Demand Forecasting at Zara: A Look at Seasonality, Product Lifecycle and Cannibalization

by

Jose M Garcia

B.S. Industrial and Systems Engineering,
Georgia Institute of Technology, 2008

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

Master of Science in Engineering Systems
and
Master of Business Administration

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

June 2014

Signature redacted

Signature of Author ______________________
MIT Engineering Systems Division, MIT Sloan School of Management
May 9, 2014

Signature redacted

Certified by ______________________
Josef Oehmen, Thesis Supervisor
Associate Professor, Department of Management Engineering
Technical University of Denmark

Signature redacted

Read by ______________________
Warren Seering, Thesis Reader
Weber-Shaughness Professor of Mechanical Engineering and Engineering Systems
MIT School of Engineering

Signature redacted

Certified by ______________________
Georgia Perakis, Thesis Supervisor
William F. Pounds Professor of Operations Research and Operations Management
MIT Sloan School of Management

Signature redacted

Accepted by ______________________
Richard C. Larson
Mitsui Professor of Engineering Systems
Chair, Engineering Systems Division Education Committee

Signature redacted

Accepted by ______________________
Maura Herson, Director of MIT Sloan MBA Program
MIT Sloan School of Management

This page intentionally left blank

Demand Forecasting at Zara: A Look at Seasonality, Product Lifecycle and Cannibalization

by

Jose M Garcia

Submitted to the MIT Engineering Systems Division and the MIT Sloan School of Management on May 9, 2014 in Partial Fulfillment of the Requirements for the Degrees of Master of Science in Engineering Systems and Master of Business Administration.

Abstract

Zara introduces 10,000 new designs every year and distributes 5.2 million clothing articles per week to a network of over 1925 stores in more than 86 countries. Their high product mix and vast global network makes demand forecasting for Zara a challenging endeavor. This thesis sets out to incorporate the effects from seasonality, product lifecycle, and cannibalization into a long term aggregate demand forecast and a short term SKU replenishment forecast.

For seasonality, there are two categories of events that are explored in detail: 1) Macro patterns, which are the year to year sales patterns that remain fairly consistent, such as rising sales in spring; and, 2) Specific Events, which refers to events that have an impact on demand but shift dates from one year to the next, such as Easter or Ramadan. These two factors are used to forecast short and long term aggregated store demand by using regression that leverages historical demand with dummy variables for specific events.

Product lifecycle and cannibalization are incorporated in the SKU demand forecast. Products at Zara experience a majority of their sales in the first few weeks in the store. For this reason, when forecasting demand for replenishment purposes, it is of paramount importance to understand: 1) How long the item has been in a store; and, 2) how many new items are being displayed for the first time at the store on the week in question. This thesis details a methodology that successfully uses regression to incorporate both of those components.

In addition to detailing the methods for forecasting demand this thesis also covers: an overview of the current forecasting methodology and the special characteristics of Zara's demand; a results section which detail reductions in forecast error from 21% to 17%. This has the potential to reduce lost sales by 24%; lastly, it details implementation efforts at Zara.

Thesis Supervisor: Josef Oehmen
Associate Professor, Department of Management Engineering, Technical University of Denmark

Thesis Reader: Warren Seering
Weber-Shaughness Professor of Mechanical Engineering and Engineering Systems, MIT

Thesis Supervisor: Georgia Perakis
William F. Pounds Professor of Operations Research and Operations Management, MIT

This page intentionally left blank

Acknowledgements

The opportunity to work on this thesis was made possible by the committed partnership that Zara and MIT LGO have developed since 2006. First and foremost I would like to thank Zara's distribution department for giving me the opportunity to work on a problem that is so central to their daily operations. From the first day, I received the full support from the distribution and information systems personnel assigned to the project and was given the liberty and luxury to focus solely on the project. The opportunity to spend 6 months working at one of the world's most revered operations companies in the world while immersing myself in the rich Galician culture is an experience I will cherish for the rest of my life.

I am forever grateful to my supervisor and friend Martín Nóvoa Vidal who carefully explained the intricacies of Zara's distribution process with the aid of countless hand-drawn, and spectacular, pictures. I would also like to praise the attention and effort that Alberte Dapena gave to the project as he provided every bit of data I requested and reorganized it as many times as it was necessary in order for me to be able to process it on my work station. Special thanks are also warranted to Jose Manuel Corredoira Corras for his ingenuity in finding data sources and for his programming prowess to help me automate my programs. Endless thanks to Ivan Escudero Rial for championing my project and providing every resource I needed throughout the project and pushing me to develop the best model possible. Many thanks to Miguel Astor for his insights on which calendar events had significant impact on demand. And, for graciously explaining the many aspects of the local culture that were foreign to me and earnest efforts in ensuring I felt comfortable in my new environment. It is an understatement to say that everyone I met at Zara welcomed me as another one of their colleagues, especially Begoña Valle Vilela, Sergio Marquez, Eva Ruano, and Gonzaga de Bastida.

This project would not have been possible without the support and infrastructure provided by the LGO Program. I thank Dr. Georgia Perakis for meeting with me on a weekly basis throughout the project to provide guidance and encouragement throughout the emotional rollercoaster that is an LGO internship. I also thank Dr. Josef Oehmen for also dedicating countless hours to the project and for teaching me the value of stakeholder management and the immeasurable value in setting expectations early and often. I also thank all of the previous LGO interns at Zara for leaving such a strong legacy that has sustained the partnership and made it very easy for me to establish credibility.

I would like to thank my family and friends for their support and encouragement during this time. In particular, I thank my wonderful wife for supporting my decision to pursue this opportunity and for pushing me every day to reach for my dreams. And finally, I would like to dedicate this thesis to my late grandfather Silvio Fornell who passed away while I was working on this project. From my earliest memories as a child to the last time I saw him, he made sure to remind me of how limitless and successful he thought my future would be.

Table of Contents

List of Figures

List of Tables

List of Equations

Glossary & Subscript Convention

<u>Key Terms (Subscript)</u>

Cycle (c): Time unit referring to either M-W or Th-Su. Zara's distribution decisions are made at the beginning of each cycle

Week (w): Time unit encompassing two consecutive cycles.

MCCT: Refers to the SKU level identifier. M references the model, first C references the fabric or supplier, second C references color, and T references size

MCC (m): MCCTs aggregated without consideration to size. Most granular unit used for modeling

Subfamily (s): High level of product aggregation that is normally broken out by design team such as Basic Pants or TRF T-Shirts

Family (f): Highest level of product aggregation irrespective of design team such as Pants or Coats

Store (t): Store, or tienda in Spanish, is used in this study to group all products that fall under women clothing.

1 Zara Background

Zara is the world's largest fashion retailer (Hansen, 2012) and is the flagship brand of parent company Industria de Diseño Textil (Inditex) representing roughly 66% of the group's total sales (Inditex Annual Report, 2012). In addition to leading the group in sales it also serves as the group's hub for operational innovations. Most projects to improve distribution efficiency or in-store operations are first developed at Zara's headquarters in Arteixo, Spain, then tested and validated through Zara's store network and finally implemented across the other brands. This is a large reason why most published work associated with Inditex refers to improvements implemented at Zara.

The rest of this chapter will attempt to explain the key characteristics of Zara's business operations as they pertain to this study: long and short term demand forecasting. For more details on its rich history as it has evolved from its humble beginnings to a global brand the author recommends reading Hansen's 2012 article in The New York Times (Hansen, 2012).

1.1 Fast Fashion

Fast fashion retailers are characterized by their ability to respond quickly to market trends. Some of the better known players in this space include Zara and H&M but it is important to recognize that fast and traditional fashion models are ends on a spectrum. Many retailers that are hesitant to make the large investments required to be at the forefront of fast fashion still take steps be able to react to their customers, and hence fall somewhere in the middle of the spectrum (Fast Fashion and the Responsive Supply Chain, 2013). The strategies to achieve fast responses to customer demands vary significantly by company, but they generally combine at least two components (Cachon & Swinney, 2011):

- Short production and distribution lead times, enabling a close matching of supply with uncertain demand
- Highly fashionable, or trendy, product designs

Following this strategy, as Zara and its competitors do, may appear at first glance to significantly increase operational costs but the benefits from better matching supply and demand usually outweigh the costs of maintaining an agile supply chain (see, e.g., Fisher and Raman 1996, Eppen and Iyer 1997, Caro and Gallien 2010. One obvious benefit is that it helps reduce the instances where large volumes of inventory are sold at significant discounts in order to clear inventory of unpopular models. For example, Ghemawat and Nueno (2003) were able to show that fast fashion retailers such as Zara could have as low as 15-20% of sales come from discounted items while traditional retailers have rates in the 30%-40% range.

An additional benefit identified by Cachon and Swinney (2009) comes with the presence of strategic customers. Strategic customers are those who not only decide whether to buy an item but also when to buy it given mark-down expectation. Cachon and Swinney found that a quick response strategy provides an average of 67% additional benefit when there are strategic customers. The share of strategic customers continues to grow as more prices and availability information are easily accessible online. At Zara, for example, popular items rarely go on sale because they sell out before the mark down period begins, which incentivizes strategic buyers to purchase at full price.

1.2 Zara's Scale

As the world's largest fashion retailer Zara's operations are impressive by both its size and agility. With 1,925 stores in 86 countries, Zara amassed €10.5B in revenue in 2012, which represented an 18% increase over 2011 (Inditex Annual Report, 2012). Zara operates this network in a very centralized manner: it owns almost all its stores, it develops most of its software and systems in-house and most strategic decisions are made at headquarters. Operating in this manner gives them tight control over their supply chain (Ferdows, Lewis, & Machuca, 2004) but it also significantly increases the breadth of responsibilities managed at headquarters. For example, Zara introduces roughly 10,000 new articles throughout their store network every year (Kelley, 2013). This high product variety coupled with unparalleled twice weekly replenishments results in the distribution department having to ship 5.2 million units per week and to make over 1MM replenishment decisions per week. Given their scale and need for agility, any analytical tool developed must be fast and easy to understand so that it can be run and evaluated quickly.

1.3 Business Cycle and Sourcing

1.3.1 Two Selling Seasons

Similar to most clothing retailers, Zara plans two major seasons per year: Summer and Winter. The summer season runs from February through June and winter season runs from September to December. At the end of each season they offer mark downs in order to exhaust remaining inventory; these are the only periods when they offer discounted merchandize. Zara also has cross-over collections which serve as previews for the new season and usually start right before the mark down periods and run through the beginning of the new season. For example, summer cross-over collection would likely be introduced in mid-December even though the summer selling season has not begun. The purpose for these cross-over collections is to have season-appropriate merchandize before fully switching over. For example, cross over merchandize introduced in mid-December is not likely to include heavy volume of shorts.

1.3.2 Four Tier Sourcing

Zara is different than many of its competitors in that it still manufactures some of its products in order to be able to respond to customer trends in season. However, it still outsources most of its manufacturing and strategically balances long lead time/low cost sourcing with short lead time/high cost sourcing. Zara garnered attention when it began its rapid expansion in the early 2000s for having a large share of its supply base located in Europe, as opposed to Asia where it was cheaper. At the time Zara believed that it made more sense to produce closer, and not cheaper, because it allowed them to have a more efficient supply chain (Tokatli, 2008). It has since shifted to having more suppliers outside of Europe as the agility and quality of non-European suppliers has improved. Table 1.1 shows the distribution of suppliers by region which shows a wide supplier base in both Europe and Asia (Inditex Annual Report, 2012).

Table 1.1 - Supplier Distribution by Region

Region	Suppliers in 2012
Asia	672
European Union	446
Non-EU Europe	136
Africa	112
America	68
Total	1434

Zara's sourcing operations fall into four major categories which are described below, however, for more detail the author suggests a review of Bonnefoi's work on this topic (Bonnefoi, 2012):

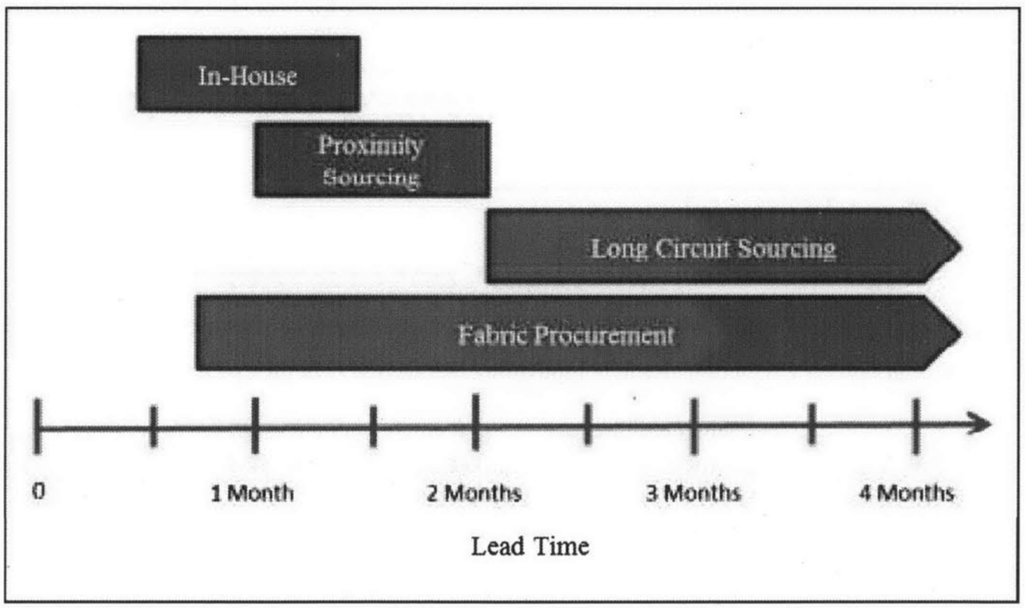
Long Circuit Sourcing: These are long lead time suppliers to whom Zara might provide a design to manufacture. This sourcing channel is used for high volume, low fashion articles such as cardigans or solid color sweaters. This is the lowest cost channel for Zara, but offers the lowest amount of control and flexibility as these suppliers are mostly in Asia (China, India and Vietnam) and the lead time can be as high as 12 weeks from order to delivery.

Proximity Sourcing: Suppliers with strong and long standing relationships with Zara which are usually located in nearby markets such as Portugal, Morocco, Turkey, or even Spain. This channel is more costly than Long Circuit but it is arguably Zara's most important channel because it offers fast lead times of just 4 to 6 weeks while still offering relatively low prices. Zara is able to minimize the cost and lead time for this channel by directly procuring the fabrics and producing the designs for the items it orders. This generates savings to the manufacturer which can be directly passed on to Zara. Maintaining close relationships with these manufacturers is important to ensure quality and fast turnarounds whenever Zara detects a growing fashion trend.

In-House Manufacturing: The share of this channel has continued to decrease in an effort to maintain high margins. However, it is still a very important sourcing channel because it allows Zara to design and introduce new products mid- season. In the event that Zara misses a major trend within a collection, it can use this channel to fill the void. For this channel, Zara produces the design and cuts the fabric before outsourcing to a network of nearly 400 small firms located near its headquarters. This channel is also used for their most complex designs which require a high level of skill.

Fabric Procurement: Zara also engages in large volume purchases of fabrics in order to internalize savings due to economies of scale. Throughout the year, Zara identifies key fabrics that they believe could be at risk of either going up in price or will be in low supply in the short to medium term. This inventory of fabric allows Zara to reduce the lead times and prices quoted by its suppliers.

Figure 1.1 - Sourcing Channel vs. Lead Time



1.4 Organizational Structure

While the company as a whole is very centralized because most decisions are made at headquarters, the office is divided into a series of pseudo-independent departments. It has separate departments for Women, which is by far the largest, Men and for Kids. Each of these departments has its own functional support and design teams. Women's sales account for more than 66% of Zara's sales and as such it is where most improvement projects begin. For this reason, this study focuses almost entirely on the Women department. The women's department has over 300 employees responsible for the design and procurement of its sub-brands (Woman, Basic and TRF), distribution, and financial performance of the group (Garro, 2011).

A couple of important functional teams within each group that help Zara sense the pulse of its customers while also controlling the look and feel of its stores across the world are: 1) Country Managers, or "Comerciales de Paises"; and, 2) Merchandizers.

Country Managers: Each member of this team analyzes and interprets sales figures for roughly 40 stores in a geographic area. Country managers are the main link between stores and headquarters. The goal of every country manager is to help each store sell as much as possible and, in fact have part of their compensation tied to the performance of the stores they oversee. Stores report high selling items and slow selling items to the store managers who can then communicate the feedback to both the designers and the distribution department (Corsi, Dessain, & Ton, 2010). This feedback loop is very important for Zara to be able to adjust its merchandize to adapt to the customer demand at each store.

Interior Designers: This is a small team which is in charge of designing the interior layouts for all stores. While each store has some autonomy as to which products it has on display, this team sends pictures taken at a Test Store detailing the preferred layout arrangement for the current collection (Corsi, Dessain, & Ton, 2010).

1.5 Product Hierarchy

The SKU-level identifier for tacking clothing articles at Zara is called "MCCT", where M refers to the Model, the first C refers to Calidad (Quality) which is used to differentiate identical articles made by multiple suppliers or have different fabric composition, the second C is for Color, and T stands for Talla (size). For this project, however, the lowest level of granularity forecasted was MCC. Nonetheless, it is still important to understand the distinction between MCC and MCCT level as the raw data is tracked at the MCCT level before it is aggregated for modeling purposes.

In addition, each MCC is also associated with a product family, product subfamily, and a set of attributes.

Product Family: Family is the widest level of aggregation and usually refers to categories such as pants or long coats, among others. This level of aggregation is mostly an IT classification and it is seldom used by Zara employees when discussing different products. We nonetheless found it to be useful for modeling purposes.

Product Subfamily: Subfamilies attempt to be one level deeper of granularity by usually combining the designer house for the MCC and a descriptive term such as B.VESTIDO (for Basic Dress) or W.FALDA (for Woman Skirt). This is the most common aggregation used by Zara, but can be difficult to predict sales volume from season to season because a particular designer could choose to exclude an entire subfamily from a collection or make it the feature item.

Attributes: This description is meant to associate each MCC with a set of attributes that can help describe the item. This value could be very valuable for grouping similar MCCs across families and subfamilies; however, it is not a standardized or mandatory designation. Every designer can choose freely which attributes to assign to an MCC and more disappointingly, many designers choose not to use this classification tool; making it a poor data point for modeling purposes.

2 Problem Introduction

2.1 Problem Motivation

The focus of this project is to develop an automated methodology to account for seasonality when forecasting demand. The ability to properly forecast demand is particularly important for Zara because one of its core strategies is to keep low inventory at stores and replenish them twice weekly based on demand expectations. Zara's presence in over 86 countries complicates seasonality forecasting because each store is impacted by global trends, such as winter and summer, and also by local trends, such as Black Friday in the U.S. or Ramadan in Saudi Arabia. Zara is interested in understanding seasonality for two major purposes:

1. Forecast aggregated Store-level demand with 4-6 month time horizon
2. Forecast item-level demand for twice weekly replenishments

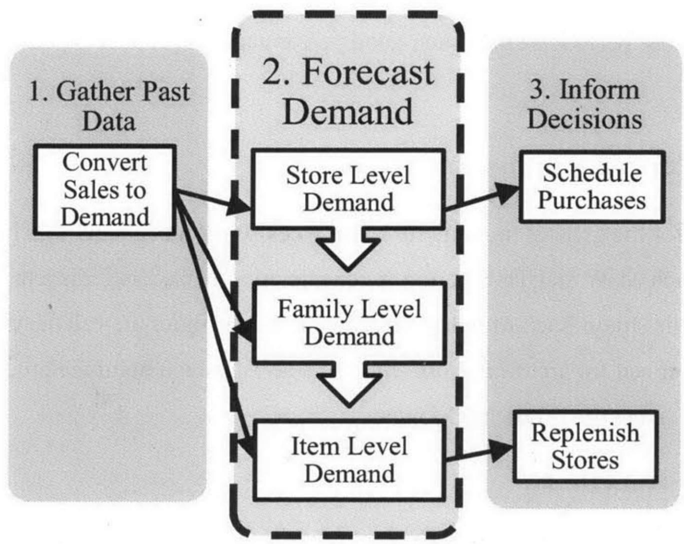
The first portion of this study focuses on forecasting store level demand, which refers to the total number of units, or MCCTs, a store is expected to sell for a given time period. Understanding peaks and troughs in demand with 4 to 6 months of anticipation for each store allows Zara to understand aggregated global demand. This knowledge would allow them to align supplier deliveries with customer demand. Currently, Zara relies mostly on prior year performance and historical demand patterns based mostly on Spain's fashion demand cycle. This study is one of the first attempts to understand demand on a global level in order to use in procurement planning. Unfortunately, Zara does not have a metric to quantify how well it matches its long term procurement orders with demand, however their spectacular growth over the past decade serves as evidence that accuracy is not its biggest problem. Instead, their major pain point is the fact that the current process is highly subjective and depends heavily on the expertise of its managers, this project attempts to deliver an objective approach that incorporates much of the knowledge that Zara managers have regarding factors that impact demand trends.

The second portion of this study focuses on forecasting short term demand quantities for every MCC at every store. Answering this question helps them determine the optimal in-store stock levels for every MCC, while taking into consideration cannibalization from other products and the length of time since product was introduced. Zara alleviates a lot of the perils that come from inaccurate forecasting by replenishing twice per week and relying on observed demand as the main input. To improve upon their methods, this study looked at input variables in addition to observed demand to capture effects from seasonality and cannibalization, among others.

2.2 MIT9 Project

There are three major areas of Zara's operations that influenced this project. Figure 2.1 shows the three areas and highlights the focus of this project. The first area is accurately translating point of sale data to demand data. This translation is necessary because whenever an item stocks out, it leads to the point of sales data being censored. In theory, more units would have been sold if the item had not stocked out. Garro (2011) and Bonnefoi (2012) looked at this problem and developed good methodologies for converting point of sales data to demand data. The method designed by Garro had been implemented at Zara by the time this study began and was therefore used as the estimate for historical demand.

Figure 2.1 - Distribution Decision Sequence and Project Focus



The second area can be thought of as the actual modeling which tells Zara how many units they can expect to sell in a given time period. This is the main focus of this study and will be discussed in depth in Chapters 4 and 5.

The third phase is the step where operational decisions are made based on the demand forecasts. This is mostly a question of service level, which is to ask, to what extent does Zara want to avoid stockouts? When considering overage and underage costs, Zara has different approaches for procurement and for distribution. In procurement, Zara seeks to reduce overage costs by purchasing low quantities of each item they introduce, which leads to frequent stock outs of popular items. Their replenishment policy, however, is very aggressive in avoiding stockouts. This means that they attempt to avoid any stock out at stores while there is inventory available in their warehouses. This last decision-making phase of their operations falls outside the scope of this study.

2.3 Current Processes

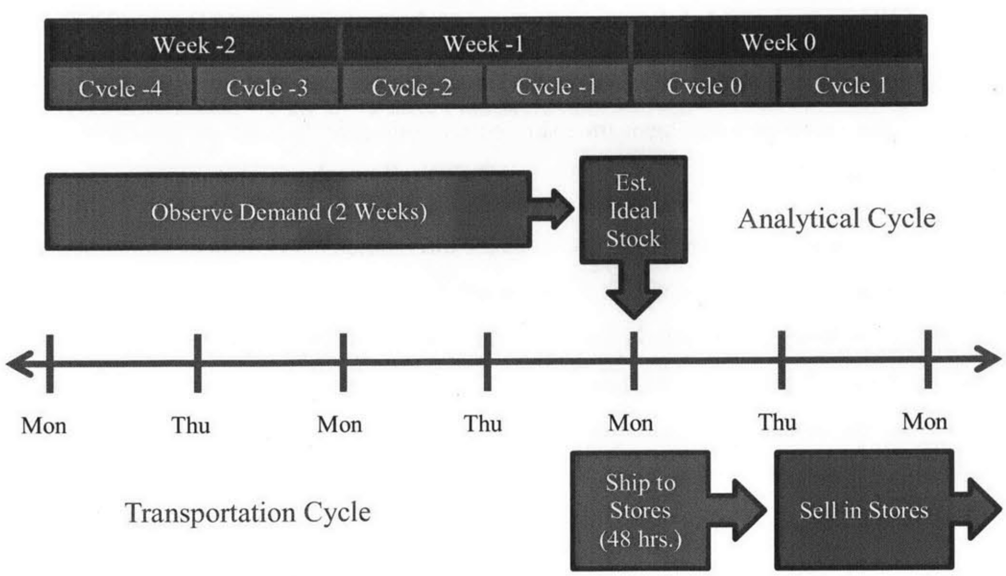
2.3.1 Long Term Purchase Planning

For long term planning Zara focuses on historical performance with particular attention on prior year performance. Zara has historically prepared their purchases so that the bulk of their merchandize is displayed at stores in line with Easter for Summer Collection and near the end of September for Winter Collection. This pattern stems from the consumer behavior in Spain and in most of Europe that experiences similar climate and where Easter is a major holiday. In other areas of the world, this pattern has led to a self-fulfilling prophecy which is difficult to isolate. Zara's decision to introduce its best merchandize at certain times of the year leads to those times being the best performing times of the year. Would those periods display such good performance if product introduction were planned differently?

2.3.2 Short Term Replenishment

Figure 2.2 illustrates the general timeline for twice weekly replenishments which occur on Mondays and Thursdays. Each M-W and Th-S period is considered a "cycle" and it is a unit of time frequently used throughout this study. Each Monday, replenishment quantities are calculated for each MCC in each store to be shipped for arrival at stores by Thursday, at which point the process is repeated for replenishments to arrive the following Monday.

Figure 2.2 - Replenishment Timeline



To determine the replenishment quantity for each store, Zara first calculates an average daily demand rate, which usually estimated as the daily demand from previous two weeks (= 4 cycles). Next, it sets a target for days of inventory which it normally sets to 21 in order to reduce the probability of stock outs. The product of these two quantities determines what they consider to be their "Objective Stock" or optimal in-store inventory level. They then ship the difference between the optimal inventory level and the store's current inventory level. At this point, it should be clear that the optimal inventory level is dynamic and fluctuates as demand for a given item goes up or down.

Equation 2.1 - Replenishment Quantity General Approach

$$Replenish_{mcc} \approx \left(\boldsymbol{Avg.Daily\ Demand_{mcc}^{past\ 2\ wks}} \times Days\ of\ Inv\right) - InStore\ Inv_{mcc}$$

There are, however, several instances when using the observed average daily demand is either not feasible or not accurate. One such instance is when an item is first introduced because there is no history available. Another instance, and a focus of this study, is when there are exogenous factors affecting demand for an item, such as a holiday. Below are the two main adjustments that Zara makes:

1. Calculate Daily Demand for New Items:
   When introducing new items Zara uses comparable items to estimate initial demand. Zara's buyers, designers, and country managers identify comparables based on fashion similarities. This process can be highly subjective. Nonetheless, it is a quick and effective way to manage this process given the large number of new items that Zara introduces each year. Once comparables are identified, Zara identifies the top three selling weeks for each comparable item and uses the average daily demand over those three weeks as the average daily demand for the new item. One significant shortcoming of this method is that it does not account for seasonal factors for neither the comparables nor the item being introduced. For example, if a comparable's best weeks occurred over Christmas (high demand) and a new item is being introduced in early February (low demand) then there is a clear mismatch in sales expectations for the two items, which is independent of the quality of the items[1]. Here again, the target days of inventory is also set to 21 days.

   Equation 2.2 - Initial Shipment Quantity General Approach

   $$Initial\ Shipment_{mcc} \approx Avg.Daily\ Demand_{comps}^{Top\ 3\ Wks} \times Days\ of\ Inv$$

2. Adjustments to Daily Demand for Seasonal Trends:
   Zara currently relies on country managers to alert the distribution department of upcoming holidays and events that could impact demand. Upon this notification the distribution department evaluates the change in demand from prior years and assumes similar behavior for

[1] Garro (2011) focuses on the distribution decision process and covers the current methodology in greater detail.

current year. The change in demand is evaluated on a per MCC basis, which is to say that it calculates, total demand per "MCCs in Store", before, during and after the event in question. With this understanding of change in demand they calculate a trend factor which is used to adjust the average daily demand.

Equation 2.3 - Trend Calculation General Approach

$$Trend_{p1,p2} \approx \frac{Demand_{p2}/M_{p2}}{Demand_{p1,}/M_{p1}}$$

Equation 2.4 - Trend-Adjusted Daily Demand General Approach

$$Adj\ Daily\ Demand_{mcc,t2} \approx Observed\ Daily\ Demand_{mcc,t1} \times Trend_{t1,t2}$$

where p1 refers to a baseline period preceding a seasonal event, p2 refers to the period when demand is affected by the event, and M refers to the number of MCCs present in the store in a given time period. The biggest shortcomings of this trend methodology are that: 1) it requires a case by case evaluation; and, 2) it relies heavily on country managers alerting the distribution department of upcoming events.

2.4 Proposed Solution

2.4.1 Stakeholders and Deliverables

This project addresses the needs of three distinct stakeholders which are defined as:

1. Distribution Department: As the owners of the forecasting models and sponsors of this project the distribution department is the lead stakeholder. They own all of the automated processes that calculate the replenishment quantities and have final say on whether to accept a country manager's recommendation to make an adjustment. In addition to developing a better forecast the output from this project should make it easier for them to incorporate new demand-impacting events as they discover them. Therefore the solution should help them be more accurate and save them time.
2. Designers/Buyers: Part of the goal for this project is to understand the truly significant events across the world. Understanding when these shocks occur on a global level will help buyers plan their purchases so that there is enough inventory to satisfy all major selling spikes across the globe.
3. Country Managers: It is currently very cumbersome for country managers to keep track of all relevant events and factors that impact sales in their respective countries. The new model should make it easier for them to test whether an event is relevant as well as incorporate the relevant ones into the automated baseline estimates. Additionally, country managers are

constantly monitoring the replenishment needs for all of their stores and therefore serve as great sources of input to validate and improve the models as time goes on.

Given the stakeholders and their respective interests, the deliverable for this project is a series of models that can automatically and systematically forecast demand while accounting for seasonality and other factors exogenous to the quality of an item. The models would serve as the basis for developing three new tools:

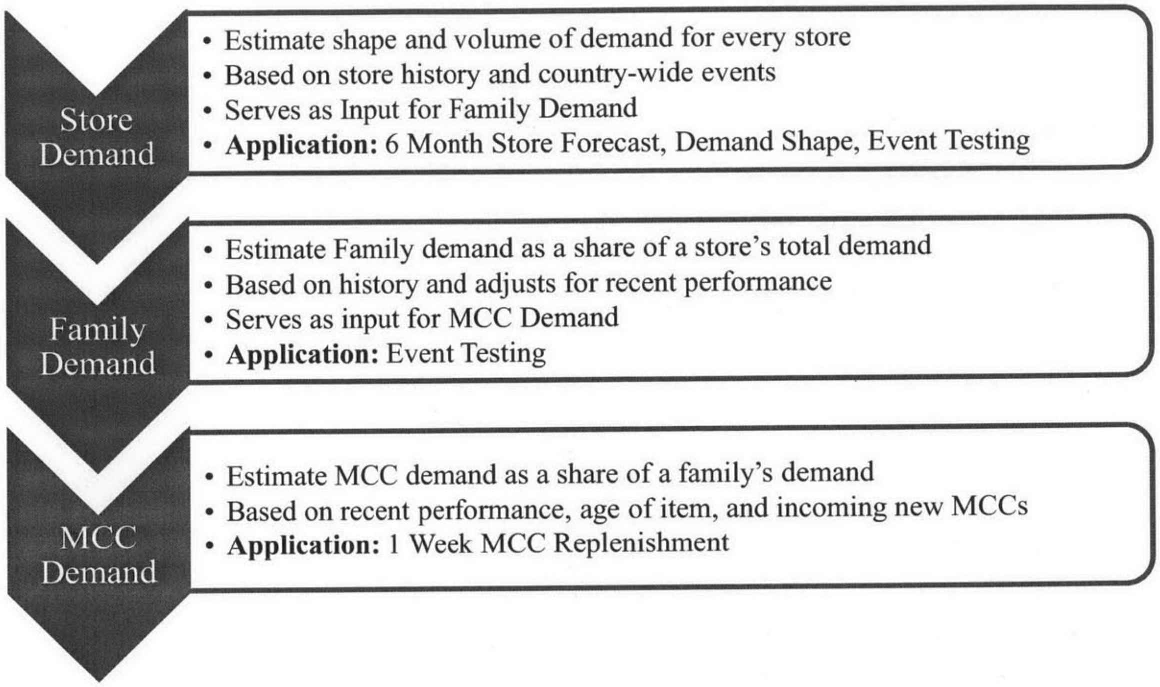
1. Aggregate Demand Forecast Tool: This tool would project demand for as far as 6 months that can be aggregated from as low as the total demand of one single store, to the total demand in a region, or for the entire network of stores. This tool would mostly help designers and buyers plan their purchases.
2. Seasonal Event Identifier: This tool refers to a systematic way to test whether a seasonal event, such as Ramadan or Father's Day, does in fact have a significant impact on demand.
3. Replenishment Quantity Estimator: This tool would serve to provide the distribution department with an estimate that automatically incorporates the demand shocks that result from events identified by tool #2.

2.4.2 Solution Framework

In evaluating the various modeling approaches for this study it was determined that regression models with dummy variables would be most appropriate. For one, it is simple to implement, easy to understand and can be modified with relative ease. Secondly, Chu and Zhang (2003) showed that "seasonal dummy variables can be very useful in developing effective regression models for predicting retail sales." While they concluded that neural networks would be the best approach it was not implemented here for its increased level of complexity and difficulty. Additionally, understanding the driving factors for a given output is much more nebulous because the model arrives at it through a series of learning iterations, and thus it can be hard to interpret a result in cases where time is limited.

The models were built in three stages. The first stage was to estimate the store level aggregate demand, which corresponds to the total number of units a store would have sold if it did not experience any stockouts. The next level was to estimate the percentage of each store's demand that could be attributed to each product family. The output of this second stage is a percentage that can be converted to units by multiplying by the estimated store aggregate demand. The last level was to estimate the percentage of each product family that should be attributed to each MCC. Figure 3 below introduces the high level components of each stage, with more details about derivation and performance in Chapter 5.

Figure 2.3 - Solution Framework



2.4.3 **Summary of Contributions**

This project makes several contributions to Zara which will be further detailed in subsequent sections. However, at a high level this project helps Zara on several distinct fronts. The first is that it defines a framework of distinct components that Zara can work on to improve its forecasting capabilities going forward. The framework is best defined by Figure 2.1, whereby Zara can decide to focus on any of the three stages of forecasting, one is improving its methodology for estimating demand data, the second is better accuracy if forecasting that input data, and lastly how to distribute based on those forecasts.

Secondly, it delivers forecasting models that significantly improve upon their current methods. In forecasting total units of demand at the store level, this project reduces the forecasting error between 5% and 12% when considering a 1 week time horizon and between 2% and 11% when considering a 6 month time horizon, as shown in Table 2.1. The errors here are calculated using a weighted MAPE approach which is detailed in Section 5.1.2.

Table 2.1 - Store Level Forecast Improvements in wMAPE

	Estimation Errors (Stores)					
Country	Baseline 1 Week	MIT9 1 Week	Difference	Baseline 24 Weeks	MIT9 24 Weeks	Difference
Spain	17.6%	11.6%	-6.0%	21.9%	15.2%	-6.7%
Belgium	17.3%	11.4%	-5.9%	14.0%	12.5%	-1.5%
China	17.9%	12.9%	-5.0%	22.1%	19.0%	-3.1%
Saudi Arabia	31.3%	19.7%	-11.6%	31.2%	20.2%	-11.0%

Thirdly, and also related to forecasting, is the potential to help Zara reduce its lost sales by 24% which would lead to an increase in total annual sales of 1.8%. These gains are achieved through improved forecasting of MCC level forecasts which help match supply with demand on a weekly basis.

2.4.4 Thesis Organization

The remaining sections of this thesis details the effort undertaken to solve the problem described thus far. Chapter 3 covers the literature reviewed for this project; first it covers the work done by past LGOs and other academics to improve Zara's operations and it ends with an overview of sound practices for forecasting demand.

Chapter 4 describes the data available for this study as well as some high level characteristics to help the reader understand some of the nuances that make this problem difficult. Chapter 5 has an extensive and detailed presentation of the models developed for this thesis as well as the accuracy of each model.

Chapter 6 discusses Zara's buy-in and current state of their implementation efforts for the models developed in this thesis. The chapter ends with the latest modifications that Zara has made to the models in order to incorporate them into their systems. Chapter 7 summarizes the contributions that this work makes to Zara and to the retail industry. Lastly, Chapter 8 makes recommendations for the future direction that Zara could take as it looks to improve upon the work presented here.

3 Literature Review

There is a considerable amount of literature regarding Zara specifically and also regarding demand forecasting in highly seasonal environments. This chapter reviews relevant works upon which this study attempts to build. First, an overview of past LGO thesis is presented. Then, several works by Caro and Gallien (2007, 2010, 2012), who have done extensive research relating to Zara's operations, are covered. Lastly, other works relating to demand forecasting are covered.

3.1 Past LGO Master's Theses

The theses of Correa (2007), Garro (2011), and Bonnefoi (2012) address the topic of demand forecasting. Correa's work is the most relevant to this study as it provided a methodology for forecasting demand at the MCCT level for 1 week into the future, and hence the basis for determining replenishment quantities. Garro and Bonnefoi estimated demand from an MCC-centric point of view prior to the products being introduced. Garro's focus was in determining the initial shipments of new items and Bonnefoi's focus was on determining the initial quantities to purchase for a given item based on a lifetime demand forecast. These works, however, fail to address three items which are developed in this study: 1) Impact on demand from seasonal events that shift in the calendar from year to year such as Easter and Ramadan; 2) Methodology for calculating a long term forecast for planning purposes; and 3) Impact of demand cannibalization from other items in store.

3.2 Caro and Gallien

Felipe Caro and Jérémie Gallien have co-authored several papers detailing improvement projects for Zara's operations. Gallien has also advised multiple MIT LGO theses and is therefore an excellent starting point for reviewing literature as it pertains to Zara. Caro and Gallien (2007) presents a strategy for optimizing product assortment based on observed demand. The paper focuses mostly on dynamically adjusting the tradeoff between using floor space for products with observed high demand versus using the space for new items in order to learn their selling potential. However, it does suggest that in terms of forecasting it would be worthwhile to use "new (demand) information...which may be combined with historical information to select a future assortment." This is essentially the data gathering approach employed in this thesis, where recent demand data is combined with historical behavior at the Store, Product Family, and MCC level to forecast demand. One deviation from Caro and Gallien 2007 is on the assumption of stationarity of demand which this thesis does not assume because, as shown in section 4.3.6, demand rate per item decreases as the time in the store increases.

Caro and Gallien (2010) builds upon the forecasting work presented in Correa (2007) and makes the assertion that more in-store inventory of a particular item leads to additional sales for that item. It is of course expected to reach a saturation point, but it is difficult to observe in the data given that Zara is mindful of over-flooding stores with inventory. This assertion is tested throughout the models developed in this thesis and is included in the final model for family level demand. Lastly, Caro and Gallien (2012) details a forecasting approach which feeds into a clearance price optimization program. Even though this thesis focuses exclusively on full price items, there were two components of the forecasting model presented by Caro and Gallien that proved to be useful for this thesis: 1) Age of the article measured as time since introduced at stores is a critical component of the MCC level demand model, and 2) Previous Period Demand is also a variable used in both the family and MCC level models as the data shows some level of autocorrelation.

3.3 Forecasting Approach

Large motivation for this study is to develop an objective estimate of the impact of seasonal impacts as it is currently heavily influenced by human experts. Onkal et. al (2009) address this topic directly and present that strong biases are in fact introduced whenever humans provide estimates, this in large part is why the models in this study do not incorporate human input to generate the forecast. Instead, it presents a methodology for testing whether human intuitions are in fact validated by past data. Chu and Zhang (2003) present various methods for forecasting seasonal demand and demonstrate that using regression with dummy variables can yield good results. This study successfully uses that approach for forecasting long term demand.

This thesis also leverages the demonstrated evidence that demand for fashion items shows strong correlation across periods (Fisher & Raman, 1996) and (Fisher, Rajaram, & Raman, 2001). To make use of this information, the models presented here consistently use auto-regressive terms for estimating demand. One strong deviation from both of these works however is the fact that in our case we assume that order quantity is fixed and cannot be updated mid-season. While in some cases Zara is able to reorder mid-season, the work presented here is designed for the distribution department which simply operates with an available inventory and does not have the ability to reorder.

Two other demand forecasting topics explored in this thesis are: the impact of demand cannibalization from new items arriving at stores; and, the impact of weather on demand across time and across products. On product cannibalization, Shah and Avittathur (2007) found that increases in demand cannibalization led to decrease in profitability. They tested this assertion by estimating profit at various levels of cannibalization, in this thesis we determine that factor by using regression to determine how much demand for an item falls with the introduction of new items. The impact of weather was less conclusive in this study and was not incorporated into the final solution. However, the impact of weather was extensively explored based on the expectation that demand across products and across time would be change when abnormal weather conditions are present. Starr-McCluer (2000) found that several effects impact retail sales, first weather can have a "convenience effect" whereby customers go shopping on very hot days to avoid being outdoors and conversely, cold temperatures and precipitation might hinder travel and therefore keep people away from stores. A second effect is that weather complements certain activities such as golfing, hiking, or going to the beach. Adding to these effects is the impact that weather might have on future spending, for example, if coats are purchased early in the season due to low temperatures it is fair to expect that demand for these goods will be lower later in the season.

4 Demand Forecasting - Preliminary Analysis

4.1 Data Input

The base dataset used for analysis contained all transactions of clothing articles going in and out of stores in Spain, Belgium, China and Saudi Arabia. This subset of countries was selected because it represents a diverse and meaningful subset of their network.

Before going into more detail, a few definitions are in order. A transaction occurs when an article of clothing is sent to a store, or when it exits the store via a purchase, a transfer to another store, or a recall back to central warehouses. The data is unique by date, store ID, and MCCT where each row has information about the transactions that occurred on each date. For purposes of this study, the data was cleaned and modified in the following four ways:

1. Removed Closed Stores: Any store that was closed for 1 or more cycles between January 2009 and December 2012 was removed. Most common causes for stores being closed was to either accommodate for renovations or because the store was launched after 2009. This filtering eliminated the following number of stores:

   Table 4.1 - Records Removed for Closed Stores

Country	Total Stores	Stores Removed	Percentage Removed
Spain	200	10	5%
Belgium	40	3	8%
Saudi Arabia	50	8	16%
China	100	30	30%

2. Removed "Dead" MCCs:
   At Zara, store managers frequently remove items from the selling floor when it stops selling well. However, this action is not tracked in any of its databases. Dead MCCs were removed in an attempt to only track MCCs that are competing with each other, i.e. live MCCs. An MCC was considered dead when it experienced 0 sales over a four week period.
3. Removed Marked Down MCCs:
   Items very seldom go on sale at Zara and generally sales start out as store wide for a few weeks at the end of each season. After a few weeks the store is transitioned from mostly marked down merchandize to mostly full price across several weeks. Furthermore, the marked down items are not replenished in the same regular fashion as the in-season items. Stores usually receive large shipments at the beginning of the mark down period to

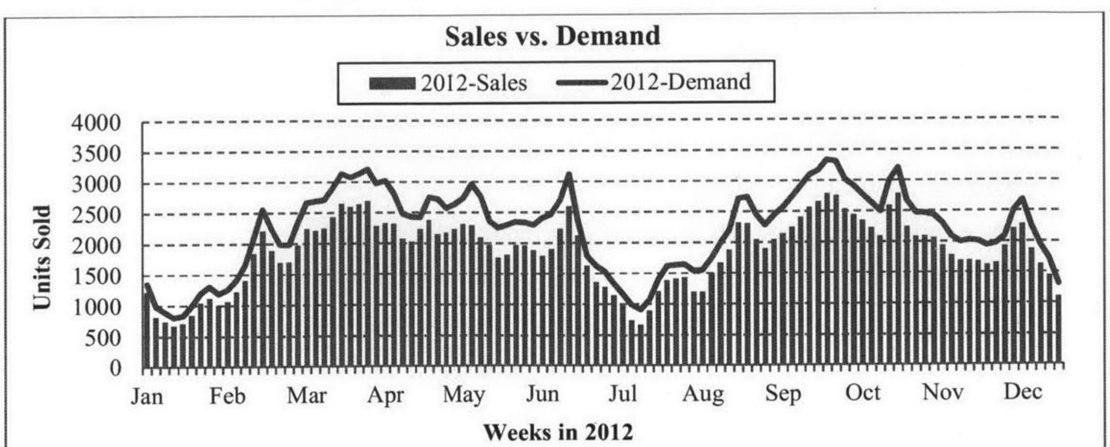
supplement their own left over inventory. Each store attempts to sell their on-sale inventory as fast as possible. For this reason, the behavior exhibited by on-sale items was excluded when studying full price items. Excluding these sales changes the annual demand curves by essentially reducing sales to 0 during January and July, which are the heavy on sale periods.

4. Aggregate by Cycle: Lastly, the dataset was aggregated by cycle, which is to say that each row contained the total transactions (inflows and outflows) for a cycle (M-W or Th-S) instead of having a row for each day. This aggregation had two important advantages. First, it greatly reduced the number of records managed which helped with analysis run time. Secondly, it reduced variance of data points by observing pooled demand versus daily demand. This aggregation does not limit the accuracy of the data because Zara already manages its replenishments by cycles, which means that it needs to know how much demand there will be in a given cycle, independent of the intra-cycle distribution.

4.2 Calculating Demand from Point of Sale (P.O.S) Data

A common problem in forecasting demand is the issue of sales data versus demand data. Sales data represents the portion of the demand that was captured and lacks the unobserved demand that occurs during stockouts. Therefore, the exercise to convert sales data to demand data consists of estimating what sales would have been if there had been infinite inventory, or no stockouts. Garro (2011) and Bonnefoi (2012) developed methodologies to resolve this problem. The two methods were relatively similar; however, since Garro's method was already integrated into the base transactional dataset so it was chosen for this study. Figure 4.1 shows the difference between P.O.S. and Demand data for one store in Belgium. The rest of this section provides an overview of this method, however the author recommends reviewing Section 2.1 of Garro's thesis. Please note that the notation in this section differs from notation in other sections in an effort to remain consistent with Garro (2011).

Figure 4.1- Point of Sale vs. Demand for Store 376 in Belgium



4.2.1 General Framework to Calculate Demand

The idea is to modify the sales data on days where stockouts occurred with an estimate of what sales would have been if the stockout had not occurred (i.e. demand). A good source of information for this estimate is the sales information for days when the item did not stock out. The method implemented took into account the day of the week when the stockout occurred and also the sales distribution across sizes for a particular item.

4.2.2 Adjustment Day-of-Week Cyclicality

The idea that demand varies by day of the week is commonly accepted by Zara managers. For example, we would expect to see higher sales on the weekends than on weekdays. Adjusting for the day of week would lead to the following estimation of demand:

Equation 4.1 - P.O.S to Demand, Day-of-Week Adjustment

$$D_t \approx \left(\sum_{p \in P} S_{t,p}\right) \Big/ \left(\sum_{p \in P} W_p\right)$$

where D_t is the demand for a size t in a given week, P is the set of days in the week with available inventory for sale, $S_{t,p}$ is the sales observed for size t on day p, and W_p is the proportion of weekly sales that occur on day p. W_p was found to vary significantly between stores that open on Sundays and stores that do not and as a result a different reference table is used for each type of store as shown in Table 4.2:

Table 4.2 - Proportion of Sales by Day of Week, Wp

Day	Closed Sunday	Open Sunday
Monday	14%	13%
Tuesday	15%	13%
Wednesday	15%	14%
Thursday	16%	14%
Friday	18%	16%
Saturday	22%	21%
Sunday	N/A	9%
	100%	100%

4.2.3 Adjustment for Size Curve

Another commonly accepted idea in fashion retail is the variation in demand by size. Size medium is the most popular size in most countries while extra-small and extra-large are less common. Therefore, sales data of size Medium inform demand of stocked-out Extra-Large by a different factor than sales of Extra-Large would inform sales of stocked out Medium. Equation 4.2 shows the calculation for making the demand adjustment based on size curve.

Equation 4.2 - Size Curve Demand Adjustment

$$D_t \approx \sum_{p \in P} S_{t,p} + \sum_{p \notin P} W_t \left(\sum_{t \in T} S_{t,p} \right) \Big/ \left(\sum_{t \in T} W_t \right)$$

where $p \notin P$ represents days with stock outs, and W_t represents proportion of sales for each size, and T represents the set of all sizes offered for a given item. The first part of the equation accounts for the observed sales for a given size t, while the second part adds up the sales of all other sizes while size t is stocked out and divides by their respective sales proportion to extrapolate a total sale for the item. Lastly, it multiplies the extrapolated value by the weight of the stocked out size to use as an estimate of demand for size t on the days it was stocked out.

W_t varies by product depending on the number of available sizes for the product. For example, some items are sold in sizes XXS, XS, S, M, L, XL, XXL while others are only sold in one size (usually M). Table 4.3 below shows values of W_t based on the number of sizes offered for a given item. For example an item offered in three sizes (S, M, L) would experience 43% of its sales in the middle size.

Table 4.3 - Proportion of Sales by Size

	% Demand by Size							
Number of Sizes	1	2	3	4	5	6	7	8
1	100%							
2	63%	37%						
3	30%	43%	26%					
4	28%	37%	22%	12%				
5	10%	23%	31%	25%	12%			
6	13%	25%	27%	20%	11%	4%		
7	10%	21%	26%	20%	13%	7%	2%	
8	6%	21%	26%	20%	13%	7%	6%	1%

4.2.4 Merging Day-of-Week and Size Curve Adjustments

Garro (2011) found that adjusting for size was the most accurate method, however it had serious limitations. Namely, it was impossible to use for items that were only offered in one size and also it did not account for situations when the store might choose to remove the item from the floor because a major size has stocked out (which is common practice at Zara stores). Therefore, the final methodology averages the results between the two methods to estimate demand.

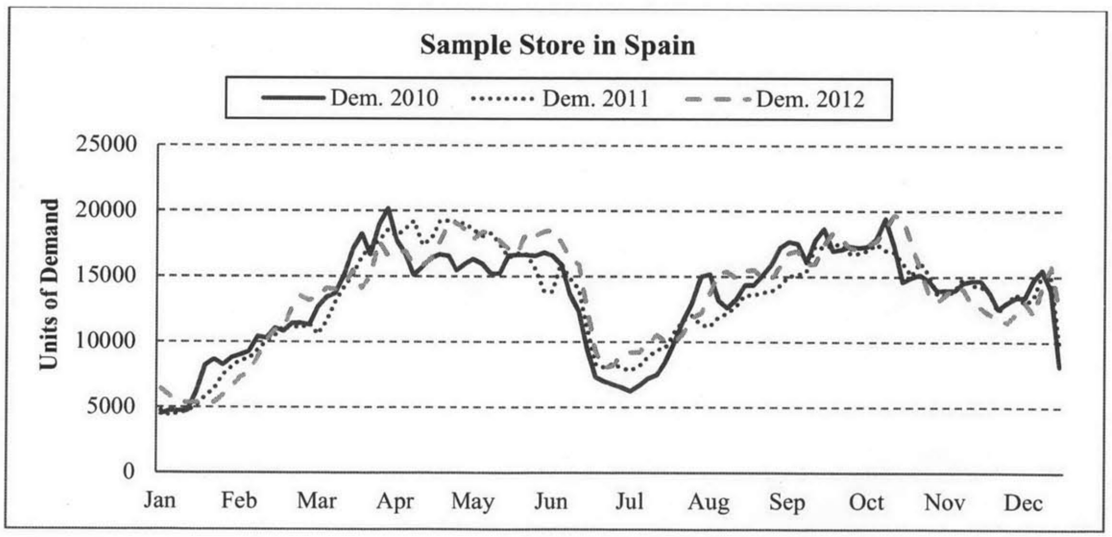
Ultimately, it was determined that the demand estimation process was not a critical piece of this study and therefore the work of Garro (2011) was carried forward. The forecasting models developed in this study and presented in Chapter 5 uses this data as the independent variable to forecast. Therefore, changes to the process for converting point of sale data into demand data would require that the coefficients for the forecasting models be recalibrated. However, the structure of the models should remain valid.

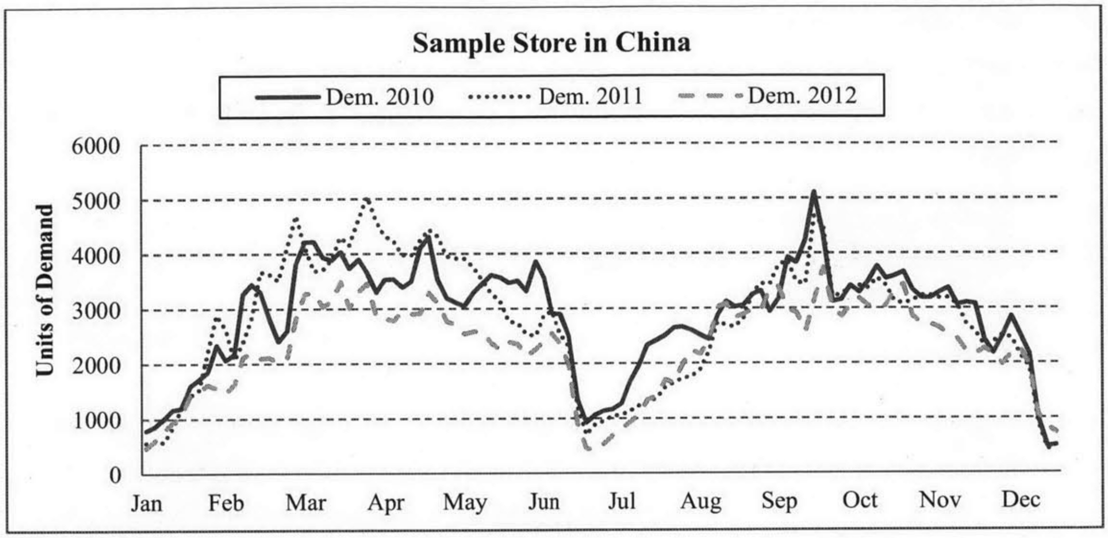
4.3 Demand Characteristics

4.3.1 Seasonality

Zara's demand cycles, in the northern hemisphere, are marked by two major seasons: summer and winter. Summer runs from February to June and experiences a peak in most countries around early April. This peak is driven by the fact that in most parts of the world customers begin to change their wardrobe from winter to summer around that time. However, Zara also plans to have their best products during that time and therefore it is difficult to isolate the effect of product mix from customer behavior. Winter runs from September to December and experiences a similar peak towards the end of September. The end of each season is marked by a store-wide sale which is intended to clear through as much inventory as possible to make room for the new season. Zara has little to no discounting outside of these end-of-season sales that run in January for winter sale and in July and August for summer sale. Figure 4.2 shows the general "m" shape of demand, where there are two clear peaks in spring and fall, mat major stores in Spain and China for years 2010 to 2012.

Figure 4.2 - Examples of Annual Demand Pattern



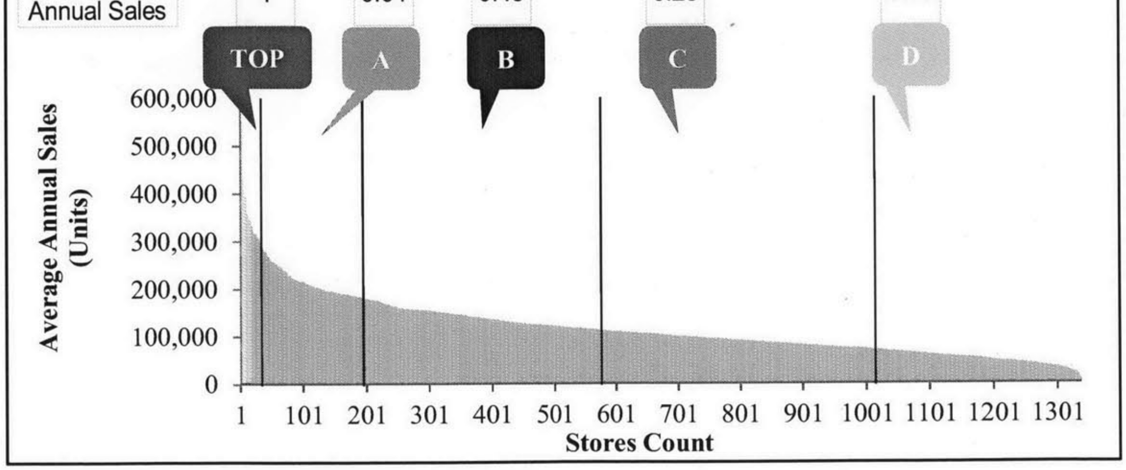


4.3.2 Large Stores vs Small Stores

Another difficult aspect of Zara's demand is that there is a wide range in sales volume between high volume stores and low volume stores. Zara classifies its stores as TOP, A, B, C, and D such that the top 15% of stores account for about 30% of sales as shown in Figure 4.3. This presents a challenge in modeling demand because the scale of demand from a small store is much different than that of a large store. In order to account for this, a relative measure of demand must be used in order to be able to use a single model across multiple stores. Chapter 5 covers this in greater detail.

Figure 4.3 - Average Annual Sales by Store Category (Top, A, B, C, D)

	TOP	A	B	C	D
Stores	52	144	381	437	324
% Sales:	11%	29%	35%	25%	10%
Normalized Annual Sales	1	0.64	0.43	0.28	0.15



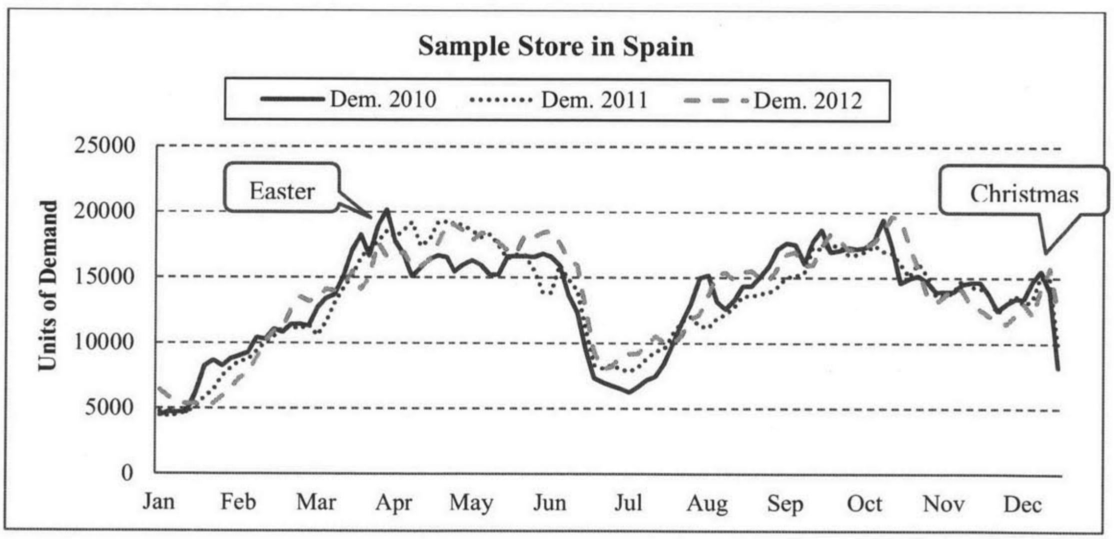
4.3.3 Fixed Calendar Events: Christmas, NYE, Mother's Day

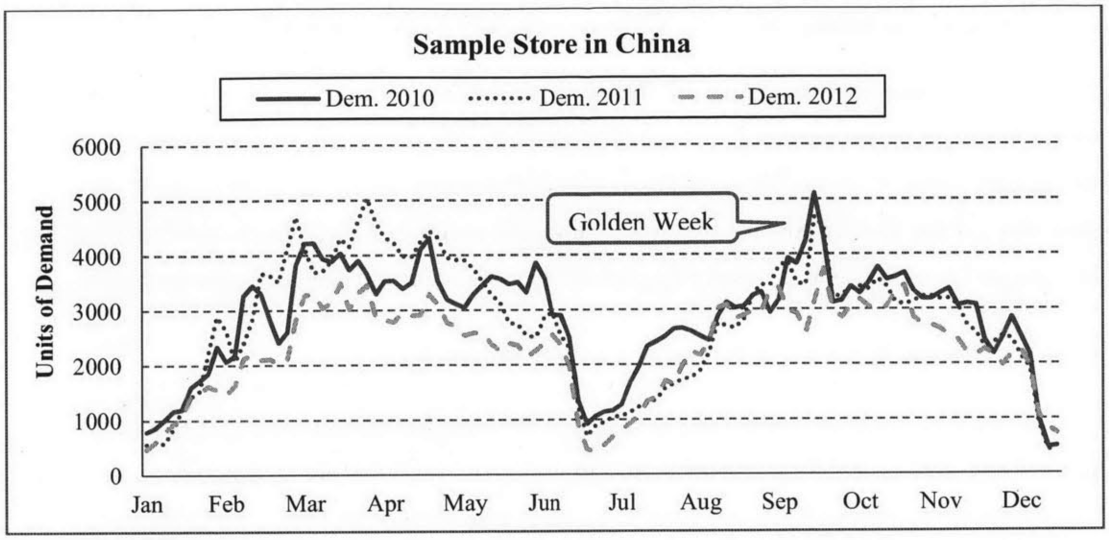
Throughout the year there are calendar events such as Christmas or Mother's day that can lead to predictable increases or decreases in demand. Furthermore, the reaction to a given calendar event is not uniform across all stores. Many events are regional in nature, for example Christmas is very important in most western countries but it is irrelevant in countries such as Saudi Arabia and China. Accounting for these events on a store by store basis is important to develop an accurate model.

4.3.4 Variable Calendar Events: Easter, Ramadan

Even more problematic are events that occur on a yearly basis but move around in the calendar. Easter, Thanksgiving and Ramadan are perfect examples of events that are known to have an impact for many stores however they do not occur at the same time every year. This makes it difficult to compare across years. Figure 4.4 shows examples of demand variations for both fixed and variable calendar events. In the figure, it can be observed that the peaks for Christmas across years occur at the same time. In the other hand, the peaks for Easter and for Golden Week vary from year to year.

Figure 4.4 - Demand Impact of Fixed and Variable Events

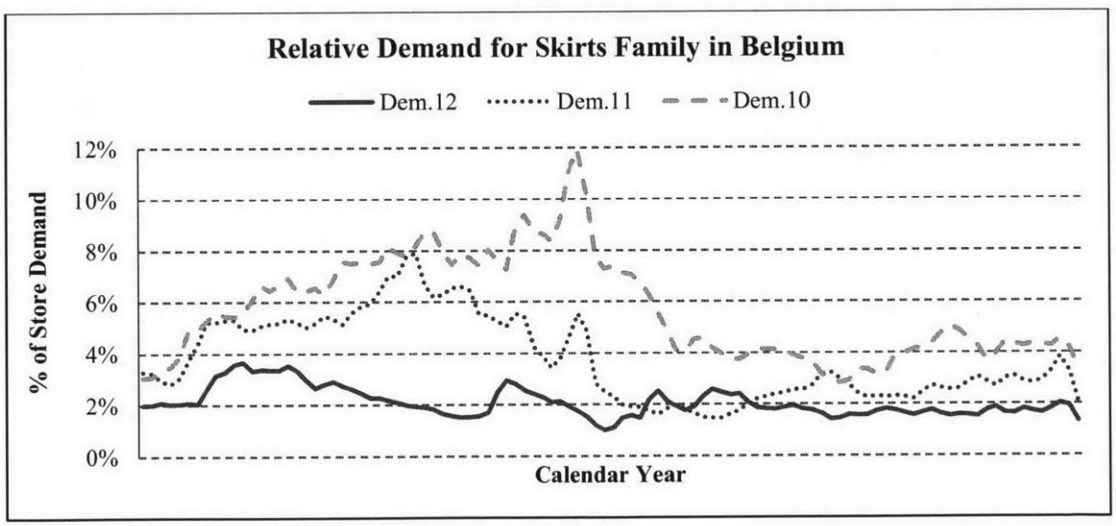


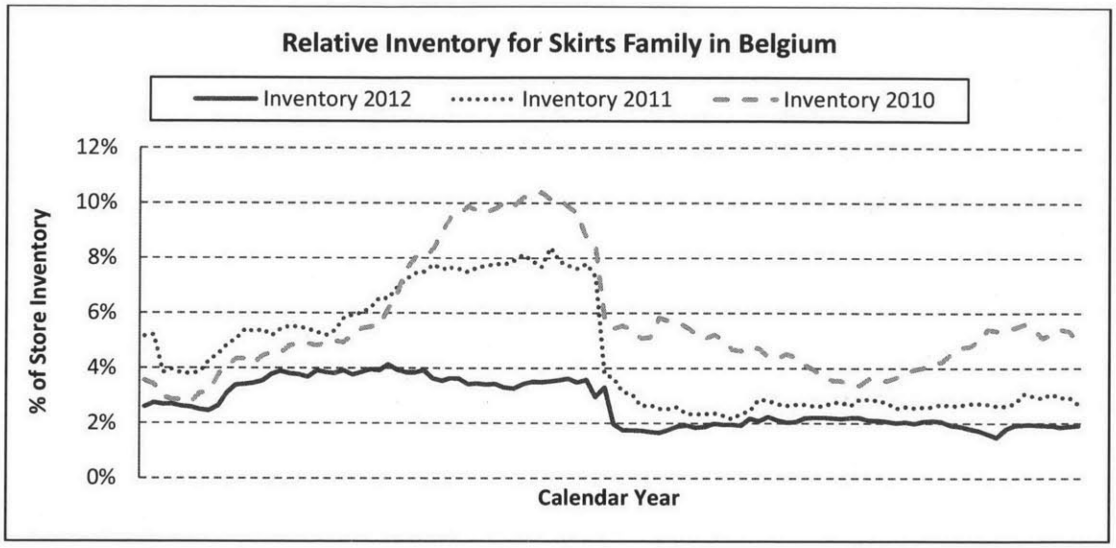


4.3.5 Product Mix Variations

When evaluating demand aggregated at the family or sub-family level it becomes difficult to isolate seasonal impact from product mix. For example, if Zara chooses not to focus on a particular product family then the sales for that item will be depressed independent of seasonal factors. As Figure 4.5 shows, stores in Belgium experienced very low demand for Skirts in 2012; however, the in-store inventory for that same year shows that skirts formed a very small part of the overall collection.

Figure 4.5 - Example of Product Mix Variation

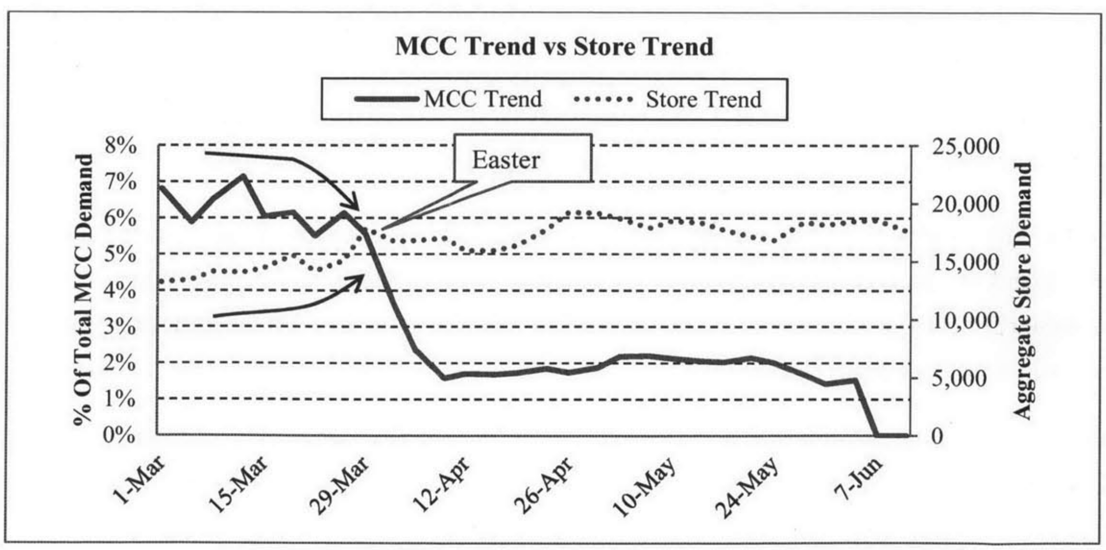




4.3.6 Product Lifecycle vs Seasonal Trends

MCC level demand and Store level demand follow different patterns. Each MCC follows a decreasing sales lifecycle which is independent of season, i.e. sales are strongest when item is first introduced and decreases over subsequent weeks. Aggregate store demand, however, goes up and down based on the season. This complicates how to make use of store level or even family level demand estimates for short term MCC replenishment decisions because even if store demand is expected to go up by 10% during a given week, it is inappropriate to increase expected demand for each MCC by 10%. An illustration of this effect is shown in Figure 4.6 where the average percentage of cumulative demand for all items introduced on March 1st show a clear decline trend as it approaches Easter. The total store demand, however, goes up during this same time period.

Figure 4.6 - MCC Trend vs Store Trend for Easter in Spain Store 160



5 Demand Forecasting - Methods and Results

This chapter covers all of the demand models developed in this study. It begins with the performance metrics used to compare models, it then details the models for Store Demand, Family Demand and MCC demand. Within each of those sections the models are detailed, then results are presented, and lastly variables that were tested but were not successfully incorporated into the final models are presented.

5.1 Performance Metrics

The model building process for this study depended on multiple iterations which required a standard and robust methodology for comparing performance across models. Three levels of tests were used to evaluate each model. The first level was testing the statistical validity of the regression models by looking at t-tests for each coefficient, and multicolinearity between independent variables, among others. The second level was to test the forecast accuracy which was done by calculating the Mean Absolute Percentage Error (MAPE) and the weighted version of the metric (wMAPE). The third level was to measure the operational gains that Zara could achieve from the models. This last metric was only implemented for MCC level forecasts.

5.1.1 Performance Metrics: Statistical Validation

The tests detailed below serve as the first pass of validation for the models developed throughout this thesis. Before considering the econometric interpretation of each model or its forecasting accuracy, the tests below were applied to verify that the formulation was sound.

T-Test: The t test is used to test the single hypothesis that a coefficient is not equal to zero. The test is carried out by calculating a t-statistic which is defined as the estimated coefficient divided by its estimated standard deviation, or standard error. Then the t-statistic is compared directly to critical values in the t-table (Kennedy, 2003). Throughout this thesis coefficients were considered to be significant if the t-test yielded a probability below .10 that the coefficient could be 0.

F-Test: The F test is a common test used in regression analysis to test the joint hypothesis that all coefficients in regression model are non-zero (Kennedy, 2003). When the test shows that the probability of all regressors being zero is very small it indicates to the user that the parameters for the entire model did not occur by chance.

Adjusted R^2: This calculation is a measure of how much of the data's variance is explained by the regression model while adjusting for the number of variables used, or degrees of freedom (Kennedy, 2003). It is difficult to determine what a good Adj. R^2 value is prior to exploring a given set of data.

However, it can be a very valuable measure in determining whether a new variable is adding considerable explanatory power to a model specification.

VIF: Variance Inflation Factor is an indicator of multicolinearity which helps identify independent variables that could be estimated using a linear combination of all other independent variables. This test is better for identifying redundant variables than the more common evaluation of pairwise correlations.

5.1.2 Performance Metrics: Forecast Accuracy

The basis for measuring forecast performance was based on the Mean Absolute Percent Error which is recommended for cases where all data are positive and simplicity is sought (Hyndman & Koehler, 2006). However, the measure was modified to account for the level of demand forecasted. In essence, Zara finds it more damaging to miss a forecast by 1% on 1000 units of demand than it does a 10% error on 10 units of demand. To account for this, the errors were weighted to account for the size of demand being forecasted. Both measures are discussed below.

Equation 5.1 - MAPE Calculation

$$MAPE \triangleq \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|Y_i - \hat{Y}_i|}{Y_i}\right)$$

Equation 5.2 - Weighted MAPE, or wMAPE, Calculation

$$wMAPE \triangleq \left(\sum |Y_i - \hat{Y}_i|\right) / \left(\sum Y_i\right)$$

$where:$

$Y_i: Actual\ Value\ of\ Dependent\ Variable$

$\hat{Y}_i: Estimated\ Value\ of\ Dependent\ Variable$

$N: Total\ Records\ in\ Sample$

The unweighted version calculates the mean value of all the percentage absolute errors. The weighted version, on the other hand, adds up all of the absolute errors before dividing by the sum of all the observed data. This in essence gives the same weight to each unit of error. After much deliberation with Zara's managers, it was determined that wMAPE was the better measure of performance because the units of mismatched demand were more important than the percentage mismatch. For example, an error of 1% during a week with 1000 units of demand would be worse than an error of 5% during a week where demand was 100 units. The wMAPE was chosen specifically to classify the former situation as being worse than the latter.

5.1.3 **Performance Metrics: Operational Gains**

Given that the end goal of this forecasting effort is to improve Zara's operations, this measure is used as a first level estimate of the potential gains they could achieve by implementing the approach presented here. The attempt was made to measure the change in lost sales between their current methodology and the proposed methodology.

Lost sales are defined as the difference between demand and available stock. This number is positive whenever demand exceeds the store's inventory. Otherwise, it is 0 because the available inventory was enough to satisfy all demand.

$$L_{m,w} = Max(0, D_{m,w} - I_{m,w})$$

$Where:$

$L_{m,w}: Quantity\ of\ Lost\ Sales\ for\ MCC\ m\ in\ week\ w$

$D_{m,w}: Estimated\ Demand\ for\ MCC\ m\ in\ week\ w$

$I_{m,w}: Inventory\ for\ MCC\ m\ in\ week\ w$

This formula was calculated for two scenarios: 1) Baseline Forecast, where Zara's current approach for forecasting demand was used to calculate an ideal level of inventory, and 2) Proposed Method, where the forecasts developed through the models presented in this thesis were used to calculate the level of inventory. This baseline approach had a limitation in terms of comparing across methods because each method could suggest a different total amount of inventory. For example, if one method proposed excessive amounts of inventory, it would have very low lost sales, but would clearly be considered inefficient and uneconomical. To account for this, the sum of inventory across all weeks for each MCC was held constant across both methods at a quantity that matched the total amount actually sent to the store.

5.2 Store Level Demand Model

This section is divided into three subsections. The first one defines the model used to forecast demand data and shows the regression outputs for the final model. The second section covers the results from applying the model on out-of-sample data to produce forecasts. The last section discusses the variables that were tested for this model but ultimately were not included in final specification.

5.2.1.1 Definition

In order to avoid having separate model structures for stores of different sales levels (see Figure 4.3) store demand was decomposed into two components: Aggregate Annual Demand and Weekly Percentage of Annual Demand. The product of the two yields a store's demand for a given week.

Equation 5.3 – General Store Demand Framework

$$D_{store,week}^{weekly\ units} = \sum_{w \in W} D_{s,w}^{WU} \times \frac{D_{s,w}^{WU}}{\sum_{w \in W} D_{s,w}^{WU}} = D_s^{AU} \times P_{s,w}$$

where, D stands for demand, WU stands for Weekly Units, AU stands for Annual Units, P stands for Weekly Proportion, and $\boldsymbol{W}$ the set of all weeks in a year. $D_{s,w}^{WU}$ refers to the demand in weekly units experienced at store *s* in week *w*. The sum of all weekly demands in a year is simply the store's annual demand and is identified by D_s^{AU}, and weekly demand divided by annual demand is simply the proportion of annual demand experienced in a given week, which is identified by $P_{s,w}$. D_s^{AU} and $P_{s,w}$ are modeled separately.

Given that replenishment decisions are made at the beginning of each cycle (Mondays and Thursdays), a week is defined as the seven days following the start of a cycle. This provides two significant benefits: 1) it smoothes out demand by looking at an entire week instead of separating weekends and weekdays; 2) it matches Zara's forecasting needs for replenishment making it easy to implement.

5.2.1.2 Annual Demand, D_s^{AU}

Annual Demand was forecasted heuristically because initial approaches yielded excellent results and the major focus of this thesis was to explore the intra-year fluctuations of demand. The model is as follows:

Equation 5.4 - Annual Demand Final Equation

$$D_s^{AU} = \left(\sum_{w=-16}^{-1} D_{s,w}^{WU} \right) \Bigg/ \left[\frac{1}{N} \sum_{w=-16}^{-1} D_{s,w}^{WP} \right]$$

where, the numerator is the observed demand data over the past 16 weeks, or roughly 4 months, and the denominator is the expected proportion during that time. To arrive at this result the expected proportion was calculated using two distinct methods across look-back periods ranging from 1 to 25 weeks. The first method for estimating expected proportion was to use the historical 3 year average for the weeks in look-back period (Figure 5.1). The second approach to estimate expected proportion was to use the forecasted demand calculated using model in Equation 5.5. The reasoning for the latter is that the forecast properly accounts for events that occurred within the look-back period. For example, if an event such as Easter usually occurs within the look-back period, the event would inflate the expected proportion in years when Easter did not fall within the look-back period. Figure 5.2 shows the results for the second methodology.

Figure 5.1 - Annual Demand Estimate: Historical Proportion

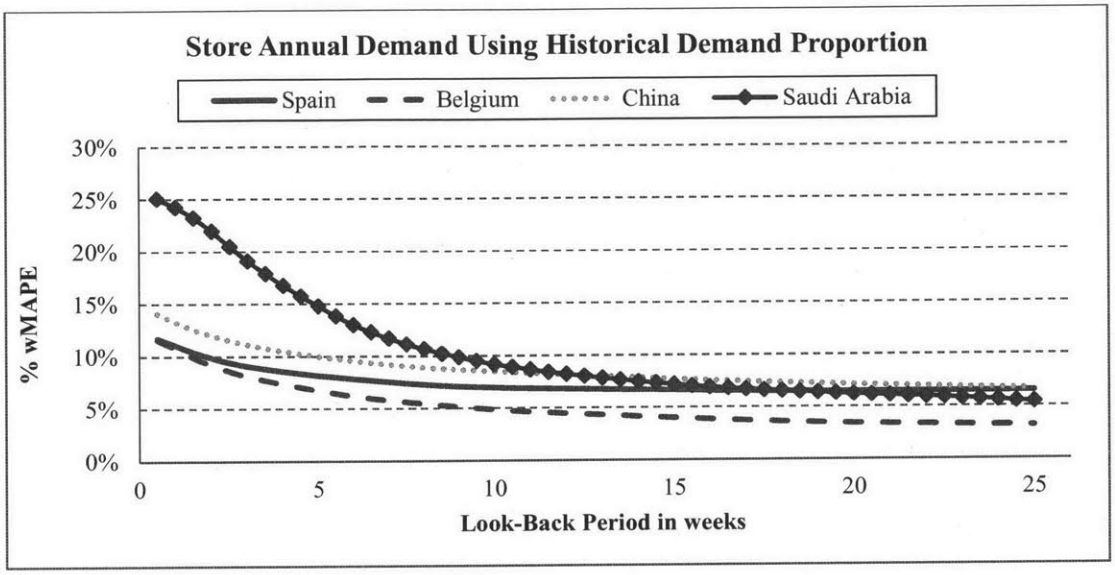
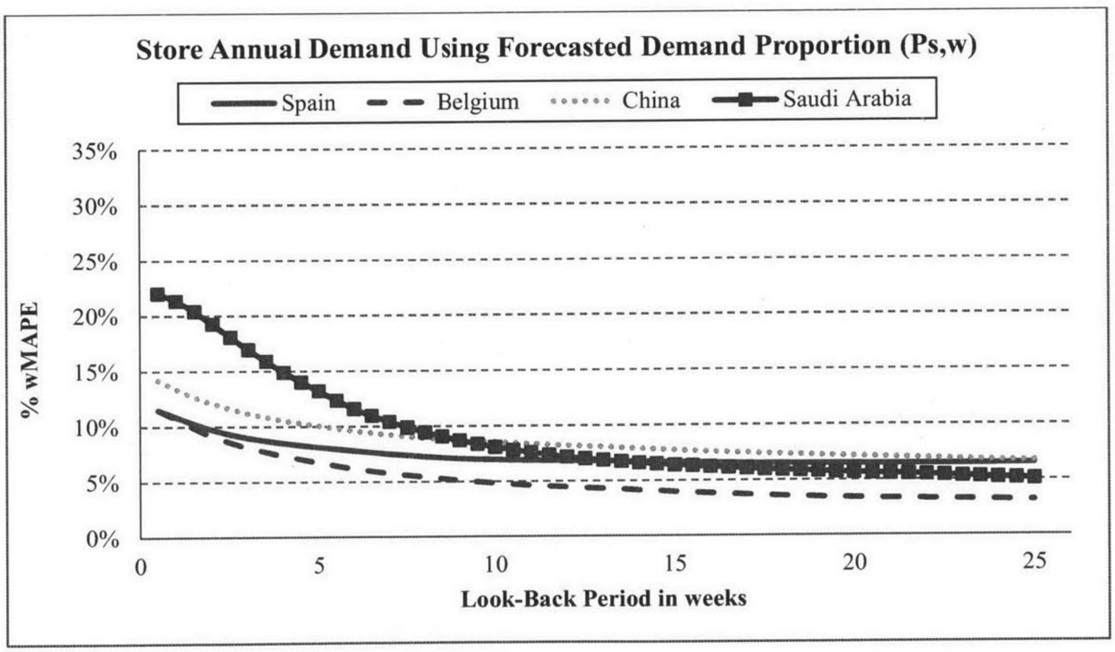


Figure 5.2 - Annual demand Estimate: Forecasted Proportion



Evaluating the results from both methods yields relatively similar results in that the error decreases as look-back period increases until leveling out. With the help of Zara managers, the look back period was selected to be 16 weeks because by this point the error for all 4 countries tested reaches steady state under both methods. The method that uses forecasted values of $D_{s,w}^{WP}$ was selected because it had provided superior performance for Saudi Arabia while maintaining similar performance for other countries. It is believed that countries where very important events move drastically in calendar from year to year, such as Ramadan in Saudi Arabia, would benefit from using the forecasted demand proportion values because the forecast more accurately accounts for the impact from the event.

5.2.1.3 Demand as Weekly Proportion, $D_{s,w}^{WP}$

As mentioned previously, the second component of the store level demand is the weekly proportion of annual demand. This model is the main tool delivered to Zara and is where the impacts of seasonality are explored in most depth. The major components of the final regression model are historical patterns of demand and dummy variables to flag events that shift dates from year to year. Equation 5.5 represents the regression model used to estimate the weekly proportion.

Equation 5.5 - Final Equation for Weekly Proportion of Store Demand

$$P_{s,w} = b_0 + b_1 H_{s,w} + \boldsymbol{B}_2 \boldsymbol{F}_{e,s,w} + \varepsilon$$

$Where:$

$P_{s,w}: Proportion\ of\ annual\ demand\ at\ store\ s\ in\ week\ w$

$H_{s,w}: 3\ year\ historical\ average\ proportion\ observed\ for\ store\ s\ in\ week\ w$

$\boldsymbol{F}_{e,s,w}: Set\ of\ dummy\ variables\ with\ value\ 1\ if\ event\ e\ has\ an\ impact\ on\ store\ s\ in\ week\ w$

An event refers to holidays such as Easter and Ramadan, but it is also used to flag the weeks prior to and after the actual event occurs. These periods are also considered given that demand tends to ramp up or be significantly depressed before and after holidays.

Table 5.1 shows the regression results when trained on data from all stores in Spain for years 2009-2011.

Table 5.1 - Regression Results: Spain Store Weekly Demand

Variable	Coefficient	Std. Error	Significance
(Intercept)	0.0001	0.000028	*
Hist. Avg: $D_{(s,w)}$	0.9942	0.001413	***
$F_{(e,s,w)}$: Easter - 6 Cycle prior	0.0005	0.000092	***
$F_{(e,s,w)}$: Easter - 5 Cycle prior	0.0001	0.000092	
$F_{(c,s,w)}$: Easter - 4 Cycle prior	0.0002	0.000093	*
$F_{(e,s,w)}$: Easter - 3 Cycle prior	0.0014	0.000093	***
$F_{(c,s,w)}$: Easter - 2 Cycle prior	0.0038	0.000093	***
$F_{(c,s,w)}$: Easter - 1 Cycle prior	0.0011	0.000093	***
$F_{(e,s,w)}$: Easter - 1st Cycle	(0.0019)	0.000093	***
$F_{(c,s,w)}$: Easter - 2nd Cycle	(0.0005)	0.000093	***
$F_{(e,s,w)}$: Easter - 1 Cycle After	(0.0006)	0.000093	***
$F_{(e,s,w)}$: Easter - 2 Cycle After	(0.0013)	0.000093	***
$F_{(e,s,w)}$: Easter - 3 Cycle After	(0.0008)	0.000093	***
$F_{(c,s,w)}$: Easter - 4 Cycle After	0.0002	0.000093	*
$F_{(c,s,w)}$: Easter - 5 Cycle After	0.0011	0.000093	***
$F_{(e,s,w)}$: Easter - 6 Cycle After	0.0013	0.000093	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

From the results from

Table 5.1, we see that the average demand from the past three years is the most important forecasting component of this model as it suggests a coefficient of nearly 1. Secondly, we note how the coefficients are positive for cycles leading up to Easter which corresponds to the observed ramp up in demand that Zara experiences every year. We also see that as we pass Easter the coefficients suggest a drop in demand and later show another up-trend after the effects of Easter appear to have passed. It should be noted that the demand being forecasted is for the 'upcoming week' which is to say, for the two following cycles. This is why the coefficients corresponding to the week of Easter are negative, because they represent the demand for the second cycle during Easter and the first cycle after Easter. Furthermore, the reader should be reminded that the dependent variable in this regression is a value between 0 and 1 representing the percentage of annual demand experienced in the week being forecasted, and therefore, the coefficients on dummy variables are to be interpreted as the percentage changes between normal weeks and weeks impacted by an event.

Table 5.2 - Regression Metrics

Metric	Result
F-Test	<.0001
Std. Error	0.0023
Adj. R2	0.8985
Max VIF	1.11

By evaluating the results in Table 5.2, it is clear that the model fits the data relatively well, with an adjusted R^2 above .8 and a max VIF which is well below the suggested cutoff of 10 (Kutner, Nachtsheim, & Neter, 2004). The following section explores how well this model preforms when forecasting future demand.

5.2.2 Results: Store Level Forecasts

The first step in testing the model results was to evaluate the out of sample R^2 for proportional demand estimates. The coefficients for the model shown in Equation 5.5 were trained using data from 2009 to 2011. The test results shown in this section are generated using data from 2012. Table 5.3 shows values for out of sample R^2 that are relatively close to the in-sample values which helps validate that the model performs well even on out of sample data and is likely to perform well for forecasting.

Table 5.3 - Out Sample R^2 for Store Demand

Country	Out-Sample R^2
Spain	0.8258
Belgium	0.7828
China	0.8009
Saudi Arabia	0.8258

The next step was to measure the model accuracy in terms of units predicted and compare to a baseline approach that closely resembles Zara's current approach. The baseline estimate is obtained by using a two week trailing average which is akin to Zara's current process for forecasting demand. Their current process essentially assumes that the demand for each item over the upcoming week will closely match the demand observed over the previous two weeks. To test the error of a 1-week forecast, the coefficients from Equation 5.5 are applied to every week in 2012. This application of the coefficients generates the percentage of annual demand that is expected for every week in 2012. The final step is to multiply these percentages by the estimates of annual demand which are updated for each week given the observed demand in the look-back period of 16 weeks. In addition to the 1-week forecast, however, it was also necessary to estimate the errors from forecasting for a time horizon as long as 6 months. To do this, the process for the one week forecast was modified to use annual demand estimates produced in weeks prior to the week being forecasted. Since the annual demand estimate depends on the characteristics of the look-back period that continues to shift throughout the year, the annual demand estimate is not the same for every week. Therefore, the demand forecast for horizons longer than one week are the result of multiplying the annual demand estimated in present week and multiplying it by the percentage of annual demand expected during the week in the future. Since the percentage estimate only depends on data from previous years and event dummy variables, the estimate does not change from week to week.

Table 5.4 shows remarkable improvements in accuracy by using the models outlined in this thesis. Especially telling is the fact that the 1-week forecasts outperform the current methodology in every case, both under MAPE and wMAPE evaluations. Secondly, the 6 month forecasts outperform the current 1-week baseline method for all but one case.

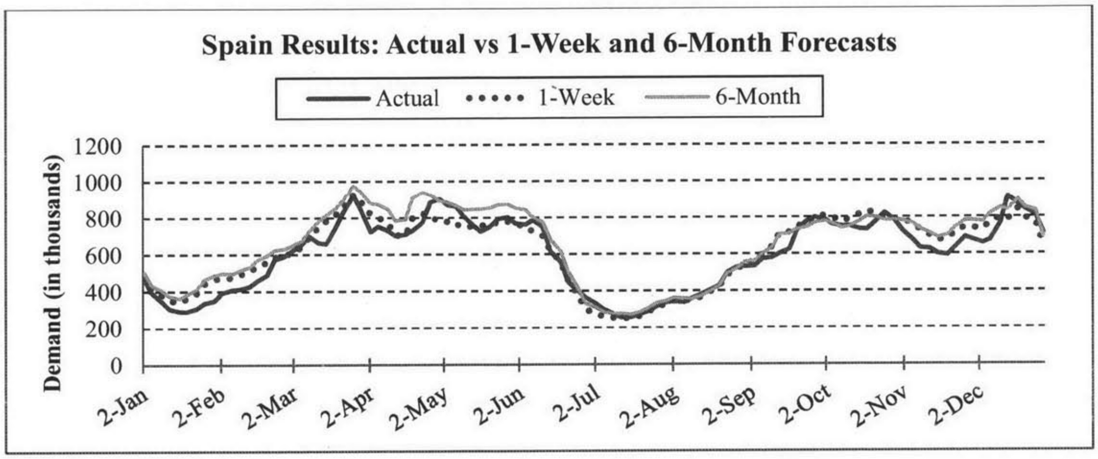
Table 5.4 - Summary of Forecast Errors for Store Demand

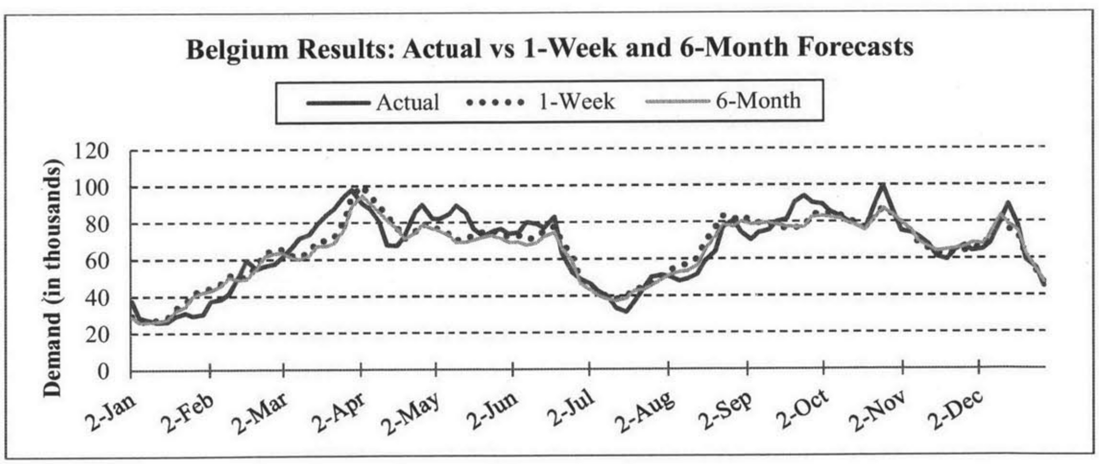
	Baseline	wMAPE by Forecast Horizon			
Country	Error	1 week	1 month	3 month	6 months
Spain	17.6%	11.7%	13.2%	14.5%	15.2%
Belgium	17.3%	11.4%	11.9%	12.2%	12.5%
China	17.9%	12.9%	15.1%	17.9%	19.0%
Saudi Arabia	31.3%	19.7%	21.1%	20.9%	20.2%

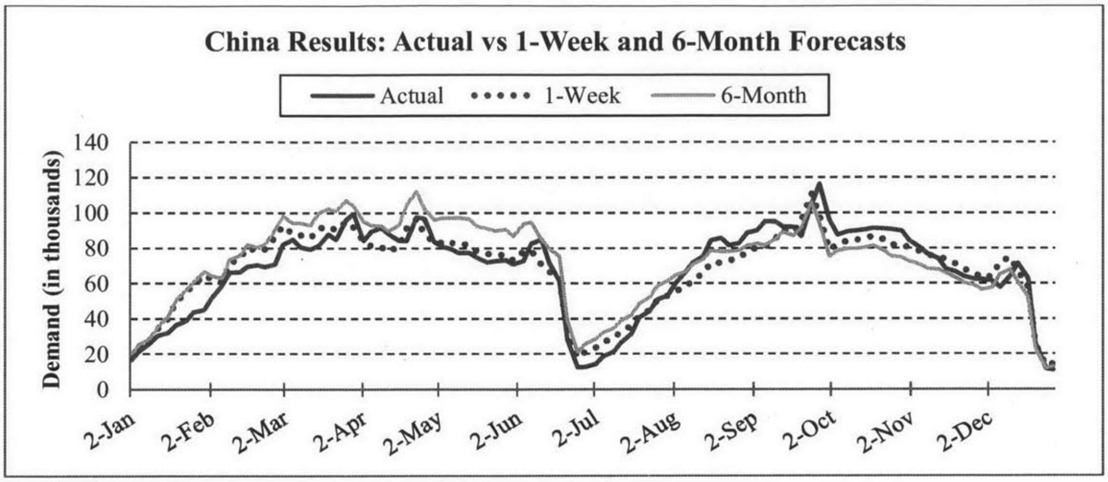
	Baseline	MAPE by Forecast Horizon			
Country	Error	1 week	1 month	3 month	6 months
Spain	20.8%	13.4%	15.0%	16.2%	17.4%
Belgium	20.2%	13.3%	13.8%	13.7%	13.9%
China	39.1%	18.8%	21.2%	23.7%	27.2%
Saudi Arabia	42.8%	25.3%	26.3%	25.4%	24.9%

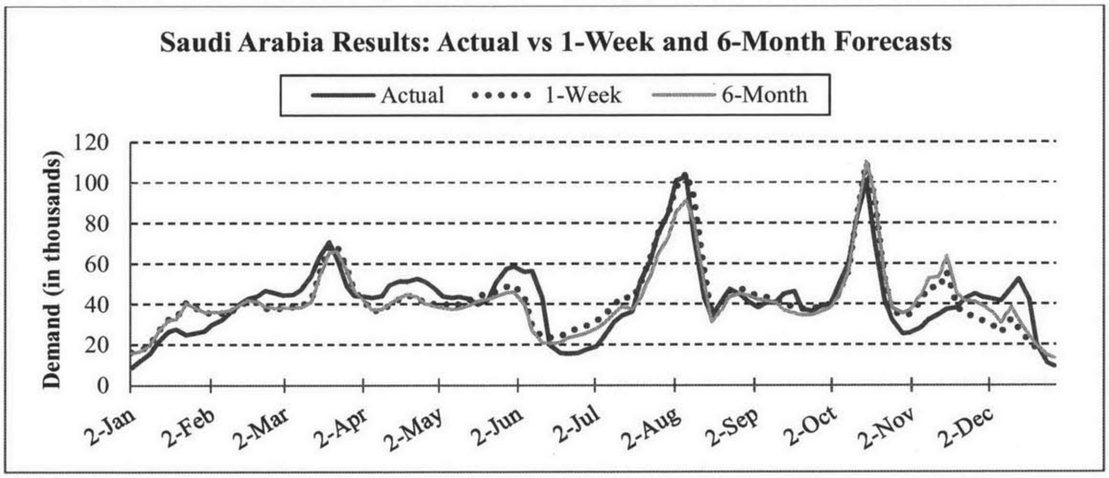
The plots in Figure 5.3 further validate how the forecasts closely match the observed demand. There are some instances where the forecast doesn't match the actual values as well as would be desired, such as the peak around April 2nd for Belgium, but it is believed that as the nature of the events causing these peaks are better understood the estimates will also improve. The following section provides an overview of additional variables that were tested but were ultimately not iincluded into the final models.

Figure 5.3 – Graphical Results for Short and Long Term Store Demand









5.2.3 Failed Variables: Store Level Forecasts

Arriving at the model in the previous section came after extensive exploration of variables. The variables described here are those that were tested but not included in the final model either because they did not yield coefficients that matched the economic expectation or because they improved forecast accuracy by so little that it did not warrant adding it to the model.

Variety and Age of Models in Store

One factor believed to have a positive impact on sales were to have a lot of new items at the store as it would entice more customers to buy the new "hot" items. The second factor would be having a lot of different models in the store, as high variety would lead to more opportunities for customers to find something they like. Therefore, the following variables were tested:

Equation 5.6 - Ratio of Models in Current Year vs Previous Year

$$MC_{s,w}^{Ratio} = MC_{s,w}^{CY} / MC_{s,w}^{PY}$$

Equation 5.7 - Ratio of New Models in Current year vs Previous Year

$$New.MC_{s,w}^{Ratio} = New.MC_{s,w}^{CY}/New.MC_{s,w}^{PY}$$

Equation 5.8 - Difference in Percentage of New Models in Current Year vs Previous Year

$$New.MC_{s,w}^{Delta} = New.MC_{s,w}^{CY}/MC_{s,w}^{CY} - New.MC_{s,w}^{PY}/MC_{s,w}^{PY}$$

Equation 5.9 - Different in Percentage of On-Sale Models in Current Year vs Previous Year

$$Sale.MC_{s,w}^{Delta} = Sale.MC_{s,w}^{CY}/MC_{s,w}^{CY} - Sale.MC_{s,w}^{PY}/MC_{s,w}^{PY}$$

where MC refers to a unique design, CY stands for Current Year, PY stands for Previous Year, New.MC refers to models that have been in store for less than one week, and Sale.MC refer to models that have been marked down.

Table 5.5 shows the coefficient estimated for each one of these variables when adding each one individually to the model defined by Equation 5.5. While all of the coefficients turn out to be significant, none of them improve the out of sample R^2 which tells us that adding the variable does not improve the forecast accuracy of the model. Furthermore, evaluating the change in wMAPE we see that even the best performing variable has a negligible improvement on forecast accuracy. In an effort to keep the model as simple as possible, variables that did not improve forecast accuracy were not included in the final model.

Table 5.5 - Results for Failed Variables: MCs in Store

Variable	Coefficient	Std. Error	Significance	Δ in Out-Sample R^2	Δ in wMAPE
MC Ratio	0.0020	0.00007	***	-0.0019	-0.0016
New MC Ratio	(0.0002)	0.00004	***	-0.0006	0.0003
New MC Pct. Delta	(0.0098)	0.00052	***	-0.0031	0.0011
On-Sale MC Pct. Delta	0.0007	0.00006	***	-0.0013	0.0006

Variety and Age of Inventory in Store

Similar to the exercise of looking at the models, or MCs, present in the store the distribution of inventory was also considered and the following variables were tested:

Equation 5.10 - Ratio of New Inventory in Current year vs Previous Year

$$New.Inv_{s,w}^{Ratio} = New.Inv_{s,w}^{CY}/New.Inv_{s,w}^{PY}$$

Equation 5.11 - Difference in Percentage of New Inventory in Current Year vs Previous Year

$$New.Inv_{s,w}^{Delta} = New.Inv_{s,w}^{CY}/MC_{s,w}^{CY} - New.Inv_{s,w}^{PY}/MC_{s,w}^{PY}$$

where, New.Inv refers to the units of inventory that have been in store for less than one week. Again, we see results, in Table 5.6, where the variables are significant but the added variable has little to no impact on improving the forecast accuracy as measured by the out of sample R^2.

Table 5.6 - Results for Failed Variables: MCs in Store

Variable	Coefficient	Std. Error	Significance	Inc. Out-Sample R^2	Δ in wMAPE
New Inv. Ratio	0.00089	0.00004	***	0.0012	-0.0009
New Inv. Pct. Delta	0.00323	0.00044	***	-0.0017	-0.0008

Variables to Capture Effects of Temperature (Weather)

Another category of variables that was expected to have great explanatory power was centered around weather effects. The thinking being that when the weather is nice people are more likely to go out and shop. Furthermore, since the fashion season switches from winter to summer and from summer to winter before the weather reflects the season, it is believed that a long winter would hurt sales because the summer items would do really poorly from February-March. Variables that tracked average, maximum, and minimum temperature, in degrees Celsius, for the current week were added to the model in Equation 5.5. Each variable was tested one at a time to determine if it would add explanatory power to the model, i.e. increase the out of sample R^2. As

Table 5.7 shows, these variables were significant but had such small coefficients that their contribution to the demand estimate was negligible. Secondly, the out of sample R^2 actually went down with the introduction of each of these variables, suggesting that the fit to the data was not suitable for forecasting. A second set of temperature variables were tested, this second group attempted to compare the recently experienced weather to the historical norm for that time of the year. The thinking behind these variables was to capture if irregularities in weather would have a larger effect on demand than simply the temperature level. For example, 20 degrees Celsius might be warm for January, but considered cold in June. The three variables developed were as follows:

Equation 5.12 - Weather Variable: Current Weekly Avg. vs. Historical Weekly Avg.

$$\text{Avg T (week) vs. Hist Avg T (Week)} = Avg.Temp\ (Week) - Hist.Avg.Temp(Week)$$

Equation 5.13 - Weather Variable: Current Weekly Avg. vs. Historical Monthly Avg.

$$\text{Avg T (week) vs. Hist Avg T (Month)} = Avg.Temp\ (Week) - Hist.Avg.Temp(Month)$$

Equation 5.14 - Weather Variable: Current Weekly Avg. vs. Historical Monthly Avg. in March and April

$$\text{Avg T (week)vs. Hist Avg T (Month)} = [Avg.Temp\ (Week) - Hist.Avg.Temp(Month)] * W$$

where Avg. T (week) refers to the average temperature observed during the latest week of data available for the demand forecast, Hist. Avg. T (week/month) refers to the historical average temperature observed in years 2009 to 2011 for the week or month that matches the latest week of data available. And, W is a binary variable that is set to 1 for the months of March and April. This last variable is attempting to capture the effect of temperature differences during the months when warm weather is expected to have the biggest impact, which is the time when the season is changing from winter to spring. The results of testing each one of these three variables is also shown in

Table 5.7 and yields results similar to the simple temperature variables, where they are statistically significant but end up having a small negative impact on the out of sample R^2.

Table 5.7 - Regression Results from Temperature Variable Tests for Store Demand

Variable	Coefficient	Std. Error	Significance	Inc. Out-Sample R^2	Δ in wMAPE
Avg. Temp	0.00002	0.000002	***	-0.0063	-0.0007
Min. Temp	0.00002	0.000001	***	-0.0062	-0.0005
Max. Temp	0.00002	0.000002	***	-0.0060	-0.0007
Avg. vs Hist. (week)	0.00008	0.000006	***	-0.0065	-0.0001
Avg. vs Hist. (month)	0.00005	0.000004	***	-0.0060	-0.0003
Avg. vs Hist. (March & April)	0.00013	0.000013	***	-0.0095	0.0000

Given the disappointing results shown in

Table 5.7, none of these variables were included in the final model. It is nonetheless believed that weather does play a role in store demand, but perhaps a different specification of the variables would yield a more useful relationship. Some candidate variables could be the number of days with precipitation, or the number of "sunny" days.

5.3 Family Level Demand Models

5.3.1 Definition

The models for estimating family demand use regression to estimate the percentage of store demand that will be attributed to a given family. As a result, the outputs from the store level demand models are needed as inputs for the family level models. The general formula is as follows:

Equation 5.15 - Family Demand Model Framework

$$D_{f,w} = D_{s,w} \times P_{f,w}$$

where $D_{f,w}$ is demand, in units, for family f in week w, $D_{s,w}$ is demand, in units, for store s in week w and $P_{f,w}$ is the proportion of the store's demand that can be attributed to family f. The goal is to understand the dynamics of the store during a given week combined with historical demand pattern for each family in order to determine an expectation of demand for a product family.

Data from the top 13 product families were evaluated to develop a model to estimate $P_{f,w}$. These 13 families comprise 92% of total volume and are the most important for Zara.

Table 5.8 shows the families used and the percentage of total demand for a major store in Barcelona. The percentage of total demand varies from store to store, but these families are the most important across most stores and it is extremely unlikely that a store would have a family outside of this grouping that comprises major weight.

Table 5.8 - Product Families Used to Build Forecast Models

Family Name	English Name	Family ID	% of Total Sales[2]
Camiseta	T-Shirt	83	30%
Pantalon	Pants	73	16%
Camisa	Shirt	76	10%
Jersey	Sweater	82	7%
Chaqueta	Jacket	81	7%
Vestido	Dress	74	7%
Blasier	Blazer	75	3%
Blusa	Blouse	77	3%
Cazadora	Sporty (Sp.) Jacket	2791	2%
Panoleta/Foulard	Scarf	78	2%
Bermuda	Shorts	98	2%
Leggings	Leggings	92	2%
Abrigo	Coat	2794	1%
		Total	92%

The approach for estimating $P_{f,w}$ was to find a single model framework with parameters that were adjusted for each family and subset of stores. It is believed that different regions have different preferences for each family. For example, stores in northern Spain have a higher propensity to sell sweaters and jackets than stores in southern Spain because it is much colder and it rains more. Therefore the subgroups of stores were picked for areas that were expected to have similar purchasing patterns while also maintaining a reasonable level of stores to make the data processing manageable. The groupings selected for model development were as follows:

Table 5.9 - Store Subgroups for Family Models

Region[3]	Stores	Records
Barcelona A & B	5	70,686
Belgium A	5	68,786
China A & B	6	80,854
Saudi Arabia A & B	6	75,682

Regression was used to find the parameters for Equation 5.16 to estimate $\hat{P}_{f,w}$. The model was purposely built without an intercept term in order for the results to have a more intuitive interpretation. As structured, the historical performance serves as a baseline estimate which is then adjusted based on the other factors in the model. Alternatively, if modeled with an intercept, it would be harder to interpret the meaning of the coefficient for the auto-regressive term H.

[2] Totals for a sample store in Barcelona

[3] Letters A and B refer to the store classification of stores included where A stores are the top selling stores.

Equation 5.16 - Family Demand Proportion Regression Model

$$\hat{P}_{f,w} = b_1 H_{f,w} + b_2 E_{f,w-1} + b_3 I_{f,w} + \varepsilon$$

where $\boldsymbol{H_{f,w}}$ is the historical average proportion for the current week and the previous week, E is the error term for the previous week's forecast[4], and $\boldsymbol{I_{f,w}}$ is the difference in inventory between current year and prior year for the given week. The regression was trained on data from 2009 to 2011 for each of the subgroups listed in Table 5.9.The coefficients and fit statistics for each family in the Barcelona region are shown in tables Table 5.10 & Table 5.11.

Table 5.10 - Family Level Demand Coefficients for A and B Stores in Barcelona

	$H_{f,w}$: Historical Avg			$E_{f,w-1}$: Lagged Error			$I_{f,w}$: Inventory Diff		
Family	Coeff	S.E.	Sig	Coeff	S.E.	Sig	Coeff	S.E.	Sig
T-Shirt	1.002	0.002	***	0.619	0.022	***	0.060	0.007	***
Pants	0.999	0.002	***	0.603	0.023	***	0.039	0.005	***
Shirt	1.001	0.003	***	0.800	0.019	***	0.013	0.002	***
Sweater	0.995	0.004	***	0.693	0.020	***	0.007	0.001	***
Jacket	0.997	0.004	***	0.625	0.022	***	0.020	0.002	***
Dress	1.000	0.003	***	0.679	0.022	***	0.011	0.002	***
Blazer	1.004	0.004	***	0.750	0.019	***	0.004	0.001	***
Blouse	1.007	0.004	***	0.712	0.019	***	0.007	0.001	***
Sp. Jacket	1.011	0.006	***	0.604	0.027	***	0.017	0.001	***
Scarf	1.001	0.006	***	0.667	0.022	***	0.004	0.001	***
Shorts	0.996	0.006	***	0.713	0.019	***	0.003	0.000	***
Leggings	1.005	0.005	***	0.671	0.022	***	0.003	0.001	***
Coat	0.996	0.006	***	0.815	0.032	***	0.001	0.001	

Table 5.11 - Family Demand Model Validation Statistics

Family	F-Test	Std. Error	Adj. R2	Max VIF
T-Shirt	<.0001	0.024	0.994	1.41
Pants	<.0001	0.015	0.991	1.33
Shirt	<.0001	0.012	0.986	1.91
Sweater	<.0001	0.011	0.977	1.42
Jacket	<.0001	0.011	0.981	1.85
Dress	<.0001	0.010	0.984	1.25
Blazer	<.0001	0.007	0.975	1.31
Blouse	<.0001	0.005	0.975	1.68
Sp. Jacket	<.0001	0.009	0.973	1.27
Scarf	<.0001	0.005	0.957	1.89
Shorts	<.0001	0.005	0.956	1.07
Leggings	<.0001	0.006	0.958	1.42
Coat	<.0001	0.006	0.974	1.20

[4] $E_{f,w-1} = P_{f,w-1} - \hat{P}_{f,w-1}$

Interpreting the coefficients for each model it is clear that historical performance is the top predictor for current performance with values very close to 1. The error correction term has values in the 0.6 to 0.8 range which is a significant adjustment and helps the model adjust when it is far off. The last term has relatively small impact on the final estimates but it helps fine tune the result to account for differences in inventory levels in current year from previous year. The fit statistics show strong fit for the model and that there is little multicolinearity between the independent variables. The R^2 values are very high and as we show in the results section they seem to indicate that the model is over-fitting the data and as we show in Section 5.3.2 the out of sample R^2 and forecasting power are not as strong as the R^2 here would suggest.

5.3.2 **Results: Family Level Forecasts**

To measure the accuracy of the models, we tested them on an out of sample dataset which included data from year 2012. The first step taken to test the performance was to measure the out-of-sample R^2 for the $\hat{P}_{f,w}$ estimates. Table 5.12 shows the values for out of sample R^2 which in many cases drops below 0.500 which means that the models explain less than half of the variance of the observed demand.

Table 5.12 - Family Model Out of Sample R^2

Family	Barcelona A & B	Belgium A	Saudi Arabia A & B	China A & B
T-Shirt	0.889	0.839	0.360	0.588
Pants	0.703	0.574	0.763	0.130
Shirt	0.699	0.668	0.871	0.693
Sweater	0.887	0.905	0.856	0.901
Jacket	0.416	0.301	0.755	0.597
Dress	0.663	0.499	0.754	0.577
Blazer	0.263	-0.093	0.513	0.285
Blouse	0.699	0.768	0.327	0.494
Sp. Jacket	0.757	0.746	0.747	0.747
Scarf	0.720	0.582	0.142	0.692
Shorts	0.467	0.447	0.246	0.178
Leggings	0.864	0.771	0.568	0.821
Coat	0.803	0.856	0.578	0.594

While the R^2 results were somewhat discouraging, the next step was to evaluate whether the models would outperform the baseline methodology in terms of forecasting units of demand. The baseline methodology resembles Zara's current methodology of assuming performance from past two weeks predicts performance for upcoming week. To do this comparison, the forecast for each week is estimated using the parameters from the regression outputs for Equation 5.16 using all observed data that occurred prior to the week being forecasted. Therefore, each calculated demand value can be considered a 1-week forecast because it only uses information that would be available in real life when making a one week forecast. Figures Figure 5.4 and Figure 5.5 show that even though the models exhibit poor R^2 results, they generally produce better forecasting results than the baseline methodology. In particular, the model performs generally well except for families 2791 and 2794 where the forecast is worse than the baseline methodology, and these families only account for 3% of total demand.

Figure 5.4- Reduction in wMAPE for Family Demand Model

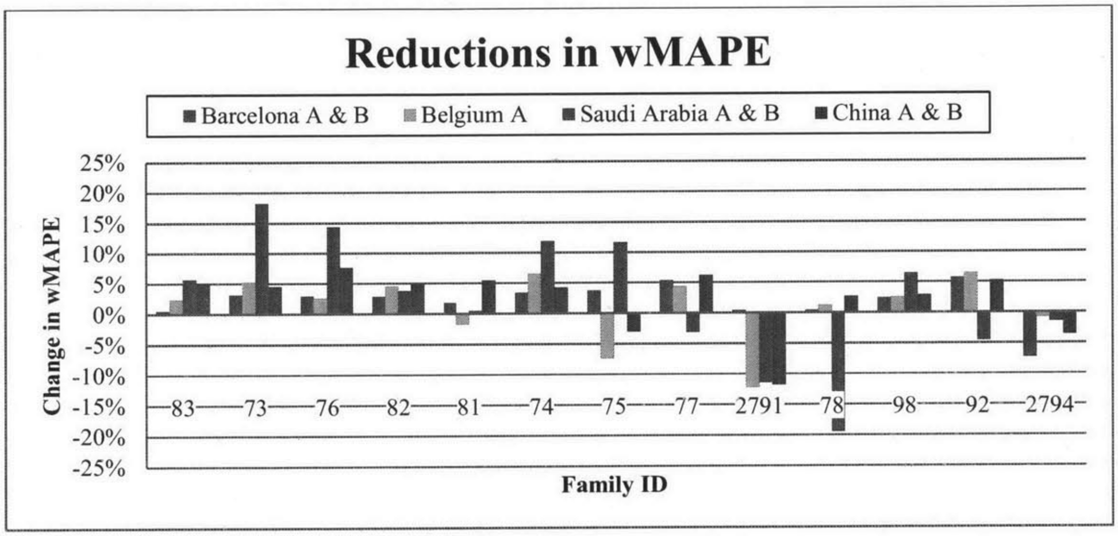
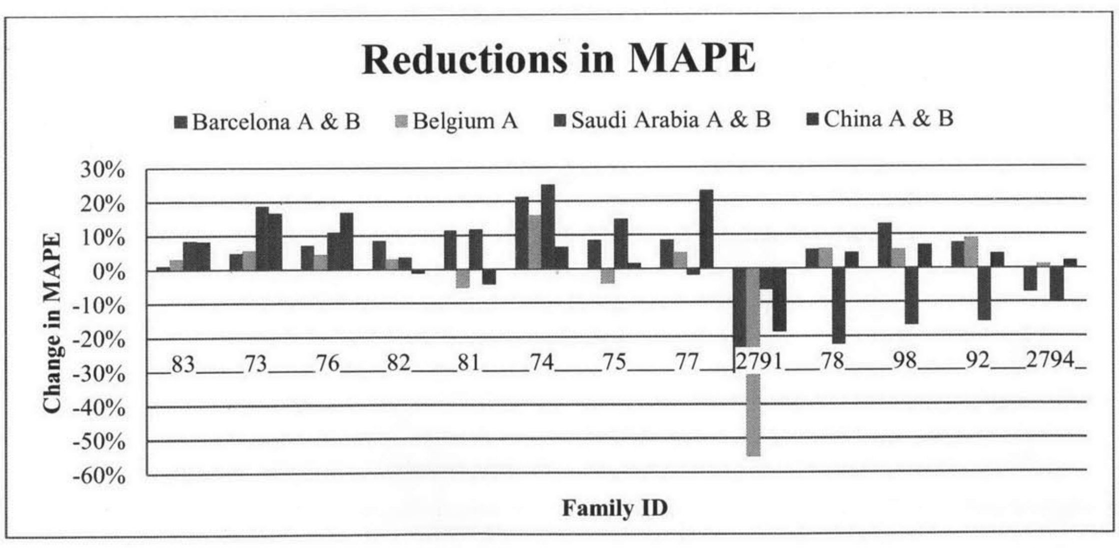


Figure 5.5 - Reduction in MAPE for Family Demand Model



The mixed results in the family demand model stems from the fact that Zara designers may decide to change how strongly they feature a given family at any given moment. As discussed in section 4.3.5, this phenomenon is very difficult to account for by using historical demand behavior. The mixed results in performance are magnified when forecasting demand for several weeks into the future. Figures Figure 5.6 and Figure 5.7 show how the forecasting power of these models can vary widely by family. However, the results were strong enough to be used as inputs for developing models for the MCC level demand. Those models are discussed in the following section and would of course also improve with an improvement in the family level models.

Figure 5.6 - Family Long Term Forecast Accuracy: Good Example (Pants)

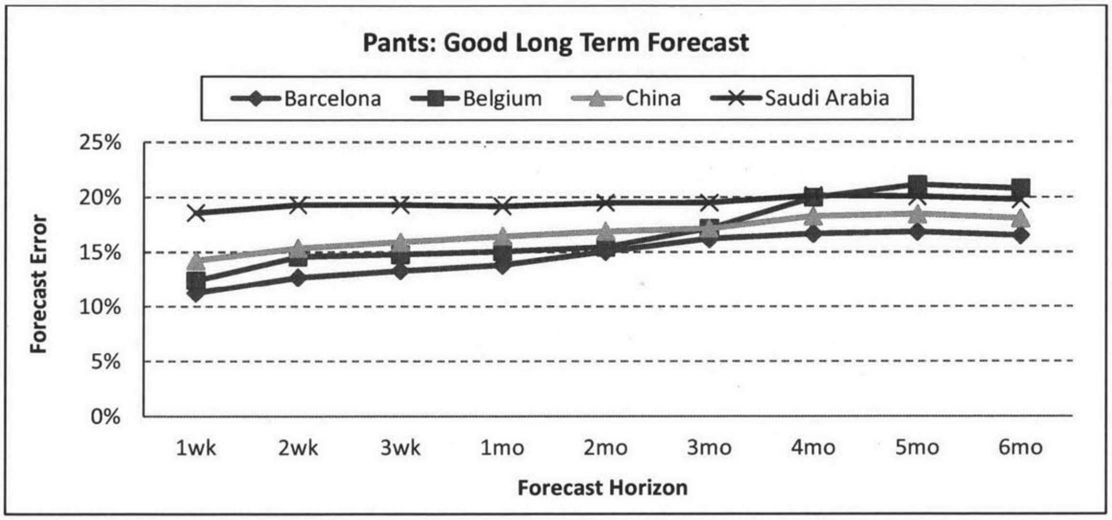
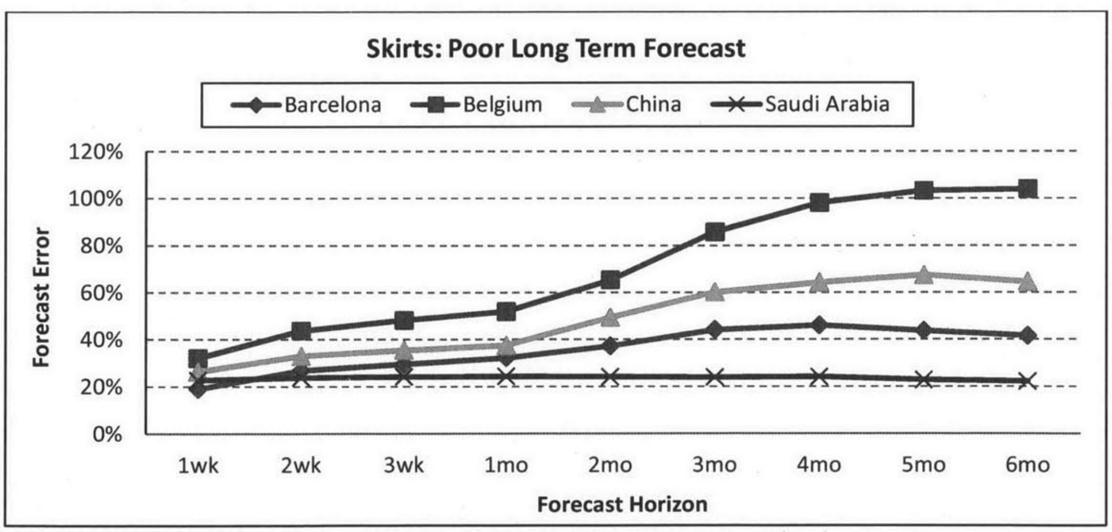


Figure 5.7 - Family Long Term Forecast Accuracy: Poor Example (Skirts)



5.3.3 Failed Variables: Family Level Forecasts

Similar to the effort for developing the store level demand model, many variables were tested in addition to the ones that made it to the final model. Those variables are detailed here as well as the regression results from attempting to incorporate them to the models. Again, these variables were excluded because they either did not yield a coefficient that matched the real world economic intuition or because they resulted in such a small improvement that they did not warrant the increase in complexity.

Variety and Age of Models in Store

One factor believed to have a positive impact on sales were to have a lot of new items at the store as this could entice more customers to buy the new "hot" items. The second factor could be having a lot of different models in the store, as high variety would lead to more opportunities for customers to find something they like. Therefore, variables that attempt to capture the increased desirability of a given family were developed.

Equation 5.17 – Year over Year Change in Family Weight within Store

$$F_{f,w}^{Ratio} = F_{f,w}^{CY} / F_{f,w}^{PY}$$

Equation 5.18 – Year over Year Change in Weight of New Items in Family within Store

$$New.F_{f,w}^{Ratio} = New.F_{f,w}^{CY} / New.F_{f,w}^{PY}$$

Equation 5.19 – Year over Year Change in Weight of New Items in Family within Store

$$New.FF_{f,w}^{Ratio} = New.FF_{f,w}^{CY} / New.FF_{f,w}^{PY}$$

where F refers to the percentage of MCs within a store that are part of family f in week w, New.F refers to the percentage of MCs within a store that are NEW and are part of family f in week w. New.FF refers to the percentage of MCs within a family that are new in week w. Superscripts CY and PY refer to current year and historical average.

Table 5.13 - Family Results for MC Variables

Family	%MC Difference			%MC (New) Difference			%MC (New in Fam.) Difference		
	Coeff.	Sig.	Δ in Out-Sample R^2	Coeff.	Sig.	Δ in Out-Sample R^2	Coeff.	Sig.	Δ in Out-Sample R^2
T-Shirt	0.0071		(0.003)	0.3832	**	(0.003)	0.0470	*	(0.000)
Pants	(0.0083)		(0.005)	0.3747	***	0.009	0.0731	***	0.007
Shirt	0.0074	***	0.008	0.6106	***	0.028	0.0631	***	0.020
Sweater	(0.0003)		(0.001)	0.4240	**	0.009	0.0211	***	0.001
Jacket	(0.0029)	*	(0.004)	0.3214	*	0.014	0.0069		0.008
Dress	0.0131	***	0.038	0.4231	***	0.010	0.0478	***	(0.002)
Blazer	(0.0035)	*	(0.028)	0.1756	*	0.031	0.0144	**	0.034
Blouse	0.0053	***	0.002	0.2984	***	0.007	0.0109	***	0.001
Sp. Jacket	0.0004		(0.000)	0.5353	**	0.022	0.0042		0.004
Scarf	0.0014	*	(0.006)	0.9296	***	(0.036)	0.0220	***	(0.022)
Shorts	(0.0034)	**	0.014	0.3132	***	(0.016)	0.0158	***	(0.010)
Leggings	0.0042	**	(0.005)	0.9941	***	0.013	0.0160	***	0.006
Coat	(0.0001)		0.000	1.4248	***	(0.007)	0.0073	***	(0.001)

Table 5.13 clearly shows that none of these three variables had consistent contributions to the out of sample R^2 across all families. The same model was desired for all families in order to reduce the number of specifications needed to maintain the system needed to forecast all families.

Variables to Capture Effects of Temperature (Weather)

The same variables for weather that were defined in section 5.2.3 were tested for family demand. It is widely believed that temperature should have an even greater impact when observed at a regional level and for a particular product family. For example, as winter turns to spring, it would be expected that T-Shirts would begin to sell more as customers look for lighter clothing. Therefore the weather variables were tested for all subfamilies in the Barcelona region to see if it would improve the models. Table 5.14 shows the results for family 83 and the results are even less encouraging than they were for the store models. Most of them are not significant, and even the ones that are, don't have a positive impact on out of sample R^2 and thus making them poor variables for forecasting.

Table 5.14 - Results for Weather Variables on Family 83 in Barcelona

Variable	Coefficient	Std. Error	Significance	Δ in Out-Sample R^2
Avg. Temp	-0.00016	0.000087	.	-0.0001
Min. Temp	-0.00012	0.000081		-0.0001
Max. Temp	-0.00012	0.000084		-0.0002
Avg. vs Hist (week)	0.00102	0.000392	**	-0.0033
Avg. vs Hist (month)	0.00074	0.000287	*	-0.0008
Avg. vs Hist (March & April)	0.00122	0.000778		-0.0016

5.4 MCC Level Demand Models

5.4.1 Definition

The models for estimating MCC demand uses regression to estimate the percentage of family demand (within one store) that will be attributed to a given MCC. As a result, the outputs from the family level demand models are needed as inputs for the MCC level models. The general formula is as follows:

$$D_{m,w} = D_{s,w} \times P_{f,w} \times P_{m,w}$$

where $D_{m,w}$ represents the demand for MCC m in week w, and $P_{m,w}$ refers to the percent of demand for MCC m within its product family. $D_{s,w}$ is the total demand for the store which is defined in section 5.2 and $P_{f,w}$ is the proportion of family demand which is defined in section 5.3. $P_{m,w}$ is estimated using regression based on data from 2009 to 2011 and validated on 2012 data. Given the large number of MCCs that Zara introduces each year, the data was partitioned by store and family, such that parameters were estimated for each store-family combination. Table 5.15 shows the number of records available for each of the subsets modeled.

Table 5.15 - Data Records by Store-Family Combination

		MCC Records by Family ID						
Country	Store ID	T-Shirt	Pants	Shirt	Sweater	Jacket	Dress	Blazer
Spain	"A"	481,000	269,860	23,662	159,032	147,572	206,994	103,367
Spain	"B"	399,150	189,923	179,328	125,411	111,750	171,737	85,032
Belgium	"C"	432,171	208,168	208,058	139,707	135,001	201,255	93,896
China	"D"	316,086	143,451	134,580	94,621	86,871	121,662	59,607
Saudi Arabia	"E"	326,773	154,796	157,123	76,239	70,371	140,360	50,331

Similar to the model for family demand, this model develops a regression to estimate $P_{m,w}$. In order to limit the number of specifications the goal was to find a single specification that would work relatively well for all store-family pairings. The equation below shows the regression model and Table 5.16-Table 5.20 show the resulting parameters.

$$\hat{P}_{m,w} = b_1 L_{m,w} + b_2 R_{m,w} + b_3 L_{m,w} R_{m,w} + b_4 Age0 + b_5 Age1 + b_6 Age2 + b_7 Age6 + b_8 Age11 + Age21 + \varepsilon$$

Where L represents the average demand for MCC m observed during the three cycles preceding week w, R represents the ratio of new MCCs in week w to total MCCs with sales in the preceding w, and the Age variables are binary variables that identify if MCC m will be new (0), 1 week old, 2-5 weeks old, 6-10 weeks old, 11-20 weeks old or older than 20 weeks during week w.

The expected coefficient for each of these variables is perhaps less intuitive than for the previous models. Variable P is used as the baseline estimate for each MCC, which is to say that the model can be thought of as starting with the observed average demand over the previous 3 cycles. Variable R represents the cannibalization effect of new items arriving at a store. We would expect a lot of new arrivals to depress the demand for other items previously in the store. The interaction of these two variables captures the compound effect of how much observed average demand should be depressed when large volumes of new items enter. For example, we would expect the depression to be larger when there are a lot of arrivals than when there few. We would also expect that MCCs with large share of family demand would be more impacted by new items than items with small shares. We can think of it as hot items facing competition for first time. The last set of binary variables, capture the drop in demand that MCCs experience as their time in stores grows. Table 5.16 shows the coefficient values for the model when run for the largest store in Spain.

Table 5.16 - MCC Demand Coefficients for Store 160 in Spain

Family	b1	b2	b3	b4	b5	b6	b7	b8	b9	R2
T-Shirt	0.8729	0.0001	-0.3374	0.0007	0.0002	0.0001	0.0002	0.0001	0.0001	0.6876
Pants	0.8397	0.0004	-0.6207	0.0016	0.0009	0.0008	0.0004	0.0003	-0.0002	0.7194
Shirt	0.7720	0.0002	-0.5005	0.0025	0.0015	0.0010	0.0005	0.0002	-0.0002	0.6581
Sweater	0.8326	0.0001	-0.1448	0.0030	0.0016	0.0008	**0.0000**	**0.0001**	-0.0003	0.6879
Jacket	0.8235	0.0001	-0.1548	0.0027	0.0014	0.0009	0.0001	0.0005	**-0.0002**	0.6914
Dress	0.6927	0.0001	-0.4157	0.0028	0.0019	0.0011	0.0003	0.0001	**-0.0001**	0.5796
Blazer	0.7532	0.0005	-0.4943	0.0043	0.0025	0.0021	0.0008	0.0004	**0.0003**	0.6192

Insignificant (p >.10) in bold

Evaluating the coefficients, we see that b1, b2, and b3 show the results that we would expect. First, b1 is positive and is less than one. This indicates that the proportion of demand observed in the previous week is a good baseline estimator for current week proportion, but that it must be discounted. Then, we see that b_2 and b_3 have a net negative effect on demand, which says that high volumes of new MCCs can significantly decrease the expected proportion in the upcoming week. Then the age coefficients b_4-b_9 for the most part decrease in size as the item gets older, which is what we would expect given the sharp decline in demand that Zara's product experience as their time in the store increases. The R^2s are a bit lower than is generally desired, but given soundness of the rest of the coefficients and the high variability of demand across different MCCs and across different weeks the models are quite strong. Furthermore the results section shows a performance measure that ties close to the operational impact that this model could have. Table 5.17 to Table 5.20 show the regression results for the other stores where the model was tested.

Table 5.17 - MCC Demand Coefficients for Store 30 in Spain

Family	b1	b2	b3	b4	b5	b6	b7	b8	b9	R2
T-Shirt	0.8141	0.0001	-0.2925	0.0007	0.0004	0.0002	0.0003	0.0002	0.0002	0.6398
Pants	0.8091	0.0004	-0.4592	0.0015	0.0011	0.0009	0.0006	0.0004	-0.0001	0.6626
Shirt	0.7338	0.0001	-0.2538	0.0025	0.0016	0.0012	0.0004	0.0002	-0.0002	0.6229
Sweater	0.8415	0.0002	-0.1364	0.0031	0.0016	0.0006	0.0002	-0.0002	0.0007	0.6538
Jacket	0.7902	0.0002	-0.1407	0.0022	0.0012	0.0013	0.0003	0.0007	0.0005	0.5945
Dress	0.6968	0.0001	-0.3172	0.0029	0.0019	0.0010	0.0003	0.0001	**-0.0002**	0.5594
Blazer	0.8585	0.0006	-0.5168	0.0036	0.0013	0.0012	**0.0001**	-0.0002	**-0.0001**	0.6081

Insignificant (p >.10) in bold

Table 5.18 - MCC Demand Coefficients for Store 376 in Belgium

Family	b1	b2	b3	b4	b5	b6	b7	b8	b9	R2
T-Shirt	0.7044	0.0002	-0.2049	0.0012	0.0005	0.0004	0.0005	0.0004	0.0007	0.5269
Pants	0.7930	0.0005	-0.3744	0.0017	0.0009	0.0009	0.0006	0.0003	**0.0000**	0.6516
Shirt	0.6293	0.0003	-0.2159	0.0033	0.0018	0.0014	0.0009	0.0005	-0.0002	0.5159
Sweater	0.7878	0.0002	-0.1107	0.0027	0.0015	0.0015	0.0008	0.0004	**0.0001**	0.6549
Jacket	0.9126	0.0003	-0.2115	0.0016	0.0005	0.0004	0.0002	**0.0000**	-0.0002	0.7381
Dress	0.7377	0.0001	-0.1339	0.0020	0.0008	0.0005	0.0002	-0.0001	-0.0004	0.5963
Blazer	0.6227	0.0006	-0.1637	0.0055	0.0038	0.0029	0.0018	0.0013	0.0006	0.5144

Insignificant (p >.10) in bold

Table 5.19 - MCC Demand Coefficients for Store 3832 in China

Family	b1	b2	b3	b4	b5	b6	b7	b8	b9	R2
T-Shirt	0.5684	0.0001	-0.0800	0.0020	0.0012	0.0008	0.0007	0.0006	0.0003	0.4539
Pants	0.6186	0.0002	-0.0352	0.0044	0.0028	0.0021	0.0014	0.0008	**0.0001**	0.5357
Shirt	0.5241	0.0003	-0.1446	0.0068	0.0049	0.0033	0.0020	0.0013	**0.0001**	0.435
Sweater	0.4868	0.0001	0.0224	0.0071	0.0064	0.0038	0.0028	0.0030	0.0066	0.2366
Jacket	0.3660	0.0002	-0.0057	0.0080	0.0078	0.0044	0.0040	0.0041	0.0046	0.2817
Dress	0.5731	0.0003	-0.1003	0.0065	0.0039	0.0020	0.0009	**0.0000**	**-0.0002**	0.4535
Blazer	0.5232	0.0008	-0.0798	0.0105	0.0079	0.0047	0.0034	0.0015	0.0025	0.4149

Insignificant (p >.10) in bold

Table 5.20 - MCC Demand Coefficients for Store 3161 in China

Family	b1	b2	b3	b4	b5	b6	b7	b8	b9	R2
T-Shirt	0.7591	0.0001	0.0138	0.0014	0.0004	0.0002	0.0001	0.0001	0.0002	0.5871
Pants	0.7162	0.0006	-0.1521	0.0034	0.0017	0.0015	0.0009	0.0006	0.0002	0.5201
Shirt	0.6371	0.0006	-0.1740	0.0053	0.0030	0.0020	0.0012	0.0006	**0.0001**	0.4677
Sweater	0.5178	0.0006	-0.0130	0.0100	0.0069	0.0046	0.0029	**0.0001**	**0.0013**	0.1896
Jacket	0.6226	0.0001	-0.0059	0.0078	0.0036	0.0038	0.0020	0.0014	**0.0000**	0.4675
Dress	0.5886	0.0002	-0.0676	0.0062	0.0032	0.0017	0.0009	0.0004	**0.0004**	0.4349
Blazer	0.8911	0.0016	-0.1919	0.0056	0.0020	0.0001	0.0010	0.0008	-0.0017	0.6797

Insignificant (p >.10) in bold

5.4.2 **Results: MCC Level Forecasts**

Lost sales, as defined in Section 4.2, was used to measure the difference between using a baseline model that closely matches the current approach to forecasting demand to the method described in Section 5.4.1. The baseline method was in line with the other baseline methods used in store and family demand forecasts, which is to use the average demand from the previous two weeks as the estimate of demand for upcoming week.

The coefficients from Section 5.4.1 were estimated using data from 2009 to 2011 and tested on data from 2012. The models were used to estimate the proportion of demand that each MCC would comprise within its respective family, then the forecasted results for the family demand were used to transform the forecasted MCC proportion to a unit value. Then, an ideal inventory level was calculated using the MCC demand forecast. The ideal inventory was calculated as the necessary level required for three weeks of demand, which is an approach that is in line with current methodology (see Section 2.3.2). Table 5.21 shows the net change in lost sales that would have resulted if the models in this thesis had been used to replenish Store 160 throughout year 2012.

Table 5.21 - Estimated Net Change in Lost Sales for Store 160 in 2012

Family	2 Week Avg	MIT9	Net Change
T-Shirt	13,748	10,859	-2,889
Pants	7,177	4,728	-2,449
Shirt	8,645	7,032	-1,613
Sweater	7,151	6,635	-516

Jacket	3,712	3,488	-224
Dress	4,449	4,024	-425
Blazer	2,234	1,689	-545
Total	47,115	38,454	-8,661

The reductions in lost sales are significant across all major product families and for store 160, the total reduction in lost sales of 8,661 units represents close to 1% of the store's total annual demand. Furthermore, the reductions in lost sales were even more signification for stores in other countries. Table 5.22 shows a summary of lost sales across the stores tested and breaks out the changes in units per family, units per store, and the percentage that the change represents relative to the store's annual demand.

Table 5.22 - Summary of Reductions in Lost Sales Across Stores and Families

Family	BCN - 160	BCN - 30	BEL-376	CHN - 3832	ARA - 3161	Total
T-Shirt	-2,889	-4,340	-2,542	-2,922	-1,329	
Pants	-2,449	-2,454	-2,153	-1,887	-2,122	
Shirt	-1,613	-1,404	-1,325	-1,674	-2,059	
Sweater	-516	59	-886	-1,374	-915	
Jacket	-224	-265	24	-759	-443	
Dress	-425	-775	-389	-931	-1,032	
Blazer	-545	-542	-259	-363	-405	
Total Δ in Lost Sales	-8,661	-9,721	-7,529	-9,910	-8,305	-44,126
Store Demand	979,792	789,580	397,070	178,012	160,854	2,505,308
Drop in Lost Sales as % of Store Demand	0.88%	1.2%	1.9%	5.6%	5.2%	1.8%

The potential drops in lost sales show strong incredible promise for Zara's operations. Especially encouraging is the fact that stores in China and Saudi Arabia could potentially stand to gain the most from this new forecasting methodology. The result is also consistent with intuition given that Zara's existing systems have been developed with Spain and European markets in mind. As Zara continues to expand, however, it could clearly benefit from having a more robust and sophisticated method to forecasting demand.

6 Implementation

6.1 Potential Tools

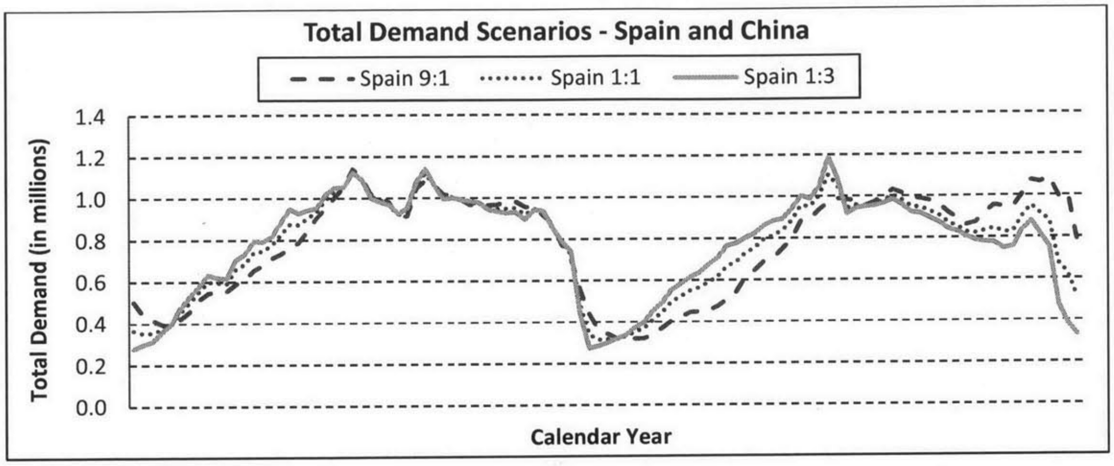
6.1.1 Store Demand Forecasting

One of the biggest advantages of the simple model framework for store demand described in sections 5.2.1.2 & 5.2.1.3 is that they can be easily forecasted. The first component, annual demand estimate, is a single number that reflects the expected annual sales given recent performance which only requires observed data as an input and is easy to calculate. The second component, relative demand, is even more convenient for forecasting given that all of its independent variables are known at the beginning of the year. The holiday events are known, and the historical demand averages are calculated from prior calendar years. This model could clearly help them gain visibility for medium and long term time horizons.

In the short to medium term, 1 to 6 weeks, Zara can use the forecasts from this model to estimate the expected trend in demand. Which means they can estimate how much demand will increase or decrease over the next several weeks and adjust their replenishments accordingly. For events where they anticipate sharp peaks, Zara sends extra inventory in the weeks preceding the event in order to avoid having to send one unreasonably large shipment to the stores. This model could help them anticipate those shocks and plan accordingly.

In the long term, 4-6 months, Zara can use this forecast to map out its global demand. By aggregating forecasted demand from all stores it becomes possible to identify peaks in demand that are outside of the normal peaks they prepare for (Easter, late September, Christmas) and instead understand if the demand in China for Golden Week or Chinese New Year is large enough to impact their global purchasing plans. Figure 6.1 shows a simulated transition where Spain:China demand ratio changes from 9:1 to 1:3. The end of year peak clearly loses importance and fall peak becomes more pronounced.

Figure 6.1 - Simulated Transition in Global Demand as China's Volume Grows



A tool of this nature would be the first of its kind at Zara and would allow them to view demand in a global scale for the first time. In addition, this model also lends itself to identifying events that have significant impact on demand. Any events that result in significant coefficients when used as independent variables in the model can be classified as important. These dates and events can be documented and used to alert Zara's staff of upcoming events.

6.1.2 MCC Demand Forecasting Tool

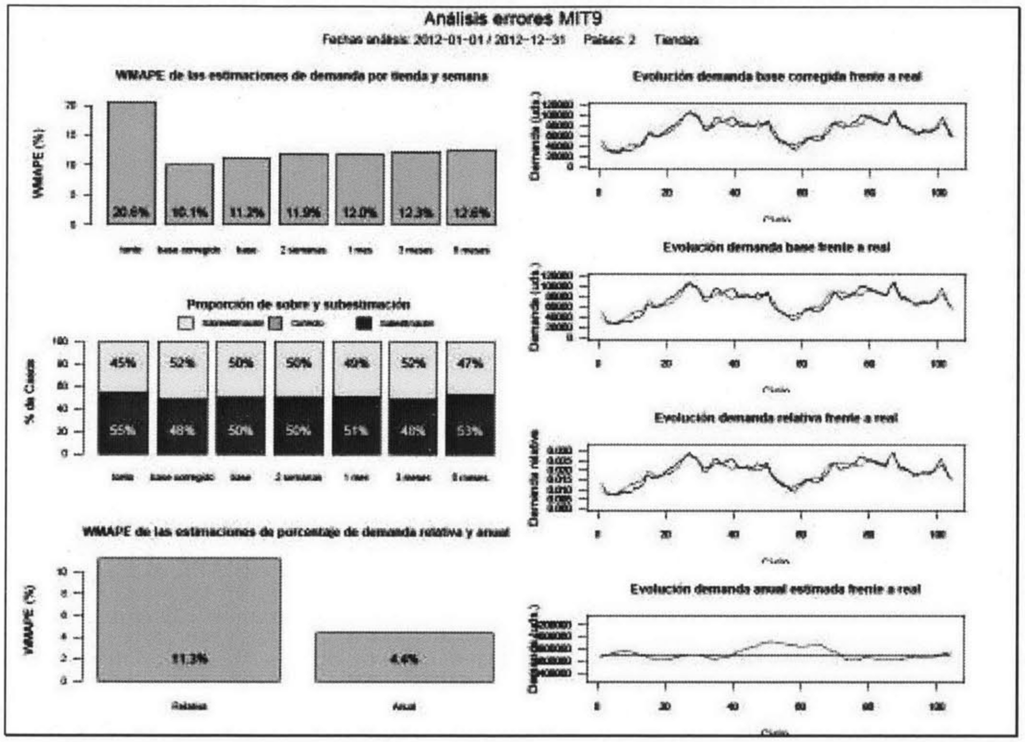
The model framework proposed in section 5.4 could easily be incorporated into Zara's current process. Currently, they use past two weeks of demand as an indication of how much demand to expect in the following week and then adjusting the value depending on user expectations for seasonal impacts (see Section 2.3.2). Instead, Zara could use the model proposed in this thesis to calculate expected demand in one step and it could just plug in the value into its current processes.

6.2 Stakeholder Buy-in

The results documented in this thesis were presented in August 2013 to a large audience of stakeholders at Zara which included managers from Distribution, Financial Controllers, Country Managers and IT. The collective group showed enthusiasm about the results and from there the Distribution department has taken the lead in further testing and improving the methods in this thesis.

In similar fashion to the development of this project, the implementation work has begun with the store level models and will later move to the family and MCC models. As of this writing, Zara has prepared its data systems to generate a six month store demand forecast in real time. Their next move is to monitor the results as demand shocks occur around the world and if the results are promising then they can incorporate them into their automated replenishment processes. Figure 6.2 shows the report that Zara generates when evaluating the forecast for a set of stores. The report includes short and long term wMAPE errors, over and underestimation percentages, and visual representations of the predicted versus actual demand values for weekly and annual demand.

Figure 6.2 - Evaluation Report for Store Demand Forecasts



The first application envisioned for this model is to estimate the trends in near term store demand, and subsequently implement models for product families and MCC demand. The distribution department has taken ownership of the models and is the most direct beneficiary of the project, if successful. However, the applications could later be expanded to designers and buyers so they can better match their product introductions with customer demand.

6.3 Model Modifications

Since the completion of this project, Zara has modified the model to forecast the weekly proportion of store demand. The approach is mostly the same, except that instead of including historical average and dummy variables for cycles surrounding an important event, they take the historical average demand as the baseline demand and then use regression only for measuring the impact of important events. The model is defined as follows:

Equation 6.1 - Zara Modified Store Demand Model

$$P = H + S$$

$$\hat{S} = \sum_{1}^{N} b_i c_i + \epsilon$$

$Where:$

$P: Proportional\ Demand\ as\ percentage\ of\ annual\ demand$

$H: Historical\ Average\ proportion$

$S: Seasonal\ Component\ of\ Demand$

$bi: Parameter\ determined\ through\ regression$

$ci: Cycle\ adjacent\ to\ the\ event$

$\epsilon: Error\ Term$

This modification to the model takes advantage of the fact that the coefficient for historical average demand in the original model had values very close to 1 for most of the countries tested. By isolating the seasonal impact of each event, this model can more accurately isolate the changes in demand for time periods surrounding an important event. This modification improved the wMAPE overall error and also improved the alignment of forecasted vs actual demand during important events. For example, the wMAPE error for Belgium went from 13.3% down to 11.8% when measured during the weeks surrounding Easter.

7 Contributions

7.1 Improvements to Zara's Operations

This study developed new forecasting methods for demand on three levels: Store level, family level, and MCC level. The methods developed were tested across four important markets: Spain, Belgium, China and Saudi Arabia. The results showed marked improvements over Zara's current forecasting methods: For example forecasting error for store level demand could be reduced between 6% and 11% for 1 week horizons (see Table 7.1); Family forecasts could be improved for 11 out of 13 largest product families (see Figure 5.4); And MCC level forecasts could help reduce lost sales by 24% (see Table 7.2).

Measuring the impact of having better forecasts is difficult because many of the potential applications would be completely new. For example, Zara does not use any long term forecasts when planning purchases so having a forecast that they can trust would undoubtedly be helpful but quantifying the savings due to better purchase planning is very difficult. Calculating an impact in dollars is a slightly easier exercise for measuring the impact of the MCC level demand models, because in this case estimates of lost sales can be translated into potential revenue.

Table 7.1 shows forecast errors for the store-level demand models developed in this study, MIT9, compared to a baseline methodology. The error is measured as the weighted mean absolute percentage error, or wMAPE, which is detailed in Chapter 5. The errors for the MIT9 models are shown for 1 week and 24 week forecasts which represent the errors from forecasting demand for 1 week or 24 weeks in the future. The 1-week baseline model refers to simply using performance from the past two weeks as the expectation for the upcoming week, which resembles the current methodology. The 24-week baseline model uses prior year demand for week in question. The results show significant improvement for 1-week forecasts and 24-week MIT9 forecasts for all countries evaluated.

Table 7.1- Store Level Forecast Improvements in wMAPE

	Estimation Errors (Stores)					
Country	Baseline 1 Week	MIT9 1 Week	Difference	Baseline 24 Weeks	MIT9 24 Weeks	Difference
Spain	17.6%	11.6%	-6.0%	21.9%	15.2%	-6.7%
Belgium	17.3%	11.4%	-5.9%	14.0%	12.5%	-1.5%
China	17.9%	12.9%	-5.0%	22.1%	19.0%	-3.1%
Saudi Arabia	31.3%	19.7%	-11.6%	31.2%	20.2%	-11.0%

Table 7.2 shows the estimated reduction in lost sales that could result from using the models outlined in this study to calculate replenishment quantities. The estimated replenishment quantities are the result of following the modeling sequence of forecasting store demand, then family level demand and finally the MCC level demand. Due to the large volume of data that needs to be processed in order to calculate these values, the author selected five representative stores to test the potential of the models. The reductions are calculated by comparing the differences between demand and in-store inventory observed versus the differences that would result from using the proposed models. The proposed models were constrained to use the same total replenishment inventory as was sent in reality in order to avoid situations where one approach had lower lost sales but higher total inventory, which would make it ambiguous to determine which model performed better; unless, overage and underage costs are calculated. Here again, we see significant improvements across all the stores tested.

Table 7.2 - Potential Reduction in Lost Sales

Store	Reduction in Lost Sales	Reduction as % of Store Demand
BCN - 160	-18%	0.9%
BCN - 30	-22%	1.2%
BEL - 376	-23%	1.9%
CHN - 3832	-39%	5.6%
ARA - 3161	-24%	5.2%
Total	-24%	1.8%

Lastly, Table 7.3 shows an approximation to the revenue represented by the improvement in lost sales. To calculate this value, the average unit price by product family was multiplied by the total lost sales that could be reduced in one year (as shown in

Table 5.22). This value serves as a first order approximation to the potential revenue that Zara could gain from implementing MIT9 in the 5 stores tested.

Table 7.3 - Potential Revenue from Reductions in Lost Sales in 5 example stores for one year

Family	Avg. Price Per Unit[5]	Reduction in Lost Sales (units)	Gain in Revenue
T-Shirt	X1	14,021	Y1
Pants	X2	11,065	Y2
Shirt	X3	8,075	Y3
Sweater	X4	3,633	Y4
Jacket	X5	1,667	Y5
Dress	X6	3,553	Y6
Blazer	X7	2,114	Y7
Total			$1,094,170

[5] Unit prices are omitted for confidentiality reasons

7.2 Potential Applications to Retail Industry

The methods outlined in this thesis can be easily replicated in any retail environment where sales data is carefully maintained, yearly demand follows a cyclical pattern with slow to moderate year over year same store growth, and sells a product where missed forecasts could lead to high mark downs.

The models detailed in Chapter 5 cover how to produce 4 to 6 month demand forecasts while accounting for events that cause sharp peaks in demand that shift across the year. Many retailers experience this demand pattern for events such as "Back To School", or "Black Friday", or "Mother's Day". These are events for which the dates of occurrence are known well in advance but the exact dates shifts in the calendar from year to year making it difficult to use historical data to forecast demand during those periods.

Furthermore, the issue of producing short term forecasts is becoming more imperative as more companies attempt to follow Zara in the fast fashion model. Accounting for cannibalization and product lifecycle demand as this thesis does could help many retailers account for the effect that new arrivals will have on items that are currently available for sale.

8 Future Direction

At the core of Zara's strategic advantage is its ability to adjust its product assortment to match customer demand. A central component for this strategy is the ability to use the observed customer behavior to properly inform their replenishment decision. Up to this point, Zara has been extremely successful at matching customer demand by maintaining simple forecast models and keeping an unparalleled awareness of what its customers' want. This is the first attempt at systematically incorporating historical demand patterns, especially with respect to events that shift in calendar dates from year to year. For the first time, Zara is now able to confidently produce a forecast to inform demand shocks on a 4 to 6 month time horizon. In addition to the long term forecasts, this study is also the first attempt at Zara to incorporate the cannibalization effect from new products when making short term forecasts for replenishment decisions.

The models in this thesis were developed using linear regression in an effort to maintain simplicity and to ensure that it could be easily explained and understood across the organization. Furthermore, the framework for each forecast quantity (Store, Product Family, or MCC) was the same across all stores and products. The parameters for the model were trained separately across regions and products, but the independent variables were the same. This again made it simpler to explain, but also has a strong implementation advantage in that once a model is setup for one situation, it is simple to expand to Zara's entire network.

While the contributions made in this thesis are significant and worthy of implementation, as presented throughout Chapter 6, there is still ample opportunity for Zara to further their research efforts in demand forecasting and replenishment optimization. In particular, the author suggests:

1. <u>Annual Store Demand:</u> The approach to estimate annual demand presented in this thesis assumes that a store's annual demand rate would not vary significantly from one year to the next. However, for stores where demand is changing monotonically this approach would lag reality. Zara could spend more time fine tuning the model so that it can adjust faster to monotonic trends.
2. <u>Distributional Estimates:</u> Zara could determine distributional forecasts to pair with the point estimates developed here. Paying particular attention to the changes in demand variability that occur throughout the year. This could help them tremendously in identifying periods with high uncertainty when they should build bigger buffers and periods with low uncertainty when lower buffers would be needed to maintain high service levels.
3. <u>Optimize In-Store Stock Levels:</u> In conjunction with distributional forecasts, Zara could look to tie its ideal stock level formulations to match target service levels (SL) and fill rates (FR).

They could then resort to optimization to either reduce inventory used to maintain same SL and FR, or use the same inventory to maximize these metrics.

4. Weather Factors: While this study was unsuccessful in its attempts to incorporate weather information to inform demand forecasts, the author nonetheless believes that weather does play a factor. A closer and more focused look at weather could help identify the proper relationship with demand.

Bibliography

Bonnefoi, T. (2012). *Demand Forecast for Short Life Cycle Products: Zara Case Study.* Massachusetts Institute of Technology.

Cachon, G. P., & Swinney, R. (2009, March). Purchasing, Pricing, and Quick Response in the Presence of Strategic Consumers. *Management Science*, pp. 497-511.

Cachon, G. P., & Swinney, R. (2011, April). The Value of Fast Fashion: Quick Response, Enganced Design, and Strategic Consumer Behavior. *Management Science*, pp. 778-795.

Caro, F., & Gallien, J. (2007, February). Dynamic Assortment with Demand Learning for Seasonal Consumer Goods. *Management Science*, pp. 276-292.

Caro, F., & Gallien, J. (2010, April). Inventory Management of a Fast-Fashion Retail Network. *Operations Research*, pp. 257-273.

Caro, F., & Gallien, J. (2012, December). Clearance Pricing Optimization for a Fast-Fashion Retailer. *Operations Research*, pp. 1404-1422.

Caro, F., & Martinez-de-Albeniz, V. (2010). The Impact of Quick Response in Inventory-Based Competition. *Manufacturing & Service Operations Management*, 409-429.

Chu, C., & Zhang, G. P. (2003). A comparative Study of Linear and Nonlinear Models for Aggregate Retail Sales Forecasting. *International Journal of Production Economics*, 217-231.

Correa, J. (2007). *Optimization of a Fast-Response Distribution Network.* Massachusetts Institute of Technology.

Corsi, E., Dessain, V., & Ton, Z. (2010). Zara: Managing Stores for Fast Fashion. *HBS Case 610-042.*

Eppen, G. D., & Iyer, A. V. (1997). Improved Fashion Buying with Bayesian Updating. *Operations Research*, 805-819.

Fast Fashion and the Responsive Supply Chain. (2013, July). *Apparel Magazine*, pp. 3-7.

Ferdows, K., Lewis, M. A., & Machuca, J. A. (2004, November). Rapid Fire Fulfillment. *Harvard Business Review*, pp. 104-110.

Fisher, M., & Raman, A. (1996). Reducing the Cost of Demand Uncertainty Through Accurate Response to Early Sales. *Operations Research*, 87-99.

Fisher, M., Rajaram, K., & Raman, A. (2001). Optimizing Inventory Replenishment of Retail Fashion Products. *Manufacturing & Service Operations Management*, 230-242.

Garro, A. (2011). *New Product Demand Forecasting and Distribution Optimization: A Case Study at Zara.* Massachusetts Institute of Technology.

Ghemawat, P., & Nueno, J. L. (2006). ZARA: Fast Fashion. *Case Study, Harvard Busines School.*

Hansen, S. (2012). How Zara Grew Into the World's Largest Fashion Retailer. *The New York Times.*

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

(2012). *Inditex Annual Report.*

Kelley, R. A. (2013). *Inventory Redistribution Optimization in the Fast fashion Industry.* Massachusetts Institute of Technology.

Kennedy, P. (2003). *A Guide to Econometrics, 5th Edition.* Cambridge, MA: The MIT Press.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Liner Regression Models.* McGraw-Hill Irwin.

Onkal, D., Goodwin, P., Thomson, M., Gonul, S., & Pollock, A. (2009). The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments. *Journal of Behavioral Decision Making*, 390-409.

Shah, J., & Avittathur, B. (2007). The Retailer Multi-Item Inventory Problem with Demand Cannibalization and Substitution. *International Journal of Production Economics*, 104-114.

Starr-McCluer, M. (2000). *The Effects of Weather on Retail Sales.* Federal Reserve Board of Governors.

Tokatli, N. (2008). Global sourcing: insights from the global clothing industry-the case of Zara, a fast fasion retailer. *Journal of Economic Geography*, 21-38.