ANALYTICAL EVALUATION OF TOP-DOWN VERSUS BOTTOM-UP FORECASTING STRATEGIES IN A PRODUCTION PLANNING ENVIRONMENT

HANDIK WIDIARTA

SCHOOL OF MECHANICAL AND AEROSPACE ENGINEERING

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2005

ABSTRACT

In today's ever-emerging marketplace, most manufacturers are forced to continuously seek for opportunities to improve their core competencies in order to satisfy consumers' demands at the right time, at the right place, and in the right quantity. Among many strategies proposed by practitioners and academics, product proliferation has grown to be one of the most adopted strategies in many diverse industries. The main idea of product proliferation is to increase market share and brand awareness with a minimum risk. However, from the production manager's point of view, this strategy requires more and more storage space to store the growing number of product varieties and thus might adversely affect the accuracy of the demand forecasts.

Many studies have been conducted in the past four decades to investigate how the demand forecasts should be done in such a case. Two forecasting strategies are generally proposed: top-down (TD) strategy and bottom-up (BU) strategy. The TD strategy, as its name implies, uses the history of the product family demand and disaggregates the result based on the historical demand proportion of each product in the family. The BU strategy, on the contrary, forecasts the product demand individually and combines the result to obtain the forecast for the family demand.

Unfortunately, among all of the papers published during this period, only a few were done under a production planning framework. This dissertation, therefore, attempts to address the issue of evaluating the relative effectiveness of TD and BU strategies for forecasting the demand at the subaggregate (product) level as well as the demand at the aggregate (family) level under the abovementioned framework. Two of the most commonly used time series forecasting techniques in this framework, exponential smoothing and moving average, are adopted for both strategies. Various forecasting situations meant to cover a broad spectrum of actual situations and several time series models representing many demand processes in a production planning environment are considered. In addition, this dissertation also deals with different statistical parameters of the demand time series such as, demand correlation between individual products, relative demand proportion of a product in the family, coefficient of the serial correlation term, and degree of product substitutability. In practice, all these parameters may change over time. Therefore, by varying these parameters, this dissertation attempts to evaluate how such

changes will affect (1) the difference in the variance of forecast error between the two strategies and (2) the preference to a particular strategy.

This dissertation addresses three stationary time series models, namely white noise, first-order moving average [MA(1)], and first-order autoregressive [AR(1)] processes. The analysis is carried out analytically for the case when the coefficient of the serial correlation term of demand time series for all individual products is identical. A simulation study is then performed for the case when this coefficient is not identical. One of the main conclusions of these studies is that the difference in the variance of forecast error between TD and BU strategies under white noise or MA(1) process is relatively insignificant.

The study is further extended by considering the case when the individual products in the family are *substitutable*. Product substitutability is a common phenomenon in retail merchandising industry, where the demand for a particular product is partly driven by the inventory level of another product with similar characteristics. In addition to the three stationary time series models used in the earlier parts, it also considers an integrated moving average of order one [IMA(1, 1)] process, a non-stationary time series model in which the application of exponential smoothing is known to be optimal. The key finding of this study is that the TD strategy outperforms the BU strategy in most of the cases and the margin tends to increase with the degree of product substitutability.

ACKNOWLEDGEMENTS

Many people have helped me in completing this doctoral dissertation. First and foremost, I wish to express my deepest gratitude to my supervisors A/P Rajesh Piplani and A/P S. Viswanathan, who have spent the past few years supporting my dissertation work. Dr. Vish continuously challenged me, critically questioned all parts of the research, and improved my analytical and writing skills along the way. He sets a rather high standard for me, for which I am thankful. I am also indebted to Dr. Rajesh for his patience, guidance, and irresistible encouragement over my tenure at Nanyang Technological University. He has inspired me to aim high and become a rigorous scholar. I feel honored to have been able to work and learn from both professors. Without their attention to every phase of my research, it would have been impossible to complete this dissertation as scheduled.

I also owe a great deal to my internal examiner, Ast/P Jiao Jianxin, and the two anonymous external examiners, whose useful comments and thoughtful suggestions have significantly improved the quality of this dissertation. Many thanks to Ms. Leong Yuit Lin and all my colleagues at the Centre for Supply Chain Management for their support over the past few years. It would not have been the same without them. I further acknowledge the doctoral program at the School of Mechanical and Aerospace Engineering as well as Nanyang Business School for providing me with the financial support and facilities throughout the study.

Last but not least, I would like to extend my sincerest gratitude to my family and in particular to my wife, Amelia, without whose patience, the completion of this thesis would have been impossible. She lifted up my confidence when I was down and helped me accomplish things that I believed impossible. I also wish to thank my parents for their unconditional love, pray, and support throughout this endeavor. My family has always stood by me; their pride in my success has made everything worthwhile.

Singapore, 1st October 2005

Handik Widiarta

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

Over the past decades, intense competition, rapid advancement of retailing technology, and changing consumer tastes have forced most manufacturers to rethink and redesign their business strategy. In order to survive in today's ever-emerging marketplace, leading manufacturers need to continuously seek for opportunities to improve their core competencies to satisfy consumers' demands at the right time, at the right place, and in the right quantity.

In addition to many strategies proposed by practitioners and academics such as, just-in-time, flexible manufacturing systems, responsive product development, etc., product proliferation has grown to be one of the most adopted strategies in many diverse industries. The practice of product proliferation is visibly evident, considering the number of new product introductions and available product varieties in the market[1]. Most consumer product manufacturers offer hundreds of stock keeping units (SKUs) with different features (e.g. flavor, size, color, etc.) to appeal to different buying motives. L. L. Bean, Inc. (Schleifer, 1993), Levi's & Co. (Achabal, McIntyre, Smith, and Kalyanam, 2000), and Procter & Gamble (Kotler and Armstrong, 2001), etc., are some of the leading companies that produce an array of products, each of which is tailored to accommodate the needs of a specific market segment.

The benefits of product proliferation look promising, at least from the marketing managers' point of view. Apart from the fact that it can stimulate sales from different

---

[1] As reported by Fisher, Hammond, Obermeyer, and Raman (1994), the number of new product introductions in the United States food industry alone has exploded from 2,000 in 1980 to 18,000 in 1991.

market segments, it is also recommended as an alternative means to fully utilize the excess production capacity, to generate short-term potential income, and to increase the market share and brand awareness, with minimum risk (Kahn, 1998a, and Chong, Ho, and Tang, 2002). However, the negative effects of product proliferation may also be staggering. The requirement for smaller storage space to store the growing numbers of SKUs has long been one of the major deterrents for companies to increase their product varieties[2]. Against this backdrop, more and more companies nowadays are finding it difficult to forecast demand accurately for all of these new varieties and thus are in greater danger of having high remnants and buffer inventories that are incurred not only in their own vicinities[3], but also in other companies' vicinities in the supply chain (Fisher *et al*., 1994 and Lee, 2004).

Several strategies have been proposed for companies to deal with this particular problem. They can generally be divided into two categories based on the way the improvement is made, namely external and internal operations. In the external operation, as its name implies, the improvement is generally concentrated on the value chain outside the companies' business process. This is done by initiating such strategic partnerships within the supply chains as synchronized replenishment, information sharing or collaborative forecasting. The rationale for strategic partnership is that by sharing critical information (such as end consumers' demands information from point-of-sales), the forecast accuracy at the upper echelon (i.e. manufacturers) can be improved and thus, lowering the buffer stock requirement and reducing the chance of inventory obsolescence. On the other hand, in the internal operation the improvement is generally focused on the companies' existing business process. For example, this can be done by improving the forecasting technique. Accurate forecasts allow schedulers to use capacity more efficiently,

---

[2] As revealed by Quelch and Kenny (1994), while the number of SKUs grew 16% each year from 1985 to 1992, the retail shelf space expanded by only 1.5% each year.

[3] Empirical studies on US seed companies found that shortening the product life cycle and expanding the product line increased the total inventory costs by 120.8%, increased the average inventory level (primarily due to added safety stock) by 56.2%, and increased the cost of carryover, stockout cost, and safety stock cost by 143%, 165%, and 119%, respectively (Dooley and Kurtz, 2001).

reduce customer response times, and cut inventories (Badinelli, 1990, Gardner, 1990, Lee, 1990, Makridakis and Hibon, 2000, and Moon, Mentzer and Thomas, 2000).

The focus of this research is on the internal operation of such a company that deals with a variety of products under the same product line/family[4]. In particular, this research attempts to address the issue of evaluating the relative performance of top-down (TD) and bottom-up (BU) forecasting strategies belonging to the so-called hierarchical planning system (HPS).

1.2 Hierarchical Planning System – A Brief Review

HPS is a decision support system whose development is motivated by the fact that different planning levels in an organization usually require different types of forecast information. HPS helps in providing forecast information customized to specific users in different management levels and organizational functions more effectively, in terms of lower forecast errors (Fliedner, 1999, 2001). The planning levels in HPS can generally be divided into two layers, namely aggregate and disaggregate (van Donselaar, 2002). The aggregate layer concerns with long-term (usually quarterly or annual) and large-scale resource allocations. The input for this layer is an aggregate forecast, which may represent the requirement for a common raw material that is shared by a group of products. The disaggregate layer, on the other hand, concerns with the details (mostly on daily or weekly basis) on how, where, and in what quantity each particular product should be produced. Thus, the input for this layer is an individual (disaggregate) demand forecast for each product in the family.

[4] Several products are said to be in one family if most of the materials or components used to build these products are identical. The example includes consumer products such as soap or shampoo. In this case, it is not uncommon to find that the manufacturers introduce a range of products with different features, such as fragrance. The main ingredients, and thus the price and the quality, however, remain the same.

The top-down (TD) and bottom-up (BU) forecasting strategies are part of the HPS. The TD forecasting strategy uses the history of the product family demand data[5] and disaggregates the result based on the historical demand proportion of each individual product in the family. The BU strategy forecasts each product demand individually and combines the results to obtain the forecast for the product family demand. Common intuition suggests that the TD strategy would be more accurate than the BU strategy because the average of a number of items is less variable than that of individual items (*risk-pooling*). It takes advantage of the statistical fact that the variance of the aggregated data is equal to the sum of the variances and covariances of the individual data. However, as will be discussed later in this dissertation, there are certain conditions under which the variance of forecast error for the BU strategy can be the same as or considerably lower than the variance of forecast error for the TD strategy.

1.3 Problem Definition and Purpose

This dissertation addresses the issue of evaluating the relative effectiveness of TD and BU strategies for forecasting the demand at the subaggregate (product) level as well as the demand at the aggregate (family) level under various forecasting situations meant to cover a broad spectrum of actual situations. This research also considers several time series models representing many demand processes in a production planning environment and includes both analytical and empirical investigations, with the empirical study using a simulated aggregate and disaggregate data.

This research is motivated by the following facts. First, the number of technical publications in the production planning literature is rather limited as compared to those in

[5] In practice, the data aggregation can also be done over *time periods* (by combining the data of a product during a certain time period), instead of over *products* (by combining the data for a particular period that belong to different products) as is considered in this thesis. For more details, see Theil (1954), Lutkepohl (1986), and Armstrong (2001, p. 316).

the economics literature. Second, most of the studies in the production planning literature were carried out using a simulation approach. Next, with the exception of Fliedner (1999), none of the earlier studies in the production planning literature have considered other stochastic time series models which are correlated over time, such as first-order autoregressive [AR(1)] or integrated moving average of order one [IMA(1, 1)] process[6]. Finally, very few studies have provided conclusive evidence and clear guidelines as to which forecasting strategy would render more accurate estimates under particular circumstances. This is primarily because the majority of the prior studies were done under various strict assumptions on the following aspects, such as product hierarchy, number of products in the family, underlying product/family demand process, level of demand knowledge, performance measure, etc.

Therefore, the main purpose of this dissertation is to fill these gaps as well as to provide in-depth analyses on the relative effectiveness of TD over BU forecasting strategy in the production planning domain by considering various time series models as well as forecasting scenarios. In attempting to analyze the forecast performance of the two strategies, both analytical and simulation approaches will be adopted. The expected outcomes from this research are (1) the conditions under which one forecasting strategy results in a lower forecast error than the other and (2) the impact of a particular problem parameter (e.g. number of products, demand correlation, demand proportion, etc.) on the preference to a forecasting strategy.

To date, no studies have been carried out to extensively analyze the relative effectiveness of top-down (TD) over bottom-up (BU) strategies with forecasting conditions and time series models similar to those presented in this thesis. The modeling assumptions

---

[6] In fact, AR(1) and IMA(1, 1) time series models have been widely used by researchers to study supply chain management issues (Graves, 1999, Lee, So, and Tang, 2000, and Chen, Drezner, Ryan, and Simchi-Levi, 2000), as they are believed to emulate many demand processes in real-life.

and methodologies used in this research are unique and lead to the advancement of knowledge in forecasting.

1.4 Layout and Limitations

The remainder of this thesis is organized as follows. Chapter 2 surveys existing literature on TD versus BU forecasting strategies. Tables summarizing the details of some important papers are also presented in this chapter.

Chapter 3 evaluates the performance of TD and BU strategies for forecasting the demand for the product family when it is composed of N individual products (for $N > 1$) that may be correlated with each other. The demand for all individual products is assumed to follow a specific stationary (time invariant) time series model with zero or non-zero autocorrelation. Three time series models are considered in the study, namely white noise, first-order moving average [MA(1)], first-order autoregressive [AR(1)] processes. As is common in a production planning environment, simple exponential smoothing (Brown, 1963) and weighted moving average are used as the forecasting techniques under both strategies. The investigation is carried out analytically for the case when the coefficient of the serial correlation term of the demand time series for all individual products is identical. A simulation study is then performed to investigate the case when this coefficient is not identical.

Chapter 4 compares the performance of TD and BU strategies for forecasting the demand of a product that belongs to a family. As in Chapter 3, it also considers three different stationary time series models to represent the product demands and uses both exponential smoothing and moving average as the forecasting techniques.

Chapter 5 empirically investigates the relative effectiveness of TD over BU strategies for forecasting the demand of a *substitutable* product that belongs to a family as well as the demand of a product family, under different types of family demand processes.

Product substitutability is a common phenomenon in the retail merchandising industry, where the demand for a particular product is partly driven by the inventory level of another product with similar characteristics. In addition to the three stationary time series models considered earlier, an integrated moving average of order 1 [IMA(1, 1)] process, which is non-stationary (time variant) and can be forecasted optimally by simple exponential smoothing (Muth, 1960), is also employed.

Finally, Chapter 6 summarizes the major research findings with the emphasis on the practical contribution made. This chapter also identifies a framework and avenues for further study on TD versus BU forecasting strategies.

1.5 Concluding Remarks

This chapter provided an introduction to the research issues being examined in this dissertation. Product proliferation, a popular strategy in retail merchandising industry, was identified as a major problem in demand forecasting. Improving the forecasting strategy under an HPS was identified as one viable approach to overcome this problem. Then, the aims and motivation of this dissertation were defined and the major research issues studied in this research were outlined.

CHAPTER 2

LITERATURE REVIEW

This chapter surveys prior research on top-down (TD) and bottom-up (BU) forecasting strategies in detail. The research on TD versus BU forecasting strategies can be traced as far back as 1954, when Theil addressed the problem of aggregation over individual microequations in the econometric field. Specifically, Theil (1954) developed the mathematical conditions under which bias was introduced when aggregating from microeconomic to macroeconomic relations. Since then, a stream of research in the economics as well as the production planning literature have been reported with various conclusions depending on the underlying modeling assumptions, such as forecasting techniques, time series models, value of predictor variables, etc. There seems to be no basic agreement as to which forecasting strategy is preferable under a particular situation. As noted by Ballou (1999, p. 296):

> "*Research on the subject (of TD versus BU forecasting strategies) has not provided a definitive answer, as to which approach is better.*"

With a similar notion, Weatherford, Kimes, and Scott (2001) also stated that:

> "*It is often difficult to determine whether aggregation or disaggregation will render a more accurate forecast, because of possible differences in model specifications, possible offsetting errors as data are aggregated, and by aggregation and pooling bias.*"

A considerable disagreement exists in the economics literature, which accounts for the majority of the publications that are presented here. In this domain, the proponents of the BU strategy include Edwards and Orcutt (1969), Dunn, William, and Spivey (1971), Dunn, William, and DeChaine (1976), Tiao and Guttman (1980), and Weatherby (1984), whereas the proponents of the TD strategy include Schwarzkopf, Tersine, and Morris (1988) and Chen, Kanetkar, and Weiss (1994).

Similarly, in the production planning literature, although the majority of the studies tend to support the TD strategy over the BU strategy (Barnea and Lakonishok, 1980, Muir, 1983, Tersine, 1985, p. 420, Fogarty, Hoffman, and Blackstone, 1991, p. 83, DeLurgio, 1998, p. 729, and Fliedner, 1999), there are a considerable number of papers that favor the BU strategy instead (Dangerfield and Morris, 1992, Weatherford *et al*., 2001, and Choi and Kimes, 2002).

This chapter is organized as follows. Section 2.1 discusses the prior research (primarily from the economics literature) which assumes that the underlying process of the time series is known with certainty. Section 2.2 deals with the papers (mainly from the production planning literature) which assume that the underlying process is completely unknown. At the end of every section, a detailed summary of select papers based on various attributes (e.g. method of study, level of forecast, type of data, number of streams, etc.) is provided. Finally, the time series models considered in this study are briefly summarized in the Appendices.

2.1 Known Time Series Process

This section discusses the literature which assumes that the statistical properties of the data generating process are perfectly known *a priori* and that the available data are free from measurement errors. This assumption is normally considered in the marketing and economics domains, whose major interest is to predict the sales/economic activity levels for national, regional, and local market segments. According to Chambers, Mullick, and Smith (1971) and Wei (1993, p. 86), the ideal forecasting technique in this case is to use its own data function, such as regression analysis or autoregressive integrated moving average (ARIMA) (Box and Jenkins, 1976). This is due to the fact that these techniques, also known as causal (explanatory) methods, allow the user to specifically adapt the forecasting model that best fits the time series process so that the resulting one-step ahead forecast errors are independently and identically distributed[7].

2.1.1 Forecasting Aggregate Time Series

Some important research in the economics literature has been empirical in nature. Barring one exception (Chen, *et al*., 1994[8]), the general consensus made by most of the researchers in this domain is that the BU strategy could not do worse than the TD strategy for forecasting the aggregate time series. The research in this area was initiated by Theil (1954), who assumed complete knowledge of the time series and considered the following macro-equation:

$$y(t) = \sum_{i=1}^{I} y_i(t) \tag{2.1}$$

where,

$$y_i(t) = a_i + \sum_{\lambda=1}^{N} \beta_{\lambda i} x_{\lambda i}(t) + \varepsilon_i(t) \tag{2.2}$$

I is the number of micro-equations ($I \geq 2$), $y_i(t)$ is endogenous variable of product i (micro-equation), $x_{\lambda i}(t)$ is exogenous variable, a_i and $\beta_{\lambda i}$ are micro-parameters, $\varepsilon_i(t)$ is random disturbance with zero mean, and N is the number of exogenous variables. Deducing the above expressions, the author then developed the mathematical conditions for the bias when aggregating from microeconomic to macroeconomic relations.

This assumption of complete knowledge on the behavior of the subaggregate series was later relaxed by Grunfeld and Griliches (1960). They argued that, in practice,

---

[7] Non-zero correlations among the forecast errors would indicate that there is information in the data that has not yet been captured or used.

[8] In the marketing context, Chen *et al*. (1994) examined the effectiveness of TD strategy in predicting the market share of a brand consisting of several product types. In contrast to other papers in the economics literature, the authors found that the TD strategy was consistently better than the BU strategy.

practitioners may not be able to perfectly identify the series behavior. Thus, the model specification may be less subject to errors at the aggregate level, rather than at the subaggregate level as assumed earlier by Theil (1954). This condition was illustrated by the authors via a simulation study using actual data. Similar to Theil, the authors then analytically derived the conditions when the TD strategy produced an aggregation 'gain'[9] instead of an aggregation 'bias'. Based on an ARMA time series model, an analogous conclusion was also reported by Lutkepohl (1984) and Pesaran, Pierse, and Kumar (1989). Specifically, the two studies demonstrated that if the underlying subaggregate processes were not known and had to be estimated on the basis of the available sample information, the resulting mean squared error (MSE) for the TD strategy might be lower than that for the BU strategy. The performance of the two strategies, nevertheless, was equal if the disaggregate variables were uncorrelated and had identical stochastic structures.

In another study, Orcutt, Watts, and Edward (1968) considered three levels of aggregation in order to evaluate the impact of the degree of aggregation on the relative effectiveness of TD and BU forecasting strategies. The authors assumed that the micro-equations were homogeneous and that the least-squared regression was used as the forecasting technique under both strategies. Using MSE as the performance measure, it was discovered that estimation prior to aggregation (BU) always yielded substantially greater precision than estimation after the aggregation (TD). In the worst scenario, the BU strategy would only be as good as the TD strategy. The probable reason for this finding was related to the ability of the TD strategy to respond to changes in the system, which was weaker than the BU strategy. This conclusion was supported in later studies by Zellner (1969), Edwards and Orcutt (1969) (who relaxed the assumption of homogeneous micro-equations), Aigner and Goldfeld (1973) (who also reported that the relative performance of

---

[9] This situation was defined when the macro-equation fitted the aggregate observations better than the sum of the micro-equations.

the TD strategy further deteriorated if the predictor variables were negatively correlated across equations), Dunn *et al*. (1976) (who empirically found that the BU strategy could reduce the forecast error for estimating the telephone demand in an exchange area by as much as 25.6%), Rose (1977) (who considered an ARIMA process), Tiao and Guttman (1980) (who considered MA(q) and IMA(1, q) processes), and Kohn (1982) and Weatherby (1984) (who considered an ARMA process and suggested that the gain in forecasting accuracy realized by using the BU strategy would generally exist for shorter forecast lead times).

Aigner and Goldfeld (1974) examined the influence of disaggregate measurement error which would tend to cancel out at the aggregate level. The authors considered macro- and micro-equations that were similar to (2.1) and (2.2), with $I = 2$ and identical and non-identical $\beta_{\lambda i}$. The findings were somewhat intuitive. That is, when the measurement error reduced to zero at the aggregate level, the aggregate model (TD) provided superior forecasts. However, when $\beta_{\lambda i}$ was identical, there was only a small probability that the TD strategy would outperform the BU strategy.

In a different estimation context, Binkley and Nelson (1990) examined the relative efficiency of *parameter* estimates that were obtained by aggregating or disaggregating the component data. They analytically showed that a negative correlation between errors and a positive correlation between predictor variables (both measured across equations for the disaggregate data) would improve the efficiency of the aggregate (TD) forecasting strategy.

2.1.2 Forecasting Disaggregate Time Series

Papers evaluating the relative effectiveness of TD and BU strategies for forecasting disaggregate time series are rather limited in number as compared to the case of aggregate time series. The first comprehensive study on this issue was done by Shlifer and Wolff

(1979), who analytically examined the relative performance of TD and BU forecasting strategies based on the variance of forecast error. It was assumed that the standard deviation of the forecast error for the BU strategy was a function of a market size and of the time period *t* into the future for which the forecast was made. Mathematically, it can be written as:

$$\sigma_i = c + a(s_i, y)s_i^{b(s_i,t)} \tag{2.3}$$

where s_i is the size (expected demand) of segment (item) *i*, *a* and *b* are functions of s_i and *t*, and *c* is a constant. The authors argued that (2.3) was relevant for a wide range of applications, such as for forecasting sales of lube oil, sneakers, and spare parts. By varying the value of *a*, *b*, and *c* systematically, the authors then developed the theoretical conditions under which one strategy would be preferred over the other. In conclusion, it was reported that the BU strategy was likely to outperform the TD strategy and the margin increased with the number of market (disaggregate) segments and with the decrease in the forecast lead-time. This finding was, in fact, similar to the earlier empirical study done by Dunn *et al*. (1971), who assessed the accuracy of demand forecast for telephones by using mean absolute deviation (MAD) and MSE.

Schwarzkopf *et al*. (1988, 1989) compared the performance of TD and BU strategies based on two criteria, namely cost (i.e. processing time and storage space) and accuracy (i.e. MSE). In terms of cost, no difference in performance could be identified between the two forecasting strategies. However, in terms of accuracy, the authors showed that the TD strategy would consistently outperform the BU strategy provided that the item (product) demands were free from bias and outliers. Moreover, the relative superiority of the TD strategy increased as neither item dominated the family class.

A rather different approach was pursued by van Donselaar (2002) and Dekker, van Donselaar, and Ouwehand (2004). In both studies, the authors considered the case where the demand aggregation is carried out in order to obtain the aggregate seasonal indices.

These seasonal indices are then used to derive forecasts at the product level. In the first paper, the author assumed that the product demands followed an identical seasonal pattern without a trend and that the deseasonalized demand followed a gamma distribution. In addition, Winter's exponential smoothing was used as the forecasting technique for both strategies. It was analytically found that the resulting inventory level from the TD strategy was substantially lower than the BU strategy by as much as 50% and that the relative superiority of the TD strategy increased with the number of products in the family and the variability of individual product demand. In the later paper, the authors studied the same issue empirically using real-data from two prominent Dutch wholesalers. In addition to the conclusion that the TD strategy outperforms the BU strategy in all cases, it was also revealed that the combination of product-aggregation and combined forecasts was able to give a reduction of 12%-59% in MSE.

Detailed summary of some influential papers assuming that the true process of the subaggregate time series is perfectly known is provided in Table 2.1.

2.2 Unknown Time Series Process

This section summarizes the literature on TD and BU forecasting strategies which assume that the forecasters are not aware of the true process of the time series. Thus, instead of examining and defining the real characteristics of the data generating process prior to determining the forecasting model (as is done in the earlier case), they simply use an on-hand extrapolative forecasting method which is relatively simple and easy to implement, such as moving average or exponential smoothing. This assumption is usually considered in the production planning domain, where the main interest is to estimate the demand for individual products or a product line/family. Unlike in the economics domain, there are not many studies done in this domain as it was only started in 1980 by Barnea and Lakonishok.

Table 2. 1: Current State of Research on Hierarchical Forecasting (True Process of the Subaggregate Time Series Perfectly Identified)

References	Forecast Level	Solution Type	Data Type	Subaggregate Time Series	No. of Streams	Independency btw Sources	Forecasting Technique	Performance Measure	Preference (Conclusion)	Notes
Theil (1954, 1966)	Family	Ana	-	Not Specified	∞	Independent	Regression	Theil's U^2	BU	Assumed that (1) regressors for subaggregate series were stochastic and (2) the microequations were perfectly specified
Grunfeld & Griliches (1960)	Family	Ana, Emp	Real	Not Specified	∞	Independent	Regression	Composite R^2	-	Relaxed the assumptions made by Theil (1954) above
Orcutt *et al.* (1968), Edwards & Orcutt (1969)	Family	Emp	Comp	Not Specified	> 2	Correlated	Regression	MSE	BU	- Considered 3 level of aggregations - Assumed identical and non-identical micro-components
Aigner and Goldfeld (1973, 1974)	Family	Ana, Emp	Comp	i.i.d.	2	Correlated	Regression	MSE	Tendency on BU	Assumed identical and non-identical $\beta_{\lambda i}$
Dunn *et al.* (1976)	Family	Emp	Real	Sea, Trend	> 2	Not Specified	ARIMA	MAD, RMSE	BU	- Forecast lead-time = 12 months - The highest (lowest) improvement from using BU was 25.6% (4.7%)
Rose (1977)	Family	Ana	-	ARIMA Sea & Trend	∞	Independent	ARIMA	MSE	Tendency on BU	No difference in the performance of TD and BU when the subaggregate components were identical
Shlifer & Wolff (1979)	Family & Item	Ana	-	Not Specified	∞	Independent	Not Specified	VFE	Family: Mixed Item: BU	Distribution of the forecast errors were assumed to be known
Tiao & Guttman (1980)	Family	Ana, Emp	Real	MA(q), IMA(1,q)	∞	Independent & Correlated	ARIMA	MSE	BU	- Assumed a multivariate ARIMA for the BU strategy - The relative superiority of BU decreased as the forecast lead time increased

Continued

References	Forecast Level	Solution Type	Data Type	Subaggregate Time Series	No. of Streams	Independency btw Sources	Forecasting Technique	Performance Measure	Preference (Conclusion)	Notes
Lutkepohl (1984)	Family	Ana Emp	Comp, Real	ARMA	≥ 2	Independent & Correlated	ARIMA	MSE	BU	- Assumed a multivariate ARIMA - No difference in the performance of TD and BU when the subaggregate components were identical and uncorrelated
Dunn *et al.* (1971)	Item	Emp	Real	Sea, Trend	= 2	Not Specified	AES, ARIMA	MAD, RMSE	BU	ARIMA tended to produce a more accurate forecast than AES
Schwarzkopf *et al.* (1988, 1989)	Item	Ana	-	i.i.d.	= 2	Independent & Correlated	Expected value	MSE	TD	- The forecast was based on the expected value of the demand process - BU might be more accurate than TD when demand bias and robustness were introduced
van Donselaar (2002)	Item	Ana, Emp	Comp	Gamma, Sea	≥ 2	Not Specified	WA with no trend	MAD, AIL	TD	Used the aggregate seasonal indices to derive forecasts at the product level

Abbreviations: AES=Adaptive exponential smoothing; AIL=Average inventory level; Ana=Analytically; Comp: Computer generated; Emp=Empirically; MAD=Mean absolute deviation; MSE=Mean squared error; Sea=Seasonality; RMSE=Root of mean squared error; VFE=Variance of forecast error; WA=Winters' additive exponential smoothing

2.2.1 Forecasting Family Demand

Barring one exception (Kahn, 1998b[10]), most of the studies suggested that the TD strategy is superior to the BU strategy for forecasting the family demand, which is comprised of several statistically correlated item demands (Chambers *et al*., 1971, Muir, 1983, Tersine, 1985, p. 420, McLeavey and Narasimhan, 1985, p. 67, Fogarty *et al*., 1991, p. 83, DeLurgio, 1998, p. 729, and Ballou, 1999, p. 296). This is primarily due to risk-pooling, which smoothes the overall variability of the aggregate demand (see §1.2). As stated by DeLurgio (1998, p. 729):

> "*In general, forecasts of groups are more accurate than forecasts of individual items. This principle states that the error in forecasting a specific series is normally greater than the error in forecasting the group from which this series was selected*".

Other studies that support the TD strategy instead of the BU strategy include Miller, Berry, and Lai (1976) and Barnea and Lakonishok (1980). In both of the studies, the authors concentrated on deriving a predictive equation to determine the theoretical conditions under which one strategy would perform better than the other. The aggregate demand was assumed to consist of multiple correlated components, whose variance of demand process and forecast errors were known with certainty. Theil's (1966) U^2 statistics, which was defined as the ratio of the variance of the forecast error to the variance of the underlying forecasted variable, was used in the analysis. Assuming that Δ was the difference in the variances of forecast errors:

$$\Delta = \hat{\sigma}_{TD}^2 - \hat{\sigma}_{BU}^2 \quad (2.4)$$

Barnea and Lakonishok (1980) then provided the mathematical expression for Δ when the number of subaggregate components is more than 2. The analysis was further extended by

[10] Considering three forecast levels (i.e. 7 locations at the lowest level, 2 items at the middle level, and 1 brand at the highest level), Kahn (1998b) empirically found that both TD and BU strategies performed equally well for forecasting the brand, regardless of the forecasting technique used under both strategies.

simulation, in which the simple exponential smoothing (SES) was used as the forecasting technique under both strategies. It was concluded that the TD strategy was superior to the BU strategy in 70% of the cases and that this preference depended upon (i) the correlation between the individual variables, (ii) the correlation between the forecast errors, and (iii) the accuracy of the forecasting technique measured by Theil's U^2 statistics.

In another study, Fliedner (1999) considered two subaggregate components whose demands were generated using a computer algorithm and were assumed to follow a first-order univariate moving average process, MA(1) (Appendices). The two strategies, TD and BU, were examined under two different forecasting techniques (i.e. exponential smoothing and moving average) and five levels of systematically controlled correlation between the two subaggregate components. The author reported that the TD strategy consistently outperformed the BU strategy by as much as 10%, irrespective to the forecasting technique and the performance criteria used. It was also shown that the forecast performance with either forecasting strategy was improved under conditions of high negative and high positive correlations between the two subaggregate components.

2.2.2 Forecasting Item Demands

The literature in forecasting the item demands is generally inconclusive and has no consensus favoring one strategy over the other. It appears that the preference is dependent upon the modeling assumptions.

Gross and Sohl (1990) addressed the issue of selecting an appropriate disaggregation (apportioning) technique for the TD strategy and identified the situations under which the BU strategy would outperform the TD strategy. In total, 21 disaggregation techniques and 14 forecasting methods were evaluated empirically using actual data. The root of mean squared error (RMSE) was used as the performance measure and the two forecasting strategies were compared with a simple model similar to (2.4). In conclusion, the TD strategy (with a proper disaggregation technique) was found to provide better estimates than the BU strategy in two out of three product lines examined. In addition, simple averaging process was the most promising disaggregation technique.

Beside this paper, other proponents of the TD strategy include Kahn (1998b), who used mean absolute percent error (MAPE) as the performance measure of the two forecasting strategies, and Achabal *et al*. (2000), who studied the issue in a vendor-managed inventory (VMI) system.

However, one of the proponents of the BU strategy, Lapide (1998), claimed that in the case when the product family demand was composed of several competing products that potentially cannibalize each other's demand, the BU strategy would be preferable in most cases. The TD strategy only made sense if and only if the demand patterns of the lower-level components were the same. This finding was in agreement with Dangerfield and Morris (1988, 1992), Giesberts (1993), and Diebold (1998, p. 188), to name a few.

Dangerfield and Morris (1992) used a compilation of actual data, which is also known as M1-competition database[11], in their simulation study. This database contains 1001 time series, from which only 192 data sets were actually used in the study. In order to eliminate the potential bias in the interpretation of the summary statistics, the performance of the two forecasting strategies was compared by the natural log of the ratios of MAPE as given below:

$$\ln\left[MAPE(TD)/MAPE(BU)\right] \tag{2.5}$$

The value of $\ln\left[MAPE(TD)/MAPE(BU)\right]>0$ implies that the BU strategy outperforms the TD strategy, while $\ln\left[MAPE(TD)/MAPE(BU)\right]<0$ indicates otherwise. Winter's exponential smoothing was adopted as the forecasting technique under both TD and BU strategies. Regardless of whether the smoothing constants in the Winter's exponential

---

[11] Available online: http://www.ms.ic.ac.uk/iif/data/mcomp/mcomp.htm (last accessed: 01/05/2004)

smoothing were obtained randomly or through optimization, the BU strategy was found to outperform the TD strategy in nearly three out of four cases (74%) and the relative superiority of the BU strategy was more pronounced when the correlation between the subaggregate components increased and/or when one item increasingly dominated the aggregate series. These findings are also backed by other studies from the same authors (Dangerfield and Morris, 1988, and Gordon, Morris and Dangerfield, 1997).

Weatherford *et al*. (2001) extended the investigation on TD versus BU forecasting strategies in yield management by using the data from Marriott Hotel. There were, in total, four different levels of aggregation (two of which were hybrid mechanisms that involved both TD and BU strategies) and four forecasting techniques considered in the study. It was reported that the pure BU strategy strongly outperformed the other three forecasting strategies in terms of MAD, MAPE, and median relative absolute error (MdRAE), a conclusion which was also supported by a later study by Choi and Kimes (2002). Further observation revealed that the relative superiority of the BU strategy did not change significantly for different forecasting techniques.

Table 2.2 shows the detailed summary of the technical papers which assume that the true process of the subaggregate time series is unknown.

2.3 Concluding Remarks

In this chapter, existing studies on top-down (TD) versus bottom-up (BU) forecasting strategies in both economics and production planning literature were surveyed in detail. The findings that emerge from this review provide the necessary guidelines for investigating the performance of the two forecasting strategies.

The next three chapters evaluate the relative effectiveness of TD versus BU forecasting strategies under various modeling assumptions. Subsequently, the conclusions, contributions, and directions for future research are presented.

Table 2. 2: Current State of Research on Hierarchical Forecasting (True Process of the Subaggregate Time Series Unknown)

References	Forecast Level	Solution Type	Data Type	Subaggregate Time Series	No. of Streams	Independency btw Sources	Forecasting Technique	Performance Measure	Preference (Conclusion)	Notes
Miller *et al.* (1976), Barnea and Lakonishok (1980)	Family	Ana, Emp	Real	Not Specified	Ana: ∞ Emp: = 2	Correlated	SES	Theil's U^2	TD (70%) BU (30%)	Forecast errors were assumed to be known
Kahn (1998)	Family & Item	Emp	Real	Seasonality	> 2	Not Specified	7 methods (incl. SES)	MAPE	Family: TD/BU Item: TD	- Considered 3 levels of forecast - Proposed hybrid approach
Fliedner (1999)	Family	Emp	Comp	MA(1)	= 2	Independent & Correlated	WMA, SES	SD, MAPE	TD	The relative superiority of TD was greater with non-zero demand correlation
Gross & Sohl (1990)	Item	Emp	Real	Not Specified	> 2	Independent	SES, DES, Holt, WA, WM, WQ	RMSE, CD	TD (75%) BU (25%)	- Evaluated 21 disaggregation method for the TD approach - Simple disaggregation method outperformed other more sophisticated methods
Dangerfield & Morris (1992)	Item	Emp	M1	Not Specified	= 2	Independent & Correlated	WM	MAPE	TD (26%) BU (74%)	The conclusion was not affected by the way the smoothing constants were obtained
Gordon *et al.* (1997)	Item	Emp	M1	Not Specified	= 2	Independent & Correlated	WM, judgmental	MAPE	BU	- TD and BU were relatively identical with judgmental forecasting - Statistical forecasting method outperformed the judgmental method
Weatherford *et al.* (2001), Choi and Kimes (2002)	Item	Emp	Real	Not Specified	> 2	Not Specified	C-P, MA, SES, LR, RW	MAD, MAPE, MdRAE	BU	The conclusion was not affected by the choice of forecasting technique

Abbreviations: Ana=Analytically; CD=Composite root mean square error differential; Comp=Computer generated; C-P=Classical pickup (Weatherford and Kimes, 2003); DES=Double exponential smoothing; Emp=Empirically; LR=Linear regression; M1=M1-competition data (Makridakis *et al*., 1982); MA=Moving average; MAD=Mean absolute deviation; MAPE=Mean absolute percent error; MdRAE=Median relative absolute error (Armstrong and Collopy, 1992); MSE=Mean squared error; RW=Random walk; SD=Standard deviation; SES=Simple exponential smoothing (Brown, 1963); WA=Winters' additive exponential smoothing; WM=Winters' multiplicative exponential smoothing; WQ=Winters' quadratic exponential smoothing

CHAPTER 3

FORECASTING FAMILY-LEVEL DEMAND

3.1 Introduction

This chapter analytically evaluates the relative effectiveness of top-down (TD) and bottom-up (BU) forecasting strategies for estimating the aggregate (family) demand comprising of several statistically correlated subaggregate (item) demands. The demand data for each item (product) is assumed to follow a white noise, a first-order univariate moving average [MA(1)], or a first-order univariate autoregressive [AR(1)] process. The investigation is further extended via a simulation study in order to gain a better understanding of the impact of: (i) demand correlation between the items, (ii) relative proportion of an item in the family demand, and (iii) coefficient of the serial correlation term of the item demand time series, on the relative performance of TD over BU forecasting strategy.

The remainder of this chapter is organized as follows. Section 3.2 describes the notations and assumptions used in the rest of the chapter. Section 3.3 presents the analytical evaluation of the TD and BU strategies. It is assumed that the coefficient of the serial correlation term for the MA(1) and AR(1) item demand processes is identical (i.e. $\theta_1 = \theta_2 = ... = \theta_N$ and $\phi_1 = \phi_2 = ... = \phi_N$). Since the mathematical model cannot be easily solved for the opposite cases (i.e. $\theta_1 \neq \theta_2$ and $\phi_1 \neq \phi_2$), the analysis is done by simulation; the results are presented in Section 3.4. Finally, some concluding remarks and possible directions for future research are discussed in Section 3.5.

3.2 Notation and Assumptions

The following notation is used for the analysis:

α = Smoothing constant used in exponential smoothing technique to forecast the aggregate demand, $0 < \alpha \leq 1$

α_i = Smoothing constant used to forecast the demand for item *i*, $0 < \alpha_i \leq 1$

$d_{i,t}$ = Demand for item *i* in period *t*, in units

D_t = $\sum_{i=1}^{N} d_{i,t}$ = Aggregate family demand in period *t*, in units

$\varepsilon_{i,t}$ = Error term for the demand process for item *i* in period *t*, normally distributed with zero mean $[E(\varepsilon_{i,t}) = 0]$, variance σ_i^2, and zero autocovariance $[Cov(\varepsilon_{i,t}, \varepsilon_{i,(t-k)}) = 0]$, for all $k \neq 0$

$\hat{\varepsilon}_t$ = $\sum_{i=1}^{N} \varepsilon_{i,t}$ = Error term for the aggregate demand in period *t*, where $Cov(\hat{\varepsilon}_t, \hat{\varepsilon}_{(t-k)}) = 0$, for all $k \neq 0$

$F_{i,t}$ = Demand forecast for item *i* in period *t*, in units

F_t = Forecast of aggregate demand in period *t*, in units

γ_{ij} = Coefficient of correlation between the demand of items *i* ($d_{i,t}$) and *j* ($d_{j,t}$), $|\gamma_{ij}| \leq 1$

μ = Expected value of the aggregate demand for the family, in units

μ_i = Expected value of the demand for item *i*, in units

N = Total number of items in the family

ρ_{ij} = Coefficient of correlation between the error term of the demand process for items *i* ($\varepsilon_{i,t}$) and *j* ($\varepsilon_{j,t}$), $|\rho_{ij}| \leq 1$

$\overline{\rho}_{ik}$ = Coefficient of autocorrelation of the demand process for item *i*, between the current observation at time *t* and the observation at time $t \pm k$, where $k \geq 0$, $|\overline{\rho}_{ik}| \leq 1$ when $k > 0$, and $\overline{\rho}_{ik} = 1$ when $k = 0$

T = Historical periods used in moving average technique to forecast the aggregate demand, $T > 0$ and integer.

T_i = Historical periods to forecast the demand for item i, $T_i > 0$ and integer.

The forecasting techniques adopted for both TD and BU forecasting strategies are simple exponential smoothing (SES) and weighted moving average (WMA). The primary reasons for choosing SES and WMA are as follows. First, previous studies showed that the complex methods (such as ARIMA) do not necessarily produce more accurate forecasts than the simpler ones (Makridakis *et al*., 1982, Makridakis, Chatfield, Hibon, Lawrence, Mills, Ord, and Simmons, 1993, Chatfield, 1998, de Leeuw, van Donselaar, and de Kok, 1998, and Makridakis and Hibon, 1979, 2000). Second, prior surveys indicate that, despite the development of sophisticated forecasting algorithms, managers continue to prefer relatively simple methods (Sanders and Manrodt, 1994, and Klassen and Flores, 2001). Finally, both SES and WMA is more efficient than other more sophisticated methods, in terms of processing time, implementation cost, implementation complexity, and human intervention, especially when there are a large number of series to forecast (Chambers *et al*., 1971, Tersine, 1985, p. 444, and Dekker *et al*., 2004).

It is important to note that SES and WMA give unbiased forecasts as the expected value of the demand forecast ($E[F_t]$) is equal to the expected value of the observed demand ($E[D_t]$) in the long run (Nahmias, 2001, p. 106).

3.3 Analytical Evaluation of Forecasting Strategies

In this section, the performance of the two forecasting strategies, TD and BU, is analytically evaluated based on the variance of forecast error. This performance measure is adopted as both forecasting strategies provide unbiased estimates, meaning that in the long run, their expected forecast errors are equal to zero, i. e.:

$$E(D_t - F_t) = E\left[D_t - \sum_{i=1}^{N} F_{i,t}\right] = 0 \tag{3.1}$$

In the BU strategy, the variance of forecast error (V_{BU}) is denoted as:

$$V_{BU} = Var\left[D_t - \sum_{i=1}^{N} F_{i,t}\right]$$

$$= \sum_{i=1}^{N} Var(d_{i,t} - F_{i,t}) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(d_{i,t} - F_{i,t}, d_{j,t} - F_{j,t}) \tag{3.2}$$

whereas, in the TD strategy, the variance of forecast error (V_{TD}) is determined as:

$$V_{TD} = Var(D_t - F_t) = Var(D_t) + Var(F_t) - 2Cov(D_t, F_t) \tag{3.3}$$

The analytical evaluation begins with the case when simple exponential smoothing is used under both strategies.

3.3.1 Simple Exponential Smoothing (SES) Forecasting Technique

Under the BU strategy, the forecast of the family demand for period t (F_t) is derived as the sum of the item-level forecasts, i.e.

$$F_t = \sum_{i=1}^{N} F_{i,t} \tag{3.4}$$

where $F_{i,t}$ is mathematically denoted as:

$$F_{i,t} = \alpha_i d_{i,(t-1)} + (1-\alpha_i)F_{i,(t-1)} \tag{3.5}$$

or, in an infinite form:

$$F_{i,t} = \sum_{k=1}^{\infty} \alpha_i (1-\alpha_i)^{k-1} d_{i,(t-k)} \tag{3.6}$$

Under the TD strategy, the forecast of the family demand for period t is defined as:

$$F_t = \alpha D_{t-1} + (1-\alpha)F_{t-1} \tag{3.7}$$

3.3.1.1 White Noise Demand Process

The mathematical representation of the white noise demand process is given by (2A.2). The variance of forecast error for the bottom-up forecasting strategy (V_{BU}) is first computed. To do that, it is necessary to first derive the relation between the demands of the two products as follows:

$$Cov(d_{i,t}, d_{j,t}) = Cov\left[\mu_i + \varepsilon_{i,t}, \mu_j + \varepsilon_{j,t}\right] = Cov(\varepsilon_{i,t}, \varepsilon_{j,t}) = \rho_{ij}\sigma_i\sigma_j \quad (3.8)$$

where $i \le N$ and $j \le N$. Note that the above equation is valid for all $t > 0$, as a white noise time series is time invariant. Recall that $Var\left[d_{i,(t-k)}\right] = \sigma_i^2$ and $Cov(d_{i,t}, d_{i,(t-k)}) = 0$ for all $k \ne 0$. Hence, the variance of forecast for the item i $\left[Var(F_{i,t})\right]$ can be derived:

$$Var(F_{i,t}) = Var\left[\sum_{k=1}^{\infty}\alpha_i(1-\alpha_i)^{k-1}d_{i,(t-k)}\right] = \sum_{k=1}^{\infty}\alpha_i^2(1-\alpha_i)^{2(k-1)}Var(d_{i,(t-k)})$$

$$= \alpha_i^2\sigma_i^2\sum_{k=1}^{\infty}(1-\alpha_i)^{2(k-1)} = \frac{\alpha_i\sigma_i^2}{2-\alpha_i} \quad (3.9)$$

By using (3.8) and (3.9), and the fact that $Cov(d_{i,t}, F_{i,t}) = 0$, the covariance between the two item forecasts $\left[Cov(F_{i,t}, F_{j,t})\right]$ and their variance of forecast errors $\left[Var(d_{i,t} - F_{i,t})\right]$ can be computed:

$$Cov(F_{i,t}, F_{j,t}) = Cov\left[\sum_{k=1}^{\infty}\alpha_i(1-\alpha_i)^{k-1}d_{i,(t-k)}, \sum_{k=1}^{\infty}\alpha_j(1-\alpha_j)^{k-1}d_{j,(t-k)}\right]$$

$$= \alpha_i\alpha_j\rho_{ij}\sigma_i\sigma_j\left(\sum_{k=1}^{\infty}\left[(1-\alpha_i)(1-\alpha_j)\right]^{k-1}\right)$$

$$= \frac{\alpha_i\alpha_j\rho_{ij}\sigma_i\sigma_j}{1-(1-\alpha_i)(1-\alpha_j)} \quad (3.10)$$

and,

$$Var(d_{i,t} - F_{i,t}) = Var(d_{i,t}) + Var(F_{i,t}) - 2Cov(d_{i,t}, F_{i,t})$$

$$= \sigma_i^2 + \frac{\alpha_i \sigma_i^2}{2-\alpha_i} = \frac{2\sigma_i^2}{2-\alpha_i} \tag{3.11}$$

Note that the smoothing constant (α_i) is a parameter normally set to the optimal value by practitioners (based on minimization of the forecast error during the model fitting period). Consequently, it is of interest to find the optimal α_i which minimizes the value of $Var(d_{i,t} - F_{i,t})$. (Note that under the BU strategy, the optimized item-level forecasts are aggregated to obtain the family-level forecast.) Let the optimal smoothing constant that minimizes $Var(d_{i,t} - F_{i,t})$ in (3.11) be denoted as α_i^*. From (3.11), it is obvious that $Var(d_{i,t} - F_{i,t})$ is monotonically increasing in α_i. Hence, $\alpha_i^* = \min\{\alpha_i\}$. Since $\alpha_i^* > 0$ (Nahmias 2001, p. 69), $\alpha_i^* = \min\{\alpha_i\} = 0.01$.

The analysis now proceeds to the derivation of $Cov(d_{i,t} - F_{i,t}, d_{j,t} - F_{j,t})$, which can be written as:

$$\begin{aligned} Cov(d_{i,t} - F_{i,t}, d_{j,t} - F_{j,t}) &= Cov(d_{i,t}, d_{j,t}) - Cov(d_{i,t}, F_{j,t}) \\ &\quad - Cov(F_{i,t}, d_{j,t}) + Cov(F_{i,t}, F_{j,t}) \end{aligned} \tag{3.12}$$

By substituting (3.8) and (3.10) into (3.12), and using the result that $Cov(d_{i,t}, F_{j,t}) = Cov(F_{i,t}, d_{j,t}) = 0$, it can be shown that:

$$Cov(d_{i,t} - F_{i,t}, d_{j,t} - F_{j,t}) = \rho_{ij}\sigma_i\sigma_j \left[1 + \frac{\alpha_i \alpha_j}{1-(1-\alpha_i)(1-\alpha_j)} \right] \tag{3.13}$$

Finally, the variance of forecast error for the BU strategy (V_{BU}) can be obtained by substituting (3.11) and (3.13) into (3.2):

$$V_{BU} = 2\sum_{i=1}^{N} \frac{\sigma_i^2}{2-\alpha_i} + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \left[1 + \frac{\alpha_i \alpha_j}{1-(1-\alpha_i)(1-\alpha_j)} \right] \tag{3.14}$$

The variance of forecast error for the TD strategy is derived next. Let D_t and $\hat{\varepsilon}_t$ be the product family demand and the error term at time t. Since $D_t = \sum_{i=1}^{N} d_{i,t}$, the aggregate family demand also follows a white noise process. Therefore, D_t can be written as $D_t = \mu + \hat{\varepsilon}_t$, where $Var(\hat{\varepsilon}_t) = \sum_{i=1}^{N} Var(\varepsilon_{i,t}) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(\varepsilon_{i,t}, \varepsilon_{j,t})$. By using (3.8), $Var(D_t)$ can be derived as:

$$Var(D_t) = \sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \tag{3.15}$$

Subsequently, from (3.15) and the fact that $Cov(D_t, D_{t-k}) = 0$ for all $k \neq 0$, $Var(F_t)$ can be shown as:

$$Var(F_t) = Var\left(\sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} D_{t-k}\right) = \frac{\alpha}{2-\alpha}\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j\right) \tag{3.16}$$

The variance of forecast error for the TD strategy (V_{TD}) is obtained by substituting (3.15) and (3.16) into (3.3), and using the result that $Cov(D_t, F_t) = 0$,

$$V_{TD} = \frac{2}{2-\alpha}\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j\right) \tag{3.17}$$

Similar to the earlier case, the smoothing constant (α^*) that minimizes the variance of forecast error needs to be determined. From (3.17), it is obvious that V_{TD} monotonically increases with α. Hence, $\alpha^* = \min\{\alpha\} = 0.01$.

The effectiveness of TD and BU strategies can now be compared by evaluating the ratio of variance of their forecasting errors (i.e. V_{TD}/V_{BU}). The ratio of V_{TD}/V_{BU} is used because, as shown in (3.1), the long run expected forecast errors for both TD and BU strategies is equal to zero. The value of $V_{TD}/V_{BU} > 1$ implies that the BU strategy is superior to the TD strategy, whereas the value of $V_{TD}/V_{BU} < 1$ implies otherwise. From (3.14) and (3.17),

$$V_{TD}/V_{BU} = \frac{\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j}{(2-\alpha)\left(\sum_{i=1}^{N}\frac{\sigma_i^2}{2-\alpha_i} + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\left[1+\frac{\alpha_i\alpha_j}{1-(1-\alpha_i)(1-\alpha_j)}\right]\right)} \tag{3.18}$$

<u>Theorem 3.1:</u> *If the time series of the item demands follows a white noise process, the performance of the TD and BU strategies is strictly identical ($V_{TD} = V_{BU}$) as long as the smoothing constants (α) used for forecasting both item demands and aggregate family demand are set to optimum or equal to each other.*

Proof: Suppose that the optimal smoothing constants are used for both TD and BU forecasting strategies. By substituting $\alpha_i^* = \alpha^* = \min\{\alpha\} = 0.01$, for i = 1, 2, …, N, into (3.18), it is straightforward to show that $V_{TD}/V_{BU} = 1$. Now suppose that $\alpha_1 = \alpha_2 = ... = \alpha_N = \alpha$ for $0 < \alpha \le 1$. By substituting this into (3.18), it can also be easily shown that $V_{TD}/V_{BU} = 1$.

3.3.1.2 First-Order Moving Average [MA(1)] Demand Process

The mathematical representation of the first-order moving average demand process is given by (2A.3). It is assumed that $\theta_1 = \theta_2 = ... = \theta_N$ because, when $\theta_1 \ne \theta_2 \ne ... \ne \theta_N$, the detailed statistical properties of the aggregate family demand can only be obtained through approximation (Granger and Morris, 1976) and a precise analytical evaluation of the forecast error value would be difficult. This assumption, however, is relaxed in the simulation study to be presented in Chapter 4, which considers both identical ($\theta_1 = \theta_2$) and non-identical ($\theta_1 \ne \theta_2$) coefficients. Note that when $\theta_i = 0$, the demand process corresponds to the special case where the demand for each time period is i.i.d. (Section 3.3.1).

The analysis for V_{BU} begins by deriving the following relation:

$$Cov(d_{i,t},\varepsilon_{i,(t-k)}) = Cov(\mu_i + \varepsilon_{i,t} - \theta_i\varepsilon_{i,(t-1)}, \varepsilon_{i,(t-k)})$$

$$= \begin{cases} \sigma_i^2 & k=0 \\ -\theta_i\sigma_i^2 & k=1 \\ 0 & k>1 \end{cases} \tag{3.19}$$

Furthermore, the covariance between the two item demands is defined below:

$$Cov(d_{i,t},d_{j,(t-k)}) = Cov\left[\mu_i + \varepsilon_{i,t} - \theta_i\varepsilon_{i,(t-1)}, \mu_j + \varepsilon_{j,(t-k)} - \theta_j\varepsilon_{j,(t-k-1)}\right] \tag{3.20}$$

Since $Cov(\varepsilon_{i,(t-k)},\varepsilon_{j,(t-k)}) = \rho_{ij}\sigma_i\sigma_j$ for all k, (3.20) can be transformed to:

$$Cov(d_{i,t},d_{j,(t-k)}) = \begin{cases} (1+\theta_i\theta_j)\rho_{ij}\sigma_i\sigma_j & k=0 \\ -\theta_i\rho_{ij}\sigma_i\sigma_j & k=1 \\ 0 & k>1 \end{cases} \tag{3.21}$$

Consequently, from (2A.4) and (3.21), the coefficient of correlation between the two item demands (γ_{12}) can be determined as follows:

$$\gamma_{ij} = \frac{Cov(d_{i,t},d_{j,t})}{\sqrt{Var(d_{i,t})\times Var(d_{j,t})}} = \frac{(1+\theta_i\theta_j)\rho_{ij}}{\sqrt{\left(1+\theta_i^2\right)\left(1+\theta_j^2\right)}} \tag{3.22}$$

Note that due to the linear relationship between γ_{ij} and ρ_{ij} as shown in (3.22), in the rest of the thesis the terms "coefficient of correlation between the *error terms* of the item demand processes" and the "coefficient of correlation between the item demands" will be used interchangeably.

Based on (2A.5), the covariance between one item demand and its forecast can be easily derived:

$$Cov(F_{i,t},d_{i,t}) = Cov\left[\sum_{k=1}^{\infty}\alpha_i(1-\alpha_i)^{k-1}d_{i,(t-k)}, d_{i,t}\right]$$

$$= \alpha_i \sum_{k=1}^{\infty} (1-\alpha_i)^{k-1} Cov(d_{i,(t-k)}, d_{i,t})$$

$$= -\alpha_i \theta_i \sigma_i^2 \tag{3.23}$$

Also, for the variance of the subaggregate forecast:

$$Var(F_{i,t}) = Var\left[\sum_{k=1}^{\infty} \alpha_i (1-\alpha_i)^{k-1} d_{i,(t-k)}\right]$$

$$= \lim_{M\to\infty} \left(\begin{array}{l} \alpha_i^2 \sum_{k=1}^{M} (1-\alpha_i)^{2(k-1)} Var(d_{i,(t-k)}) \\ +2\alpha_i^2 \left[\sum_{k=0}^{(M-2)} \sum_{l=1}^{(M-k-1)} (1-\alpha_i)^{2k+l} Cov(d_{i,(t-1-k)}, d_{i,(t-1-k-l)}) \right] \end{array} \right)$$

Taking the limit, $M \to \infty$,

$$Var(F_{i,t}) = \frac{\alpha_i \sigma_i^2 \left[(1-\theta_i)^2 + 2\alpha_i \theta_i \right]}{2-\alpha_i} \tag{3.24}$$

Finally, the variance of forecast error $\left[Var(d_{i,t} - F_{i,t}) \right]$ can be obtained by substituting (2A.4), (3.24), and (3.23) into the following equation:

$$Var(d_{i,t} - F_{i,t}) = Var(d_{i,t}) + Var(F_{i,t}) - 2Cov(d_{i,t}, F_{i,t})$$

$$= \frac{\sigma_i^2 (\theta_i^2 + \alpha_i \theta_i + 1)}{1 - 0.5\alpha_i} \tag{3.25}$$

From (3.25), it is obvious that $Var(d_{i,t} - F_{i,t})$ is monotonically increasing in α_i. Hence, the optimal smoothing constant $\alpha_i^* = \min\{\alpha_i\} = 0.01$.

Next, the covariance between the demand of one product and the forecast for the other product, denoted as $Cov(d_{i,t}, F_{j,t})$, is evaluated. From (3.21) and through some algebraic manipulations, it can be shown that,

$$Cov(d_{i,t}, F_{j,t}) = Cov\left[d_{i,t}, \sum_{k=1}^{\infty} \alpha_j (1-\alpha_j)^{k-1} d_{j,(t-k)} \right]$$

$$= \alpha_j \sum_{k=1}^{\infty} (1-\alpha_j)^{k-1} Cov(d_{i,t}, d_{j,(t-k)})$$

$$= -\alpha_j \theta_i \rho_{ij} \sigma_i \sigma_j \tag{3.26}$$

From (3.21) and (3.26), $Cov(F_{i,t}, F_{j,t})$ can be derived as follows:

$$Cov(F_{i,t}, F_{j,t}) = Cov\left[\alpha_i d_{i,(t-1)} + (1-\alpha_i)F_{i,(t-1)}, \alpha_j d_{j,(t-1)} + (1-\alpha_j)F_{j,(t-1)}\right]$$

$$= \alpha_i \alpha_j Cov(d_{i,(t-1)}, d_{j,(t-1)}) + \alpha_i (1-\alpha_j) Cov(d_{i,(t-1)}, F_{j,(t-1)}) +$$
$$\alpha_j (1-\alpha_i) Cov(d_{j,(t-1)}, F_{i,(t-1)}) + (1-\alpha_i)(1-\alpha_j) Cov(F_{i,(t-1)}, F_{j,(t-1)})$$

$$= \alpha_i \alpha_j (1+\theta_i \theta_j) \rho_{ij} \sigma_i \sigma_j - \alpha_i (1-\alpha_j) \alpha_j \theta_i \rho_{ij} \sigma_i \sigma_j$$
$$-\alpha_j (1-\alpha_i) \alpha_i \theta_j \rho_{ij} \sigma_i \sigma_j + (1-\alpha_i)(1-\alpha_j) Cov(F_{i,(t-1)}, F_{j,(t-1)})$$

Note that since $Cov(F_{i,t}, F_{j,t}) = Cov(F_{i,(t-k)}, F_{j,(t-k)})$ for all $k \neq 0$ (properties of stationary time series) and the fact that $\alpha_1^* = \alpha_2^* = ... = \alpha_N^*$, the above expression can further be simplified as:

$$Cov(F_{i,t}, F_{j,t}) = \frac{\alpha_N^* \rho_{ij} \sigma_i \sigma_j \left[(1+\theta_i \theta_j) - (1-\alpha_N^*)(\theta_i + \theta_j)\right]}{2-\alpha_N^*} \tag{3.27}$$

It immediately follows from (3.21), (3.26), and (3.27) that:

$$Cov(d_{i,t} - F_{i,t}, d_{j,t} - F_{j,t}) = Cov(d_{i,t}, d_{j,t}) - Cov(d_{i,t}, F_{j,t}) - Cov(d_{j,t}, F_{i,t}) + Cov(F_{i,t}, F_{j,t})$$

$$= \frac{\rho_{ij} \sigma_i \sigma_j}{2-\alpha_N^*}\left[2(1+\theta_i \theta_j) + \alpha_N^* (\theta_i + \theta_j)\right] \tag{3.28}$$

Finally, by substituting (3.25) and (3.28) into (3.2), V_{BU} can be determined as:

$$V_{BU} = \frac{\sum_{i=1}^{N} \sigma_i^2 (\theta_i^2 + \alpha_N^* \theta_i + 1) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij} \sigma_i \sigma_j \left[2(1+\theta_i \theta_j) + \alpha_N^* (\theta_i + \theta_j)\right]}{1-0.5\alpha_N^*} \tag{3.29}$$

Assuming that $\theta_1 = \theta_2 = ... = \theta_N = \hat{\theta}$, (3.29) can be simplified as:

$$V_{BU} = \frac{\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij} \sigma_i \sigma_j\right)(\hat{\theta}^2 + \alpha_N^* \hat{\theta} + 1)}{1-0.5\alpha_N^*} \tag{3.30}$$

The variance of forecast error for the top-down (TD) strategy is now derived. As previously discussed, it is assumed that $\theta_1 = \theta_2 = ... = \theta_N$. Accordingly, the aggregate family demand will also follow an MA(1) process with the same coefficient value (Lutkepohl, 1984). Let the series be defined as:

$$D_t = \mu + \hat{\varepsilon}_t - \hat{\theta}\hat{\varepsilon}_{t-1} \tag{3.31}$$

where $\theta_1 = \theta_2 = ... = \theta_N = \hat{\theta}$ and $Var(\hat{\varepsilon}_t) = \sum_{i=1}^{N} Var(\varepsilon_{i,t}) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(\varepsilon_{i,t}, \varepsilon_{j,t})$. The variance and autocovariance of D_t are then described as:

$$Var(D_{t-k}) = (1+\hat{\theta}^2)\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right), \text{ for all } k \tag{3.32}$$

$$Cov(D_t, D_{t-k}) = \begin{cases} -\hat{\theta}\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right) & |k| = 1 \\ 0 & |k| > 1 \end{cases} \tag{3.33}$$

Based on (3.32) and (3.33), the covariance between the product family demand and its forecast can be derived as:

$$Cov(F_t, D_t) = Cov\left[\sum_{k=1}^{\infty}\alpha(1-\alpha)^{k-1}D_{t-k}, D_t\right]$$

$$= -\alpha\hat{\theta}\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right) \tag{3.34}$$

and the variance of the aggregate forecast,

$$Var(F_t) = \left[(1+\hat{\theta}^2) - 2(1-\alpha)\hat{\theta}\right]\alpha^2\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)\left[\sum_{k=0}^{\infty}(1-\alpha)^{2k}\right]$$

$$= \frac{\alpha\left[(1+\hat{\theta}^2) - 2\hat{\theta}(1-\alpha)\right]\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)}{2-\alpha} \tag{3.35}$$

Next, the variance of forecast error for the TD strategy $\left[V_{TD} = Var(D_t - F_t)\right]$ is computed. By substituting (3.32), (3.34), and (3.35) into (3.3), it can be shown that,

$$V_{TD} = \frac{(\hat{\theta}^2 + \alpha\hat{\theta} + 1)\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j\right)}{1 - 0.5\alpha} \tag{3.36}$$

It is straightforward to demonstrate from (3.36) that the optimal smoothing constant (α^*) for the product family demand is again the lowest possible value of α. That is, $\alpha^* = \min\{\alpha\} = 0.01$.

The effectiveness of TD and BU strategies can now be compared by evaluating the ratio of variance of their forecasting errors. From (3.30) and (3.36), it is found to be:

$$V_{TD}/V_{BU} = \left[\frac{\hat{\theta}^2 + \alpha^*\hat{\theta} + 1}{1 - 0.5\alpha^*}\right]\left[\frac{1 - 0.5\alpha_N^*}{\hat{\theta}^2 + \alpha_N^*\hat{\theta} + 1}\right] \tag{3.37}$$

<u>Theorem 3.2:</u> *If the time series of the two item demands follows an MA(1) process with $\theta_1 = \theta_2 = ... = \theta_N$, the performance of the TD and BU strategies is strictly identical ($V_{TD} = V_{BU}$) as long as the smoothing constants (α) used for forecasting both item demands and aggregate family demand are set to optimum or equal to each other.*

Proof: By substituting $\alpha_N^* = \alpha^* = 0.01$ into (3.37), it is easy to demonstrate that $V_{TD}/V_{BU} = 1$. Now suppose that $\alpha_1 = \alpha_2 = ... = \alpha_N$, for $0 < \alpha \le 1$. Again, by substituting this into (3.37), it can easily be shown that $V_{TD}/V_{BU} = 1$.

3.3.1.3 First-Order Autoregressive [AR(1)] Demand Process

The mathematical representation of the first-order autoregressive demand process is given by (2A.8). Similar to the case of MA(1) process, for simplicity it is assumed that the coefficient of the serial correlation term for the item demand processes is identical (i.e. $\phi_1 = \phi_2 = ... = \phi_N$). However, in the simulation study, the cases of identical (i.e. $\phi_1 = \phi_2$) and non-identical (i.e. $\phi_1 \ne \phi_2$) lag-1 autocorrelations are both considered.

The derivation begins by defining the following relation:

$$Cov(F_{i,t},d_{i,t}) = Cov\left[\sum_{k=1}^{\infty}\alpha_i(1-\alpha_i)^{k-1}d_{i,(t-k)},(1-\phi_i)\mu_i + \phi_i d_{i,(t-1)} + \varepsilon_{i,t}\right]$$

$$\begin{aligned} &= \alpha_i\phi_i\sum_{k=1}^{\infty}(1-\alpha_i)^{k-1}Cov(d_{i,(t-k)},d_{i,(t-1)}) \\ &+ \alpha_i\sum_{k=1}^{\infty}(1-\alpha_i)^{k-1}Cov(d_{i,(t-k)},\varepsilon_{i,t}) \end{aligned} \quad (3.38)$$

Using the fact that $Cov(d_{i,(t-1)},d_{i,(t-1)}) = Var(d_{i,(t-1)})$ and $Cov(d_{i,(t-k)},\varepsilon_{i,t}) = 0$ for all $k \geq 1$, and substituting (2A.9) and (2A.10) into (3.38), it can be solved as:

$$Cov(F_{i,t},d_{i,t}) = \frac{\sigma_i^2\alpha_i\phi_i}{(1-\phi_i^2)}\left[\sum_{k=0}^{\infty}(1-\alpha_i)^k\phi_i^k\right] = \frac{\sigma_i^2\alpha_i\phi_i}{(1-\phi_i^2)(1-\phi_i+\alpha_i\phi_i)} \quad (3.39)$$

The variance of forecast for the item *i* can be derived as:

$$Var(F_{i,t}) = Var\left[\alpha_i d_{i,(t-1)} + (1-\alpha_i)F_{i,(t-1)}\right]$$

$$= \alpha_i^2 Var(d_{i,(t-1)}) + (1-\alpha_i)^2 Var(F_{i,(t-1)}) + \alpha_i(1-\alpha_i)Cov(d_{i,(t-1)},F_{i,(t-1)}) \quad (3.40)$$

By substituting (2A.9) and (3.39) into (3.40) and using the fact that $Var(F_{i,t}) = Var(F_{i,(t-1)})$ [as the demand process for the items is time invariant], it becomes:

$$Var(F_{i,t}) = \frac{\alpha_i\sigma_i^2(1+\phi_i-\alpha_i\phi_i)}{(1-\phi_i^2)(1-\phi_i+\alpha_i\phi_i)(2-\alpha_i)} \quad (3.41)$$

Now, the variance of forecast error for the item *i*,

$$Var(d_{i,t}-F_{i,t}) = Var(d_{i,t}) + Var(F_{i,t}) - 2Cov(d_{i,t},F_{i,t}) \quad (3.42)$$

By substituting (2A.9), (3.39), and (3.41) into (3.42) and simplifying the resulting expression:

$$Var(d_{i,t}-F_{i,t}) = \frac{2\sigma_i^2}{(1+\phi_i)(1-\phi_i+\alpha_i\phi_i)(2-\alpha_i)} \quad (3.43)$$

Solving the first order condition of $Var(d_{i,t}-F_{i,t})$ results in:

$$\alpha_i^* = \frac{3\phi_i - 1}{2\phi_i} \quad (3.44)$$

It is found that (3.43) is a convex function only when $1/3 < \phi_i < 1$; for this range of ϕ_i, the optimal smoothing constant is given by (3.44). When $-1 < \phi_i \le 1/3$, (3.43) monotonically increases with α_i, for $0 < \alpha_i^* \le 1$. Hence, $\alpha_i^* = \min\{\alpha_i\} = 0.01$. Combining these two cases:

$$\alpha_i^* = \begin{cases} (3\phi_i - 1)/2\phi_i & 1/3 < \phi_i < 1 \\ \min\{\alpha_i\} = 0.01 & -1 < \phi_i \le 1/3 \end{cases} \tag{3.45}$$

The covariance between the two item demands $\left[Cov(d_{i,t}, d_{j,(t-k)})\right]$ is derived next. For this purpose, it is necessary to first calculate the value of $Cov(d_{i,t}, \varepsilon_{j,(t-k)})$, which is derived below:

$$Cov(d_{i,t}, \varepsilon_{j,(t-k)}) = Cov\left[(1-\phi_i)\mu_i + \phi_i d_{i,(t-1)} + \varepsilon_{i,t}, \varepsilon_{j,(t-k)}\right]$$

$$= Cov\left[(1-\phi_i)\mu_i + (1-\phi_i)\phi_i\mu_i + \phi_i^2 d_{i,(t-2)} + \phi_i\varepsilon_{i,(t-1)} + \varepsilon_{i,t}, \varepsilon_{j,(t-k)}\right]$$

$$= \phi_i^2 Cov(d_{i,(t-2)}\varepsilon_{j,(t-k)}) \tag{3.46}$$

Using (3.46) recursively:

$$Cov(d_{i,t}, \varepsilon_{j,(t-k)}) = \phi_i^k Cov(d_{i,(t-k)}, \varepsilon_{j,(t-k)}) \tag{3.47}$$

Now, note that,

$$Cov(d_{i,(t-k)}, \varepsilon_{j,(t-k)}) = Cov\left[(1-\phi_i)\mu_i + \phi_i d_{i,(t-k-1)} + \varepsilon_{i,(t-k)}, \varepsilon_{j,(t-k)}\right] \tag{3.48}$$

Since $Cov(d_{i,(t-k-1)}, \varepsilon_{j,(t-k)}) = 0$ and $Cov(\varepsilon_{i,(t-k)}, \varepsilon_{j,(t-k)}) = \rho_{ij}\sigma_i\sigma_j$,

$$Cov(d_{i,(t-k)}, \varepsilon_{j,(t-k)}) = \rho_{ij}\sigma_i\sigma_j \tag{3.49}$$

Subsequently, by substituting (3.49) back into (3.47),

$$Cov(d_{i,t}, \varepsilon_{j,(t-k)}) = \phi_i^k \rho_{ij}\sigma_i\sigma_j \tag{3.50}$$

Hence,

$$Cov(d_{i,t}, d_{j,(t-k)}) = Cov\left[(1-\phi_i)\mu_i + \phi_i d_{i,(t-1)} + \varepsilon_{i,t}, (1-\phi_j)\mu_j + \phi_j d_{j,(t-k-1)} + \varepsilon_{j,(t-k)}\right]$$

$$= \phi_i \phi_j Cov(d_{i,(t-1)}, d_{j,(t-k-1)}) + \phi_i Cov(d_{i,(t-1)}, \varepsilon_{j,(t-k)})$$

$$= \phi_i \phi_j Cov(d_{i,(t-1)}, d_{j,(t-k-1)}) + \phi_i^k \rho_{ij} \sigma_i \sigma_j$$

$$= \phi_i \phi_j \left[\phi_i \phi_j Cov(d_{i,(t-2)}, d_{j,(t-k-2)}) + \phi_i^k \rho_{ij} \sigma_i \sigma_j \right] + \phi_i^k \rho_{ij} \sigma_i \sigma_j$$

$$= (\phi_i \phi_j)^2 Cov(d_{i,(t-2)}, d_{j,(t-k-2)}) + \phi_i^k \rho_{ij} \sigma_i \sigma_j (1 + \phi_i \phi_j)$$

$$= (\phi_i \phi_j)^M Cov(d_{i,(t-M)}, d_{j,(t-k-M)}) + \phi_i^k \rho_{ij} \sigma_i \sigma_j \left(\sum_{k=1}^{M} (\phi_i \phi_j)^{k-1} \right)$$

Taking the limit, $M \to \infty$,

$$Cov(d_{i,t}, d_{j,(t-k)}) = \frac{\phi_i^k \rho_{ij} \sigma_i \sigma_j}{1 - \phi_i \phi_j}, \text{ for } k \geq 0 \tag{3.51}$$

It follows from (2A.9) and (3.51) that the coefficient of correlation between the demands for the two items (γ_{ij}) is,

$$\gamma_{ij} = \frac{Cov(d_{i,t}, d_{j,t})}{\sqrt{Var(d_{i,t}) \times Var(d_{j,t})}} = \frac{\rho_{ij} \sqrt{(1-\phi_i^2)(1-\phi_j^2)}}{1 - \phi_i \phi_j} \tag{3.52}$$

Note that due to the linear relationship between γ_{ij} and ρ_{ij} (3.52), in the rest of the thesis the terms "correlation between the *error terms* of the item demand processes" and the "correlation between the item demands" will be used interchangeably.

The next step is to compute the value of $Cov(d_{i,t}, F_{j,t})$, which is given by:

$$Cov(d_{i,t}, F_{j,t}) = Cov\left[(1-\phi_i)\mu_i + \phi_i d_{i,(t-1)} + \varepsilon_{i,t}, \alpha_j d_{j,(t-1)} + (1-\alpha_j) F_{j,(t-1)} \right]$$

$$= \phi_i \alpha_j Cov(d_{i,(t-1)}, d_{j,(t-1)}) + \phi_i (1-\alpha_j) Cov(d_{i,(t-1)}, F_{j,(t-1)}) \tag{3.53}$$

By substituting (3.51) into (3.53):

$$Cov(d_{i,t}, F_{j,t})$$

$$= \frac{\phi_i \alpha_j \rho_{ij} \sigma_i \sigma_j}{1 - \phi_i \phi_j} + \phi_i (1-\alpha_j) \left[\frac{\phi_i \alpha_j \rho_{ij} \sigma_i \sigma_j}{1 - \phi_i \phi_j} + \phi_i (1-\alpha_j) Cov(d_{i,(t-2)}, F_{j,(t-2)}) \right]$$

Again, by recursive substitution of the above expression up to *M* terms:

$$Cov(d_{i,t},F_{j,t})=\left[\sum_{k=1}^{M}\phi_i^{k-1}(1-\alpha_j)^{k-1}\right]\frac{\phi_i\alpha_j\rho_{ij}\sigma_i\sigma_j}{1-\phi_i\phi_j}+\left[\phi_i(1-\alpha_j)\right]^M Cov(d_{i,(t-M)},F_{j,(t-M)})$$

Taking the limit, $M\rightarrow\infty$, it can further be simplified to:

$$Cov(d_{i,t},F_{j,t})=\frac{\phi_i\alpha_j\rho_{ij}\sigma_i\sigma_j}{(1-\phi_i\phi_j)(1-\phi_i+\alpha_j\phi_i)} \tag{3.54}$$

By using (3.51) and (3.54), and through recursive substitution, the expression for the covariance between the demand forecasts for the two items $\left[Cov(F_{i,t},F_{j,t})\right]$ can be derived as follows:

$$Cov(F_{i,t},F_{j,t})=Cov\left[\alpha_i d_{i,(t-1)}+(1-\alpha_i)F_{i,(t-1)},\alpha_j d_{j,(t-1)}+(1-\alpha_j)F_{j,(t-1)}\right]$$

$$=\alpha_i\alpha_j Cov(d_{i,(t-1)},d_{j,(t-1)})+\alpha_i(1-\alpha_j)Cov(d_{i,(t-1)},F_{j,(t-1)})$$
$$+\alpha_j(1-\alpha_i)Cov(d_{j,(t-1)},F_{i,(t-1)})+(1-\alpha_i)(1-\alpha_j)Cov(F_{i,(t-1)},F_{j,(t-1)})$$

$$=\frac{\alpha_i\alpha_j\rho_{ij}\sigma_i\sigma_j\left[1+\phi_i\phi_j(\alpha_i+\alpha_j)-\phi_i\phi_j(1+\alpha_i\alpha_j)\right]}{(1-\phi_i\phi_j)(1-\phi_i+\alpha_j\phi_i)(1-\phi_j+\alpha_i\phi_j)\left[1-(1-\alpha_i)(1-\alpha_j)\right]} \tag{3.55}$$

Similarly, using both (3.51), (3.54), and (3.55), it can be shown that,

$$Cov(d_{i,t}-F_{i,t},d_{j,t}-F_{j,t})$$
$$=Cov(d_{i,t},d_{j,t})-Cov(d_{i,t},F_{j,t})-Cov(F_{i,t},d_{j,t})+Cov(F_{i,t},F_{j,t})$$

$$=\frac{\rho_{ij}\sigma_i\sigma_j\left[(\alpha_i+\alpha_j)(1-\phi_i-\phi_j+\phi_i\phi_j)+\alpha_i\alpha_j(\phi_i+\phi_j-2\phi_i\phi_j)\right]}{(1-\phi_i\phi_j)(1-\phi_i+\alpha_j\phi_i)(1-\phi_j+\alpha_i\phi_j)(\alpha_i+\alpha_j-\alpha_i\alpha_j)} \tag{3.56}$$

Finally, the variance of forecast error for the BU forecasting strategy (V_{BU}) can be computed by substituting (3.43) and (3.56) into (3.2):

$$V_{BU}=2\left(\sum_{i=1}^{N}\frac{\sigma_i^2}{(1+\phi_i)(1-\phi_i+\alpha_i\phi_i)(2-\alpha_i)}\right)$$
$$+2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\rho_{ij}\sigma_i\sigma_j\left[(\alpha_i+\alpha_j)(1-\phi_i-\phi_j+\phi_i\phi_j)+\alpha_i\alpha_j(\phi_i+\phi_j-2\phi_i\phi_j)\right]}{(1-\phi_i\phi_j)(1-\phi_i+\alpha_j\phi_i)(1-\phi_j+\alpha_i\phi_j)(\alpha_i+\alpha_j-\alpha_i\alpha_j)} \tag{3.57}$$

Equation (3.57) can further be simplified by substituting the smoothing constants with the optimal ones as shown in (3.45). However, since the value of the optimum smoothing constant depends on the value of ϕ_i, V_{BU} is derived for different ranges of ϕ_i below:

Case 1: $1/3 < \phi_i < 1$. For this case, $\alpha_i^* = (3\phi_i - 1)/2\phi_i$

Substituting from (3.44) into (3.57):

$$V_{BU} = 8\left[\begin{array}{l}\displaystyle\sum_{i=1}^{N}\frac{\phi_i\sigma_i^2}{(1+\phi_i)^3} \\ \displaystyle +\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\rho_{ij}\sigma_i\sigma_j\phi_i\phi_j\left[\phi_i\phi_j(8+\phi_i+\phi_j-6\phi_i\phi_j)-(\phi_i^2+\phi_j^2)-(\phi_i+\phi_j)\right]}{(1-\phi_i\phi_j)\left[\phi_i\phi_j(5+\phi_i+\phi_j+\phi_i\phi_j)-2(\phi_i^2+\phi_j^2)\right](3\phi_i\phi_j+\phi_i+\phi_j-1)}\end{array}\right] \tag{3.58}$$

If $\phi_1 = \phi_2 = ... = \phi_N = \hat{\phi}$, then:

$$V_{BU} = \frac{8\hat{\phi}\left(\displaystyle\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)}{(1+\hat{\phi})^3} \tag{3.59}$$

Case 2: $-1 < \phi_i \le 1/3$. For this case, $\alpha_i^* = \min\{\alpha_i\} = 0.01$

Substituting the above into (3.57):

$$V_{BU} = 2\left[\begin{array}{l}\displaystyle\sum_{i=1}^{N}\frac{\sigma_i^2(1-\phi_i)}{(1-\phi_i^2)(1-\phi_i+\alpha_i^*\phi_i)(2-\alpha_i^*)} \\ \displaystyle +\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\sigma_i\sigma_j\rho_{ij}\left[(\alpha_i^*+\alpha_j^*)(1-\phi_i-\phi_j+\phi_i\phi_j)+\alpha_i^*\alpha_j^*(\phi_i+\phi_j-2\phi_i\phi_j)\right]}{(1-\phi_i\phi_j)(1-\phi_i+\alpha_j^*\phi_i)(1-\phi_j+\alpha_i^*\phi_j)(\alpha_i^*+\alpha_j^*-\alpha_i^*\alpha_j^*)}\end{array}\right] \tag{3.60}$$

If $\phi_1 = \phi_2 = ... = \phi_N = \hat{\phi}$, then:

$$V_{BU} = \frac{2}{(1+\hat{\phi})}\left[\begin{array}{l}\displaystyle\sum_{i=1}^{N}\frac{\sigma_i^2}{(1-\hat{\phi}+\alpha_i^*\hat{\phi})(2-\alpha_i^*)} \\ \displaystyle +\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\sigma_i\sigma_j\rho_{ij}\left[(\alpha_i^*+\alpha_j^*)(1-\hat{\phi})+2\hat{\phi}\alpha_i^*\alpha_j^*\right]}{(1-\hat{\phi}+\alpha_j^*\hat{\phi})(1-\hat{\phi}+\alpha_i^*\hat{\phi})(\alpha_i^*+\alpha_j^*-\alpha_i^*\alpha_j^*)}\end{array}\right] \tag{3.61}$$

Next, the focus is on deriving the variance of forecast error for the TD strategy. Again, for this analysis it is assumed that the coefficient of the serial correlation term for the two item demand processes is identical (i.e. $\phi_1 = \phi_2 = ... = \phi_N$). Consequently, the aggregate family demand will also follow an AR(1) process with the same value of the coefficient of the serial correlation term ($\phi_1 = \phi_2 = ... = \phi_N = \hat{\phi}$) (Lutkepohl, 1984). Let D_t and $\hat{\varepsilon}_t$ be the family demand and the error term at time t. D_t can then be defined as:

$$D_t = (1-\hat{\phi})\mu + \hat{\phi}D_{t-1} + \hat{\varepsilon}_t \tag{3.62}$$

where $Var(\hat{\varepsilon}_t) = \sum_{i=1}^{N} Var(\varepsilon_{i,t}) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(\varepsilon_{i,t}, \varepsilon_{j,t})$. The variance of the family demand (D_t) can easily be derived as:

$$Var(D_t) = Var\left[(1-\hat{\phi})\mu + \hat{\phi}D_{t-1} + \hat{\varepsilon}_t\right] = \frac{\sigma_i^2 + \sigma_j^2 + 2\rho_{ij}\sigma_i\sigma_j}{1-\hat{\phi}^2} \tag{3.63}$$

In addition, it is also necessary to compute the value of $Cov(D_t, \varepsilon_{i,(t-k)})$:

$$Cov(D_t, \varepsilon_{i,(t-k)}) = Cov\left[(1-\hat{\phi})\mu + \hat{\phi}D_{t-1} + \hat{\varepsilon}_t, \varepsilon_{i,(t-k)}\right]$$

$$= \hat{\phi}Cov(D_{t-1}, \varepsilon_{i,(t-k)}) = \hat{\phi}^k Cov(D_{t-k}, \varepsilon_{i,(t-k)})$$

$$= \hat{\phi}^k \left[\sum_{j=1}^{N} Cov(\varepsilon_{j,(t-k)}, \varepsilon_{i,(t-k)})\right]$$

Recall that $Cov(\varepsilon_{i,(t-k)}, \varepsilon_{i,(t-k)}) = Var(\varepsilon_{i,(t-k)}) = \sigma_i^2$ and $Cov(\varepsilon_{i,(t-k)}, \varepsilon_{j,(t-k)}) = \rho_{ij}\sigma_i\sigma_j$. Hence,

$$Cov(D_t, \varepsilon_{i,(t-k)}) = \hat{\phi}^k \sigma_i \left(\sum_{j=1}^{N} \rho_{ij}\sigma_j\right), \text{ for all } k > 0 \tag{3.64}$$

The autocovariance of the family demand,

$$Cov(D_t, D_{t-k}) = Cov\left[(1-\hat{\phi})\mu + \hat{\phi}D_{t-1} + \hat{\varepsilon}_t, (1-\hat{\phi})\mu + \hat{\phi}D_{t-k-1} + \hat{\varepsilon}_{t-k}\right]$$

$$= \hat{\phi}^2 Cov(D_{t-1}, D_{t-k-1}) + \hat{\phi}\sum_{i=1}^{N} Cov(D_{t-1}, \varepsilon_{i,(t-k)}) \tag{3.65}$$

By substituting (3.64) into (3.65) and through recursive substitution,

$$Cov(D_t, D_{t-k})$$
$$= \hat{\phi}^2 Cov(D_{t-1}, D_{t-k-1}) + \hat{\phi}^k \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{ij}\sigma_i\sigma_j \right)$$
$$= \lim_{M\to\infty} \left(\hat{\phi}^M Cov(D_{t-M/2}, D_{t-k-M/2}) + \hat{\phi}^k \left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \right) \left(\sum_{l=0}^{(M-2)/2} \hat{\phi}^{2l} \right) \right)$$

Taking the limit as $M \to \infty$,

$$Cov(D_t, D_{t-k}) = \frac{\hat{\phi}^k \left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \right)}{1-\hat{\phi}^2}, \text{ for all } k \neq 0 \tag{3.66}$$

Next, the covariance between the family demand and its forecast is derived,

$$Cov(D_t, F_t) = Cov\left[(1-\hat{\phi})\mu + \hat{\phi}D_{t-1} + \hat{\varepsilon}_t, \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} D_{t-k} \right]$$

$$= \hat{\phi}\alpha \left[\sum_{k=1}^{\infty} (1-\alpha)^{k-1} Cov(D_{t-1}, D_{t-k}) \right] + \alpha \left[\sum_{k=1}^{\infty} (1-\alpha)^{k-1} Cov(\hat{\varepsilon}_t, D_{t-k}) \right] \tag{3.67}$$

By substituting (3.63) and (3.66) into (3.67), and using the fact that $Cov(\hat{\varepsilon}_t, D_{t-k}) = 0$ for all $k > 1$:

$$Cov(D_t, F_t) = \frac{\hat{\phi}\alpha \left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \right)}{1-\hat{\phi}^2} \left[\sum_{k=1}^{\infty} (1-\alpha)^{k-1} \hat{\phi}^{k-1} \right]$$

$$= \frac{\hat{\phi}\alpha \left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \right)}{(1-\hat{\phi}^2)(1-\hat{\phi}+\alpha\hat{\phi})} \tag{3.68}$$

Since the TD strategy also uses the SES method, in which $F_t = \alpha D_{t-1} + (1-\alpha)F_{t-1}$, the variance for the aggregate forecast can be derived as follows:

$$Var(F_t) = \alpha^2 Var(D_{t-1}) + (1-\alpha)^2 Var(F_{t-1}) + \alpha(1-\alpha) Cov(D_{t-1}, F_{t-1})$$

From (3.63) and (3.68), and through recursive substitutions, it is found:

$$Var(F_t) = \frac{\alpha \left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \right)(1+\hat{\phi}-\alpha\hat{\phi})}{(2-\alpha)(1-\hat{\phi}^2)(1-\hat{\phi}+\alpha\hat{\phi})} \tag{3.69}$$

Finally, the variance of forecast error for the TD strategy (V_{TD}) can be obtained by substituting (3.63), (3.68), and (3.69) into (3.3):

$$V_{TD} = \frac{2\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \right)}{(1+\hat{\phi})(1-\hat{\phi}+\alpha\hat{\phi})(2-\alpha)} \tag{3.70}$$

As earlier, the smoothing constant (α^*) that minimizes the error variance can be determined. From the first order conditions of V_{TD} in (3.70):

$$\alpha^* = \frac{3\hat{\phi}-1}{2\hat{\phi}} \tag{3.71}$$

Note the similarity of the above expression to (3.44). When $1/3 < \hat{\phi} < 1$, it is straightforward to show that (3.70) is a convex function and the optimal α is given by (3.71). When $-1 < \hat{\phi} \le 1/3$, (3.70) is monotonically increasing in α, for $0 < \alpha^* \le 1$; the optimal smoothing constant in this case would be its lowest possible value. Hence, by combining these two cases:

$$\alpha^* = \begin{cases} (3\hat{\phi}-1)/2\hat{\phi} & 1/3 < \hat{\phi} < 1 \\ \min\{\alpha^*\} = 0.01 & -1 < \hat{\phi} \le 1/3 \end{cases} \tag{3.72}$$

The variance of forecast error for the TD strategy (V_{TD}) is now computed and compared with the variance of forecast error for the BU strategy (V_{BU}) under different cases of ϕ_i (recall that $\phi_1 = \phi_2 = ... = \phi_N = \hat{\phi}$).

<u>*Case 1*</u>: $1/3 < \hat{\phi} < 1$. For this case, $\alpha^* = (3\hat{\phi}-1)/2\hat{\phi}$

Substituting from (3.71) into (3.70):

$$V_{TD} = \frac{8\hat{\phi}\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)}{(1+\hat{\phi})^3} \tag{3.73}$$

From (3.59) and (3.73),

$$V_{TD}/V_{BU} = 1 \tag{3.74}$$

<u>*Case 2*</u>: $-1 < \hat{\phi} \le 1/3$. For this case, $\alpha^* = \min\{\alpha\}$

Substituting the above into (3.70):

$$V_{TD} = \frac{2\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)}{(1+\hat{\phi})(1-\hat{\phi}+\alpha^*\hat{\phi})(2-\alpha^*)} \tag{3.75}$$

From (3.61) and (3.75),

$$V_{TD}/V_{BU} = \frac{\dfrac{\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)}{(1-\hat{\phi}+\alpha^*\hat{\phi})(2-\alpha^*)}}{\left[\begin{array}{l}\sum_{i=1}^{N}\dfrac{\sigma_i^2}{(1-\hat{\phi}+\alpha_i^*\hat{\phi})(2-\alpha_i^*)} \\ +\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\dfrac{\sigma_i\sigma_j\rho_{ij}\left[(\alpha_i^*+\alpha_j^*)(1-\hat{\phi})+2\hat{\phi}\alpha_i^*\alpha_j^*\right]}{(1-\hat{\phi}+\alpha_i^*\hat{\phi})(1-\hat{\phi}+\alpha_j^*\hat{\phi})(\alpha_i^*+\alpha_j^*-\alpha_i^*\alpha_j^*)}\end{array}\right]} \tag{3.76}$$

Substituting $\alpha_1^* = \alpha_2^* = ... = \alpha_N^* = \alpha^* = \min\{\alpha\} = 0.01$ into (3.76) and simplifying the resulting equation:

$$V_{TD}/V_{BU} = 1 \tag{3.77}$$

<u>Theorem 3.3:</u> *If the time series of the two item demands follows an AR(1) process with $\phi_1 = \phi_2 = ... = \phi_N$, the performance of TD and BU strategies is strictly identical ($V_{TD} = V_{BU}$), as long as the smoothing constants (α) used for forecasting both item demands and aggregate family demand are set to optimum or equal to each other.*

Proof: Expressions (3.74) and (3.77) above immediately provide the result when the smoothing constants are optimal. When the smoothing constants are not optimal and

$\alpha_1 = \alpha_2 = ... = \alpha_N = \alpha$, it can be easily proved that $V_{TD}/V_{BU} = 1$ by simplifying (3.57) with $\phi_1 = \phi_2 = ... = \phi_N = \hat{\phi}$ and comparing the result with (3.70).

3.3.2 Weighted Moving Average (WMA) Forecasting Technique

Under the BU strategy, $F_{i,t}$ is mathematically denoted as:

$$F_{i,t} = \frac{\sum_{l=1}^{T_i} d_{i,(t-l)}}{T_i} \tag{3.78}$$

whereas under the TD strategy, the forecast of the family demand for period *t* is given as:

$$F_t = \frac{\sum_{l=1}^{T} D_{t-l}}{T} \tag{3.79}$$

To avoid repetition, the detailed analytical evaluation is omitted in this dissertation and only the summary of the end results is presented (see Table 3.1).

3.4 Simulation Study

This section reports on a simulation study to verify the results from the analytical evaluation and also to investigate the relative performance of TD and BU forecasting strategies when $\theta_1 \neq \theta_2$ and $\phi_1 \neq \phi_2$. As the main purpose of this experiments is to gain better understanding on the impact of: (i) demand correlation between the items, (ii) relative proportion of an item in the family demand, and (iii) coefficient of the serial correlation term of the item demand time series, on the relative performance of TD over BU forecasting strategy, the number of products in the family is restricted to two. The simulation program is written in MS Visual C++ 6.0. The specific algorithm used to generate a particular demand process for the two products that are correlated with each other is provided in the Appendices.

Table 3. 1: Summary of Analytical Evaluation of TD versus BU Strategies for Forecasting Family Level Demand Using Weighted Moving Average Forecasting Technique

	White Noise Demand Process	MA(1) Demand Process	AR(1) Demand Process
Optimal T_i (T_i^*)	$T_i^* = \max\{T_i\}$	$T_i^* = \max\{T_i\}$	$T_i^* = \begin{cases} 1 & 0.43 < \phi_i < 1 \\ \max\{T_i\} & -1 < \phi_i \le 0.43 \end{cases}$
V_{BU}	$\left(\frac{1}{T_i}+1\right)\left[\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right]$	$\left[\left(\frac{1}{T}+1\right)(1+\hat{\theta}^2)+\frac{2\hat{\theta}}{T^2}\right]\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)$ †	$\left(\frac{1}{1-\hat{\phi}^2}\right)\left[1+\frac{1+\hat{\phi}}{T(1-\hat{\phi})}+\frac{2\hat{\phi}(1-\hat{\phi}^T)}{T(\hat{\phi}-1)}\left(1-\frac{1}{T(\hat{\phi}-1)}\right)\right]\sum_{i=1}^{N}\sigma_i^2 + \frac{2}{1-\hat{\phi}^2}\left(1-\frac{\hat{\phi}(1-\hat{\phi}^T)}{T^2(1-\hat{\phi})^2}-\frac{1}{T(1-\hat{\phi})}\left[2\hat{\phi}(1-\hat{\phi}^T)-1\right]\right)\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l$ ‡
Optimal T (T^*)	$T^* = \max\{T\}$	$T^* = \max\{T\}$	$T^* = \begin{cases} 1 & 0.43 < \hat{\phi} < 1 \\ \max\{T\} & -1 < \hat{\phi} \le 0.43 \end{cases}$
V_{TD}	$\left(\frac{1}{T_i}+1\right)\left[\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right]$	$\left[\left(\frac{1}{T_i}+1\right)(1+\hat{\theta}^2)+\frac{2\hat{\theta}}{T_i^2}\right]\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)$	$\frac{1}{1-\hat{\phi}^2}\left[1+\frac{T(1-\hat{\phi}^2)-2\hat{\phi}(1-\hat{\phi}^T)}{T^2(\hat{\phi}-1)^2}-\frac{2\hat{\phi}(1-\hat{\phi}^T)}{T(1-\hat{\phi})}\right]\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right)$
V_{TD}/V_{BU}	1	1	$\dfrac{\left[1+\frac{T(1-\hat{\phi}^2)-2\hat{\phi}(1-\hat{\phi}^T)}{T^2(\hat{\phi}-1)^2}-\frac{2\hat{\phi}(1-\hat{\phi}^T)}{T(1-\hat{\phi})}\right]\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right)}{\left(\left[1+\frac{1+\hat{\phi}}{T(1-\hat{\phi})}+\frac{2\hat{\phi}(1-\hat{\phi}^T)}{T(\hat{\phi}-1)}\left(1-\frac{1}{T(\hat{\phi}-1)}\right)\right]\sum_{i=1}^{N}\sigma_i^2 + 2\left(1-\frac{\hat{\phi}(1-\hat{\phi}^T)}{T^2(1-\hat{\phi})^2}-\frac{1}{T(1-\hat{\phi})}\left[2\hat{\phi}(1-\hat{\phi}^T)-1\right]\right)\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right)}$
Conclusion	**Theorem 3.4:** If the time series of the item demands follows a white noise process, the performance of the TD and BU strategies is strictly identical ($V_{TD} = V_{BU}$) as long as the historical period (T) used for forecasting both item demands and aggregate family demand are set to optimum or equal to each other.	**Theorem 3.5:** If the time series of the two item demands follows an MA(1) process with $\theta_1 = \theta_2 = ... = \theta_N$, the performance of the TD and BU strategies is strictly identical ($V_{TD} = V_{BU}$) as long as the historical period (T) used for forecasting both item demands and aggregate family demand are set to optimum or equal to each other.	**Theorem 3.6[ξ]:** If the time series of the two item demands follows an AR(1) process with $\phi_1 = \phi_2 = ... = \phi_N$ and $0.43 < \phi_N < 1$, the performance of the TD and BU strategies is strictly identical ($V_{TD} = V_{BU}$) as long as the historical period (T) used for forecasting both item demands and aggregate family demand are set to optimum or equal to each other. **Theorem 3.7[ξ]:** If the time series of the two item demands follows an AR(1) process with $\phi_1 = \phi_2 = ... = \phi_N$ and $-1 < \phi \le 0.43$, the performance of the TD and BU strategies becomes more identical as the historical period (T) increases and the number of products in the family (N) decreases.

† Based on the assumption that $\theta_1 = \theta_2 = ... = \theta_N = \hat{\theta}$; ‡ Based on the assumption that $\phi_1 = \phi_2 = ... = \phi_N = \hat{\phi}$; ξ Numerical results are presented in Table A.1 in the Appendix.

The smoothing constants used in the study are set to "near-optimum" as determined by the grid search algorithm (Dangerfield and Morris, 1992, and Fliedner, 1999). It is called "near-optimum" as it depends on the grid size used in the search. The smaller the grid size, the closer the smoothing constants would be to the "exact" optimum value. However, as revealed by Farnum and Stanton (1989), the forecast error function is relatively flat in the vicinity of the optimum value, so knowing the "exact" optimum value is not critical. The search base used in this experiment is from 0.01 to 0.99, while the grid size is set to 0.01. The criterion used for optimizing the smoothing constants is root of mean squared error (RMSE).

RMSE is also used to evaluate the performance of the two forecasting strategies. For each experiment, RMSE is calculated as $\sqrt{\left(\sum_{t=365}^{884}(d_{i,t}-F_{i,t})^2\right)\Big/520}$. The relative benefit of one forecasting strategy over the other is measured by:

$$\nabla = RMSE\ (TD) / RMSE\ (BU) \tag{3.80}$$

The value of $\nabla < 1$ implies that the TD strategy is superior to the BU strategy, whereas the value of $\nabla > 1$ implies otherwise.

The simulation is conducted on a period-to-period basis. In each experiment, the simulation is carried out for a total of 884 periods. The first 104 periods are used to obtain the initial parameter of $F_{i,(t-1)}$ (3.2) or F_{t-1} (3.4). The second part containing 260 periods is used to determine the optimal smoothing constant (note that for the case of BU forecasting, the smoothing constants are optimized for each component item demand individually). Finally, the last 520 periods are used to compile the performance statistics for the two forecasting strategies. The TD and BU forecasts for each period are calculated using the methods explained earlier (Section 3.3.1). Table 3.2 shows the values of various statistical parameters that are used in the experiment.

Table 3. 2: Problem Parameters Used in the Experiment

μ (units)	σ_1^2 and σ_2^2 (units^2)	p_1*	θ_1 and θ_2	ϕ_1 and ϕ_2	ρ_{12}
200	400	0.1 – 0.9 (in increments of 0.1)	-0.9 – 0.9 (in increments of 0.2)	-0.9 – 0.9 (in increments of 0.2)	-0.9 – 0.9 (in increments of 0.1)

* p_i, for i = 1, 2, denotes the relative proportion of the value of the demand for item i over the value of the aggregate family demand, where $p_1 + p_2 = 1$.

Note that due to the similarity in the simulation results, only those obtained by simple exponential smoothing is presented. Moreover, the results for the white noise demand process are not shown as no significant difference is found between the two forecasting strategies; a finding which is consistent with Theorem 3.1. For a similar reason, the results for the cases when $\theta_1 = \theta_2$ (for the MA(1) process) and $\phi_1 = \phi_2$ (for the AR(1) process) are also not presented. These findings have been verified with a paired difference comparisons test (t-test) carried out using SPSS 12.0. At 0.05 level of significance, the observed t-value is larger than the projected t-value; thus leading to the acceptance of the null hypothesis (i.e. $H_0 : \nabla = 1$). For the cases of non-identical coefficients ($\theta_1 \neq \theta_2$ and $\phi_1 \neq \phi_2$), however, the null hypothesis is rejected at 0.05 significance level. Thus, it can generally be concluded that when $\theta_1 \neq \theta_2$ and $\phi_1 \neq \phi_2$, there is indeed a significant difference between the two forecasting strategies. The next two sections discuss the simulation results for these cases.

3.4.1 First-Order Moving Average [MA(1)] Demand Process

An analysis of variance (ANOVA) test is first performed to evaluate the impact of certain parameters on the performance ratio of the two forecasting strategies (∇). The results are exhibited in Table 3.3. As can be seen, there is sufficient evidence to statistically confirm that subaggregate proportion (p_i), intercorrelation (ρ_{12}), and coefficient θ_i have significant impact on ∇. The only exceptions are for the 2-way interaction between p_i and θ_i and the

3-way interaction between p_i, ρ_{12}, and θ_i, whose p-values are found to be considerably higher than the 0.05 level of significance.

Table 3. 3: Summary of Analysis of Variance (ANOVA) Test with Non-Identical Coefficients ($\theta_1 \neq \theta_2$)

Source	Sum of Squares	df	Mean Square	F-value	Sig. Level
p_i	3.47×10^{-3}	2	1.74×10^{-3}	3.91	0.02
θ_i	9.47×10^{-3}	9	1.05×10^{-3}	2.37	0.01
ρ_{12}	4.55×10^{-2}	18	2.53×10^{-3}	5.70	< 0.0001
$p_i * \theta_i$	2.71×10^{-3}	9	3.01×10^{-4}	0.68	0.73
$p_i * \rho_{12}$	3.06×10^{-2}	33	9.27×10^{-4}	2.09	< 0.0001
$\theta_i * \rho_{12}$	0.15	162	8.96×10^{-4}	2.02	< 0.0001
$p_i * \theta_i * \rho_{12}$	7.31×10^{-2}	148	4.94×10^{-4}	1.11	0.21
Error	0.17	378	4.44×10^{-4}		
Total	0.48	760			

Detailed analysis of the performance ratio, ∇, under different experimental scenarios is discussed next. The results of this study are plotted in Figures 3.1 and 3.2. In both of the figures, the performance ratio ∇ is plotted on y-axis and is studied with respect to two particular parameters. Figure 3.1 plots the impact of θ_2 and γ_{12} on the value of ∇ when θ_1 and p_1 are held constant at -0.3 and 0.3, respectively. The results for the other values of p_1 are not shown due to the similarity in the pattern to Figure 3.1. For a similar reason, the results for other negative values of θ_1 between -1 and 0 are also not presented.

While intuition suggests that the difference in performance between the two forecasting strategies in this case is not significant, it is found that a considerable difference between the two strategies exists, especially when the error terms of the item demands (ρ_{12}) are negatively correlated. When both of the item demands have negative coefficients of the serial correlation term, it can be seen that the reduction in the value of ρ_{12} (or γ_{12}) reduces the relative benefit of the BU strategy (as compared to the TD strategy) by as much as 6%. However, when the coefficient of the serial correlation term for the two item demands are of different sign, the RMSE of the BU strategy is lower than the TD strategy by as much as 4%. This implies that when the item demands are negatively correlated ($\gamma_{12} < 0$), the TD strategy tends to outperform the BU strategy as long as the coefficient for both of the item demands are negative (i.e. $-1 < \theta_1 < 0$ and $-1 < \theta_2 < 0$). This preference, however, is reversed for the case of $-1 < \theta_1 < 0$ and $0 < \theta_2 < 1$.

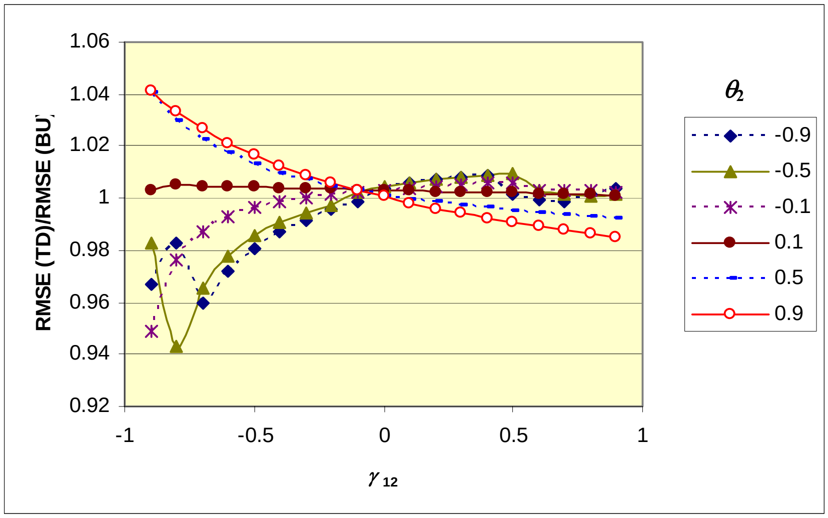


Figure 3. 1: Impact of γ_{12} and θ_2 on Relative Performance of TD and BU Strategies (θ_1 = -0.3 and p_1 = 0.3)

From Figure 3.1, it can also be seen that the difference between the TD and BU strategies diminishes when both of the item demands are independent (i.e. $\gamma_{12} = 0$), and tends to increase in the opposite direction when they are positively correlated. This leads to the conclusion that the decision to select a particular forecasting strategy should be looked at more cautiously, especially when the demand correlation between the products in the family is (positively or negatively) high. The stronger the correlation between the item demands, especially when they are negatively correlated, the greater the likelihood that the performance of the two forecasting strategies would be different. However, when no correlation exists between the two item demands (i.e. γ_{12} = 0), evidence shows that both TD and BU strategies perform equally well, irrespective to the values of θ_1 and θ_2.

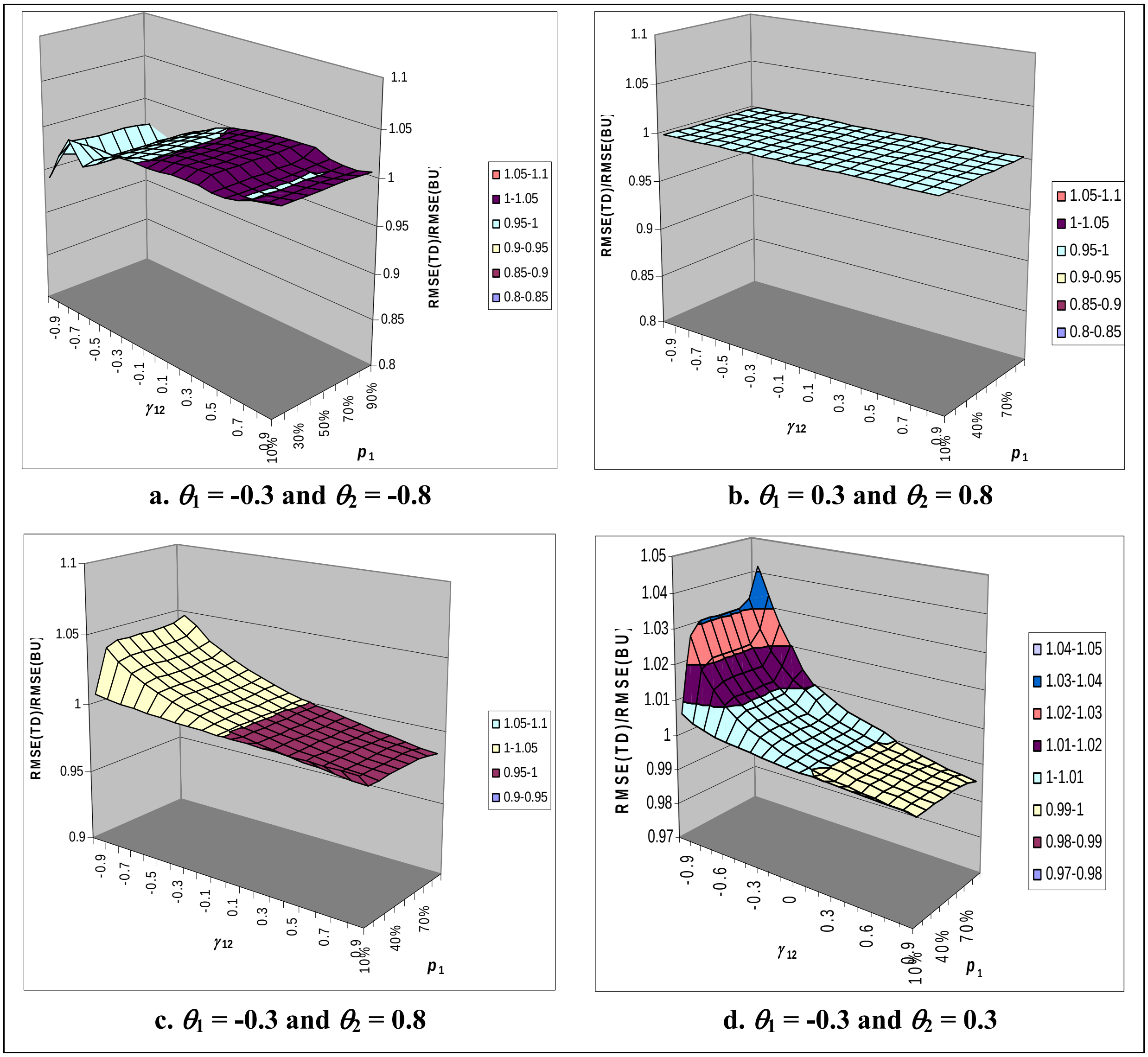


Figure 3. 2: Impact of γ_{12}, p_1 and p_2 on Relative Performance of TD and BU Strategies under Different Values of θ_1 and θ_2

The effect of the item demand's proportion in the family (p_i) and the correlation between the error terms of the item demands (ρ_{12}) on the value of ∇ is illustrated next in Figure 3.2. This figure consists of 4 different graphs representing, namely, (θ_1 = -0.3, θ_2 = -0.8), (θ_1 = 0.3, θ_2 = 0.8), (θ_1 = -0.3, θ_2 = 0.8) and (θ_1 = -0.3, θ_2 = 0.3). As can be seen from Figure 3.2a, when both θ_1 and θ_2 are negative, it is found that varying p_i does not significantly change the value of ∇. Furthermore, when both θ_1 and θ_2 are positive, ∇ is consistently equal to 1 [$RMSE(TD) = RMSE(BU)$], regardless of the values of p_i and ρ_{12} (or γ_{12}) (see Figure 3.2b).

Figures 3.2c and 3.2d, which illustrate the case when θ_1 and θ_2 are of the opposite signs, show that the performance of the BU strategy is better than the TD strategy in majority of the cases. For the case of negatively correlated item demands, the variance of forecast error for the TD strategy is higher than the BU strategy by as much as 4%, whereas for the case of positively correlated item demands, the difference is relatively negligible. In addition, while p_i seems to have a minimum impact on ∇ when the value of γ_{12} is moderate to high ($0 \le \gamma_{12} \le 1$), the performance of the TD strategy generally deteriorates more rapidly than the BU strategy (or ∇ increases) as the item demand with negative coefficient θ increases in proportion. (Note that in this experiment, p_1 corresponds to the demand proportion for item 1 with θ_1 = -0.3.)

3.4.2 First-Order Autoregressive [AR(1)] Demand Process

Similar to the earlier case, an ANOVA test is carried out in order to evaluate the impact of certain parameters on the value of ∇ when the demand follows an AR(1) process. As shown in Table 3.4, there is sufficient evidence to statistically confirm that item demand proportion (p_i), intercorrelation (ρ_{12}), and coefficient ϕ_i significantly affect the relative performance of TD over BU forecasting strategy. The only exception is the two-way interaction between p_i and ϕ_i, whose p-value (0.88) is found to be considerably higher than the 0.05 level of significance.

Table 3. 4: Summary of Analysis of Variance (ANOVA) Test with Non-Identical Coefficients ($\phi_1 \neq \phi_2$)

Source	Sum of Squares	Df	Mean Square	*F*-value	Sig. Level
p_i	6.28×10^{-3}	1	6.28×10^{-3}	4.93	0.03
ϕ_i	0.69	9	7.70×10^{-2}	60.38	< 0.0001
ρ_{12}	0.14	18	7.68×10^{-3}	6.03	< 0.0001
$p_i * \phi_i$	5.68×10^{-3}	9	6.31×10^{-4}	0.50	0.88
$p_i * \rho_{12}$	7.84×10^{-2}	18	4.35×10^{-3}	3.42	< 0.0001
$\phi_i * \rho_{12}$	1.02	162	6.30×10^{-3}	4.95	< 0.0001
$p_i * \phi_i * \rho_{12}$	0.29	162	1.77×10^{-3}	1.39	0.01
Error	0.48	380	1.27×10^{-3}		
Total	2.88	760			

The detailed analysis of the performance ratio ∇ under different experimental scenarios is discussed next. The results of this simulation study are plotted in Figures 3.3 to 3.5. Figure 3.3 illustrates the impact of demand correlation (γ_{12}) and coefficient ϕ_2 on the relative benefit of TD over BU strategy, when coefficients ϕ_1 and p_1 are kept constant at 0.8 and 0.3, respectively. The results for other positive values of ϕ_1 are not presented due to the similarity of the pattern to Figure 3.3.

In contrast to the case when the lag-1 autocorrelation for all the item demands is the same ($\phi_1 = \phi_2$) (in which the forecast performance of both TD and BU strategies is found to be strictly identical ($\nabla = 1$), Figure 3.3 demonstrates a considerable difference in the RMSE of the two forecasting strategies, especially when the two item demands are negatively correlated. In a majority of the cases, the BU strategy outperforms the TD strategy with the variance of forecast error for the BU strategy being lower than the TD strategy by as much as 80%. This finding is quite surprising as prior research in the production planning domain has reported that the TD strategy generally outperforms the BU strategy (Barnea and Lakonishok, 1980, and Fliedner, 1999).

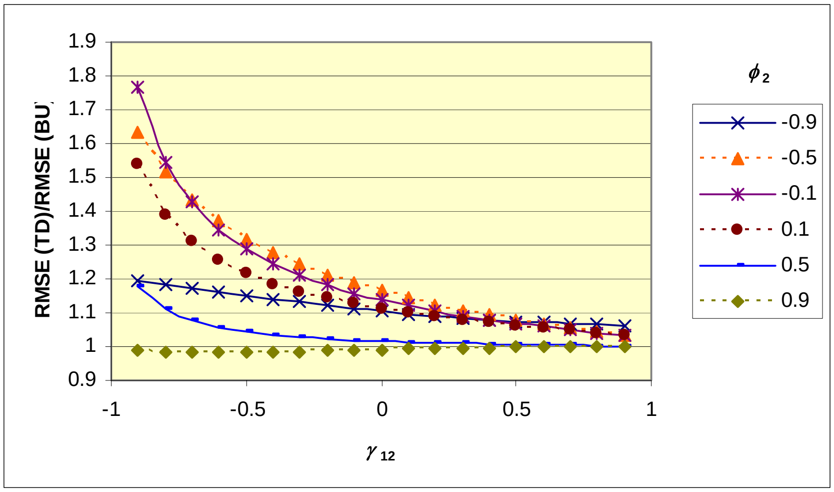


Figure 3. 3: Impact of γ_{12} and ϕ_2 on Relative Performance of TD and BU Strategies (ϕ_1 = 0.8 and p_1 = 0.3)

Moreover, the experiment also reveals that $\nabla > 1$ when the lag-1 autocorrelation of the two item demand processes are of the opposite sign. This result can also be seen in Figure 3.4, which illustrates the impact of γ_{12} and coefficient ϕ_2 on the value of ∇ when coefficient ϕ_1 is set negative and kept constant at -0.8 and p_1 is kept fixed at 0.3. If the lag-1 autocorrelation for one of the products is positive ($0 < \phi_1 < 1$) while the other one is negative ($-1 < \phi_2 < 0$), the experiment shows that the BU strategy outperforms the TD strategy by even a greater margin.

This, however, is not the case when the demand time series for both of the products has negative lag-1 autocorrelation ($-1 < \phi_1 < 0$ and $-1 < \phi_2 < 0$). Regardless of whether or not ϕ_1 and ϕ_2 are identical, the performance of both forecasting strategies is nearly equivalent ($\nabla \approx 1$). One reason for this phenomenon could be due to the fact that when the item demand has a very low lag-1 autocorrelation, the underlying demand pattern tends to be highly fluctuating (large demand interspersed with low demand). Consequently, according to (3.45), the optimal value of α_i in this case should be set as low as possible. This would, in return, affect the performance of the BU strategy adversely as its best estimate now would include very little adjustment over time.

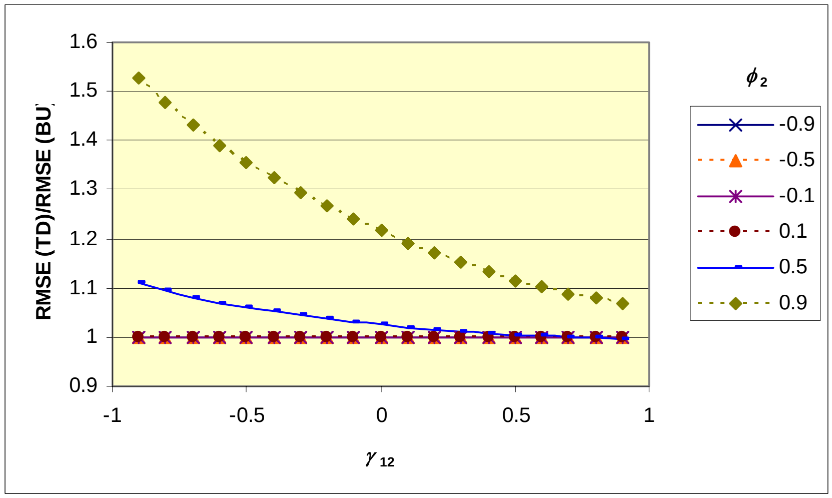
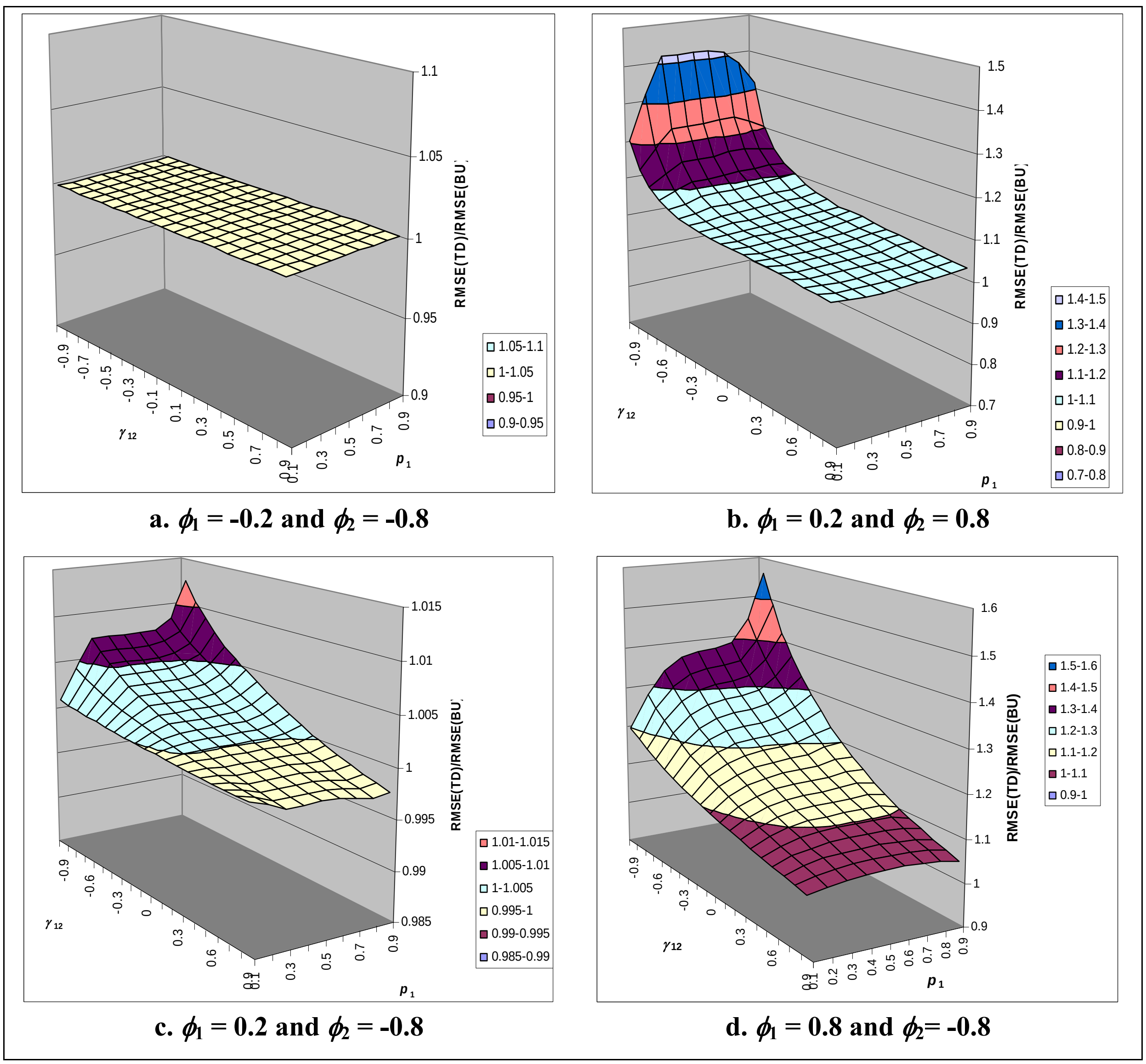


Figure 3. 4: Impact of γ_{12} and ϕ_2 on Relative Performance of TD and BU Strategies (ϕ_1 = -0.8 and p_1 = 0.3)

Finally, when both coefficients ϕ_1 and ϕ_2 are positive, the difference in performance between TD and BU strategies seems to depend on how close the values of the lag-1 autocorrelation of the two item demand processes are. That is, the closer the lag-1 autocorrelation value of the two item demand processes (i.e. ϕ_1 approaches ϕ_2), the greater the likelihood for the TD and BU strategies to perform equally well. All these findings hold true even when the relative proportion of one item in the family (p_i) is varied in the simulation study.

From Figures 3.3 and 3.4, it is further discovered that the performance ratio of TD over BU forecasting strategy tends to increase exponentially as the correlation between the two item demand processes (γ_{12}) decreases. This finding, which also applies to other values of p_1, is in contradiction to some of the earlier studies, such as Fliedner (1999) who argued that a lower demand correlation helps to stabilizing the aggregate demand and thus, affects the performance of the TD strategy positively. Hence, it can be concluded that regardless of the demand proportion of each product in the family (p_1 and p_2) and the values of ϕ_1 and ϕ_2, the difference in the performance of the two strategies diminishes as the item demands become more positively correlated (i.e. γ_{12} approaches 1).

The next experiment investigates how the relative proportion of one product in the family (p_i) and demand correlation (γ_{12}) affect the relative performance of TD and BU strategies when the values of both ϕ_1 and ϕ_2 are held constant. The results when the values of both ϕ_1 and ϕ_2 are either negative or positive are shown in Figures 3.5a and 3.5b, respectively. Apparently, varying the value of p_i in these cases do not have a significant impact on ∇. The value of ∇ is greatly influenced by the change in p_1 only when both ϕ_1 and ϕ_2 are of different sign.



a. ϕ_1 = -0.2 and ϕ_2 = -0.8

b. ϕ_1 = 0.2 and ϕ_2 = 0.8

c. ϕ_1 = 0.2 and ϕ_2 = -0.8

d. ϕ_1 = 0.8 and ϕ_2= -0.8

Figure 3. 5: Impact of γ_{12}, p_1 and p_2 on Relative Performance of TD and BU Strategies under Different Values of ϕ_1 and ϕ_2

Figures 3.5c and 3.5d exhibit the impact of p_1 and p_2 on ∇ when (ϕ_1 = 0.2 and ϕ_2 = -0.8) and (ϕ_1 = 0.8 and ϕ_2 = -0.8). The results for other values of ϕ_1 and ϕ_2 (when they are of different sign) are not shown due to the similarity in the performance graphs to these figures (see also the results of the ANOVA test in Table 3.4). In agreement with the earlier figures, the difference in the performance of the two forecasting strategies appears to be more pronounced as the correlation between the two products decreases (i.e. γ_{12} approaches -1), regardless of the value of p_1. Furthermore, as p_1 increases, the relative superiority of the BU strategy (as compared to the TD strategy) improves, especially when the demand correlation between the two items is highly negative. Hence, to summarize, as long as the lag-1 autocorrelation of the two item demand processes are of different sign, the relative superiority of the BU strategy increases as the proportion of the demand for the item with positive lag-1 autocorrelation increasingly dominates the family demand.

3.5 Conclusions

The effectiveness of TD and BU strategies for forecasting the aggregate family demand in a production planning environment was compared in this chapter. When SES was used as the forecasting technique under both strategies, the study showed that the desirability of a strategy depended on the parameters. Some parameter combinations were particularly suited to the TD strategy, whereas for others the payoff was marginal. The following are the summary of the findings:

- When the item demands followed a white noise, MA(1), or AR(1) process with identical coefficients of the serial correlation term (i.e. $\theta_1 = \theta_2 = ... = \theta_N$ and $\phi_1 = \phi_2 = ... = \phi_N$), no significant difference in the variance of forecast error was found between TD and BU strategies. This conclusion also held true for the case when the item demands were uncorrelated ($\gamma_{ij} = 0$);

- When both of the item demands followed an MA(1) process, $N = 2$ and $\theta_1 \neq \theta_2$, the maximum difference between the RMSE of the two forecasting strategies was only 6%. This implies that there was not much difference in the relative effectiveness of the two forecasting strategies (Table 3.5);
- When both of the item demands followed an AR(1) process, $N = 2$ and $\phi_1 \neq \phi_2$, the variance of forecast error (or RMSE) of the BU strategy tended to be lower than the TD strategy by as much as 80%. Moreover, the relative superiority of the BU strategy increased as the correlation between the two item demand processes (γ_{12}) decreased and as the proportion of the item demand with positive lag-1 autocorrelation increasingly dominated the family demand. An identical performance between TD and BU strategies was expected when both ϕ_1 and ϕ_2 were negative (Table 3.6).

Table 3. 5: Preferred Forecasting Strategy under Different States of Intercorrelation (γ_{12}) and Coefficient θ_i

<table>
<tr><th rowspan="2">θ_1</th><th rowspan="2">θ_2</th><th colspan="3">Dominating Strategy</th></tr>
<tr><th>$\gamma_{12}^* < 0$</th><th>$\gamma_{12} = 0$</th><th>$\gamma_{12} > 0$</th></tr>
<tr><td>Negative</td><td>Negative</td><td>TD</td><td rowspan="4">Identical Performance</td><td>BU</td></tr>
<tr><td>Negative</td><td>Positive</td><td>BU</td><td>TD</td></tr>
<tr><td>Positive</td><td>Positive</td><td>Identical Performance</td><td>Identical Performance</td></tr>
<tr><td>Positive</td><td>Negative</td><td>BU</td><td>TD</td></tr>
</table>

* $-1 \leq \gamma_{12} \leq 1$

When moving average is used as the forecasting technique, on the other hand, both TD and BU strategies performed equally well. The only exception was for the case when the product demands followed an AR(1) time series process with $-1 < \phi \leq 0.43$, in which the difference in the performance of TD and BU strategies increased with the number of products in the family (N) and as the historical period (T) decreased.

Table 3. 6: Preferred Forecasting Strategy under Different States of Coefficients ϕ_1 and ϕ_2

	Negative ϕ_1	**Positive ϕ_1**
Negative ϕ_2	Identical Performance	BU[a]
Positive ϕ_2	BU[a]	Mixed[b]

[a] $\nabla \to 1$ when $\phi_1 \approx \phi_2$

[b] Depends on specific values of ϕ_1 and ϕ_2, though BU dominates in most of the cases.

The next chapter evaluates the relative performance of TD over BU strategy for forecasting the demand of an individual product that belongs to a product family. Similar to this chapter, it considers three stationary time series models and utilizes both simple exponential smoothing and weighted moving average as the forecasting techniques.

CHAPTER 4

FORECASTING ITEM-LEVEL DEMAND

In this chapter, the performance of TD and BU strategies for forecasting the demand of an individual product belonging to a product family is compared. Similar to Chapter 3, it is assumed that the demand for each item follows a white noise, a first-order moving average [MA(1)], or a first-order autoregressive [AR(1)] process and that the forecasting techniques used are exponential smoothing and moving average.

4.1 Analytical Evaluation of Forecasting Strategies

This chapter uses the same notation and assumptions listed in Chapter 3. The performance of the two forecasting strategies is evaluated based on the variance of forecast error as both of them produce unbiased estimates $\left[\text{i.e. } E(d_{i,t} - F_{i,t}) = E(d_{i,t} - p_i F_t) = 0\right]$, where $p_i = \mu_i / \mu$ is the relative proportion of item i in the family's expected aggregate demand. The variance of the forecast error for the BU strategy (V_{BU}) is defined as:

$$V_{BU} = \sum_{i=1}^{N} \left[Var(d_{i,t} - F_{i,t}) \right] \tag{4.1}$$

while the variance of forecast error for the TD strategy (V_{TD}) is determined as:

$$V_{TD} = \sum_{i=1}^{N} \left[Var(d_{i,t} - p_i F_t) \right] \tag{4.2}$$

4.1.1 Simple Exponential Smoothing (SES) Forecasting Technique

Under the BU strategy, the demand forecast for the item i in period t is given by (3.5). Under the TD strategy, on the other hand, it is mathematically defined as:

$$F_{i,t} = p_i F_t \tag{4.3}$$

where F_t is given by (3.4). It is assumed that the relative proportion (p_i) is known with certainty or can be estimated with high accuracy using the historical demand data.

4.1.1.1 White Noise Demand Process

The variance of forecast error for the BU forecasting strategy (V_{BU}) can be easily obtained by substituting (3.11) into (4.1),

$$V_{BU} = 2\sum_{i=1}^{N} \frac{\sigma_i^2}{2-\alpha_i^*} \tag{4.4}$$

Furthermore, the variance of forecast error for the item i can be determined by:

$$Var(d_{i,t} - p_i F_t) = Var(d_{i,t}) + p_i^2 Var(F_t) - 2p_i Cov(d_{i,t}, F_t) \tag{4.5}$$

By substituting (3.16) into (4.5), and using $Var(d_{i,t}) = \sigma_i^2$ and $Cov(d_{i,t}, F_t) = 0$, $Var(d_{i,t} - p_i F_t)$ can be solved as follows:

$$Var(d_{i,t} - p_i F_t) = \sigma_i^2 + \frac{p_i^2 \alpha}{2-\alpha}\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j)\right) \tag{4.6}$$

The variance of forecast error for the TD strategy (V_{TD}) can then be obtained by substituting (4.6) into (4.2) and simplifying the resulting expression to,

$$V_{TD} = \sum_{i=1}^{N} \sigma_i^2 + \frac{\alpha\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j)\right)\left(\sum_{i=1}^{N} p_i^2\right)}{2-\alpha} \tag{4.7}$$

As in the BU strategy, the optimal smoothing constant that minimizes the variance of forecast error above needs to be determined. The difference with the BU forecasting, however, is that the optimization of the smoothing constant here should be done at the family level, instead of at the item level [see (4.3)]. Therefore, the optimal smoothing constant for the TD strategy (α^*) is computed by minimizing α in (3.17). It is straightforward to show that $Var(D_t - F_t)$ is monotonically increasing in α, for $0 < \alpha \leq 1$ (Nahmias 2001, p. 69). Hence, $\alpha^* = \min\{\alpha\} = 0.01$.

Therefore, from (4.4) and (4.7),

$$V_{TD}/V_{BU} = \frac{(2-\alpha^*)\left(\sum_{i=1}^{N}\sigma_i^2\right) + \alpha^*\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j)\right)\left(\sum_{i=1}^{N}p_i^2\right)}{2(2-\alpha^*)\left(\sum_{i=1}^{N}\frac{\sigma_i^2}{2-\alpha_i^*}\right)} \tag{4.8}$$

<u>Theorem 4.1:</u> *If the time series of the item demands follows a white noise process, the BU strategy always outperforms the TD strategy for N = 2. The difference in the variance of forecast error between the two forecasting strategies further increases with the number of products in the family.*

Proof: Recall that $\alpha_1^* = \alpha_2^* = ... = \alpha_N^* = \alpha^* = 0.01$. Hence, (4.8) can be simplified to:

$$V_{TD}/V_{BU} = 1 - 0.005\left[1 - \sum_{i=1}^{N}p_i^2\right] + \frac{0.01\left(\sum_{i=1}^{N}p_i^2\right)\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)}{\sum_{i=1}^{N}\sigma_i^2} \tag{4.9}$$

By using Solver provided in MS Excel, the above equation is solved under different number of products in the family (see Table 4.1). Obviously, the variance of forecast error of the TD strategy is always higher than that of the BU strategy when $N = 2$, regardless of the values of p_i, ρ_{ij}, and σ_i. Moreover, the performance difference between the two strategies increases with N.

Table 4. 1: Upper and Lower Bounds of V_{TD}/V_{BU} for Different Number of Products in the Family under White Noise Demand Process

N	LB	UB	UB - LB
2	1.00	1.01	0.01
3	0.99	1.01	0.02
4	0.99	1.02	0.02
5	0.99	1.02	0.03

4.1.1.2 First-Order Moving Average [MA(1)] Demand Process

Similar to the earlier section, the variance of forecast error for the BU forecasting strategy (V_{BU}) can be obtained easily by substituting (3.25) into (4.1),

$$V_{BU} = \sum_{i=1}^{N} \frac{\sigma_i^2(\theta_i^2 + \alpha_i^* \theta_i + 1)}{1 - 0.5\alpha_i^*} \tag{4.10}$$

As $\alpha_1^* = \alpha_2^* = ... = \alpha_N^*$ (see 3.3.1.2) and by assuming that $\theta_1 = \theta_2 = \hat{\theta}$, (4.10) can further be simplified to:

$$V_{BU} = \frac{(\hat{\theta}^2 + \alpha_N^* \hat{\theta} + 1)}{1 - 0.5\alpha_N^*} \sum_{i=1}^{N} \sigma_i^2 \tag{4.11}$$

The variance of forecast error for the top-down (TD) strategy is now derived. The expressions for $Cov(d_{i,t}, D_{t-k})$ and $Cov(d_{i,t}, F_t)$ are first solved below:

$$Cov(d_{i,t}, D_{t-k}) = Cov\left[p_i\mu + \varepsilon_{i,t} - \hat{\theta}\varepsilon_{i,(t-1)}, \mu + \hat{\varepsilon}_{t-k} - \hat{\theta}\hat{\varepsilon}_{t-k-1} \right]$$

$$= \begin{cases} -\hat{\theta}\sigma_i \sum_{j=1}^{N} \rho_{ij}\sigma_j & |k| = 1 \\ 0 & |k| > 1 \end{cases} \tag{4.12}$$

and,

$$Cov(d_{i,t}, F_t) = Cov\left[d_{i,t}, \alpha \sum_{k=1}^{\infty} (1-\alpha)^{k-1} D_{t-k} \right]$$

$$= -\alpha\hat{\theta}\sigma_i \sum_{j=1}^{N} \rho_{ij}\sigma_j \tag{4.13}$$

The variance of forecast error for the product i can thus be computed by substituting (2A.4), (3.35), and (4.13) into (4.5). Hence,

$$Var(d_{i,t} - p_i F_t) = (1+\hat{\theta}^2)\sigma_i^2 + 2p_i\alpha\hat{\theta}\sigma_i \sum_{j=1}^{N} \rho_{ij}\sigma_j + \frac{p_i^2\alpha(1+\hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta}\alpha)\left(\sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \right)}{2-\alpha} \tag{4.14}$$

Finally, by substituting (4.14) into (4.2), the variance of forecast error for the TD strategy can be determined as follows:

$$V_{TD} = \sum_{i=1}^{N} \left[\begin{array}{l} (1+\hat{\theta}^2)\sigma_i^2 + 2p_i\alpha^*\hat{\theta}\sigma_i \sum_{j=1}^{N} \rho_{ij}\sigma_j \\ + \dfrac{p_i^2\alpha^*\left(1+\hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta}\alpha^*\right)\left(\sum_{l=1}^{N}\sigma_l^2 + 2\sum_{l=1}^{N-1}\sum_{j=l+1}^{N}\rho_{lj}\sigma_l\sigma_j\right)}{(2-\alpha^*)} \end{array} \right] \tag{4.15}$$

Based on (3.36), it is straightforward to show that the optimal smoothing constant (α^*) for the aggregate family demand is again the lowest possible value of α. That is, $\alpha^* = \min\{\alpha\} = 0.01$. As (4.15) is evaluated against (4.11) and using $\alpha_1^* = \alpha_2^* = ... = \alpha_N^* = \alpha^* = \min\{\alpha\} = 0.01$, it can be shown that:

$$V_{TD}/V_{BU} = \frac{(1+\hat{\theta}^2)(2-\alpha^*) + \varpi \Big/ \sum_{i=1}^{N}\sigma_i^2}{2(\hat{\theta}^2 + \alpha^*\hat{\theta} + 1)} \tag{4.16}$$

where,

$$\begin{aligned} \varpi = {} & 2(2-\alpha^*)\alpha^*\hat{\theta}\sum_{i=1}^{N}\left(p_i\sigma_i \sum_{j=1}^{N}\rho_{ij}\sigma_j \right) \\ & + \alpha^*\left(1+\hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta}\alpha^*\right)\left(\sum_{i=1}^{N} p_i^2\right)\left(\sum_{l=1}^{N}\sigma_l^2 + 2\sum_{l=1}^{N-1}\sum_{j=l+1}^{N}\rho_{lj}\sigma_l\sigma_j\right) \end{aligned} \tag{4.17}$$

<u>Theorem 4.2:</u> *If the time series of the item demands follows an MA(1) process with* $\theta_1 = \theta_2 = ... = \theta_N$, *the difference in the variance of forecast error between TD and BU forecasting strategies for N = 2 is always lower than or equal to 1% . The difference further increases with the number of products in the family.*

Proof: Substituting $\alpha^* = \min\{\alpha\} = 0.01$, (4.16) and (4.17) can be simplified to:

$$V_{TD}/V_{BU} = \frac{1.99(1+\hat{\theta}^2) + \varpi \Big/ \sum_{i=1}^{N}\sigma_i^2}{2(\hat{\theta}^2 + 0.01\hat{\theta} + 1)} \tag{4.18}$$

where,

$$\varpi = 0.04\hat{\theta}\sum_{i=1}^{N}\left(p_i\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j\right) + 0.01\left(\hat{\theta}^2 - 1.98\hat{\theta} + 1\right)\left(\sum_{i=1}^{N}p_i^2\right)\left(\sum_{l=1}^{N}\sigma_l^2 + 2\sum_{l=1}^{N-1}\sum_{j=l+1}^{N}\rho_{lj}\sigma_l\sigma_j\right) \quad (4.19)$$

The upper and lower bounds of V_{TD}/V_{BU} are then determined by using Solver; the results of which are presented in Table 4.2. As can clearly be seen, when N = 2, the difference between the two forecasting strategies is relatively negligible, meaning that in the worst case scenario, the variance of forecast error of one strategy would be higher than the other by merely 1%. This performance difference, however, increases with N.

Table 4. 2: Upper and Lower Bounds of V_{TD}/V_{BU} for Different Number of Products in the Family under MA(1) Demand Process

N	LB	UB	UB - LB
2	0.99	1.01	0.02
3	0.99	1.02	0.04
4	0.98	1.03	0.05
5	0.98	1.04	0.06

4.1.1.3 First-Order Autoregressive [AR(1)] Demand Process

The derivation of the variance of forecast error for the BU forecasting strategy (V_{BU}) can be obtained easily by substituting (3.43) into (4.1),

$$V_{BU} = 2\sum_{i=1}^{N}\frac{\sigma_i^2}{(1+\phi_i)(1-\phi_i+\alpha_i^*\phi_i)(2-\alpha_i^*)} \quad (4.20)$$

Note that the above expression can be further simplified by substituting the smoothing constants with the optimal ones shown in (3.45). The derivation of V_{BU} for different ranges of ϕ_i is given below:

<u>*Case 1*</u>: $1/3 < \phi_i < 1$. In this case, $\alpha_i^* = (3\phi_i - 1)/2\phi_i$

Substituting from (3.44) into (4.20) and simplifying the resulting expression:

$$V_{BU} = 8\sum_{i=1}^{N} \frac{\phi_i \sigma_i^2}{(1+\phi_i)^3} \tag{4.21}$$

If $\phi_1 = \phi_2 = \hat{\phi}$, then:

$$V_{BU} = \frac{8\hat{\phi}\sum_{i=1}^{N}\sigma_i^2}{(1+\hat{\phi})^3} \tag{4.22}$$

<u>*Case 2*</u>: $-1 < \phi_i \le 1/3$. In this case, $\alpha_i^* = \min\{\alpha_i\}$

Substituting the above into (4.20):

$$V_{BU} = 2\sum_{i=1}^{N} \frac{\sigma_i^2}{(1+\phi_i)(1-\phi_i+\alpha_i^*\phi_i)(2-\alpha_i^*)} \tag{4.23}$$

If $\phi_1 = \phi_2 = \hat{\phi}$, then:

$$V_{BU} = \frac{2}{(1+\hat{\phi})}\sum_{i=1}^{N} \frac{\sigma_i^2}{(1-\hat{\phi}+\alpha_i^*\hat{\phi})(2-\alpha_i^*)} \tag{4.24}$$

For the TD strategy, the derivation begins by first solving for $Cov(d_{i,t}, D_{t-k})$:

$$\begin{aligned}
Cov(d_{i,t}, D_{t-k}) &= Cov\left[(1-\hat{\phi})\mu_i + \hat{\phi}d_{i,(t-1)} + \varepsilon_{i,t}, D_{t-k}\right] \\
&= \hat{\phi}Cov(d_{i,(t-1)}, D_{t-k}) + Cov(\varepsilon_{i,t}, D_{t-k}) \\
&= \hat{\phi}^k Cov\left[(1-\hat{\phi})\mu_i + \hat{\phi}d_{i,(t-k-1)} + \varepsilon_{i,(t-k)}, (1-\hat{\phi})\mu + \hat{\phi}D_{t-k-1} + \hat{\varepsilon}_{t-k}\right] \\
&= \hat{\phi}^k\left[\hat{\phi}^2 Cov(d_{i,(t-k-1)}, D_{t-k-1}) + Cov(\varepsilon_{i,(t-k)}, \hat{\varepsilon}_{t-k})\right]
\end{aligned}$$

Recall that $Cov(\varepsilon_{i,(t-k)}, \varepsilon_{i,(t-k)}) = Var(\varepsilon_{i,(t-k)}) = \sigma_i^2$, $Cov(\varepsilon_{i,(t-k)}, \varepsilon_{j,(t-k)}) = \rho_{ij}\sigma_i\sigma_j$ and $Cov(\varepsilon_{i,t}, D_{t-k}) = 0$ (for all $k > 0$). Through recursive substitutions as shown above, it becomes:

$$Cov(d_{i,t}, D_{t-k}) = \lim_{M\to\infty}\begin{pmatrix} \hat{\phi}^k\left[\hat{\phi}^M Cov(d_{i,(t-k-M/2)}, D_{t-k-M/2})\right] \\ +\hat{\phi}^k\sigma_i\left(\sum_{j=0}^{(M-2)/2}\hat{\phi}^{2j}\right)\sum_{j=1}^{N}\rho_{ij}\sigma_j \end{pmatrix}$$

Taking the limit, $M \to \infty$,

$$Cov(d_{i,t}, D_{t-k}) = \sigma_i \left(\frac{\hat{\phi}^k}{1-\hat{\phi}^2} \right) \sum_{j=1}^{N} \rho_{ij} \sigma_j \text{ , for all } k \geq 0 \tag{4.25}$$

The value of $Cov(d_{i,t}, F_t)$ can be solved as follows:

$$Cov(d_{i,t}, F_t) = Cov\left[d_{i,t}, \sum_{k=1}^{\infty} \alpha (1-\alpha)^{k-1} D_{t-k} \right]$$

$$= \alpha \sum_{k=1}^{\infty} (1-\alpha)^{k-1} Cov(d_{i,t}, D_{t-k}) \tag{4.26}$$

By substituting (4.25) into (4.26),

$$Cov(d_{i,t}, F_t) = \frac{\alpha \hat{\phi} \sigma_i \sum_{j=1}^{N} \rho_{ij} \sigma_j}{(1-\hat{\phi}^2)(1-\hat{\phi}+\alpha\hat{\phi})} \tag{4.27}$$

The variance of forecast error for the item *i* can thus be computed by substituting (2A.9), (3.69), and (4.27) into (4.5). Hence,

$$\begin{aligned} Var(d_{i,t} - p_i F_t) = {} & \frac{\sigma_i^2}{1-\hat{\phi}^2} + \frac{p_i^2 \alpha (1+\hat{\phi}-\alpha\hat{\phi}) \left(\sum_{l=1}^{N} \sigma_l^2 + 2 \sum_{l=1}^{N-1} \sum_{j=l+1}^{N} \rho_{lj} \sigma_l \sigma_j \right)}{(1-\hat{\phi}^2)(1-\hat{\phi}+\alpha\hat{\phi})(2-\alpha)} \\ & - \frac{2 p_i \alpha \hat{\phi} \sigma_i \sum_{j=1}^{N} \rho_{ij} \sigma_j}{(1-\hat{\phi}^2)(1-\hat{\phi}+\alpha\hat{\phi})} \end{aligned} \tag{4.28}$$

Finally, the variance of forecast error for the TD strategy (V_{TD}) is derived below. It is then followed by the comparison of V_{TD} with the variance of forecast error for the BU strategy (V_{BU}) by calculating the ratio V_{TD}/V_{BU}. Note that the *optimal* smoothing constant (α^*) for the aggregate forecast is given by (3.72).

<u>*Case 1*</u>: $1/3 < \hat{\phi} < 1$. In this case, $\alpha^* = (3\hat{\phi}-1)/2\hat{\phi}$

Substituting (4.28) into (4.2) and using the optimal smoothing constant above,

$$V_{TD} = \frac{1}{1-\hat{\phi}^2}\sum_{i=1}^{N}\left(\begin{array}{l}\sigma_i^2 + \dfrac{p_i^2(3\hat{\phi}-1)(3-\hat{\phi})\left(\sum_{l=1}^{N}\sigma_l^2 + 2\sum_{l=1}^{N-1}\sum_{j=l+1}^{N}\rho_{lj}\sigma_l\sigma_j\right)}{(1+\hat{\phi})^2} \\ -\dfrac{2p_i(3\hat{\phi}-1)\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j}{(1+\hat{\phi})}\end{array}\right) \quad (4.29)$$

Using (4.22) and (4.29), and with some further simplifications, it results in:

$$\begin{aligned} V_{TD}/V_{BU} = & \frac{(1+\hat{\phi})^2}{8\hat{\phi}(1-\hat{\phi})} + \frac{(3\hat{\phi}-1)(3-\hat{\phi})\left(\sum_{i=1}^{N}p_i^2\right)\left(\sum_{l=1}^{N}\sigma_l^2 + 2\sum_{l=1}^{N-1}\sum_{j=l+1}^{N}\rho_{lj}\sigma_l\sigma_j\right)}{8\hat{\phi}(1-\hat{\phi})\left(\sum_{i=1}^{N}\sigma_i^2\right)} \\ & -\frac{(3\hat{\phi}-1)(1+\hat{\phi})\sum_{i=1}^{N}\left(p_i\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j\right)}{4\hat{\phi}(1-\hat{\phi})\sum_{i=1}^{N}\sigma_i^2} \end{aligned} \quad (4.30)$$

<u>*Case 2*</u>: $-1 < \hat{\phi} \leq 1/3$. In this case, $\alpha^* = \min\{\alpha\} = 0.01$

Substituting (4.28) into (4.2) and using the optimal smoothing constant above,

$$V_{TD} = \sum_{i=1}^{N}\left(\begin{array}{l}\dfrac{\sigma_i^2}{1-\hat{\phi}^2} + \dfrac{p_1^2\alpha^*(1+\hat{\phi}-\alpha^*\hat{\phi})\left(\sum_{l=1}^{N}\sigma_l^2 + 2\sum_{l=1}^{N-1}\sum_{j=l+1}^{N}\rho_{lj}\sigma_l\sigma_j\right)}{(2-\alpha^*)(1-\hat{\phi}^2)(1-\hat{\phi}+\alpha^*\hat{\phi})} \\ -\dfrac{2p_i\alpha^*\hat{\phi}\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j}{(1-\hat{\phi}^2)(1-\hat{\phi}+\alpha^*\hat{\phi})}\end{array}\right) \quad (4.31)$$

Using (4.24) and (4.31), and the fact that $\alpha_1^* = \alpha_2^* = ... = \alpha_N^* = \alpha^* = \min\{\alpha\}$,

$$V_{TD}/V_{BU} = \frac{\left[\begin{array}{l}(2-\alpha^*)(1-\hat{\phi}+\alpha^*\hat{\phi})\sum_{i=1}^{N}\sigma_i^2 \\ -2\alpha^*\hat{\phi}(2-\alpha^*)\sum_{i=1}^{N}\left(p_i\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j\right) \\ +\alpha^*(1+\hat{\phi}-\alpha^*\hat{\phi})\left(\sum_{i=1}^{N}p_i^2\right)\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right)\end{array}\right]}{2(1-\hat{\phi})\sum_{i=1}^{N}\sigma_i^2} \quad (4.32)$$

From the above expressions, the following important results can be obtained.

Theorem 4.3: *If the time series of the item demands follows an AR(1) process with $\phi_1 = \phi_2 = ... = \phi_N$, where $-1 < \phi_i \le 1/3$ for i = 1, 2, …, N, the difference in the variance of forecast error between TD and BU forecasting strategies for N = 2 is always lower than or equal to 1%. . The difference further increases with the number of products in the family.*

Proof: The proof begins by substituting $\alpha^* = \min\{\alpha\} = 0.01$ in (4.32):

$$V_{TD}/V_{BU} = \frac{\begin{bmatrix} (1.99 - 1.97\hat{\phi})\sum_{i=1}^{N}\sigma_i^2 \\ -0.04\hat{\phi}\sum_{i=1}^{N}\left(p_i\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j\right) \\ +0.01(1+\hat{\phi})\left(\sum_{i=1}^{N}p_i^2\right)\left(\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\sigma_i\sigma_j\right) \end{bmatrix}}{2(1-\hat{\phi})\sum_{i=1}^{N}\sigma_i^2} \tag{4.33}$$

Table 4. 3: Upper and Lower Bounds of V_{TD}/V_{BU} for Different Number of Products in the Family under AR(1) Demand Process ($-1 < \phi_i \le 1/3$)

N	LB	UB	UB - LB
2	0.99	1.01	0.02
3	0.99	1.02	0.04
4	0.98	1.03	0.05
5	0.98	1.04	0.06

Using Solver, the results show that the difference in the variance of forecast error between TD and BU strategies is relatively insignificant when the number of products in the family (*N*) is equal to 2. This implies that in the worst case scenario, one strategy would outperform the other by only 1% margin.

Theorem 4.4: *If the time series of the item demands follows an AR(1) process with $\phi_1 = \phi_2 = ... = \phi_N$, where $1/3 < \phi_i < 1$ for i = 1, 2, …, N, there is a very high probability that* *the TD strategy performs significantly worse than the BU strategy. The difference in performance further increases with the number of products in the family (N).*

Proof: This theorem can be proved by computing (4.30) in Solver; the results of which are presented in Table 4.4. As can be seen, when $N = 2$, the lower bound is only 0.92, meaning that in the worst case scenario, the variance of forecast error of the BU strategy would only be 8% higher than that of the TD strategy. The upper bound, however, indicates otherwise. That is, the variance of forecast error of the TD strategy can be higher than that of the BU strategy by as much as 100 times. Further observation reveals that the difference between the two forecasting strategies increases with N.

Table 4. 4: Upper and Lower Bounds of V_{TD}/V_{BU} for Different Number of Products in the Family under AR(1) Demand Process ($1/3 < \phi_i < 1$)

N	LB	UB	UB - LB
2	0.92	100.00	99.08
3	0.93	149.99	149.07
4	0.93	199.99	199.06
5	0.94	249.99	249.05

4.1.2 Weighted Moving Average (WMA) Forecasting Technique

Under WMA technique, the BU strategy is represented by (3.78), whereas the TD strategy is given by (4.3). To avoid repetition, only the summary of the end results is presented in Table 4.5.

Table 4. 5: Summary of Analytical Evaluation of TD versus BU Strategies for Forecasting Item Level Demands Using Weighted Moving Average Forecasting Technique

	White Noise Demand Process	MA(1) Demand Process	AR(1) Demand Process
Optimal T_i (T_i^*)	$T_i^* = \max\{T_i\}$	$T_i^* = \max\{T_i\}$	$T_i^* = \begin{cases} 1 & 0.43 < \phi_i < 1 \\ \max\{T_i\} & -1 < \phi_i \le 0.43 \end{cases}$
V_{BU}	$\frac{1+T^*}{T^*}\sum_{k=1}^{N}\sigma_k^2$	$\left[\frac{(1+\hat{\theta}^2)(1+T^*)}{T^*}+\frac{2\hat{\theta}}{T^{*2}}\right]\sum_{i=1}^{N}\sigma_i^2$ †	$\left(\frac{1}{1-\hat{\phi}^2}\right)\left[1+\frac{1+\hat{\phi}}{T_i(1-\hat{\phi})}+\frac{2\hat{\phi}(1-\hat{\phi}^{T_i})}{T_i(\hat{\phi}-1)}\left(1-\frac{1}{T_i(\hat{\phi}-1)}\right)\right]\sum_{i=1}^{N}\sigma_i^2$ ‡
Optimal T (T^*)	$T^* = \max\{T\}$	$T^* = \max\{T\}$	$T^* = \begin{cases} 1 & 0.43 < \hat{\phi} < 1 \\ \max\{T\} & -1 < \hat{\phi} \le 0.43 \end{cases}$
V_{TD}	$\sum_{k=1}^{N}\sigma_k^2+\frac{1}{T^*}\left[\sum_{i=1}^{N}\sigma_i^2+2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right]\left(\sum_{k=1}^{N}p_k^2\right)$	$\sum_{i=1}^{N}\left[\begin{array}{l}(1+\hat{\theta}^2)\sigma_i^2+p_i^2\left(\frac{T(1+\hat{\theta}^2)-2\hat{\theta}(T-1)}{T^2}\right) \\ \left[\sum_{i=1}^{N}\sigma_i^2+2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right] \\ +\left(\frac{2\hat{\theta}}{T}\right)p_i\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j\end{array}\right]$	$\left(\frac{1}{1-\hat{\phi}^2}\right)\sum_{i=1}^{N}\sigma_i^2+\left(\sum_{i=1}^{N}p_i^2\right)\left(\frac{1}{T^2(1-\hat{\phi}^2)}\right)\left(\sum_{i=1}^{N}\sigma_i^2+2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right)$ $\left(\frac{T(1-\hat{\phi}^2)-2\hat{\phi}(1-\hat{\phi}^T)}{(\hat{\phi}-1)^2}\right)-\frac{2(\hat{\phi}^{T+1}-\hat{\phi})}{T(1-\hat{\phi}^2)(\hat{\phi}-1)}\sum_{i=1}^{N}\left(p_i\sigma_i\sum_{l=1}^{N}\rho_{il}\sigma_l\right)$
V_{TD}/V_{BU}	$\frac{T^*}{1+T^*}+\frac{\frac{1}{(1+T^*)}\left[\sum_{i=1}^{N}\sigma_i^2+2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right]\left(\sum_{k=1}^{N}p_k^2\right)}{\sum_{k=1}^{N}\sigma_k^2}$	$\frac{\left(\begin{array}{l}(1+\hat{\theta}^2)\sum_{i=1}^{N}\sigma_i^2+\left[\sum_{i=1}^{N}p_i^2\right]\left(\frac{T(1+\hat{\theta}^2)-2\hat{\theta}(T-1)}{T^2}\right) \\ \left[\sum_{i=1}^{N}\sigma_i^2+2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right] \\ +\left(\frac{2\hat{\theta}}{T}\right)\sum_{i=1}^{N}\left[p_i\sigma_i\sum_{j=1}^{N}\rho_{ij}\sigma_j\right]\end{array}\right)}{\left[\frac{(1+\hat{\theta}^2)(1+T)}{T}+\frac{2\hat{\theta}}{T^2}\right]\sum_{i=1}^{N}\sigma_i^2}$	$\left[1+\frac{1+\phi_i}{T_i(1-\phi_i)}+\frac{2\phi_i(1-\phi_i^{T_i})}{T_i(\phi_i-1)}\left(1-\frac{1}{T_i(\phi_i-1)}\right)\right]^{-1}$ $\left[1+\frac{\left(\sum_{i=1}^{N}p_i^2\right)\left(\frac{T(1-\hat{\phi}^2)-2\hat{\phi}(1-\hat{\phi}^T)}{T^2(\hat{\phi}-1)^2}\right)\left(\sum_{i=1}^{N}\sigma_i^2+2\sum_{i=1}^{N-1}\sum_{l=i+1}^{N}\rho_{il}\sigma_i\sigma_l\right)}{\left(\sum_{i=1}^{N}\sigma_i^2\right)} - \frac{\frac{2(\hat{\phi}^{T+1}-\hat{\phi})}{T(\hat{\phi}-1)}\sum_{i=1}^{N}\left(p_i\sigma_i\sum_{l=1}^{N}\rho_{il}\sigma_l\right)}{\left(\sum_{i=1}^{N}\sigma_i^2\right)}\right]$
Conclusion[ξ]	**Theorem 3.8:** If the time series of the item demands follows a white noise process, the performance of the TD and BU strategies becomes more identical as the historical period (T) increases and the number of	**Theorem 3.9:** If the time series of the item demands follows an MA(1) process with $\theta_1 = \theta_2 = ... = \theta_N$, the performance of the TD and BU strategies becomes more identical as the	**Theorem 3.10:** If the time series of the item demands follows an AR(1) process with $\phi_1 = \phi_2 = ... = \phi_N$ and $0.43 < \phi_N < 1$, there is a very high probability that the TD strategy may perform worse than the BU strategy. The difference in performance further

	products in the family (N) decreases.	historical period (T) increases and the number of products in the family (N) decreases.	increases with the number of products in the family (N). **<u>Theorem 3.11:</u>** If the time series of the item demands follows an AR(1) process with $\phi_1 = \phi_2 = ... = \phi_N$ and $-1 < \phi_N \leq 0.43$, the performance of the TD and BU strategies becomes more identical as the historical period (T) increases and the number of products in the family (N) decreases.

† Based on the assumption that $\theta_1 = \theta_2 = ... = \theta_N = \hat{\theta}$; ‡ Based on the assumption that $\phi_1 = \phi_2 = ... = \phi_N = \hat{\phi}$; ξ Numerical results are presented in Tables A.2 to A.5 in the Appendix.

4.2 Simulation Study

This section reports the results of the simulation study which was conducted to evaluate the relative effectiveness of TD over BU strategy for forecasting the individual product demands under various time series models. The simulation study is carried out for different values of the various statistical parameters provided in Table 3.2. The number of products in the family is restricted to two and the contribution achieved by each strategy is compared by a simple ratio shown in (3.80). Due to the similarity in the simulation results, only those obtained by SES is presented.

4.2.1 White Noise and First-Order Moving Average [MA(1)] Demand Processes

The results of the simulation study under white noise and MA(1) demand processes are illustrated in Figure 4.1. It plots the impact of varying the coefficient of correlation between the error terms of the item demand processes (ρ_{12}) on the average performance ratio of TD and BU forecasting strategies (∇) for different values of θ_1 and θ_2. (Note that $\theta_1 = \theta_2 = 0$ corresponds to a white noise time series model.) Each data point plotted in the graph is obtained by computing the average value of ∇ across all possible combinations of p_1, θ_1, and θ_2 as shown in Table 3.2.

In general, Figure 4.1 shows that the difference in performance between the two forecasting strategies under all time series models is remarkably small, with the value of ∇ very close to 1. This is in agreement with the analytical findings which suggest that the performance difference between TD and BU strategies is insignificant. Interestingly, this finding also holds true for the case of MA(1) demand process with $\theta_1 \neq \theta_2$. In this case, the performance of the two forecasting strategies is even closer as compared to the other cases when $\theta_1 = \theta_2 = 0$ (i.e. white noise process) or $\theta_1 = \theta_2 \neq 0$.

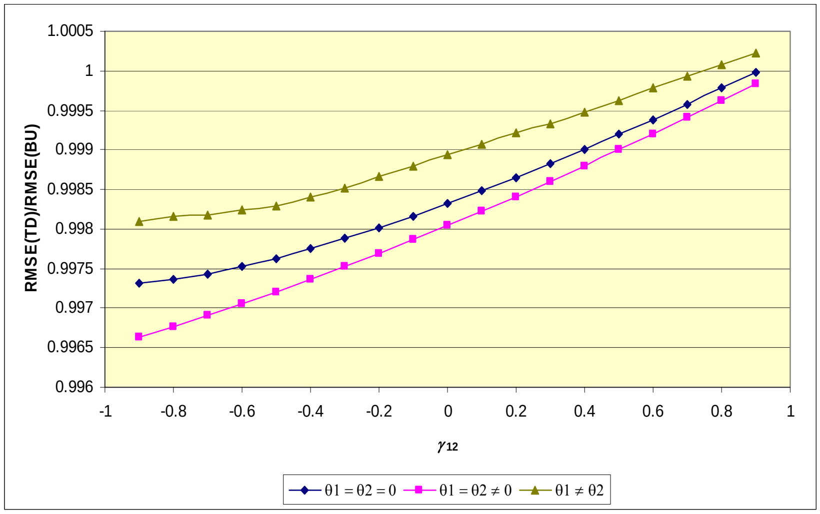


Figure 4. 1: Average Performance Ratio between TD and BU Strategies under Different Values of γ_{12}, θ_1 and θ_2

Furthermore, it can be observed that even though the performance difference between TD and BU strategies is negligible across all values of demand correlation and all time series models, the variance of forecast error for the TD strategy appears to be smaller than the BU strategy in a majority of the cases. This is, again, consistent with the analytical finding (especially for the case when $\theta_1 = \theta_2 = 0$; see Theorem 4.1). This might probably be caused by the error cancellation effect which makes the family demand to be less volatile, as positive error values of one product tend to cancel out negative error values from the other product (risk-pooling). The BU strategy, however, does not experience such a phenomenon as, in this case, the forecast is carried out individually for each of the item demand.

4.2.2 First-Order Autoregressive [AR(1)] Demand Process

The discussion of the simulation results begins with the case of $\phi_1 = \phi_2 = \hat{\phi}$. An initial investigation using a paired difference comparisons test (*t*-test) suggests that the performance of the two forecasting strategies is indeed different across all values of p_i, ϕ_i, and ρ_{12}, as their *p*-values are consistently lower than 0.05 significance level. Furthermore, based on the results of the ANOVA test shown in Table 4.6, p_i, ϕ_i, and ρ_{12} seems to significantly affect the performance ratio ∇.

Table 4. 6: Summary of Analysis of Variance (ANOVA) Test with Identical Coefficients ($\phi_1 = \phi_2$)

Source	Sum of Squares	Df	Mean Square	*F*-value	Sig. Level
p_i	1.14×10^{-2}	1	1.14×10^{-2}	27.06	< 0.0001
ϕ_i	17.31	9	1.92	4571.36	< 0.0001
ρ_{12}	0.18	18	9.82×10^{-3}	23.35	< 0.0001
$p_i * \phi_i$	8.86×10^{-2}	9	9.84×10^{-3}	23.39	< 0.0001
$p_i * \rho_{12}$	5.84×10^{-2}	18	3.24×10^{-3}	7.71	< 0.0001
$\phi_i * \rho_{12}$	1.19	162	7.32×10^{-3}	17.40	< 0.0001
$p_i * \phi_i * \rho_{12}$	0.18	162	1.11×10^{-3}	2.65	< 0.0001
Error	0.16	380	4.21×10^{-4}		
Total	19.17	759			

The detailed results are illustrated in Figures 4.2. In this figure (and also in Figure 4.3 which will be discussed later), the performance ratio ∇ is plotted on *y*-axis and is studied with respect to two particular parameters. The results plotted in Figure 4.2 assume that the value of p_1 is kept fixed at 0.3. The results for the other values of p_1 are not presented here due to their similarity in the pattern to Figure 4.2. As can be seen, for the case when $-1 < \hat{\phi} \leq 0.3$, the values of ∇ are generally close to one, meaning that no significant difference in performance is discovered between TD and BU strategies. This, however, changes for $0.3 < \hat{\phi} < 1$, as the corresponding ∇ values are significantly higher than one across all values of γ_{12}. These results are in accordance to the analytical findings (see Theorems 4.3 and 4.4).

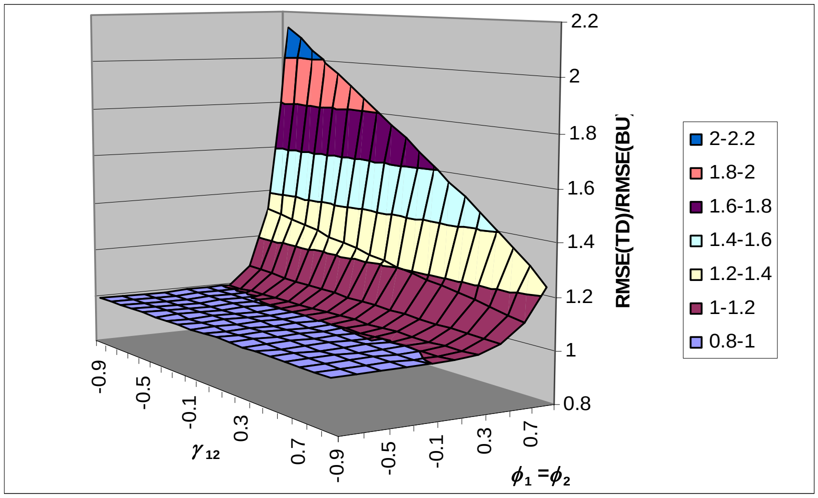


Figure 4. 2: Impact of γ_{12}, ϕ_1 and ϕ_2 on Relative Performance of TD and BU Strategies ($\hat{\phi} = \phi_1 = \phi_2$ and $p_1 = 0.3$)

Table 4. 7: Summary of Analysis of Variance (ANOVA) Test with Non-Identical Coefficients ($\phi_1 \neq \phi_2$)

Source	Sum of Squares	Df	Mean Square	*F*-value	Sig. Level
p_i	2.72×10^{-2}	1	2.72×10^{-2}	46.05	< 0.0001
ϕ_i	1.42	9	0.16	266.68	< 0.0001
ρ_{12}	2.60×10^{-2}	18	1.44×10^{-3}	2.44	< 0.0001
$p_i * \phi_i$	0.13	9	1.45×10^{-2}	24.46	< 0.0001
$p_i * \rho_{12}$	2.8×10^{-2}	18	1.56×10^{-3}	2.63	< 0.0001
$\phi_i * \rho_{12}$	0.30	162	1.88×10^{-3}	3.18	< 0.0001
$p_i * \phi_i * \rho_{12}$	9.47×10^{-2}	162	5.85×10^{-4}	0.99	0.53
Error	0.23	380	5.92×10^{-4}		
Total	2.72	760			

The next experimental findings are obtained for $\phi_1 \neq \phi_2$. The results of the ANOVA test in Table 4.7 show statistical evidences that both p_i, ρ_{12}, and ϕ_i have a significant influence to the relative benefit of TD over BU forecasting strategy. Figure 4.3 plots the impact of demand correlation (γ_{12}) and component ϕ_2 on the relative performance of TD over BU forecasting strategy (∇) when the value of components ϕ_1 and p_1 are kept fixed at 0.5 and 0.3, respectively. The value of ϕ_1 = 0.5 is arbitrarily selected as the pattern in Figure 4.3 does not vary significantly for other values of ϕ_1 between 1/3 to 1.

For a similar reason, the other results of ϕ_1 between -1 and 1/3 are also not reported in this dissertation, as the pattern is similar to that in Figure 4.2. This, therefore, leads to the conclusion that when the lag-1 autocorrelation of one of the item demands satisfies $-1 < \phi_1 \leq 1/3$, the RMSE value of the BU strategy is identical or close to the RMSE value of the TD strategy, so long as the lag-1 autocorrelation for the other item demand also satisfies $-1 < \phi_2 \leq 1/3$. This finding holds true across all values of γ_{12} and p_i, and whether or not the demand process of the two items has the same level of lag-1 autocorrelation. One possible explanation for this phenomenon is that when the item demand has a very low lag-1 autocorrelation, the underlying demand pattern tends to be highly fluctuating (large demand interspersed with low demand). As a result, according to (3.2), the optimal value of α_i in this case should be set as low as possible. This would, in turn, affect the performance of the BU strategy negatively as its best estimate would involve very little adjustment over time.

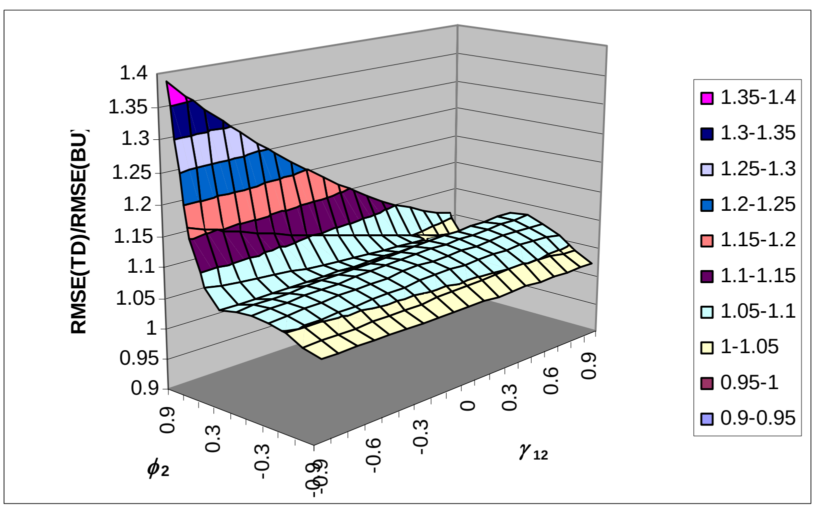


Figure 4. 3: Impact of γ_{12} and ϕ_2 on Relative Performance of TD and BU Strategies ($\phi_1 = 0.5$ and $p_1 = 0.3$)

Another important observation from Figure 4.3 is that the BU strategy consistently outperforms the TD strategy under all experimental conditions. Regardless of how the two item demands interact with each other in terms of the correlation value, γ_{12}, the BU strategy always produces more accurate estimates than the TD strategy. The impact of γ_{12} on the performance ratio ∇ appears to be negligible when the coefficient ϕ_2 ranges between -1 and 0.3. However, as the value of ϕ_2 increases and becomes greater than 0.3, the impact of γ_{12} becomes more significant. That is, a reduction in the value of γ_{12} improves the relative superiority of BU strategy over TD strategy. This is in contradiction to some of the earlier studies (Dangerfield and Morris, 1992, and Gordon *et al.*, 1997), which argued that a lower demand correlation helps to stabilize the aggregate family demand and thus, positively affects the performance of the TD strategy. Finally, for high values of ϕ_2 and low values of γ_{12}, the RMSE of the TD strategy is higher than the RMSE of the BU strategy by as much as 40%.

4.3 Conclusions

This chapter evaluates the performance of TD and BU strategies for forecasting the product demands belonging to a product family. It has identified the conditions under which one forecasting strategy would be preferred over the other. The findings are summarized below:

- When simple exponential smoothing is employed as the forecasting technique under both strategies and N = 2, the difference in the variance of forecast error between TD and BU strategies was relatively insignificant. For instance, when the item demands follow a white noise or an MA(1) process, the difference between the two strategies would never exceed 1%, regardless of the item's proportion in the family (p_i), the coefficient of correlation between the demands of the two items (γ_{12}), and the coefficient of the correlation term (θ_i);
- The above finding also held true for the case of AR(1) demand process with $-1 < \phi_i \leq 1/3$ (for i = 1, 2). However, if the lag-1 autocorrelation (ϕ_i) of the demand

for at least one of the items was greater than 1/3, there was a very high probability that the TD strategy might perform worse than the BU strategy, irrespective to p_i and γ_{12}. The superiority of the BU strategy further increased with the value of lag-1 autocorrelation and with the decrease in correlation between the demand of the two items;

- When $N > 2$, the difference in performance between the two strategies increased with N and as T decreased. This finding was valid for both exponential smoothing and moving average forecasting techniques.

The next chapter examines, through a simulation study, the relative effectiveness of TD and BU strategies for forecasting the demand of a substitutable product (that belongs to a family) as well as the demand of a product family under different types of family demand processes. In addition to the three stationary time series models used in this chapter, one non-stationary time series model, known as integrated moving average of order 1 [IMA(1, 1)], is also considered.

CHAPTER 5

FORECASTING SUBSTITUTABLE PRODUCTS

5.1 Introduction

This chapter studies the relative benefit of top-down (TD) over bottom-up (BU) forecasting strategy when the individual products in the family are substitutable. *Substitutable* product is defined as a case where the demand for a particular product is no longer driven by its unique characteristics, but also by the inventory level of another product with similar characteristics[12]. Product substitutability has long been a major concern in a production planning environment due to its negative effect on the accuracy of the point-of-sale (POS) data. Companies have to make sure that they do not over-forecast the demand for a product, when its observed demand might be inflated because of the additional demand from the other similar product, and vice versa[13]. Another complexity associated with the product substitutability is that even if the companies, by any means, know which products are substitutable, it is still difficult to measure the consumer's willingness to buy another product in case his preferred product is out of stock. Many researchers have studied this issue from an inventory management perspective. The most recent ones include Anupindi, Dada, and Gupta (1998), Smith and Agrawal (2000) and Rajaram and Tang (2001).

This chapter, however, addresses the problem of managing substitutable products from a different perspective. Specifically, the relative effectiveness of top-down (TD) and bottom-up (BU) forecasting strategies in estimating the demand of a substitutable product

---

[12] A survey conducted by Food Marketing Institute (1993) reported that the majority of the typical shoppers (82% - 88%) would be willing to buy another size of the same brand, or switch brands, if their favorite brand-size was not available on a shopping trip; while the rest indicated that they would not buy any items.
[13] A survey by Harvard/Wharton Merchandising Effectiveness Project revealed that the average forecast error in the retail industry is 50% or more (Fisher and Raman, 1996).

(that belongs to a family) as well as the demand of a product family under different types of family demand processes is examined through a simulation study. As is common in practice, it is assumed that the analysts responsible for forecasting have no means to know the *actual* demand of each product. Thus, only the information on the *observed* demand data is available.

The main difference between this study and all the previous research in forecasting is that the focus of investigation is on the impact of (i) the degree of product substitutability and (ii) the variability of demand quantity of each product in the family, on the relative performance of the two forecasting strategies. To the best of knowledge, there has been no study done on this topic. DeLurgio and Bhame (1991, p. 122), for example, suggested that when substitutions occur, it is important to record the demand history in such a way as to not distort the demand forecasts that will use that history. However, the authors did not provide any clear conclusion as to which forecasting strategy should be adopted when the information on the actual demand process is impossible to obtain. In another study, Lapide (1998) suggested that the BU strategy is preferred over the TD strategy when the product family is composed of several competing items that potentially cannibalize each other's demand. However, the author did not elaborate further how the change in the degree of product substitutability would affect this preference. The companion studies and other research by Gross and Sohl (1990), Dangerfield and Morris (1992), and Zellner and Tobias (2000), to name a few, did not consider substitutable products in their modeling framework at all.

Another contribution of this study is the assumption that the product family demand (instead of the product demands as in other papers) follows a certain time series process, and the product demands are derived from the product family demand. This assumption is motivated by the fact that in the retail merchandising industry, the individual product demands are usually more erratic than the product family demand. The fluctuations of several product demands tend to cancel each other out when aggregated together, thus resulting in a smoother family demand. Hence, it is common to find in reality where the product family demand may be identified to follow a particular time series model, while the individual product demands are simply derived from the product family demand based upon a certain proportion.

The remainder of this chapter is organized as follows. In Section 5.2, the design of the simulation experiments is discussed. Section 5.3 analyzes the results of the simulation study. Finally, Section 5.4 presents some concluding remarks.

5.2 Design of Simulation Experiments

As mentioned previously, the focus of this research is to examine the impact of the product demand variability and the degree of product substitutability on the performance of TD and BU forecasting strategies. Therefore, in order to obtain meaningful insights, the number of products in the family is restricted to two. Most of the earlier studies have also considered two components in the family (see for instance, Schwarzkopf *et al*., 1988, and Dangerfield and Morris, 1992). In addition, the demand process for the product family is assumed to follow a particular time series, whereas the demand for each individual product is derived from the product family demand. The demand proportion for each product in the family, p_i ($0 \le p_i \le 1$), is assumed to be uniformly distributed, $p_i \sim U(LB,UB)$, where $UB > LB$ and $\tau = UB - LB$. In the rest of the study, τ, the range of variability of the product's proportion in the family, is also referred to as the product demand variability because of its positive correlation to the level of demand uncertainty of each individual product.

Four time series (i.e. 3 stationary and 1 non-stationary) are considered to represent the demand process of the product family. Let $D(t)$ be the family demand in period t. The four processes for the family demand are as follows:

(i) White noise: $D(t) = \mu + \varepsilon(t)$ (5.1)

(ii) First-order moving average [MA(1)]: $D(t) = \mu + \varepsilon(t) - \theta\varepsilon(t-1)$ (5.2)

(iii) First-order autoregressive [AR(1)]): $D(t) = (1-\phi)\mu + \phi D(t-1) + \varepsilon(t)$ (5.3)

(iv) Integrated moving average of order 1 [IMA(1, 1)]:

$$D(t) = D(t-1) + \varepsilon(t) - \theta\varepsilon(t-1) \quad (5.4)$$

where $|\theta| < 1$, $|\phi| < 1$, μ is the expected demand, and $\varepsilon(t)$ is a random variable that is normally distributed with zero mean $\left(E\left[\varepsilon(t)\right] = 0\right)$, variance σ^2, and zero autocovariance $\left(Cov\left[\varepsilon(t), \varepsilon(t-k)\right] = 0\right)$, for all $k \neq 0$. Note that coefficients θ and ϕ in (5.2) to (5.4) introduce the serial correlation of the demand process, i.e. demand for period t is dependent on demand for the earlier periods. Therefore, in the rest of the study, θ and ϕ are identified as the coefficient of the serial correlation term. Also note that in period $t = 0$, equations (5.2) to (5.4) are simplified to (5.1). The simple exponential smoothing (which is known for its wide usage in the industry due to its simplicity and robustness (Dekker *et al.*, 2004)) is used as the forecasting technique under both the TD and BU strategies.

The degree of product substitutability (or the substitutability ratio) for product *i* is denoted as β_{ij}, where $0 \leq \beta_{ij} \leq 1$. This represents a deterministic portion of the unsatisfied / excess demand for product *i* which is passed to product *j*, provided that product *j* has sufficient inventory. Clearly, $\beta_{ij} = 0$ implies that none of the excess demand for product *i* would be passed to and fulfilled by product *j*. This represents the case when the consumer would rather wait for his preferred product than buy the alternative. For simplicity, it is assumed that the two products have same level of importance, quality, and consumer perception, meaning that the degree of substitutability from products *i* to *j* is equal to that from products *j* to *i* (i.e. $\beta_{ij} = \beta_{ji} = \beta$).

The detailed explanation on the behavior of the consumers to purchase substitutable products is provided in the Appendices. In summary, product substitution takes place only if part of the substituted demand can be fulfilled with the on-hand inventory of the alternative product. If this condition is not satisfied, no product substitution happens and thus, the entire unmet demand will be backordered to the respective product. In marketing management, this mechanism is known as *consumer-controlled* substitution. Note that no part of the unmet demand is lost. In other words, all excess demands are assumed to be completely backordered. This option is adopted due to the complexity in estimating the unobservable lost sales. Unless the consumer's intention can be clearly identified, it is almost impossible to accurately estimate the probability of lost sales by only using sales information from POS (Tan and Karabati, 2000).

The sequence of activities for the retailer in each time period of the simulation run is as follows. First, the shipments from the outside supplier (shipped L periods ago, where L is the replenishment lead-time) are received into the inventory. Next, the demand for the two products arises and is immediately satisfied, if there are adequate stocks, or backlogged, otherwise. Then, the backlogged demand from the previous periods, if any, is satisfied, if there is stock. It is assumed that partial fulfillment of demand is allowed. The demand forecasts (based on either TD or BU strategies) are then updated, followed by computation of the safety stock[14] and the net requirement. Finally, the order schedule using the lot-for-lot method is calculated and orders corresponding to the current period are placed with the outside supplier, which has infinite supply of products.

Note that for the case of forecasting the product demands, simple exponential smoothing is used to apportion the aggregate forecast [$F(t)$] from the TD strategy to each

---

[14] Safety stock (s) is required in this study because of its influence to the extent of the impact of β on the performance of TD and BU strategies. When s is set too high (e.g. to achieve 99% service level), the number of stockout occurrence would be minimal and therefore, the role of β (for $\beta > 0$) to the performance of the two strategies would be of no value. Intuitively, the use of safety stock becomes less important when $\beta = 0$.

individual product (McLeavey and Narasimhan, 1985, pp. 69-71). Mathematically, it is written as follows:

$$F_1(t) = \left[\hat{\alpha} \frac{D_1(t-1)}{D_1(t-1) + D_2(t-1)} + (1-\hat{\alpha}) \frac{F_1(t-1)}{F(t-1)} \right] F(t) \tag{5.5}$$

$$F_2(t) = F(t) - F_1(t) \tag{5.6}$$

where $F_i(t)$ is the demand forecast for product i at time t (for i = 1, 2), $F(t-1)$ is the aggregate (family) forecast at time t-1, and $\hat{\alpha}$ is a pre-determined (non-optimized[15]) smoothing constant. Notice that $\hat{\alpha}$ used in the disaggregation or allocation method here is different from α_i (for i = 1, 2) and α used in the forecasting process. α_i is the smoothing constant used for forecasting the demand for products 1 and 2 under the BU strategy, while α is the smoothing constant used for forecasting the product family demand under the TD strategy.

In each experiment, the simulation is carried out for a total of 1040 periods. The first 52 periods are used to obtain the initial parameters. The second part containing 260 periods is used to determine the optimal smoothing constants α_i and α (by minimizing the root of mean squared error (RMSE) for this part). Note that in the BU strategy, the smoothing constant α_i is optimized for each product demand individually. Finally, the last 728 periods are used to compile the performance statistics of the two forecasting strategies.

The study is carried out in MS Visual C++ 6.0 and for a wide range of statistical parameters. Values of problem parameters used are provided in Table 5.1.

---

[15] The preliminary experiment showed that, as long as the product demands were randomly drawn from the product family demand, the performance of the TD strategy for forecasting the product demands was not significantly affected by the decision of whether or not $\hat{\alpha}$ was optimized.

Table 5. 1: Problem Parameters Used in the Experiment

Parameter	Value
Expected demand of the product family (μ)	500 units
Lower bound[a] of the product's proportion in the family (LB)	0.3
Delivery lead-time from the external supplier to the retailer (L)	1 week
Grid size used to optimize the smoothing constant	0.01
Smoothing constant used in the proration mechanism ($\hat{\alpha}$)	0.2
Actual fill rate for each product (FR)	95%
Variability[b] of the product's proportion in the family (τ)	(i) From 0.05, 0.1, to 0.7 in increments of 0.1 (ii) From 0.05 to 0.4 in increments of 0.05
Demand variance of the product family (σ^2)	(i) 900 units^2 (ii) 90000 units^2
Coefficient for the MA(1) and IMA(1, 1) processes (θ)	From -0.9 to 0.9 in increments of 0.2
Coefficient for the AR(1) process (ϕ)	From -0.9 to 0.9 in increments of 0.2
Degree of product substitutability (β)	From 0.0 to 1.0 in increments of 0.2

[a] The upper bound of the product's proportion (UB) = $LB + \tau$

[b] There are two sets of data for each category in order to ensure that the simulation study encompasses all possible values of demand correlation between the two products. The combination of low demand variance and high variability of the product's proportion [i.e. (i) – (i)] results in a negative demand correlation between the two products, and vice versa.

For all the problems, the forecast error (RMSE) values of both TD and BU strategies are computed and compared using the ratio:

$$\nabla = RMSE(TD)/RMSE(BU) \tag{5.7}$$

The value of $\nabla < 1$ implies that the TD strategy is superior to the BU strategy, whereas the value of $\nabla > 1$ indicates the contrary. For a particular set of parameters, the simulation is basically run for different values of the safety stock (s) (to be specific, the value of z in $s = z \times RMSE \times \sqrt{L}$ is raised in increments of 0.01 during the model fitting period) until the safety stock that achieves 95% fill rate (FR = 95%) is obtained. The value of ∇ is then calculated corresponding to this level of safety stock (s). This involved running the simulation repeatedly for a particular parameter set until the safety stock level for the required fill rate is determined.

5.3 Analysis of the Results

In this section, the behavior of the performance ratio ∇ under different experimental scenarios is discussed in detail. The explanation is divided into two sub-sections. Section 5.3.1, which refers to Figures 5.1 to 5.3, evaluates the relative effectiveness of TD and BU strategies for forecasting the product demands; whereas Section 5.3.2, referring to Figures 5.4 to 5.6, discusses the case of forecasting the product family demand. Each figure consists of several graphs, each of which plots the simulation results for a specific demand process. In all the graphs, the average performance ratio ∇ is plotted on *y*-axis and is studied with respect to two particular parameters. Note that BO denotes backorders and represents a special case where the excess demand for a particular product is completely backlogged and is not satisfied by the other product (i.e. $\beta = 0$). Therefore, in this case the observed demand used in the forecasting process is always equal to the actual demand $\left[\text{i.e. } \tilde{D}_i(t) = D_i(t)\right]$.

5.3.1 Forecasting Product Level Demands

When the effectiveness of the TD and BU strategies for forecasting the product demands is compared, it is generally found from all the graphs in Figure 5.1 that the relative superiority of the TD strategy over the BU strategy increases with the degree of product substitutability (β). In Figure 5.1d, for instance, the average of the performance ratio ∇ under an IMA(1, 1) demand process appears to decrease by as much as 8.5% as the two products becomes more substitutable with each other.

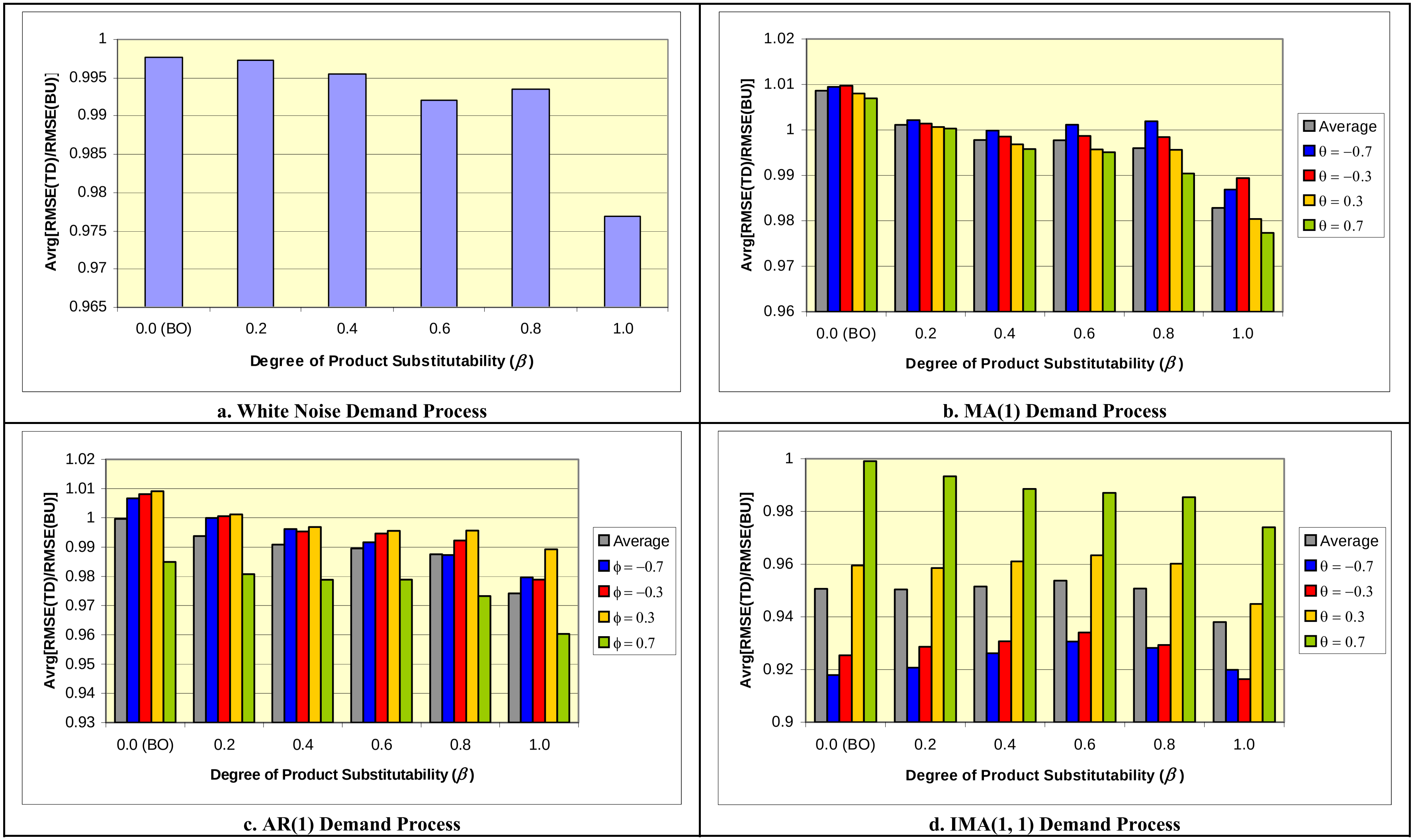
The possible explanation for this phenomenon is that when the two products are highly substitutable, the distortion of the *observed* demand from the *real* demand for each individual product becomes more significant. This is primarily due to the fact that any excess demand for a particular product, say product *i*, would not be immediately visible to the retailer as the consumer might try to satisfy this excess demand (within a certain portion β) by buying the alternative product *j*. The higher the substitutability ratio of a product, the lower the exact amount of the excess demand or the excess inventory known to the retailer. Therefore, the consequence of this information distortion to the performance of the BU strategy is very obvious: for the above example, the retailer is likely to under-forecast the demand for product *i* and to over-forecast product *j*.

The performance of the TD strategy, on the other hand, is not really affected by the information distortion caused by the product substitution, as it uses the historical demand of the product family, which is independent to the degree of product substitutability (β). One may argue, though, that it would still be difficult for the TD strategy to properly allocate the aggregate demand forecast to each individual product, especially when the degree of product substitutability is high. This is because the retailer may not be able to identify the actual demand quantity of each individual product. In this experiment, however, the exponential smoothing (with $\hat{\alpha} = 0.2$) is used as the disaggregation method for the TD strategy [see (5.5) and (5.6)]. The advantage of this method is that it considers the most recent distribution of demands, in addition to taking into account the historical distribution of demands. Therefore, the prediction of the product's proportion in the TD strategy is significantly improved.

The impact of product demand variability (τ) on the performance ratio ∇ is illustrated in Figure 5.2. Each value of ∇ plotted in the graph is obtained by averaging the value of ∇ across different values of ϕ or θ, depending on the time series process. As can be seen, the impact of τ on the value of ∇ depends on the degree of product substitutability (β). Specifically, when the value of β is at low-to-medium level, the increase in τ is found to negatively affect the relative effectiveness of the TD strategy over the BU strategy. For a special case when $\beta = 0$, the RMSE of the TD strategy is even higher than the RMSE of the BU strategy by as much as 4%. This is to be expected as the value of τ has a direct influence on the volatility level of the product demand. As τ increases, the amplification of the product demand also increases, thus making the allocation process of the aggregate forecast in the TD strategy more difficult.

The negative impact of the increase of τ on the TD strategy's performance, however, reduces as β increases. For a special case when $\beta = 1$, increasing τ is actually found to improve the TD strategy's performance. Intuition suggests that this might be due to the "mitigation effect" of β, which stabilizes the individual product demand (when τ is high). This means that the higher the degree of product substitutability (β), the larger the portion of the excess demand for a particular product that would be passed to the other product whose on-hand inventory is in surplus. As a result, the overall product demand variability decreases, thus improving the ability of the TD strategy to accurately distribute the aggregate forecast.

Note that when τ is low ($\tau < 0.2$), the change in β has a little impact on the relative effectiveness of TD over BU strategies. This can be explained with easily. When τ is low, the product demand becomes more stable and easier to estimate. This results in the reduction in the number of stockout occurrences as well as the (potential) number of product substitution occurrences. Note that in Figure 5.2, the result for the white noise demand time series is not presented as it is found that τ has no clear relationship to ∇.



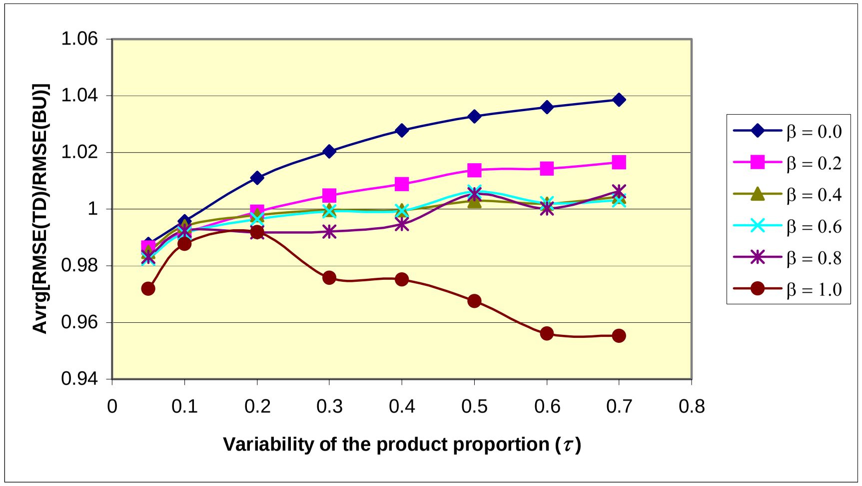
a. White Noise Demand Process

b. MA(1) Demand Process

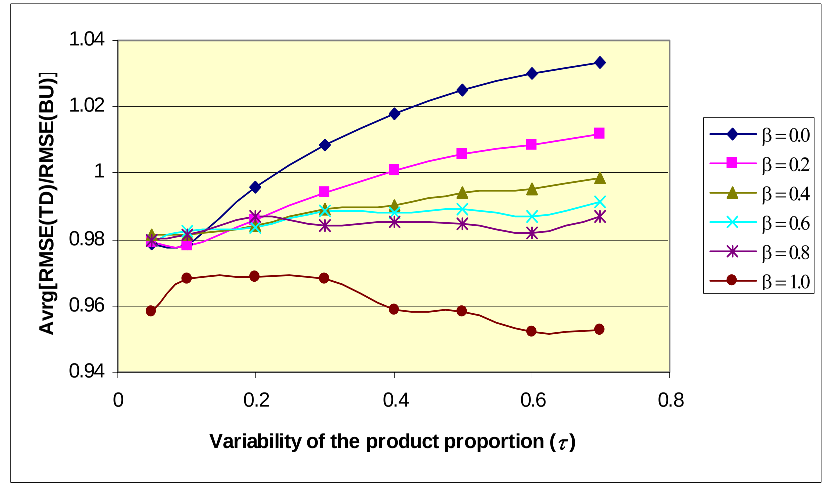
c. AR(1) Demand Process

d. IMA(1, 1) Demand Process

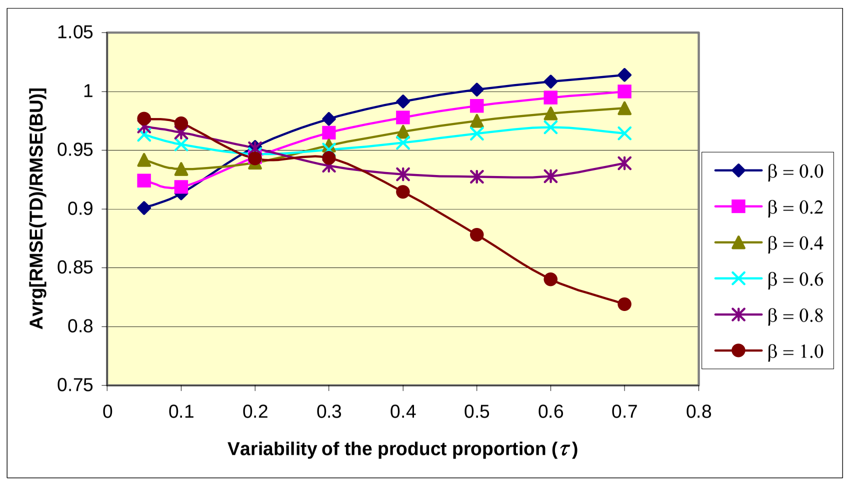
Figure 5. 1: Relative Performance of TD over BU Strategies for Forecasting Product Demands (Different β)



a. MA(1) Demand Process

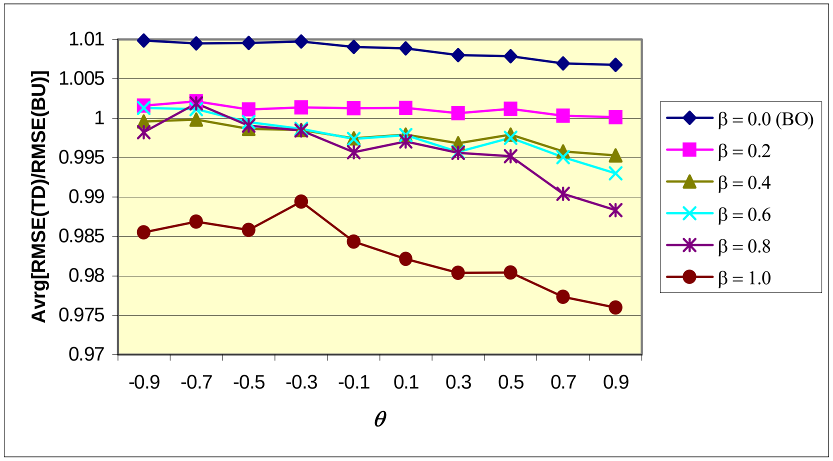


b. AR(1) Demand Process

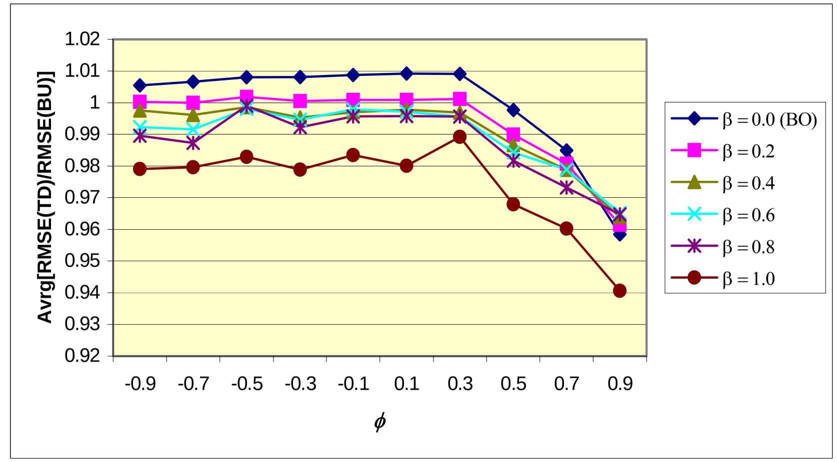


c. IMA(1, 1) Demand Process

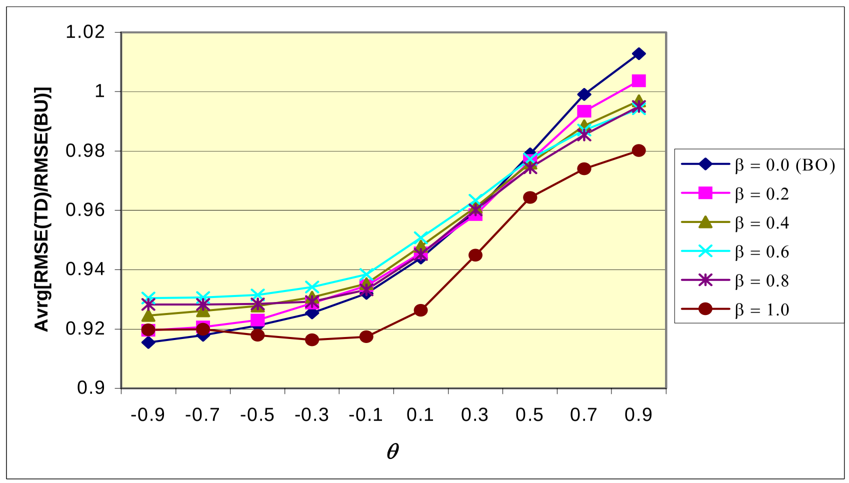
Figure 5. 2: Relative Performance of TD over BU Strategies for Forecasting Product Demands (Different τ)



a. MA(1) Demand Process



b. AR(1) Demand Process



c. IMA(1, 1) Demand Process

Figure 5. 3: Relative Performance of TD over BU Strategies for Forecasting Product Demands (Different Values of Coefficient of the Serial Correlation Term)

Figure 5.3 exhibits the impact of the coefficient of the serial correlation term of the product family demand (i.e. θ and ϕ) on the relative benefit of TD over BU forecasting strategies. Similar to the earlier figures, each value of ∇ plotted in the graph is obtained by averaging the value of ∇ across different values of τ. Except for the case of IMA(1, 1) demand process [Figure 5.3c], the increase in the coefficient of the serial correlation term is found to positively impact the relative performance of the TD strategy over the BU strategy, especially when the value of θ and ϕ are at medium-to-high levels. This can be seen, for instance, in the MA(1) demand process, where ∇ apparently decreases with the increase in θ when $\theta > -0.3$ [Figure 5.3a]. Similarly for the AR(1) process, the decrease in ∇ becomes more obvious when $\phi > 0.3$, regardless of the degree of product substitutability [Figure 5.2b].

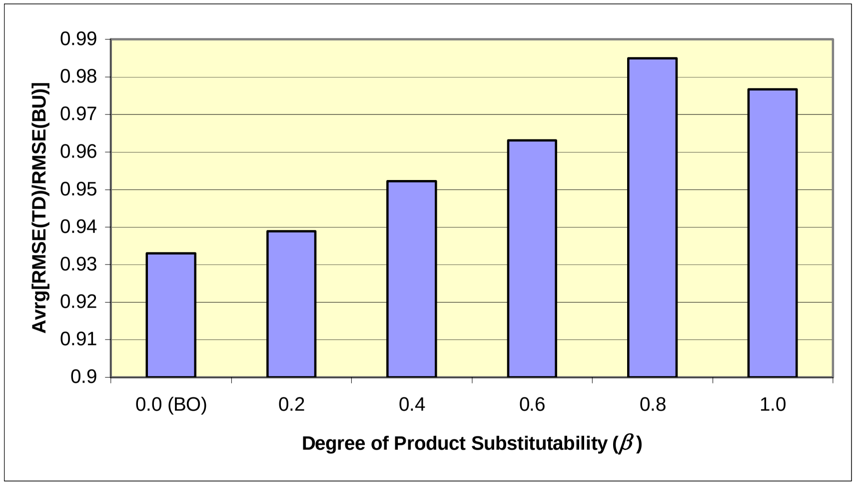
5.3.2 Forecasting Family Level Demand

In this section, the performance of TD and BU strategies for forecasting the product family demand is compared. Unlike the earlier case, from all the figures it is discovered that the TD strategy consistently outperforms the BU strategy by as much as 52%, regardless of the degree of product substitutability (β), the product demand variability (τ), and the coefficient of the serial correlation term (θ and ϕ). Although under some scenarios it is found that $\nabla > 1$ [Figures 5.4b and 5.6a], the relative superiority of the BU strategy over the TD strategy is considered insignificant as it never exceeds by more than 1%. One reason for the superior performance of the TD strategy over the BU strategy may be due to risk-pooling. Risk-pooling takes advantage of the statistical fact that the variance of the aggregated demand is equal to the sum of the variances and covariances of the individual product demand. In this way, the fluctuation of demand from one source may be offset by that from other sources and thus, resulting in a lower forecast error.

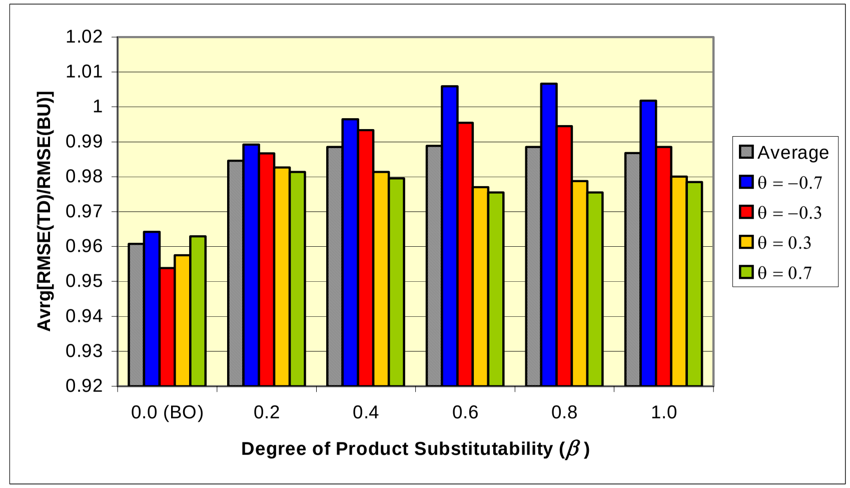
In addition, with the exception of the MA(1) demand time series, Figure 5 shows that the value of ∇ tends to decrease as τ increases. This confirms the general intuition as a highly volatile product demand reduces the ability of the BU strategy to forecast accurately. That is, the higher the uncertainty of the product demand, the higher the forecast error incurred by the BU strategy. The performance of the TD strategy, however, is unaffected by the changes in τ and β due to the distinct approach on how the product demands are generated in this study (see Section 5.2).

The impact of the degree of product substitutability (β) on the value of ∇ can be observed in all the figures. It is generally found that ∇ increases with β, especially when τ is high [Figures 5.5b and c]. This might be due to the mitigation effect of β, which reduces the fluctuation of the product demand and thus, improves the accuracy of the BU strategy (as in the previous case). When τ is low, however, the positive influence of β on the performance of the BU strategy is not that significant as most of the demands from the consumers are fairly predictable.

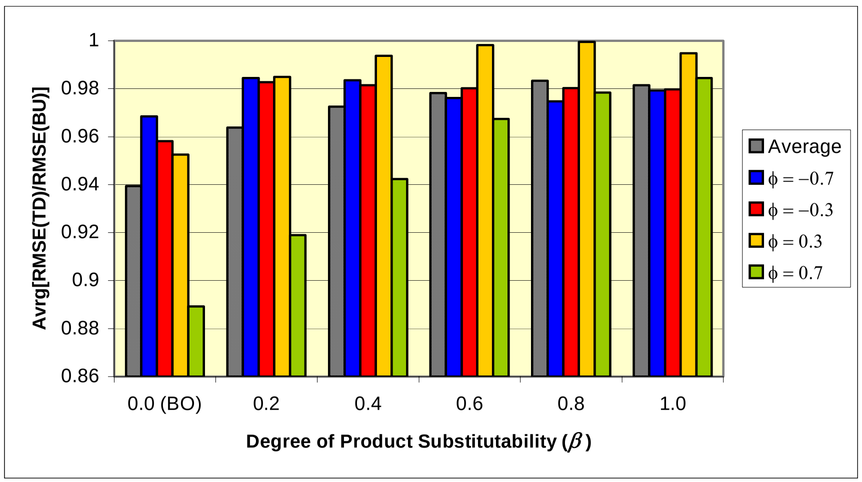
Finally, Figure 5.6 plots the impact of the coefficient of the serial of correlation term of each demand process on the performance ratio ∇. Consistent with the earlier results, it is found that for the case of MA(1) and AR(1) demand processes, the modification of the coefficient of the serial of correlation term significantly affects the relative domination of the TD strategy over the BU strategy only when $\theta > -0.3$ and $\phi > 0.3$, respectively [Figures 5.6a and b]. Furthermore, when the product family demand follows an IMA(1, 1) process, Figure 5.6c suggests that the impact of θ on ∇ becomes more pronounced when $\beta \leq 0.8$; in which case, the RMSE of the TD strategy is lower than the BU strategy by as much as 34%.



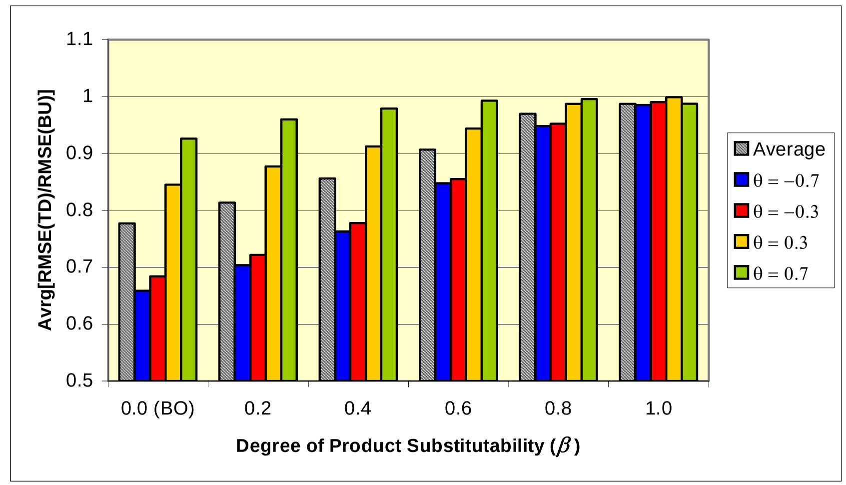
a. White Noise Demand Process



b. MA(1) Demand Process

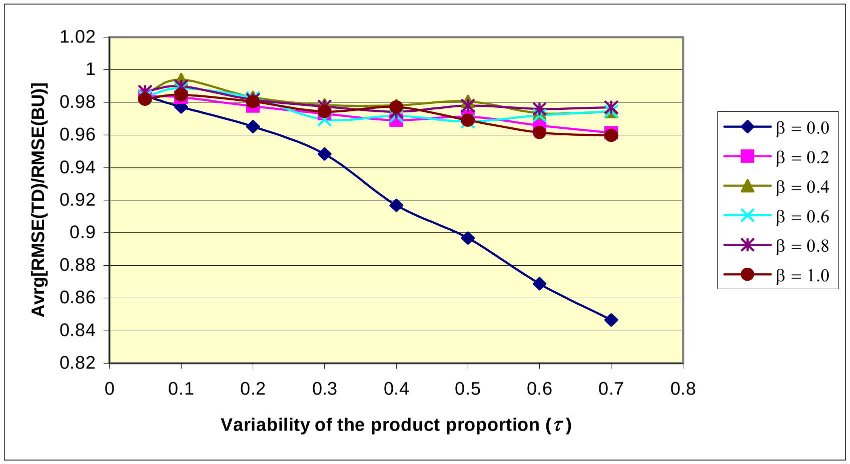


c. AR(1) Demand Process

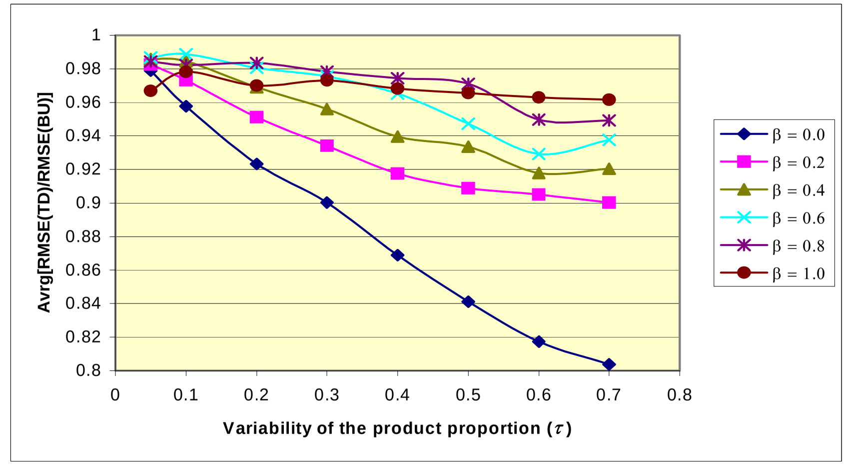


d. IMA(1, 1) Demand Process

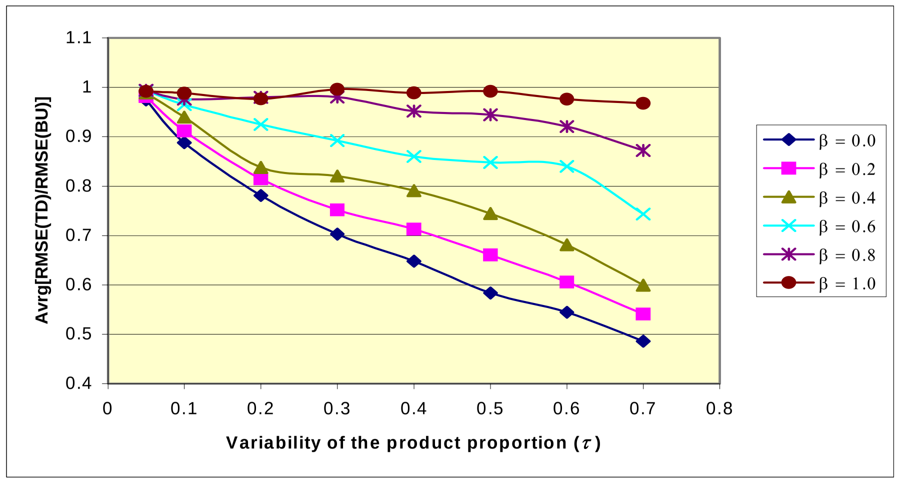
Figure 5. 4: Relative Performance of TD over BU Strategies for Forecasting Family Demand (Different β)



a. MA(1) Demand Process

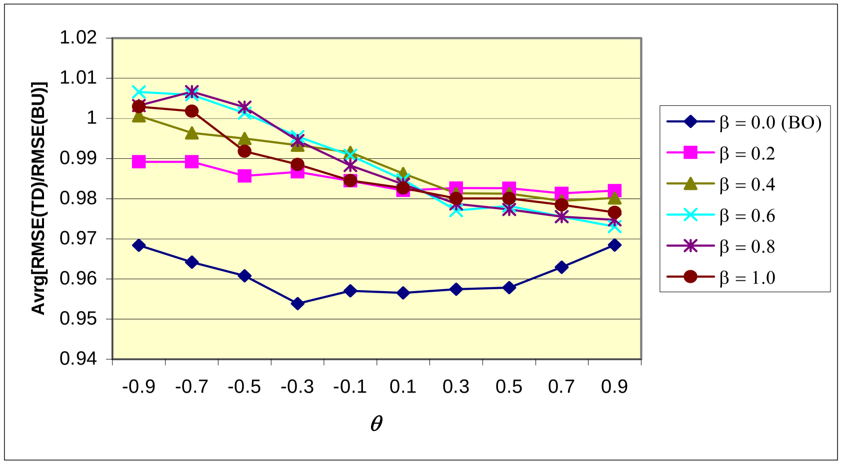


b. AR(1) Demand Process

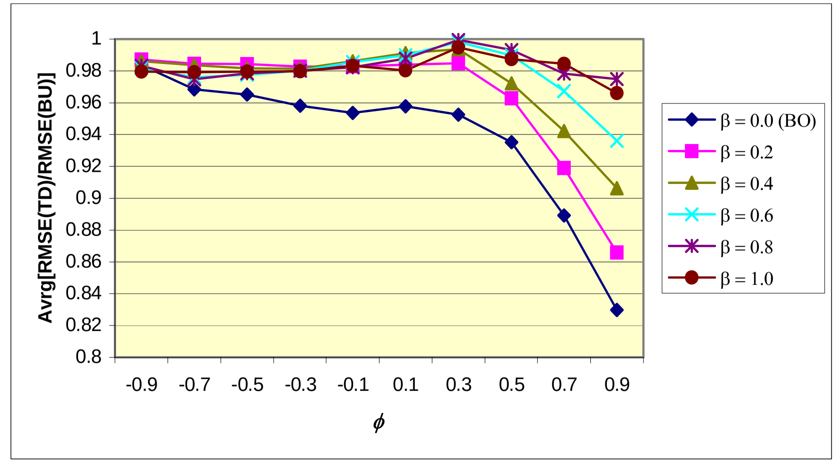


c. IMA(1, 1) Demand Process

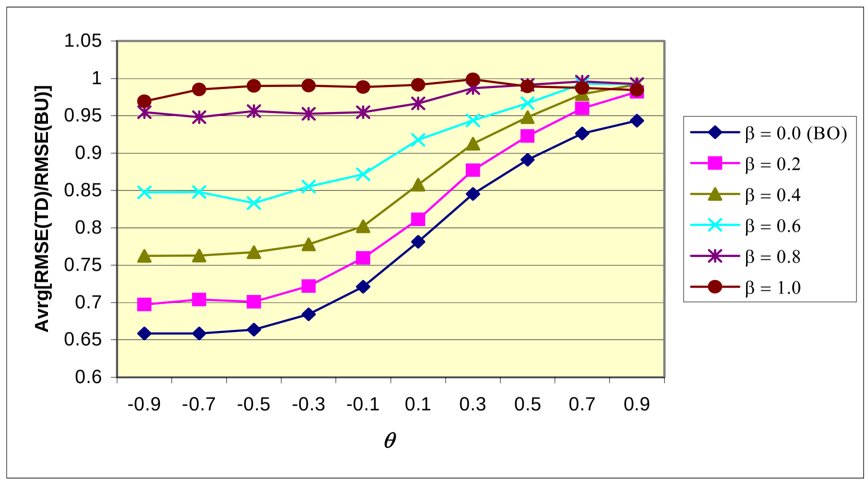
Figure 5. 5: Relative Performance of TD over BU Strategies for Forecasting Family Demand (Different τ)



a. MA(1) Demand Process



b. AR(1) Demand Process



c. IMA(1, 1) Demand Process

Figure 5. 6: Relative Performance of TD over BU Strategies for Forecasting Family Demand (Different Values of Coefficient of the Serial Correlation Term)

5.4 Conclusions

In this chapter, the performance of TD and BU strategies for forecasting the demand of a product that belongs to a family and the demand of a product family was evaluated. As is common in the retail merchandising industry, the products may be substituted with each other to a certain degree. The demand of the product family was assumed to follow a particular time series process, whereas the demand of each individual product was derived from the product family. Four demand time series were considered: white noise, first-order moving average [MA(1)], first-order autoregressive [AR(1)], and integrated moving average of order one [IMA(1, 1)]. Using a simulation approach, the impact of (i) the degree of product substitutability and (ii) the variability of demand quantity of each product in the family, on the relative performance of the two forecasting strategies were investigated.

Simple exponential smoothing, which is known for its reasonable accuracy, is used as the forecasting technique under both strategies and as the disaggregation method for the TD strategy. The forecast performance was measured by the root of mean squared error (RMSE) and the contribution achieved by each strategy was compared by a simple ratio.

The study revealed that the TD strategy, benefited by risk-pooling, consistently outperformed the BU strategy for forecasting the product family demand by as much as 52%. The superiority of the TD strategy over the BU strategy further improved as the variability of the product's proportion in the family increased and the degree of substitutability between the two products decreased. This phenomenon, however, was not always true when the performance of the two strategies for forecasting the product demands was compared. It was found that there were a few cases where the BU strategy outperformed the TD strategy by as much as 4%, particularly when the variability of the product's proportion was high and the degree of product substitutability was low.

CHAPTER 6

SUMMARY, IMPLICATIONS, AND FUTURE DIRECTIONS

This dissertation studied the relative effectiveness of top-down (TD) and bottom-up (BU) strategies for forecasting the demand of a product (belonging to a single family) as well as the demand of the product family under various time series models and forecasting scenarios meant to cover a broad spectrum of actual conditions.

The important findings from Chapters 3 to 5 are summarized in Table 6.1 and their implications to the practitioners are discussed in the next section. It is then followed by a discussion on some possible directions for future research.

6.1 Summary and Practical Implications

6.1.1 Forecasting Family-Level Demand

This study considers a two-level product structure, where the lower level comprises of N items, where $N \geq 2$. The demand for all individual items is assumed to follow three time series models, namely white noise, first-order moving average [MA(1)], first-order autoregressive [AR(1)] processes. As is common in a production planning environment, simple exponential smoothing and weighted moving average are used as the forecasting techniques under both strategies. The investigation is carried out analytically for the case when the coefficient of the serial correlation term of the demand time series for all individual products is identical. A simulation study is then performed to investigate the case when this coefficient is not identical.

Judging from the results of this study, one should consider the TD strategy for forecasting product family demand when the item demands follow a white noise or an

MA(1) processes. The actual cost of utilizing the BU strategy may be considerably higher than the TD strategy, especially when the number of products in the family is large. This is due to the fact that in the BU strategy, the demand forecast (and the optimization of the smoothing constants [α_i] or the historical period [T_i]) is carried out individually for each individual product before being aggregated to obtain the family demand forecast. Consequently, as the product variety proliferates, the processing time required to obtain the family demand forecast increases. Therefore, while the results seems to indicate that there is no significant difference in performance between the two strategies, one should take into account the cost involved when using the BU strategy for forecasting the demand of a product family which may comprise of hundreds (or even thousands) of individual products.

For the case of AR(1) item demand process, however, the decision should be taken more carefully, especially when the number of products in the family (N) is large. When $\phi_1 \neq \phi_2$, the BU strategy is apparently a better alternative than the TD strategy. This is because, in a majority of the cases, the RMSE value of the BU strategy is significantly lower than the TD strategy (see Figure 3.3).

6.1.2 Forecasting Item-Level Demand

Using the same set of assumptions and problem parameters, this study investigates the performance of TD and BU strategies for forecasting the demand of an individual product belonging to a product family.

The results of this study provide insights into the flexibility of adopting a forecasting strategy, particularly when there is a small number of products in the family and the demand of both products follows a white noise or an MA(1) process. In contrast to the case of forecasting the product family demand, the processing cost between TD and

BU strategies in this case was found to be the same (Schwarzkopf *et al*., 1988). Furthermore, the analytical study done in this research indicates that the difference in the variance of forecast error between the two strategies is relatively negligible. Hence, as far as the forecast error and the processing cost are concerned, no particular forecasting strategy is superior to the other. However, when the number of products in the family is large, more detailed analysis is required as the performance difference between the two forecasting strategies might be significant.

For the AR(1) item demand process, the finding is not straightforward. One needs to first compute the lag-1 autocorrelation (ϕ_i) value of the demand for each individual product. When its value is found to be highly positive, say $\phi_i \geq 0.4$, there is a very high probability that the BU strategy would produce more accurate estimates than the TD strategy (see Figures 4.2 and 4.3 and Table A.4).

Table 6. 1: Summary of the Research Findings

Forecast Level	Solution	Subaggregate Series	No. of Streams (N)	Independency btw Sources	Performance Measure	Forecasting Technique	Forecasting Strategy Preference	Other Insights
Family	Ana, Emp	WN	∞	Independent & Correlated	VFE, RMSE	SES	TD ≈ BU	• Holds true when the item demands are uncorrelated ($\gamma_{ij} = 0$)
						WMA	TD ≈ BU	• Holds true when the item demands are uncorrelated ($\gamma_{ij} = 0$)
		MA(1)	• ∞ (for $\theta_1 = \theta_2 = \ldots = \theta_N$) • 2 (for $\theta_1 \neq \theta_2$)	Independent & Correlated	VFE, RMSE	SES	TD ≈ BU	• Holds true when the item demands are uncorrelated ($\gamma_{ij} = 0$) • When $\theta_1 \neq \theta_2$, max. difference is ±6%
						WMA	TD ≈ BU	• Holds true when the item demands are uncorrelated ($\gamma_{ij} = 0$) • When $\theta_1 \neq \theta_2$, max. difference is ±6%
		AR(1)	• ∞ (for $\phi_1 = \phi_2 = \ldots = \phi_N$) • 2 (for $\phi_1 \neq \phi_2$)	Independent & Correlated	VFE, RMSE	SES	• TD ≈ BU (for $\phi_1 = \phi_2 = \ldots = \phi_N$) • Towards BU (for $\phi_1 \neq \phi_2$)	• TD ≈ BU still holds when the item demands are uncorrelated ($\gamma_{ij} = 0$)
						WMA	• TD ≈ BU (for $\phi_1 = \phi_2 = \ldots = \phi_N$) • Towards BU (for $\phi_1 \neq \phi_2$)	• The performance of TD and BU becomes more identical as the historical period (T) increases and the number of products in the family (N) decreases
Item	Ana, Emp	WN	∞	Independent & Correlated	VFE, RMSE	SES	TD ≈ BU	• When N = 2, the difference in performance ±1% • The performance of TD and BU becomes more identical as N decreases
						WMA	TD ≈ BU	• The performance of TD and BU becomes more identical as T increases and N decreases
		MA(1)	• ∞ (for $\theta_1 = \theta_2 = \ldots = \theta_N$) • 2 (for $\theta_1 \neq \theta_2$)	Independent & Correlated	VFE, RMSE	SES	TD ≈ BU	• When N = 2, the difference in performance ±1% • The performance of TD and BU becomes more identical as N decreases
						WMA	TD ≈ BU	• The performance of TD and BU becomes more identical as T increases and N decreases

		AR(1)	• ∞ (for $\phi_1 = \phi_2 = \ldots = \phi_N$) • 2 (for $\phi_1 \neq \phi_2$)	Independent & Correlated	VFE, RMSE	SES	Towards BU	• The superiority of the BU strategy increased with the lag-1 autocorrelation and with the decrease in correlation between the two item demands • The performance of TD and BU becomes more identical as *N* decreases
						WMA	Towards BU	• The performance of TD and BU becomes more identical as *T* increases and *N* decreases
Family & Item	Emp	WN, AR(1), MA(1), IMA(1, 1)	= 2	Not Specified	RMSE	SES	• Family: TD • Item: TD ≈ BU	• Studied the impact of product substitutability • The superiority of the TD strategy increased with the variability of the product's proportion in the family and with the decrease in the degree of product substitutability

<u>Abbreviations:</u> Ana=Analytically; AR(1)=Autoregressive of Order 1; Emp=Empirically; MA(1)=Moving Average of Order 1; MAPE=Mean absolute percent error; RMSE=Root of mean squared error; SES=Simple exponential smoothing (Brown, 1963); VFE=Variance of forecast error; WMA=Weighted Moving Average; WN=White Noise

6.1.3 Forecasting Substitutable Products

This study restricts the number of products in the family to two. The demand process for the product family is assumed to follow a particular time series, whereas the demand for each individual product is derived from the product family demand. Four time series (i.e. 3 stationary and 1 non-stationary) are used as the demand process of the product family. The degree of substitutability between products *i* and *j* is represented by β_{ij}, where $0 \leq \beta_{ij} \leq 1$.

The results from this study indicate the existence of an important relationship between product substitutability and the relative performance of TD over BU strategy. When the products within the family are substitutable, the TD strategy appears to perform better than the BU strategy for forecasting the product demands. This is primarily due to the inability of the BU strategy to separate the *true* demand from the total demand observed in a particular point of time. As a result, the BU strategy tends to under-forecast the demand of a product which is out of stock and to over-forecast the demand of a product whose on-hand inventory is in excess. The higher the degree of substitutability of a product (e.g. consumer products such as, shampoo, toothpaste, etc.), the higher the forecast error of the BU strategy would be in predicting the demand of an individual product (see Figures 5.1 to 5.3).

6.2 Directions for Future Research

This research represents an initial effort to investigate the relative performance of TD and BU strategies in a production planning environment. It leads practitioners a step closer to understanding the issues related to the selection of an appropriate forecasting strategy under various stochastic time series models and forecasting situations. Despite many insights gained by the current research, several important issues remain to be investigated. Some directions for future research are presented below:

1. Current study assumes that the demand for all the products in the family is homogeneous, meaning that the time series model of the demand for all the products is identical. While most of the studies in this issue make a similar assumption, this may render the results not generalizable under certain real-world conditions, where the demand process for one product variety may be different from the demand process for the other varieties. (This can be seen, for instance, in general household products such as detergent, which is normally introduced under several brands, each of which is dedicated to a specific market segment, purpose or type of material.) Hence, future study needs to consider different time series model for every product demand. The analytical investigation should be accompanied with an empirical study using real-data from companies in order to gain more complete insights.
2. For the case of substitutable products (Chapter 5), it was assumed that all excess demand is completely backordered or delayed. In reality, this is not always the case as part of the excess demand may be lost, depending on the type of products. Therefore, in order to develop a more comprehensive policy in choosing the best forecasting strategy, one needs to involve lost sales in the analysis. One difficulty in modeling this replenishment process, however, is the technique to measure the probability of lost sales, as this data may not be readily available. This is an opportunity for future research, not only in the field of forecasting, but also in the field of multi-echelon inventory systems.
3. With the exception of Chapter 5, all the problems presented in this dissertation used time series models which were stationary over time. However, since the demand of a product is often influenced by various other factors (e.g. season, promotion, weather, etc.), future work might need to take into account non-stationary demand processes or even stationary demand processes with trend and seasonality.

Beverages and fashion products are some of the typical examples of this phenomenon, as their demands are highly dependent on many factors, such as trend, seasons and locations.

BIBLIOGRAPHY

Abraham, B. and Ledolter, J. (1983). *Statistical Methods for Forecasting*, New York: John Wiley & Sons.

Achabal, D. D., McIntyre, S. H., Smith, S. A., Kalyanam, K. (2000). A Decision Support System for Vendor Managed Inventory, *Journal of Retailing*, 76(4), 430-454.

Aigner, D. J. and Goldfeld, S. M. (1973). Simulation and Aggregation: A Consideration, *The Review of Economics and Statistics*, 55(1), 114-118.

Aigner, D. J. and Goldfeld, S. M. (1974). Estimation and Prediction from Aggregate Data When Aggregates are Measured More Accurately than Their Components, *Econometrica*, 42(1), 113-134.

Anupindi, R., Dada, M. and Gupta, S. (1998). Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products, *Marketing Science*, 17(4), 406-423.

Armstrong, J. S. and Collopy, F. (1992). Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons, *International Journal of Forecasting*, 8(1), 69-80.

Armstrong, J. S. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Massachusetts: Kluwer Academic Publishers.

Badinelli, R. D. (1990). The Inventory Costs of Common Mis-specification of Demand-Forecasting Models, *International Journal of Production Research*, 28(12), 2321-2340.

Ballou, R. H. (1999). *Business Logistics Management*, 4th ed., New Jersey: Prentice-Hall.

Barnea, A. and Lakonishok, J. (1980). An Analysis of the Usefulness of Disaggregated Accounting Data for Forecasts of Corporate Performance, *Decision Sciences*, 11(1), 17-26.

Binkley, J. K. and Nelson, C. H. (1990). How Much Better is Disaggregate Data?, *Economics Letter*, 32, 137-140.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed., San Francisco: Holden-Day.

Brown, R. G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*, New Jersey: Prentice-Hall.

Chambers, J. C., Mullick, S. K. and Smith, D. D. (1971). How to Choose the Right Forecasting Technique, *Harvard Business Review*, 49(4), 45-74.

Chatfield, C. (1998). What is the Best Method of Forecasting?, *Journal of Applied Statistics*, 15, 19-38.

Chen, F., Drezner, Z., Ryan, J. K. and Simchi-Levi, D. (2000). Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information, *Management Science*, 46(3), 436-443.

Chen, Y., Kanetkar, V. and Weiss, D. L. (1994). Forecasting Market Shares with Disaggregate or Pooled Data: A Comparison of Attraction Models, *International Journal of Forecasting*, 10(2), 263-276.

Choi, S. and Kimes, S. E. (2002). Electronic Distribution Channels' Effect on Hotel Revenue Management, *Cornell Hotel and Restaurant Administration Quarterly*, 43(3), 23-31.

Chong, J-K, Ho, T-H. and Tang, C. S. (2002). Demand Modeling in Product Line Trimming: Substitutability and Variability, Working Paper, National University of Singapore.

Dangerfield, B. J. and Morris, J. S. (1988). An Empirical Evaluation of Top-Down and Bottom-Up Forecasting Strategies, *Proceedings of the 1988 Meeting of Western Decision Sciences Institute*, 322-324.

Dangerfield, B. J. and Morris, J. S. (1992). Top-Down or Bottom-Up: Aggregate versus Disaggregate Extrapolations, *International Journal of Forecasting*, 8(2), 233-241.

de Leeuw, S. L. J. M., van Donselaar, K. H. and de Kok, A. G. (1998). Forecasting Techniques in Logistics, in *Advances in Distribution Logistics*, van Nunen, J., Staehly, P., Fleischmann, B. Eds. Berlin: Springer.

Dekker, M., van Donselaar, K. H. and Ouwehand, P. (2004). How to Use Aggregation and Combines Forecasting to Improve Seasonal Demand Forecasts, *International Journal of Production Economics*, 90, 151-167.

DeLurgio, S. A. (1998). *Forecasting Principles and Applications*, 1st ed., Missouri: Irwin/McGraw-Hill.

DeLurgio, S. and Bhame, C. (1991). *Forecasting Systems for Operations Management*, Illinois: Business One Irwin.

Diebold, F. X. (1998). *Elements of Forecasting*, Ohio: International Thomson Publishing.

Dooley, F. and Kurtz, M. M. (2001). The Effect of a Changing Market Mix in Seed Corn on Inventory Costs, Working Paper, Purdue University.

Dunn, D. M., Williams, W. H. and Spivey, W. A. (1971). Analysis and Prediction of Telephone Demand in Local Geographic Areas, *Bell Journal of Economics and Management Science*, 2(2), 561-576.

Dunn, D. M., Williams, W. H. and DeChaine, T. L. (1976). Aggregate versus Subaggregate Models in Local Area Forecasting, *Journal of the American Statistical Association*, 71(353), 68-71.

Edwards, J. B. and Orcutt, G. H. (1969). Should Aggregation Prior to Estimation be the Rule?, *The Review of Economics and Statistics*, 51(4), 409-420.

Farnum, N. R. and Stanton, L. W. (1989). *Quantitative Forecasting Methods*, Boston: PWS-Kent Publishing Company.

Fisher, M. L., Hammond, J. H., Obermeyer, W. R. and Raman, A. (1994). Making Supply Meet Demand in an Uncertain World, *Harvard Business Review*, 72(3), 83-93.

Fisher, M. L. and Raman, A. (1996). Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales, *Operations Research*, 44(1), 87-99.

Fliedner, G. (1999). An Investigation of Aggregate Variable Time Series Forecast Strategies with Specific Subaggregate Time Series Statistical Correlation, *Computers & Operations Research*, 26(10-11), 1133-1149.

Fliedner, G. (2001). Hierarchical Forecasting: Issues and Use Guidelines, *Industrial Management & Data Systems*, 101(1), 5-12.

Fogarty, D. W., Hoffman, T. R. and Blackstone, Jr. J. H. (1991). *Production and Inventory Management*, Cincinnati: South-Western Publishing Co.

Food Marketing Institute. (1993). Variety or Duplication: A Process to Know Where You Stand. The Research Department, Food Marketing Institute, Washington, DC.

Gardner, E. S. J. (1990). Evaluating Forecast Performance in an Inventory Control System, *Management Science*, 36(4), 490-499.

Giesberts, P. M. J. (1993). Master Planning: Dealing with Flexibility. PhD Dissertation, Eindhoven University of Technology, The Netherlands.

Gordon, T. P., Morris, J. S. and Dangerfield, B. J. (1997). Top-Down or Bottom-Up: Which is the Best Approach to Forecasting?, *The Journal of Business Forecasting Methods & Systems*, 16(3), 13-16.

Granger, C. W. J. and Morris, M. J. (1976). Time Series Modelling and Interpretation, *Journal of the Royal Statistical Society: Series A*, 139(2), 246-257.

Graves, S. C. (1999). A Single-Item Inventory Model for a Nonstationary Demand Process, *Manufacturing & Service Operations Management*, 1(1), 50-61.

Gross, C. W. and Sohl, J. (1990). Disaggregation Methods to Expedite Product Line Forecasting, *Journal of Forecasting*, 9(3), 233-254.

Grunfeld, Y. and Griliches, Z. (1960). Is Aggregation Necessarily Bad?, *The Review of Economics and Statistics*, 42(1), 1-13.

Kahn, B. (1998a). Variety: From the Consumer's Perspective, in Ho, T. H. and Tang, S. C., *Product Variety Management Research Advances*, Dordrecht: Kluwer Academic Publishers, 19-37.

Kahn, K. B. (1998b). Revisiting Top-Down versus Bottom-Up Forecasting, *The Journal of Business Forecasting Methods & Systems*, 17(2), 14-19.

Klassen, R. D. and Flores, B. E. (2001). Forecasting Practices of Canadian firms: Survey results and comparisons, *International Journal of Production* Economics, 70, 163-174.

Kohn, R. (1982). When is an Aggregate of a Time Series Efficiently Forecast by Its Past?, *Journal of Econometrics*, 18(3), 337-349.

Kotler, P. and Armstrong, G. (2001). *Principles of Marketing*, 9th ed., New Jersey: Prentice Hall.

Lapide, L. (1998). New Developments in Business Forecasting, *The Journal of Business Forecasting Methods & Systems*, 17(2), 28-29.

Lee, A. O. (1990). Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process. PhD Dissertation, Massachusetts Institute of Technology.

Lee, H. L., So, K. C. and Tang, C. S. (2000). The Value of Information Sharing in a Two-Level Supply Chain, *Management Science*, 46(5), 626-643.

Lee, H. L. (2004). The Triple-A Supply Chain, *Harvard Business Review*, 102-112.

Lutkepohl, H. (1984). Forecasting Contemporaneously Aggregated Vector ARMA Processes, *Journal of Business & Economic Statistics*, 2(3), 201-214.

Lutkepohl, H. (1986). Comparison of Predictors for Temporally and Contemporaneously Aggregated Time Series, *International Journal of Forecasting*, 2(4), 461-475.

Makridakis, S., Anderson, S., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition, *Journal of Forecasting*, 1(2), 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. and Simmons, L. F. (1993). The M-2 Competition: A Real-Time Judgmentally Based Forecasting Study, *International Journal of Forecasting*, 9, 5-23.

Makridakis, S. and Hibon, M. (1979). Accuracy of Forecasting: An Empirical Investigation (With Discussion), *Journal of the Royal Statistical Society A*, 142, 97-145.

Makridakis, S. and Hibon, M. (2000). The M3-Competition: Results, Conclusions and Implications, *International Journal of Forecasting*, 16(4), 451-476.

McLeavey, D. W. and Narasimhan, S. L. (1985). *Production Planning and Inventory Control*, Boston: Allyn and Bacon, Inc.

Miller, J. G., Berry, W. and Lai, C. Y. F. (1976). A Comparison of Alternative Forecasting Strategies for Multi-Stage Production Inventory Systems, *Decision Sciences*, 7(4), 714-724.

Moon, M. A., Mentzer, J. T. and Thomas, Jr., D. E. (2000). Customer Demand Planning at Lucent Technologies: A Case Study in Continuous Improvement Through Sales Forecast Auditing, *Industrial Marketing Management*, 29(1), 19-26.

Muir, J. W. (1983). Problems in Sales Forecasting Needing Pragmatic Solutions, *APICS Conference Proceedings*, 4-7.

Muth, J. F. (1960). Optimal Properties of Exponentially Weighted Forecasts, *American Statistical Association Journal*, 55(290), 299-306.

Nahmias, S. (2001). *Production and Operations Analysis*, 4th ed., Singapore: McGraw-Hill Companies.

Orcutt, G., Watts, H. W. and Edwards, J. B. (1968). Data Aggregation and Information Loss, *American Economic Review*, 58, 773-787.

Pesaran, M. H., Pierse, R. G. and Kumar, M. S. (1989). Econometric Analysis of Aggregation in the Context of Linear Prediction Models, *Econometrica*, 57(4), 861-888.

Quelch, J. and Kenny, D. (1994). Extend Profits, not Product Lines, *Harvard Business Review*, 72(5), 153-160.

Rajaram, K. and Tang, C. S. (2001). The Impact of Product Substitution on Retail Merchandising, *European Journal of Operational Research*, 135, 582-601.

Rose, D. E. (1977). Forecasting Aggregates of Independent ARIMA Processes, *Journal of Econometrics*, 5(3), 323-345.

Sanders, N. R. and Manrodt, K. B. (1994). Forecasting Practices in US Corporations: Survey Results, *Interfaces*, 24, 92-100.

Schleifer, Jr., A. (1993). L. L. Bean Inc.: Item Forecasting and Inventory Management, *Harvard Business School Case (9-893-003)*.

Schwarzkopf, A. B., Tersine, R. J. and Morris, J. S. (1988). Top-Down versus Bottom-Up Forecasting Strategies, *International Journal of Production Research*, 26(11), 1833-1843.

Schwarzkopf, A. B. and Tersine, R. J. (1989). Robustness of Forecast Aggregation Procedure: Reality of Mirage, *Proceedings: Decision Sciences Institute Annual Conference*, 391-393.

Shlifer, E. and Wolff, R. W. (1979). Aggregation and Proration in Forecasting, *Management Science*, 25(6), 594-603.

Smith, S. A. and Agrawal, N. (2000). Management of Multi-Item Retail Inventory Systems with Demand Substitution, *Operations Research*, 48(1), 50-64.

Tan, B. and Karabati, S. (2000). An Order-Up-to-Level Updating Mechanism for Inventory Systems with Unobserved Lost Sales, Working Paper, Graduate School of Business, Koc University, Turkey.

Tersine, R. J. (1985). *Production/Operations Management: Concepts, Structure, and Analysis*, 2nd ed., Amsterdam: North-Holland.

Theil, H. (1954). *Linear Aggregation of Economic Relations*, Amsterdam: North-Holland Publishing Company.

Theil, H. (1966). *Applied Economic Forecasting*, Chicago: Rand McNally.

Tiao, G. C. and Guttman, I. (1980). Forecasting Contemporal Aggregates of Multiple Time Series, *Journal of Econometrics*, 12(2), 219-230.

van Donselaar, K. H. (2002). Demand Forecasting Based on Product-Aggregation, Internal Working Report, Eindhoven University of Technology, The Netherlands.

Weatherby, G. (1984). Aggregation, Disaggregation, and Combination of Forecasts. PhD Dissertation, Georgia Institute of Technology.

Weatherford, L. R., Kimes, S. E. and Scott, D. (2001). Forecasting for Hotel Revenue Management: Testing Aggregation against Disaggregation, *Cornell Hotel and Restaurant Administration Quarterly*, 42(4), 63-64.

Weatherford, L. R. and Kimes, S. E. (2003). A Comparison of Forecasting Methods for Hotel Revenue Management, *International Journal of Forecasting*, 19(3), 401-415.

Wei, W. W. S. (1993). *Time Series Analysis: Univariate and Multivariate Methods*, California: Addison-Wesley Publishing Company.

Wright, D. J. (1986). Forecasting Data Published at Irregular Time Intervals Using an Extension of Holt's Method, *Management Science*, 32(4), 499-510.

Yule, G. U. (1926). Why do We Sometimes Get Nonsense Correlations between Time Series? – A Study in Sampling and the Nature of Time Series, *Journal of the Royal Statistical Society*, 89(1), 1-63.

Yule, G. U. (1927). On a Method for Investigating Periodicities in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers, *Philosophical Transactions of the Royal Society of London: Series A*, 226, 267-298.

Zellner, A. (1969). On the Aggregation Problem: A New Approach to a Troublesome Problem, in K. A. Fox *et al*., eds., *Economic Models, Estimation and Risk Programming*, New York: Springer-Verlag, 365-374.

Zellner, A. and Tobias, J. (2000). A Note on Aggregation, Disaggregation and Forecasting Performance, *Journal of Forecasting*, 19(5), 457-469.

APPENDICES

Appendix to Chapter 2: *Summary of the time series models used in the study*

The time series models considered in this thesis are of stochastic type. According to Yule (1927), a time series whose successive values are correlated can always be represented as a linear combination of a sequence of uncorrelated random variables (also known as *linear filter*). The mathematical representation of the linear filter is given by (Abraham and Ledolter, 1983, p. 197):

$$d_{i,t} - \mu_i = \varepsilon_{i,t} + \psi_{i,1}\varepsilon_{i,(t-1)} + \psi_{i,2}\varepsilon_{i,(t-2)} + \ldots = \sum_{j=0}^{\infty} \psi_{i,j}\varepsilon_{i,(t-j)} \qquad (2A.1)$$

where $d_{i,t}$ is the observation (or demand) for item *i* at time *t*, μ_i is the expected value, $\psi_{i,j}$ is a constant ($\psi_{i,0} = 1$ and $\sum_{j=0}^{\infty} \left|\psi_{i,j}\right| < \infty$), and $\varepsilon_{i,t}$ is the random variable with zero mean $E(\varepsilon_{i,t}) = 0$, variance $Var(\varepsilon_{i,t}) = \sigma_i^2$, and zero autocovariance $Cov(\varepsilon_{i,t}, \varepsilon_{i,(t-k)}) = 0$ for all $k \neq 0$. (Unless mentioned otherwise, $\{\varepsilon_{i,t}\}$ is assumed to be normally distributed (Gaussian) with zero mean throughout the study.)

The equation (2A.1) is also known as a moving average representation of a process as it is constructed by both permanent and temporary components. Specifically, the previous error terms correspond to the permanent component of the series, while the current error term represents the temporary component (or noise) of the series. As will be discussed later, all the time series models considered in this thesis are essentially derived from (2A.1).

There are, in total, four time series models used in this research. Three of them [i.e. white noise, AR(1), MA(1)] are of *stationary* type, meaning that their mean, variance, and autocovariances are time invariant. The last one, IMA(1, 1), is of *non-stationary* type, meaning that its mean and variance are non-constant (time variant). The white noise process, which is the simplest of all time series generating mechanisms, is first reviewed below.

White Noise [ARIMA(0, 0, 0)] Time Series Model

Also known as an independently and identically distributed (i.i.d.) time series, a white noise process comprises of a sequence of uncorrelated random variables from a fixed distribution, usually normal, with constant mean $\left(E\left[d_{i,(t-k)}\right]=\mu_i\right)$ and constant variance $\left(Var\left[d_{i,(t-k)}\right]=\sigma_i^2\right)$ for all k, and zero autocovariances $\left(Cov(d_{i,t}, d_{i,(t-k)})=0\right)$ and zero autocorrelation $\left(\overline{\rho}_{ik}=0\right)$ for all $k \neq 0$. From the definition of a stochastic time series process in (2A.1), for the white noise time series model, $\psi_{i,j}=0$ for all $j \geq 1$. Therefore, (2A.1) can be rewritten as:

$$d_{i,t}=\mu_i+\varepsilon_{i,t} \tag{2A.2}$$

where $E(\varepsilon_{i,t})=0$ and $Var(\varepsilon_{i,t})=\sigma_i^2$. It is assumed that σ_i is significantly smaller than μ_i, so that the probability of a negative data value is negligible.

First-Order Moving Average [ARIMA(0, 0, 1)] Time Series Model

Also known as an MA(1) process, this time series model was first introduced by Yule in 1926, who addressed the issue of serial correlation within the time series. This time series is useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time (Wei, 1993, p. 46). From the definition of a stochastic time series process in (2A.1), for the MA(1) time series model, $\psi_{i,1}=-\theta_i$ and $\psi_{i,j}=0$ for all $j>1$. Hence, mathematically, the MA(1) time series is formulated as:

$$d_{i,t}=\mu_i+\varepsilon_{i,t}-\theta_i\varepsilon_{i,(t-1)} \tag{2A.3}$$

where θ_i is identified as the coefficient of the serial correlation term.

In order to obtain a converging auto-regressive representation (also known as *invertibility* condition[16]), the moving average parameter θ_i is restricted such that $|\theta_i| < 1$. It immediately follows from (2A.3) that (the proofs are straightforward and thus omitted):

$$Var(d_{i,(t-k)}) = (1+\theta_i^2)\sigma_i^2 \text{, for all } k \qquad \text{(2A.4)}$$

$$Cov(d_{i,t}, d_{i,(t-k)}) = \begin{cases} (1+\theta_i^2)\sigma_i^2 & k = 0 \\ -\theta_i\sigma_i^2 & |k| = 1 \\ 0 & |k| > 1 \end{cases} \qquad \text{(2A.5)}$$

$$\overline{\rho}_{ik} = \begin{cases} 1 & k = 0 \\ \dfrac{-\theta_i}{1+\theta_i^2} & |k| = 1 \\ 0 & |k| > 1 \end{cases} \qquad \text{(2A.6)}$$

From (2A.5), it is apparent that the observations one step apart are correlated, while those separated more than one step are uncorrelated. In addition, (2A.6) implies that the MA(1) process is *always* stationary and that $|\overline{\rho}_{ik}| < 0.5$, for $|k| = 1$.

<u>*First-Order Autoregressive [ARIMA(1, 0, 0)] Time Series Model*</u>

Also known as an AR(1) process, this time series was first introduced by Yule (1927) to describe the phenomena of sunspot numbers. It is one of many statistical processes whose error term $\varepsilon_{i,t}$ (t = 1, 2, …) is serially correlated, which in turn implies that the observations $d_{i,t}$ (t = 1, 2, …) are also serially correlated[17]. From the definition of a stochastic time series process in (2A.1), for the AR(1) time series model, $\psi_{i,j} = \phi_i^j$. Therefore, this leads to a

---

[16] Invertibility condition means that the MA(1) process can be transformed into an autoregressive representation $d_{i,t} - \mu_i = -\theta_i(d_{i,(t-1)} - \mu_i) - \theta_i^2(d_{i,(t-2)} - \mu_i) - ... + \varepsilon_{i,t}$. Apparently, $-\theta_i^j$ converges (i.e. the past observations decrease with their age) only if $|\theta_i| < 1$.

[17] According to Abraham and Ledolter (1983, p. 192), this serial correlation can be expected if the data are collected sequentially in time.

process in which the deviation from the mean at time t is regressed on itself, but with a lag of one time period. Mathematically, it is written as:

$$d_{i,t} - \mu_i = \varepsilon_{i,t} + \phi_i \varepsilon_{i,(t-1)} + \phi_i^2 \varepsilon_{i,(t-2)} + \phi_i^3 \varepsilon_{i,(t-3)} + \ldots = \sum_{j=0}^{\infty} \phi_i^j \varepsilon_{i,(t-j)} \qquad (2A.7)$$

By taking ϕ_i out from the right-hand side of (2A.7) and simplifying the resulting expression:

$$d_{i,t} = (1-\phi_i)\mu_i + \phi_i d_{i,(t-1)} + \varepsilon_{i,t} \qquad (2A.8)$$

where $|\phi_i| < 1$ and is constant at all times. The restriction of $|\phi_i| < 1$ is necessary as it ensures the stationarity condition of the time series (see Abraham and Ledolter, 1983, p. 200, for more detail). The following properties can be obtained, provided that $|\phi_i| < 1$ (the proofs are straightforward and thus omitted):

$$Var(d_{i,(t-k)}) = \frac{\sigma_i^2}{1-\phi_i^2}, \text{ for all } k \qquad (2A.9)$$

$$Cov(d_{i,t}, d_{i,(t-k)}) = \frac{\phi_i^k \sigma_i^2}{1-\phi_i^2}, \text{ for all } k \geq 0 \qquad (2A.10)$$

$$\bar{\rho}_{ik} = \frac{Cov(d_{i,t}, d_{i,(t-k)})}{\sqrt{Var(d_{i,t}) \times Var(d_{i,(t-k)})}} = \phi_i^k, \text{ for all } k \geq 0 \qquad (2A.11)$$

Due to the direct relationship between components ϕ_i and $\bar{\rho}_{i1}$ as shown in (2A.11), in the rest of this thesis component ϕ_i is identified as a lag-1 autocorrelation.

Integrated Moving Average of Order One [ARIMA(0, 1, 1)] Time Series Model

Also known as an IMA(1, 1) process, the mean and variance of this time series model vary over time. Muth (1960) has shown that the simple exponential smoothing is optimal under this time series model. From the definition of a stochastic time series process in (2A.1), for the IMA(1, 1) time series model, $\psi_{i,j} = 1-\theta_i$ for all j. Hence, mathematically, the IMA(1, 1) time series is expressed as:

$$d_{i,t} = d_{i,(t-1)} + \varepsilon_{i,t} - \theta_i \varepsilon_{i,(t-1)} \tag{2A.12}$$

where $d_{i,0} = \mu_i + \varepsilon_{i,0}$. Note that in the IMA(1, 1) process, the parameter μ_i plays a different role. This is because for a very large t, this term can become very dominating so that it forces the series to follow a deterministic pattern. Hence, in practice, unless it is clear from the historical data that a deterministic trend term exists, μ_i is normally set to zero for $t > 0$, as shown in (2A.12). It immediately follows from (2A.12) that (the proofs are straightforward and thus omitted):

$$Var(d_{i,(t-k)}) = \left[1 + (t-k-n_0-1)(1-\theta_i^2)\right]\sigma_i^2 \tag{2A.13}$$

$$Cov(d_{i,t}, d_{i,(t-k)}) = \left[(1-\theta_i) + (t-k-n_0-1)(1-\theta_i)^2\right]\sigma_i^2 \text{, for all } k > 0 \tag{2A.14}$$

$$\overline{\rho}_{ik} = \frac{(1-\theta_i) + (t-k-n_0-1)(1-\theta_i)^2}{\sqrt{\left[1 + (t-k-n_0-1)(1-\theta_i)^2\right]\left[1 + (t-n_0-1)(1-\theta_i)^2\right]}} \tag{2A.15}$$

where n_0 ($t > n_0$) is the time origin from which the observation starts. From (2A.13) to (2A.15), it is obvious that $Var(d_{i,(t-k)})$, $Cov(d_{i,t}, d_{i,(t-k)})$, and $\overline{\rho}_{ik}$ become unbounded as $t \to \infty$.

Appendix to Chapters 3 and 4: *Results of the numerical study when both the TD and BU strategies use weighted moving average (WMA) technique*

The numerical study is done using Mathematica 5.1. In Table A.1, $T = 1$ column represents the case when $0.43 < \hat{\phi} < 1$, whereas the remaining columns represents the case when $-1 < \hat{\phi} \le 0.43$ (see Table 3.1 for more details).

Table A. 1: Upper and Lower Bounds of V_{TD}/V_{BU} under AR(1) Demand Process in Forecasting Family Demand

N	T = 1			T = 2			T = 4			T = 6		
	LB	UB	UB-LB	LB	UB	UB-LB	LB	UB	UB-LB	LB	UB	UB-LB
2	1	1	0	0.88	1.05	0.17	0.94	1.05	0.11	0.97	1.04	0.07
3	1	1	0	0.86	1.07	0.21	0.92	1.07	0.15	0.95	1.06	0.11
4	1	1	0	0.84	1.09	0.25	0.91	1.09	0.18	0.93	1.07	0.15
5	1	1	0	0.82	1.11	0.29	0.89	1.11	0.21	0.91	1.09	0.18

All the results presented in the following Tables are obtained by using expression in Table 4.5. For Tables A.2 and A.3, since $T_i^* = T^* = \max\{T\}$, where $T > 1$ and integer, the analysis is conducted under different value of T = 2, 4 and 6.

Table A. 2: Upper and Lower Bounds of V_{TD}/V_{BU} under White Noise Demand Process in Forecasting Item Demand

N	T = 2			T = 4			T = 6		
	LB	UB	UB-LB	LB	UB	UB-LB	LB	UB	UB-LB
2	0.67	1.33	0.67	0.80	1.20	0.40	0.86	1.14	0.29
3	0.54	1.67	1.12	0.73	1.40	0.67	0.80	1.29	0.48
4	0.42	2.00	1.58	0.65	1.60	0.95	0.75	1.43	0.68
5	0.30	2.33	2.03	0.58	1.80	1.22	0.70	1.57	0.87

Table A. 3: Upper and Lower Bounds of V_{TD}/V_{BU} under MA(1) Demand Process in Forecasting Family Demand

N	T = 2			T = 4			T = 6		
	LB	UB	UB-LB	LB	UB	UB-LB	LB	UB	UB-LB
2	0.67	1.31	0.64	0.80	1.07	0.27	0.86	1.14	0.28
3	0.59	1.74	1.15	0.74	1.23	0.49	0.82	1.22	0.40
4	0.51	2.17	1.66	0.69	1.39	0.71	0.79	1.30	0.51
5	0.42	2.60	2.17	0.63	1.55	0.93	0.76	1.38	0.62

Finally, in Tables A.4 and A.5, the results are based on the following criteria:

$$T_i^* = \begin{cases} 1 & 0.43 < \phi_i < 1 \\ \max\{T_i\} & -1 < \phi_i \le 0.43 \end{cases}.$$

Table A. 4: Upper and Lower Bounds of V_{TD}/V_{BU} under AR(1) Demand Process in Forecasting Item Demand ($0.43 < \phi_N < 1$)

N	LB	UB	UB - LB
2	0.76	100	99.24
3	0.24	150	149.76
4	0.14	200	199.86
5	0.08	250	249.92

Table A. 5: Upper and Lower Bounds of V_{TD}/V_{BU} under AR(1) Demand Process in Forecasting Item Demand ($-1 < \phi_N \leq 0.43$)

N	*T* = 2			*T* = 4			*T* = 6		
	LB	UB	UB-LB	LB	UB	UB-LB	LB	UB	UB-LB
2	0.63	1.74	1.11	0.78	1.47	0.69	0.84	1.35	0.51
3	0.33	2.40	2.07	0.58	1.90	1.32	0.67	1.66	1.00
4	0.03	3.07	3.03	0.38	2.33	1.95	0.50	1.98	1.49
5	-0.27	3.73	3.99	0.18	2.76	2.58	0.33	2.30	1.98

Appendix to Chapter 3: *Algorithm to generate two sets of random variables that follow a particular time series model and are statistically correlated with each other*

The algorithm begins by creating two sets of random variables ($\varepsilon_{1,t}$ and $\varepsilon_{2,t}$) which are normally distributed with zero means, variances of σ_i^2 (for i = 1, 2), zero autocorrelations, and intercorrelated with each other with correlation coefficient of ρ_{12}, $|\rho_{12}| \leq 1$. This is done by firstly generating two sets of independent random data ($X_{1,t}$ and Y_t) with the following distribution: $N \sim (0,1)$. Then, with a predetermined value of ρ_{12}, the new data set of $X_{2,t}$ is computed by:

$$X_{2,t} = \rho_{12} X_{1,t} + Y_t \sqrt{1-\rho_{12}^2} \tag{3A.1}$$

Subsequently, the two random variables ($\varepsilon_{1,t}$ and $\varepsilon_{2,t}$) can be obtained by adjusting the variances of $X_{1,t}$ and $X_{2,t}$ linearly:

$$\varepsilon_{i,t} = X_{i,t} \times \sigma_i \tag{3A.2}$$

Finally, the actual item demands ($d_{1,t}$ and $d_{2,t}$) are then generated by substituting (3A.2) into the equations for the time series models shown in (2A.2), (2A.8), and (2A.3).

Appendix to Chapter 5: *Algorithm for substitutable products*

Define the following notation:

β : Degree of product substitutability, $0 \leq \beta \leq 1$

$D_i(t)$: *Actual* (real) demand for product *i* in period *t* (where *i* = 1, 2), in units

$\tilde{D}_i(t)$: *Observed* demand for product *i* in period *t*, in units

$I_i(t)$: On-hand inventory for product *i* in period *t* (after satisfying the direct demand for product *i*), in units

$I_i^B(t)$: Beginning inventory for product *i* in period *t* (before adding with the order quantity shipped *L* periods ago), in units (equivalent to $I_i^E(t-1)$)

$I_i^E(t)$: Ending inventory for product *i* in period *t*, in units (equivalent to $I_i^B(t+1)$)

$\tilde{I}_i(t)$: Available inventory for product *i* in period *t* (before satisfying the direct demand for product *i*), in units

$q_{i\to j}$: Excess demand for product *i* which is satisfied by product *j*, in units

$SR_i(t)$: Scheduled receipt for product *i* in period *t*, in units

Note that negative inventory level represents backorders. The explanation of the inventory logic used in the simulation begins with the case when $\beta = 0$.

Case 1: The Individual Products are Not Substitutable (i.e. $\beta = 0$)

The inventory logic for this case is relatively straightforward as no excess demand for a particular product is passed to and satisfied by the other product. At time *t*, the available inventory $\left[\tilde{I}_i(t)\right]$ is first computed by summing the incoming supply, which was delivered *L* periods ago $\left[SR_i(t)\right]$, and the ending inventory from the previous period $\left[I_i^E(t-1)\right]$. This available inventory is then used to satisfy any (real) demand requested by a consumer

$[D_i(t)]$. Whenever the quantity of the demand is larger than the available inventory $\left[D_i(t) > \tilde{I}_i(t)\right]$, the remaining demand is completely backlogged and no part of this excess demand will be substituted by the other product.

Case 2: The Individual Products are Substitutable (i.e. 0 < $\beta \le 1$)

The pseudocode of the inventory logic with two substitutable products is given below:

Step 1: Update the parameters for product 1.

$\tilde{I}_1(t) = I_1^B(t) + SR_1(t)$	*Update the available inventory*
$I_1(t) = \tilde{I}_1(t) - D_1(t)$	*Update the on-hand inventory*
$\tilde{D}_1(t) = D_1(t)$	*Copy the value of the actual demand to the observed demand*
If $(I_1(t) > 0)$	*If the on-hand inventory is positive*
{	
Set $q_{1\to2} = 0$	*No excess demand is passed to product 2*
$I_1^E(t) = I_1(t)$	*Update the ending inventory*
}	
Else	*If the on-hand inventory is negative*
{	
If $(\tilde{I}_1(t) < 0)$	*If the available inventory is negative*
$q_{1\to2} = \lceil \beta \times D_1(t) \rceil$	*Compute the demand quantity to be passed to product 2*
Else	*If the available inventory is positive*
$q_{1\to2} = \lceil \beta \times I_1(t) \times (-1) \rceil$	*Compute the demand quantity to be passed to product 2*
$\tilde{D}_1(t) = \tilde{D}_1(t) - q_{1\to2}$	*Update the observed demand*
$I_1^E(t) = I_1(t) + q_{1\to2}$	*Update the ending inventory*
}	

Step 2: Update the parameters for product 2.

$\tilde{I}_2(t) = I_2^B(t) + SR_2(t)$	*Update the available inventory*
$I_2(t) = \tilde{I}_2(t) - D_2(t)$	*Update the on-hand inventory*
$\tilde{D}_2(t) = D_2(t)$	*Equalize the actual and the observed demand*
If $(I_2(t) > 0)$	*If the on-hand inventory is positive*
{	
Set $q_{2\to1} = 0$	*No excess demand is passed to product 1*
If $(q_{1\to2} > 0)$	*If there is an excess demand from product 1*
{	

$I_2^E(t) = I_2(t) - q_{1\to2}$	*Update the ending inventory*
$\tilde{D}_2(t) = \tilde{D}_2(t) + q_{1\to2}$	*Update the observed demand*
Return $q_{1\to2} = 0$	
}	
Else	
$I_2^E(t) = I_2(t)$	*Update the ending inventory*
}	
Else	*If the on-hand inventory is negative*
{	
If $\left(\tilde{I}_2(t) < 0\right)$	*If the available inventory is negative*
$q_{2\to1} = \lceil \beta \times D_2(t) \rceil$	*Compute the demand quantity to be passed to product 1*
Else	*If the available inventory is positive*
$q_{2\to1} = \lceil \beta \times I_2(t) \times (-1) \rceil$	*Compute the demand quantity to be passed to product 1*
$\tilde{D}_2(t) = \tilde{D}_2(t) - q_{2\to1}$	*Update the observed demand*
$I_2^E(t) = I_2(t) + q_{2\to1}$	*Update the ending inventory*
If $\left(q_{1\to2} > 0\right)$	*If there is an excess demand from product 1*
{	
$\tilde{D}_2(t) = \tilde{D}_2(t) + q_{2\to1}$	*Update the observed demand*
$I_2^E(t) = I_2^E(t) - q_{2\to1}$	*Update the ending inventory*
Return $q_{2\to1} = 0$	
$\tilde{D}_1(t) = \tilde{D}_1(t) + q_{1\to2}$	*Update the observed demand for product 1*
$I_1^E(t) = I_1^E(t) - q_{1\to2}$	*Update the ending inventory for product 1*
Return $q_{1\to2} = 0$	
}	
Else	
{	
$\tilde{D}_1(t) = \tilde{D}_1(t) + q_{2\to1}$	*Update the observed demand for product 1*
$I_1^E(t) = I_1(t) - q_{2\to1}$	*Update the on-hand inventory for product 1*
Return $q_{2\to1} = 0$	
}	
}	

The above inventory logic can be explained as follows. Suppose that at time t, the sum of the beginning inventory for a particular product, say product 1 $\left[I_1^B(t)\right]$, and the scheduled receipt $\left[SR_1(t)\right]$, is not sufficient to fully satisfy the (real) demand from a consumer $\left[D_1(t)\right]$. A portion of this unsatisfied (real) demand $\left(\text{i.e. } q_{1\to2} = \lceil \beta \times \min\{D_1(t), I_1(t) \times (-1)\} \rceil\right)$ (where β is assumed to be known and deterministic, and $\lceil x \rceil$ is the smallest integer $\geq x$) would then be fulfilled from the on-hand inventory of the alternative product 2 $[I_2(t)]$, provided that $I_2(t) > 0$. In this case, if it is found that $q_{1\to 2} > I_2(t)$, the remaining unsatisfied demand $[\text{i.e. } q_{1\to 2} - I_2(t)]$ would be backordered to product 2. This mechanism is adopted in order to avoid duplication of demands and also from the fact that it may be impractical for the consumer to have orders for both products 1 and 2 if the two products are qualitatively comparable. The computation of the ending inventory and the new observed demand for the alternative product 2 are shown in code lines 26 and 27, respectively.

However, if the on-hand inventory of the alternative product 2 is also not enough to even satisfy its own demand $(I_2(t) \leq 0)$, then the excess demand for each individual product would not be substituted with the other product. In other words, substitution to the other product takes place only if at least part of the substituted demand can be satisfied in the current period. The computation of the ending inventory and the new observed demand for the two products under this situation are shown in code lines 43 to 48.