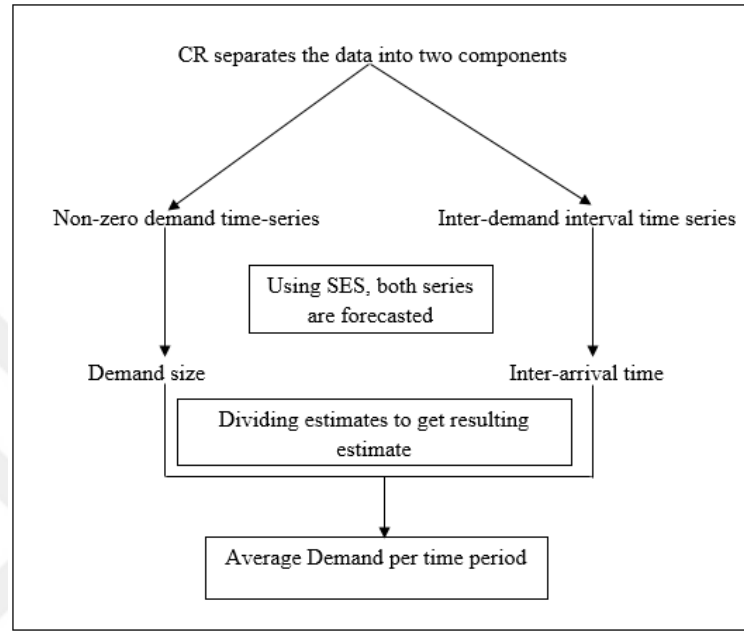
In this study, the Syntetos-Boylan-Croston (SBC) scheme (proposed by Syntetos et al. (2005)), is used. SBC scheme uses average inter-demand interval (p) and squared coefficient of variation (CV^2) to categorize the demand patterns into four groups. These groups are smooth, intermittent, erratic, and lumpy. Authors defined the demand patterns, and they established the regions of superior performance for EWMA, CR, and SBA. The results showed that the cut off points are $p = 1,32$ and $CV^2 = 0,49$. These values and the appropriate forecasting methods suggested for each demand pattern can be found in Section 1.2.4 (*Table 1*). This scheme is selected since one of the most recent studies conducted by Fildes et al. (2019) achieved good results by using this scheme, and it is easy to understand and interpret.

3.3. BUILDING A STATISTICAL FORECAST MODEL

After preparing the data and examining the data patterns by using an appropriate categorization scheme, the third stage of the study is building a statistical forecast model. There are several methods in the literature that are proposed for intermittent demand forecasting. The details for these methods are explained in *Section 1.3*. This study covers three of the given methods: SES, CR, and SBA. SES is the most used traditional method by the practitioners for SKUs that have both smooth and non-smooth demand pattern. It is straightforward and robust. Since it is widely used in real world applications and it is easier to implement and understand, this method is selected for evaluating its performance on intermittent demand forecasting context. *Equation 6* in *Section 1.3.3* shows the general formulation for SES. However, Syntetos et al. (2015: 1747) stated that *"In SES, data that is more recent weights more heavily. Thus, just after a demand occurs, it gives forecasts that are biased high while it gives forecasts that are biased low just before a demand"*. This situation results in high replenishments and excessive stock levels. Thus, CR as the first method, is proposed,

especially for the SKUs that have intermittent demand pattern. Figure 14 schematizes the CR forecasting process.

Figure 14: Forecasting Process of Croston's Method



Source: Nagaria, 2017

If there is no demand in a review period, the estimates remain unchanged. That means if in period t , the demand is zero, the method only increments the count of periods since the last positive demand (Croston, 1972: 295). Only after the occurrence of positive demand, the forecasts are updated. If demand occurs every period, the forecasts that CR generate are same with forecasts that SES generate. So, it is possible to use this method both for intermittent and smooth demand pattern.

Several studies showed the superiority of CR over traditional methods (see Willemain et al., 1994; Johnston and Boylan, 1996; Ghobbar and Friend, 2003). The details for CR are explained in *Section 1.3.4*. Due to its proved success in forecasting the demand for the SKUs with intermittent demand and its applicability without any additional costs for the organizations, this method is also used while generating the statistical forecasts.

Last forecasting method used in this study is SBA, which is the most widely used variant of CR. This method is developed by Syntetos and Boylan (2001) for fixing

the bias caused by a mathematical derivation error in CR. Several studies in the literature showed that SBA outperforms other methods such as SES, CR, and MA (see Eaves and Kingsman, 2004; Syntetos and Boylan, 2005; Syntetos et al., 2005; Gutierrez et al., 2008). This method is selected since it is shown that it improves the quality of forecasts that are generated by CR. Also, recently it has been used by many researchers, mostly having successful results. It is also easy to implement and understand as other selected methods which meets the expectations of this study that aims to propose a forecasting model that do not require any highly advanced knowledge about statistics. Another issue is that, neither of these methods cause any additional cost for the companies since it is easy to apply them by using simple software like MS Excel or RStudio. Due to the success and prevalence of these methods both in literature and practice for intermittent demand forecasting, they are selected for this study. However, any other method given in *Section 1.3* could be chosen depending on the purpose of the practitioner.

While building a forecast model, the datasets are generally divided into two; train and test sets. The train set is used to estimate the parameters of a forecasting method, and the test set is used to evaluate the model's accuracy (Hyndman and Athanasopoulos, 2018).

All three of the selected methods use a smoothing constant α for forecasting and initial values for the forecasts. For CR, Schultz (1987: 457) suggested to use two different smoothing parameters for both estimates, whereas Croston assumed a common smoothing parameter between the range of 0,05-0,20. In the literature, it is suggested to use a small α value for these methods between 0,05 – 0,30. Thus, the α values in this range incrementing in 0,05 is used. Another parameter that need to be determined before starting the forecasting process is the initial values. For SES, one initial value is required to start the forecasting process. For CR and SBA, two initial values, one for demand size and one for the interval between non-zero demands estimate, are required. These initial values can be determined by using the naïve method (one-step ahead), calculated by taking the mean of the related values or determined by the practitioner judgmentally. Likewise, instead of using fixed α and initial values, the practitioners can prefer to use optimal parameters. In the optimization process, the parameters that minimize the given cost functions are

selected. MAR (mean absolute rate), MAE (mean absolute error), MSR (mean squared rate) and MSE (mean squared error) are the cost functions that are used in this study for optimizing the parameters. Cost (or loss) functions are used for the optimization process, and they measure the fit of a model to the actual data. MSE is the most used cost function in SES case. Kourentzes (2014:183) mentioned that “Since this equation is based on the squares, it is more robust to extreme values”. The formula for MSE function is given in *Equation 25* in *Section 3.6* As an alternative to MSE, MAE is also used for optimization purposes. For the equation for MAE, see *Equation 24* in *Section 3.6*.

CR and its variants calculate a demand rate instead of a point forecast. Kourentzes (2014:183) suggested that measuring the difference in demand rate forecast from the raw time series data is not meaningful since they have different units. Moreover, instead of comparing the average demand for each period with the realized demand, it can be compared with the in-sample mean demand over time. Where r_i is the rate, n is the number of in-sample observations, y_t is a time series and $\hat{y}_t$ is the forecasts at time t , *Equation 18* and *Equation 19* represent the MSR and MAR, respectively.

$$r_i = \hat{y}_i - i^{-1} \sum_{j=1}^i y_j \quad (17)$$

$$MSR_n = \sum_{i=1}^n r_i^2 \quad (18)$$

$$MAR_n = \sum_{i=1}^n |r_i| \quad (19)$$

In this work, the statistical methods are evaluated by using some of the forecast accuracy measures explained in *Section 3.6*. For evaluating the statistical forecasting methods, scaled mean absolute error (sMAE), scaled mean squared error (sMSE), root mean squared error (RMSE), symmetric mean absolute percentage error (S-MAPE), mean absolute scaled error (MASE), scaled cumulative error (sCE), and scaled periods in stock (sPIS) are used. Scaling the errors turn the accuracy measure into scale-

independent form. Scale-independent errors can be compared across the series. That is why sMAE, sMSE and MASE are selected. Also, MASE is commonly used in evaluating intermittent demand forecasts, since it is appropriate to use for this problematic demand pattern. RMSE is a scale-dependent error measure, and it is widely used in practice. So, it is included in the comparison process as well. S-MAPE shows accuracy as a percentage. Also, it overcomes the division by zero problems that MAPE has in intermittent demand context. These measures are not capable of determining whether there is a systematic error, bias, or not. To determine if the forecasting methods generate results which are not biased, it is essential to use bias measures together with other accuracy measures. sCE and SPIS are selected as bias measures. Since these measures are scaled, it is possible to use them when comparing across the series.

After calculating the forecast accuracies for each SKU, the results are compared considering different measures. The final forecasts are generated by using the best-performing methods and parameters for each SKU.

3.4. BUILDING JUDGMENTAL FORECAST MODEL

In parallel to the statistical model building in Stage 3, judgmental forecasting process can also be implemented. In real-world applications, practitioners commonly use judgmental forecasting. These forecasts are based on the knowledge and the experience of the experts in the institutions. It has some advantages for the companies in situations like when the historical data is absent; there are changes and unexpected situations in the environment that the statistical forecasting methods are not capable of catching or time-series are highly variable. Also, since it does not need any mathematical background, it is easier to implement. Beside its advantageous points, it also has some disadvantages of the bias. Judgmental forecasting methods are useful if there is no data to use for statistical forecasting.

There are some judgmental forecasting methods such as jury of executive opinion, Delphi method, analogies, and scenario forecasting. A brief explanation for these methods can be found in *Section 1.4*. Using such structured methods is suggested for this stage to obtain more accurate forecasts - a study conducted by Wing Lee et al.

(2007) that uses structured analogies for building a forecasting support system showed that this judgmental input significantly improves forecast accuracy.

3.5. COMBINING STATISTICAL AND JUDGMENTAL FORECASTS

Combining forecasts generated by using different methods plays an important role in improving the overall accuracy. Each technique adds different information to the forecasting process, which a single technique is not capable. Blattberg and Hoch (1990:896) stated that the final forecasts generated by combining judgmental and statistical forecasts are more accurate than the constituent forecasts. The opinions of the experts add contextual information, which explains the unexpected issues in the data. Also, it helps forecasters to understand the future events which would affect the historical data. There are some studies about the effectiveness of combining statistical and judgmental forecast. In their study, Lawrence et al. (1986: 1524) examined the effectiveness of combining forecasts for time series that have different forecasting difficulty levels and seasonality. They found out that for series with low MAPE, the combination is more effective and seasonality has no influence on the benefits of the combination. Working on the effects of difficulty levels of series with a coefficient of variance instead of MAPE, Sanders, and Ritzman (1990) also found out that combining forecasts is best for easier series. For harder series, judgmental forecasts are more accurate than statistical forecasts; thus, combined forecasts have less accuracy than judgmental forecasts. Weinberg (1986) worked on the forecasts for attendance at events. In his study, by using an econometric model, ex-ante forecasts were generated. Also, managers generated forecasts by using their judgment. Results showed that the econometric model provided more accurate results than managers' judgment. However, the accuracy of the combination of the econometric model and judgment was superior to both model and judgment alone.

Statistical forecasts can be combined with judgment in various ways. The details for all explained in *Section 1.5*. In his study, Armstrong (2001: 419) proposed the following procedures for combining the forecasts.

- *Use different data or different methods.* By using different data, forecasters can have additional useful information. This additional information and using different methods can decrease the forecast error.

- *When possible use at least five forecasts.* When it is possible, combining forecasts from five different methods leads to an improvement in accuracy.
- *Use formal procedures to combine forecasts.* To achieve higher accuracy and avoid biases, combining should be done mechanically, and all procedure should be fully described. In the case of using judgment, it should be in a structured way, and details of the procedure should be recorded. If there is a lack of information about the relative accuracy of alternative sources of forecasts, judgmental weights should be avoided. In the case that subjects have good feedback about the accuracy of the sources, a judgmental weighting is more accurate (Fischer and Harvey, 1999: 228).
- *Use equal weights unless there is strong evidence to support unequal weighting of forecasts.* The study of Clemen (1989) showed that using equal weights for the constituent forecasts provide accurate results. Armstrong (2001: 422) stated that, when the practitioner is uncertain about the forecasting methods to be used, each forecast should be weighted equally.
- *Use trimmed means.* Due to reasons like miscalculations, data errors, and misunderstandings; one may observe larger errors in individual forecasts. For avoiding this situation, it may be useful to throw high and low forecasts. It is recommended to use trimmed means when there are at least five forecasts. Since the effects of trimming not studied, this principle is speculative.
- *Use the track record to vary the weights if the evidence is strong.* If the forecaster has good evidence supporting that a particular method is more accurate than others, when the method has been tested in a situation, forecaster should give it a heavier weight.
- *Use domain knowledge to vary the weights on methods.* If someone is familiar with the situation, s/he may be able to make useful judgments about which methods are appropriate most. Also using an expert's domain knowledge can be beneficial. For example, they can be asked for the seasonality, expectation about trends, or if they expect any discontinuity or not.

As mentioned earlier, combining the forecasts provide an improvement in the forecast accuracy since it brings different and independent information together instead of choosing only the best performing method. It also prevents the loss of valuable information. One of the mostly used combining method, judgmental

adjustments, cause bias, and combining the forecasts helps forecaster to avoid such problem. Combining can perform even better when there is uncertainty in forecasting environment. Unexplained significant variations in the past or on expected volatile changes in the future can be judged as uncertain (Armstrong, 2001: 426). For the spare parts, this is the case. Generally, the demand size and the time between two non-zero demands have variations. To overcome the adverse effects of these variations, practitioners can integrate judgmental forecasts with the statistical ones. It is expected that the combination would lead to higher forecast accuracy.

Combination procedure can range from mechanical methods such as taking a simple or weighted average of the constituent forecast, to using judgment to determine how to forecasts should be combined (Lawrence, et al., 2006: 508). The simplest approach is to combine both statistical and judgmental forecasts using equal weights. This research focuses on combining by using weighted averages. The weights that provide the best forecast accuracy should be chosen.

3.6. EVALUATION

The last stage of this research is evaluating the performances of statistical, judgmental, and combined forecasts to find out the best performing method for each SKU. Besides producing the most accurate forecasts, the chosen forecasting method should also produce a forecast that is timely and understood by management so the forecast can help produce better decisions (Hanke and Wichern, 2014: 4).

For evaluating accuracy levels of statistical, judgmental, and combination forecasts and choosing the best one for each SKU, the appropriate accuracy measures that are used. Evaluation process can be done by using various forecast accuracy measures that are appropriate for SKUs that have intermittent demand.

Forecast accuracy is the most important criterion when deciding whether to use a forecasting method or not. There are many forecasting methods proposed in the literature which generate different forecast values for the same data. However, in practice, the generated forecasts are rarely perfect. To decide which forecasting method is the best, it is important to see the performance of these methods. To do so, forecast accuracy measures are used. Generally, the forecasting method that generates

the smallest forecast error, is the best. Forecast error is the difference between the actual and the forecast value for the same term. Equation 20 shows the calculation of forecast error (e_t).

$$e_t = A_t - F_t \quad (20)$$

where;

e_t = the forecast error in period t

A_t = the actual value in period t

F_t = the forecast value for period t

The e_t values which are close to 0 shows that the forecast is close to the actual. Positive e_t ($e_t > 0$) value shows that the forecasting method used underestimates and generates forecasts which are smaller than the actual. Negative e_t ($e_t < 0$) values show the overestimation since the forecasting method generates forecasts that are larger than the actuals.

The error measures are classified into four groups as scale-dependent, scale-independent, percentage, and relative. Table 4 shows the classification of the accuracy measures explained in this section.

Scale-dependent Errors:

Scale-dependent errors have the same scale as the data, and they used for comparing different forecasting methods applied on the same data. It is not appropriate to use such measures for comparing across datasets that have a different scale (Hyndman and Koehler, 2006: 682).

Table 4: Performance Measures

Scale-Dependent Error Measures	Scale-Independent Error Measures
• Mean Error • Cumulative Error • Periods-in-Stock • Mean Absolute Error • Mean Square Error • Root Mean Square Error	• Mean Absolute Scaled Error • Scaled Cumulative Error • Scaled Periods in Stock • Scaled Mean Absolute Error • Scaled Mean Square Error
Percentage Error Measures	Relative Error Measures
• Mean Absolute Percentage Error • Symmetric Mean Absolute Percentage Error	• Relative Geometric Root Mean Square Error • Percentage Better and Percentage Best

Mean Error (ME) is calculated by taking the arithmetical average of forecast errors. It simply determines the bias of a forecasting method. The small value of ME does not mean that the error is small, but it shows biases. If the forecasting method underpredicts the actual, ME takes a negative value, and if it overpredicts, the ME takes a positive value.

$$ME = \frac{1}{n} \sum_{t=1}^n e_t \quad (21)$$

where;

e_t = the forecast error in period t

Cumulative Error (CE) is a widely used bias measurement when evaluating forecast performance. This method sums up the errors to determine the bias as in Equation 22.

$$CE = \sum_{t=1}^n e_t \quad (22)$$

Periods in stock (PIS) is a bias measure which is proposed in the study of Wallström and Segerstedt (2010: 629). This measure shows the total number of periods

in stock or stock-out periods of the forecasted item. By using double cumulative summation, one can compute the error measure. To exemplify how does this measure work, let us assume that our forecast horizon is three days (periods) and demand for each day is 1 unit. At the beginning of the first period, 1 item is delivered to the stock. If no demand occurs, then the result is plus one PIS. If a demand occurs in the same period, then the value of PIS equals to zero in the first period. In the case of no demand in all three days, then the PIS equals to +6 since the item from day one spends 3 days, from day two spends 2 days and from day three 1 day in the stock (Wallström and Segerstedt, 2010: 629). Equation 23 shows the formula of PIS for series i as follows.

$$PIS_i = - \sum_{h=1}^H \sum_{j=1}^h (Y_{N+j} - \hat{Y}_j) \quad (23)$$

where H is the length of the required forecast horizon. If the value of PIS is positive, this means there is stock which is left over. Negative values refer to stock-outs.

Mean Absolute Error (MAE) calculates the forecast accuracy by averaging the absolute values of the forecast errors. MAE shows the errors regardless of under or over forecast. This method gives an average of error measurement irrespective of the direction. Also, it eliminates the canceling out problem caused in ME method. This method shows how large the average error can be.

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (24)$$

Mean Square Error (MSE) squares the forecast errors and averages them to measure overall forecast error. Like MAE, it eliminates the canceling out problem. This approach penalizes the large forecasting errors since the errors are squared (Hanke and Wichern, 2014: 36). MSE is related to the standard deviation of forecast errors. Thus, it is appropriate to use it for mathematical operations. The formulation for MSE is given in Equation 25:

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (25)$$

This method is widely used by the practitioners to compare the accuracy levels of different methods. MSE value close to 0 is preferable.

Root Mean Squared Error (RMSE) is a scale-dependent measure which is calculated by taking the square root of MSE. Lower values of RMSE is better. The formula for calculating RMSE is stated in Equation 26:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (26)$$

Scale-independent Errors:

In scale independent error measures, the error is scaled, and it becomes independent of the scale of the data. These error measures can be used to compare forecasting methods both on a single series and between different series (Hyndman, 2006: 43).

Mean absolute scaled error (MASE) is a scale-free forecast accuracy measure that was proposed by Hyndman and Koehler (2006: 685). It can be used for comparing forecast errors both on a single series and between series. Also, Hyndman and Koehler (2006: 682) mentioned that using this method for intermittent series is appropriate since it does not give infinite or undefined values. In MASE, authors scaled the errors based on the in-sample MAE from the naïve forecasting method. By using the naïve method, one-period-ahead forecasts are generated from each data point in the sample. Scaled error is defined in Equations 27 and 28.

$$q_t = \frac{e_t}{\frac{1}{n-1} \sum_{t=2}^n |A_t - A_{t-1}|} \quad (27)$$

and

$$MASE = \text{mean}(|q_t|) \quad (28)$$

If the value of scaled error is less than one, it means that the actual forecast performance is better than the naïve method. Conversely, if q_t is higher than one, it can be interpreted as the actual forecasting method performs worse than the naïve method.

In MASE, division by zero only occurs if all the values in time series are equal. This measure is symmetrical and robust to outliers. Hyndman (2006: 46) also stated that this measurement is the best accuracy metric for intermittent demand studies and beyond.

To average the error measures across series, scaling is a useful option. Petropoulos and Kourentzes (2014: 921) suggested scaling out-of-sample CE of each series by using the mean value of all in-sample periods. The formulation for *scaled CE* (sCE) is given in Equation 29.

$$sCE = \frac{\sum Y_{N+h} - F_h}{\frac{1}{N} \sum_{t=1}^N Y_t} \quad (29)$$

where N is the number of in-sample observations, Y_{N+h} is the h^{th} out-of-sample period and F_h is the h -steps ahead forecasts. sCE can be used for comparing the performance of different methods on different datasets.

To turn PIS into a scale-independent form, *Scaled PIS* ($sPIS$) is calculated. The formula for $sPIS$ is:

$$sPIS_i = \frac{PIS_i}{\frac{1}{N} \sum_{t=1}^N Y_t} \quad (30)$$

Like in sCE , MAE can be scaled to be able to average the measures across series. The out-of-sample mean error of each series is scaled using the mean value of all in-sample periods. *Scaled MAE* ($sMAE$) is calculated as the average of scaled absolute error across all series and horizons (Petropoulos and Kourentzes, 2014:921). Equation 31 demonstrates scaled absolute error (sAE) for the series i , and horizon h is as follows:

$$sAE_{i,h} = \frac{|Y_{N+h} - F_h|}{\frac{1}{N} \sum_{t=1}^N Y_t} \quad (31)$$

where N is the number of in-sample observations, Y_{N+h} is the h^{th} out-of-sample period and F_h is the h -steps ahead forecasts.

Since it is scale dependent, using MSE across different time series is not suitable as well. However, it is possible to have scale independent values by using scaled MSE. sMSE is the mean of scaled squared error across all series and horizons. The formula is given in Equation 32.

$$sSE_{i,h} = \left(\frac{Y_{N+h} - F_h}{\frac{1}{N} \sum_{t=1}^N Y_t} \right)^2 \quad (32)$$

where N is the number of in-sample observations, Y_{N+h} is the h^{th} out-of-sample period and F_h is the h -steps ahead forecasts.

Percentage Error Measures:

Percentage error measures are also independent of scale. In practice, these measures are widely used for comparing the forecasting performance of various methods across different datasets.

Mean Absolute Percentage Error (MAPE) shows the accuracy of a forecasting method as a percentage. It is a scale-independent method so that it can be used for comparing forecasting performance across different data sets. The formula for calculating MAPE is in Equation 33.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{A_t} \right| \times 100 \quad (33)$$

Application of MAPE is simple. However, it has some disadvantages. First, if there are zero values in the denominator (actual demand), which is mostly the case for intermittent demand, there will be division by zero problems. Thus, the percentage calculation will be undefined. Because of this, using MAPE is not appropriate for items

with intermittent demand. Another drawback is the influence of low valued denominator. If this is the case, the percentage error inflates and causes outliers. Lastly, if the forecast is larger than the actual demand, MAPE provides a larger error than if the forecast is lower than the actual demand.

Symmetric Mean Absolute Percentage Error (S-MAPE) is also an accuracy measure based on percentage errors. This method is proposed by Makridakis (1993: 529) to avoid asymmetry caused by the application of MAPE. S-MAPE is defined as follows:

$$S - MAPE = \frac{2}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t + F_t} \times 100 \quad (34)$$

Unlike MAPE, S-MAPE has a lower and an upper bound. The formula above provides results between 0 and 200%. However, since it is easier to interpret the percentage error between 0 and 100%, the formula in Equation 35 is used:

$$S - MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t + F_t} \times 100 \quad (35)$$

Relative Error Measures:

Relative error measures are used to determine the best performing method by dividing the forecast errors obtained by using different forecasting methods to each other. This division scales the measures (Hyndman and Koehler, 2006: 683).

Relative Geometric Root Mean Square Error (RGRMSE) is one of the mostly used relative error measurements, and it was firstly proposed by Fildes (1992: 84). Syntetos (2001: 207) recommended using this accuracy measure for intermittent demand. Syntetos and Boylan (2005: 313) also mentioned that RGRMSE is a very well-behaved accuracy measure to use in intermittent demand. Equation 36 is given for calculating the RGRMSE for methods A and B in a particular time series:

$$RGRMSE = \frac{(\prod_{t=1}^n (A_t - F_{A,t})^2)^{\frac{1}{2n}}}{(\prod_{t=1}^n (A_t - F_{B,t})^2)^{\frac{1}{2n}}} \quad (36)$$

where subscripts A and B refers to the forecasting methods, A_t is the actual demand in period t, F_t is the estimate of demand in period t obtained by any of the forecasting methods considered and n is the number of demand periods considered for comparison (Syntetos and Boylan, 2005: 308).

Fildes (1992: 83) assumed that the squared errors produced by a particular method at different points in time for a particular series are of the form:

$$(A_t - F_{M,t})^2 = e_{M,t}^2 = \varepsilon_{M,t}^2 u_t \quad (37)$$

In the equation above, error in a particular period has two parts. Firstly, u_t refers to the errors due to the particular period affecting all methods equally. Secondly, $\varepsilon_{M,t}^2$ is the specific errors of method M. Fildes (1992: 84) showed that using RGRMSE is independent of the u_t . Thus, this accuracy measure focuses only on the relative merits of the methods (Gutierrez et al., 2008: 414).

Percentage of times better (PB) is defined by Syntetos and Boylan (2005: 307) as “an accuracy measure that summarizes the performance of one method relative to another method across a set of time series where the forecast error of the methods for each individual series is assessed using one of the absolute accuracy measures”. It is the percentage of times that one method under consideration performs better than one another based on any performance criteria like mean deviation or geometric mean square error. It can easily be calculated and interpreted. Also, it is robust to large forecast errors.

PB is used for comparing two estimators. When more than two estimators are compared, *Percentage Best (PBt)* is used rather than PB. In PBt, the proportion of times that one method performs better than all other methods are calculated (Syntetos and Boylan, 2005: 308).

In the final evaluation, MASE, RMSE, S-MAPE, and bias are calculated. Bias is calculated by averaging the difference between the actual and predicted value. This value should be close to zero if the forecast is unbiased. The positive value refers to underestimation, and the negative value means that the forecasting method overestimated the values.

CHAPTER 4

DEMONSTRATION OF THE PROPOSED FRAMEWORK

In order to demonstrate the proposed framework, the stages explained in *Chapter 3* are implemented on two real-world data that contains demand information for after-sales spare parts from companies that operate in automotive and small home appliances industries. Spare parts are typical examples for the SKUs that have intermittent demand pattern. They are vital for companies, especially for the ones that serve in industries such as automotive, aerospace, and aviation. Even though these parts do not play a big role in the sales of a company, they may establish up to 60% of the total stock value (Johnston et al., 2003: 834). So, even small improvements in spare parts management result in significant cost savings. Many companies hold spare parts which are strategic for operations. These are interchangeable parts that companies keep in inventory to use for repairing or replacing a failed unit. In case of a breakdown of a component or, if a replacement is required, a need for these SKUs arises. They are used for after-sales purposes and preventive and corrective maintenance processes as well. Companies need to hold a wide range of spare parts to prevent stock-outs. However, some spare parts may be very expensive, and holding excessive inventory may cause high holding costs. If the required spare part is not in the inventory, the processes may limp. Stock-outs may cause lost revenues and dissatisfaction of the customers, which may result in high costs. Also, holding an unnecessary amount of spare parts cause high costs and obsolescence. To avoid these kinds of problems, managing the inventory of these SKUs effectively is crucial for a company.

Forecasting the demand accurately is an essential element of inventory management. Unlike finished goods, spare parts have an intermittent demand pattern which turns forecasting into a robust process. For the forecasting process of the parts that have these patterns, using traditional forecasting methods proposed for smooth demand do not provide a high accuracy level. To overcome this problem and generate more accurate forecasts, a methodological framework that combines statistical and judgmental forecasts is proposed in *Chapter 3*.

First, the demand data is prepared for the forecasting process by eliminating inappropriate information from the dataset. In the second stage, the demand data is

categorized into subgroups (smooth, intermittent, lumpy, and erratic) by using SBC categorization scheme. In the third stage the statistical forecasting methods, SES, CR, and SBA are applied on each SKU of both datasets and the best performing method for each SKU is chosen based on forecast accuracy measures MASE, sMSE, sMAE, S-MAPE, RMSE, sCE and sPIS. Parallel to this process, the judgmental forecasts are obtained from the companies that provide the data. Next, the statistical and judgmental forecasts are combined by using weighted averages. Finally, the accuracy measures MASE, RMSE, and bias are calculated for statistical, judgmental, and combined forecasts. The results are evaluated, and the best performing method for each SKU in both datasets is selected.

In this chapter, the implementation processes of the proposed framework for combining statistical and judgmental forecasts are briefly explained step by step.

4.1. DATA PREPARATION

In order to demonstrate the proposed framework, this study uses two real datasets from different companies, which include historical demand information for after-sales spare parts. These datasets cover monthly customer demands. All the data are analyzed by using RStudio (version 3.6.0) and MS Excel 2019.

Dataset A: Automotive Spare Parts

The first dataset is provided by an international company operating in the automotive industry, which is placed in Izmir. This dataset includes 36 months after-sales spare parts demanded from 2016 to 2018. The final dataset is restricted to 301 SKUs, 10.836 data in total, after eliminating 14 SKUs that do not have any demand in the first 12 months. For anonymity, the company did not provide any information about the details of the SKUs, and all SKUs have symbolic names. The descriptive statistics for the automotive spare parts dataset are given in Table 5.

Table 5: Descriptive Statistics for Dataset A

301 SKUs	Demand Intervals		Demand Sizes		Demand per Period	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Minimum	1,000	0,000	8,385	48,431	0,000	0,000
25%ile	1,000	0,000	41,280	212,444	13,290	45,261
Median	1,110	0,329	73,967	359,549	49,409	189,558
75%ile	1,694	0,874	123,180	532,862	101,836	410,342
Maximum	4,306	2,217	391,575	1443,764	391,575	1443,764

The descriptive statistics table is constructed for intervals between non-zero demands, non-zero demand sizes and demand per period regardless of the demand size. The results show that the average non-zero demand sizes and demand per period are highly variable for this dataset.

Dataset B: Small Home Appliances Spare Parts

The second dataset is provided by a company that operates in the electronics industry, manufacturing small home appliances. It covers after-sales spare parts demanded for 60 months, starting from 2014 and ending in 2018. Like the automotive spare parts dataset, this company did not provide any information about details except the amounts of the demanded SKUs and all SKU names are changed for the anonymity. There are 8.498 SKUs in this dataset. Some SKUs not demanded in 60 months are eliminated. Also, the SKUs that have no demand or too few demand occurrences in the first and last 24 months; the ones that have demand for less than ten times in 60 months and the products that have no demand between months 24 and 36 are eliminated as well. Finally, 2.431 SKUs are left, which is equal to 145.860 data points in total. The descriptive statistics for small home appliances spare parts dataset are given in Table 6.

Table 6: Descriptive Statistics for Dataset B

2431 SKUs	Demand Intervals		Demand Sizes		Demand per Period	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Minimum	1,000	0,000	1,111	0,788	0,626	6,237
25%ile	1,074	0,255	5,239	19,509	5,387	28,609
Median	1,523	0,850	9,842	37,236	10,653	48,101
75%ile	2,632	2,068	17,831	64,468	19,053	75,091
Maximum	7,299	4,923	44,689	149,373	50,194	169,438

Descriptive statistics table shows that the demand intervals for Dataset B are more variable than the Dataset A, which means that the number of consecutive zero demands are higher. However, the variety in non-zero demand sizes and demand per period in Dataset B is less than the Dataset A.

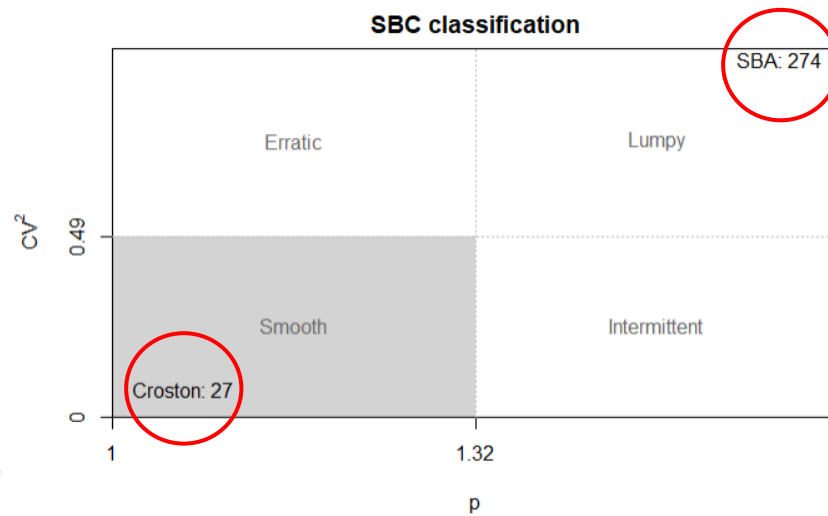
4.2. CATEGORIZATION OF THE DEMAND DATA

Syntetos-Boylan-Croston (SBC) categorization scheme is used for both datasets to categorize the demand data into four groups (smooth, erratic, intermittent, and lumpy) by using p and CV^2 values of each product. In order to categorize the whole dataset, RStudio's "tsintermittent" package (Kourentzes & Petropoulos, Package 'tsintermittent', 2016), the *idclass* function is used. However, this function does not provide any information about the specific demand pattern of each SKU. So according to their p and CV^2 values, SKUs are classified into related demand patterns in MS Excel.

Categorization of Dataset A:

For automotive spare part data (Dataset A), SBC classification can be seen in Figure 15.

Figure 15: SBC Classification for Dataset A



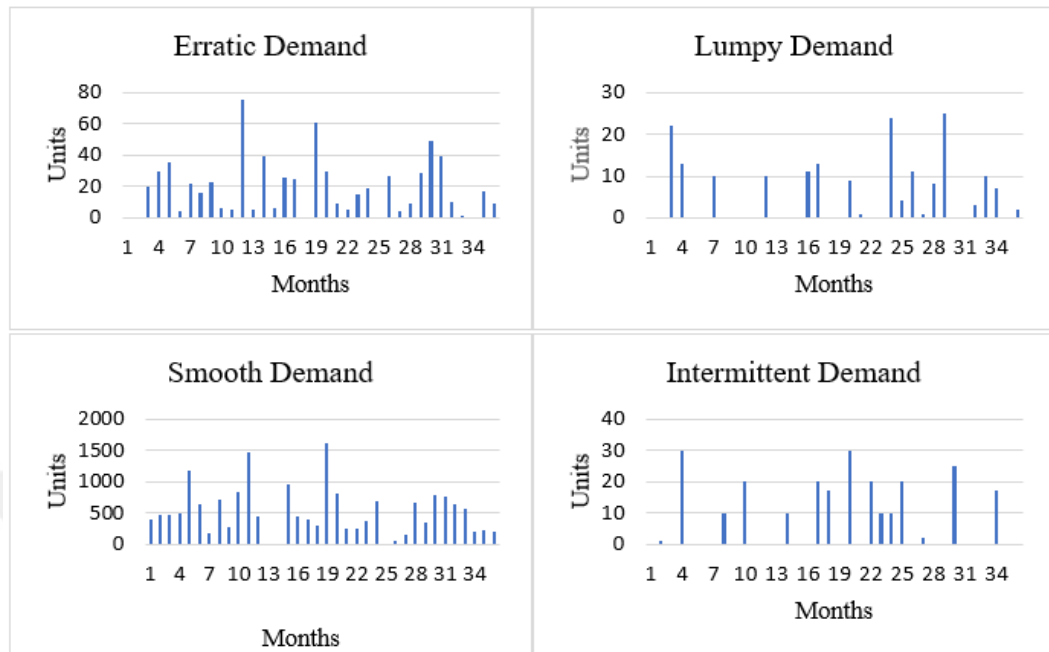
Syntetos et al. (2005: 499) suggested using CR for smooth and SBA for other patterns. Thus, the circles in Figure 15 states that 27 of 301 SKUs have smooth and 274 SKUs in total have erratic, lumpy, and intermittent demand patterns. Appendix 1 shows the R output for classification of the SKUs in Dataset A. By using the p and CV^2 values in this output, SKUs are assigned to the relevant groups in MS Excel. The number of SKUs belong to each category, and their percentages are shown in Table 7.

Table 7: Demand Patterns for Dataset A

Pattern	Condition	Quantity	Percentage
Smooth	$p \leq 1,32$ and $CV^2 \leq 0,49$	27	9,0
Erratic	$p \leq 1,32$ and $CV^2 > 0,49$	111	36,9
Intermittent	$p > 1,32$ and $CV^2 \leq 0,49$	25	8,3
Lumpy	$p > 1,32$ and $CV^2 > 0,49$	138	45,8

Majority of the SKUs in Dataset A have lumpy and erratic demand pattern while the SKUs that have smooth and intermittent pattern form only the 17,3%. Figure 16 illustrates real examples of each demand category based on three years demand history of Company A.

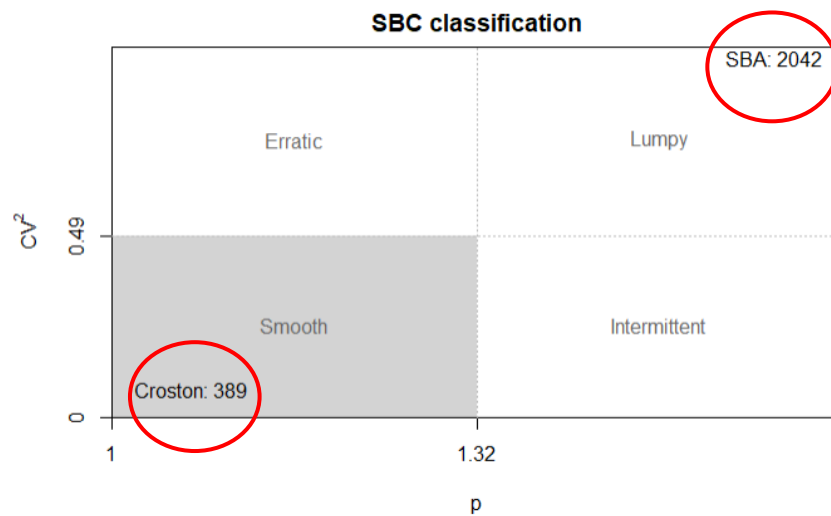
Figure 16: Visual Representation of Demand Patterns Selected from Dataset A



Categorization of Dataset B:

As the first step, the small home appliances spare parts dataset is categorized on R by using SBC categorization scheme. Figure 17 represents the categorization scheme.

Figure 17: SBC Classification for Dataset B



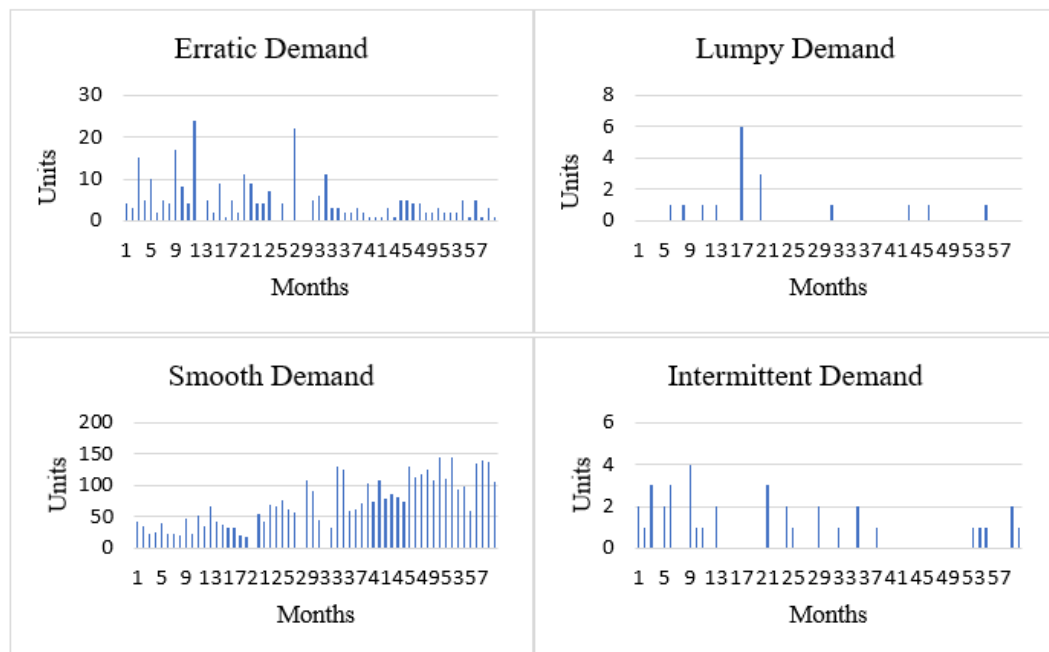
As shown in Figure 17, 389 SKUs in Dataset B belong to smooth demand pattern, and CR should be used for forecasting these SKUs. The rest 2.042 SKUs belong to erratic, lumpy and intermittent demand patterns, and they should be forecasted by using SBA. The SKUs are classified in detail, as given in Table 8. This table indicated that majority of the SKUs in dataset B have intermittent demand while the least number of SKUs (16%) have a smooth demand pattern.

Table 8: Demand Patterns for Dataset B

Pattern	Condition	Quantity	Percentage
Smooth	$p \leq 1,32$ and $CV2 \leq 0,49$	389	16,0
Erratic	$p \leq 1,32$ and $CV2 > 0,49$	582	23,9
Intermittent	$p > 1,32$ and $CV2 \leq 0,49$	802	33,0
Lumpy	$p > 1,32$ and $CV2 > 0,49$	658	27,1

The following figures are the examples of each demand category considering 60 months real data of after-sales spare parts of Company B.

Figure 18: Visual Representation of Demand Patterns Selected from Dataset B



4.3. BUILDING A STATISTICAL FORECAST MODEL

After categorizing the SKUs based on their demand patterns, the third stage of this study is building a statistical forecasting model. In this study, statistical forecasting methods that generate accurate forecasts and which are easy to implement and interpret for the practitioners, without any additional costs to the company are selected. For this purpose, one of the traditional forecasting methods, SES, the pioneer method that is proposed for intermittent demand forecasting, CR, and one of its variants, SBA are selected. These methods are frequently used in the literature and practice for forecasting the intermittent demand. Studies showed that they generate accurate forecasts without adding complexity to the forecasting process (see Willemain et al. (1994), Boylan (1996), Ghobbar and Friend (2003), Eaves and Kingsman (2004) Syntetos and Boylan (2005), Syntetos et al. (2005) and Gutierrez et al. (2008)). Also, the users do not need additional software to apply the selected methods.

Firstly, the datasets are split into two subsets as train and test sets. The Dataset A has 36 months demand information for 301 SKUs. The first 24 months is separated as the train set while the rest 12 months form the test set. The Dataset B, which has demand information of 2.431 SKUs for 60 months, is separated as the first 42 months train set and rest 18 months test set.

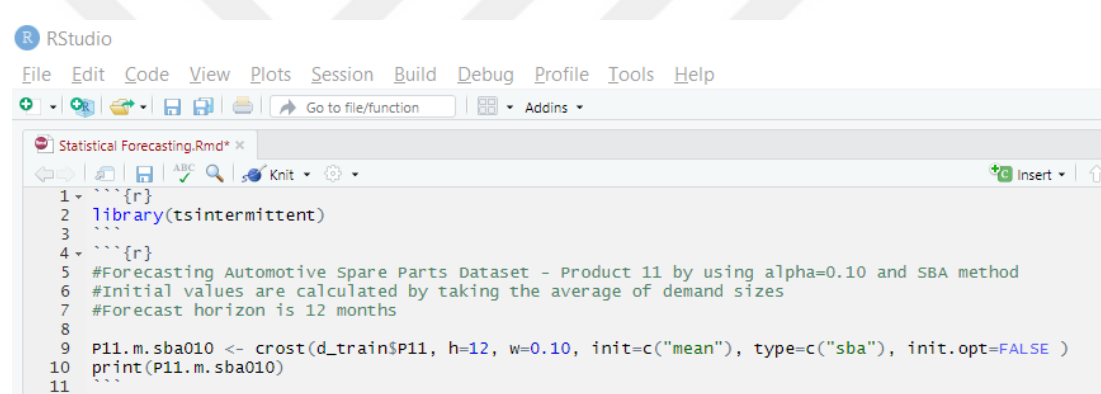
For generating the forecasts, “*tsintermittent*” package of RStudio is used. It offers functions for SES, CR, and SBA. For SES, “*sexsm*” function and for CR and SBA *crost* function are used. The formulations and brief explanations for these functions are given in Appendix 2 and Appendix 3, respectively.

The performance of these forecasting techniques is directly related to the choice of α parameter and initial values. Croston (1972: 290) suggests that smoothing constant α should be between 0,1 - 0,2. Syntetos and Boylan (2005:309) also suggested that low smoothing constant α values between 0,1 and 0,20 are realistic. Based on this information, for all SKUs in both datasets, α values are chosen between 0.05 - 0.30 incrementing in 0.05. The “*crost*” function offers an option to use single or two optimal α parameters for CR and SBA. This parameter α is calculated automatically by choosing one of the predetermined cost functions: mean absolute rate (MAR), mean absolute error (MAE), mean squared error (MSE), mean squared rate

(MSR). Besides the preset α values; forecasts are generated by using all combinations of single and double α values that optimize each cost function. For the preset forecasts, the calculations are repeated for the mean and naïve initial values. In total, for SBA and CR, 56 different forecasts are generated for each product. The forecast horizons are equal to the length of test sets; 12 months for Dataset A and 18 months for Dataset B.

Figure 19 shows an example for the application of the forecasting methods by using R. In this example, Product 11 Dataset A is forecasted by using SBA method for a forecast horizon of 12 months ($h=12$), $\alpha(w)=0.10$, and the mean initial values of the demand size and demand interval. The output for this example is given in Figure 20.

Figure 19: RStudio – *tsintermittent* SBA Codes for Product 11 (From Dataset A)

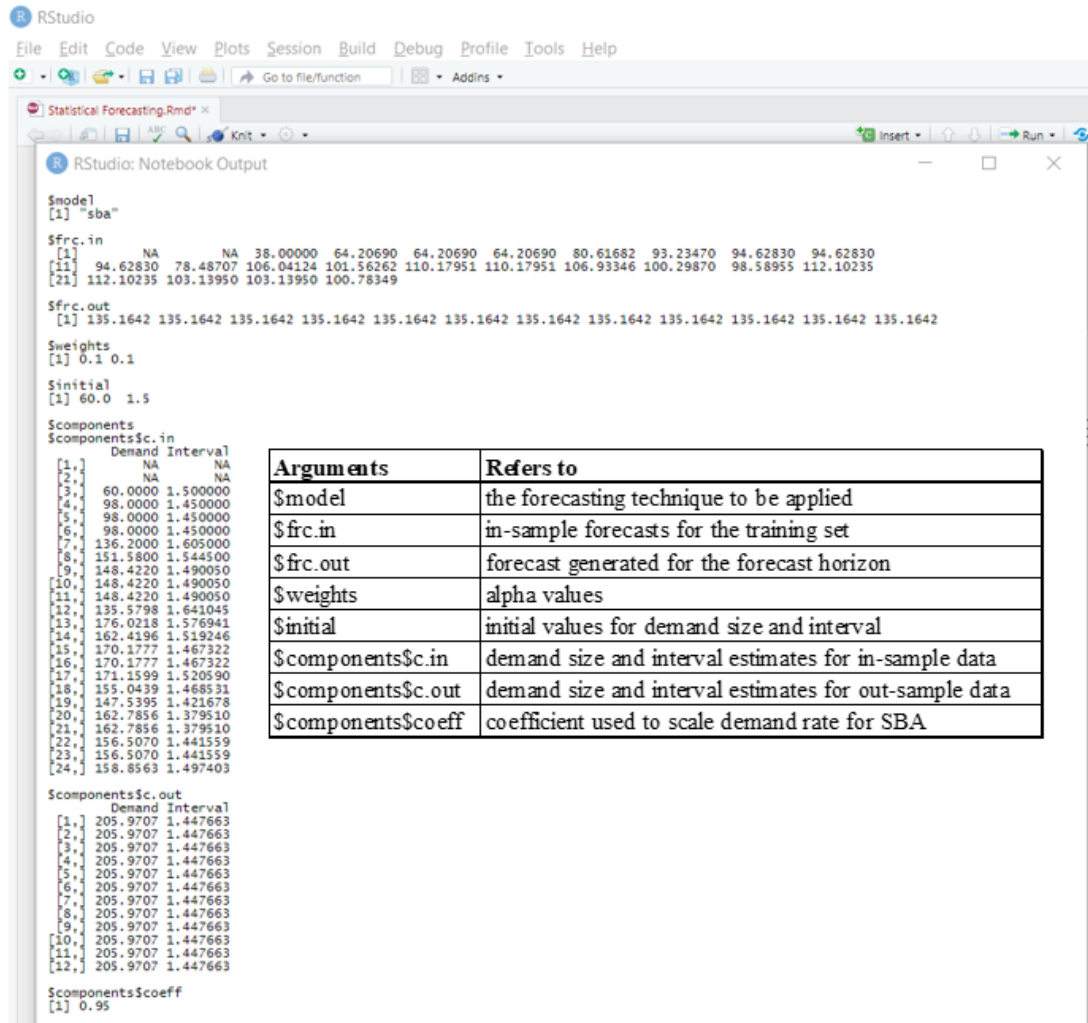


```

1  {r}
2  library(tsintermittent)
3  {r}
4  {r}
5  #Forecasting Automotive Spare Parts Dataset - Product 11 by using alpha=0.10 and SBA method
6  #Initial values are calculated by taking the average of demand sizes
7  #Forecast horizon is 12 months
8
9  P11.m.sba010 <- crost(d_train$P11, h=12, w=0.10, init=c("mean"), type=c("sba"), init.opt=FALSE )
10 print(P11.m.sba010)
11

```

Figure 20: RStudio – Output of SBA Forecasts for Product 11 (From Dataset A)



For example, the SBA forecasts for Product 11 can be calculated based on Equation 14. Argument \$weights in Figure 20 refers to the α value, demand column both in \$components\$in and \$components\$out to Z_t and interval column both in \$components\$in and \$components\$out to P_t in Equation 14. Z_t estimates the demand size in units, and P_t estimates the mean interval between nonzero demands in periods, months for the datasets used in this study. When forecasting the demand for the first month of the out-sample (train) data (25th month) by using SBA, the calculation is as following:

$$Y_{25} = \left(1 - \frac{0.1}{2}\right) \frac{205,9707}{1,447663} = 135.1642$$

When the demand is forecasted by using CR in RStudio, the values of the arguments except for \$frc.in and \$frc.out are the same as Figure 20. Forecasts using CR are calculated by using Equation 7. Same as SBA, demand columns refer to Z_t and interval columns to P_t . The forecast for Product 11 by using CR for 25th months is calculated as:

$$Y_{25} = \frac{205,9707}{1,447663} = 142,28$$

CR and SBA do not provide a point forecast. Instead, they give the average demand per period. For example, if the demand size estimate is 205.9707, and the interval between nonzero demands estimate is 1.447663. It is likely to get demand for 205,97 units for Product 11 in the next 1,45 months. The forecast obtained from CR is 142,28 units per month. Thus, it can be said that for the next 12 months, the demand would likely to be $142,28 \times 12 = 1.707,36$ units.

For generating forecasts for all SKUs in both Dataset A and Dataset B, RStudio codes are written for running SES, CR, and SBA methods. For some selected parameters, see Appendix 4.

Once after the forecasts are generated, the best performing method and parameters for each SKU in both datasets are chosen based on the forecast accuracies obtained by using “smooth” and “Metrics” packages of RStudio. “Accuracy” function of “smooth” package provides accuracy measures as sMAE, sMSE, sCE and sPIS while the MASE, S-MAPE and RMSE values can be calculated by the help of “Metrics” package.

All these accuracies obtained from the relevant RStudio functions are organized via Excel for evaluation. For Dataset A, an accuracy table including 16.856 rows for 301 SKUs and seven columns for selected accuracy measures is formed. Similarly, for Dataset B, an accuracy table with 136.136 rows and seven columns is constructed.

Table 9 shows partially this organized accuracy table in MS Excel for Dataset A. The method column of this table refers to the codes created by the author. Each code is constructed as having four components: SKU number (P1), accuracy function (acc), initial value (m for mean, n for naïve and o for optimum), forecast method (CR, SBA or SES) and α value (005 – 030) respectively. As mentioned earlier, optimum parameters that minimize the selected cost function (MAR, MAE, MSR, and MSE) can be calculated by using “*tsintermittent*” package. This optimization process can be applied both by using a single α parameter or two different α parameters for demand size and interval estimates in CR and SBA. mar, mae, msr, or mse represents the cost function that is used for the optimization of the parameters, and 1 or 2 shows the number of α parameters used. Accuracies of forecasting methods for each SKU are controlled one by one. Based on different measures, the best performing method and parameters are selected (i.e., α and initial values) for both datasets. For example, in Table 9, the best performing forecasting method for the SKU coded as P1 is SBA with an α value of 0,05 and the mean initial values. This selection is made based on considering all of the accuracy measures since the best performing method is not always the same for all of them. Except for sCE and sPIS, smaller values show that the forecast is more accurate and for sCE and sPIS, which are bias measures, the values that are close to 0 are preferable.

Table 9: Partial Accuracy Table for Dataset A

METHOD	MASE	SMAPE	RMSE	sMAE	sMSE	sCE	sPIS
P1.acc.m.cr005	1,050	0,711	435,704	0,379	0,151	1,693	17,430
P1.acc.m.cr010	1,160	0,692	536,686	0,419	0,229	3,753	30,818
P1.acc.m.cr015	1,361	0,734	600,953	0,491	0,287	4,737	37,213
P1.acc.m.cr020	1,475	0,757	635,869	0,533	0,321	5,232	40,430
P1.acc.m.cr025	1,540	0,770	656,199	0,556	0,342	5,511	42,246
P1.acc.m.cr030	1,585	0,779	670,595	0,572	0,357	5,706	43,511
P1.acc.m.ses005	1,204	0,700	554,475	0,435	0,244	4,038	32,669
P1.acc.m.ses010	1,447	0,751	627,182	0,523	0,312	5,111	39,643
P1.acc.m.ses015	1,735	0,809	713,035	0,627	0,403	6,266	47,153
P1.acc.m.ses020	2,068	0,871	808,020	0,747	0,518	7,468	54,965
P1.acc.m.ses025	2,427	0,931	914,726	0,876	0,664	8,762	63,376
P1.acc.m.ses030	2,821	0,990	1035,802	1,019	0,851	10,186	72,631
P1.acc.m.sba005	1,049	0,717	429,068	0,379	0,146	1,487	16,089
P1.acc.n.sba010	1,116	0,685	518,036	0,403	0,213	3,441	28,788
P1.acc.n.sba015	1,189	0,697	549,366	0,429	0,239	3,957	32,145
P1.acc.n.sba020	1,211	0,701	556,751	0,437	0,246	4,073	32,900
P1.acc.n.sba025	1,199	0,699	552,629	0,433	0,242	4,009	32,480
P1.acc.n.sba030	1,176	0,694	543,808	0,424	0,235	3,868	31,568
P1.acc.o.crmse2	7,209	1,374	2468,770	2,603	4,835	26,028	175,609
P1.acc.o.sbamse1	1,096	0,681	509,602	0,396	0,206	3,294	27,832
P1.acc.o.sbamse2	9,150	1,464	3117,436	3,304	7,710	33,037	221,165
P1.acc.o.sesmsr1	1,142	0,689	529,018	0,412	0,222	3,627	29,996

4.4. BUILDING A JUDGMENTAL FORECASTING MODEL

As stated earlier, judgmental forecasts are widely used in practice. In this study, judgmental forecasts which are generated by the company by using executive opinions are used. Both companies rely on the judgmental forecasts generated by the executives instead of using any of the statistical techniques. The subjective opinions of the experts from a single or different department are collected. These forecasts may be directly

used or averaged to get the final value. Executive opinions are easy to use and do not need any statistical background. If the opinions are collected from a group of executives, there may be some problems like the pressure of strong leadership to other participants. Their ideas may be affected by the ideas of others as well.

Company A judgmentally forecasts the demand relying on the experiences of its sales executive. Based on the feedback from their customers, they forecast the annual demand for the SKUs that they are responsible for. Monthly demand is calculated by dividing the annual demand into 12. The company had judgmental forecasts only for 12 months of 2018.

Judgmental forecasts of Company B are generated by the purchasing specialists. Using their past experience, they decide the number of spare parts to order from their suppliers. The purchasing records of the previous 18 months from July 2017 to December 2018 for 3.639 SKUs are obtained from the company. However, only 1.205 of these records had the demand data. Thus, the remainder steps of this forecasting process use 1.205 SKUs for Dataset B.

4.5. COMBINING STATISTICAL AND JUDGMENTAL FORECASTS

To generate more accurate forecasts for the SKUs with intermittent demand pattern, this study aims to combine statistical forecasts that are generated by using the proposed methods in the literature for SKUs with such pattern (*Section 3.3*) with the judgmental forecasts of the companies which are based on the expert opinions (*Section 3.4*). It is appropriate to use judgmental forecasts when the time series is highly variable, which is the case in SKUs with intermittent demand pattern. Thus, it is expected that this combination process will increase forecast accuracy. When combining the constituent forecasts for these datasets, weighted averages procedures is used as this simple method performs better than the sophisticated combination strategies (see Clemen (1989), Sanders and Ritzman (2004).) Even using equal weights for the constituent forecasts (0,5 for statistical forecast and 0,5 for judgmental forecast) provides rather good performance (Clemen, 1989: 559). However, to see the effect of the different weights on accuracy level, weights ranging between 0.1 and 0.9 are used. Thus, nine different combinations are tried.

In combining stage, all calculations are done in MS Excel environment. As the judgmental forecasts provided by two companies (A, B) include the last 12 and 18 months, respectively, the combination procedure is applied to only this range of both datasets. Table 10 shows an example for combining statistical and judgmental forecasts of two SKUs (P1 and P2) from Dataset B based on the weighted average procedure. In this example, the weights for the statistical and judgmental forecasts are 0,6 and 0,4, respectively. The marked values refer to statistical, judgmental, and combined forecasts for P1, respectively. The combined forecast for P1, for instance, is calculated as: $(0,941 * 0,6) + (10 * 0,4) = 4,564$

All of the resulting combination tables are imported to RStudio to calculate the forecast accuracy for each SKU. For the best performing statistical methods, judgmental forecasts provided by the companies and all combinations with different weights; MASE, RMSE, and bias are calculated again by using “*Metrics*” package of RStudio. These results are evaluated for both datasets. Finally, for each SKU, the forecasting method that provides the highest accuracy is selected among the combined, statistical, and judgmental forecasts.

Table 10: An Example of Combining Forecasts for Dataset B

	Statistical Forecasts		Judgmental Forecasts		Combination (0,6S+0,4J)	
Date	P1	P2	P1	P2	P1	P2
07.2017	0,941	103,598	0	0	0,564	62,159
08.2017	0,941	103,598	0	0	0,564	62,159
09.2017	0,941	103,598	0	0	0,564	62,159
10.2017	0,941	103,598	0	0	0,564	62,159
11.2017	0,941	103,598	0	0	0,564	62,159
12.2017	0,941	103,598	0	0	0,564	62,159
01.2018	0,941	103,598	0	0	0,564	62,159
02.2018	0,941	103,598	5	300	2,564	182,159
03.2018	0,941	103,598	0	223	0,564	151,359
04.2018	0,941	103,598	10	0	4,564	62,159
05.2018	0,941	103,598	0	232	0,564	154,959
06.2018	0,941	103,598	0	0	0,564	62,159
07.2018	0,941	103,598	0	0	0,564	62,159
08.2018	0,941	103,598	0	0	0,564	62,159
09.2018	0,941	103,598	0	0	0,564	62,159
10.2018	0,941	103,598	0	400	0,564	222,159
11.2018	0,941	103,598	0	400	0,564	222,159
12.2018	0,941	103,598	0	0	0,564	62,159

4.6. RESULTS/FINDINGS

In the previous chapters, the proposed framework for combining statistical and judgmental forecasting methods and the implementation process of this framework on two real-world spare parts datasets are discussed. First, the data is prepared for the forecasting process, and it is categorized into demand patterns (smooth, intermittent, erratic and lumpy) by using SBC scheme. In the next stage, three different methods which are widely used for forecasting the demand of SKUs with intermittent demand, CR, SBA, and SES, are used to generate the statistical forecasts. The performance of these methods is evaluated to find out the best performing method and parameters for each SKU in both datasets. In the evaluation process, accuracy measures MASE, S-MAPE, RMSE, sMAE, sMSE and bias measures sCE, and sPIS are used. The judgmental forecasts are obtained from the companies. Later, the forecasts that are generated by using the selected statistical methods for each SKU and the judgmental forecasts are combined by using weighted averages. Finally, the accuracies of the selected statistical forecasting methods, the judgmental forecasts, and the combination forecasts are evaluated to determine the best performing model within others. The final evaluation is made based on MASE, RMSE, and bias.

In this chapter, the results and findings of the application process will be explained and analyzed for Dataset A and Dataset B.

In the proposed framework for combining forecasts, the stage after preparing the data for the forecasting purposes is the categorization of the demand data into demand patterns by using SBC categorization scheme. This method suggests to group SKUs into four classes (smooth, intermittent, lumpy, and erratic) based on their p and CV^2 values. By using MS Excel, p , and CV^2 values for each SKU in both datasets are calculated. Based on the conditions given in Table 1 in *Section 1.2.4*, the SKUs in both datasets are categorized. Table 11 shows the number of SKUs that belong to each demand category according to the SBC categorization.

Table 11: SBC Categorization for Datasets A and B

Number of SKUs in Each Demand Pattern		
Demand Pattern	Dataset A (301 SKUs)	Dataset B (2.431 SKUs)
Erratic	111	582
Intermittent	25	802
Lumpy	138	658
Smooth	27	389

After categorizing the demand data, the statistical forecasts are generated by using SES, CR, and SBA with different α and initial value parameters. By using the accuracy metrics MASE, S-MAPE, RMSE, sMAE, sMSE, sCE and sPIS the accuracy of these methods is compared for both datasets. For each SKU, the best performing forecasting method and parameters are selected based on these measures. This selection process is explained in *Section 3.3* in detail. The results for each dataset are analyzed based on the demand pattern of the SKUs. The number of SKUs that perform best for each forecasting method and parameter is used to assess the performance of these methods and parameters.

First, the initial values are considered based on SBC demand patterns. In total, 106 SKUs performed best with mean, 87 SKUs with naïve and 108 SKUs with optimal initial parameters. The results showed that all initial parameters perform similarly. Thus, it can be said that none of them outperforms the other.

Table 12 visualizes the number and percentage of the SKUs that perform best for each demand pattern for Dataset A based on each initial value type (naïve or mean) and optimum parameters. Majority of the SKUs that have erratic and lumpy demand generate the most accurate forecasts by using optimal parameters. Except for the SKUs with smooth demand, mean initials perform better than the naïve ones. In total, 106 SKUs performed best with mean, 87 SKUs with naïve and 108 SKUs with optimal initial parameters. The results showed that all initial parameters perform similarly. Thus, it can be said that none of them outperforms the other.

Table 12: Evaluation: Initial Values for Demand Patterns (Dataset A)

Initials	Number of SKUs	% of SKUs
ERRATIC (111)		
Mean	39	35%
Naive	33	30%
Optimum	39	35%
INTERMITTENT (25)		
Mean	12	48%
Naive	6	24%
Optimum	7	28%
LUMPY (138)		
Mean	49	36%
Naive	36	26%
Optimum	53	38%
SMOOTH (27)		
Mean	6	22%
Naive	12	44%
Optimum	9	33%

Next, the performance of three forecasting methods used in the implementation process is evaluated based on the demand patterns. Table 13 summarizes the number and percentage of best-performing forecasting methods according to each demand pattern for Dataset A. Syntetos et al. (2005: 499), recommended to use SBA for intermittent, lumpy, and erratic demand patterns and CR for smooth demand pattern. Table 13 demonstrate that for erratic, intermittent, and lumpy demand the majority of the best performing statistical forecasting method is SBA. Also, 63% of the SKUs that have smooth demand achieved the highest accuracy when they are forecasted by using CR. These results are consistent with the results of Syntetos et al. (2005: 499). However, this is not valid for all the SKUs in a demand pattern. For example, 24 of the 138 SKUs that have lumpy demand generated the most accurate forecasts by using CR even though the authors suggested to use SBA for this pattern.

Table 13: Evaluation: Statistical Methods for Demand Patterns (Dataset A)

Forecast Methods	Number of SKUs	% of SKUs
ERRATIC (111)		
CR	8	7%
SBA	94	85%
SES	9	8%
INTERMITTENT (25)		
CR	4	16%
SBA	17	68%
SES	4	16%
LUMPY (138)		
CR	24	17%
SBA	97	70%
SES	17	12%
SMOOTH (27)		
CR	17	63%
SBA	8	30%
SES	2	7%

Finally, the α parameter values that performed best for each demand type in Dataset A are investigated.

Table 14 demonstrates the numbers and percentages of α values selected when using predetermined parameters (between 0,05-0,30) and cost function (MAE, MAR, MSE, MSR) and the number of α values (1 or 2) used when forecasting with optimum parameters. In the literature, the suggested interval for smoothing constant α is 0,05-0,20 for the SKUs with intermittent demand (see Croston (1972), Syntetos and Boylan, (2005)). Most of the SKUs have the highest accuracy level with α values within the range (193 SKUs – 66,1%).

Table 14: Evaluation: Alpha Values for Demand Patterns (Dataset A)

Alpha Value	Number of SKUs	% of SKUs	Alpha Value	Number of SKUs	% of SKUs
ERRATIC (111)			LUMPY (138)		
0,05	38	34%	0,05	32	23%
0,1	10	9%	0,1	17	12%
0,15	2	2%	0,15	11	8%
0,2	7	6%	0,2	5	4%
0,25	2	2%	0,25	7	5%
0,3	13	12%	0,3	13	9%
α^*_{mae1}	12	11%	α^*_{mae1}	12	9%
α^*_{mae2}	14	13%	α^*_{mae2}	9	7%
α^*_{mar1}	3	3%	α^*_{mar1}	6	4%
α^*_{mar2}	2	2%	α^*_{mar2}	3	2%
α^*_{mse1}	3	3%	α^*_{mse1}	5	4%
α^*_{mse2}	4	4%	α^*_{mse2}	10	7%
α^*_{msr1}	1	1%	α^*_{msr1}	5	4%
			α^*_{msr2}	3	2%
INTERMITTENT (25)			SMOOTH (27)		
0,05	5	20%	0,05	11	41%
0,1	2	8%	0,1	2	7%
0,15	2	8%	0,2	1	4%
0,2	2	8%	0,3	4	15%
0,25	3	12%	α^*_{mae1}	1	4%
0,3	4	16%	α^*_{mae2}	2	7%
α^*_{mae1}	1	4%	α^*_{mar2}	2	7%
α^*_{mae2}	1	4%	α^*_{mse2}	4	15%
α^*_{mar1}	1	4%			
α^*_{mar2}	2	8%			
α^*_{mse2}	2	8%			

α^* refers to optimal smoothing parameters calculated by minimizing the cost functions

The results in Table 14 demonstrates that $\alpha=0,05$ is the most frequently used (86 SKUs – 28,6%). However, some SKUs achieved the highest accuracy level for α values that are above 0,20. Larger values of smoothing constant put more weight on the most recent observation, in other words, the model pays more attention to the most recent data, and the values near zero allow past observations to have a significant influence. For the optimum parameters, α values are even larger than 0,30. 108 of the 301 SKUs performed best with optimal α parameters. This may be caused by the selected α range which is 0,05-0,30, since the α value that shows the best performance is not in this given range.

The same analyses are applied to Dataset B as well. Table 15 shows the number and percentage of each initial value for each demand type. For the majority of SKUs that have erratic, intermittent, and lumpy demand patterns, mean initial values perform better.

Table 15: Evaluation: Initial Values for Demand Patterns (Dataset B)

Initials	Number of SKUs	% of SKUs
ERRATIC (582)		
Mean	284	49%
Naive	258	44%
Optimum	40	7%
INTERMITTENT (802)		
Mean	362	45%
Naive	342	43%
Optimum	98	12%
LUMPY (658)		
Mean	356	54%
Naive	257	39%
Optimum	45	7%
SMOOTH (389)		
Mean	150	39%
Naive	204	52%
Optimum	35	9%

For this dataset, 1.002 SKUs (41,2%) performed better with mean initial, 1.061 SKUs (43,6%) performed better with naïve initials and 368 SKUs (15,2%) performed better with optimal parameters. In smooth demand pattern, naïve initials are better. These outcomes are similar to results obtained from Dataset A. However, like in Dataset A; there is no single best performing initial value option for any of the demand patterns.

Next, the performance of the statistical forecasting methods used is evaluated based on each demand pattern. Table 16 demonstrates the number and percentages of forecasting methods selected for each demand pattern. Majority of the SKUs in each group have the best accuracy values when they are forecasted by using SBA (1.493 SKUs – 61,4%). However, it is not the only method that performs well. Surprisingly, SES is the second method that provided the best results (720 SKUs – 29,6% of the whole dataset) for the SKUs in this dataset while CR is the worst performing one (218 SKUs – 9%). The results obtained for Dataset B are not consistent with the suggestions of the SBC categorization scheme that the authors propose to use CR for the SKUs with smooth demand pattern.

Table 16: Evaluation: Statistical Methods for Demand Patterns (Dataset B)

Forecast Methods	Number of SKUs	% of SKUs
ERRATIC (582)		
CR	48	8%
SBA	446	77%
SES	88	15%
INTERMITTENT (802)		
CR	69	9%
SBA	452	56%
SES	281	35%
LUMPY (658)		
CR	45	7%
SBA	314	48%
SES	299	45%
SMOOTH (389)		
CR	56	14%
SBA	281	72%
SES	52	13%

Lastly, the performances of the α parameter values are evaluated. Table 17 represents the number and percentage of each α value that performs the best for each demand type. Opposite to the literature that suggests using small α values; for this dataset larger ones performed better for most SKUs. For all demand types, $\alpha=0,30$ generated the forecasts with the highest accuracy for the majority of the SKUs.

Table 17: Evaluation: Alpha Values for Demand Patterns (Dataset B)

Alpha Value	Number of SKUs	% of SKUs	Alpha Value	Number of SKUs	% of SKUs
ERRATIC (582)			LUMPY (658)		
0,05	41	7%	0,05	88	13%
0,1	33	6%	0,1	47	7%
0,15	37	6%	0,15	47	7%
0,2	42	7%	0,2	67	10%
0,25	69	12%	0,25	61	9%
0,3	320	55%	0,3	303	46%
mae1	9	2%	mae1	19	3%
mae2	2	0%	mae2	11	2%
mar1	1	0%	mar2	3	0%
mse1	8	1%	mse1	2	0%
mse2	16	3%	mse2	6	1%
msr1	1	0%	msr1	2	0%
msr2	3	1%	msr2	2	0%
INTERMITTENT (802)			SMOOTH (389)		
0,05	145	18%	0,05	54	14%
0,1	87	11%	0,1	49	13%
0,15	78	10%	0,15	35	9%
0,2	81	10%	0,2	51	13%
0,25	56	7%	0,25	35	9%
0,3	257	32%	0,3	130	33%
mae1	29	4%	mae1	9	2%
mae2	29	4%	mae2	6	2%
mar1	7	1%	mar1	2	1%
mar2	4	0%	mar2	1	0%
mse1	7	1%	mse1	3	1%
mse2	12	1%	mse2	7	2%
msr1	8	1%	msr1	2	1%
msr2	2	0%	msr2	5	1%

α^* refers to optimal smoothing parameters calculated by minimizing the cost functions

Next stage of the proposed framework is the combination process of the selected statistical forecasting methods that are selected for all SKUs in both datasets with the judgmental forecasts which are provided by Company A and Company B. Forecast accuracy measures MASE, RMSE and bias are calculated for each combination obtained by assigning different weights on statistical and judgmental forecasts, best performing statistical forecasts and the judgmental forecast. To decide which forecasting method performs better, given accuracy measures are compared for all SKUs. Table 18 is the partial representation of the accuracy table used selecting the best method for each SKU in Dataset A. The column “statistical” shows the accuracy measures for selected statistical forecasts, “judgmental” shows the accuracy of the judgmental forecasts given by the company and others show the accuracy of combined forecasts. For instance, “0,4S+0,6J” means that for the statistical forecast, the weight of 0,4 and for the judgmental forecast weight of 0,6 is assigned. The example SKUs given in Table 18 performs best with different combinations and methods. When deciding on which method to choose, all measures are considered together. For instance, for P215, MASE has the lowest value for the combination of 2S + 8J. However, RMSE and bias have the lowest value for the combination of equal weights. Thus, the combined forecast with equal weights is selected.

Table 18: Partial Accuracy Table: All Forecasts (Dataset A)

SKU Number	Accuracy Measure	Statistical	Judgmental	0,1S - 0,9J	0,2S - 0,8J	0,3S - 0,7J	0,4S - 0,6J	0,5S - 0,5J	0,6S - 0,4J	0,7S - 0,3J	0,8S - 0,2J	0,9S - 0,1J
P22	MASE	0,886	0,813	0,814	0,720	0,705	0,705	0,705	0,708	0,727	0,763	0,814
	RMSE	589,172	549,354	522,834	502,242	488,329	481,673	482,576	490,995	506,555	528,628	556,437
	Bias	340,069	265,167	204,643	144,119	83,596	23,072	37,451	97,975	158,499	219,022	279,546
P80	MASE	0,968	0,738	0,938	0,729	0,756	0,787	0,817	0,847	0,877	0,907	0,938
	RMSE	26,570	24,215	24,074	24,019	24,050	24,169	24,372	24,659	25,025	25,469	25,985
	Bias	11,364	3,083	1,639	0,194	1,251	2,696	4,140	5,585	7,030	8,474	9,919
P129	MASE	0,666	0,705	0,669	0,693	0,688	0,683	0,680	0,677	0,674	0,671	0,669
	RMSE	9,431	10,034	9,966	9,899	9,833	9,770	9,709	9,649	9,592	9,536	9,483
	Bias	3,168	4,667	4,517	4,367	4,217	4,067	3,918	3,768	3,618	3,468	3,318
P131	MASE	0,807	0,803	0,807	0,803	0,804	0,804	0,805	0,805	0,806	0,806	0,807
	RMSE	8,567	8,434	8,447	8,461	8,474	8,487	8,500	8,513	8,527	8,540	8,554
	Bias	4,982	4,750	4,773	4,796	4,820	4,843	4,866	4,889	4,912	4,935	4,959
P215	MASE	0,645	0,575	0,631	0,564	0,568	0,575	0,584	0,593	0,604	0,618	0,631
	RMSE	8,753	8,727	8,678	8,641	8,615	8,600	8,597	8,606	8,626	8,657	8,699
	Bias	1,643	1,500	1,186	0,871	0,557	0,243	0,072	0,386	0,700	1,015	1,329

Table 19 summarizes the number of selected SKUs for each method. For comparison, average p and CV^2 values are calculated for each method. These results show that Company A's judgmental forecasts which are generated by the sales executives performed best only for 40 SKUs out of 301 (13,3%). Selected statistical forecasts among SES, CR and, SBA perform best for 155 SKUs (51,5%) and, the combined forecasts generated the most accurate forecasts for 106 SKUs (35,2%). Most of the SKUs in this dataset have the highest accuracy when the forecasts are generated by using statistical methods. For Dataset A, the performance of the combined forecasts is below expectations. The reason of this situation may be the unstructured way that the company generates its forecasts. More qualified judgmental forecasting methods would increase the accuracy of combination forecasts.

Table 19: Performance of Combined Forecasts Based on p and CV^2 (Dataset A)

Method	# of	Average	
	SKUs	Average p	CV^2
0,1S - 0,9J	11	1,377	1,025
0,2S - 0,8J	6	1,274	0,934
0,3S - 0,7J	4	1,191	0,838
0,4S - 0,6J	15	1,516	0,964
0,5S - 0,5J	14	1,630	0,903
0,6S - 0,4J	11	1,405	1,094
0,7S - 0,3J	14	1,333	0,894
0,8S - 0,2J	18	1,477	0,822
0,9S - 0,1J	13	1,345	0,720
Judgmental	40	1,502	0,988
Statistical	155	1,478	0,858
Average	301	1,463	0,892

To inspect the relationship between the weights assigned to statistical forecasts and values of the average p and CV^2 parameters, regression analysis is applied by using MS Excel. The detailed regression tables are demonstrated in Appendix 5.

Firstly, the regression analysis is held for determining the relationship between the weights assigned to statistical forecasts, and average CV^2 values of the SKUs that

belong to each forecasting method. The results showed that the Multiple R value for this regression is 0,5681. It is found out that this shows a negative relationship by calculating the correlation coefficient. R^2 equals to 0,3227, which means that CV^2 can explain 32,3% of the changes in the weights assigned to the statistical forecasts. Significance F value is 0,0683, which shows a significant relationship under 0,90 confidence interval.

Secondly, the regression between the weights and average p values of the SKUs for each group is investigated. The results showed that there is no significant relationship between these variables for 0,95 confidence interval (Multiple R=0,1292, $R^2= 0,0167$ and p-value=0,7049).

Lastly, both the average p and CV^2 are included in the regression analysis. The multiple R value is 0,6144, and R^2 is 0,3775. Significance F value is 0,1502. There is no significant relationship between p and CV^2 and the weights assigned to the statistical forecasts in 0,95 confidence interval for the Dataset A.

Table 20 is a partial representation of the table which is used for selecting the best-performing methods for Dataset B. By evaluating three accuracy measures, the methods that generate the most accurate forecasts are selected for each SKU.

Table 20: Partial Accuracy Table: All Forecasts (Dataset B)

SKU Number	Accuracy Measure	Statistical	Judgmental	0,1S - 0,9J	0,2S - 0,8J	0,3S - 0,7J	0,4S - 0,6J	0,5S - 0,5J	0,6S - 0,4J	0,7S - 0,3J	0,8S - 0,2J	0,9S - 0,1J
P65	MASE	0,839	1,894	1,686	1,478	1,277	1,112	0,988	0,915	0,868	0,832	0,821
	RMSE	10,294	22,751	20,767	18,841	16,992	15,248	13,650	12,255	11,138	10,392	10,097
	Bias	2,013	20,389	18,149	15,908	13,668	11,428	9,188	6,948	4,707	2,467	0,227
P114	MASE	0,960	1,483	1,280	1,077	0,874	0,735	0,690	0,671	0,679	0,737	0,823
	RMSE	6,785	9,784	8,618	7,515	6,508	5,647	5,009	4,686	4,742	5,165	5,876
	Bias	5,407	6,833	5,609	4,385	3,161	1,937	0,713	0,511	1,735	2,959	4,183
P775	MASE	0,791	2,258	2,053	1,847	1,641	1,435	1,230	1,024	0,897	0,821	0,795
	RMSE	2,606	6,642	6,070	5,511	4,967	4,446	3,955	3,507	3,122	2,824	2,644
	Bias	0,153	6,111	5,485	4,858	4,232	3,605	2,979	2,353	1,726	1,100	0,473
P934	MASE	2,385	2,397	1,936	1,511	1,150	0,893	0,799	0,912	1,178	1,523	1,937
	RMSE	7,936	7,972	6,651	5,404	4,296	3,461	3,127	3,444	4,269	5,372	6,617
	Bias	7,294	7,333	5,871	4,408	2,945	1,482	0,020	1,443	2,906	4,369	5,831
P1135	MASE	19,765	0,810	1,465	3,305	5,363	7,420	9,478	11,535	13,593	15,650	17,708
	RMSE	16,298	1,054	1,313	2,842	4,491	6,165	7,848	9,535	11,224	12,914	14,606
	Bias	16,277	0,667	1,028	2,722	4,416	6,111	7,805	9,500	11,194	12,888	14,583

In Table 21, the number of SKUs selected for each method, and the average p and CV^2 values of these methods are given. These results show that, for Dataset B, the majority of the SKUs do perform better if judgmental and statistical forecasts are combined. Nevertheless, the combinations that assign heavier weight on the statistical forecasts result in better accuracy levels for more SKUs. For Dataset B, the company's judgmental forecasts, which are generated by the purchasing specialists, perform the best for 29 SKUs (2,4%) of 1.205 SKUs. The statistical forecasts are selected as the best-performing ones are the best for 402 SKUs (33,4%) of the dataset. For the 772 SKUs (64,2%), the most accurate forecasts are generated by using the combination process. The methods that put the weight of more and equal to 0,5 on the statistical forecasts performed better for 666 SKUs (55,3%).

Table 21: Performance of Combined Forecasts Based on p and CV^2 (Dataset B)

Method	# of SKUs	Average	
		Average p	CV^2
0,1S - 0,9J	10	1,755	0,805
0,2S - 0,8J	19	2,193	0,690
0,3S - 0,7J	38	1,786	0,980
0,4S - 0,6J	41	1,886	0,709
0,5S - 0,5J	120	1,866	0,728
0,6S - 0,4J	68	1,357	0,842
0,7S - 0,3J	109	1,374	0,736
0,8S - 0,2J	162	1,304	0,744
0,9S - 0,1J	207	1,270	0,583
Judgmental	29	2,442	0,791
Statistical	402	1,438	0,555
Average	1205	1,4884	0,6636

The regression analysis is applied to Dataset B as well to investigate the relationship between the weights assigned to statistical forecasts and average p and CV^2 values of the SKUs that take place in each group. The detailed regression tables can be found in Appendix 6.

First, the relationship between the weights assigned to statistical forecasts and CV^2 values are inspected. The multiple R value is 0,5824. Calculating the coefficient correlation showed that there is a moderate negative relationship between the variables. R^2 is 0,3392, which means that 33,92% of the changes in the weights are related to the values of the CV^2 . The significance F value in the ANOVA table is 0,06, and there is a significant relationship between the weight and the CV^2 value for 0,90 confidence interval.

Next, the regression calculations are discussed for the relationship between the weights assigned to statistical forecasts and average p values. Multiple R value equals to 0,8547 and, the sign of this relationship is determined as negative. The R^2 value is 0,7305, which means that 73,05% of the changes in weights assigned to statistical forecasts are related to the average p value of the SKUs. Also, significance F =0,0008 shows that this is a significant relationship in 0,95 confidence interval.

Finally, the regression analysis is applied for the weights assigned to statistical forecasts, and both average CV^2 and p values. R^2 (0,8791) shows that these independent variables cause 87,91% of the changes in the dependent variable. P-values indicated that both average p and CV^2 are significant. Finally, significance F (0,0002) shows that this relationship is significant when the confidence interval is determined as 0,95.

CONCLUSION

In order to survive in today's competitive world, effective management of the inventory is crucial for the institutions. Accurate forecasting is one of the most important aspects of inventory management. Companies hold different SKUs on hand which have different demand patterns. Most of the finished goods and work-in-process items have smooth demand pattern while some SKUs, like spare parts, have intermittent demand. Due to the high number of zeros and variability in the size of non-zero demand in the historical demand data of a SKU, it is hard to generate accurate forecasts for SKUs with intermittent demand. Previous studies showed that traditional statistical forecasting methods such as MA, SES, ARIMA, etc. perform poorly on the SKUs that have intermittent demand. In order to generate accurate forecasts, there are numbers of methods proposed in the literature for intermittent demand forecasting (details are given in *Section 1.3*). On the other hand, judgmental forecasting methods are widely used by the companies in practice. These methods are useful in situations that the historical data is absent and the forecasting environment and the time series are variable (see *Section 1.4* for the details of judgmental forecasting methods). Also, statistical and judgmental forecasting methods can be combined by using various methods to generate more accurate forecasts. In the literature, there are some studies that covered the issue of judgmentally adjusting the statistical forecasts. However, combining the statistical and judgmental forecasts for the SKUs with intermittent demand pattern is a field that the researchers did not focus as far as we know.

The aim of this study is to conduct a forecasting model that combines statistical and judgmental forecasting methods in order to generate more accurate forecasts for the SKUs that have intermittent demand pattern. Proposing a forecasting model which is easy to implement and interpret for the practitioners who do not have a strong statistical background was the secondary goal of the study.

To achieve the goals of this study, a framework that combines statistical and judgmental forecasts is proposed. Structure of proposed framework includes categorization of the demand according to the demand pattern by using appropriate categorization schemes, building a statistical forecasting model and judgmental forecasting model in parallel, combining the selected statistical forecasts with the

judgmental forecasts and evaluating the results for determining the best performing method for each SKU among statistical, judgmental and combined forecasts.

First, the demand data is categorized by using SBC categorization scheme which is proposed by Syntetos et al. (2005: 499). The results also showed that, the proposed methods for this categorization scheme do not work completely for the datasets that are used in the study. Results for Dataset A showed that majority the SKUs in each category have the best forecast accuracy level for the forecasting methods that are suggested by Syntetos et al. (2005: 499). However, there are many SKUs in each demand pattern that have better accuracies with the other methods. For Dataset B, the majority of the SKUs with smooth demand pattern performed better with SBA, not with CR as suggested. Also, unexpectedly SES generated accurate forecasts for the SKUs in all categories. For the categorization phase, it can be concluded that SBC categorization scheme provided conflicted results for Dataset A and Dataset B. Other schemes could provide better results. Also, to get more accurate forecasts, instead of using the suggested forecasting methods based on the demand patterns, the forecasting methods can be selected for each SKU individually.

After the categorization, the forecasting performance of each SKU is inspected. In the literature, there are some studies which defend that SBA and CR outperforms SES. However, for both datasets, SES generated quite accurate forecasts. For Dataset A, 10,6% and Dataset B, for 30% of the 2.431 SKUs SES is selected as the best performing method. Though using SES is not suggested in the literature, it has a good performance for these datasets. That is why this method should not be overlooked. The performance of all forecasting methods that are in consideration should be evaluated. Another issue is the suggested α value range of 0,05-0,30 in the literature. For 62,3% of the Dataset A, α values are selected within this range. 15,2% of the 301 SKUs generates the best forecasts by using $\alpha=0,25$ and $\alpha=0,30$. 37,7% of the SKUs perform best with the optimal parameters. The optimal α values change between very small values close to 0 and very large values close to 1. For Dataset B, 91% of the 2.431 SKUs have the α value within the range. However, 51% of the dataset have the best performance with $\alpha=0,25$ and $\alpha=0,30$. 8% of the SKUs have the best performance by using optimal parameters. These results show that, though most of the SKUs have high accuracy level for the α values that are within the range of 0,05-0,30, the α value to be

used in the forecasting process should be selected individually instead of using the same value for all SKUs in the dataset. Because, the best performing α values are not always small (0,05-0,20) as it is suggested in the literature. Using small α values may result in a decrease in the accuracy of the forecasts.

In this study, judgmental forecasts are provided by the companies. For Company A and Company B, these forecasts are generated without using any structured method but only using expert opinions. Using structured methods may provide more accurate forecasts.

After building statistical and judgmental forecasting models, forecasts obtained from each method are combined for both datasets. Weighted averages are used in the combination process, which is a simple method that generates accurate forecasts. Considering the Dataset A, unfortunately the combination of the forecasts did not meet the expectations. For 51,5% of the SKUs, statistical forecasting outperformed the others. 13% of the SKUs performed best with the judgmental forecasts of the company. Combined forecasts are better for 35,5% of the SKUs. For this dataset, the statistical forecasts and the combination that have higher or equal weight on statistical forecasts perform better. The results showed that, both the statistical forecasts and combined forecasts perform better than the company's forecasts. Combined forecasts are evaluated for the Dataset B as well. Statistical forecasting methods perform best for 33,4% of the 1.205 SKUs. Also, the combinations that have equal or higher weight on statistical forecasts (55,3%) have good performance for the SKUs. The judgmental forecasts provided by the company perform best only for 2,4% of the 1205 SKUs. For Dataset B, the combined forecasts showed the best performance among others, which meets with the expectations.

In order to investigate the relationship between the weights assigned to statistical forecasts and p and CV^2 values, regression analysis is held for both datasets. The results for Dataset A show that weight assigned to statistical forecast and CV^2 are negatively related (Multiple $R = 0,5681$ and $R^2 = 0,3227$). There is no significant relationship between the weights and p . The relation between the weights and both p and CV^2 is analyzed and the results showed that there is no significant relationship between the variables. For the Dataset B, the regression analysis showed that under a negative relationship between assigned weights and CV^2 (Multiple $R = 0,5824$ and $R^2 = 0,3391$).

=0,3392), between the weights and p value (Multiple R=0,8547 and $R^2=0,7305$) and between weight and p and CV^2 (Multiple R=0,9376 and $R^2=0,8791$). It can be concluded for both datasets that if the CV^2 is higher, less weight should be assigned to the statistical forecast. Also, Dataset B shows that for higher p values, lower weights should be preferred and both p and CV^2 should be considered when specifying the weight of statistical forecasts.

In the literature there are number of studies that focuses on statistical forecasting methods for intermittent demand data, judgmental adjustments of the statistical forecasts and combining forecasts for the demand data with normal pattern. However, this study is the first that focuses on the combination of statistical and judgmental forecasts for non-normal demand patterns. The results for both datasets showed that combining both forecasts may result in better accuracy level.

This study also showed that higher α values may result in high forecast accuracy level opposite to the predetermined values in the literature. Instead of using the fixed values, the best performing values should be determined and used for each SKU. Using optimal parameters can be an option to save time in statistical forecasting process. Another result of this study shows that, though the methods that are developed for intermittent demand forecasting performs quite well, SES is also successful in this context. So, instead of using a single method, different methods should be tried and the best performing one should be chosen.

The results of this study must be interpreted in the light of some limitations. First, neither Company A nor Company B provided any details for the SKUs in the dataset like what is the SKU, if it is a new one or not used anymore, if it should be grouped with other SKUs or not, etc. Thus, all the calculations are done assuming that the SKUs are independent from each other and they are actively used for the given periods. More detail about the SKUs would help to categorize and arrange the dataset for forecasting process, which would provide better accuracy levels.

Second limitation is for Dataset A which is occurred due to the size of Dataset A. This dataset included demand information for 36 months. For time-series forecasting, data that contains information for longer periods help to provide better forecasts.

The third limitation of the study appeared due to one of the primary goals of this study, which was conducting a forecasting model which is easy to apply and interpret by the users in the real-world companies who do not have strong statistical background. Also, it is aimed that the model would not cause any additional costs for the companies. Thus, instead of using more sophisticated methods like ANN, which generates more accurate forecasts, simpler methods that also show good performance (SES, CR and SBA) are used. These methods do not require any software to apply.

Fourth limitation is the way that the judgmental forecasts are generated both by Company A and Company B. Neither of these companies use a structured method, like Delphi Method or jury of executive opinion, for forecasting. Their forecasts, which are most likely to be biased, were based on the experience of an executive. In most companies, the both purchasing and sales executives have pressure on them while they make their projections. This situation causes the bias in their forecasts. However, using an organized method that includes more executives would generate better judgmental forecasts. Due to the limitation in time, we were not able to guide to companies to generate these forecasts by using such structured methods.

For the future research, different schemes can be tried for the categorization process of the demand data. For the statistical forecasting part, in order to generate a simple forecasting model for the practitioners, easier methods that generate accurate forecasts are used. Also, since there are no studies that focusing on the combining forecasts for SKUs with intermittent demand, using simple methods would be helpful to understand the effect of combining for such pattern without any additional complexity. In the future works, more advanced methods such as bootstrapping or artificial neural networks can be used in statistical forecasting stage. These methods may provide more accurate forecasts. Judgmental forecasts which are generated by using an organized method like Delphi, analogies or jury of executive opinion can be used. If quality of the judgmental forecasts increases, the quality of the combined forecasts will increase as well. Studies may include the process of generating structured judgmental forecasts with companies with the lead of the researcher. Also, the inventory control perspective is crucial for inventory management. In the future, instead of using forecast accuracy measures, the performances can be compared by

using stock control perspective. This will provide a more realistic analysis of statistical, judgmental and combined forecasts.



REFERENCES

- Ahmet, M. (2016). Seven reasons why you need to forecast in supply chain. Retrieved January 25, 2019, from <https://www.supplychaindigital.com/top-10/seven-reasons-why-you-need-forecast-supply-chain>
- Armstrong, J. S. (1985). Long-range forecasting: from crystal ball to computer (Second edition ed.). New York: Wiley-Interscience.
- Armstrong, J. S. (2001). Combining Forecasts. In J. S. Armstrong, Principles of Forecasting: A Handbook for Researchers and Practitioners (pp. 417-439). Dordrecht: Kluwer Academic Publishers.
- Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1), 69-80.
- Babiloni, E., Cardós, M., Albarracín, J. M., & Palmer, M. E. (2010). Demand categorisation, forecasting, and inventory control for intermittent demand items. *South African Journal of Industrial Engineering*, 21(2), 115-130.
- Bartezzaghi, E., & Kalchschmidt, M. (2011). The impact of aggregation level on lumpy demand management. In N. Altay, & L. A. Litteral (Eds.), *Service Parts Management* (First Edition ed., pp. 89-104). London: Springer-Verlag.
- Batchelor, R., & Kwan, T. Y. (2007). Judgemental bootstrapping of technical traders in the bond market. *International Journal of Forecasting*, 23(3), 427-445.
- 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.

Boylan, J., & Syntetos, A. A. (2007). The accuracy of a modified Croston procedure. *International Journal of Production Economics*, 107(2), 511-517.

Bunn, D. W., & Salo, A. A. (1993). Forecasting with scenarios. *Journal of Operational Research*, 291-303.

Bunn, D., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods: issues and analysis. *Management Science*, 37(5), 501-518.

Callegaro, A. (2010). Forecasting methods for spare parts demand. Three-Year Degree Thesis . Padova: Universita' Degli Studi di Padova Engineering Faculty.

Caniato, F., Kalchschmidt, M., & Ronchi, S. (2009). Integrating quantitative and qualitative forecasting approaches: organizational learning in an action research case. *Journal of the Operational Research Society*, 62(3), 413-424.

Carbone, R., Andersen, A., Corriveau, Y., & Corson, P. P. (1983). Comparing for different time series methods the value of technical expertise individualized analysis, and judgmental adjustment. *Management Science*, 29(5), 559-566.

Clemen, R. T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559-583.

Constantino, F., Di Gravio, G., Patriarca, R., & Petrella, L. (2018). Spare parts management for irregular demand items. *Omega*, 81, 57-66.

Cook, T., Falchi, P., & Mariano, R. (1984). An urban allocation model combining time series and analytic hierarchical methods. *Management Science*, 30(2), 198-208.

Croston, J. D. (1972, September). Forecasting and stock control for intermittent demands. *Operational Research Quarterly* (1970-1977), 23(3), 289-303.

Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: the case of judgmental adjustments to sku-level demand forecasts. *International Journal of Forecasting*, 29(3), 510-522.

Eaves, A. H. (2002, July). Forecasting for the ordering and stock-holding of consumable spare parts. PhD Thesis. Lancaster, UK: Department of Management Science, The Management School, Lancaster University.

Eaves, A. H., & Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55(4), 431–437.

Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.

Elgers, P. T., Lo, M. H., & Murray, D. (1995). Note on adjustments to analysts' earning forecasts based upon systematic cross-sectional components of prior-period errors. *Management Science*, 41(8), 1392-1396.

Eroğlu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, 26(1), 116-133.

Fildes, R., Ma, S., & Kolassa, S. (2019, October). Retail forecasting: research and practice. Retrieved July 02, 2019, from https://mpra.ub.uni-muenchen.de/89356/1/MPRA_paper_89356.pdf

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, 8(1), 81-98.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for

improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3-23.

Fischer , I., & Harvey, N. (1999). Combining forecasts: what information do judges need to outperform the simple average? *International Journal of Forecasting*, 3(15), 227-246.

Franses, P. H., & Legerstee, R. (2011). Combining SKU-level sales forecasts from models and experts. *Expert Systems with Applications*, 38(3), 2365-2370.

Gardner, E. S., & Koehler, A. B. (2005). Comments on a patented bootstrapping method for forecasting intermittent demand. *International Journal of Forecasting*, 21(3), 617-618.

Gelders, L. F., & van Looy, P. M. (1978). An inventory policy for slow and fast movers in a petrochemical plant: a case study. *The Journal of the Operational Research Society*, 29(9), 867-874.

Ghobbar , A. A., & Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30(14), 2097-2114.

Ghobbar, A. A., & Friend, C. H. (2002). Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8(4), 221-231.

Goodwin, P. (2000). Correct or combine? Mechanically integrating judgmental forecasts with statistical methods. *International Journal of Forecasting*, 16(2), 261-275.

Gönül, M. S. (2013). A popular and important tool for forecast users: judgmental adjustments. *H.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi*, 31(2), 73-79.

Graefe, A., Armstrong, J. S., Jones Jr., R. J., & Cuzan, A. G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, 30(1), 43-54.

Green, K. C., & Armstrong, J. S. (2007). Structured Analogies for Forecasting . *International Journal of Forecasting*, 365-376.

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

Hanke, J. E., & Wichern, D. (2014). *Business Forecasting* (Ninth Edition ed.). Essex: Pearson Education Limited.

Hasni, M., Aguir, M. S., & Jemai, Z. (2018). Spare parts demand forecasting: a review on bootstrapping methods. *International Journal of Production Research*, 1-14.

Hasni, M., Babai, M. Z., Aguir, M. S., & Jemai, Z. (2017). An investigation on bootstrapping forecasting methods for intermittent demands. *International Journal of Production Economics*, In Press, 1-10.

Heinecke, G., Syntetos, A. A., & Wang, W. (2013). Forecasting-based SKU classification. *International Journal of Production Economics*, 143(2), 455-462.

Heizer, J., & Render, B. (2014). Forecasting Demand. In J. Heizer, & B. Render, *Operations Management* (pp. 137-188). Essex: Pearson Education Limited.

Hu, Q., Boylan, J. E., Chen, H., & Labib, A. (2018). OR in spare parts management: A review. *European Journal of Operational Research*, 266(2), 395-414.

Hua, Z. S., Zhang, B., Yang, J., & Tan, D. S. (2007). A new approach of forecasting intermittent demand for spare parts inventories in the process industries. *Journal of the Operational Research Society*, 58(1), 52-61.

Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight*(4), 43-46.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. (Second Edition), 323-348. Melbourne: Otexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.

Johnston, F. R., & Boylan, J. E. (1996). Forecasting for items with intermittent demand. *The Journal of the Operational Research Society*, 47(1), 113-121.

Johnston, F. R., Boylan, J. E., & Shale, E. A. (2003). An examination of the size of orders from customers, their characterisation and the implications for inventory control of slow moving items. *Journal of the Operational Research Society*, 54(8), 833-837.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Science*, 39(1), 17-31.

Kaya, G. O. (2015, February). *Forecasting intermittent demand using COX process*. PhD Thesis . İstanbul: Graduate School of Sciences and Engineering, Fatih University.

Kostenko, A. V., & Hyndman, R. J. (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society*, 57(10), 1256-1257.

Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. *International Journal of Production Economics*, 143(1), 198-206.

Kourentzes, N. (2014). On intermittent demand model optimisation and selection. *International Journal of Production Economics*, 156(1), 180-190.

Kourentzes, N., & Petropoulos, F. (2016, March 10). Package 'tsintermittent'. Retrieved from The Comprehensive R Archive Network: <https://cran.r-project.org/web/packages/tsintermittent/tsintermittent.pdf>

Lawrence, M., Goodwin, P., O'Connor, M., & Önköl, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), 493 – 518.

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, 23(3), 377-390.

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

Lim, J. S., & O'Connor, M. (1995). Judgmental adjustment of initial forecasts: its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149-168.

Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., & Gucci, S. (2017). Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics*, 118, 116-128.

Mahmoud, E. (1989). Combining Forecasts: Some Managerial Issues. *International Journal of Forecasting*(5), 599-600.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4), 527-529.

Mancuso, A. C., & Werner, L. (2013). Review of combining forecasts approaches. *Independent Journal of Management & Production*, 4(1), 248-277.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology*, 4(4), 268-273.

Nagaria, P. (2017, December 7). Forecasting intermittent demand: Traditional smoothing approaches versus the Croston method. Retrieved from Strata Data Conference: <https://conferences.oreilly.com/strata/strata-sg/public/schedule/detail/62953>

Porras, E., & Dekker, R. (2008). An inventory control system for spare parts at a refinery: an empirical comparison of different re-order point methods refinery. *European Journal of Operational Research*, 184(1), 101-132.

Pour, A. N., Tabar, B. R., & Rahimzadeh, A. (2008). A hybrid neural network and traditional approach for forecasting lumpy demand. *International Journal of Industrial and Manufacturing Engineering*, 2(4), 1028-1034.

Rao, A. V. (1973). A comment on: forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 24(4), 639-640.

Reinmuth, J. E., & Geurts, M. D. (1972). A bayesian approach to forecasting the effects of atypical situations. *Journal of Marketing Research*, 9(3), 292-297.

Sanders, N. R. (1992). Accuracy of judgmental forecasts: A comparison. *Omega*, 20(3), 353-364.

Sanders, N. R., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, 5(1), 39-52.

Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 405-416). Dordrecht: Kluwer Academic Publishers.

Sanders, N. R., & Ritzman, L. P. (2004). Integrating judgmental and quantitative forecasts: methodologies for pooling marketing and operations information. *International Journal of Operations & Production Management*, 24(5), 514-529.

Sani, B., & Kingsman, B. G. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48(7), 700-713.

Schultz, C. R. (1987). Forecasting and inventory control for sporadic demand under periodic review. *The Journal of the Operational Research Society*, 38(5), 453-458.

Silver, E. A., Pyke, D. F., & Peterson, R. (1998). *Inventory management and production planning and scheduling* (Third Edition ed.). New York: John Wiley & Sons, Inc.

Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, 28(2), 485-496.

Snyder, R., Koehler, A., & Ord, J. (2002). Forecasting for inventory control with exponential smoothing. *Journal of Forecasting*, 18(1), 5-18.

Song, H., Gao, B. Z., & Lin, V. S. (2013). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, 29(2), 295-310.

Syntetos, A. A. (2001). *Forecasting of intermittent demand*. PhD Thesis. West London, UK: Brunel University, Business School Buckinghamshire Chilterns University College.

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

Syntetos, A. A., Babai, M. Z., & Gardner Jr., E. S. (2015). Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. *Journal of Business Research*, 68(8), 1746–1752.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, 56(5), 495-503.

Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., & Goodwin, P. (2009). The effects of integrating management judgement into intermittent demand forecasts. *Int. J. Production Economics*, 118(1), 72-81.

Şahin, M. (2011, July). Forecasting aviation spare parts demand. Master Thesis. İstanbul, Turkey: The Graduate Institute of Sciences and Engineering, Fatih University.

Tavares, L. V., & Almeida, L. T. (1983). A binary decision model for the stock control of very slow moving items. *Journal of the Operational Research Society*, 34(3), 249-252.

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

Teunter, R., & Duncan, L. (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, 60(3), 321-329.

Thomasson, S. A. (2017, July). Improving forecast performance of LPG demand. Master Thesis. Twente: University of Twente - Industrial Engineering and Management & Production and Logistics Management.

Turner, D. S. (1990). The role of judgement in macroeconomic forecast. *Journal of Forecasting*, 9(4), 315-345.

van Kampen, T. J., Akkerman, R., & van Donk, D. P. (2012). SKU classification: a literature review and conceptual framework. *International Journal of Operations & Production Management*, 32(7), 850-876.

Viswanathan, S., & Zhou, C. X. (2008). A new bootstrapping based method for forecasting and safety stock determination for intermittent demand items. Working paper. Singapore, Singapore: Nanyang Business School, Nanyang Technological University.

Waller, D. (2015, February 20). Methods for Intermittent Demand Forecasting. Retrieved from Lancaster University: http://www.lancaster.ac.uk/pg/waller/pdfs/Intermittent_Demand_Forecasting.pdf

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, X., & Petropoulos, F. (2016). To select or to combine? The inventory performance of model and expert forecasts. *International Journal of Production Research*, 54(17), 5271-5282.

Webby, R., & O'Connor, M. (1996). Judgmental and Statistical Time Series Forecasting: A Review of the Literature. *International Journal of Forecasting*(12), 91-118.

Willemain, T. R. (1989). Graphical adjustment of statistical forecasts. *International Journal of Forecasting*, 5(2), 179-185.

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.

Willemain, T. R., Smart, C. N., Shockor, J. H., & DeSautels, P. A. (1994). Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, 10(4), 529-538.

Williams, T. M. (1984). Stock control with sporadic and slow-moving demand. *The Journal of the Operational Research Society*, 35(10), 939-948.

Wilson, J. H., & Keating, B. (2008). Introduction to Business Forecasting. In J. H. Wilson, & B. Keating, *Business Forecasting with ForecastX* (pp. 1-55). New York: McGraw Hill-Irwin.

Wolfe, C., & Flores, B. (1990). Judgmental adjustment of earnings forecasts. *Journal of Forecasting*, 9(4), 389-405.

Worthen, B. (2003). Contrary to what vendors tell you, computer systems alone are incapable of producing accurate forecasts. Retrieved January 16, 2019, from CIO: https://www.cio.com.au/article/168757/future_results_guaranteed/

Wright, T. C. (2013, August 25). Real Life Examples of Qualitative Forecasting. Retrieved from Chron: <https://smallbusiness.chron.com/real-life-examples-qualitative-forecasting-72990.html>

Yüksel, S. (2007). An integrated forecasting approach to hotel demand. *Mathematical and Computer Modelling*, 46(7-8), 1063-1070.

Zhang, G., Patuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14(1), 35-62.

Zhou , C., & Viswanathan, S. (2011). Comparison of a new bootstrapping method with parametric approaches for safety stock determination in service parts inventory systems. *International Journal of Production Economics*, 133(1), 481-485.

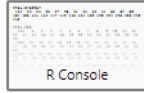




APPENDICES

APPENDIX 1: R Output of SBC Categorization Scheme for Dataset A

RStudio: Notebook Output



```
Sidx.croston
[1] 23 25 26 27 30 31 35 52 54 82 87 89 104 108 111 112 147 149 158 165 189 190 193 198 208 230 250

Sidx.sba
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 24 28 29 32 33
[28] 34 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 53 55 56 57 58 59 60 61 62 63
[55] 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 83 84 85 86 88 90 91 92 93
[82] 94 95 96 97 98 99 100 101 102 103 105 106 107 109 110 113 114 115 116 117 118 119 120 121 122 123 124
[109] 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 148 150 151 152 153
[136] 154 155 156 157 159 160 161 162 163 164 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182
[163] 183 184 185 186 187 188 191 192 194 195 196 197 199 200 201 202 203 204 205 206 207 209 210 211 212 213 214
[190] 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 231 232 233 234 235 236 237 238 239 240 241 242
[217] 243 244 245 246 247 248 249 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
[244] 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297
[271] 298 299 300 301

Sidx.ses
NULL

Scv2
[1] 0.6265026 3.7161740 1.2211573 0.5028845 0.8348111 0.7440952 1.0135987 1.3267810 1.8822985 1.2748309
[11] 0.6508599 0.8665376 0.9212822 0.8604162 1.9036091 0.5603244 0.7690552 0.4559415 0.9891721 2.2558754
[21] 0.9582156 1.5825286 0.2413482 0.5355848 0.3642042 0.3676949 0.4599247 1.3085472 0.5842931 0.3360880
[31] 0.3319062 0.5236105 0.7008672 0.5393284 0.3739748 0.9882828 0.9825872 0.5349683 1.0741197 1.2620809
[41] 0.6279906 0.7409061 1.4718444 1.2500344 0.9339229 1.6806016 0.7059378 0.7964847 0.5157993 0.6987387
[51] 2.0374823 0.4541468 1.0967034 0.4276996 1.7954805 1.1735404 0.8106745 1.1087977 2.0810347 1.5381385
[61] 2.1486826 1.1752363 0.5429066 1.0017023 0.9578592 0.7780249 0.6889535 0.7039341 2.4439092 0.5015444
[71] 2.5609804 0.7397302 1.2000135 1.9512268 0.9374474 1.0225786 1.1983235 1.3886316 0.5181390 0.5588156
[81] 0.9854854 0.3556553 0.6957302 0.8466738 0.7291190 1.9083659 0.4581162 0.6897410 0.4182595 0.7374379
[91] 0.7695039 0.8132596 0.6616004 0.8790078 0.6063068 1.6535538 2.1805681 0.8367187 0.3784304 0.8087649
[101] 0.9010695 0.5902148 0.5709651 0.4261229 0.5476144 0.3746556 1.0041544 0.4504278 0.5279512 1.1827802
[111] 0.4325983 0.3792761 0.4057914 0.3966684 0.4927358 2.5953568 0.4874549 0.7706167 1.1384016 0.7206252
[121] 1.1500000 0.2631069 0.9960485 0.8964913 1.3383046 1.4241614 1.3081014 0.6785366 1.0486529 0.8499145
[131] 0.5768773 0.3562286 1.0641253 0.8068653 1.8223631 0.6726897 0.6972022 1.1722886 0.5564370 1.0465068
[141] 0.8187631 0.9274938 1.2049412 0.5257993 1.1505611 0.5494353 0.3687411 0.8826541 0.3402125 0.4758248
[151] 0.6382401 0.4759868 0.6082663 0.4924527 0.7337088 0.6923730 0.8048792 0.4129537 0.3755205 0.5005931
[161] 0.9413837 0.9784546 0.3347062 0.8616404 0.4064430 1.2293269 0.6999057 0.8860337 0.6617647 1.2821758
[171] 1.1749835 1.0595105 1.2286296 1.0421829 1.7975867 1.6298435 0.8178906 0.7798089 2.0639906 0.5302511
[181] 0.5216995 0.6963229 0.6082251 0.6549019 0.7301654 0.5130450 0.4432757 0.6405504 0.4456295 0.3134266
[191] 0.6396622 0.5602216 0.3600653 0.5385255 0.6014944 0.7748909 0.3028921 0.4126546 0.5221088 2.2324417
[201] 0.0663242 0.9162248 0.6346311 0.8989603 0.7906542 0.9032699 0.6800769 0.3433615 0.5899026 0.4092745
[211] 0.5860110 0.6253056 0.5185030 0.8341356 0.7628303 0.7372559 0.7583795 0.7892316 0.7036235 0.6368378
[221] 0.6896876 0.0480659 0.8313029 1.6636616 1.6094051 1.5063761 0.3710044 1.4425400 1.5093057 0.3049143
[231] 0.7148012 1.0852184 2.2710404 1.0992642 1.0635488 0.9029816 0.5369707 0.4609280 0.5051088 1.2980769
[241] 1.5373956 0.6249270 0.5290930 0.5455753 0.6277178 0.5407307 0.7723191 1.3701316 0.6534417 0.4776950
[251] 0.4915873 0.7423574 0.6131668 0.6901593 0.7489102 0.5837321 0.6155802 0.5949428 0.8971672 0.68554709
[261] 0.4269407 0.6194341 0.7225784 0.9257143 0.8823106 0.7166597 0.4601376 1.1904772 0.5117730 1.4560018
[271] 0.8710891 0.6179647 3.2291667 0.3241070 0.8214946 0.9602693 0.4897370 0.8722222 0.8670422 0.8742102
[281] 0.3903038 0.6548518 0.6134813 1.0749537 0.8000000 0.7089947 1.0447917 0.1853780 0.5113769 0.4561867
[291] 1.6866454 1.2550679 0.7494015 0.7827655 0.4956268 0.5620502 0.6915087 0.3174603 0.7273663 1.4909214
[301] 0.6387424

Sp
[1] 1.241379 1.125000 1.333333 1.714286 1.161290 1.333333 1.500000 1.440000 1.500000 1.800000 1.384615 1.440000
[13] 1.028571 1.565217 1.440000 1.500000 1.161290 1.454545 1.333333 1.636364 1.185185 1.125000 1.161290 1.031250
[25] 1.090909 1.090909 1.241379 1.125000 1.058824 1.185185 1.230769 1.103448 1.125000 1.241379 1.142857 1.172414
[37] 1.125000 1.161290 1.125000 1.200000 1.100000 1.161290 1.166667 1.230769 1.285714 1.200000 1.241379 1.200000
[49] 1.103448 1.161290 1.400000 1.090909 1.346154 1.090909 1.333333 1.107143 1.200000 1.440000 1.333333 1.241379
[61] 1.384615 1.440000 1.241379 1.200000 1.320000 1.285714 1.241379 1.192308 1.296296 1.384615 1.166667 1.161290
[73] 1.161290 1.750000 1.200000 1.416667 1.500000 1.028571 1.166667 1.166667 1.161290 1.161290 1.028571 1.894737
[85] 1.028571 1.440000 1.090909 1.200000 1.125000 1.285714 1.125000 1.090909 1.161290 1.416667 1.250000 1.090909
[97] 1.333333 4.000000 1.384615 1.129032 1.590909 1.161290 1.894737 1.166667 1.440000 2.125000 1.400000 1.241379
[109] 1.800000 1.789474 1.066667 1.200000 1.565217 1.440000 1.161290 2.117647 1.333333 1.894737 1.650000 1.458333
[121] 1.125000 1.400000 1.714286 1.636364 2.571429 1.888889 2.250000 2.400000 1.440000 2.571429 2.571429 2.461538
[133] 1.384615 1.285714 1.125000 1.206897 1.333333 1.030303 1.062500 1.241379 1.125000 1.565217 1.565217 1.619048
[145] 1.241379 1.333333 1.285714 1.333333 1.161290 1.333333 1.800000 1.346154 1.161290 1.384615 1.241379 1.058824
[157] 1.894737 1.058824 1.400000 1.285714 1.360000 1.285714 1.360000 1.166667 1.200000 1.800000 1.200000 1.200000
[169] 2.000000 1.440000 1.028571 1.161290 1.125000 1.161290 1.241379 1.333333 1.090909 1.129032 1.440000 1.028571
[181] 1.125000 1.296296 1.296296 1.200000 1.200000 1.058824 1.478261 1.333333 1.028571 1.090909 1.333333 1.384615
[193] 1.200000 1.384615 1.478261 1.161290 2.266667 1.250000 1.440000 1.125000 1.700000 1.200000 1.125000 1.090909
[205] 1.200000 1.636364 1.619048 1.307692 1.206897 1.400000 1.200000 1.440000 1.241379 1.333333 1.285714 1.714286
[217] 1.285714 1.440000 1.241379 1.636364 1.333333 1.296296 1.360000 1.500000 1.565217 1.285714 1.636364 1.440000
[229] 1.636364 1.285714 1.384615 1.285714 1.333333 1.285714 1.384615 1.894737 1.307692 1.384615 1.440000 1.333333
[241] 1.333333 1.241379 1.200000 1.714286 2.000000 1.307692 1.440000 1.714286 1.440000 1.285714 1.894737 2.250000
[253] 1.894737 1.736842 1.384615 2.400000 2.117647 2.538462 1.500000 2.250000 1.888889 1.440000 1.700000 4.125000
```

\$idx.croston: Indicates the SKUs which should be forecasted by using CR

\$idx.sba: Indicates the SKUs which should be forecasted by using SBA

\$scv2 and \$p: Indicates CV² and p of each SKU respectively

APPENDIX 2: R Formulation of *sexsm* Function (for SES)

```
sexsm(data, h=10, w=NULL, init=c("mean","naive"),  
cost=c("mar","msr","mae","mse"), init.opt=c(TRUE,FALSE),  
outplot=c(FALSE,TRUE), opt.on=c(FALSE,TRUE), na.rm=c(FALSE,TRUE))
```

data: Intermittent demand time series

h: Forecast horizon

init: Initial values for demand and intervals. This can be: 1. x - Numeric value for the initial level; 2. "naive" - Initial value is a naive forecast; 3. "mean" – Initial value is equal to the average of data.

cost: Cost function used for optimization: 1. "mar" - Mean Absolute Rate; 2. "msr" - Mean Squared Rate; 3. "mae" - Mean Absolute Error; 4. "mse" - Mean Squared Error.

init.opt: If init.opt==TRUE then initial values are optimized.

outplot: If TRUE a plot of the forecast is provided.

opt.on: This is meant to use only by the optimization function. When opt.on is TRUE then no checks on inputs are performed.

na.rm: A logical value indicating whether NA values should be remove using the method

APPENDIX 3: R Formulation of *crost* Function (for CR and SBA)

```
crost(data, h=10, w=NULL, init=c("mean","naive"), nop=c(2,1),  
type=c("croston","sba","sbj"), cost=c("mar","msr","mae","mse"),  
init.opt=c(TRUE,FALSE), outplot=c(FALSE,TRUE), opt.on=c(FALSE,TRUE),  
na.rm=c(FALSE,TRUE))
```

data: Intermittent demand time series

h: Forecast horizon

w: Smoothing parameters. If w = NULL then parameters are optimized. If w is a single parameter then the same is used for smoothing both the demand and the intervals. If two parameters are provided then the second is used to smooth the intervals.

init: Initial values for demand and intervals. This can be: 1. c(z,x) - Vector of two scalars, where first is initial demand and second is initial interval; 2. "naive"- Initial demand is first non-zero demand and initial interval is first interval; 3. "mean" - Same as "naive", but initial interval is the mean of all in sample intervals.

nop: Specifies the number of model parameters. Used only if they are optimized. 1. 1 - Demand and interval parameters are the same; 2. 2 - Different parameters.

type: Croston's method variant: 1. "croston" Croston's method; 2. "sba" Syntetos Boylan approximation; 3. "sbj" Shale-Boylan-Johnston.

cost: Cost function used for optimization: 1. "mar" - Mean Absolute Rate; 2. "msr" - Mean Squared Rate; 3. "mae" - Mean Absolute Error; 4. "mse" - Mean Squared Error.

init.opt: If init.opt==TRUE then initial values are optimized.

outplot: If TRUE a plot of the forecast is provided.

opt.on: This is meant to use only by the optimization function.

APPENDIX 4: R Codes for Forecasting the Demand

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Statistical Forecasting.Rmd x Forecasts.R* x
Source on Save Run Source
1 library(readxl)
2 library("tsintermittent")
3 library("forecast")
4 library(xlsx)
5 library(tsenssembler)
6 library(Metrics)
7 s.train_t <- read_excel("C:/Users/Gizem/Desktop/DATASETS/Delphi Train Son.xlsx")
8 s.test_t <- read_excel("C:/Users/Gizem/Desktop/DATASETS/Delphi Test Son.xlsx")
9 n.cr005 <- apply(s.train_t, 2, crost, h = 12, w = 0.05, init = c("naive"), init.opt=FALSE)
10 n.cr010 <- apply(s.train_t, 2, crost, h = 12, w = 0.10, init = c("naive"), init.opt=FALSE)
11 n.cr015 <- apply(s.train_t, 2, crost, h = 12, w = 0.15, init = c("naive"), init.opt=FALSE)
12 n.cr020 <- apply(s.train_t, 2, crost, h = 12, w = 0.20, init = c("naive"), init.opt=FALSE)
13 n.cr025 <- apply(s.train_t, 2, crost, h = 12, w = 0.25, init = c("naive"), init.opt=FALSE)
14 n.cr030 <- apply(s.train_t, 2, crost, h = 12, w = 0.30, init = c("naive"), init.opt=FALSE)
15 o.crmr1 <- apply(s.train_t, 2, crost, h = 12, nop= c(1), cost =c("mar"), init.opt=TRUE)
16 o.crmr2 <- apply(s.train_t, 2, crost, h = 12, nop= c(2), cost =c("mar"), init.opt=TRUE)
17 o.cmr1 <- apply(s.train_t, 2, crost, h = 12, nop= c(1), cost =c("msr"), init.opt=TRUE)
18 o.cmr2 <- apply(s.train_t, 2, crost, h = 12, nop= c(2), cost =c("msr"), init.opt=TRUE)
19 o.cmae1 <- apply(s.train_t, 2, crost, h = 12, nop= c(1), cost =c("mae"), init.opt=TRUE)
20 o.cmae2 <- apply(s.train_t, 2, crost, h = 12, nop= c(2), cost =c("mae"), init.opt=TRUE)
21 o.crmse1 <- apply(s.train_t, 2, crost, h = 12, nop= c(1), cost =c("mse"), init.opt=TRUE)
22 o.crmse2 <- apply(s.train_t, 2, crost, h = 12, nop= c(2), cost =c("mse"), init.opt=TRUE)
23 n.sba005<-apply(s.train_t, 2, crost,type=c("sba"), h=12, w=0.05,init=c("naive"),init.opt=FALSE)
24 n.sba010<-apply(s.train_t, 2, crost,type=c("sba"), h=12, w=0.10,init=c("naive"),init.opt=FALSE)
25 n.sba015<-apply(s.train_t, 2, crost,type=c("sba"), h=12, w=0.15,init=c("naive"),init.opt=FALSE)
26 n.sba020<-apply(s.train_t, 2, crost,type=c("sba"), h=12, w=0.20,init=c("naive"),init.opt=FALSE)
27 n.sba025<-apply(s.train_t, 2, crost,type=c("sba"), h=12, w=0.25,init=c("naive"),init.opt=FALSE)
28 n.sba030<-apply(s.train_t, 2, crost,type=c("sba"), h=12, w=0.30,init=c("naive"),init.opt=FALSE)
29 o.sbamr1<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(1),cost=c("mar"),init.opt=TRUE)
30 o.sbamr2<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(2),cost=c("mar"),init.opt=TRUE)
31 o.sbamr1<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(1),cost=c("msr"),init.opt=TRUE)
32 o.sbamr2<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(2),cost=c("msr"),init.opt=TRUE)
33 o.sbmae1<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(1),cost=c("mae"),init.opt=TRUE)
34 o.sbmae2<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(2),cost=c("mae"),init.opt=TRUE)
35 o.sbmse1<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(1),cost=c("mse"),init.opt=TRUE)
36 o.sbmse2<-apply(s.train_t, 2, crost,type=c("sba"), h=12, nop=c(2),cost=c("mse"),init.opt=TRUE)
37 n.ses005<-apply(s.train_t, 2, sexsm, h=12, w=0.05,init=c("naive"),init.opt=FALSE)
38 n.ses010<-apply(s.train_t, 2, sexsm, h=12, w=0.10,init=c("naive"),init.opt=FALSE)
39 n.ses015<-apply(s.train_t, 2, sexsm, h=12, w=0.15,init=c("naive"),init.opt=FALSE)
40 n.ses020<-apply(s.train_t, 2, sexsm, h=12, w=0.20,init=c("naive"),init.opt=FALSE)
41 n.ses025<-apply(s.train_t, 2, sexsm, h=12, w=0.25,init=c("naive"),init.opt=FALSE)
42 n.ses030<-apply(s.train_t, 2, sexsm, h=12, w=0.30,init=c("naive"),init.opt=FALSE)
43 o.sesmar1<-apply(s.train_t, 2, sexsm, h=12, cost=c("mar"),init.opt=TRUE)
44 o.sesmr1<-apply(s.train_t, 2, sexsm, h=12, cost=c("msr"),init.opt=TRUE)
45 o.sesmae1<-apply(s.train_t, 2, sexsm, h=12, cost=c("mae"),init.opt=TRUE)
46 o.sesmse1<-apply(s.train_t, 2, sexsm, h=12, cost=c("mse"),init.opt=TRUE)
47 print("finished")
```

APPENDIX 5: Regression Tables for Dataset A Weights – p and CV²

Weights Assigned to Statistical Forecasts - Average CV2

(Dataset A)

<i>Regression Statistics</i>	
Multiple R	0,5681
R Square	0,3227
Adjusted R Square	0,2475
Standard Error	2,8771
Observation	11

ANOVA

					<i>Significance</i>
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>F</i>
Regression	1	35,5015	35,5015	4,2889	0,0683
Residual	9	74,4985	8,2776		
Total	10	110			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower %95</i>	<i>Upper %95</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	21,5268	8,0273	2,6817	0,0251	3,3678	39,6858	3,3678	39,6858
Average CV2	-18,1086	8,7441	-2,0710	0,0683	-37,8892	1,6719	-37,8892	1,6719

Weights Assigned to Statistical Forecasts - Average p

<i>Regression Statistics</i>	
Multiple R	0,1292
R Square	0,0167
Adjusted R Square	-0,0926
Standard Error	3,4667
Observation	11

ANOVA

					<i>Significance</i>
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>F</i>
Regression	1	1,8368	1,8368	0,1528	0,7049
Residual	9	108,1632	12,0181		
Total	10	110			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower %95</i>	<i>Upper %95</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	0,1445	12,4641	0,0116	0,9910	-28,0513	28,3403	-28,0513	28,3403
Average p	3,4396	8,7984	0,3909	0,7049	-16,4637	23,3430	-16,4637	23,3430

Weights Assigned to Statistical Forecasts - Average CV2 and p

<i>Regression Statistics</i>	
Multiple R	0,6144
R Square	0,3775
Adjusted R Square	0,2219
Standard Error	2,9257
Observation	11

ANOVA

					<i>Significance</i>
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>F</i>
Regression	2	41,5242	20,7621	2,4256	0,1502
Residual	8	68,4758	8,5595		
Total	10	110			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower %95</i>	<i>Upper %95</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	13,8223	12,2880	1,1249	0,2933	-14,5138	42,1583	-14,5138	42,1583
Average p	6,3294	7,5455	0,8388	0,4259	-11,0706	23,7293	-11,0706	23,7293
Average CV2	-19,4567	9,0358	-2,1533	0,0634	-40,2933	1,3799	-40,2933	1,3799

APPENDIX 6: Regression Tables Dataset B Weights – p and CV²

Weights Assigned to Statistical Forecasts - Average CV2

<i>Regression Statistics</i>	
Multiple R	0,5824
R Square	0,3392
Adjusted R Square	0,2658
Standard Error	0,2842
Observation	11,0000

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1,0000	0,3731	0,3731	4,6200	0,0601
Residual	9,0000	0,7269	0,0808		
Total	10,0000	1,1000			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower %95</i>	<i>Upper %95</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	1,7244	0,5760	2,9935	0,0151	0,4213	3,0275	0,4213	3,0275
Average CV2	-1,6497	0,7675	-2,1494	0,0601	-3,3859	0,0865	-3,3859	0,0865

Weights Assigned to Statistical Forecasts - Average p

<i>Regression Statistics</i>	
Multiple R	0,8547
R Square	0,7305
Adjusted R Square	0,7005
Standard Error	0,1815
Observation	11,0000

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1,0000	0,8035	0,8035	24,3937	0,0008
Residual	9,0000	0,2965	0,0329		
Total	10	1,1			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower %95</i>	<i>Upper %95</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	1,7424	0,2574	6,7684	0,0001	1,1600	2,3247	1,1600	2,3247
Average p	-0,7320	0,1482	-4,9390	0,0008	-1,0672	-0,3967	-1,0672	-0,3967

Weights Assigned to Statistical Forecasts - Average CV2 and p

<i>Regression Statistics</i>	
Multiple R	0,9376
R Square	0,8791
Adjusted R Square	0,8489
Standard Error	0,1289
Observation	11

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	0,9670	0,4835	29,0840	0,0002
Residual	8	0,1330	0,0166		
Total	10	1,1			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower %95</i>	<i>Upper %95</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	2,4370	0,2873	8,4837	0,0000	1,7746	3,0995	1,7746	3,0995
Average p	-0,6489	0,1086	-5,9769	0,0003	-0,8992	-0,3985	-0,8992	-0,3985
Average CV2	-1,1260	0,3591	-3,1358	0,0139	-1,9540	-0,2979	-1,9540	-0,2979

