OPTIMIZATION-BASED DECISION SUPPORT SYSTEM FOR RETAIL SOURCING

by

Jalpa Patel

B.S. Electrical Engineering, Northeastern University, 2005

Submitted to the MIT Sloan School of Management and the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration

and

Master of Science in Engineering Systems

In conjunction with the Leaders for Global Operations Program at the Massachusetts Institute of Technology

June 2012

Signature of Author ....

MIT Sloan School of Management, Engineering Systems Division

May 18, 2012

Certified by ....

Stephen Graves, Thesis Advisor

Professor of Management Science and Engineering Systems, MIT Sloan School of Management

Certified by ....

David Simchi-Levi, Thesis Advisor

Co-director, Leaders for Global Operations

Professor of Civil Engineering and Environmental Engineering and Engineering Systems

Accepted by ....

Olivier L. de Weck, Chair, Engineering Systems Education Committee

Associate Professor of Aeronautics and Astronautics and Engineering Systems

Accepted by ....

Maura Herson, Director, MBA Program

MIT Sloan School of Management

This page intentionally left blank.

OPTIMIZATION-BASED DECISION SUPPORT SYSTEM FOR RETAIL SOURCING

by

Jalpa Patel

Submitted to the MIT Sloan School of Management and the Engineering Systems Division on May 18, 2012 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Engineering Systems

Abstract

Some of the biggest challenges in the retail sourcing lie in predicting demand for a new article and making purchase decisions such as quantity, source, transportation mode and time of the order. Such decisions become more complex and time consuming as the number of SKUs and suppliers increase. The thesis addresses the issue of managing retail sourcing using forecasting and optimization based decision system developed for Zara, a leading fast-fashion clothing retailer.

We started with an existing pre-season demand forecasting method that uses POS data from a comparable older article to forecast demand for a new article after adjusting for stock-outs and seasonality. We developed and compared various forecast updating methods for accuracy and found that an exponential smoothing-based model, modified to accommodate for changes in level few steps ahead, resulted in highest accuracy using Cumulative Absolute Percentage Error (CAPE).

Next, we implemented a profit-maximizing optimization model to produce explicit sourcing decisions such as quantity, time and source of orders. The model takes in distributional forecasts, supply constraints, holding cost, pricing information and outputs explicit sourcing decisions mentioned above. A prototype for forecasting and optimization code is ready and currently being evaluated to secure approval for a live pilot for Summer 2013 campaign sourcing.

Thesis Supervisor: Stephen Graves
Title: Professor of Management Science and Engineering Systems, MIT Sloan School of Management

Thesis Supervisor: David Simchi-Levi
Title: Co-director, Leaders for Global Operations
Professor of Civil Engineering and Environmental Engineering and Engineering Systems

This page intentionally left blank.

Acknowledgments

I would like to thank Zara and the Leaders for Global Operations (LGO) program for the opportunity to work on this project. Working in Spain and for Zara was a unique and rewarding experience, one that I will always remember.

I am deeply grateful for the hard work and contributions from many key individuals involved this project. I would first like to thank Professor Jérémie Gallien (London Business School) and Professor Adam Mersereau (University of North Carolina), for their advice on the technical and managerial aspects of the project. I would also like to thank Professor Stephen Graves and Professor David Simchi-Levi for their help and advice on the thesis.

Within Zara, I would like to thank Miguel Diaz and Francisco Babio Fernandez for the opportunity to work on a very crucial project for Zara. I appreciate their support, feedback and business advice. I would like to thank the IT department at Zara, in particular Rubén Melcón Fariña, Juan Jose Lema Cundins and Alberte Dapena Mora. Their hard work and collaboration have been crucial in implementing the project. I would also like to thank Diego Teijeiro Ruiz, Olaia Vazquez Sanchez and David Lorenzo González for their help. Lastly, I would like to thank the many buyers at Zara who helped me understand the sourcing process at Zara.

This page intentionally left blank.

Table of Contents

List of Figures

List of Tables

1 Introduction

This section aims to briefly introduce the thesis and related project, the problem statement, the key thesis ideas as well as the thesis structure.

1.1 Thesis Background – Zara1 Project

The thesis discusses the key topics studied during the summer-fall 2011 project at Zara as part of Leaders for Global Operations (LGO) program. Past LGO projects at Zara were targeted at the downstream operations such as distribution optimization and sales pricing. The summer-fall 2011 project was a direct follow-on project of another LGO project. Together the two projects, referred to as the Zara1 within the company, targeted the upstream demand forecasting and sourcing of new clothing articles[1].

1.2 Problem Statement

Each year Zara sells thousands of clothing designs across its three departments: man, woman and children. Many designs are available in more than one color and each colored article may be available in one or more sizes. Between the Summer 2011 and Winter 2011 selling seasons, known as campaigns, Zara sold about 11000 different clothing designs. These 11000 designs further translated into about 32000 design-color combinations and about 58000 design-color-size combinations.

Each of the three departments has a team of five to ten buyers. These fashion savvy employees are responsible to procure hundreds or thousands of articles not just for the pre-season sales but also for subsequent replenishments. They need to determine the quantity of each clothing article, their source, transportation method as well as time of order. Buyers need to make the difficult tradeoffs between cost and lead time for thousands of articles, which is very time consuming. Thus, there is a need for a scalable system to assist buyers in making such complex purchasing decisions. Also, there is a

[1] Throughout the document, we will use the word article to refer to a clothing design.

significant opportunity for the project to improve the company's profits. In 2011, Inditex's Net Sales were equal to €13.79B and total Cost of Sales was equal to €5.61B, resulting in Gross Profit of €8.2B. Zara accounted for 65% percent of the group net sales with its own net sales of €8.9B [1]. At such an enormous scale, even a one percent reduction in sourcing costs can increase the gross profits by millions of euros.

The goal of Zara1 project was to develop an analytical decision support system to provide more consistent and explicit sourcing decisions. The approach for solving this problem and results will be discussed in the subsequent chapters.

1.3 Thesis Structure

Chapter 1 provides project introduction and the problem statement. Chapter 2 provides a more detailed background on the company and the buying process. It also outlines the approach we used to develop a decision support system for sourcing at Zara. Chapter 3 contains a literature review on point forecasts and distributional forecasts, forecast error measurement metrics and inventory optimization. Chapter 4 discusses the concept and formulation of the optimization model used to generate detailed sourcing decisions. Chapter 5 discusses the method we used for updating point forecasts. Chapter 6 shows our method to derive distributional forecasts and learning curve information using the point forecast and empirical forecast errors. Chapter 7 contains sensitivity analysis for the optimization model. Chapter 8 provides our recommendations for future work and thesis conclusion.

2 Project Background

Section 2.1 provides a brief company background including the group history and the key features of Zara's business model. Section 2.2 provides details on the sourcing process at Zara relevant for the project.

2.1 Company Background

The Inditex Group is the largest fast fashion chain, by revenue, in the world with revenues of €12.5B in 2010 and €13.79B in 2011. The group consists of eight different store formats: Zara, Pull & Bear, Massimo Dutti, Bershka, Stradivarius, Oysho, Zara Home and Uterqüe. As of March 2012, the group owned 5402 stores in 78 different countries [1], [2], [3].

2.1.1 History

Armancio Ortega Gaona, founder and chairman of the Inditex Group, began textile manufacturing operations in 1963 and by 1974 he owned several factories. In 1975, he opened the first Zara store in downtown A Coruña, a remote coastal city in the northwest part of Spain. Thus, Zara is the first retail store format for the parent company and till this day remains the flagship brand at Inditex [4]. Over the next decade, several Zara stores were launched in major cities in Spain. In 1985, Inditex was "founded as the holding company of the group of businesses operating at the time" [4].

Since 1985, the Inditex group grew both organically, by expanding in European and international markets and launching brands such as Pull & Bear, Bershka, Oysho, Zara home and Uterqüe, organically as well inorganically, by acquiring brands such as Massimo Dutti and Stradivarius [4]. Of all brands, Zara accounts for roughly 65% of the group's net sales [1].

2.1.2 Business Model

Zara's fast-fashion retailing model aims at bringing the latest fashion trends from the runway to the stores fast. It can bring a product from design to store in as little as 15 days. Zara stores create "a sense of tantalizing exclusivity" by displaying only a few units per design [5]. This creates a sense of urgency amongst the customers, who are tempted to buy the article before it runs out. Zara also displays thousands of designs each year. Articles that are not selling well after few weeks of introduction are taken off the display. Large number of new designs, lower stock per design and limited time to buy motivate the customer to visit the store more frequently compared to other stores. Zara relies solely on its products, store location, store appearance and word of mouth advertising to promote sales and spends only about 0.3% of its sales on advertising, compared to 3-4% that the competitors spend [5].

The above model requires all the aspects of the business from design to store delivery to respond fast to the changing trends and customer tastes. For Zara, the challenge is further heightened as it operates in multiple markets with different customer preferences. Zara has a successful formula to tackle these challenges and we would like to briefly mention some of these well-documented and important factors involved in its success.

Zara has developed a "superresponsive" supply chain by maintaining excellent communication through all layers of the organization, by adopting and adhering to a strict routine for information and product flows and by investing in a flexible supply chain [5].

The physical and organizational setup at Zara drives the closed loop communication. For its three main product sections – Man, Woman and Children, Zara has different departments, all located at the headquarters in Coruña. Each department has its own designers, market representatives and buyers. The designers are mainly responsible for creating new garment designs. The market representatives specialize in certain regions of Zara's sales network and collect the fashion trend information from their regions. The buyers are sourcing experts and work with suppliers to order the garments. They make decisions regarding total quantity, cost, shipping method and time of order for finished garments. While the three departments are separated with walls, each department has open sitting arrangement, which allows for fast and direct communication within the department. There is constant communication amongst the different organizational layers: between stores and the market representatives at the headquarters (via PDAs and telephone conversations), between the market representatives and the buyers, between the buyers and the designers. Such fast, regular and direct communications channels allow Zara to stay in touch with both the product and the customer, which allows it to react fast to market needs [5].

Zara's processes are designed to allow for timely information and product flows. For example, the stores place orders to the headquarters bi-weekly and also receive shipments from one of the two distribution centers (DC) at the same frequency. Such predictability of shipments aids capacity planning at the DCs, shipment scheduling with the carriers and product display planning at the stores. The bi-weekly order fulfillment schedule allows Zara to quickly react to demand variations and fluctuations in the different markets [5].

Zara's focus on maintaining fast response supply chain, allows it to make supply chain decisions that are sometimes contrary to general practice. Although Zara operates in five continents, it has only two DCs both located in Spain, one next to the group headquarters in Arteixo and one in Zaragoza. This arrangement results in higher transportation costs, especially because of the bi-weekly shipment model, but also allows greater flexibility in responding to different market demands. Using pre-scheduled trucks and airfreight deliveries, inventory is shipped from the DCs to European stores in 24 hours, to US stores in 48 hours and Japanese stores in 72 hours. Also, excess capacity at the warehouses allows fast-response while facing high variability in demand [5].

2.2 Sourcing at Zara

Each of the three departments at Zara has its own buying team, consisting of five to ten buyers. They determine sourcing details such as the quantity, location and time for order as well as shipping method via which to bring the items into one of the two main DCs. Once the order has arrived at the DC, it is then distributed to stores around the world. The quantity arriving at the DC determines how much can be distributed to the stores, so it is crucial for Zara to optimally manage the quantity and timing of these orders.

2.2.1 Buying Timeline

Zara divides its annual buying budget into two main campaigns – Summer and Winter. Regular summer campaign runs from late January until early June, after which the remaining articles are sold at a markdown price during the clearance sales. Similarly, regular winter campaign runs from July to December, after which the remaining winter articles are sold at a markdown price during the clearance sales.

Figure 1. Buying Timeline for Summer 2012 Campaign

Figure 1 shows the buying timeline for Summer 2012 campaign. Zara buys from three main supply regions, known as circuits. Long circuit suppliers are located in Asia. Zara uses two different transportation modes – ocean and air, to ship from long circuit suppliers. Medium circuit suppliers are typically located in Europe or Middle East, while short circuit suppliers are factories in Portugal or Zara's own factories next to the headquarters in Arteixo. Articles from the medium and short circuit suppliers are typically shipped via ground transportation. Supply lead time, including production and transportation time, varies from 90 days for long circuit to 15 days for the local factories.

With a few exceptions, Zara uses a multi-vendor strategy, allowing it to procure the same article from multiple vendors often located in different circuits. These multiple sourcing options pose a trade-off between lead time and cost. In order to take advantage of the low cost Asian suppliers, Zara would need to start placing orders at least three to four months in advance of the selling period. On the other hand, they can place an order at the local factory just two to three weeks before the selling season, but at a higher cost. Shorter lead times translate to more accurate demand information and reduced inventory costs, but higher purchase costs. Longer lead times translate to reduced purchase cost, but higher inventory costs tied up as working capital.

Due to sufficient planning time for pre-season orders, long circuit suppliers are often a better choice than short circuit suppliers for such orders. However, the pre-season orders are based on new article forecast, made using point of sales (POS) data for similar article, known as a 'comparable'. A comparable is an article sold in previous campaign and used as a sales reference for the new article based on certain attributes such as department (Man vs. Woman), style, fabric, etc. The sales of the new article may differ significantly from its pre-season forecast due to changes in consumer tastes triggered by changing trends, changes in macro-economic factors such a recession, etc. As a result the first week of a new article introduction is often an important predictor of large changes in article demand from the pre-season forecast. Buyers often determine whether to replenish an article based on the first week sales. If the new article demand is high compared to the forecast, the buyers need to replenish the article fast in order to ensure article availability during regular season, making medium or short circuit suppliers a better option compared to the long circuit suppliers. We aimed to design an optimization-based system to help optimize such trade-offs.

2.3 Solution Approach

The goal of the Zara1 project was to develop a forecasting and optimization based support system to aid Zara's sourcing decisions. The project was divided into two sub-projects. The first sub-project Zara1a aimed at developing an article-level pre-season demand forecast, which is used to place the pre-season orders well in advance of the selling period. Details of Zara1a project can be found in

Tatiana M. Bonnefoi's thesis titled 'Demand Forecast for Short Life Cycle Products: Zara Case Study' [6]. The key outputs from the first project used for the second project, Zara1b, were:

- Method to derive pre-season demand forecast using point-of-sale and stock-out data
- Method to calculate seasonality factors using the sales data

In Zara1b, we aimed to implement an optimization model to produce explicit purchasing decisions for clothing articles. The linear optimization model, developed by Professor Jérémie Gallien and Professor Adam Mersereau, takes as input the distributional forecasts and learning curve information along with supply and pricing information for the article, and outputs optimum quantity, source, time and transportation mode for the order. We developed a method to create distributional forecast from a given point forecast and historical forecasting errors and also a method to estimate a learning curve, which shows expected improvement in forecasts for various update periods. When running the optimization model prior to selling season, we use the initial demand forecast developed in Zara1a project and converted it to distributional forecast. Once the season starts, the initial forecast can then be adjusted according to actual demand[2] of the article. This updated demand forecast is converted to updated distributional forecast, which is supplied to the optimization model to produce updated replenishment orders for the remaining weeks in the season. We developed various forecast updating methods and selected one method based on forecast accuracy metric selected via literature review on forecast error metrics as described in Section 3.1.

From the implementation perspective, forecast updating, distributional forecast and learning curve and optimization method programming was modular. The forecast updating code can update a given point forecast, pre-season or previous update, by comparing the input forecast to the actual demand. The distributional forecast code can take any point forecast, pre-season or updated, and related historical errors and output distributional forecast. Similarly, code for the optimization model is

[2] We refer to demand of an article as the quantity calculated after adjusting sales for stock-outs. The details for sales to demand calculation can be found in Bonnefoi's thesis [6].

independent of internal details of the previous phases. Appendix A shows a more detailed process map for the project.

3 Literature Review

The main goal of forecast updating method is to take into account any forecast errors in the previous forecasts after observing newer sales data and make adjustments in the forecasts for the remainder of product life. In this section we would like to review some of the literature in relation to forecast accuracy metrics and forecast updating methods.

3.1 Forecast Accuracy Measurement Literature

Demand forecasts are represented by time series as expected demand of the article by the calendar week. Each series starts at the earliest week that can be predicted, but for which the real sales have not been observed yet. For example, the pre-season forecast series starts from the first week of the selling period. The first forecast update series, made after observing one week of real sales starts from the second week of the selling period. The last week of the forecast horizon for a series is the week when the series demand settles down to zero or the last week before the markdown period, whichever comes first. Demand forecasts can be made at country level or aggregated across many countries or entire sales network.

Comparing two demand forecasts for the same article meant comparing two different time series. The goal was to find error measurement metrics that would allow us to compare against two demand series spanning multiple weeks. There were many factors that made traditional error measurement methods such as Root Mean Squared Error (RMSE) non-applicable for comparing two time series representing demand forecasts.

First, instead of averaging weekly forecast errors, we used cumulative errors over the remaining weeks in the forecast horizon because ordering decisions are made based on cumulative demand forecasts for the remaining weeks in the forecast horizon. Thus, we needed to develop updating methods, which would reduce the cumulative errors over the forecasting horizon, rather than reduce weekly errors. Also, as per [7], "One advantage of the cumulative-horizon error is simplicity. Using this single measure for calibration would be preferable to examining the error measure for each forecast horizon."

Second, in order to compare updating methods for a subfamily[3] of articles, it was important to summarize the forecast accuracy results from multiple articles, each with a different demand. This meant that the error summary method should be unbiased towards demand series with different scales. RMSE, which has been favored by many academics as well as industry in the 1980s, has the same dimensions as the series and therefore does not work well summarizing errors across series with different scales. Same argument can be applied for other dimensional error methods such as Mean Absolute Error (MAE) or Absolute Error (AE).

Armstrong and Collopy recommend that one "way to control for scale is to use percentage errors; that is, to calculate the error as a percentage of the actual value" [7]. One percent error method that is commonly is Mean Absolute Percent Error (MAPE). As we measure the percent error over the horizon instead of averaging over weeks, we called it Cumulative Absolute Percent Error (CAPE). Another way to avoid scale issue is to use relative errors to "compare the forecast errors from a given model against those from another model" [7]. However we found it very hard to base decisions using relative error such as Relative Absolute Error (RAE), which is simply the ratio of Absolute Error from one forecast divided by Absolute Error from another forecast. Our literature review supported this observation. According to [7], "Relative error measures do not relate closely to the economic benefits associated with the use of a particular forecasting method. For example, for the RAE, progressive reductions of errors from 40 to 20, then from 20 to 10, then from 10 to 5, and so forth would all be

[3] A subfamily is a group of articles from same department and similar physical characteristics. For example women's t-shirts form a subfamily.

judged to be of equal importance. Thus, relative error measures are typically inappropriate for managerial decision-making." While according to [8], "The MAPE can give us a very good indication of how well the forecasting method is performing. Because it is dimensionless, it facilitates communication to management." Flores also approves using MAPE as a method to compare different forecasting methods, by stating, "The comparison between forecasting methods is quite simple with this measure. It can also be used to compare results across time series" [8]. This was especially important as we represented our demand forecasts in the form of time series. Although Hyndman and Koehler noted disadvantages with using MAPE such as being biased towards lower forecasts and being skewed or undefined for close-to-zero or zero actual demand, they also recommend "If all data are positive and much greater than zero, the MAPE may still be preferred for reasons of simplicity" [9]. Note, that for comparing different forecasting methods, we used CAPE and therefore had very few cases where denominator was zero or close to zero. Also, as close to zero demand usually represented very small percentage of the total demand for the article, we were safely able to filter out those errors from our calculations. If we had not been able to filter such cases, CAPE would have in fact not been a useful metric for measuring forecast accuracy.

Yet another method to summarize errors across articles with different demand levels is Mean Absolute Scaled Error (MASE) proposed by Hyndman and Koehler [9]. Like, the relative error metrics, we found that MASE does not translate well to managerial insights. If the MASE of one method is greater than that of another by 0.1, what implications does it have on inventory? Also, MASE did not fit our first criteria of using a cumulative error metric as opposed to error metrics that are averages of weekly errors.

To conclude, no single error metric seemed to fit perfectly for all the criteria related to comparing updating methods. We decided to use CAPE as our primary metric for comparison. Since we are interested in forecasting only the regular campaign period and not the markdown period, the article demand is normally greater than zero. 'Divide by small number' cases could arise towards the end of article life cycle and before markdown period, when article demand is much lower than the first few weeks. In such cases, if weekly demand or forecast value is less than 2.5% of the sum of article demand and forecast over the article life, that week's CAPE is filtered so as to not bias CAPE across the articles. In the absence of other known data filtration techniques that could have worked, we proposed using this technique to avoid biasing error summaries on account of less relevant, yet hard to forecast weeks in the article life cycle.

3.2 Forecasting Literature

The pre-season forecast method to derive demand from sales information was covered under the previous project. Here we discuss some of the literature we came across while finding and refining the updating method for in-season forecasting.

When searching for updating methods for short-term forecasts, most papers point to time-series forecasting methods such as exponential smoothing. Taylor, de Menezes and McSharry compared the "accuracy of six univariate methods for short-term electricity demand forecasting" [10]. The demand data included weekly as well as daily seasonality patterns. The methods studied included Double seasonal ARMA model, Exponential Smoothing for double seasonality, Artificial neural network, A regression method with principal component analysis (PCA) and two simple benchmark methods - näive benchmark method, which is seasonal version of the random walk and näive benchmark with error model, which was a refined version of the seasonal random walk method [10].

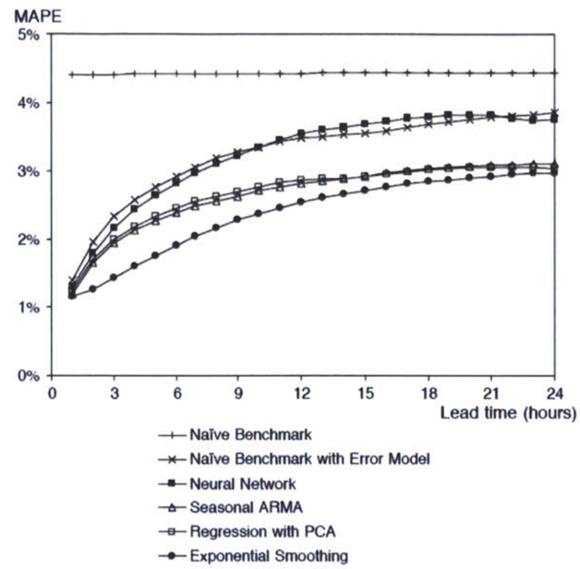


Figure 2. MAPE results plotted against lead time for 10-week post-sample period [10]

Figure 2, shows the results from the six methods for forecasts of over 10 weeks, with lead times of up to 1 day (24 hours). As seen in the figure and concluded by the authors, "overall, the best results were achieved with the exponential smoothing method, leading us to conclude that simpler and more robust methods, which require little domain knowledge, can outperform more complex alternatives" [10]. In his 1985 study of univariate forecasting methods, Gardner had supported exponential smoothing stating "Exponential smoothing is frequently the only reasonable time series methodology in large forecasting systems" [11]. In his 2006 paper Gardner confirmed, "Simple smoothing (N-N) is certainly the most robust forecasting method and has performed well in many types of series" [12].

Considering such empirical support for exponential smoothing methods and the fact that the methods developed will be used for predicting thousands of articles in any given campaign, we decided to develop our updating method based on exponential smoothing.

3.2.1 Distributional Forecasting Literature

It is desirable to provide a point forecast with related uncertainty in the form of an interval forecast or distributional forecast. Interval forecasts give the user an idea of the best and worse case scenarios [13]. They also aid in inventory planning decisions such as safety stock estimation. Forecast intervals are referred to in literature sometimes as confidence intervals but more widely as prediction intervals and "usually consists of an upper and lower bound with prescribed probability" [14]. Despite their importance, many companies do not use prediction intervals due to reasons described by Chatfield in [14]. Some of his explanations for this are cited below.

- The topic has been not covered very well in literature; textbooks provide little guidance on this topic except for regression and Autoregressive integrated moving average (ARIMA) modeling [14]. Taylor and Bunn show a theoretical method to estimate variance for exponential smoothing assuming that the optimal ARIMA is the true underlying model, but also point out that the problem with theoretical methods is that the variance estimation is "far from correct" if the optimum underlying model assumption is incorrect [15]. This could lead to bad estimations of spread of the distributional forecasting, thereby causing errors in inventory decisions made there-of.
- "There is no generally accepted method of calculating PI's except for procedures based on fitting a probability model for which the variance of forecast errors can be readily evaluated" [14]. We found many methods assuming normal distribution, but since our forecasting errors were not normally distributed the direct application of such methods was not possible.
- "Empirically based methods for calculating PI's are not widely understood and their properties have been little studied" [14]. We could not find an empirical based method for estimating prediction interval that worked well for cumulative errors or errors summarized over series with different levels.

The usual assumption underlying the generation of prediction intervals is that the forecast errors are normally distributed [13], [15]. Using this assumption, a prediction interval of k steps ahead forecast can be given as,

Equation 1

$$F_{t+k} \pm z\sqrt{MSE_k}$$

where F_{t+k} is the k-step ahead point forecast made at time t and MSE_k is Mean Squared Error of the k-step ahead forecasts. It is assumed that MSE_k is normally distributed with zero mean. The value z determines the width and probability of the interval. So z = 1.96 gives a 95% prediction interval, i.e. if z= 1.96, the probability that the true value of the quantity forecasted will lie within the above interval is 95% [13]. Other values for z for different probability of prediction interval can be found using normal inverse table such as one shown in Table 1.

z	Probabilty
0.674	0.50
1.000	0.68
1.150	0.75
1.282	0.80
1.645	0.90
1.960	0.95
2.576	0.99

Table 1. Values of z and Probability of the Prediction Interval [13]

In order to avoid the assumptions about validity of the model or the distribution of the forecasting errors, Gardner proposed a method based on Chebyshev's inequality. Here the variance and standard errors of fitted errors are calculated for different leadtimes. Then a multiplier, based on Chebyshev's inequality is applied to each standard error, to yield the desired prediction interval. Suppose *Y* is a random variable with mean μ and standard deviation σ, the interval is given as

Equation 2

$$P\left[\left|\frac{Y-\mu}{\sigma}\right| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}$$

The above expression states that the probability of an observation falling beyond ε standard errors from the mean is at most $1/\varepsilon^2$ [16]. Later empirical studies suggest that this method is of little practical use due to very wide prediction intervals in certain cases [15]. Very wide prediction intervals may lead to large safety stocks and left over inventory, which is not desirable for.

Williams and Goodman suggest a method to construct prediction intervals using the percentiles of empirical distribution of the post-sample forecasting errors [17]. Later studies find it computationally intensive and not used too often [15]. It is unclear if the two observations are causally related. Williams and Goodman used absolute errors, which we could not use as we combined errors from series with different levels. However, we borrowed the concept of using percentiles of historical forecast errors to construct prediction interval from point forecast as described in Section 6.1.

3.3 Optimization Review

The underlying concepts behind the optimization of the purchase orders come from the traditional inventory management theory. In this section we review literature on some of these concepts and terminology (where possible) as described in [18]. Chapter 4 describes a specific optimization model used for the Zara1 project.

3.3.1 Economic Order Quantity

The most basic of the inventory management concepts is that of the Economic Order Quantity (EOQ). The EOQ is the quantity which minimizes the total relevant costs of an article. Relevant costs include all costs that are related to the size of the order Q. The basic formula for the EOQ for a replenishment order for an article with constant demand is given as:

Equation 3

$$EOQ = \sqrt{\frac{2AD}{vr}}$$

D: Known demand rate for an article in units / unit time

A: Fixed component of replenishment ordering cost measured in €/order

v: Variable component of the replenishment ordering cost measured in €/unit

r: carrying charge i.e. cost of having one euro of the article tied up in inventory for a unit time, measured in €/€/unit time

Based on the above order quantity, the total relevant costs per unit time are given as,

Equation 4

$$TRC = \frac{AD}{Q} + \frac{Qvr}{2}$$

The EOQ method does not apply for ordering at Zara as the demand is usually not constant over time. The next section describes an extension of the above concept for probabilistic demand case.

3.3.2 Inventory Policy for Probabilistic Demand

When the demand is probabilistic, there are risks of stockouts (underages) or overages. Stockouts occur when demand is greater than inventory on hand and overages occur when the demand is lower than the inventory on hand; the excess inventory gets carried over, sometimes until the end of the season into the markdown period.

One of the inventory policies is to deal with these uncertainties is to maintain a safety stock by "specifying (explicitly or implicitly) a way of costing a shortage and then minimizing total cost" [18]. One way of doing this is to specify a shortage cost per shortage occasion. This cost is independent of number of units short, so it works well in cases where there are costs associated with not fulfilling an order, for example costs of expedited shipping. But in many retail setups there is an opportunity cost associated with each unit short. In such cases, a way of specifying a cost of shortage is by charging a fraction of unit cost of the article for each unit short. Under this consideration the total relevant ordering costs are given by,

Equation 5

$$TC = \frac{AD}{Q} + \left(\frac{Q}{2} + k\sigma_L\right) vr + \frac{B_2 v \sigma_L G_u(k) D}{Q}$$

TC: Total expected annual costs

$\frac{AD}{Q}$: Fixed ordering costs

$\left(\frac{Q}{2} + k\sigma_L\right) vr$: Inventory Holding Costs, Q/2 is the average inventory on hand per unit time and $k\sigma_L$is the safety stock.

$\frac{B_2 v \sigma_L G_u(k) D}{Q}$: Shortage costs,

$\sigma_L G_u(k)$is shortage costs per replenishment cycle

D/Q is number of replenishment cycles per year

$B_2 v$ is the cost per unit short

Silver, Pike and Peterson provide a detailed derivation of the method (pg. 263 of [18]). Direct application of this method for purchase optimization at Zara would be difficult as each article typically has multiple sourcing options, each with a different lead time L and a unit cost v. As a result the optimization method takes in to consideration multiple combinations of unit costs and lead times in deciding order quantity for replenishments. Still the above concepts are presented to relate the optimization formula described in Chapter 4 with theoretical inventory policies.

4 Optimization Model

A supply network such as Zara, which has multiple supply options for same article, involves one or more trade-offs amongst the various options. In the case of Zara, there is a trade-off between unit cost for the article and the order lead time. For example, Zara's own factories located in Arteixo have lead times as short as 15 days but higher unit costs as compared to factories in Turkey, which have relatively lower unit costs but higher lead times for delivery to one of the two DC's in Spain. Similarly, Asian suppliers have the longest lead times and the lowest unit costs.

Most articles have more than one supply option. So when deciding where to buy from, the buyers need to balance the uncertainty involved in ordering from long lead time suppliers with the associated cost benefits. To help evaluate such complex trade-offs, we implemented a linear optimization model developed by Professor Jérémie Gallien and Professor Adam Mersereau. We first explain the model concept in Section 4.1 followed by the optimization formulation in Section 4.2. Model implementation and sample outputs are included in Section 4.3.

4.1 Model Concept

A thorough explanation of the model concepts is in [19]. Key ideas necessary for understanding the model details are reiterated in this section.

4.1.1 Model Terminology

Let t index time in weeks. Important weeks are:

- t = 0 : the earliest week in which Zara can make a purchase order for the article.
- t = 1 : the first week of sales in the store network.
- t = T : the first week at which the article will be sold at a discount. This will mark the end of our planning horizon.

Supply Options:

S = {1, ..., J} represents the set of all supply options, where j indexes for the set. Each supply option S_j is defined by the combination of a vendor and a transportation mode. For example, if a supplier can ship via air or via sea, it has two separate S_j elements for, one for each transportation mode.

c_j: sum of unit production and unit (inbound) transportation costs when supply option **j** is used.

l_j: lead time of supply option **j**, defined as the maximum number of weeks required between the communication of an order to all stakeholders involved in coordinating supply option **j** (manufacturer, carrier and other providers of logistical services, any receiving and finishing operations at the warehouse, etc.) and the availability of the garments for distribution to Zara stores.

Selling Price:

p : average network-wide full selling price for the article, that is up until (and excluding) week **T**

$\mathbf{p}_m$: average network-wide markdown/clearance price for any article still unsold at the beginning of week **T** [19].

h: Holding cost per unit per week

4.1.2 Demand Scenarios

Figure 3 shows an example of eight demand scenarios for an article. "We assume that, viewed from the perspective of time t = 0, each of the scenarios occurs with a known probability. (The probability of scenario w being the true demand scenario is denoted by π_w)" [19].

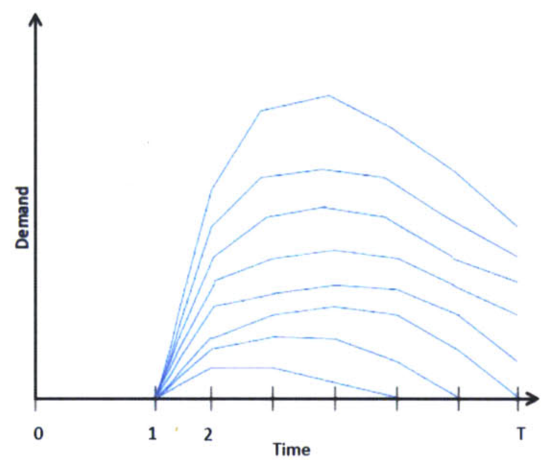


Figure 3. Demand Scenarios for the pre-season forecast

In order to model the evolution of demand forecast over time, we define 'information sets' for each week. "Scenarios are assumed to be indistinguishable from others within the same set in the same week" [19]. Figure 4 shows how information sets change over time. Important assumptions underlying the information sets are:

- The number of information sets is increasing in t. That is, as time passes the decision maker's uncertainty surrounding the true prevailing scenario is resolved.
- Information sets are nested in the sense that, if two scenarios are in different sets in week t, they must be in different sets in week t+1.

- At time t=1, all scenarios are in the same information set. That is, no learning takes place before sales begin.
- At week T, there are as many information sets as there are scenarios. That is, at the end of the selling season the full demand scenario has been unambiguously revealed [19].

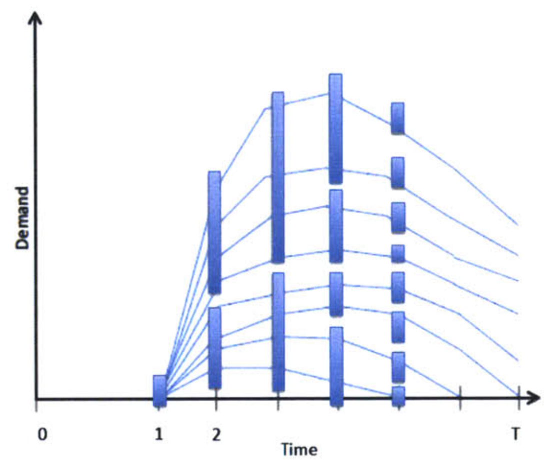


Figure 4. Information Sets Changes Over Time

In our implementation the information sets increase in order of powers of 2 i.e. (t) = 2^x (0,1,2,3,4,5) or n(t) = No. of demand scenarios. So, one information set splits into two and two into 4. This is one way to implement information sets, but it is recommended to explore other methods of forming information sets within the above assumptions to allow more granularity in information sets.

Model Notations

The following definitions support the above concepts.

W: Set of all scenarios.

|W|: Number of scenarios.

d_t^w: Demand in week t under scenario w.

$W_{k,t}$: Information set k in week t.

λ(t): Number of information sets in week t.

$\lambda(t) = 1$ for $t < t_s$ and $\lambda(t_m) = |W|$

4.2 Model Formulation

Decision Variables

$q^t_{w,j}$: The purchase quantity ordered from supply option j in week t (to be received in week t + l_j) under scenario w.

I^w_t: The inventory on-hand at the beginning of week t under scenario w.

x^w_t: Lost sales during week t under scenario w

Objective Function

$$\textbf{Minimize:} \qquad \sum_{w \in W} \pi_w \left(\sum_{t=0}^{T-1} \sum_{j \in S} c_j\, q^w_{t,j} + \sum_{t=1}^{T-1} p\, x^w_t + \sum_{t=0}^{T-1} h\, \mathrm{I}^w_t - p_m\, I^w_T \right)$$

Constraints

$$s.t. \quad I^w_{t+1} = I^w_t - d^w_t + x^w_t + \sum_{j \in S} q^w_{t-l_j,j} \quad for\ all\ w \in W, t \in [0, T)$$

$$q^w_{t,j} = q^{w'}_{t,j} \qquad for\ any\ j \in S, \quad \mathrm{t} \in [0, T)\ and\ k = 1, \dots, \lambda(t)\ such\ that\ (w, w') \in W_{k,t}$$

$$I^w_t, x^w_t, q^w_{t,j} \geq 0 \qquad for\ all\ w \in W, t \in [0, T], j \in S$$

[19]

The objective includes expected purchasing costs plus the expected lost revenues arising from lost sales plus the expected holding costs minus expected clearance revenue. Thus, the objective of the model is to maximize the profits by minimizing the costs and lost sales [19]. Note the original documentation did not contain the term $\sum_{t=0}^{T-1} h\, \mathrm{I}^w_t$ for holding costs, but it was included in the original code.

The first constraint equation ensures that the inventory sums are correct. i.e. under any scenario w, the on-hold inventory for article j in week t+1 is sum of inventory in week t and expected inventory arriving in week t from previous orders minus the sales in week t. Note sales are simply demand minus the lost sales. The second constraint function shows that there is only one purchase decision per information set, so same purchase decisions are made for all demand scenarios in the same information set [19].

If a given update week t has only one information set, average expected demand for future weeks is the average demand across all scenarios. The model evaluates the inventory on hand at time t against this average expected demand to make purchasing decisions. Whereas if there are two information sets in week t, the model calculates two different expected demand values. One average is taken across the high demand scenarios that fall in the top information set and the other average is taken across the low demand scenarios that fall in the bottom information set. Each of these two average expected demands is compared against inventory on hand at time t and then two different purchasing decisions are made. If the last observed demand is the bottom information set, the purchasing decision is made using average expected demand of all scenarios in the bottom information set is used. This purchase quantity may turn out to be small if the inventory on hand is greater than the average expected demand of this set of scenarios. If the observed demand falls in the top information set, then the purchasing decision are made using the average expected demand of all scenarios in the top information set. The resulting purchase quantity will be closer to the actual demand when there are two information sets as opposed to one. The same logic can be applied for further larger number of information sets.

Certain articles at Zara are single-order articles; these articles are only ordered once, before the selling season begins. For such articles, the number of information set for each week is equal to one. The resulting purchase quantity is optimized based on average expected demand of all scenarios, selling price and unit cost. Thus, in such cases, the model is similar to the newsboy model.

4.3 Model Implementation

We first used Microsoft Excel to input (output) the data to (from) the optimization, but later moved on to interfacing directly with Zara's SQL database. The following inputs are supplied to the model before any optimization run.

- The instantaneous demand scenarios for the new article under test based on the latest point forecast and historical error distribution.
- The information set array related to that optimization run. Information sets are updated before each run.
- For each supplier and transportation mode,
    - Unit Cost (euros)
    - Lead Time (weeks)
- Holding costs per unit (euro).
- Regular selling price and average clearance price for the article.
- Maximum Clearance Quantity – Maximum clearance quantity that can be sold at clearance price. In our implementation we always set clearance quantity to zero, as it was decided to only optimize regular season profits.

Table 2 shows an example of inputs supplied to demonstrate the model implementation. For illustration purposes, we used only two supply options – Supplier 1 and Supplier 2. We also created 100 uniformly distributed demand scenarios between the low and high scenarios as input to the model. Note these inputs, including demand and pricing information are created for demonstration purposes.

Supply Information		
Supply Option	Lead-Time (weeks)	Unit Cost (€/unit)
Supplier 1	10	6.25
Supplier 2	6	8.00

Pricing Information	
Price	€/unit
Regular Season	16.00
Clearance	6.00

Other Inputs		
Max Clearance Quantity	0	units
Holding Costs	0.001	(€/unit/week)

Demand			Learning
Time	Low Demand Scenario	High Demand Scenario	# Info Sets
0	0	0	1
1	0	0	1
2	0	0	1
3	0	0	1
4	0	0	1
5	0	0	1
6	0	0	1
7	0	0	1
8	0	0	1
9	0	0	1
10	224	568	1
11	381	969	2
12	229	581	2
13	188	478	2
14	132	336	2
15	76	194	4
16	66	168	4
17	81	207	4
18	76	194	4
19	76	194	4
20	71	181	10
21	61	155	10
22	41	103	20
23	30	78	20
24	10	26	50
25	5	13	100

Table 2. Input Data for Model Demo

The introduction date for the new article is t=10. The first run of the model is at time t=0, allowing 10 weeks of lead time for initial orders placed with Supplier 1. In practice, the first run could be performed earlier to accommodate longer lead times. Figure 5 and Table 3 show the average expected model output, after running it with above settings at time t=0. The green line in Figure 5 shows a average demand scenario. The yellow bars show the orders placed from Supplier 1 and the pink bars show the orders placed from Supplier 2. The black curve represents the average inventory on hand at the end of a given week. Red curve represents average lost sales for the week. For an outcome close to average expectation, the model recommends buying most units from Supplier 1 and average lost sales and clearance inventory as shown in Table 3. Since there is enough time for Supplier 1 pre-season order to arrive before sales start and Supplier 1 costs are lower than Supplier 2, there is no pre-season order placed with Supplier 2. There are multiple pre-season orders as we have holding costs but no fixed costs of ordering.

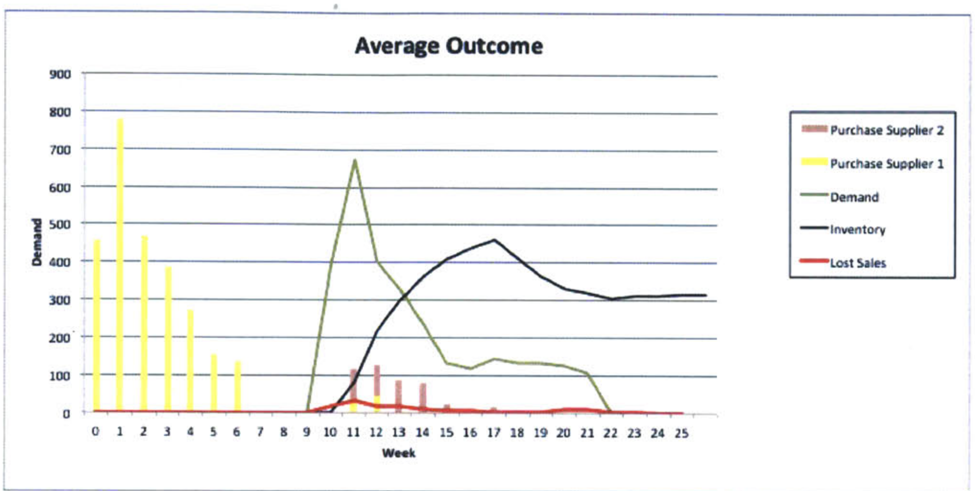


Figure 5. Plot of Average Model Outcome

Average Overall Performance			
		Units	Euros
	Season Demand	2934	46936
Revenues	Season Sales	2799	44786
	Clearance Sales	0	0
Costs	Lost Sales	134	2150
	Purchase Supplier 1	2717	16979
	Purchase Supplier 2	397	3175
	Total Purchase	3114	20154
	Inventory Costs		5
Profit			24627
	Clearance Inventory		314

Table 3. Average Model Outcome

Under the average scenario, the model recommends to place replenishment orders after learning increases in week 2. In practice, only the pre-season orders will be placed at time t=0 as actual demand can differ from the initial forecast.

In order to understand how the ordering between the two suppliers may differ under a different realization of actual demand, Figure 6 shows outcome of scenario with lower than average expected demand. Table 4 shows the related numerical results. Note that in this scenario, the model recommends only initial order from Supplier 1 and does not recommend replenishments orders from either supplier. This is because the model detects low actual demand compared to the left over inventory from the initial purchase. There is no need to place replenishment from either supplier. There are no lost sales in this scenario as demand is low, and the clearance inventory at the end of season is 175 units.

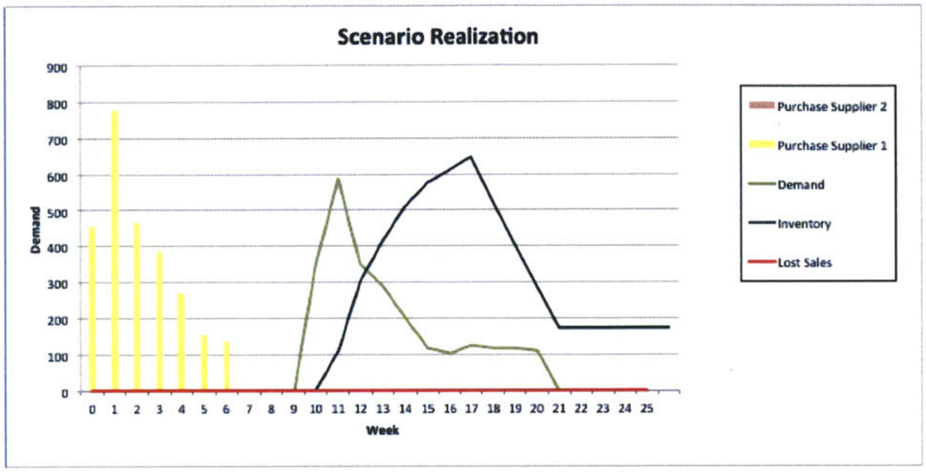


Figure 6. Plot of Model Outcome - Low Demand Scenario

Scenario to plot		36	
		Units	Euros
	Season Demand	2473	
Revenue	Season Sales	2473	39573
	Clearance Sales	0	
Costs	Lost Sales	0	
	Purchase Supplier 1	2648	16552
	Purchase Supplier 2	0	0
	Total Purchases	2648	16552
Profit			23020
	Clearance Inventory	175	

Table 4. Model Outcome – Low Demand Scenario

Figure 7 and Table 5 show the results for a scenario where demand is higher than expected demand. Here the model orders much higher replenishment quantity (compared to average scenario) from Supplier 2 in order to adjust inventory for higher than expected demand. It also recommends placing small replenishment orders from Supplier 1 in addition to the initial order, but this quantity is small due to longer lead time compared to Supplier 2 and lower demand at the time of order arrival. Clearance inventory is zero and lost sales are higher (80 units) compared to the average scenario.

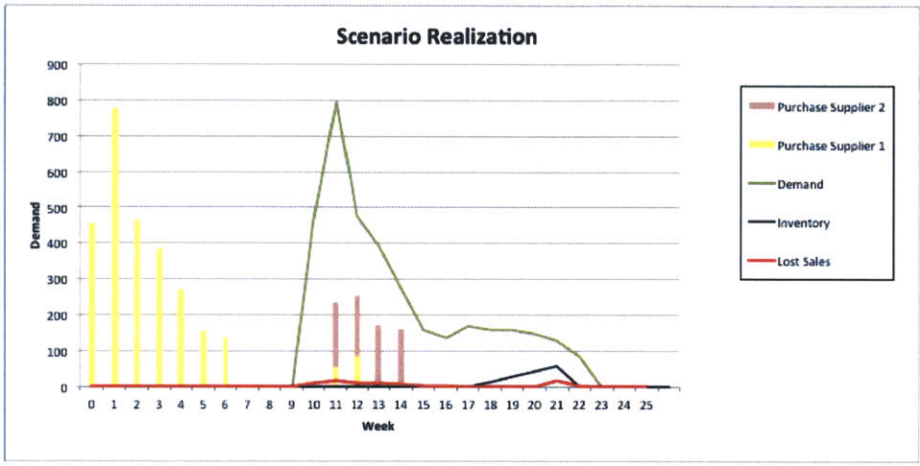


Figure 7. Plot for Model Outcome – High Demand Scenario

Scenario to plot		71	
		Units	Euros
	Season Demand	3559	
Revenue	Season Sales	3479	55668
	Clearance Sales	0	
Costs	Lost Sales	80	
	Purchase Supplier 1	2785	17405
	Purchase Supplier 2	694	5555
	Total Purchases	3479	22961
Profit			32708
	Clearance Inventory	0	

Table 5. Model Outcome – High Demand Scenario

The above demonstrations show that though model shows the recommendations for replenishment at the time of initial order placement, these recommendations can vary substantially as the actual demand is realized. Therefore, for any order period, the buyers should order quantities recommended for arrival before the next suggested run time. The next suggested run time in the demonstration was week t=2, but it depends on the distribution of historical forecast errors for the subfamily under consideration. Further information on model run-time procedure is included in 6.2.3.

5 Forecast Updating – Point Forecast

The main purpose of the forecasting updating process was to make adjustments in the pre-season forecast based on observed sales and to prepare a new forecast for the remaining weeks in the season. We compared seven main formulas, all adaptations of exponential or Holt-Winter's smoothing and tried many small variations such as different method to optimize forecasting parameters and removing seasonality factors to effectively result in about 20 different methods. With each exponential smoothing based method, we tried to accomplish the best possible way to combine the observed data and the pre-season forecast for the article using a weighted average. Holt-Winter's methods, as implemented in the theoretical form, used pre-season forecast only to initiate the series.

Seasonality factor changes were required to improve the accuracy of the pre-season forecast and are very specific to the data at Zara, so we will not discuss that in detail. We will, however, present our results on the top three methods as determined by CAPE results and further refinement of the final method that we are using for in-season forecasting at Zara.

5.1 Updating Methods Formulation

For each forecast updating methods, we start with a group of 'new' articles within the same subfamily and same campaign. Note, in order to analyze the forecasting errors, we need real sales information for the entire campaign, for both new and old articles. Therefore the new articles were selected from the latest completed summer campaign while the comparables were selected from the previous summer campaign. For each 'new' article, we find one or more comparable articles from previously finished campaign/s, based on similarity of attributes such as department (man vs. woman), style, fabric and if possible color. The comparable POS data and other buyer selected information such as article introduction date and sales network for new article is used to derive the new article demand forecast as described in [6]. This forecast is referred to as the pre-season (initial) forecast. This forecast is used for pre-season buying, but as the article is introduced and real sales are observed, we need to update the pre-season forecast to determine the replenishment order quantities.

We start by randomly dividing the new articles in to two equal sets: fitting and forecasting. The updating method is applied first to fitting group articles. For each update period t, the updated forecast is calculated, using one of the methods shown below, over a range of forecasting parameters. The forecasting results are compared against the demand of new article (derived from sales using methods from [6]). Next, the fitting group results are compared to find the best forecasting parameter from the range, as discussed in the individual method descriptions later. The updating formulas are then applied, using the selected parameter values, to the articles in the forecasting group. The methods are then compared based on their performance against the forecasting group.

For the forecasting results described in Sections 5.1.2 through 0, the dataset contained 36 article-comparable pairs found in the subfamily. An article-comparable pair contains one new article and one or more comparable articles. Two pairs can have same comparables, but not the same new article. The small data-set was due to the fact that, within the chosen pairs, many comparable and new articles contained two peaks in the demand curve and were removed from our data-set as special cases. The two (or more) peaks in sales were caused by two distribution peaks and were not considered as good representatives of typical sales curve. For 5.2.4 results, we added many articles from a different subfamily to the first data set and had total 64 article-comparable pairs.

5.1.1 Point Forecast Terminology

t: update period ; t >= 0

t = 0 refers to pre-season period and is NOT necessarily one week before t = 1. t = 0 in our calculations refers to time of pre-season forecast.

k: weeks (steps) ahead of the update period

T: Final week in the forecast series or final week of the campaign, whichever comes first.

Series Notation:

D = Actual demand series calculated over the regular selling period, prior to markdown sales.

Actual demand represents the demand for the new article derived using the POS and stock-out information. Please refer to Chapter 4 in [6] for details on demand derivation method. The start date for actual demand series is the first date of sales of the article.

$\mathbf{F_0}$ – Pre-season sales forecast series, made with 0 weeks of observed sales

$\mathbf{F_t}$ – Updated forecast series, made after observing t weeks of sales; $1 < t < T$

Forecast and demand for an individual week in the series:

$\mathbf{D_{t+k}}$ = Actual demand for a specific week t+k; $1 < t < T$ and $t+1 < k <= T-t$

$\mathbf{F_{0,\,t+k}}$ = Pre-season forecast for week t+k; $t=0$ and $1 < k <= T$;

$\mathbf{F_{t,\,t+k}}$ = Forecast for week t+k, made after observing t weeks of sales; $1 < t < T$ and $t+1 < k <= T-t$

$\boldsymbol{\alpha}$**(t):** Smoothing parameter for level used for F_t series calculations

$\boldsymbol{\beta}$: Smoothing parameter for trend

$\mathbf{S_t}$: Seasonality factor for week t; $1 < t < T$

The seasonality factor for a week t represents the strength of a given week in selling an article of given subfamily compared to that of the rest of the weeks in the season. Thus seasonality factor of week t for given subfamily of articles in a given country is calculated by normalizing the sales of the subfamily in week t with respect to the total sales of the subfamily in the country over the entire season. Further details of seasonality factor calculation can be found in [6]. *Note that if a new article is selling in Summer 2011, the weekly seasonality factors are calculated using Summer 2010 data, but we are using S_t notation instead of S_{t-52}.*

Forecast updating formulas presented next are at country level, as seasonality factors are different for different countries. Once an updated forecast series is calculated for all countries in the new article's sale network, they are aggregated by calendar week to derive the network-wide forecast.

5.1.2 Exponential (EXP2)

Equation 6

$$F_{t,t+k} = \left[\alpha(t)\frac{D_t}{S_t} + (1-\alpha(t))\left[\frac{F_{0,t+k}}{S_{t+k}}\right]\right] * S_{t+k}$$

Equation 6 above shows the simplest adaption of simple exponential smoothing we used. Forecasting parameter $\alpha(t)$ is optimized each update period t, by calculating the average CAPE from period t+5 through T for the fitting group articles over a range of $\alpha(t)$ values: 0.1, 0.2, ..., 0.9. The value of $\alpha(t)$ that gives the lowest average CAPE is then selected to be the 'optimum' value from the grid. The goal was to find updating methods that would reduce forecasting errors taking into consideration an approximate lead time. The optimized parameters are then used to make predictions for the forecasting group. The parameter selection is coded in basic programming language (java). It is recommended to explore opportunities to speed up the selection by using optimization software packages.

Equation 6 calculates forecast for week t+k, given t weeks of actual demand is available. It represents a weighted average of two terms. The first term contains latest weekly demand D_t, de-seasonalized using seasonality factor S_t and then re-seasonalized using S_{t+k} and multiplied with forecasting parameter $\alpha(t)$. The second term contains the pre-season forecast value for week t+k, multiplied by $1-\alpha(t)$.

Compared to traditional definition for exponential smoothing with no trend and multiplicative seasonality in [12], the Exponential method above uses $\frac{F_{0,t+k}}{S_{t+k}}$ instead of the level from week t-1. Note that since level is already de-seasonalized, there is no de-seasonalization required for the method described in [12]. The reason behind using this method is to base the forecast for a future week t+k, on the weighted average of the latest value of actual demand and pre-season forecast of week t+k. To calculate forecast for k weeks after the updating period t, Equation 6 uses the pre-season forecast value for week t+k rather than multiplying the level from week t with seasonality factor for week t+k.

Since we have the entire pre-season forecast for the new article, this method allows us to base each week's updated forecast on the pre-season forecast for that particular week. The traditional method's recursive formula, if applied here, would include the naïve forecast only to predict for second week in the season and not incorporate the other F_0 predictions for the future predictions at all. Since we wanted to incorporate the entire pre-season forecast, we applied the above-mentioned non-recursive formula.

5.1.3 Accumulative (ACC1)

Equation 7

$$F_{t,t+k} = \left[\alpha(t)\frac{D_t}{S_t} + (1-\alpha(t))\frac{\sum_{i=1}^{t} D_i}{\sum_{i=1}^{t} F_{0,i}}\left[\frac{F_{0,t+k}}{S_{t+k}}\right]\right] * S_{t+k}$$

Equation 7 calculates updated weekly forecast for week t+k, given that t weeks of actual demand is available. It represents a weighted average with two terms. The first term contains latest weekly demand D_t, de-seasonalized using seasonality factor S_t and then re-seasonalized using S_{t+k} and multiplied with forecasting parameter $\alpha(t)$. The second term contains the pre-season forecast value for week t+k, multiplied by $1-\alpha(t)$ and a ratio. The numerator in the ratio is sum of all *actual demand* values until week t and the denominator is the sum of all forecasted values until week t. Thus the ratio is used to compensate for any systematic over-estimation or under-estimation in the pre-season forecast series. Forecasting parameters $\alpha(t)$ are found using the same process as the one mentioned in 5.1.2.

The main intent behind this formulation was to be able to capture any systematic bias in pre-season forecast until the update period and avoid applying this bias for future updates. The ratio $r = \frac{\sum_{i=1}^{t} D_i}{\sum_{i=1}^{t} F_{0,i}}$ is used to adjust the pre-season weekly forecast for any bias observed in the previous weeks. For example, if the cumulative pre-season forecast until a week t is higher than the cumulative actual demand over the same period by a ratio r, all weekly forecasts made at time t decrease the weight of the pre-season forecast by r.

5.1.4 Holt-Winters' Method with Multiplicative Seasonality and Trend (HW3)

Equation 8: Level

$$L_t = \alpha(t)\frac{D_t}{S_t} + (1 - \alpha(t))\,(L_{t-1} + b_{t-1})$$

Equation 9: Trend

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)\,b_{t-1}$$

Equation 10: Forecast

$$F_{t,t+k} = (L_t + b_t k) * S_{t+k}$$

Initiation

Equation 11

$$L_0 = \frac{F_{0,1}}{S_1}$$

Equation 12

$$b_0 = 0$$

We wanted to adopt a smoothing formula from literature and evaluate its performance against the other methods we tested. So we formed Equation 8 - Equation 12 using the Holt-Winters' trend and multiplicative seasonality method in [13]. L_t in is the updated level of the series having observed the latest demand D_t. L_t in Equation 8 is a weighted average of D_t/S_t and the sum of last updated level and trend. Since the pre-season forecast values are not included in the L_t equation, the F_0 forecast has little effect on HW3 as opposed to the ACC1 and EXP2 methods. We initialized the series by setting the initial level L_0 to the de-seasonalized, pre-season forecast for week 1, $\frac{F_{0,1}}{S_1}$, as this was the best naïve estimate of the level available to us. The initial slope b_0 was set to zero. For each update period t, forecasting parameter $\alpha(t)$ parameter was selected, as shown in 5.1.2 to minimize the CAPE from period t+5 through T and β was set to 0.1, which allows slow but steady reaction to changes in the trend [13].

5.1.5 Updating Method Comparison

We compared all variations of updating methods we had developed using CAPE results and found that the above three methods had better results than the rest. The CAPE calculation for a forecast series F_t for a single article is as shown in Equation 13. For a forecast series, F_t, $CAPE_t$ [t+L:t] is measured from week L+1 through T, allowing for L weeks of lead time. The only exception is that pre-season sales forecast error at t=0 is measured from week 1 through week T under the assumption that this forecast is made more than L weeks ahead of the buying season.

Equation 13

$$CAPE_t[\mathrm{t+L:T}] = \frac{|\sum_{\mathrm{i=t+L}}^{\mathrm{T}} D_i - \sum_{\mathrm{i=t+L}}^{\mathrm{T}} \mathrm{F}_{\mathrm{t,i}}|}{\sum_{\mathrm{i=t+L}}^{\mathrm{T}} D_i}$$

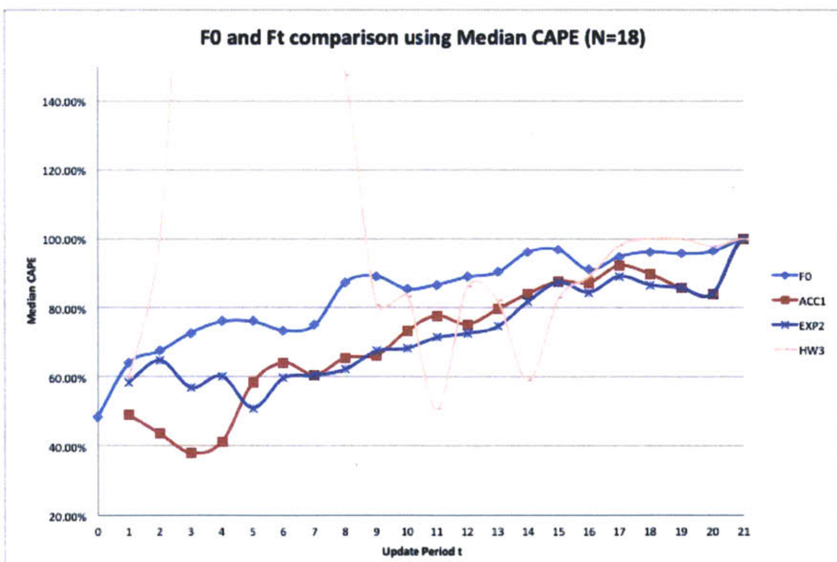


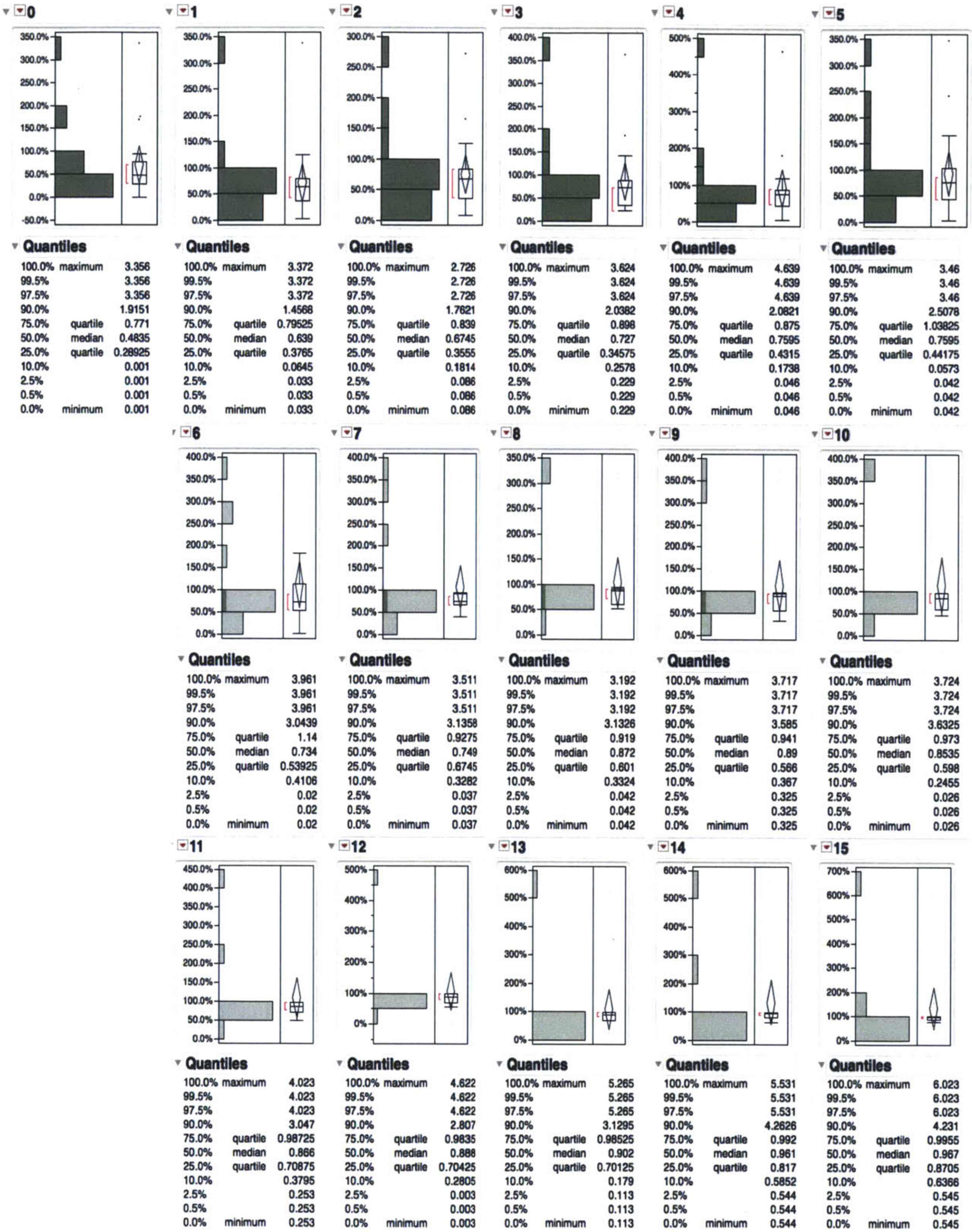
Figure 8. Updating Method Comparison based on Median CAPE

Figure 8 shows the median $CAPE_t$ [t+5:T] results from 18 Women's shirts articles from the forecasting group. The article life ranged from 11 weeks to 31 weeks, with average life of 22 weeks.

Thus, there are huge variations in the article life within the same subfamily of articles. The last 5 weeks of CAPE results in the figure are omitted as they represent 2 or fewer articles.

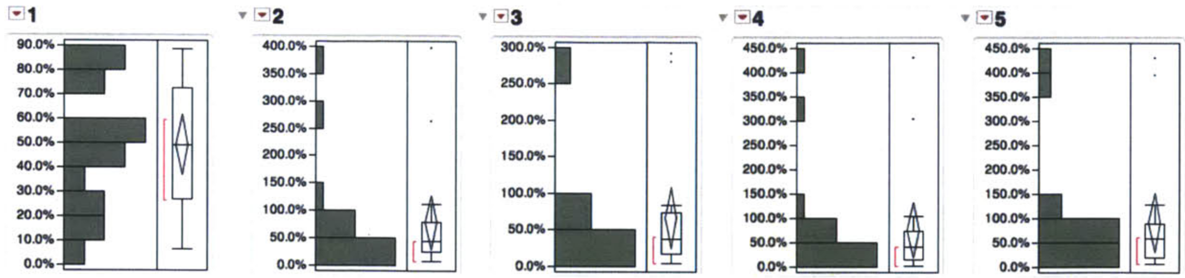
Figure 8 shows that, ACC1 was the better than the other EXP2 and HW3 as Median CAPE (MdCAPE) for ACC1 was consistently lower than F0 and HW3. Though EXP2 results were slightly better than ACC1 for few weeks updating weeks, ACC1 results were better in the first 5 weeks, which is a crucial period for determining replenishments for the rest of the season. All MdCAPE errors, including ACC1 errors, increased as the F_0 errors increased. The reason for this was that all $\alpha(t)$ values selected based on the grid values were closer to 0.1 than 0.9. HW3 performed very poorly as F_0 information was not used at all beyond initializing the series. HW3 results improved after 10 weeks of sales, which is a long period for Zara; the buyers would need to determine whether to replenish the article within few weeks of introduction.

Figure 9 through Figure 12 show distribution of CAPE errors, for pre-season forecast series F_0, ACC1, EXP2 and HW3, respectively. The column label represents the update period t. Figure 9 represents distribution for CAPE errors from the pre-season forecast series F_0; the first column shows the error $CAPE_0[5:T]$ measured on F_0, the second column shows distribution for $CAPE_1[6:T]$ measure on F_0, third columns shows distribution for $CAPE_2[7:T]$ and so on. Figure 10 through Figure 12 display CAPE errors from updated forecasts and therefore start from t=1. In Figure 10, the first column represents forecast error $CAPE_1[6:T]$ measured on updated forecast series F_1, the second column shows distribution for $CAPE_2[7:T]$ measure on F_1, third columns shows distribution for $CAPE_3[8:T]$ and so on. Figure 11 and Figure 12 are arranged in the same manner as Figure 10, except they show forecast errors from EXP2 and HW3 respectively.



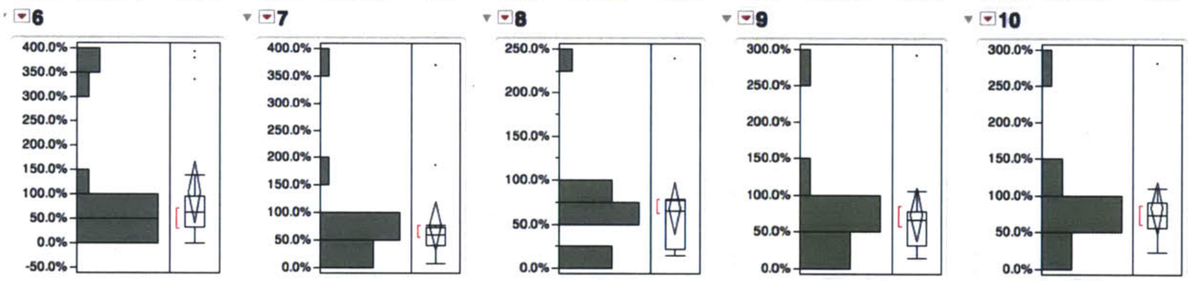
Quantile		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100.0%	maximum	3.356	3.372	2.726	3.624	4.639	3.46	3.961	3.511	3.192	3.717	3.724	4.023	4.622	5.265	5.531	6.023
99.5%		3.356	3.372	2.726	3.624	4.639	3.46	3.961	3.511	3.192	3.717	3.724	4.023	4.622	5.265	5.531	6.023
97.5%		3.356	3.372	2.726	3.624	4.639	3.46	3.961	3.511	3.192	3.717	3.724	4.023	4.622	5.265	5.531	6.023
90.0%		1.9151	1.4568	1.7621	2.0382	2.0821	2.5078	3.0439	3.1358	3.1326	3.585	3.6325	3.047	2.807	3.1295	4.2626	4.231
75.0%	quartile	0.771	0.79525	0.839	0.898	0.875	1.03825	1.14	0.9275	0.919	0.941	0.973	0.98725	0.9835	0.98525	0.992	0.9955
50.0%	median	0.4835	0.639	0.6745	0.727	0.7595	0.7595	0.734	0.749	0.872	0.89	0.8535	0.866	0.888	0.902	0.961	0.967
25.0%	quartile	0.28925	0.3765	0.3555	0.34575	0.4315	0.44175	0.53925	0.6745	0.601	0.566	0.598	0.70875	0.70425	0.70125	0.817	0.8705
10.0%		0.001	0.0645	0.1814	0.2578	0.1738	0.0573	0.4106	0.3282	0.3324	0.367	0.2455	0.3795	0.2805	0.179	0.5852	0.6366
2.5%		0.001	0.033	0.086	0.229	0.046	0.042	0.02	0.037	0.042	0.325	0.026	0.253	0.003	0.113	0.544	0.545
0.5%		0.001	0.033	0.086	0.229	0.046	0.042	0.02	0.037	0.042	0.325	0.026	0.253	0.003	0.113	0.544	0.545
0.0%	minimum	0.001	0.033	0.086	0.229	0.046	0.042	0.02	0.037	0.042	0.325	0.026	0.253	0.003	0.113	0.544	0.545

Figure 9. Distribution of CAPE Errors - F0



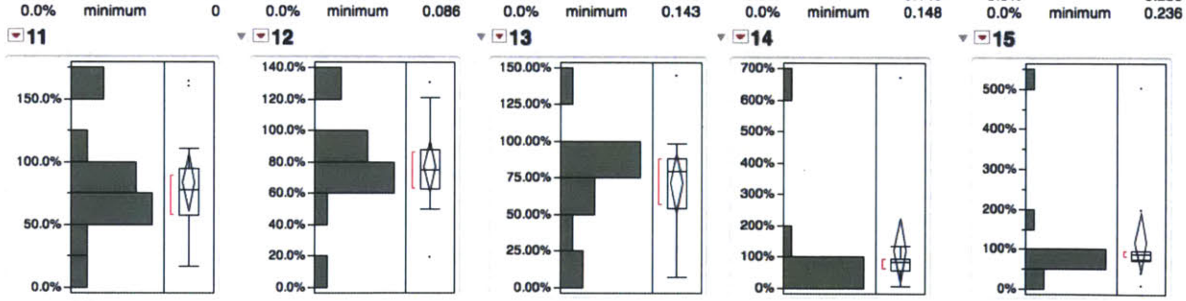
Quantiles

Quantile		1	2	3	4	5
100.0%	maximum	0.885	3.97	2.919	4.303	4.297
99.5%		0.885	3.97	2.919	4.303	4.297
97.5%		0.885	3.97	2.919	4.303	4.297
90.0%		0.8409	2.7658	2.8155	3.1618	3.9793
75.0%	quartile	0.72525	0.79125	0.73275	0.73925	0.8845
50.0%	median	0.4895	0.4355	0.38	0.413	0.584
25.0%	quartile	0.27075	0.2535	0.17075	0.1555	0.18675
10.0%		0.1549	0.0846	0.0552	0.0289	0.0615
2.5%		0.064	0.081	0.039	0.028	0.057
0.5%		0.064	0.081	0.039	0.028	0.057
0.0%	minimum	0.064	0.081	0.039	0.028	0.057



Quantiles

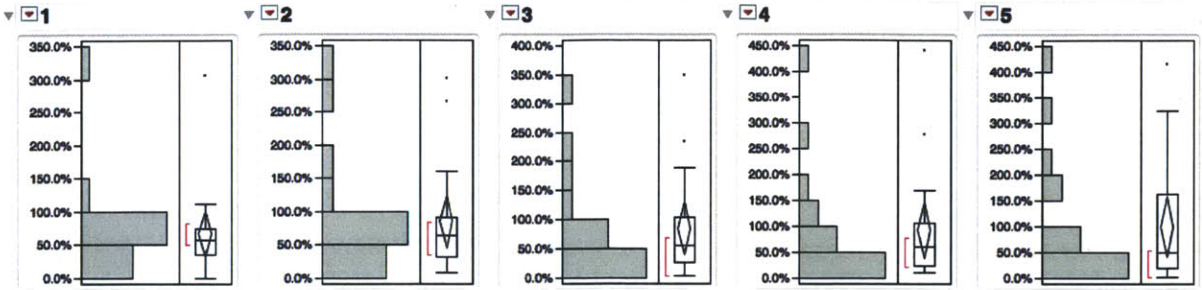
Quantile		6	7	8	9	10
100.0%	maximum	3.928	3.695	2.39	2.915	2.811
99.5%		3.928	3.695	2.39	2.915	2.811
97.5%		3.928	3.695	2.39	2.915	2.811
90.0%		3.8101	2.2278	1.4294	1.7972	1.958
75.0%	quartile	0.953	0.7425	0.772	0.787	0.9085
50.0%	median	0.6405	0.606	0.655	0.663	0.733
25.0%	quartile	0.3225	0.4035	0.222	0.321	0.56225
10.0%		0.0495	0.0892	0.164	0.2098	0.254
2.5%		0	0.086	0.143	0.148	0.236
0.5%		0	0.086	0.143	0.148	0.236
0.0%	minimum	0	0.086	0.143	0.148	0.236



Quantiles

Quantile		11	12	13	14	15
100.0%	maximum	1.639	1.305	1.446	6.699	5.013
99.5%		1.639	1.305	1.446	6.699	5.013
97.5%		1.639	1.305	1.446	6.699	5.013
90.0%		1.6195	1.2575	1.2175	4.5578	3.7962
75.0%	quartile	0.94975	0.87925	0.887	0.939	0.964
50.0%	median	0.7765	0.752	0.7955	0.841	0.876
25.0%	quartile	0.57125	0.62625	0.54575	0.5755	0.7485
10.0%		0.3275	0.346	0.1135	0.2412	0.2028
2.5%		0.168	0.193	0.07	0.084	0.076
0.5%		0.168	0.193	0.07	0.084	0.076
0.0%	minimum	0.168	0.193	0.07	0.084	0.076

Figure 10. Distribution of CAPE Errors – ACC1



Quantiles (1)

100.0%	maximum	3.058
99.5%		3.058
97.5%		3.058
90.0%		1.3057
75.0%	quartile	0.743
50.0%	median	0.582
25.0%	quartile	0.36225
10.0%		0.0556
2.5%		0.007
0.5%		0.007
0.0%	minimum	0.007

Quantiles (2)

100.0%	maximum	3.015
99.5%		3.015
97.5%		3.015
90.0%		2.6955
75.0%	quartile	0.9155
50.0%	median	0.648
25.0%	quartile	0.32825
10.0%		0.1353
2.5%		0.084
0.5%		0.084
0.0%	minimum	0.084

Quantiles (3)

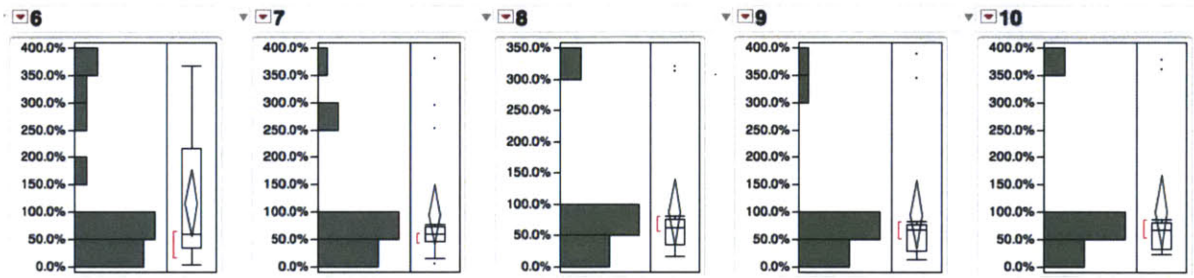
100.0%	maximum	3.499
99.5%		3.499
97.5%		3.499
90.0%		2.4595
75.0%	quartile	1.047
50.0%	median	0.5685
25.0%	quartile	0.2625
10.0%		0.053
2.5%		0.035
0.5%		0.035
0.0%	minimum	0.035

Quantiles (4)

100.0%	maximum	4.416
99.5%		4.416
97.5%		4.416
90.0%		2.9418
75.0%	quartile	1.06375
50.0%	median	0.6015
25.0%	quartile	0.2425
10.0%		0.1519
2.5%		0.106
0.5%		0.106
0.0%	minimum	0.106

Quantiles (5)

100.0%	maximum	4.153
99.5%		4.153
97.5%		4.153
90.0%		3.3268
75.0%	quartile	1.64125
50.0%	median	0.5095
25.0%	quartile	0.19325
10.0%		0.0758
2.5%		0.02
0.5%		0.02
0.0%	minimum	0.02



Quantiles (6)

100.0%	maximum	3.68
99.5%		3.68
97.5%		3.68
90.0%		3.5882
75.0%	quartile	2.168
50.0%	median	0.596
25.0%	quartile	0.339
10.0%		0.1683
2.5%		0.045
0.5%		0.045
0.0%	minimum	0.045

Quantiles (7)

100.0%	maximum	3.806
99.5%		3.806
97.5%		3.806
90.0%		3.1204
75.0%	quartile	0.7365
50.0%	median	0.604
25.0%	quartile	0.471
10.0%		0.1274
2.5%		0.053
0.5%		0.053
0.0%	minimum	0.053

Quantiles (8)

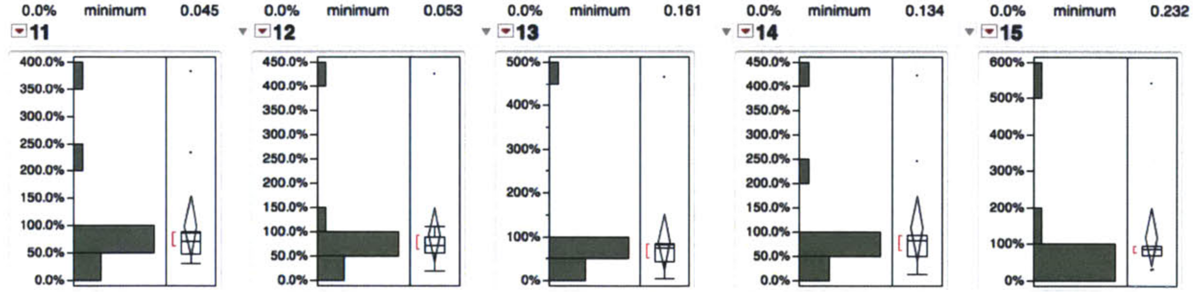
100.0%	maximum	3.202
99.5%		3.202
97.5%		3.202
90.0%		3.1564
75.0%	quartile	0.764
50.0%	median	0.622
25.0%	quartile	0.357
10.0%		0.1622
2.5%		0.161
0.5%		0.161
0.0%	minimum	0.161

Quantiles (9)

100.0%	maximum	3.888
99.5%		3.888
97.5%		3.888
90.0%		3.6222
75.0%	quartile	0.767
50.0%	median	0.674
25.0%	quartile	0.289
10.0%		0.2162
2.5%		0.134
0.5%		0.134
0.0%	minimum	0.134

Quantiles (10)

100.0%	maximum	3.78
99.5%		3.78
97.5%		3.78
90.0%		3.6925
75.0%	quartile	0.814
50.0%	median	0.6825
25.0%	quartile	0.33725
10.0%		0.242
2.5%		0.232
0.5%		0.232
0.0%	minimum	0.232



Quantiles (11)

100.0%	maximum	3.832
99.5%		3.832
97.5%		3.832
90.0%		3.0835
75.0%	quartile	0.8715
50.0%	median	0.7135
25.0%	quartile	0.49025
10.0%		0.3705
2.5%		0.31
0.5%		0.31
0.0%	minimum	0.31

Quantiles (12)

100.0%	maximum	4.256
99.5%		4.256
97.5%		4.256
90.0%		2.685
75.0%	quartile	0.88275
50.0%	median	0.7255
25.0%	quartile	0.5565
10.0%		0.224
2.5%		0.187
0.5%		0.187
0.0%	minimum	0.187

Quantiles (13)

100.0%	maximum	4.659
99.5%		4.659
97.5%		4.659
90.0%		2.751
75.0%	quartile	0.83175
50.0%	median	0.7455
25.0%	quartile	0.43825
10.0%		0.1025
2.5%		0.043
0.5%		0.043
0.0%	minimum	0.043

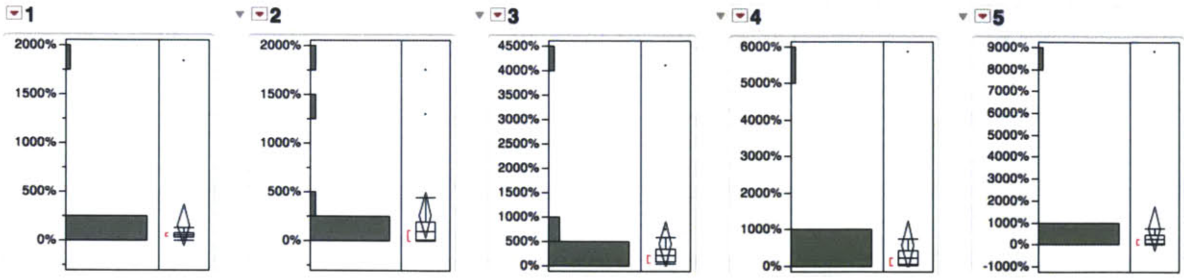
Quantiles (14)

100.0%	maximum	4.225
99.5%		4.225
97.5%		4.225
90.0%		3.5162
75.0%	quartile	0.9245
50.0%	median	0.818
25.0%	quartile	0.505
10.0%		0.2406
2.5%		0.141
0.5%		0.141
0.0%	minimum	0.141

Quantiles (15)

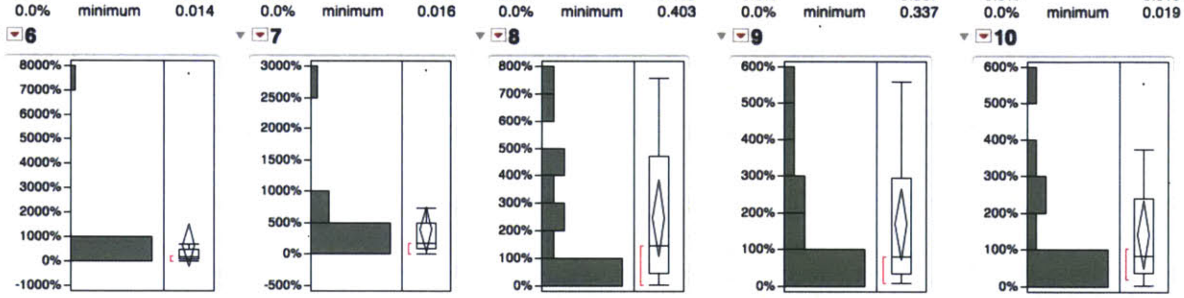
100.0%	maximum	5.428
99.5%		5.428
97.5%		5.428
90.0%		4.0032
75.0%	quartile	0.954
50.0%	median	0.873
25.0%	quartile	0.6965
10.0%		0.303
2.5%		0.299
0.5%		0.299
0.0%	minimum	0.299

Figure 11. Distribution of CAPE Errors - EXP2



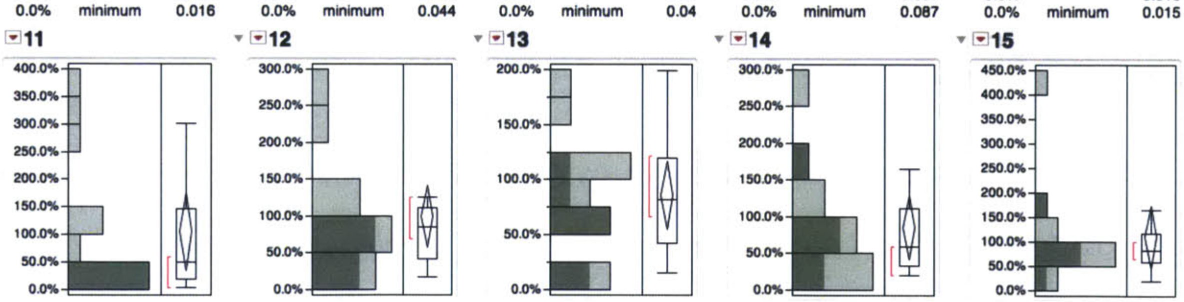
Quantiles

		1	2	3	4	5
100.0%	maximum	18.398	17.517	41.047	58.686	88.067
99.5%		18.398	17.517	41.047	58.686	88.067
97.5%		18.398	17.517	41.047	58.686	88.067
90.0%		2.9738	13.4382	10.8124	12.5637	15.6053
75.0%	quartile	0.80525	1.95825	3.4045	4.4815	4.36725
50.0%	median	0.5955	1	2.24	2.1895	2.3315
25.0%	quartile	0.36375	0.1325	0.7855	0.64525	0.56025
10.0%		0.1328	0.0214	0.5362	0.3847	0.0415
2.5%		0.014	0.016	0.403	0.337	0.019
0.5%		0.014	0.016	0.403	0.337	0.019
0.0%	minimum	0.014	0.016	0.403	0.337	0.019



Quantiles

		6	7	8	9	10
100.0%	maximum	76.726	29.27	7.594	5.586	5.537
99.5%		76.726	29.27	7.594	5.586	5.537
97.5%		76.726	29.27	7.594	5.586	5.537
90.0%		14.032	11.7804	7.114	5.0724	4.6345
75.0%	quartile	4.85125	4.952	4.705	2.971	2.39575
50.0%	median	1.728	1.815	1.475	0.809	0.834
25.0%	quartile	0.7085	0.8685	0.462	0.352	0.3805
10.0%		0.0295	0.0984	0.0658	0.1326	0.0295
2.5%		0.016	0.044	0.04	0.087	0.015
0.5%		0.016	0.044	0.04	0.087	0.015
0.0%	minimum	0.016	0.044	0.04	0.087	0.015



Quantiles

		11	12	13	14	15
100.0%	maximum	3.758	2.707	1.989	2.762	4.378
99.5%		3.758	2.707	1.989	2.762	4.378
97.5%		3.758	2.707	1.989	2.762	4.378
90.0%		3.3865	2.56	1.7775	2.318	3.288
75.0%	quartile	1.47725	1.12225	1.1985	1.1145	1.1785
50.0%	median	0.5065	0.86	0.82	0.589	0.827
25.0%	quartile	0.20075	0.42075	0.4295	0.3345	0.596
10.0%		0.0695	0.2775	0.1655	0.2294	0.2718
2.5%		0.033	0.179	0.156	0.207	0.195
0.5%		0.033	0.179	0.156	0.207	0.195
0.0%	minimum	0.033	0.179	0.156	0.207	0.195

Figure 12. Distribution of CAPE Errors - HW3

Comparing F0 CAPE variability with ACC1, EXP2 and HW3

In order to analyze variability of CAPE errors across different articles, we compared the difference between the minimum (0^{th}) and maximum (100^{th}) percentile[4] for the corresponding columns for F_0 and an update series (ACC1, EXP2 or HW3). First we compared the difference between the maximum and minimum $CAPE_t[t+5:T]$ percentiles for ACC1 with those for F_0, for update period t=1 through t=15 (see Figure 9 and Figure 10). Note T will vary by each article. By this comparison, ACC1 errors are less widely distributed compared to F_0 for most update periods except for t=2,5,7, 14. For example the difference between the maximum and minimum percentiles for F0, t=1 is $\Delta_{F0,\,t=1}$ = 3.372-0.001 = 3.371. Over the same period the difference between maximum and minimum quantiles for ACC1 F_1 update is , $\Delta_{ACC1,\,t=1}$ = 0.821. Hence for F_1 update using ACC1 method, the CAPE errors over entire update series are narrowly distributed around the median compared to the same period F_0 CAPE errors. The opposite is true for update periods t=2,5,7 and 14.

Comparing F_0 CAPE variability with that of EXP2 for different update periods, we see that EXP2 variability is lower than that of F_0 except for update periods t=2,3,5,7 and 9 using the same comparison method shown above (See Figure 9 and Figure 11).

Comparing F_0 CAPE variability with that of HW3 for different update periods, we see that HW3 variability is higher than that of F_0 except for update periods t=11 through 15 using the same comparison method as for ACC1 (See Figure 9 and Figure 12). Thus HW3 did not reduce either median CAPE or variability of CAPE compared to F_0 until week 11.

---

[4] We used JMP software to calculate the percentiles and to produce all the histograms and outlier box plots. Note that the 0^{th} and the 100^{th} percentile values were the same as the 2.5^{th} and 97.5^{th} percentile values respectively for all methods and all update periods shown in this section.

Comparing ACC1 and EXP2

Comparing ACC1 CAPE distribution with that of EXP2 for different update periods, we found that ACC1 CAPE is less variable than EXP2 CAPE except for update periods t = 2,5,6,14. However, for update periods t=2 and t=5, EXP2 results are still more variable than those for F_0, implying that there is no advantage of using EXP2 update or ACC1 update over pre-season forecast in those periods.

Based on above variability test, we decided that ACC1 reduces CAPE variability much faster than HW3. ACC1 results are also less variable than EXP2 in most update periods between t=1 and t=15[5].

As a result of this and past performances of the updating methods, we chose ACC1 over other methods. However, we still needed ACC1 errors to reduce significantly compared to the F_0 errors over time. So we adjusted it further as shown in the next section.

5.1.6 Accumulative with α-matrix (ACC15[6])

ACC1 uses an $\alpha(t)$ vector, which contains a different α value for each update period t. We realized that such one-dimensional α value was inadequate in tracking the sudden changes in in-season demand. Though seasonality factors helped capture some of the variation in demand over time, their effectiveness in tracking demand curve was limited.

In order to track the demand curve changes over the article life, we developed a two-dimensional forecasting parameter matrix, $\alpha(t, k)$ by minimizing the one-step ahead absolute errors, scaled with cumulative article demand. That is, to find an $\alpha(t, k)$ value, we minimized the sum of $|D_{t+k} - F_{t,t+k}| / \sum_t^T D$ over a range of $\alpha(t,k)$ values. This selection was done by first calculating all fitting group forecasts over a matrix of $\alpha(t, k)$ by varying t and k each, from 0.1-0.9 in steps of 0.1. Once the results

[5] We did not include variability analysis for update period t=16 onwards as there were 12 or fewer articles contributing data to weeks t>15.

[6] For code-reference at Zara, ACC15 refers to ACC1-MT5, but ACC15 is used here for brevity. Similarly ACC1 and EXP2 refer to ACC1-MT2 and EXP2-MT2 respectively.

for all $\alpha(t, k)$ combinations are calculated, for all articles in the fitting group, the best values for each t and k pair is selected by finding the one that reduces the $|D_{t+k} - F_{t,t+k}| / \sum_t^T D$. The scaling by total article was used to avoid biasing results towards articles with higher overall demand. There were other changes made simultaneously such as limiting the data to same start and end dates for the campaign so that any article sales beyond the normal campaign duration would be not be counted towards normal campaign sales. As a result of this change, the average life of articles reduced to 14 weeks with the total range of 7 weeks to 23 weeks. Also, as mentioned in Section 3.1, high article level CAPE errors appearing during weeks with last few weeks were removed from CAPE summary across multiple articles. Figure 13 shows the MdCAPE results from ACC15 method on the forecasting group containing 32 articles combined across women's shirts and t-shirts groups.

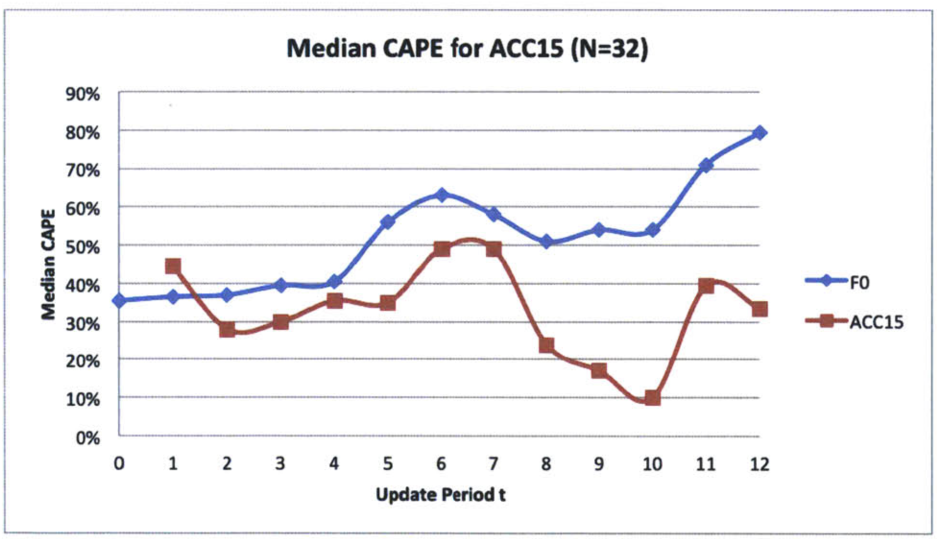
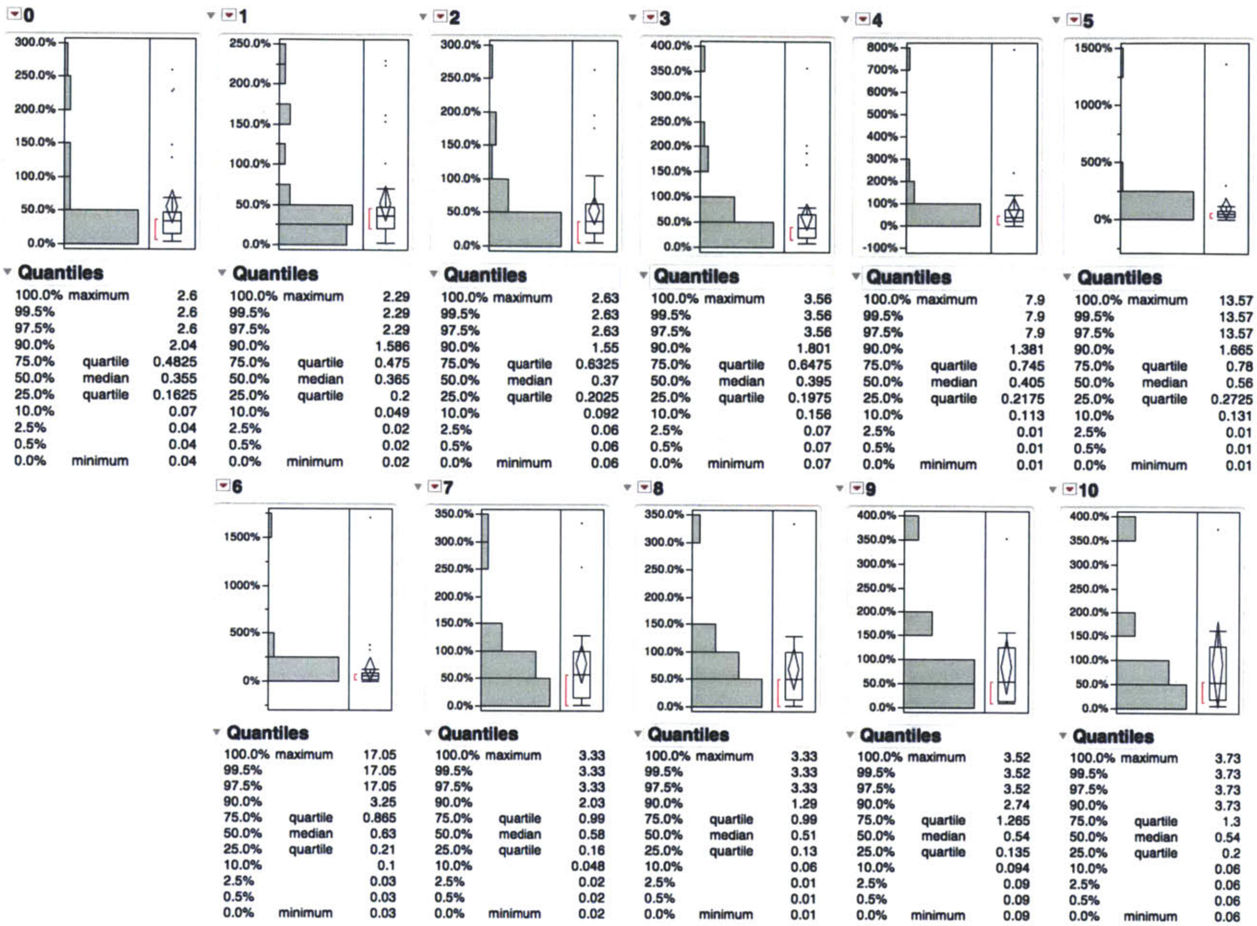


Figure 13. Median CAPE results for ACC15

Although, F_1 had higher MdCAPE than F_0 forecast over the same period, MdCAPE for F_2 was lower by 12% compared to the same period MdCAPE for F_0. Other update series (t=3, 4, ...) also showed improvements over the pre-season forecast F_0. Also, the gap between the two curves is generally increasing with t, suggesting that the forecast update errors were not increasing with the increasing F_0 errors. We decided to use ACC15 as our updating method to use with the optimization model.



Quantiles		0	1	2	3	4	5	6	7	8	9	10
100.0%	maximum	2.6	2.29	2.63	3.56	7.9	13.57	17.05	3.33	3.33	3.52	3.73
99.5%		2.6	2.29	2.63	3.56	7.9	13.57	17.05	3.33	3.33	3.52	3.73
97.5%		2.6	2.29	2.63	3.56	7.9	13.57	17.05	3.33	3.33	3.52	3.73
90.0%		2.04	1.586	1.55	1.801	1.381	1.665	3.25	2.03	1.29	2.74	3.73
75.0%	quartile	0.4825	0.475	0.6325	0.6475	0.745	0.78	0.865	0.99	0.99	1.265	1.3
50.0%	median	0.355	0.365	0.37	0.395	0.405	0.56	0.63	0.58	0.51	0.54	0.54
25.0%	quartile	0.1625	0.2	0.2025	0.1975	0.2175	0.2725	0.21	0.16	0.13	0.135	0.2
10.0%		0.07	0.049	0.092	0.156	0.113	0.131	0.1	0.048	0.06	0.094	0.06
2.5%		0.04	0.02	0.06	0.07	0.01	0.01	0.03	0.02	0.01	0.09	0.06
0.5%		0.04	0.02	0.06	0.07	0.01	0.01	0.03	0.02	0.01	0.09	0.06
0.0%	minimum	0.04	0.02	0.06	0.07	0.01	0.01	0.03	0.02	0.01	0.09	0.06

Figure 14. Distribution of CAPE Errors - F0 (N=32)

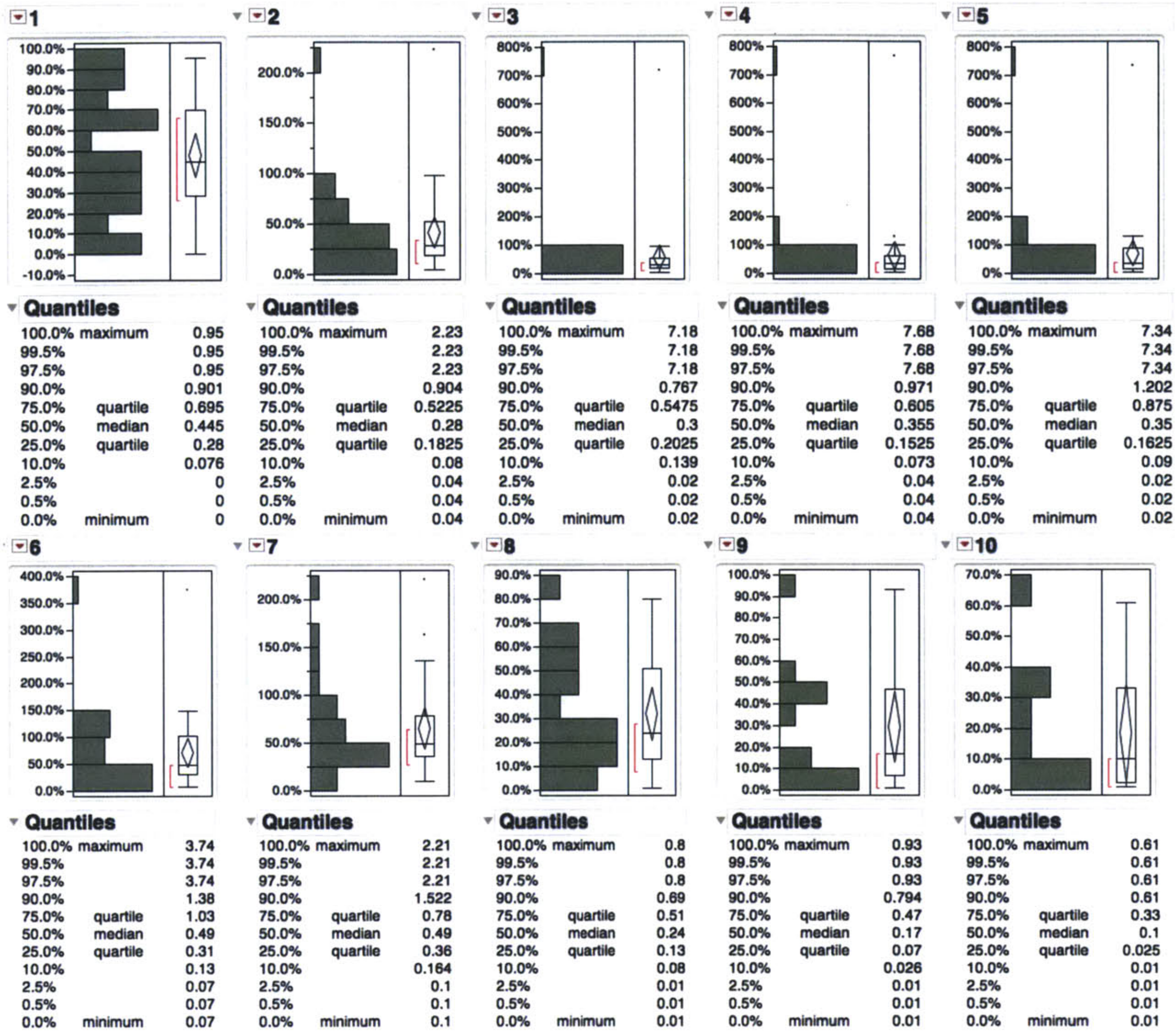


Figure 15. Distribution of CAPE Errors - ACC15 (N=32)

F_0 and ACC15 CAPE Variability Results

Figure 14 and Figure 15 show the distribution of CAPE errors for pre-season forecast F_0 and ACC15 for updating periods t=1 through t=10. For weeks t=11 and 12, there were only 8 or less data points due to many articles with short life cycles. The column labels represent the update period t. Each column contains CAPE errors for the corresponding forecast series (F_0 or ACC15), from period t+5 through T. T depends on the life of each article.

Comparing F_0 and ACC15 results on the difference between the minimum and maximum CAPE values for each update period, we see that ACC15 CAPE variability is lower than that of F_0 except for update periods t=3. Thus, ACC15 has lower median CAPE as well as lower variability in CAPE across articles compared to F_0.

5.2 Forecast Updating Results – Single Article

In this section we discuss forecasting results at article level in order to illustrate the application of the updating methods discussed in previous sections. Table 6 through Table 9 presents the results for an article for ACC1, EXP2, HW3 and ACC15 methods respectively. Note that the actual demand (D_t) and initial forecast (F_0) are identical for the first three methods, whereas there is a slight variation in these inputs for ACC15. This is due to fix in seasonality factor calculation[7] introduced before calculating results for ACC15.

5.2.1 ACC1 Results

Table 6 below shows article level forecasting results for the ACC1 method. Alpha vector represents forecasting parameter values $\alpha(t)$ for each update period t, calculated using the cumulative results from fitting group articles. The first column, t marks the week for which the forecast is being prepared. The inputs to the forecast updating method are the actual demand of new article and the initial forecast. The actual demand D_t, is shown in second column. Note that when forecasting for week 2 onwards, only actual demand for week 1 is used. The initial forecast F_0 is calculated prior to the selling period and therefore entire series is available at the time of updating.

Columns F1 through F15 represent the 15 update series made during the 16 weeks life of the article using Equation 7. D (t+1:T) represents the cumulative demand from week t+1 through T. F0 (t+1:T) represents the cumulative forecasted demand for week t+1 through T, based on initial forecast F_0. Similarly, Ft (t+1:T) represents the cumulative forecasted demand for week t+1 through T, based on updated forecast F_t. CAPE F0 represents the CAPE results from week t+1 through T, based on F_0 results.

---

[7] The fix to average by the number of stores carrying the article (SF302).

CAPE F0 represents the CAPE results from week t+1 through T, based on ACC1 results. The arrangement of columns and rows is same from Table 6 through Table 9.

Note that the $\alpha(t)$ values for the ACC1 method is mostly 0.1, implying that updated forecast will weight more on the initial forecast than the actual demand, after adjusting for the over or under estimation of demand in the previous week. Consider update F_1, forecast made after observing demand of 2 units in week 1 as opposed to predicted demand of 6 units. Since there was an over-estimation by factor of 3, the formula divides the second half in Equation 7 by 3 for all estimations from week 2 through 16 (T=16). This causes the resulting ACC1 forecast from week 2 through 16 to be low, when compared to the actual demand, resulting in a CAPE of 84%. After week 2, ACC1 improves compared to the original forecast: the reduction in ACC1 CAPE compared to that of F_0 for week 3 through 16 is 12%, while that from week 4 through week 16 is 9%. Despite the reduction, ACC1 CAPE increase as F_0 CAPE increases due to the low α values and high dependence on initial forecast.

ACC1

Alpha vector {0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.3,0.4,0.5,0.4,0.3,0.7,0.9,0.9,0.9,0.9,0.9,0.1}

t	Actual demand Dt	Initial Forecast F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
1	2	6															
2	1180	1065	339														
3	1115	1100	350	1211													
4	1237	901	287	1008	964												
5	1459	746	237	853	816	895											
6	1780	652	208	776	741	815	933										
7	1430	433	138	532	508	559	641	732									
8	861	330	105	431	411	455	523	594	643								
9	724	252	80	359	342	379	437	494	532	483							
10	381	145	46	238	226	253	292	328	350	305	296						
11	329	80	25	157	149	167	194	216	229	189	180	161					
12	215	37	12	113	106	121	142	156	162	120	109	90	93				
13	218	15	5	79	74	85	100	109	112	77	67	50	52	43			
14	108	3	1	60	56	65	77	83	84	52	42	27	29	21	24		
15	65	2	1	35	33	38	45	48	49	31	25	17	18	13	15	21	
16	52	2	1	43	40	47	55	60	61	37	31	20	21	15	18	26	35
D (t+1:T)		11157	11155	9976	8861	7624	6164	4384	2954	2092	1368	987	658	443	225	117	52
F0 (t+1:T)		5769	5763	4698	3598	2696	1951	1299	866	535	283	138	59	22	7	4	2
Ft (t+1:T)			1835	5894	4466	3878	3439	2820	2221	1293	750	366	212	92	57	48	35
CAPE F0		48%	48%	53%	59%	65%	68%	70%	71%	74%	79%	86%	91%	95%	97%	96%	96%
CAPE ACC1			84%	41%	50%	49%	44%	36%	25%	38%	45%	63%	68%	79%	75%	59%	32%

Table 6. ACC1 Results - Single Article

5.2.2 EXP2 Results

Table 7 shows results from the EXP2 method for the same input data as for ACC1 in previous section. The optimized α parameter values for EXP2 were low (0.1 or 0.2) for the first 14 update periods. As a result EXP2, like ACC1, follows the initial forecast closely. Unlike ACC1, EXP2 does not contain the ratio $\frac{\sum_{i=1}^{t} D_i}{\sum_{i=1}^{t} F_{0,i}}$ to adjust for errors in latest estimation. As a result, EXP2 follows F_0 even more closely than ACC1 does. Despite, this EXP2 improves on the initial forecast slightly as it has lower CAPE compared to F_0 for same period. For example, EXP2 CAPE is 8% lower than equivalent F_0 CAPE for week 3 through 16 and 7% lower than comparable F_0 CAPE.

While dependence on initial forecast is acceptable when the initial forecast is reliable, it challenges the need for updating. In the next section we explore the results of HW3, which depends very little on the initial forecast.

EXP2

Alpha vector {0.2,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.2,0.4,0.4,0.4,0.3,0.7,0.9,0.9,0.9,0.9,0.9,0.1}

t	Actual demand Dt	Initial Forecast F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
1	2	6															
2	1180	1065	853														
3	1115	1100	881	1109													
4	1237	901	721	925	918												
5	1459	746	597	784	777	794											
6	1780	652	522	715	707	727	753										
7	1430	433	347	492	485	501	522	531									
8	861	330	265	401	394	410	431	441	442								
9	724	252	202	335	329	345	367	378	379	317							
10	381	145	116	225	219	233	252	261	263	209	193						
11	329	80	64	149	144	156	172	179	181	136	123	103					
12	215	37	30	109	104	116	132	139	140	96	84	64	65				
13	218	15	12	77	73	83	96	102	103	67	56	39	41	31			
14	108	3	2	59	56	65	76	82	82	50	40	26	27	18	22		
15	65	2	2	35	33	37	44	47	48	29	24	15	16	11	13	14	
16	52	2	2	43	40	46	55	59	59	36	29	18	19	13	16	17	34
D (t+1:T)		11157	11155	9976	8861	7624	6164	4384	2954	2092	1368	987	658	443	225	117	52
F0 (t+1:T)		5769	5763	4698	3598	2696	1951	1299	866	535	283	138	59	22	7	4	2
Ft (t+1:T)			4616	5457	4279	3513	2899	2218	1697	941	549	265	167	75	51	31	34
CAPE F0		48%	48%	53%	59%	65%	68%	70%	71%	74%	79%	86%	91%	95%	97%	96%	96%
CAPE EXP2			59%	45%	52%	54%	53%	49%	43%	55%	60%	73%	75%	83%	77%	73%	34%

Table 7. EXP2 Results - Single Article

5.2.3 HW3 Results

Table 8 contains results for HW3 method on the inputs as ACC1 and EXP2 above. HW3 uses the initial forecast of the first week only to initialize the level for the series. As a result, HW3 uses no other weekly forecast from the F_0 series, so all future updates depend on initial forecast for first week of sales, observed weekly demand and consequent changes in level and trend.

Consider F_1 update. Since only one week of demand, equal to 2 units, is observed, the estimated demand for the entire season is very small. This estimation results in large errors in demand estimates for the later weeks, as the first week generally has lower demand than the rest of the article life. The seasonality factor adjustment on the F_0 forecast is not enough to accommodate the change in demand from week 1 through week 16. Thus, HW3 fails to predict future demand compared to other forecasting methods. Although, HW3 CAPE at 42% is much lower for week 5 through 16 compared to same period F_0 CAPE at 65%, about 45% of season demand has already been realized in the first 4 weeks, so the resulting benefits of the HW3 forecasts could only be applied to 56% of the remaining article demand. HW3 typically resulted in much higher errors during the first few weeks compared to ACC1 and EXP2 as seen in the aggregate results in Figure 8. Thus HW3 generally will not be useful in planning replenishments.

HW3

Alpha vector {0.6,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.4,0.1,0.9,0.1,0.1,0.1,0.1,0.1,0.1,0.1}

Beta 0.1

t	Actual demand Dt	Initial Forecast F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
1	2	6															
2	1180	1065	4														
3	1115	1100	4	134													
4	1237	901	3	139	242												
5	1459	746	3	149	261	369											
6	1780	652	3	182	320	454	591										
7	1430	433	2	155	272	388	506	610									
8	861	330	2	167	295	422	551	667	758								
9	724	252	2	186	329	472	619	750	855	867							
10	381	145	1	170	302	434	570	692	791	802	679						
11	329	80	1	148	263	379	499	607	694	705	594	350					
12	215	37	1	153	272	393	518	631	723	734	616	355	335				
13	218	15	1	134	239	345	456	557	639	648	542	306	287	188			
14	108	3	0	125	224	324	429	524	602	611	509	282	264	167	193		
15	65	2	0	75	134	194	256	314	361	366	305	165	154	95	111	63	
16	52	2	0	98	175	254	337	413	475	482	400	214	199	118	140	74	77
D (t+1:T)		11157	11155	9976	8861	7624	6164	4384	2954	2092	1368	987	658	443	225	117	52
F0 (t+1:T)		5769	5763	4698	3598	2696	1951	1299	866	535	283	138	59	22	7	4	2
Ft (t+1:T)			26	2014	3329	4429	5333	5764	5898	5215	3646	1672	1239	568	444	137	77
CAPE F0		48%	48%	53%	59%	65%	68%	70%	71%	74%	79%	86%	91%	95%	97%	96%	96%
CAPE HW3			100%	80%	62%	42%	13%	31%	100%	149%	166%	69%	88%	28%	97%	17%	49%

Table 8. HW3 Results - Single Article

5.2.4 ACC15 Results

Table 9 shows ACC15 CAPE results on similar pre-season forecast and demand data as that used for ACC1, EXP2 and HW3. Table 10 contains the optimized $\alpha(t,k)$ matrix found using the fitting group articles. The matrix shows that for a fixed update period t, α value generally decrease as k increases. This shows that as the forecasted week is further away from the update week, the estimates have more weight on the initial forecasts rather than on the last observed demand.

Table 9 shows that the ACC15 F_1 series produced extremely low estimates for week 2 through 16 compared to the actual demand. This was due to the fact that the week 1 actual demand was very small compared to the future demand. Additionally, the factor $D_1/F_{0,1} = 2/10$, further lowered these estimates by a factor of 5. These article level results as well as the cumulative F_1 errors for ACC15 shown in Figure 13 show that ACC15 does not improve forecast over the initial forecast after just one week of sales. However, ACC15 does lower the same period CAPE errors from F_2 onwards. For example, Table 9 shows that the CAPE from F_2 (forecast for week 3 through 16) equals 8% as compared to the same period initial forecast error of 51%. Similarly, the CAPE for F_3 forecast (week 4 through 16) is 27% while that of F_0 forecast on same period is 27%. Future updates, F_4, F_5, F_6, ... continue to improve over the same period F_0 forecasts. Unlike HW3 ACC15 shows faster improvement over F_0 estimates early in the season. Such early improvement in forecasts will lead to faster replenishment quantity decisions. Also ACC15 has larger improvements over F_0 even in later update periods. We chose to use ACC15 based on such article level and group level forecast results analysis.

ACC15

t	Actual demand Dt	Initial Forecast F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
1	2	10															
2	1159	769	138														
3	1086	901	38	1422													
4	1227	1053	24	1845	1540												
5	1422	774	18	1378	1154	1053											
6	1741	657	80	1079	940	897	1312										
7	1390	439	62	751	650	634	853	1280									
8	854	332	53	606	521	524	761	1241	1302								
9	710	307	55	618	525	471	652	888	1030	823							
10	369	168	30	405	337	296	457	530	792	692	586						
11	314	72	13	239	194	164	197	342	563	441	381	260					
12	191	41	7	202	160	132	161	188	309	285	204	188	300				
13	183	19	3	155	120	96	121	141	141	164	116	104	110	128			
14	100	3	0	107	81	63	81	95	93	59	40	46	50	69	103		
15	69	2	0	65	49	38	49	58	57	36	24	16	17	11	20	20	
16	50	0	0	48	36	28	36	42	41	25	16	9	10	12	13	18	37
D (t+1:T)		10867	10865	9706	8620	7393	5971	4230	2840	1986	1276	907	593	402	219	119	50
F0 (t+1:T)		5547	5537	4768	3867	2814	2040	1383	944	612	305	137	65	24	5	2	0
Ft (t+1:T)			521	8920	6307	4396	4680	4805	4328	2525	1367	623	487	220	136	38	37
CAPE F0		49%	49%	51%	55%	62%	66%	67%	67%	69%	76%	85%	89%	94%	98%	98%	100%
CAPE ACC15			95%	8%	27%	41%	22%	14%	52%	27%	7%	31%	18%	45%	38%	68%	26%

Table 9. ACC15 Results - Single Article

Alpha (t,k) for ACC15

α(t,k)	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11	k=12	k=13	k=14	k=15	k=16	k=17	k=18	k=19	k=20
t=1	0.1	0.8	0.9	0.9	0.4	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
t=2	0.8	0.5	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
t=3	0.7	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0
t=4	0.6	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0
t=5	0.7	0.4	0.4	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0
t=6	0.8	0.8	0.3	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
t=7	0.9	0.4	0.4	0.4	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0
t=8	0.5	0.6	0.5	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0
t=9	0.9	0.7	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0	0
t=10	0.8	0.5	0.3	0.2	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0	0	0	0	0	0	0	0	0
t=11	0.9	0.3	0.2	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0	0	0	0	0	0	0	0	0	0
t=12	0.7	0.5	0.1	0.2	0.3	0.4	0.2	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0
t=13	0.7	0.2	0.2	0.4	0.3	0.3	0.2	0.2	0	0	0	0	0	0	0	0	0	0	0	0
t=14	0.3	0.4	0.6	0.3	0.6	0.5	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0
t=15	0.7	0.9	0.3	0.8	0.6	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t=16	0.9	0.9	0.9	0.9	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t=17	0.9	0.9	0.9	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t=18	0.6	0.4	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t=19	0.7	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t=20	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 10. Alpha (t,k) Matrix for ACC15

6 Distributional Forecasts and Learning Curve

The optimization model requires, as input, distributional forecast rather than a point forecast. This chapter shows how we derive distributional forecast (demand scenario) and learning curve, which are used as inputs to the optimization model.

6.1 Point Forecast to Distributional Forecast

In this section we describe our method for deriving distributional forecast from point forecast for a new article and historical accuracy results on similar articles.

6.1.1 Distributional Forecast Terminology

While the key terminology remains the same as the point forecast, additional terms or changes specific to distributional forecast calculations are as follows.

m: index for first week in any range of weeks such as m:T

Cumulative forecast from week m through T:

Equation 14

$F_t\,[m{:}T] = \sum_{i=m}^{T} F_{t,i}$; $t < m <= T$

Cumulative demand from week m to T:

Equation 15

$D\,[m{:}T] = \sum_{i=m}^{T} D_i$; $t < m <= T$

Cumulative D/F Error: Demand/Forecast ratio showing the over or underestimation of demand

Equation 16

$\frac{D}{F_t}[\mathrm{m{:}\,T}] = \frac{\mathrm{D\,[m{:}T]}}{F_t\,\mathrm{[m{:}T]}} = \frac{\sum_{i=m}^{T} D_i}{\sum_{i=m}^{T} F_{t,i}}$; $t < m <= T$

Percentile notation:

Suppose random variable X represents the distribution for $\frac{D}{F_t}[\mathrm{m{:}\,T}]$ across multiple articles. Then, the P^{th} percentile for this distribution is the smallest value x such that the probability of the random variable assuming a value less than x is less than P/100.

$$\left\{\frac{D}{F_t}[\mathrm{m{:}\,T}]\right\}^{\mathrm{P}} = x \quad s.t. \quad Pr[X < x] \leq \mathrm{P}/100$$

Here, P=2.5 refers to the 2.5^{th} percentile and P=97.5 refers to 97.5^{th} percentile values. A numerical example is provided in the next section.

Additional terminology is added in the following sections as required.

6.1.2 D/F errors

We used Cumulative Demand/Forecast (CD/F) ratio to calculate the spread of the forecast errors and derive distributional forecasts. Equation 14 through Equation 16 show how we calculated the cumulative D/F error. Here on, we will refer to cumulative D/F ratio as CD/F or when needed, in more precise terms as D/F [m:T].

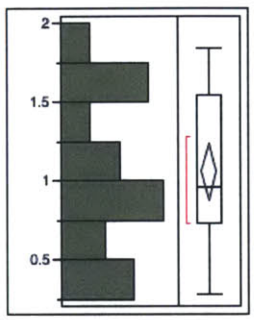


Figure 16. CD/F Distribution for Pre-season Forecast

As seen in Figure 16, a histogram of the CD/F errors from the shirt and t-shirt articles showed that the errors were not normally distributed. So we used empirical error distribution to derive the distributional forecast. We used a java function to estimate the percentile values for CD/F errors from 2.5^{th} percentile to 97.5^{th} percentile in steps of 2.5, thus resulting in 39 different percentile values, each with a probability of 1/39. Since there were only 32 data points for each percentile estimation, using a more granular percentile separation would result in more interpolated values in between actual data points. If larger datasets are available, we recommend using more refined percentile estimation. The java function emulates the same formula as that used in 'percentile' function in Microsoft Excel 2007. See Appendix B for a mathematical notation of this function.

Here, the 2.5th percentile for the CD/F error from week m through T calculated on forecast update F_t is,

$$\left\{\frac{D}{F_t}[\mathrm{m:T}]\right\}^{2.5} = 1.84$$

And 97.5th percentile is,

$$\left\{\frac{D}{F_t}[\mathrm{m:T}]\right\}^{97.5} = 0.28$$

Therefore,

$$Pr\left[0.28 < \frac{D}{F_t}[\mathrm{m:T}]\right] \leq \frac{2.5}{100} \quad \text{and}$$

$$Pr\left[\frac{D}{F_t}[\mathrm{m:T}] < 1.84\right] \leq \frac{97.5}{100}$$

$$\Rightarrow Pr\left[0.28 < \frac{D}{F_t}[\mathrm{m:T}] < 1.84\right] \leq \frac{95}{100} = 95\%$$

6.1.3 Error Distribution to Demand Distribution

This section describes our method for combining historical error distribution with the new article point forecast to find the distributional forecast. The point forecast for the new article is marked by the letter j in order to distinguish it from the historical forecast notations.

Expected Cumulative Demand Distribution:

Using the **historical cumulative** error **distribution** from forecast F_t across a group of articles and **cumulative point** forecast $F_{t,j}$ [m:T] for a **new** article j, expected cumulative demand distribution for the article j is calculated as follows:

Equation 17

$\forall$P s.t. P $\in$ {2.5, 5, 7.5,...., 97.5},

$$\{\mathrm{ED}\ [\mathrm{m:T}]\}^{\mathrm{P}}_{\mathrm{F}_{t,j}} = \left\{\frac{D}{\mathrm{F}_t}[\mathrm{m:T}]\right\}^{\mathrm{P}} * \mathrm{F}_{t,j}\ [\mathrm{m:T}]$$

Instantaneous Demand Scenarios:

Each percentile value for cumulative expected demand maps to a cumulative demand scenario. These demand scenarios need to be converted into instantaneous or weekly demand scenarios. To calculate instantaneous expected demand distribution for a given article for week m, we subtract each of the 39 cumulative demand scenarios over [m+1, T] from the respective scenario over [m, T].

Expected Instantaneous Demand Distribution for article j, for week m, at P^{th} percentile based on forecast F_t is thus given as:

$\forall P$ s.t. $P \in \{2.5, 5, 7.5, \ldots, 97.5\}$,

$$\{\mathrm{ED}\ (\mathrm{m})\}^{\mathrm{P}}_{\mathrm{F}_{t,j}} = \{\mathrm{ED}\ [\mathrm{m}: \mathrm{T}]\}^{\mathrm{P}}_{\mathrm{F}_{t,j}} - \{\mathrm{ED}\ [\mathrm{m}+1: \mathrm{T}]\}^{\mathrm{P}}_{\mathrm{F}_{t,j}}$$

After calculating all percentiles, for each week m, we get our 95 percent prediction interval or in other words our 39 instantaneous demand scenarios for article j based on forecast F_t.

Example of Instantaneous Expected Demand Derivation

Figure 17 shows the sample inputs required to estimate instantaneous demand scenarios of article for weeks 1 through 3 of the selling period. The table on the left shows historical distribution Cumulative D/F errors for periods 1:T, 2:T, 3:T and 4:T. This example demonstrates derivation for the four lowest demand scenarios, so only values for percentiles 2.5 through 10.0 will be used. The table on the right side of the figure contains the information for new article, j, to be forecasted. Column F_0 contains weekly pre-season forecast for article j. Column F0[m:T] shows the cumulative estimated forecast based on F_0 from weeks m through T (T=20).

Percentile	Cumulative D/F error distribution up to D/F [4:T]			
	D/F [1:T]	D/F [2:T]	D/F [3:T]	D/F [4:T]
97.5	1.805	1.763	1.954	2.831
95.0	1.758	1.726	1.856	2.460
92.5	1.708	1.720	1.785	2.033
90.0	1.694	1.720	1.750	1.774
87.5	1.645	1.685	1.735	1.695
85.0	1.594	1.640	1.700	1.662
82.5	1.573	1.605	1.595	1.564
80.0	1.558	1.582	1.472	1.406
77.5	1.547	1.543	1.391	1.307
75.0	1.540	1.480	1.370	1.300
72.5	1.484	1.466	1.363	1.293
70.0	1.388	1.384	1.332	1.286
67.5	1.271	1.268	1.287	1.275
65.0	1.208	1.254	1.266	1.240
62.5	1.180	1.250	1.255	1.230
60.0	1.154	1.218	1.236	1.230
57.5	1.098	1.106	1.187	1.230
55.0	1.078	1.084	1.156	1.218
52.5	1.037	1.062	1.125	1.201
50.0	0.960	1.020	1.090	1.180
47.5	0.946	1.020	1.090	1.131
45.0	0.940	0.996	1.078	1.094
42.5	0.939	0.958	1.043	1.049
40.0	0.932	0.944	0.924	0.902
37.5	0.895	0.935	0.870	0.840
35.0	0.854	0.910	0.834	0.804
32.5	0.833	0.840	0.778	0.748
30.0	0.788	0.824	0.758	0.734
27.5	0.754	0.799	0.744	0.724
25.0	0.740	0.750	0.730	0.710
22.5	0.733	0.736	0.709	0.640
20.0	0.730	0.718	0.660	0.610
17.5	0.701	0.680	0.597	0.610
15.0	0.498	0.540	0.576	0.610
12.5	0.420	0.445	0.530	0.495
10.0	0.382	0.388	0.464	0.374
7.5	0.319	0.381	0.373	0.353
5.0	0.304	0.338	0.348	0.338
2.5	0.294	0.307	0.322	0.297

Article j weekly and cumu. Forecasts (T=20)		
week m	F0	F0 [m:T]
1	225	2443
2	366	2218
3	221	1852
4	193	1631
5	187	1438
6	193	1251
7	199	1058
8	292	859
9	169	567
10	123	398
11	95	275
12	50	180
13	56	130
14	29	74
15	18	45
16	9	27
17	13	18
18	1	5
19	4	4
20	0	0

Figure 17. Inputs for Instantaneous Demand Distribution Calculation

Table 11 shows the results for the Cumulative Expected Demand values and Instantaneous Expected Demand for article j for percentiles 2.5 through 10. The highlighted values represents situations where for percentile P = -7.5, and week 1, value of Expected Instantaneous Demand for an article, our method returns a negative value of -66. One reason for this is because the 7.5th percentile for the Cumulative Error Distribution from week 1:T is lower than the corresponding value for week 2:T. Cumulative demand scenarios are assumed to decline at more or less the same slope as the each neighbouring scenario. However, this does not hold true for the actual realization of the cumulative demand scenario as seen in the graphical representation of ED[1:T] through ED[4:T] in Figure 18. In order to avoid negative instantaneous demand values due to this issue, we decided to replace the cumulative demand value of 845 with 779, which will result in 7.5th percentile of instantaneous demand to be zero; we artificially inflate the 7.5th percentile of demand for week 1. But this adjustment also reduces the weekly demand for the next week by the same amount, thus compensating for the overestimation with a slight delay.

Percentile	Cumulative Expected Demand for Article j				Expected Instantaneous Demand for Article j		
	ED [1:T]	ED [2:T]	ED [3:T]	ED [4:T]	EDj (1)	EDj (2)	EDj (3)
10.0	**933**	**861**	**859**	**610**	73	1	249
7.5	**779**	**845**	**691**	**576**	-66	154	115
5.0	**743**	**750**	**644**	**551**	-7	105	93
2.5	**718**	**681**	**596**	**484**	37	85	112

Table 11. Sample Instantaneous Demand Results

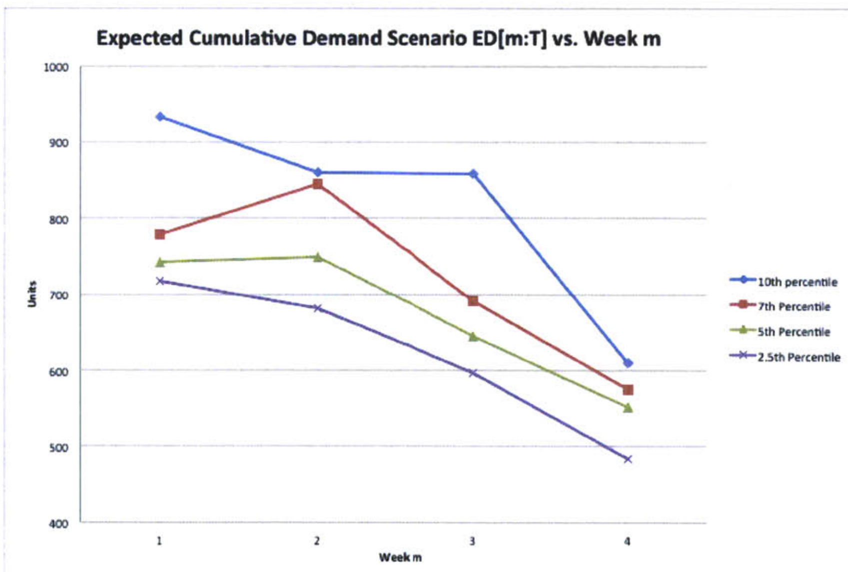


Figure 18. Demand Scenario - P=2.5th-10th, m= 1 - 4

6.2 Learning Curve

The learning curve is a linear function of time, more precisely update period. It is used to capture the forecast improvement or learning derived upon observing real sales of an article.

Specifically, it measures the relative decrease in spread of the error distribution when comparing forecast errors from two different update periods.

6.2.1 Spread of the Error Distribution

Spread of any distribution will be defined as the difference between the 97.5th percentile and 2.5th percentile values.

Spread of CD/F errors for period m:T based on forecast F_t, where t = 0,1,2,....,T, is given by:

Equation 18

$$S\left[\frac{D}{F_t}[\mathrm{m{:}\,T}]\right] = \left\{\frac{D}{F_t}[\mathrm{m{:}\,T}]\right\}^{97.5} - \left\{\frac{D}{F_t}[\mathrm{m{:}\,T}]\right\}^{2.5}$$

For each forecast, including original forecast F_0, we can calculate spreads from period [m:T], [m+1:T], [m+2:T], etc.

When running the optimization model at time t=0, we are interested in how the spread of the cumulative error distribution over period [t+1: T] from the latest F_0 compares to the same from F_t forecast. Using error information from past campaigns, we define this relationship as follows,

Equation 19

$$Spread\ Ratio, SR(t) = \frac{S\left[\frac{D}{F_0}[\mathrm{t{+}1{:}T}]\right]}{S\left[\frac{D}{F_t}[\mathrm{t{+}1{:}T}]\right]}$$

For example, the first comparison can be made between forecasts F_0 and F_1 over the period [2:T].

Equation 20

$$SR(1) = \frac{S\left[\frac{D}{F_0}[\mathrm{2{:}T}]\right]}{S\left[\frac{D}{F_1}[\mathrm{2{:}T}]\right]}$$

6.2.2 Information Sets

The learning curve, λ(t), is a step function of number of information sets versus update period t. For any t, if SR (t), is closer to 2 than 1, then we say that the number of information sets for update period t, λ(t) = 2. It means that the spread of error distribution has essentially halved compared to t=0.

To calculate the rest of information set array, we continue increasing t in Equation 20, until $SR\ (t)$ is closer to 4 rather than 2 (i.e. > 3) and for this t, the number of information sets increases to 4. Table 12 shows relation between λ(t) and SR (t).

SR (t)	λ(t) = 2^x
0 <= SR (t) < 1.5	1
1.5 <= SR (t) < 3	2
3 <= SR (t) < 6	4
6 <= SR (t) < 12	8
12 <= SR (t) < 24	16
24 <= SR (t) < 39	32
SR(t) >= 39	39

Table 12. Calculating Information Sets from Spread Ratios

Other constraints on the information sets required for the optimization model are as follows:

- λ(t+1) >= λ(t). In other words, once we split information sets, we do not combine them again.
- λ(1) = 1. Until the first week of sales is observed, all demand scenarios belong to the same infoset.
- λ(T) = Total number of scenarios i.e. at the end of the article life, we know everything about the demand evolution.

6.2.3 Learning Curve Results

Figure 19 shows CD/F results for the women's shirts and t-shirts articles calculated over periods 1:T, 2:T, ...6:T, where T is the last week in the forecasting horizon. The first histogram in Figure 20 is the CD/F ratio from pre-season forecast calculated over period 1:T. The rest of the histograms in Figure 20 show the CD/F results for the first five forecast updates (F_1 through F_5), calculated over periods t+1:T, where t is the update week. Note that we did observe some outliers but found that removing all outliers significantly reduced our already small sample size.

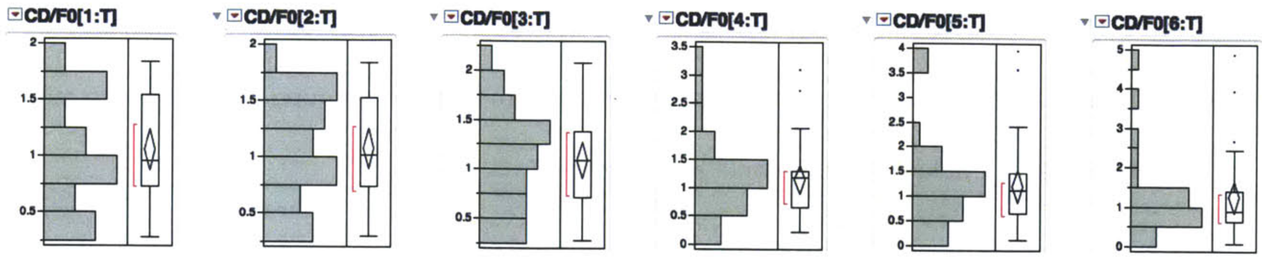


Figure 19. CD/F Results from Pre-Season Forecast (F_0)

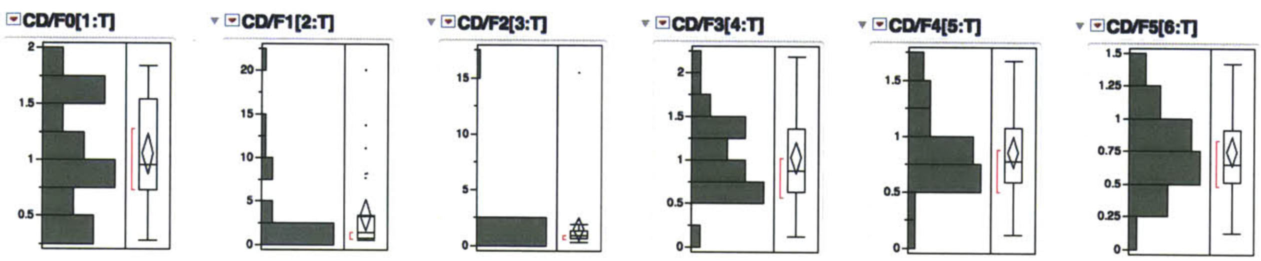


Figure 20. CD/F Results from Forecast Updates F_1 – F_5

Table 13 shows the spread ratio and infoset results. Based on these results, the 95% spread of the CD/F errors from F_1 and F_2 errors is significantly larger than that of the F_0 forecast. One explanation for this is there are many outliers in the updated forecast errors during this period. This is also the period when the demand of the article is highest and increasing, making it more volatile and susceptible to more outliers as seen in Figure 20.

T	0	1	2	3	4	5	6
S [D/F_0 [t+1:T]]	1.51	1.46	1.63	2.53	3.43	4.01	4.54
S [D/F_t [t+1:T]]	1.51	14.99	5.80	1.53	1.31	0.97	1.17
SR (t)	1.00	0.10	0.28	1.66	2.63	4.13	3.88
λ (t)	1	1	1	2	2	4	4

Table 13. Example of Infoset Derivation

Figure 21 shows a plot of resulting learning curve until t=12. For these group of articles, only five articles had more than 10 weeks of relevant[8] error data and therefore, there was no reduction in variability after week 10 for this group. We expect the results to improve over update period with more articles in the data sets yielding more sample errors for t>10.

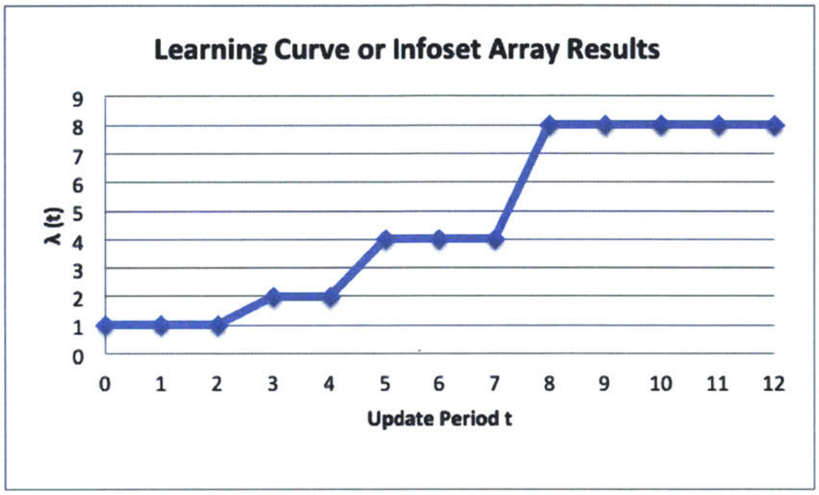


Figure 21. Learning Curve (Infoset Array) Results

The following operational procedure for the optimization model helps understand the usage of the information sets.

- The first run for the optimization model is at time t=0. At this time, we calculate the information set (infoset) array $\lambda(t)$, by comparing the **historical** F_0 forecast errors to those from an **historical** F_t forecast, for each t from 1 through T.
- The resulting information set array is the learning curve for optimization run at time t=0. The model makes optimized initial order and replenishment recommendations for a new article j, based on this initial learning curve and $F_{0,j}$ forecast for the new article. Only the orders prior to t', such that $\lambda(t')=2$, are placed. Replenishment orders are delayed in practice until t', when improved demand forecast is expected.

[8] Recall that we ignore some error data for ending weeks if the weekly demand or forecast was not significant compared to overall demand or forecast.

- The optimization model is **run again at t'**; t' is not necessarily the same as t=2. At this time, the infoset array is re-calculated by replacing F_0 by $F_{t'}$ and infoset array is re-estimated. Note that since we replaced F_0 by $F_{t'}$, the new infoset array will re-start from λ(t)=1. The new infoset array, the new article updated forecast $F_{t',j}$ and any left over inventory is plugged in to the optimization model. The model evaluates this information as well expected inventory from previous orders and makes updated purchase recommendations for the future periods. The new infoset also provides an updated t'.

7 Sensitivity Analysis

This chapter discusses the sensitivity analyses results on input parameters such as unit cost, lead time, spread of the distributional forecast and the cumulative percent error between the actual demand and estimated demand of an article. For each analysis, we changed only one parameter at a time and recorded the output of the optimization model. We present the results and related business implications, where appropriate, in the following sections.

7.1 Unit Cost

Figure 22 shows the results for the sensitivity analysis on the unit cost for an article. We used only two supply options: 1 and 2, with unit cost per articles C1 = €6.25 /unit and C2 = €8 /unit. Lead times for suppliers 1 and 2 are 10 and 6 weeks respectively. The average estimated demand for this article is 2934 units. The optimization results for these settings are recorded in the columns below Output.

Sensitivity Analysis - Unit Cost

Base Case C1 = 6.25 €/unit

Input				Output				Revenue	Costs		Profit
Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25		2717					16979		
2	6	8.00		397					3175		
			Total	3114	2799	134	314	44786	20154	5	24627

C1 = 6.50 €/unit

Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.50		2716					17655		
2	6	8.00		397					3175		
			Total	3113	2799	135	314	44783	20830	5	23948

C1 = 6.75 €/unit

Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.75		2686					18128		
2	6	8.00		402					3216		
			Total	3088	2789	145	299	44623	21344	5	23274

C1 = 7.00 €/unit

Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	7.00		2662					18632		
2	6	8.00		412					3296		
			Total	3074	2784	150	290	44536	21928	5	22603

Figure 22. Sensitivity Analysis - Unit Cost

Under the base case, we buy 2717 units from supplier 1 and 397 units from supplier 2 over the course of article life. The resulting net profit (revenue – purchase costs – holding costs) from this case is €24627. With all else remaining constant, if supplier 1 cost increases to 6.50 €/unit, the quantity ordered from supplier 1 reduces slightly to 2716 units. The resulting higher ordering costs decrease the profits to €23948. At C1 = €7.00 /unit, the quantity ordered from supplier 1 decreases significantly to 2662 units and quantity ordered from supplier 2 increases to 412 units. The net profit still decreases due to increase in purchasing costs and reduced sales. The decrease in profit from base case (C1 = €6.25 /unit) to the last (C1= € 7.00 /unit) is driven mostly by increase in the purchase costs (Δ = + €1774) rather than decrease in the revenue (Δ = - €250 euros). In comparison to these, the total inventory holding costs reduction (€1) is irrelevant.

The overall benefits resulting from supply cost reduction can be estimated before entering any negotiations using such sensitivity analysis. If the improvements are not significant, as in the case of C1 = €6.50 /unit, the buyers may find it more beneficial to negotiate other terms of contract such as lead time.

7.2 Lead Time

Figure 23 shows the sensitivity analysis results on supply lead time. We used the same article as for unit cost.

Sensitivity Analysis - Lead Time

Input | Output — Revenue | Costs | Profit

Base Case: L1 = 10 weeks

Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25		2717					16979		
2	6	8.00		397					3175		
			Total	3114	2799	134	314	44786	20154	5	24627

L1 = 9 weeks

Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	9	6.25		2804					17523		
2	6	8.00		311					2490		
			Total	3115	2800	134	315	44800	20014	5	24782

L1 = 8 weeks

Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	8	6.25		2909					18180		
2	6	8.00		208					1665		
			Total	3117	2801	132	316	44818	19846	5	24967

L1 = 7 weeks

Supply option	Lead Time (week)	Unit Cost (€/unit)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	7	6.25		3020					18875		
2	6	8.00		99					790		
			Total	3119	2802	131	317	44834	19665	5	25164

Figure 23. Sensitivity Analysis - Lead Time

In the base case, suppliers 1 and 2 have lead time of 10 weeks 6 weeks respectively. The results show total purchase of 2717 units from supplier 1 and 397 units from supplier 2 resulting in net profit of €24627. As lead time of supplier 1 decreases with all else remaining constant, supplier 1 becomes more attractive compared to supplier 2, so we order more from supplier 1 with decrease in L1. Note that changing the lead time has little effect on the total quantity purchased as the purchase costs for each supplier and expected demand remain the same in all cases. So the increase in the net profit is due to reduced costs of ordering more from supplier 1.

A change in lead time by 1 week has much higher effect on the change in order quantities than a decreasing the unit cost by €0.25. Such trade-off information is key to negotiating the right variable while entering in to supply negotiations.

7.3 Forecast Variability

Figure 24 shows the sensitivity analysis results on variability of the demand forecast. The base case distribution was created assuming that the cumulative D/F errors are uniformly distributed around the estimated demand with a standard deviation of 0.40. This high variability case results in high clearance inventory of 693 units and lost sales of 305 units. In the next case, the standard deviation reduces to 0.20, so the spread of cumulative demand distribution reduces to half of that in base case. The SR column represents the ratio of the spread of the demand distribution in the base case to the current case. When the standard deviation reduces to 0.20, SR = 2; the total quantity purchased reduces, the clearance inventory and the lost sales also reduce by about 50% of that in the base case. The profits increase to €24672. Note that this is same as the base cases in Sections 7.1 and 7.2.

When the spread ratio further increases to four, the clearance inventory and lost sales again reduce by about 50% and 64% respectively, but the sales do not increase by large percentage and therefore the profit increase (€1676 = €26303 - €24627) is not as big as that in the previous case. The increase in profit is even smaller for spread ratio of eight compared to previous cases.

Although the incremental benefit of reducing variability on the season profits is low, it has significant effect on amount of inventory left for clearance. Thus, forecast variability is a major determinant of amount of inventory left for clearance sales and should be reduced in order to reduce the clearance stock.

Sensitivity Analysis - Forecast Variability

Input | Output — Revenue | Costs | Profit

Base Case: σ = 0.40

Supply option	Lead Time (week)	Unit Cost (€/unit)	SR	Std Dev		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	1.00	0.40		2860					17875		
2	6	8.00				506					4046		
					Total	3366	2673	305	693	42763	21921	10	20832

σ = 0.20

Supply option	Lead Time (week)	Unit Cost (€/unit)	SR	Std Dev		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	2.00	0.20		2717					16979		
2	6	8.00				397					3175		
					Total	3114	2799	134	314	44786	20154	5	24627

σ = 0.10

Supply option	Lead Time (week)	Unit Cost (€/unit)	SR	Std Dev		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	4.00	0.10		2662					16638		
2	6	8.00				358					2865		
					Total	3020	2863	48	157	45809	19502	3	26303

σ = 0.05

Supply option	Lead Time (week)	Unit Cost (€/unit)	SR	Std Dev		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	8.00	0.05		2791					17445		
2	6	8.00				191					1531		
					Total	2983	2872	28	111	45952	18977	2	26974

Figure 24. Sensitivity Analysis - Forecast Variability

7.4 Forecast Accuracy

Figure 25 shows the sensitivity analysis results on the percent error of the estimated cumulative demand versus the observed cumulative demand. We chose percent error instead of the absolute percent error as the results for the negative and positive percent errors of same magnitude are significantly different as discussed below.

The base case represents the scenario where the actual demand is just 1% higher than the estimated demand. The base case has 523 units to be sold as clearance inventory. This means that even with the actual demand similar to the estimation, we will have about 17% (523/3114) more inventory than the actual demand. This extra inventory is needed to maintain a safety stock for high forecast variability which is discussed later.

Sensitivity Analysis - Cumulative Percent Error

Input | Output (Revenue, Costs, Profit)

Base Case: PE = 1%

Supply option	Lead Time (week)	Unit Cost (€/unit)	Pecent Error (%)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	1		2785					17405		
2	6	8.00			694					5555		
				Total	3479	2956	0	523	47295	22961	0	24335

PE = -10%

Supply option	Lead Time (week)	Unit Cost (€/unit)	Pecent Error (%)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	-10		2648					16552		
2	6	8.00			0					0		
				Total	2648	2648	0	0	42367	16552	0	25814

PE = -20%

Supply option	Lead Time (week)	Unit Cost (€/unit)	Pecent Error (%)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	-20		2648					16552		
2	6	8.00			0					0		
				Total	2648	2448	0	200	39174	16552	0	22621

PE = 10%

Supply option	Lead Time (week)	Unit Cost (€/unit)	Pecent Error (%)		Quantity Purchased (units)	Sales (units)	Lost Sales (units)	Clearance Inventory (units)	Revenue (€)	Purchase Costs (€)	Holding Costs (€)	Profits (€)
1	10	6.25	10		2785					17405		
2	6	8.00			694					5555		
				Total	3479	3267	0	212	52272	22961	0	29312

Figure 25. Sensitivity Analysis - Forecast Accuracy

In the case PE = -10%, where the actual demand is 10% lower than the estimated demand, the model identifies the low demand scenario and therefore there is no replenishment order from supplier 2. As a result, the sales and profits are lower than the base case. Following the same reason, in the case of PE = -20%, the order quantity and thereby the profits are lower due to lower actual demand.

When PE = 10%, the actual demand is 10% higher than the estimated demand. Due to safety stock, we are able to satisfy this demand and sell more than the base case. As a result, the net profit for PE = 10% case are in fact higher than those for the base case. The sales for PE = 20% are higher than PE = 10% case, resulting in even higher profits. However, there is also increase in the lost sales due to insufficient inventory.

While it may be desirable to sell more in high-demand scenarios, higher than expected demand for one article could also cannibalize sales of other similar products planned during the same selling period and stores, potentially reducing overall profits for the two articles. As a result, positive percent error, though profitable for one article, is not desirable for group of similar articles sharing the same demand pool.

8 Conclusions and Recommendations for Future Work

We developed and implemented a modular forecasting and optimization based system to support sourcing decisions at Zara. Our tests on forecasting group articles show that the learning increased after third week of the article introduction. This means that the CD/F errors are 50% less variable for updated forecast F_3 compared to those from pre-season forecast F_0, which is a significant improvement. Using this narrower distribution to estimate demand would result in significantly less variability in the demand scenarios and therefore reduce safety stock and lost sales through more accurate purchase recommendations. Reduction in safety stock inventory and lost sales translates directly to higher profits. The earlier such improvement in the forecast can be made, the more room there is to order replenishments from the cheaper supply options, which can further add to profit by reducing purchase costs. Our study was based on limited sample size, with about 30 articles in the forecasting group. It is recommended to do a "virtual" test of the entire system on a larger group of articles for determining the learning curve information. We expect that more data will result in better estimation of the empirical error distribution and the learning curve.

For measuring historical forecast errors, we grouped articles by subfamilies (for example women's t-shirts). However, articles in the same subfamily may exhibit very different demand patterns. For example, casual and business shirts for women fall under the same subfamily but the former are considered 'basic' articles with more predictable demand whereas the latter are considered as 'fashion' articles with less predictable and typically shorter demand curve. It is recommended to study such differences or commonality in demand behavior amongst articles and re-categorize them accordingly.

The optimization model currently calculates the inventory requirements on the basis of the demand, supply constraints, cost and pricing information. It can be used to set constraints on amount of inventory left for clearance sales. If the clearance sales quantity is set to a non-zero quantity, the model will recommend purchasing certain quantity to satisfy clearance sales requirement. After discussion with Zara's management, it was decided to optimize profits for main selling season, so the clearance sales quantity was set to zero. The business case for limiting the markdown period inventory below the profit-maximizing amount is to limit clearance inventory from flooding the stores at the cost of sales of newer items. Also an estimation of minimum display stock can be used to more accurately predict initial shipment requirements. This estimation can be added as a fixed quantity to the optimization output. Given that all inputs to the model are prepared beforehand, the model takes only about 15 seconds to provide explicit sourcing decisions for each ordering period, which is a significant improvement over current ordering methods, that do not employ dynamic optimization across different supply options. However, Zara1 methods will require significant amount work from the IT department for generating forecasting and optimization results and maintaining the database, thus increasing the resource requirement compared to the current ordering methods.

A direct comparison of forecasting accuracy and optimization outcomes with current method was not possible as there was no benchmark forecasting data available to compare against. We recommended performing a virtual test with a larger group of articles as well as a live pilot test to determine if the Zara1 system results in better forecasting and purchase outcomes compared to existing methods. For the pilot test, the buyers will provide comparable information as well as conduct benchmark testing on one set of articles (control group). Another similar set of articles (test group) will be tested using Zara1 methods. The two methods will be compared on metrics such as overall forecast accuracy and profits for the main campaign. The results of the pilot will help determine the feasibility and readiness of the project for implementation on a larger scale.

References

[1] "FY2011 Results," inditex.com, para. 1, [Online]. Available: http://inditex.com/en/shareholders_and_investors/investor_relations/quarterly_results. [Accessed: May 4, 2012]

[2] "H&M Sees Profits Dip, Market-Share Gains", online.wsj.com, para. 3, [Online]. Available: http://online.wsj.com/article/SB10001424052970204573704577184191578087190.html. [Accessed: May 4, 2012]

[3] "Our Group," inditex.com, para. 1, [Online]. Available: http://www.inditex.com/en/who_we_are/our_group. [Accessed: March 10, 2012]

[4] "Timeline," inditex.com, [Online]. Available: http://www.inditex.com/en/who_we_are/timeline. [Accessed: March 10, 2012]

[5] K. Ferdows, M. A. Lewis, and J. A. D. Machuca, "Rapid-Fire Fulfillment," Harvard Business Review, vol. 82, no. 11, pp. 104-110, Nov. 2004.

[6] T. M. Bonnefoi, "Demand Forecast for Short Life Cycle Products – Zara Case Study," LGO Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2012.

[7] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," International Journal of Forecasting, vol. 8, no. 1, pp. 69-80, June 1992.

[8] B. E. Flores, "A pragmatic view of accuracy measurement in forecasting", Omega, vol. 14, no. 2, pp. 93-98, 1986.

[9] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp. 679-688, 2006.

[10]J. W. Taylor, L. M. de Menezes, P. E. McSharry, "A comparison of univariate methods for forecasting electricity demand, up to a day ahead," International Journal of Forecasting, vol. 22, no. 1, pp. 1-16, Jan. – Mar. 2006.

[11]E. S. Gardner Jr., "Exponential smoothing: The State of the Art," Journal of Forecasting, vol. 4., no. 1, pp. 1-28, Jan. 1985.

[12]E. S. Gardner Jr., "Exponential Smoothing: The state of the art - Part II," International Journal of Forecasting, vol. 22, no. 1, pp. 637-666, Oct. 2006.

[13] S. Makridakis, S. C. Wheelwright, and R. J. *Hyndman, Forecasting: Methods and Applications*, 3rd ed., New Jersey: John Wiley & Sons, 1998, pp. 161-169.

[14] C. Chatfield, "Calculating Interval Forecasts," Journal of Business and Economic Statistics, vol. 11, no. 2, pp. 121-135, Apr. 1993.

[15] J. W. Taylor and D. W. Bunn, "A Quantile Regression Approach to Generating Prediction Intervals," Management Science, vol. 45, no. 2, pp. 225 – 237, Feb. 1999.

[16] E. S. Gardner Jr., "A Simple Method of Computing Prediction Intervals for Time Series Forecasts," Management Science, vol. 34, no. 4, pp. 541-546, Apr. 1988.

[17] W. H. Williams and M. L. Goodman, "A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts," Journal of the American Statistical Association, vol. 66, no. 336, pp. 752-754, Dec. 1971.

[18] E. A. Silver, David, F. Pyke, and R. Peterson, *Inventory Management and Production Planning and Scheduling*, 3rd ed., New Jersey: John Wiley & Sons, 1998, pp. 150-188.

[19] A. Mersereau and J. Gallien, "Zara Purchasing Optimization: Optimization Formulation," Nov. 2011.

Appendices

Appendix A: Detailed Process Map for Zara1 Project

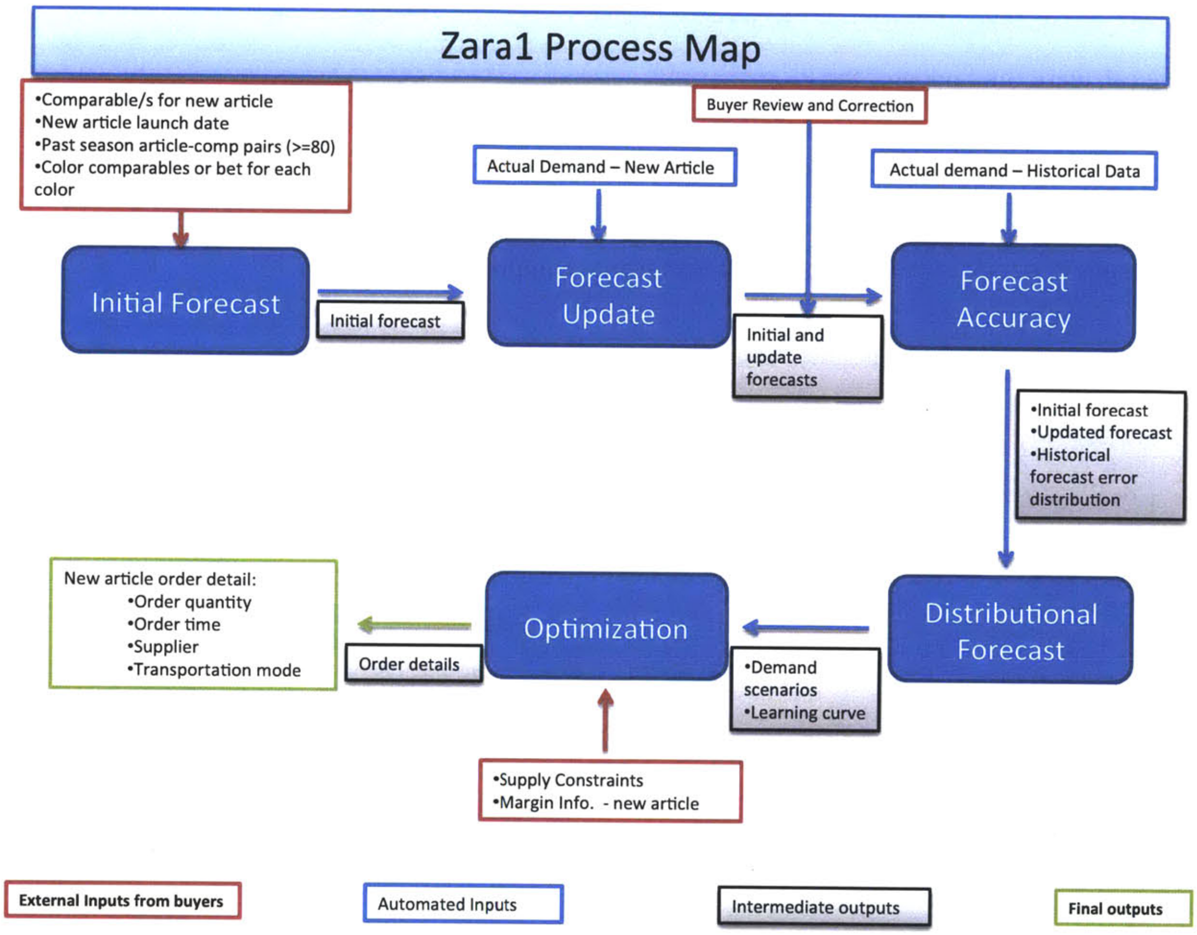


Figure 26. Detailed Process Map for Zara1 Project

Appendix B: Percentile Function

Suppose we want to find a P^th^ percentile from an array of values a_1, a_2, a_3, … , a_N arranged in an ascending order.

First we calculate the rank n of the P^th^ percentile.

$$n = \frac{P}{100}(N-1) + 1$$

n is then split into its integer component i and decimal component d such that

$$n = k + d$$

Then the value of the P^th^ percentile is calculated as

$$a^P = \begin{cases} a_1 & for\ n = 1 \\ a_N & for\ n = N \\ a_k + d * (a_{k+1} - a_k) & for\ 1 < n < N \end{cases}$$