Size Curve Optimization for Replenishment Products

by

Andrew J. Gabris
B.S.E. Civil Engineering, University of Michigan, Ann Arbor, 2008
M.S.E. Environmental Engineering, University of Michigan, Ann Arbor, 2009

Submitted to the MIT Sloan School of Management and the Institute for Data, Systems, and Society in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
and
Master of Science in Engineering Systems

In conjunction with the Leaders for Global Operations Program at the
Massachusetts Institute of Technology
June 2016

Signature of Author Signature redacted
MIT Sloan School of Management and MIT Institute for Data, Systems, and Society
May 6, 2016

Certified by Signature redacted
Stephen Graves, Thesis Supervisor
Abraham J. Siegel Professor of Management, MIT Sloan School of Management

Certified by Signature redacted
David Simchi-Levi, Thesis Supervisor
Professor of Engineering Systems, Institute for Data, Systems, and Society

Accepted by Signature redacted
John N. Tsitsiklis
Clarence J. Lebel Professor of Electrical Engineering
IDSS Graduate Officer

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

This page intentionally left blank.

Size Curve Optimization for Replenishment Products

by

Andrew J. Gabris

Submitted to the MIT Sloan School of Management and the MIT Institute for Data, Systems, and Society on May 6, 2015 in partial fulfillment of the requirements for the degrees of Master of Business Administration and Master of Science in Engineering Systems.

Abstract

Nike replenishment products (make to stock) are forecasted and planned at a style/color level and then disaggregated to a size level forecast through the use of a size curve. This method of forecasting and planning provides many advantages such as reduced effort expended on forecasting and the ability to quickly roll up data for capacity planning.

Size curves are based on historical proportions of sales. For instance, if size small sells 10% of the volume for a given style/color, the size curve would be set to 10% for small. Not surprisingly, the size curve for a given style/color sums to 100%.

In a manner similar to the forecast, size curves are used to disaggregate style/color safety stock quantities that are used to ensure target service levels (item fill rates) are met. However, this disaggregation results in lower-than-anticipated service levels for the size-level stock-keeping units (sku's), since the style/color safety stock does not account for the increased forecast error at the size level.

Additional challenges occur from the fact that the relative magnitude of the forecast error is inversely proportional to the demand level. As a consequence, fringe sizes, which account for lower volumes of sales, account for a disproportionate amount of variability within a style/color affecting service levels.

To resolve these observations, the project first attempted to improve the quality of the size curves by applying different statistical forecasting techniques in their formulation. We found that the status quo forecasting methodology was as good as or better than other methods, which suggests that there is a limit to the accuracy of size curves.

In order to increase service levels across all sizes, several recommendations have been made. First, by reducing the number of size offerings from the replenishment products, many of the more challenging sizes will be eliminated. Next, this additional size level error can be accounted by right-sizing safety stock. Finally, a manual update process for size curves is employed that leaves many facets of the process to individual planners. Standardization of the size curve process will support more consistent results.

Thesis Supervisor: Stephen Graves
Title: Abraham J. Siegel Professor of Management, MIT Sloan School of Management

Thesis Supervisor: David Simchi-Levi
Title: Professor of Engineering Systems, Institute for Data, Systems, and Society

This page intentionally left blank.

Acknowledgements

First, I would like to thank the Nike team for empowering me to deliver on this project. Jessica McCoy, I appreciate all of your patience and insight during my time at Nike. The broader team including but not limited to Susan Brown, Nikhil Soares, Eric Van Wely, Justin Jackson, Natalia Sutin, Ken Philliber and Wolf Hoffman have helped me further refine and develop both my technical skills as well as communication skills, and for that I am truly grateful. To Clara Voigt, Blair Holbrook and Amanda Lurie, your support as fellow Nike LGO interns has been invaluable in keeping me focused.

Next, I would like to thank the Leaders for Global Operations program at MIT for its support of this work. My thesis advisors, Dr. Stephen Graves and Dr. David Simchi-Levi, have guided me through this challenging process, and MIT armed with me with newfound knowledge and enthusiasm to apply at Nike. Dr. Chris Caplice, I owe much of my project's success to what I learned in your Logistics Systems course.

Thank you to my parents and sister, Cara, for their continued love and support. And finally, Molly, thank you for allowing me to switch coasts for six months and pursue my dreams at Nike. This thesis is dedicated to you.

This page intentionally left blank.

Note on Nike Proprietary Information

In order to protect information that is proprietary to Nike, Inc., the data presented throughout this thesis has been modified and does not represent actual values. Data labels have been altered, converted or removed in order to protect competitive information, while still conveying the findings of this project.

Table of Contents

List of Figures

List of Equations

List of Tables

List of Abbreviations and Definitions

AA	Always Available
CSL	Cycle Service Level
DC	Distribution Center
DSI	Days of Sales in Inventory
IFR	Item Fill Rate
LGO	Leaders for Global Operations
NPV	Net Present Value
PO	Purchase Order
RMSE	Root Mean Square Error
SKU	Stock Keeping Unit

Introduction

The purpose of this thesis is to describe the methodologies and results of forecast disaggregation and safety stock calculations in an apparel, equipment and footwear replenishment business. Research for this paper was conducted with Nike, Inc. (Nike) in their Always Available (AA) Group. This introductory chapter provides the reader with necessary context and background to understand the analysis performed. An overview of Nike and the area of research, project objectives and approach, and a summary of future chapters are provided.

Overview of Nike, Inc.

Nike was incorporated in 1967 under the State of Oregon. Founded as a running footwear company, Nike has since evolved to provide apparel and equipment in addition to footwear. Historical revenue by these three product engines can be found in Figure 1.

Product offerings focus on eight key categories: running, basketball, global football (soccer), men's training, women's training, action sports, sportswear and golf. Additionally, Nike, Inc. owns subsidiary brands Jordan, Hurley, Converse, and Nike Golf, which further extends the reach of the company.

Nike is a truly global company with approximately 55% of revenue coming from international markets leaving 45% of revenue coming from the United States. In fiscal year 2015, Nike revenue exceeded $30 billion with a net income of $3.3 billion. Revenue and net income have grown at 8.8% and 9.0%, respectively, over the past five years.

As with most companies in the retail industry, Nike manufactures its products almost exclusively through the use of contract manufacturing. Apparel is supplied from 408 factories in 39 countries with China, Vietnam, Sri Lanka, Thailand, Indonesia, Malaysia and Cambodia accounting for the largest volumes of apparel production.

Footwear, on the other hand, is manufactured in 146 factories in 14 countries. China, Vietnam and Indonesia account for the largest volumes of footwear production. An increasing emphasis has been placed on sustainability and social impact on all manufacturing. More information on this initiative can be found at nikebetterworld.com.[1]

[1] Nike, Inc. Form 10-K for the Fiscal Year Ended May 31, 2015

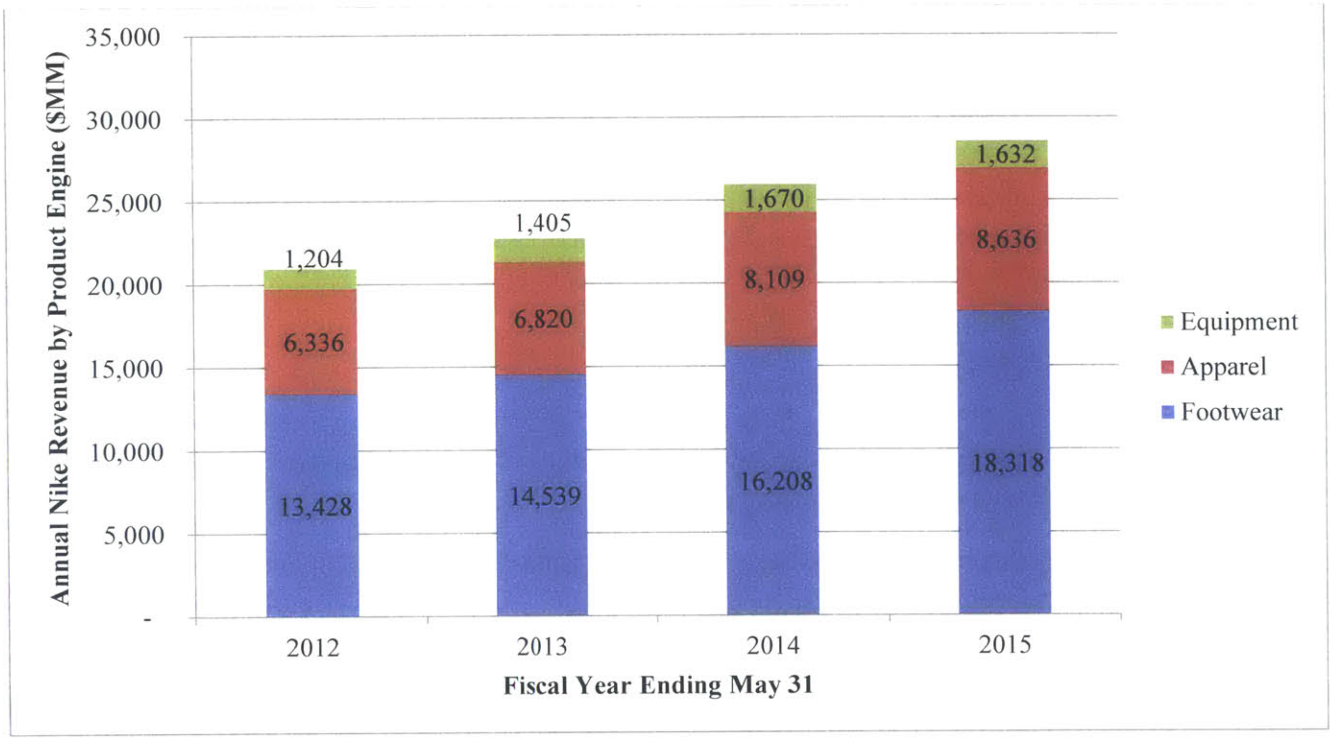

Figure 1: Nike Annual Revenue by Product Engine[2]

Problem Statement and Motivation

Nike operates a replenishment (make to stock) model called "Always Available." Here products are forecasted at the style/color level (e.g. white t-shirt) and forecasts are disaggregated to the size level using a size curve. Size curves rely on historical sales data and are typically updated between every one to three months. Similarly, safety stock is calculated first at the style/color level and then disaggregated to the style/color/size level, based on the size curve.

The challenge has been that service levels across different sizes appear to be inconsistent. In particular, fringe sizes, which compose the less common sizes such as XS or XXL in apparel or 12.5, 14 or 15 in footwear, appear to stock out more frequently than core sizes. These stockouts lead to reduced revenue for the business as well as strained relationships with Nike retailers when service levels fall below targets.

To make matters more difficult, internal processes are set up for style/color planning and little is done to manage the size level. Inventory managers are measured on their style/color forecasts and not the size curves that they set. Given that customers (retailers) and consumers (product users) buy at the size level, this layer is crucial to the success of Nike and the AA business model.

[2] Ibid, 81

Project Goals and Anticipated Outcomes

The first goal of the project was to quantify the actual service levels across sizes and the different product engines at Nike. Most of the initial information on service levels by size was purely anecdotal. Size level order and forecast data will need to be analyzed to understand the current state.

The second outcome is to provide a consistent framework for setting and measuring the service level across all sizes. As will be determined by the first outcome, the expectation is that all sizes have a near equal probability of being in stock. The success of the project will be measured through revenue generated, NPV, inventory and service level.

Project Approach

The project consisted of four distinct phases in order to reach the desired goals: current state analysis, modeling of proposed impacts, recommendation formation and implementation.

Current state analysis consisted of evaluating current Nike systems and their effects on service levels. The analysis focused on target service levels to see if the systems in place were allowing for correct service levels at the product size level.

The current system calculates forecast error at the style/color level and uses this to calculate an aggregate safety stock for each style/color. Size curves are then applied to disaggregate this safety stock to each style/color/size. This current state approach was then compared to calculating the safety stock for each style/color/size directly using size level data. With this information, the true item fill rate (IFR) target could be inferred.

With a clear understanding of the current state, the next step of the project was to analyze ways to correct any deficiencies in the current system. This focused on using historical sales data to form more accurate size curves as well as correcting safety stock quantities to match the desired service level.

Next, using the model results, recommendations were then formed to balance the analytical insights with the true business needs. Feedback was solicited from key project stakeholders across the Nike organization and included in these recommendations.

Finally, an implementation plan was developed in order to ensure that actions were taken to put recommendations into place. The author used his time on site to move the implementation forward as much as possible, but some work will require additional time.

This page intentionally left blank.

The Nike Supply Chain

Nike employs a variety of supply chain strategies such as make to stock, make to order, engineer to order and build to order. These strategies show up in the futures, responsive business/AA and NikeiD business models. Each have their own advantages and disadvantages in meeting customer and consumer needs and are outlined in more detail in the coming subsections.

The Nike supply chain consists of three district product engines: apparel, footwear and equipment. In order to unite these product engines through commonalities, Nike mixes products from the different engines into several categories such as Global Football, Men's Training, and Running.

Futures Business Model

The Futures model is a make to order business model. Customer orders are placed approximately six months in advance of delivery and the production process follows what is shown in Figure 2[3]. In this model no raw materials or finished goods are stored in anticipation of future demand. In this sense, there is little risk as the products are being manufactured and delivered to meet a certain demand beyond a negligible cost of reserving contract manufacturer capacity.

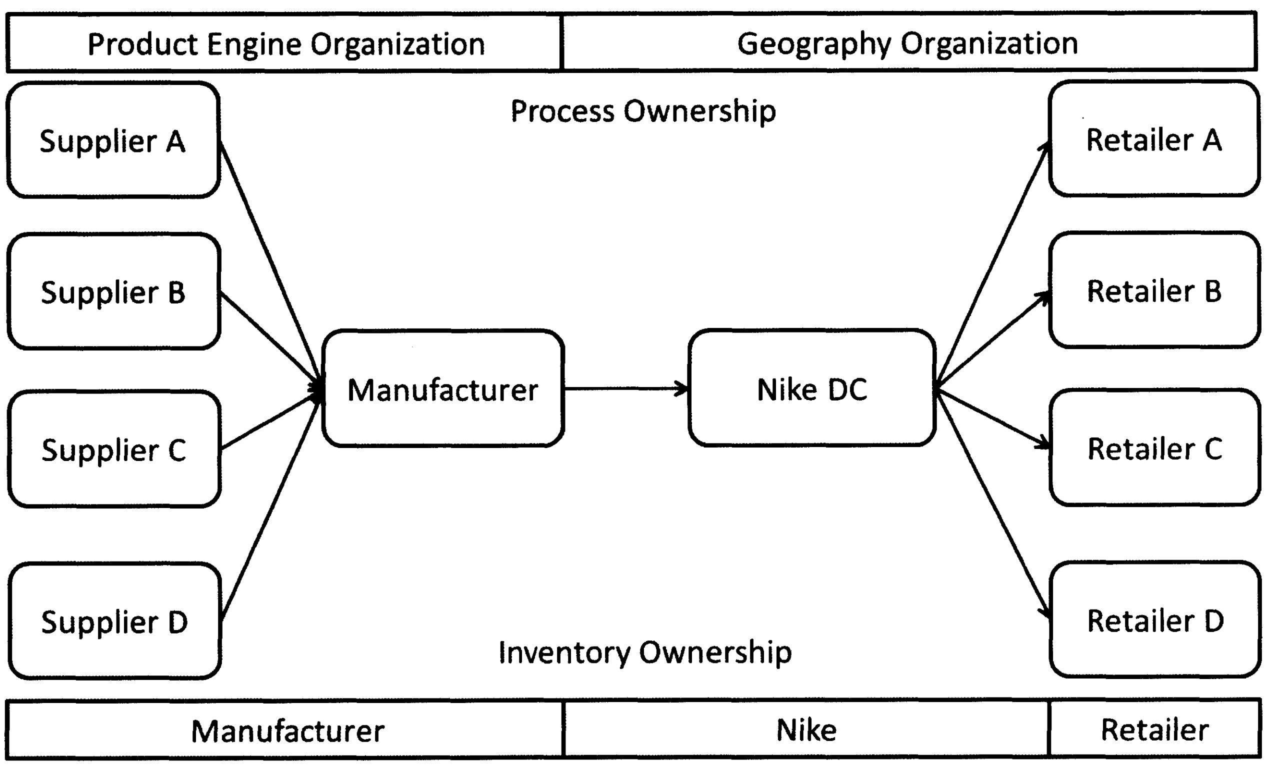

Figure 2: Nike Futures Supply Chain Overview

The information flow for the manufacturing process flows backwards. Retailers place orders with Nike geographies, which inform their demand planning. These orders are then passed along to manufacturing and sourcing whose job is to work with factory partners to execute the order.

[3] Ben Polak, Multi-Echelon Inventory Strategies for A Retail Replenishment Business Model, 21

The manufacturing organization is aligned across product engines such that, for instance, footwear orders across all geographies are rolled up and sent to factories. Raw materials are then procured, products are manufactured, and finished goods are sent to a consolidator where Nike takes ownership of the product.

The consolidated orders are then reorganized by geography and typically sent to their destination via ship. Once the product arrives at its destination port, one of two things will happen. Most products will be shipped to a Nike DC, but some product will be shipped directly to retailers bypassing the Nike DCs. This destination is largely dependent on the size of the customer and whether the container holds product from only one retailer.

Nike Always Available Supply Chain

Nike's replenishment (make to stock) supply chain model employs the same underlying supply chain as the futures (make to order) model. Product is manufactured at the same contract manufacturer facilities, grouped at the consolidator, and shipped to its port of entry.

Yet there are some key differences that exist as highlighted in Figure 3. For one, inventory levels are higher across this supply chain model, which mainly corresponds with the additional staged inventory at various locations in the supply chain.

All products have inventory staged at Nike owned distribution centers as shown by the second blue triangle. This allows customers to place regular orders for goods to replenish their current stock. The entire additional inventory held at Nike distribution centers is owned by Nike.

Some products may have either raw materials or finished goods stored at manufacturing facilities. This is done to further reduce product lead times (which reduces inventory required to be held at Nike DC) but runs some additional risk of inventory obsolescence. These modifications allow Nike to quickly respond to customer needs with a supply chain containing relatively long lead times.

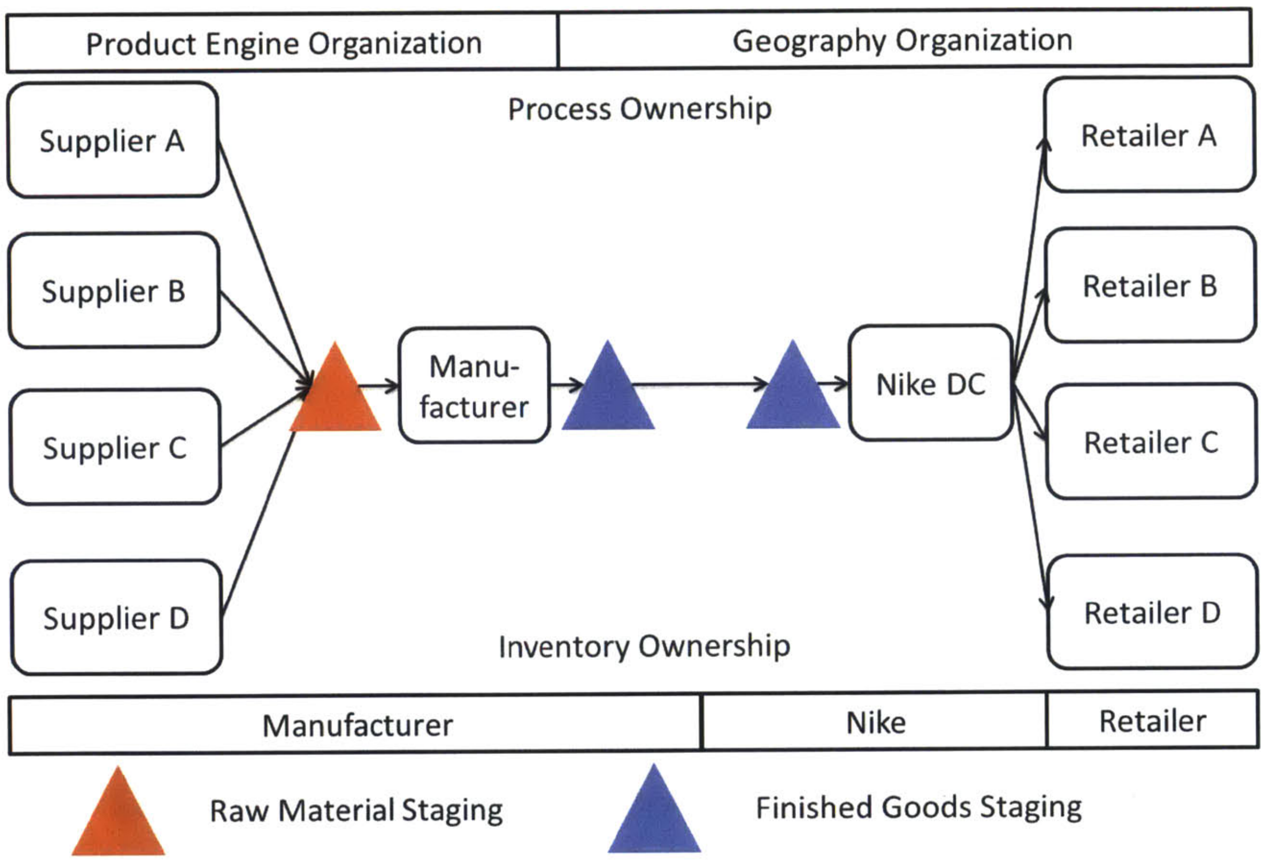

***Figure 3 : Nike Always Available Supply Chain*[4]**

Size Curves

Size curves, also referred to as disaggregation ratios, are an important facet of this project. Figure 4 shows the basic concept of how size curves are applied in the make-to-stock business. Forecasts are completed at the style/color level and size curves are applied to break the forecast down to the size level. Size curves are formed based on historical demand information. For instance, if size small accounted for 10 percent of historical sales, then the size curve would be set to 10 percent for size small. If the style/color forecast for a particular product was 1000 units, the forecast for size small of that product would simply be 1000 units * 10 percent = 100 units. To maintain unity, the size curves across all sizes must sum to 1.

$$F_{S/C/S} = s * F_{S/C}$$

Equation 1: Forecast Relationship with Size Curve

Where

$F_{S/C/S}$ = Style/color/size forecast for a given time period

$F_{S/C}$ = Style/color forecast for a given time period

[4] Polak, Multi-Echelon Inventory Strategies, 24

s = size curve for a size of a given style/color

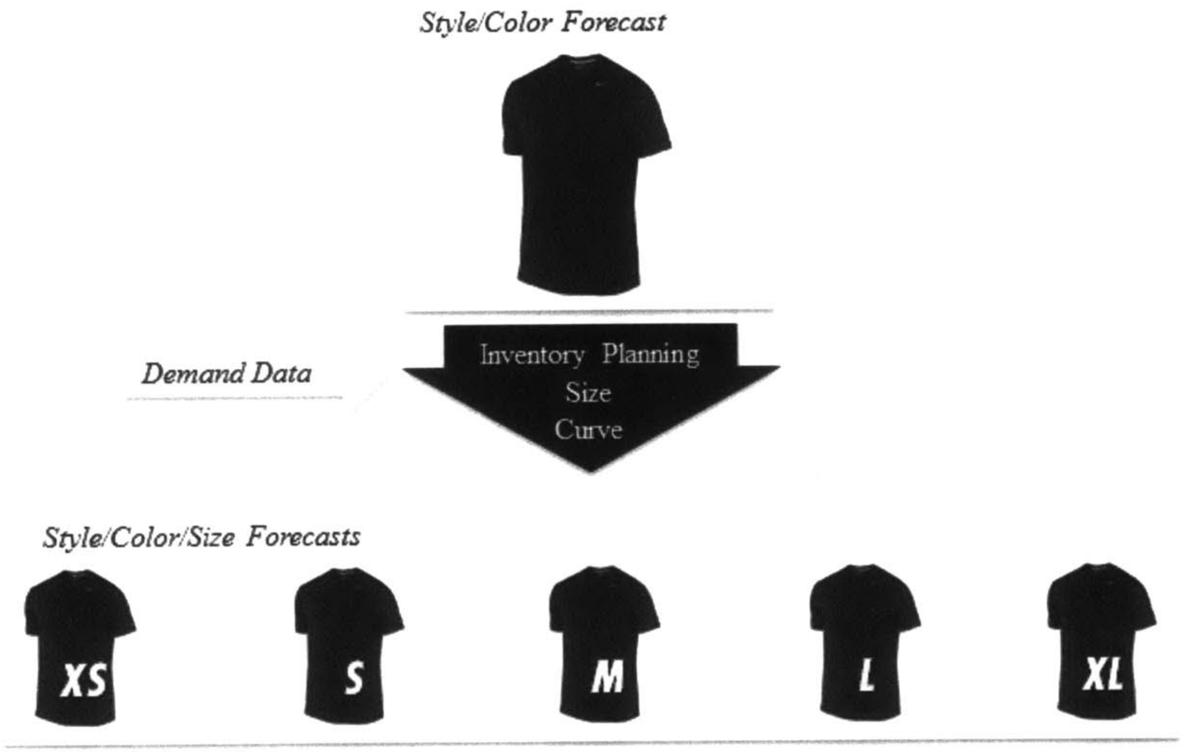

Figure 4: Size Curve Overview

Size curves also help to set safety stock targets at the size level. An inventory planning system is used, which measures the weekly style/color forecast error against actual demand (e.g., a standard deviation or MAPE). This value is then disaggregated to each size where an individual quantity of safety stock is calculated. An equivalent way to think about safety stock is that safety stock is calculated in aggregate for each style color and then proportioned through the size curve.

Size curves follow a regular update process to reflect current demand data, and the inventory planning function owns this process. For products that are new to replenishment, size curves are typically pulled from like models or historical futures sales are used to initialize size curves.

Literature Review

Introduction

The project relies on classical inventory management techniques as well as forecasting techniques. Thus, this literature review is split into these two parts. The purpose of this review is to determine methodologies and techniques that might be relevant to the successful execution of this project.

Safety Stock

In their classic inventory management textbook, *Inventory Management and Production Planning and Scheduling*, Silver, Pyke and Peterson provide the necessary background to understand the analytical methods used in this paper. First, let us begin with the general equation used to calculate safety stock.

$$Safety\ Stock = k * \sigma$$

Equation 2: General Safety Stock Calculation

Where,

k = Safety factor, determined from desired service level

σ = Standard deviation of demand over the lead time and review period[5]

Using type II inventory metrics or Item Fill Rate, the safety factor k can be derived as follows

$$IFR = \frac{E[D] - E[US]}{E[D]}$$

Equation 3: Item Fill Rate

Where

E[D] = Expected demand over the review period

E[US] = Expected units short or unmet demand in a replenishment cycle

IFR is typically set by management as a strategic objective. It can also be set at a level that maximizes a firm's profitability, accounts for aggregate considerations such as an overall budget or applies a simple approach such as a common factor.[6]

For the sake of this analysis, an R, S inventory policy will be used. Here, R is the review period in weeks and S is the order up to quantity. What this means is that every R weeks, a planner will review the inventory position (IP) and place an order equal to S – IP.

[5] Edward Silver, D. F Pyke, and Rein Peterson, *Inventory Management and Production Planning and Scheduling,* 109, 244
[6] Ibid, 245-247

If the forecast error is assumed to be normally distributed and the forecast is assumed to be unbiased, the following can be substituted into the above equation

$$E[US] = G(k) * \sigma_{Product}$$

Equation 4: Expected Units Short Under Normality Assumption

$$\sigma_D = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(D_t - F_t)^2}$$

Equation 5: Root Mean Square Error

$$\sigma_{Product} = \sqrt{E[LT + R](\sigma_D)^2 + (E[D])^2(\sigma_{LT+R})^2}$$

Equation 6: Standard Deviation with Variable Lead Time under Normality Assumption

$$G(k) = \frac{E[D]}{\sigma_{Product}}(1 - IFR)$$

Equation 7: Approximate Value of Unit Loss Function for determining k

Where

σ_D = Standard deviation of forecast error or root mean square error

σ_{LT+R} = Standard deviation of lead time plus review period

F_i = Forecast for demand in time period t, assumed to be unbiased

D_i = Demand in time period t

n = Number of observations

LT+R = Length of lead time and review period

E[D] = Expected demand over review period

It is important to note that Equation 5 is based upon the assumption that demand and lead time are independent random variables.

G(k) is known as the normal unit loss function. It is common to find textbooks such as Silver, Pyke and Peterson containing tables of G(k) with corresponding k values. Alternatively, one can use a solver such as Excel to determine the value of k, with the following formula.

$$G(k) = f(k) - k * F(k)$$

Equation 8: Unit Loss Function

Where

f(k) = Standard normal probability density function

F(k) = Standard normal cumulative distribution function, $P(x \leq k)$

Still, Equation 6 has some limitations, and should be used for large values of $E[D]/\sigma_{Product}$. For smaller values of $E[D]/\sigma_{Product}$ (less than 0.5), Equation 8 should be used:

$$G(k) - G\left(k + \frac{E[D]}{\sigma_{Product}}\right) = \frac{E[D](1 - IFR)}{\sigma_{Product}}$$

Equation 9: Approximation Used to Determine k Value for Low E[D]/σ

In addition to setting safety stock targets based on items filled, management may also use a different metric to fix safety stock targets: cycle service level (CSL). This is the probability of not stocking out in a given replenishment cycle.[7]

$$CSL = F(k)$$

Equation 10: Cycle Service Level under Normality Assumption[8]

Using CSL as a metric, it is more straight forward to derive a k value, especially when normality is assumed. Knowing the target CSL, k can be quickly taken from normal distribution tables or using the norm.s.inv function in Excel. Figure 5 shows how these k values relate to the standard normal probability density function. For a 50% CSL target, we find a k value of 0 and would carry no safety stock. This makes intuitive sense because if our forecast is unbiased and the demand is stochastic, we would expect half of our replenishment cycles to have a demand greater than our forecast and half to have a demand less than our forecast. Thus, there is no reason to hold any safety stock.

Next, if we select a CSL of 95%, this means that 95% of the time our total inventory should be enough to cover the demand over the review period. We find that $k \approx 1.64$ fits this criteria; 95% of the distributions volume is to the left of that line.

The final case worth noting is when a CSL is less than 50%. In the case below, we see that a CSL of 35% yields a k value of $\approx$ -0.385 following the logic of the paragraph above. Returning to Equation 2, we see a negative safety stock value. This means that we actually want to order less than our forecast. It may not sound right at first, but this strategy may make sense for rare cases where the cost of holding a product or inventory obsolescence is high compared to the margin to be gained from the product.

[7] Ibid. 268-269
[8] Ibid. 266-268

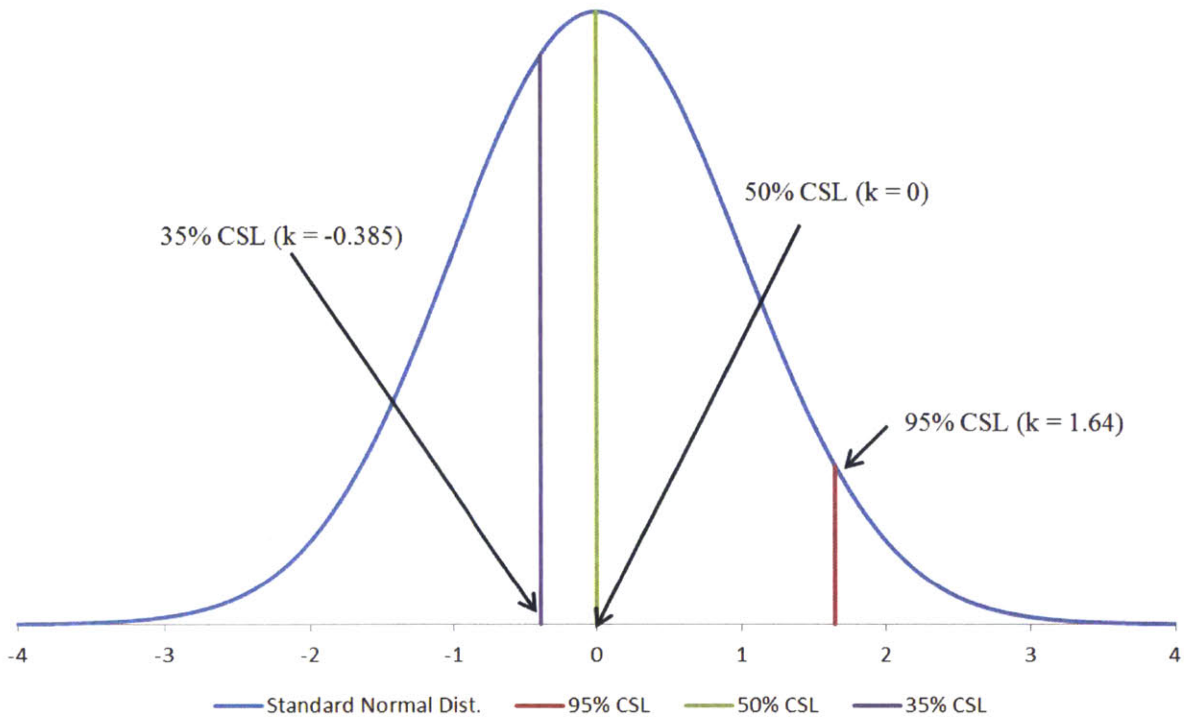

Figure 5: Standard Normal Distribution with Various CSLs

Having both our safety stock and order quantities, our average on-hand can be found as follows

$$OH\ Inventory = \frac{R * \mu}{2} + k * \sigma_{Product}$$

Equation 11: On-hand Inventory

Where

R = Review period in days

μ = Average daily demand

This formula accounts for the fact that we are expected to use up our ordered inventory, so on average half of it will be on hand during a replenishment period. Our safety stock is provided to protect us against surges in demand.

Figure 6 illustrates this effect. Our safety stock is designed to protect us when lead time delays are present or when an unanticipated surge in demand occurs. At the same time, if demand is less than forecasted, additional cycle stock will remain for that given replenishment cycle.

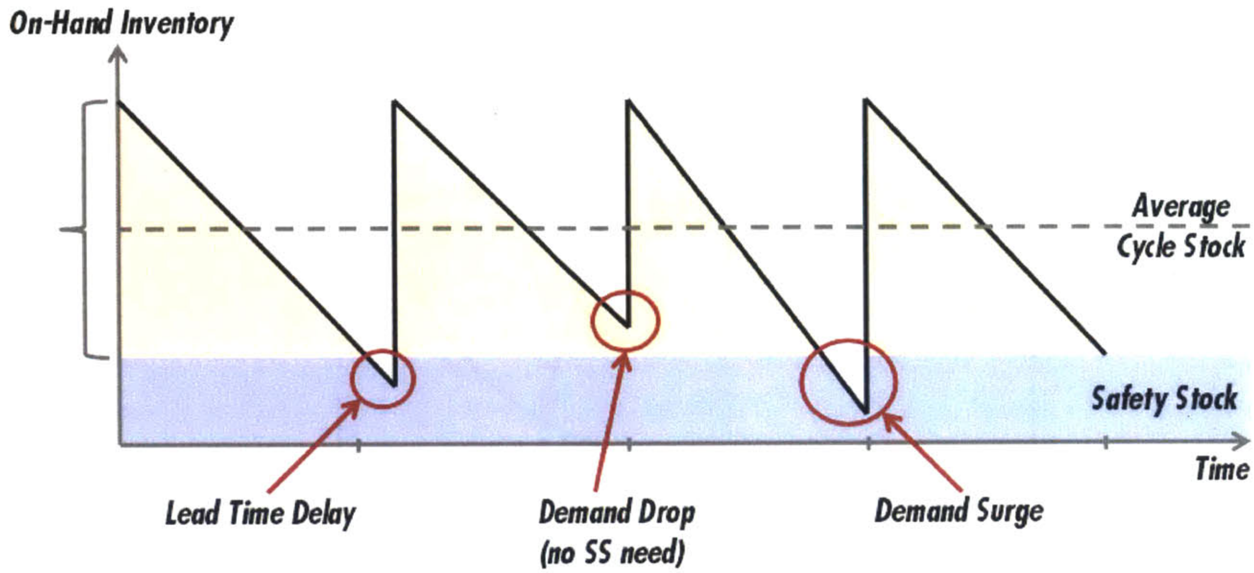

Figure 6: Diagram of Safety and Cycle Stock for an R, S Order Policy[9]

Safety stock quantities are highly dependent on target IFR. Figure 7 shows that for a given product demand and variability characteristics, ever increasing quantities of safety stock are required. Under the normality assumption, we cannot technically achieve an IFR of 100%. This is because the normal distribution extends from negative infinity to positive infinity. Thus, despite the odds being very near zero, there is still some non-zero probability of a very large demand (up to infinity).

In practice, it is possible to hit a 100% IFR if the true distribution underlying the demand is known and bounded. Still, if the distribution contains tails of lower probability occurrences like the normal distribution, we will still see the exponential slope in Figure 7.

[9] Diagram reproduced with permission of Dr. Jessica McCoy

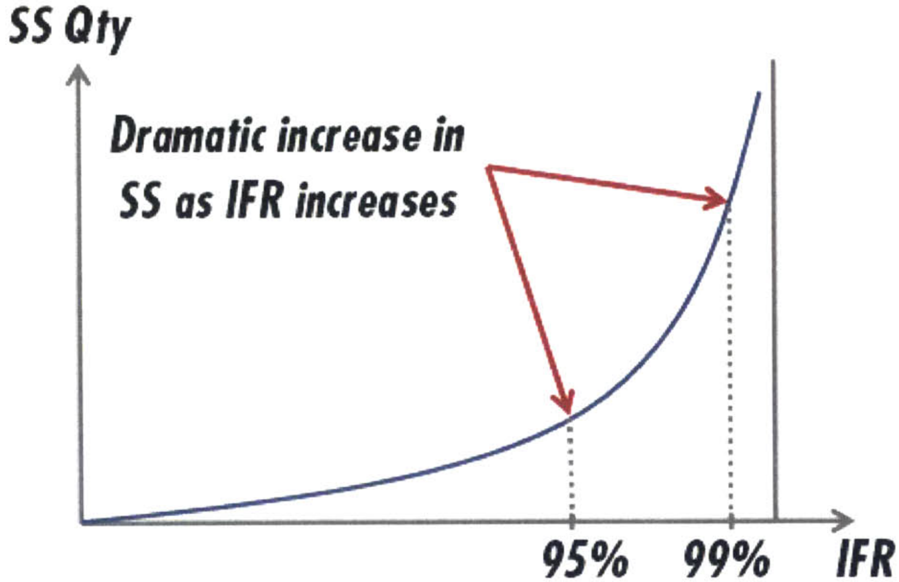

Figure 7: Relationship between IFR and Safety Stock Quantities[10]

Forecasting

Forecasting typically follows one of two approaches: bottom-up or top-down. In bottom-up approach, forecasting is done at an individual product level such as a SKU. These forecasts are then combined or rolled-up to aggregate the demand forecast. Here the top level forecasts are sums of the lower level forecasts.

Top-down forecasting is the exact opposite where forecasts are done in aggregate through the use of aggregate demand data and then disaggregated down to more discrete levels such as SKUs. There are benefits and drawbacks to each method. There is a large volume of work on which method is more accurate given certain product characteristics, but the results still seem to be inconclusive.[11]

Gross and Sohl introduce methods for disaggregating a forecast or taking a top down forecast and allocating it to sub-products. They call out numerous methods for forming disaggregation matrices and look to find the methodologies that induce the least error into the forecast. They find that the simple method of a ratio of individual product demand to overall demand performs best in most cases.

It is common for companies to use a consistent metric in measuring forecast accuracy. Equation 5 introduced root mean square error, which is often used because of its statistical significance, but it does not do a good job when comparing products that sell at different orders of magnitude. To account for this, another common metric is mean absolute percent error (MAPE) as seen in Equation 12.[12]

---

[10] Ibid

[11] Giacomo Sbrana and Andrea Silvestrini, "Forecasting Aggregate Demand: Analytical Comparison of Top-Down and Bottom-Up Approaches in a Multivariate Exponential Smoothing Framework"

[12] Silver, Pyke and Peterson, *Inventory Planning*, 109

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{D_t - F_t}{D_t}\right|$$

Equation 12: MAPE

In many cases it is necessary to deal with a seasonal demand. Here we can only calculate average RMSE and but need to adjust this for seasonality in the demand. An example of this application would be setting safety stock targets with varying seasonal demand. To help with this challenge, Brown introduces a concept that has become known as the Variance Law as seen in Equation 13. It allows us to empirically estimate the standard deviation or RMSE based on the mean demand. Typically, the forecast is used as an estimate for mean demand when predicting the future variance.[13]

$$\sigma = \alpha D^{\beta}$$

Equation 13: Variance Law Relationship

Where α and β are empirical constants that can be fit based on historical data for a product or family of products.

Finally, it is important to call out several forecasting methodologies. Moving Average, as shown in Equation 14, simply averages the demand of the n previous time periods. Moving average forecasts are typically used for data that is relatively stable and does not exhibit any trends or seasonality.

$$F_{t+1} = \frac{1}{n}\sum_{t=1}^{n} D_t$$

Equation 14: Moving Average for n time periods

Where F_{t+1} = forecast made at time t for time period t+1

There are two special cases of this formula: the naïve and the cumulative version. In the naïve model, n is equal to 1 and the forecast for period n is simply the demand in time period n-1. Not surprisingly, this tends to be a very nervous forecast that varies widely. Conversely, the cumulative version of the moving average forecast includes all relevant historical data, so the value of n increases by one with each new time period. This is going to be a much more stable forecast especially for products with large amounts of history.

For predicting a level or slowly moving demand, simple exponential smoothing is frequently used as well as seen in Equation 15. Here, α is a smoothing constant, with values ranging from 0 to 1. Higher values represent nervous and more responsive forecasts, while smaller smoothing constants lead to steady, unchanging forecasts. In practice, though, smoothing constants typically fall between 0.1 and 0.3.[14]

$$F_{t+1} = \alpha(D_t) + (1 - \alpha)F_t$$

Equation 15: Simple Exponential Smoothing Formula

[13] Robert Brown, "Estimating Aggregate Inventory Standards," 64
[14] Chris Caplice, "Introduction to Demand Forecasting and Planning"

One of the major challenges of forecasting is how to infer the true demand in the face of stockouts. If the demand is less than the available inventory it is very easy to determine the quantity demanded. But when a stockout occurs, the true demand can be equal to or greater than the quantity ordered. This is something we do not know. Ferreira et al. share some of the methodologies and algorithms that can be used to estimate the true demand signal. In essence, sales patterns of the product when it is in stock can be used to estimate demand when the product is out of stock.

Cost of a Lost Sale

To better understand and quantify the impacts of stockouts, extensive research on the cost of a lost sale was performed. Figure 8 is adapted from a Harvard Business Review article and does a good job highlighting typical consumer responses to an out of stock item. If we look at it from the retailer's perspective, buying the item at another store, not purchasing the item and delaying the purchase are all undesirable outcomes. From the viewpoint of a producer, substituting for a different brand, foregoing the purchase, and delaying the purchase are all bad outcomes. In each case, a bad outcome will occur approximately half of the time.

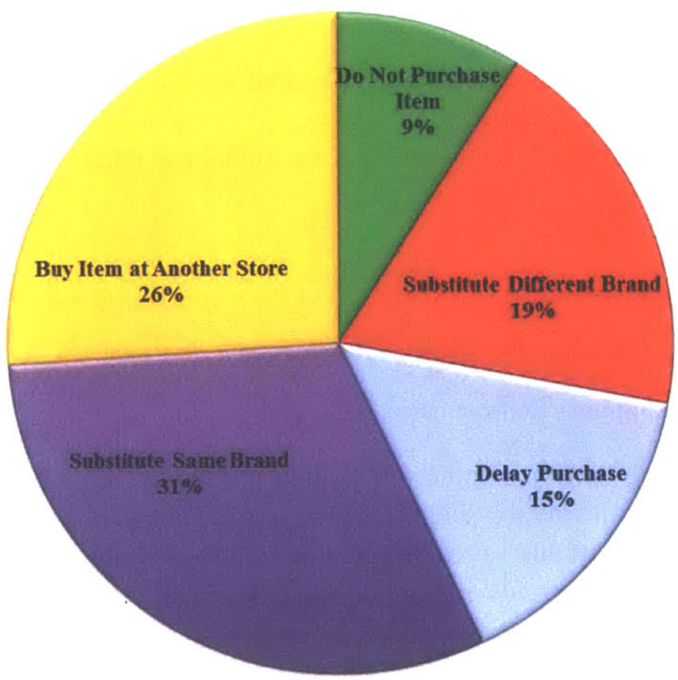

***Figure 8: Consumer Response to Stockouts*[15]**

Furthermore, Anderson, Fitzsimons and Simester; Craig, DeHoratius and Raman; and Kim and Lennon all attempt to quantify the impact of a stockout or target service level to varying degrees of success. Most of the data is anecdotal, but it is noted that stockouts need not impact the end consumer to have an impact on overall sales. If a product supplier provides poor service, this will affect their relationship with the

[15] Corsten, Daniel and Thomas Gruen, "Stock-outs Cause Walkouts"

retail buyer who ultimately control promotions, display and floor space. There is also a strong correlation between buyer order quantities and historic service levels[16].

Anderson, Fitzsimmons and Simester begin by stating that the short term and long term costs of a stockout are quite often unknown. This is not surprising given the complex relationships between supply, demand and the supply chain. It has led many retailers to abandon complex inventory models, often in favor of a set, management-driven service level[17]. Most of the impacts were found to be in the long-term due to tarnished reputations rather than in the short run. Kim and Lennon reinforce the reduced perception of the store due to stockouts. While interesting material, it did not get the author closer to a way to estimate the cost of a stockout.

---

[16] Nathan Craig, Nicole DeHoratius, and Ananth Raman, "The Impact Of Supplier Reliability Tracking On Customer Demand: Model And Estimation Methodology"
[17] Eric Anderson, Gavan Fitzsimons, and Duncan Simester, "Measuring And Mitigating The Costs Of Stockouts"

This page intentionally left blank.

Methodology

Hypothesis Formulation

With a solid foundation on the background of current research and Nike operations, it is time to introduce the hypotheses serving as the impetus for the project research. Three key hypotheses have been explored:

1. Size curves are currently optimized and best reflect sizing based on currently available data
2. Stockouts occur in the same frequency across all products, regardless of size
3. The sizes currently offered on the AA line are appropriate for the AA strategy

The remaining portion of this paper will investigate these three hypotheses in more detail and attempt to find sufficient data to reject them.

Data Collection and Analysis

In order to better understand product and demand characteristics at the size level, the project examined size level data from three different geographies: North America, Europe, and Japan. All products in the AA line over the past two years were considered in the analysis. To be included in the calculations, products needed to be on the AA line for at least 52 consecutive weeks with no more than 3 missing weeks of data.

In all, this represented 41% of the portfolio, but these products represented 91% of the sales volume within AA. Products were relatively evenly spread across the three product engines of apparel, equipment and footwear. Equipment encompasses items such as hats, gloves, bags and any other item that does not clearly fit into apparel or footwear.

For each product included, demand and forecast data were collected at the weekly level and at the geography level. This is to match the forecast error calculation to reality. Nike customers are able to order AA product on a weekly basis, so the forecast and demand should match. In order to account for the time differences between a forecast being entered and an order being placed, forecasts were lagged by the average product lead time to sync up with the demand. For instance, if a product had a lead time of 12 weeks, the forecast made 12 weeks prior is the relevant comparison to a sale made today.

Additionally, costing data was pulled into the model for calculating trade-offs. In order to calculate holding costs, a given product's landed cost, which is the cost to manufacture and ship a product to its destination DC, was used. A consistent holding cost rate was used for all products regardless of specific product attributes such as product engine or geography. This was used as a simplifying assumption. Wholesale prices were used in order to determine revenue.

Finally, lead times for each product were pulled based on the product's destination. For simplicity the maximum planned lead time for a product was used as lead times can vary depending on the manufacturing location and other factors.

Underlying Assumptions

In order to construct a representative model, several simplifying assumptions and approximations were employed. One of the key things to note is that the model is not predictive. It uses historical forecast and demand data to analyze service level and safety stock quantities. Thus, any recommendations assume that future performance is going to be similar to past results. So should something such as forecast accuracy or AA product line attributes greatly change, the model would need to be rerun with more recent data.

Next, several tactical assumptions were used to reduce the complexity of the model. These included the assumptions that forecast error follows a normal distribution and that lead time variability is zero. The normality assumption is a common assumption and the data roughly fits this. Furthermore, commercial packages such as Logility Voyager use this assumption when calculating safety stock. Lead time variability was assumed to be zero due to lack of readily available data; hence, Equation 6 will simplify to the following:

$$\sigma_{Product} = \sigma_D \sqrt{E[LT + R]}$$

Equation 16: Product RMSE over Lead Time with No Lead Time Variation

Given that Logility calculates the forecast error at a style/color level, size curves are used to get the RMSE at the size level. It is important to note that this equation only holds if s_i remains unchanged over time.

$$\sigma_{s/c/s_i} = s_i * \sigma_{product}$$

Equation 17: Relationship between RMSE and size curves

Where

s_i = the size curve fraction for size i of a given style/color

$\sigma_{product}$ = the standard deviation of the style/color of a product

$\sigma_{s/c/si}$ = the standard deviation of the style/color/size of a product

Current State and Size Curve Review Concept Model

In order to evaluate the hypothesis a model was constructed in Alteryx. The main objective of the model was to look at the current state of allocating safety stock and compare it to the safety stock requirements of each style/color. The 52 weeks of data were used to generate the forecast error for each style/color, which was then used to calculate one average safety stock quantity for each product (i.e., each style/color).

The model begins with two parallel flows: one calculating at the style/color level and another at the style/color/size level. Both models lead to quantities of safety stock using Equations 2, 5, 6 and 7 for a given service level. Equation 9 shows that there are limitations to using this methodology for low values of $E[D]/\sigma$. The model solely relies on Equation 7, so style/colors and sizes with low $E[D]/\sigma$ are omitted.

Then, the model simulates the formation of size curves. It looks at historical demand to formulate these curves. Curves are initialized with 6 months of data and updated weekly. Based on the findings of Gross and Sohl, the model uses moving average and cumulative forecasting methods in agreement with the paper. The authors found that more complicated methodologies did not yield better results. In addition to those methodologies, exponential smoothing with varying alpha parameters were included. These size curves were then used to disaggregate the safety stock calculated for the style/color to the style/color/size.

From here, a comparison could be made to determine if the allocated safety stock was sufficient. Similarly, given the safety stock allocated, one could also use Equations 2, 5, 6 and 7 to back calculate the expected service level. This allowed for the evaluation of the sufficiency of current safety stock targets.

The size curve formulation was an integral part of the model that could be used to evaluate hypothesis 1. Many alternative size curves could be formulated and compared to the status quo. Expected service level, RMSE, and safety stock discrepancy could then be used as metrics for comparing these different methodologies.

Current State and Size Curve Review Concept Example

The following section provides a more detailed example of the calculations carried out in the base model.

1. Determine safety stock quantity used in current system
    a. Calculate root mean square error for style/color over 52 week period (Equation 5)
    b. Multiply the style/color root mean square error by the given size curve of a given size to determine the style/color/size root mean square error (the given size curve is found using Equation 18 with the status quo number of time periods)
    c. Adjust the root mean square error by the product lead time to get standard deviation of demand over lead time and review period (Equation 6)
    d. Calculate k value for style/color/size (Equation 7) for given IFR
    e. Determine safety stock target set by current system (Equation 2)
2. Determine safety stock quantity needed for desired service level
    a. Calculate root mean square error for each style/color/size over 52 week period (Equation 5)
    b. Adjust the root mean square error by the product lead time to get standard deviation of demand over lead time and review period (Equation 6)
    c. Calculate k value for style/color/size (Equation 7) for given IFR.
    d. Determine actual safety stock target needed to meet service level (Equation 2)

It is important to note that the first step uses the heuristic employed by the current system, which takes the root mean square error for a style/color and disaggregates it to the sizes by the size curve. Since size data is readily available, Step 2 uses the size curve to create forecast at the style/color/size level and then can directly determine a forecast error and RMSE at the style/color/size level.

For instance, let us say that a given style/color/size has a safety stock target of 1000 units in the current system (Step 1) and the actual safety stock target needed to meet the service level requirement is found to be 1200 units (Step 2). We can infer several things.

First, the actual service level will be less than the target service level because 1200 units are needed to meet this level whereas only 1000 units are allocated. Given the actual allocated quantity of safety stock and the size level characteristics of the product, the same group of equations can be used to back calculate an expected service level.

Second, there is a mismatch in the safety stock quantities between what is given in the current system and the actual product attributes. In this case, the safety stock is off by 20%. Future scenarios would look at ways to bridge this gap while working in the current system.

It is also important to note that the safety stock allocated could be higher than what is needed. In this case, the expected service level would be higher than the target and an ideal system would reduce this quantity.

Size Curve Optimization Trials

Next, the project sought to further improve and ultimately attempt to optimize the size curves used within Nike's inventory target calculations. As sizes curves employ historical data, this analysis looked to determine if historical data can be used differently to provide a more accurate size curve. Based on the literature review and the fact that size curves were assumed to change relatively little across a given style/color (no trend or seasonality), different moving average and simple exponential smoothing techniques were employed with varying parameters. Due to the large quantities of data, it was not possible to use a solver to empirically fit these parameters. For this reason, moving averages used 4, 6, 8, 10, 12 (months) and cumulative for its parameter, and the simple exponential smoothing parameter, α, varied from 0.01 to 0.10.

$$\hat{s}_{t+1} = \frac{1}{n}\sum_{t=1}^{n} \frac{D_{S/C/S,t}}{D_{S/C,t}}$$

Equation 18: Size Curve Forecast Using Moving Average Methodology

$$\hat{s}_{t+1} = \alpha\left(\frac{D_{S/C/S,t}}{D_{S/C,t}}\right) + (1-\alpha)\hat{s}_t$$

Equation 19: Size Curve Forecast Using Exponential Smoothing Methodology

Where

$\hat{s}_t$= estimate for the size curve at time period t

$D_{S/C/S,t}$ =Demand (sales) for the style/color/size in time period t

$D_{S/C,t}$ =Demand (sales) for the style/color in time period t

Furthermore, the analysis looked at the possibility of clustering products based on attributes such as style; silhouette and product category; and silhouette and product sub category. The intuition is that sizing is unlikely to vary drastically between similar products, so clustering may reduce the noise induced into the size curves. Clustering was only used with the top performing methodologies as described below.

In order to measure the relative efficacy of the different methodologies and associated parameters, two metrics were used: mean demand weighted average of size-level root mean square error and total inventory of safety stock required as determined by Step 2 above. These metrics were chosen because a more accurate size curve should reduce forecast error leading to reductions in safety stock inventory. Both metrics were used because safety stock is the metric most pertinent to Nike, whereas forecast error is more directly traceable to the size curve accuracy. While the two metrics are interrelated, there is not a linear relationship between the two. Statistical means testing could not be implemented due to the skewness and non-normality observed within these outputs.

Safety Stock Adjustment Concept Model

Once the current state was analyzed and optimal size curves selected from the above methodology, a third model was used to apply a safety stock correction and measure the impact. The purpose of this heuristic is to correct the cases when the system sets (say) 1000 units of safety stock when (say) 1200 units of safety stock is actually needed to hit service level targets.

Before going into details, it is important to note some implementation assumptions that drove the model process. First, the project only looked at continuing to forecast at the style/color level. If forecasting and safety stock calculations were to be done at the size level, size curves would no longer be necessary and any discrepancies would disappear.

Next, it was assumed that the planning system would determine safety stock quantities at the style/color level and disaggregate them by the size curve later. This meant that any corrections would need to be implemented by a simple set of rules. For instance, the same adjustment on safety stock such as a 5% increase would need to be applied to all equipment if product engine was used to separate a set of rules.

First an increase factor (inc. factor), was calculated as the average discrepancy in inventory for a given group of products. This could be found simply by taking the total quantity of inventory needed and dividing it by the quantity of inventory supplied in the current system. It is important to note that this is just an average, so applying this correction across a group of products will not push all products above the threshold of inventory needed to meet service level targets.

To account for this, a factor of safety ranging from 0 to 3 in increments of 0.2 was then paired to the increase factors. A value of 0 corresponds to no change in the safety stock targets, whereas 0.2, 1.0, and 2.0 correspond to a 20%, 100% and 200% increase in safety stock, respectively.

This range was chosen because it provided a large range of values for comparison. The model then multiplied these two factors by the safety stock allocated in the current system for each style/color/size and compared that to the safety stock required for that style/color/size as seen in Equation 20. These safety stock quantities come from the current state model, and the increase in safety stock is the left side of Equation 20 minus the current safety stock level. It is important to recall that ss_{alloc} is found from step 1 in Current State and Size Curve Review Example, and ss_{need} is found through step 2.

For example, let us say that for a given product engine the allocated safety stock is on average 10% less than the needed safety stock. This would mean that our increase factor in Equation 20 is 0.1. Using the hypothetical product with an allocated safety stock of 1000 units that actually needs 1200 units to meet target service level. Thus, a factor of safety of at least 0.091 would be needed to ensure that system allocates enough safety stock to meet or exceed service levels.

$$[ss_{alloc}] * [inc.factor] * (1 + [Factor\ of\ Safety]) \geq ? \leq [ss_{need}]$$

Equation 20: Safety Stock Adjustment Model Formula

where

ss_{alloc} = safety stock allocated to the style/color/size in the current system

inc. factor = total safety stock required divided by total safety stock allocated for a given product category

Factor of Safety = Value ranging from 0 to 3 to further modify the inc. factor

If the value on the left side of Equation 20 is greater than the safety stock needed for a given style/color/size, a binary variable was assigned a value of 1 (the case of the factor of safety exceeding 0.091 in our example). If not, the binary value was given a value of 0. Then, the percentage of style/color sizes meeting or exceeding their service level target was given by averaging the value of this binary variable. Similarly, the percentage of product volume meeting or exceeding its target service level was given by doing a weighted average of the binary variable by a demand. These were two key outputs for future inventory decision making.

Next, the model calculated financial and inventory data needed to make business decisions around inventory. Feedback from project stakeholders indicated that inventory measured in days of sales in inventory (DSI) would be most useful. DSI can be thought of as a near inverse of inventory turns and is given in Equation 21. It is typical for DSI in the retail industry to be forward looking to forecasted sales, but in this case, we are looking at historical averages of actual sales.

$$DSI_{Increase} = \frac{[ss_{adjustment}]}{[avg.annual\ demand]} * [365]$$

Equation 21: Average Days of Sales in Inventory Increase

Then, additional annual revenue was calculated by taking the service level increase from the safety stock addition multiplied by annual demand multiplied by wholesale cost. Holding cost was calculated by looking at the average inventory increase multiplied by the average inventory holding cost multiplied by the product landed cost.

Using these values, project NPV could be calculated using Equation 22. The project duration was assumed to be five years based on feedback from project stakeholders. After which, the project would drive no additional costs or revenue. Given the percentage of products and product volume meeting or exceeded their prescribed service level and these financial and inventory metrics, tradeoff curves can then be constructed to inform future inventory decisions.

$$NPV = \left(\frac{\left(\left(\frac{[Rev]}{SL_{target}} \right) (SL_{inc}) - \left[SS_{Cost} \right] \right) [Margin](1 + [Growth])(1 - [Tax])}{[WACC] - [Growth]} \right) \left(1 - \left(\frac{1 + [Growth]}{1 + [WACC]} \right)^{[Duration]} \right)$$

Equation 22: NPV Calculation for Defined Annuity with Growth

Where

Rev. = Annual Revenue

Duration = Assumed length of project benefits (5 years)

Tax = Nike AA tax rate

WACC = Weighted average cost of capital

Growth = Assumed growth rate for Nike AA business

SS_{cost} = Additional annual inventory holding cost from project

SL_{target} = Current target service level

SL_{inc} = Increase in service level due to project

Margin = Profit divided by revenue

This page intentionally left blank.

Critique on the Mathematics of Size Curve Applications

Earlier in the paper it is reported that safety stock quantities can be thought of as being calculated for a given style/color and then disaggregated to the style/color/sizes. While this is effectively true, the execution is accomplished by proportioning the style/color standard deviation by the size curve and then calculating individual safety stock quantities for each SKU as seen in Equation 17. The underlying assumption for that relationship is given by the general rule for manipulating product variances shown below.

$$Var[aX] = a^2 Var[X]$$

Equation 23: Variance Multiplied by a Constant

It is the author's opinion that this relationship needs further modification to be appropriately applied in its current context. Equation 17 purports that we are scaling a style/color by the size curve. In this sense small, medium, large, et cetera are all just derivatives of the same style/color and behave the same way. This is clearly not the case. While there is no doubt a high correlation in the sizes within a given style/color, these are independent variables. The demand for each individual size is stochastic and not perfectly correlated to the sale of other products.

Ergo, we need to think of the style/color variance as a sum of correlated random variables. The general equation for this is given below.

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{i<j} COV[X_i, X_j]$$

Equation 24: Variance of the Sum of Correlated Random Variables

In Appendix I, it is shown that the only way to apply both Equation 17 and Equation 24 is for the products to be perfectly correlated ($\rho = 1$), which is not surprising given Equation 23. This is not to say that Equation 17 is useless. Rather, it is important for one to understand its limitations of applying it. Namely, if the products in a given style/color are not perfectly correlated, the relationship will underestimate the variance of style/color/sizes. This is due to attributing more of the style/color variance to the style/color/size covariance interaction than actually exists.

Unfortunately, Equation 24 is more complicated and does not readily yield a simple relationship to connect style/color variance, style/color/size variance and the size curve. This again points back to why Equation 17 is used by Logility.

This page intentionally left blank.

Model Results

Current State Analysis

The model showed several interesting results regarding expected service levels in AA products. For one, it is readily apparent that service level is not consistent across all sizes with the majority of the discrepancies occurring in fringe sizes. Second, we see that only products offered in a single size such as "miscellaneous" or "one-size" have adequate inventory levels to achieve desired service levels. If a hat is only offered in "one-size," there would be no sizes to disaggregate the style/color forecast to, so the safety stock would be accurately calculated.

Figure 9 further illustrates these discrepancies for apparel. As we see, all sizes fall short on their service level expectations with fringe sizes hurting the most. Figure 10 shows the same results but rolled up by product engine. Going back to our example, we know that a size that has an allocated safety stock of 1000 units but actually needs 1200 units will fall short of target service levels. We see that all sizes on average fall short of service level targets, and these effects are exacerbated in fringe sizes and footwear products.

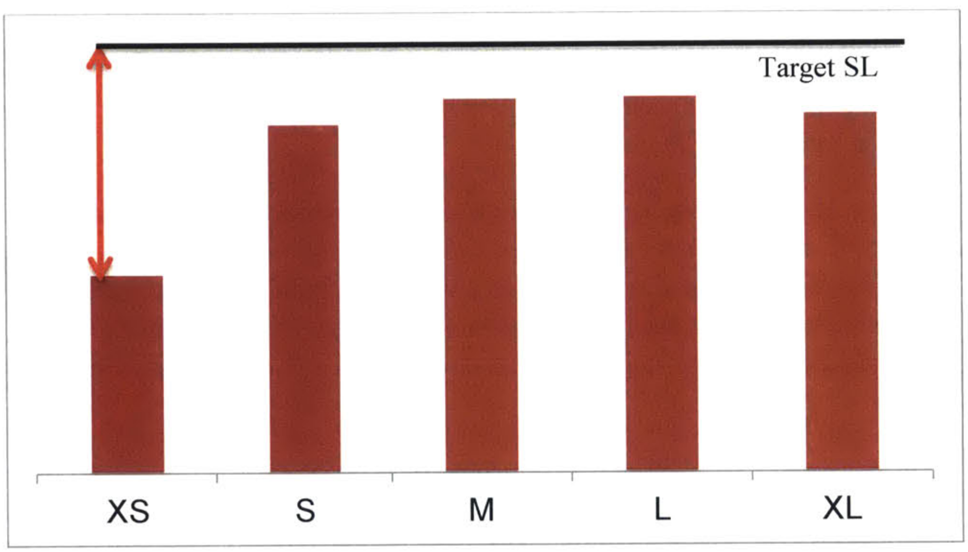

Figure 9: Expected Service Level by Size for Apparel

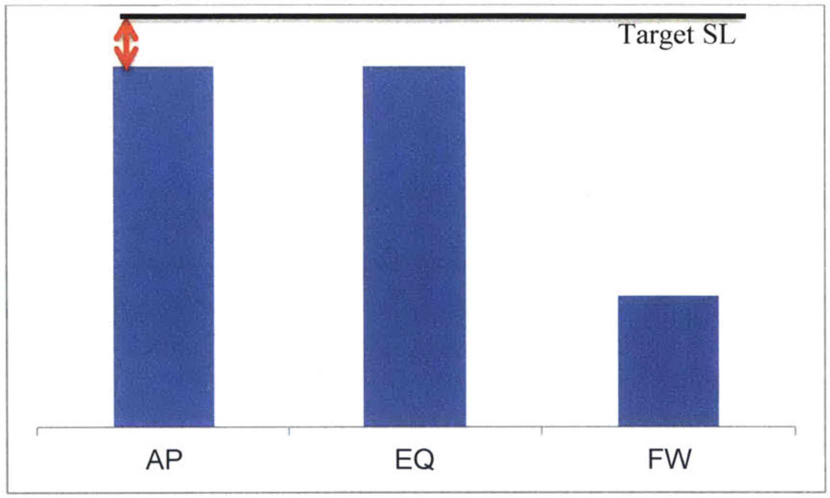

Figure 10: Expected Service Level by Product Engine

There are several ways to explain these results. The first has to do with the truisms of forecasting, which state the following:[18]

1. The forecast is always wrong.
2. Aggregate forecasts are more accurate.
3. Shorter horizon forecasts are more accurate.

While all three truisms impact the forecasting of retailers, the second truism is most germane. In the current system, forecast error and safety stock quantities are calculated at the style/color level and then disaggregated to the size level via the size curve. Since we are doing this at a more aggregated level then the actual level where the unit is sold, the forecast appears to be more accurate than it really is. This leads the system to underestimate the safety stock, so a size-level analysis finds that status quo levels of safety stock are not accurate.

[18] Caplice

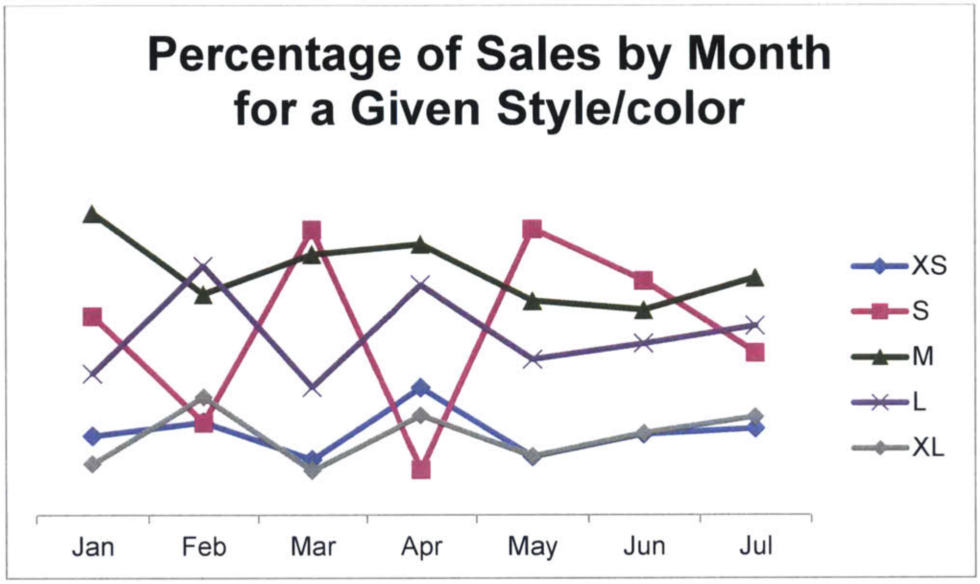

Figure 11: Percentage of Sales by Size for a Given Style/color

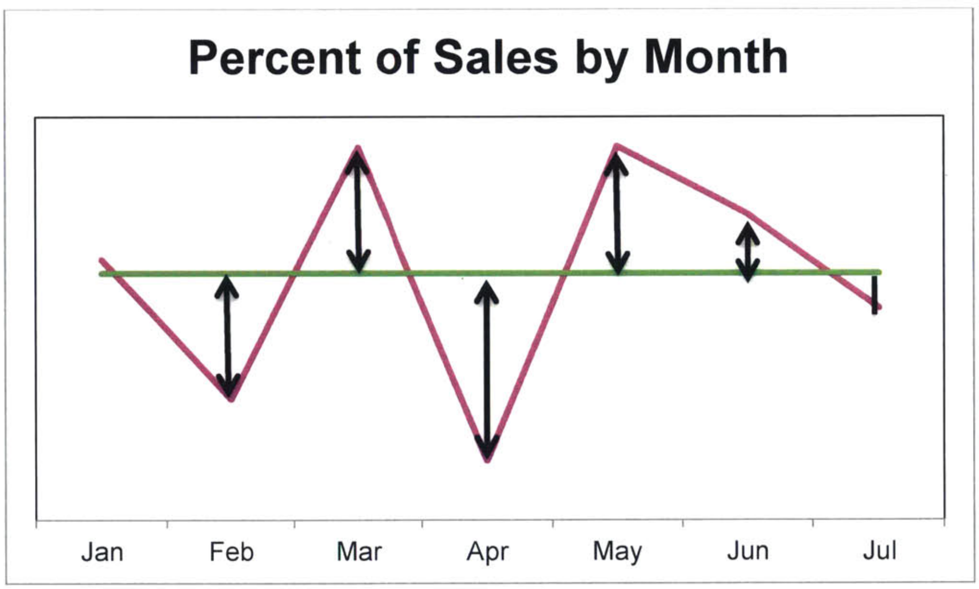

Figure 12: Percentage of Sales for One Apparel Size by Month

By not accounting for size-level variation, the current system is assuming that the size curves are 100% accurate in representing size level sales. Clearly this is not true, and Figure 11 shows an extreme example of this, and we can use Figure 12 to highlight the challenge. The pink line shows actual sales for a given size as a percentage of total sales for the style/color, and the green line signifies the size curve for this size.

Ideally, the two lines would match, but it would be naïve to assume a straight line formed off of historical consumer demand would be able to accurately predict future sales. The black lines show the difference between the predicted and actual sales represent the additional error not accounted for when determining safety stocks. This is why no size will meet target service levels.

Next, we should discuss why service levels are seen to be lowest in footwear and highest in equipment as outlined in Figure 10. The solution lies within the size offerings across the three product engines. Footwear contains the most size offerings such as 6 to 13 in half size plus 14 and 15 for Men. Equipment has the fewest size offerings with many products only being offered in one or just a handful of sizes such as a hat being offered in "One Size."

Going back to the second truism of forecasting, by aggregating across more sizes more error is reduced in footwear than in equipment. This leads to safety stock quantities within the footwear product engine to be underestimated even more so than the other product engines. Equipment on the other hand has less of this challenge due to having fewer size offerings. Products with one size are assigned proper safety stock quantities because no disaggregation is necessary.

Finally, it is important to determine why fringe sizes suffer the greatest discrepancy between their target and expected service levels. Anecdotally speaking, fringe sizes sell slower and less frequently than core sizes making them more variable and harder to predict. As discussed earlier, safety stock is allocated by the size curve, which is a function of mean demand, yet safety stock is used to protect against variability.

Size Curve Improvement

Table 1 shows us that there is a relatively tight clustering within the different methodologies within apparel. Even some of the large relative differences are seen to be relatively small when looking at absolute values, which must be masked to protect Nike's competitive information. In the tables, MA refers to the moving average trials and ES refers to trials that employed exponential smoothing techniques. The number next to the method represents the trial number not the parameter used.

Footwear and Equipment showed similar results with different methodologies proving more effective for those product engines. These results are included in Appendix II for the sake of brevity.

Method	Cluster	RMSE Improvement	Safety Stock Improvement
MA1	1	2.53%	1.80%
MA1	2	2.45%	10.14%
MA1	N/A	2.24%	8.78%
MA1	3	2.13%	0.53%
ES7	2	0.18%	-0.40%
ES7	1	0.17%	-6.67%
MA5	2	0.17%	0.63%
MA5	1	0.11%	-5.01%
MA3	2	0.01%	1.68%
Status Quo	**Status Quo**	**0.00%**	**0.00%**
MA3	1	-0.05%	-4.53%
ES6	N/A	-0.07%	-3.09%
MA4	N/A	-0.08%	-3.37%
ES7	N/A	-0.09%	-4.12%
ES5	N/A	-0.09%	-3.34%
MA6	N/A	-0.12%	1.21%
ES7	3	-0.12%	-8.00%
ES4	N/A	-0.13%	-3.12%
MA5	3	-0.14%	-5.16%
MA3	N/A	-0.15%	1.44%
ES8	N/A	-0.15%	-9.13%
MA7	N/A	-0.16%	1.49%
ES2	N/A	-0.28%	-3.99%
ES9	N/A	-0.37%	-13.34%
ES1	N/A	-0.67%	-4.69%

Table 1: RMSE and Safety Stock Increases for Size Curve Methodologies for Apparel

Given these small differences and the large effort required to implement changes to the status quo, it was decided that the size curve methodology should not be changed. Going back to Figure 12 and applying our new size curves, we can clearly see the challenge presented in Figure 13. Size curves remain relatively static, while sales for a given product will fluctuate from month to month, and no stationary line will be able to accurate predict these swings using only historical data. While there is no doubt that some of the aggregate safety stock reductions look impressive, there was no systematic rule to get this reduction across all product engines. Thus, it makes sense to keep the current methodologies, since the status quo is a top tier performer for all product engines.

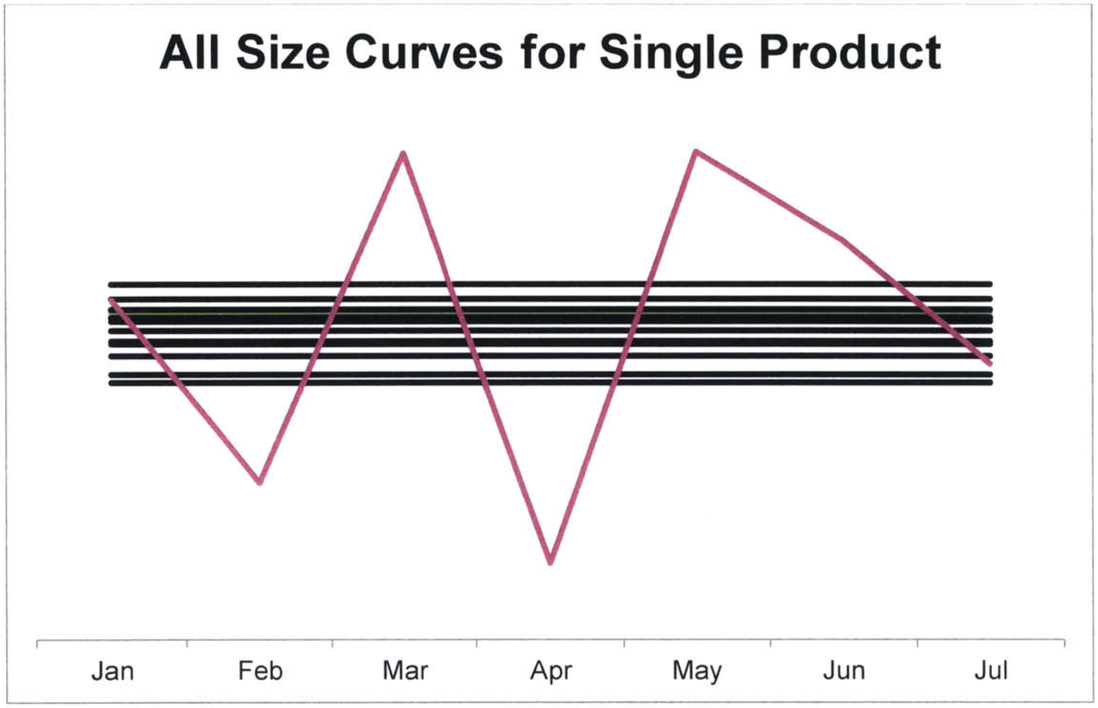

Figure 13: All Size Curve Methodologies for One Apparel Item

Right-Sizing Safety Stock

Finally, using the relationship in Equation 20, the model explored the differences in current safety stock targets versus safety stock necessary to achieve desired service levels. As expected, inventory targets were undersized in all scenarios. The results differed by geography and product engine, but the overall results are given in Figure 14 and Figure 15 with individual product engine results in Appendix II.

These curves were calculated by varying the factors of safety across the different product engines as described above. This would change the safety stock levels, which was reflected in the DSI increase over the status quo. Since additional safety stock also increases service level, more sizes will exceed the quantity of safety stock needed to meet their target service level as given in Equation 20. Furthermore, the higher service levels are assumed to lead to more captured demand resulting in higher revenues. These revenues are weighted against the additional inventory holding costs over an assumed project lifecycle of five years in order to determine the project NPV.

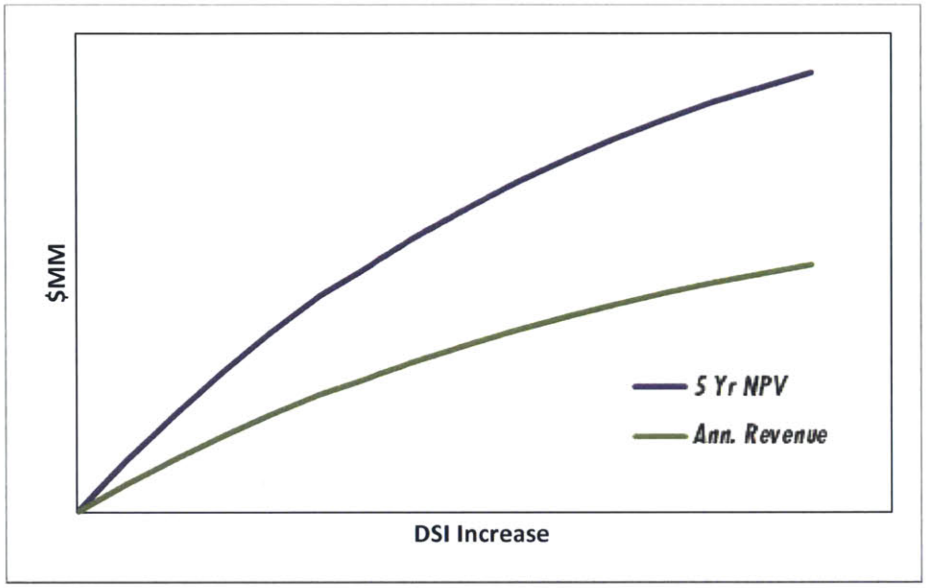

Figure 14: Overall Monetary Benefits for Given DSI Adjustment

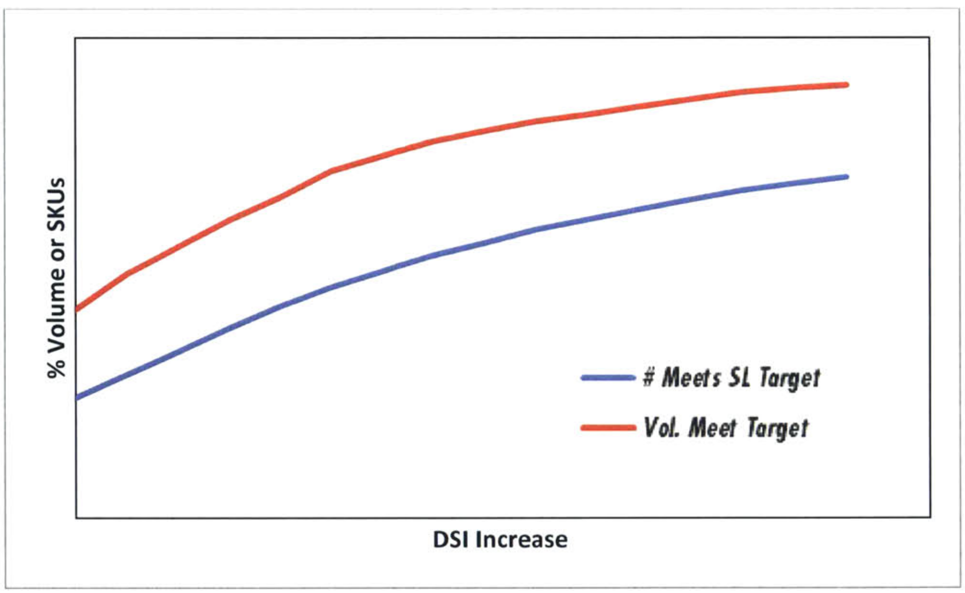

Figure 15: Overall Service Level with DSI Adjustment

In these figures, several things stand out despite the exact numbers being hidden. For one, increasing safety stock within the tested ranges always yields positive increases in NPV and revenue. This is due to the fact that holding additional inventory will increase service level, which will lead to more product sold. Then, given the margin structure at Nike, the profit outweighs the holding cost leading to positive NPV. Clearly for extremely large quantities of inventory, the NPV will go negative due to the non-linear relationship between inventory and service level, but these quantities of inventory are so large that even the most extreme factor of safety of 3.0 did not reach this point. If we were to continue to increase the factor of safety well beyond 3.0, we would expect to see the NPV to decrease due to this non-linear relationship.

Second, we notice a large increase in sizes meeting service level for a given DSI increase, which then tails off. This is again due to the nonlinear relationship between inventory and service level. We can see that initial increases in inventory will lead to gains in the number and volume of sizes meeting their service level targets, but further increases show diminishing returns (as seen by the concave nature of Figure 14).

More will be discussed in the recommendations section, but by increasing safety stock targets to account for the additional error at the size level, we can achieve size-level service level targets for the vast majority of products and sales volume, while using style/color level planning for forecasting and inventory management. Ultimately, this leads to increased profitability.

Recommendations and Implementation

Using the data collected throughout the project as well as the results of the simulations two recommendations were made. First, it was observed that the AA portfolio consists of a wide array of sizes. By reducing the size offerings on AA, the geographies will be able to focus on delivering core sizes only, and system constraints such as unaccounted size-level forecast error will be reduced due to forecasting fewer sizes. In general these fringe sizes need a disproportionately large amount of inventory, while contributed less to the bottom line. For instance, analysis showed that one geography could reduce their style/color/size offering for a product engine by 22%, while maintaining 99% of revenue. These tradeoffs would still need to be weighed against the customer service impact.

Second, it is recommended that the size curve update process be automated. Currently, the updates are performed manually, which leads the work to not be standardized. Furthermore, automating the updates will ensure that only sizes intended to be sold are included on AA, and adjustments to safety stock quantities to account for the additional size-level error can easily be incorporated into this new system. The following subsections go through these recommendations in more detail.

Size Standardization

Many sizes are offered across the AA portfolio within the three product engines. Men's footwear, for instance, can carry any sizes between 5 and 18. This can be a challenge for several reasons. For one, not all sizes are sold at the same rate. The fringe sizes are sold in much lower volumes than core sizes. As the sales team seeks to appease customers, new sizes are often added without consideration to the true costs associated with carrying an extreme fringe size. While meeting customer needs is not necessarily a bad thing, a conscious, data-driven decision should be made to ensure the best course of action is taken.

Second, these sizes are not consistently offered across all products. Going back to men's footwear, size 18 may only be carried for some products on the AA line. This creates confusion with customers as there is no clear standard, since these sizes exist for some products and not others.

Finally, more sizes mean more work across the Nike value chain. Demand and inventory planners have more unique SKUs to manage and update, and manufacturing needs to manage more unique product. Also, as many of these fringe sizes are low volume and highly variable demand, the current inventory systems described above are not able to adequately plan for them. These leads to either low service levels with no intervention or high inventory levels if inventory targets are overridden. Neither of these situations are ideal unless the customer benefit is there.

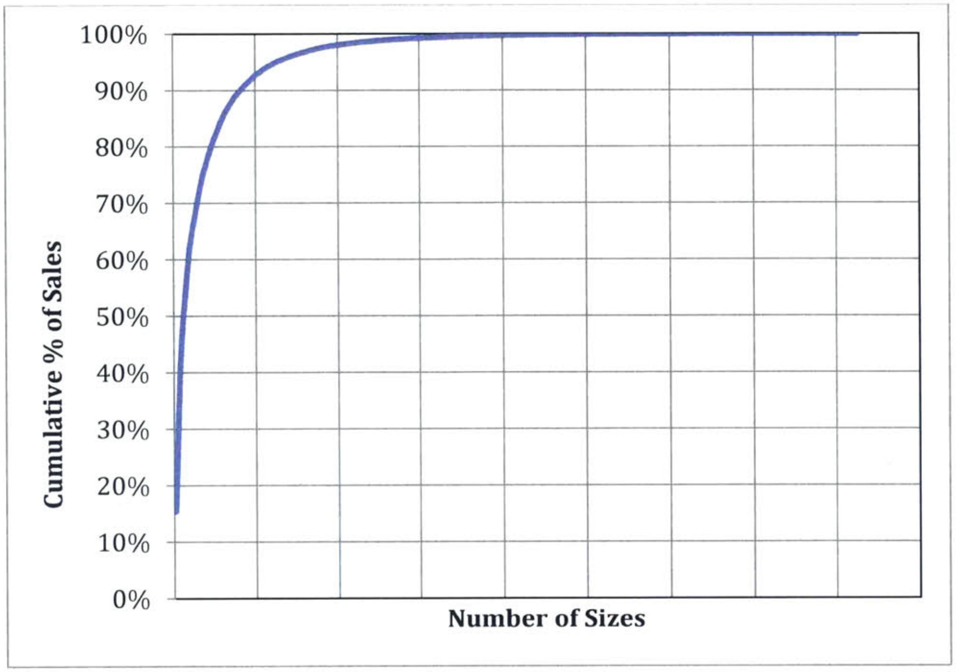

Figure 16: Size/Gender/Age as Percentage of Overall Volume

As we see in Figure 16, a small number of sizes make up a majority of sales. In fact, the bottom performing 65 percent of sizes account for less than 1 percent of sales and 2 percent of on hand inventory. Again, many sizes offered on the tail are necessary to meet customer needs, but it shows that additional scrutiny could be applied to the size portfolio. The Pareto curve was constructed by stacking up the percentage of sales from a given size from largest to lowest percentage.

To standardize the selection of sizes across the portfolio, the author proposes the following methodology:

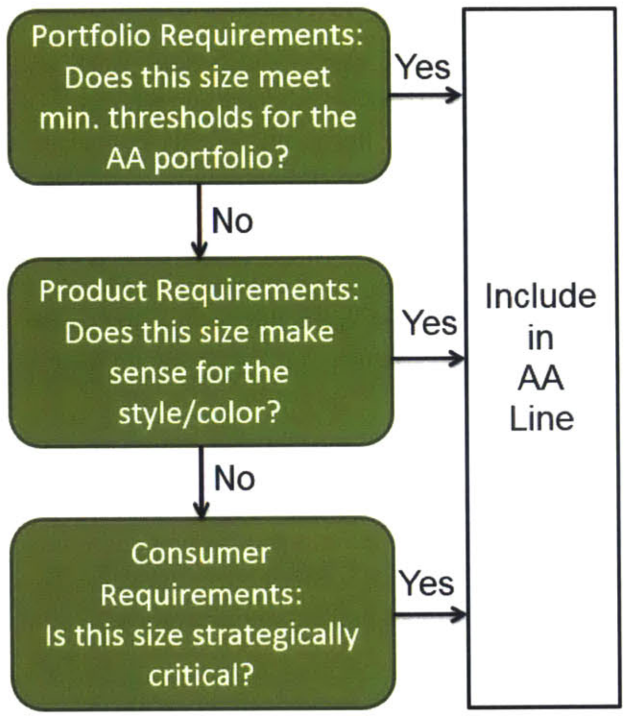

Figure 17: AA Portfolio Size Inclusion Decision Matrix

While specific criteria cannot be shared publicly, it is suggested that a three tier approach be employed in determining what sizes should be offered. First, sizing decisions should be made across a gender and age portfolio. Going back to the men's footwear example, one would first want to consider what sizes are to be offered on every men's shoe. The criteria would be based on volume and the total percentage of sales that the given size makes up.

From there, the algorithm would begin to look at individual style/colors. As some products within the portfolio will be better sellers than others, it may make sense to include additional sizes in these products. Similar to the initial step, the algorithm would look at both absolute and relative sales volume.

Finally, it remains important to incorporate any additional intangibles that would not be picked up by analytics. These could be to appease a customer or block a competitor, but the final step would call for a manual review of the size portfolio.

In order to determine specific cutoffs, a cost/benefit analysis will need to be performed. Namely the cost of holding inventory for a certain size needs to be weighed against the potential revenue benefit. There are also benefits to the network for carrying certain sizes. This is because many retailers want to be able to order a full-size run from AA, so by not holding a size 15 shoe in a given style/color, for instance, a retailer may choose to not order any shoes of that style/color. This information can be used to make data-driven decisions at the management level.

Currently, the AA line is reviewed on a regular business cadence. It is recommended that sizing be considered in these reviews.

Size Curve Automation

In order to integrate the recommendations of right-sizing safety stock and standardizing sizing, an automated size curve system is needed. This would have an algorithm resetting size curves using historical data on a regular cadence. Based on the results of this project, it is recommended that Nike keep its current methodology in place. It was observed that other methods do not consistently outperform the status quo and have the potential to over fit the data.

The current system requires manual updates of size curves by inventory planners on a regular basis, which is both time consuming and the cause of variability in the size curves. The automation would provide a further advantage in addition to standardization by freeing up the time of planners and allowing them to focus on other tasks.

The current size curve process requires inventory planners to download size curve reports that compares current size curves to historical sales for a fixed timeframe. Planners use this information to update the size curves stored in SAP. The current process recommends timing and tolerances for updates, but these recommendations are applied differently across geographies, product engines and categories. By including automation, size curve updates will be applied consistently across all products, which will lead to more predictable forecasting at the size level.

Next, looking at the recommendations made in this paper, a large amount of governance would be required to ensure implementation. To remove sizes, planners would need to zero out the size curves for these sizes. Given the fluidity of the process and the number of people involved with size curves, it is not practical to assume this can be done with the precision needed to garner the full benefits. Thus, we turn back to the automated system as the most practical way to implement the project recommendations.

Once the system is developed and tested, the new process would be rolled out to inventory planners and supervisors. It is expected that this could be accomplished within six months.

Future Opportunities

Following the work of this project, there are several things that can still be done. First, clear guidelines must be agreed on for sizing inclusion in the AA portfolio. Much like the criteria used to determine if a product should be included in AA, clear guidelines around sizing will ensure consistency across the portfolio leading to a streamlined supply chain and clear customer expectations.

Next, the system architecture for an automated size curve update process should be designed and implemented. The author lacks expertise in this area, so the task was left to IT professionals and consultants within the company. Creating a system that links the two distinct databases and a corresponding user interface is necessary for the project to continue to progress. This aspect of the project will need to include training and awareness to the different teams using size curves because at a minimum they will be providing overrides.

Finally, a decision of when to make the safety stock adjustments for size level error is still needed. The author has recommended that this be done in conjunction with an upcoming change in the safety stock system. Less training and awareness will be needed as few people are directly involved in setting safety stock quantities. Getting management buy-in on changes to inventory levels will be the remaining decision.

This page intentionally left blank.

Summary and Conclusion

Challenges exist in adequately sizing safety stock to meet desired service levels. This primarily stems from how safety stock is first calculated and then allocated across the sizes available within a given style/color. The system calculates forecast error at the style/color level, which does not account for the full error seen after size curve are applied to a forecast. Furthermore, these calculated safety stock quantities are then allocated across a style/color based on the size curve, which is not always consistent with a products forecast variability. Another way to think of this is that the current systems assumes sales of sizes within a given style/color are perfectly correlated. Since this is not the case, aggregate levels of safety stock will be underestimated.

To counter these effects, several recommendations have been made. First, by consolidating and standardizing size offerings, there will be fewer sizes taxing the current systems and planners making the sizes that are offered more likely to meet their target service level. It was seen that 65 percent of sizes account for less than 1 percent of sales, so this consolidation will not have a material impact on bottom line profits.

Next, safety stock targets need to be right-sized to account for forecast error caused by the application of a size curve. It is clear that safety stock is underestimated in most scenarios. By increasing safety stock targets for sizes based on their geography and product engine, the numbers of SKUs and sales volume meeting or exceeding their target service level can be increased. This increase is determined by weighing the tradeoffs of inventory increases (DSI), revenue increases and service level increase as seen in Figure 15 and 16 as well as Appendix III.

Finally, a system that automatically updates size curves on a regular basis is necessary to ensure proper implementation of these recommendations. There is the additional advantage that the automation will standardize the performance of all products in the AA portfolio and free up the time of inventory planners.

Nike's AA business model will remain relevant and necessary to quickly meet Nike's customer needs. By implementing these recommendations, it will be possible to continue to meet customer expectations while keeping the right amount of inventory in place.

This page intentionally left blank.

References

Anderson, Eric T., Gavan J. Fitzsimons, and Duncan Simester. 2006. "Measuring and Mitigating the Costs of Stockouts". *Management Science* 52 (11): 1751-1763. doi:10.1287/mnsc.1060.0577.

Brown, Robert G. 1963. "Estimating Aggregate Inventory Standards". *Naval Research Logistics Quarterly* 10 (1): 55-71. doi:10.1002/nav.3800100105.

Caplice, Chris. 2014. "Introduction to Demand Planning & Forecasting". Lecture, Massachusetts Institute of technology, ESD.260.

Corsten, Daniel and Thomas Gruen. 2004. "Stock-outs Cause Walkouts." *Harvard Business Review* 85(5): 26-28

Craig, Nathan, Nicole DeHoratius, and Ananth Raman. 2016. "The Impact Of Supplier Reliability Tracking On Customer Demand: Model And Estimation Methodology". *SSRN Electronic Journal.* Accessed January 31. doi:10.2139/ssrn.1679792.

Ferreira, Kris Johnson, Bin Hong Alex Lee, and David Simchi-Levi. 2015. "Analytics for an Online Retailer: Demand Forecasting and Price Optimization." *Manufacturing & Service Operations Management M&SOM*, November 13, 2015. doi:10.1287/msom.2015.0561.

Gross, Charles W., and Jeffrey E. Sohl. 1990. "Disaggregation Methods to Expedite Product Line Forecasting". *Journal Of Forecasting* 9 (3): 233-254. doi:10.1002/for.3980090304.

Hyndman, Rob J., and Anne B. Koehler. 2006. "Another Look at Measures of Forecast Accuracy". *International Journal Of Forecasting* 22 (4): 679-688. doi:10.1016/j.ijforecast.2006.03.001.

Kang, John. 2015. "Inventory Optimization Model for NIKE'S Long Lifecycle Highly Seasonal Replenishment Products". LGO Thesis, Massachusetts Institute of Technology.

Kim, Minjeong, and Sharron J. Lennon. 2011. "Consumer Response to Online Apparel Stockouts".*Psychology And Marketing* 28 (2): 115-144. doi:10.1002/mar.20383.

Nike, Inc. Form 10-K for the Fiscal Year Ended May 31, 2015. Nike Inc. website. Accessed January 31, 2016. http://s1.q4cdn.com/806093406/files/doc_financials/2015/ar/docs/nike-2015-form-10K.pdf

Polak, Ben. 2014. "Multi-Echelon Inventory Strategies for A Retail Replenishment Business Model". LGO Thesis, Massachusetts Institute of Technology.

Sbrana, Giacomo, and Andrea Silvestrini. 2013. "Forecasting Aggregate Demand: Analytical Comparison of Top-Down and Bottom-Up Approaches in a Multivariate Exponential Smoothing

Framework". *International Journal of Production Economics* 146 (1): 185-198. doi:10.1016/j.ijpe.2013.06.022.

Shlifer, E., and R. W. Wolff. 1979. "Aggregation and Proration in Forecasting". *Management Science* 25 (6): 594-603. doi:10.1287/mnsc.25.6.594.

Silver, Edward A, D. F Pyke, and Rein Peterson. 1998. *Inventory Management and Production Planning and Scheduling*. New York: Wiley.

Appendix I: Proof of Product Correlation Relationship

Looking at the relationship between correlation and the variance of the sum of correlated random variables X_i.

Given,

$$\sigma_{s/c}^2 = Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + \sum_{i>j} 2COV[X_i, X_j]$$

Assuming

All random variables are correlated equally to each other ($\rho_1 = \rho_i = \rho_n = \rho$)

$Var(X_i) = s_i^2 \sigma_{s/c}^2$

where

s_i is the size curve corresponding to product X_i such that $\sum_{i=1}^{n} s_i = 1$

Applying these assumptions

$$\sigma_{s/c}^2 = Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + \sum_{i>j} 2COV[X_i, X_j] = \sum_{i=1}^{n} Var[X_i] + 2\rho \sum_{i>j} \sigma_{X_i}\sigma_{X_j}$$

$$= \sigma_{s/c}^2 \sum_{i=1}^{n} s_i^2 + 2\rho \sum_{i>j} s_i \sigma_{S/C} s_j \sigma_{S/C}$$

$$= \sigma_{s/c}^2 \sum_{i=1}^{n} s_i^2 + 2\rho \sigma_{S/C}^2 \sum_{i>j} s_i\, s_j$$

$$= \sigma_{s/c}^2 \left(\sum_{i=1}^{n} s_i^2 + \rho \sum_{i \neq j} s_i\, s_j \right)$$

$$= \sigma_{s/c}^2 \left(\sum_{i=1}^{n} s_i^2 + \rho \sum_{i \neq j} s_i\, s_j + \rho \sum_{i=1}^{n} s_i^2 - \rho \sum_{i=1}^{n} s_i^2 \right)$$

$$= \sigma_{s/c}^2 \left(\sum_{i=1}^{n} s_i^2 + \rho \sum_{i=1}^{n} \sum_{j=1}^{n} s_i s_j - \rho \sum_{i=1}^{n} s_i^2 \right)$$

$$= \sigma_{s/c}^2 \left(\sum_{i=1}^{n} s_i^2 + \rho \sum_{i=1}^{n} s_i \sum_{j=1}^{n} s_j - \rho \sum_{i=1}^{n} s_i^2 \right)$$

Based on the unity requirement for size curves

$$= \sigma_{s/c}^2 \left(\sum_{i=1}^{n} s_i^2 + \rho - \rho \sum_{i=1}^{n} s_i^2 \right)$$

Solving for ρ

$$\rho = \frac{1 - \sum s_i^2}{1 - \sum s_i^2} = 1$$

Thus, for our assumption about the relationship between style/color variance, style/color/size variance and size curves to be accurate, all sizes must be perfectly correlated.

Appendix II: RMSE and Safety Stock Increases for Size Curve Methodologies

Method	Cluster Type	RMSE Improvement	Safety Stock Improvement
MA5	1	0.25%	-11.93%
MA5	2	0.24%	-6.96%
ES7	1	0.21%	-11.36%
MA3	1	0.16%	-8.39%
MA3	2	0.16%	-4.00%
Status Quo	**Status Quo**	**0.00%**	**0.00%**
ES7	2	-0.03%	-8.39%
MA4	N	-0.05%	-6.23%
ES1	N	-0.06%	1.62%
MA5	3	-0.09%	-35.05%
MA3	N	-0.09%	1.40%
ES6	N	-0.12%	0.16%
ES7	N	-0.15%	-1.69%
ES7	N	-0.20%	-2.25%
ES7	N	-0.29%	-1.72%
ES7	3	-0.38%	-34.95%
ES7	N	-0.39%	-1.17%
ES5	N	-0.50%	-3.22%
ES4	N	-0.66%	-6.99%
ES2	N	-1.06%	-15.35%
ES1	N	-1.84%	-48.81%
MA1	1	-2.95%	-11.80%
MA1	2	-2.95%	-2.95%
MA1	N	-3.19%	1.70%
MA1	3	-3.41%	-47.53%

Table 2: RMSE and Safety Stock Increases for Size Curve Methodologies for Equipment

Method	Cluster	RMSE Improvement	Safety Stock Improvement
ES2	N	0%	-5%
MA3	N	0%	-3%
ES6	N	0%	-3%
ES1	N	0%	-3%
ES1	N	0%	-7%
ES4	N	0%	-2%
MA3	2	0%	-4%
Status Quo	**Status Quo**	**0%**	**0%**
MA1	N	0%	-1%
ES7	2	0%	-3%
MA5	2	0%	-2%
ES7	N	0%	-2%
MA4	N	0%	2%
MA3	1	0%	-24%
ES7	N	0%	5%
MA5	3	0%	-5%
ES7	3	0%	-6%
ES5	1	0%	-24%
ES4	1	-1%	-25%
ES2	N	-1%	4%
ES1	N	-1%	5%
MA1	N	-10%	-1%
MA1	2	-11%	-11%
MA1	1	-11%	-33%
MA1	3	-11%	-12%

Table 3: RMSE and Safety Stock Increases for Size Curve Methodologies for Footwear

Appendix III: Benefits and Service Level Tradeoffs with DSI by Product Engine

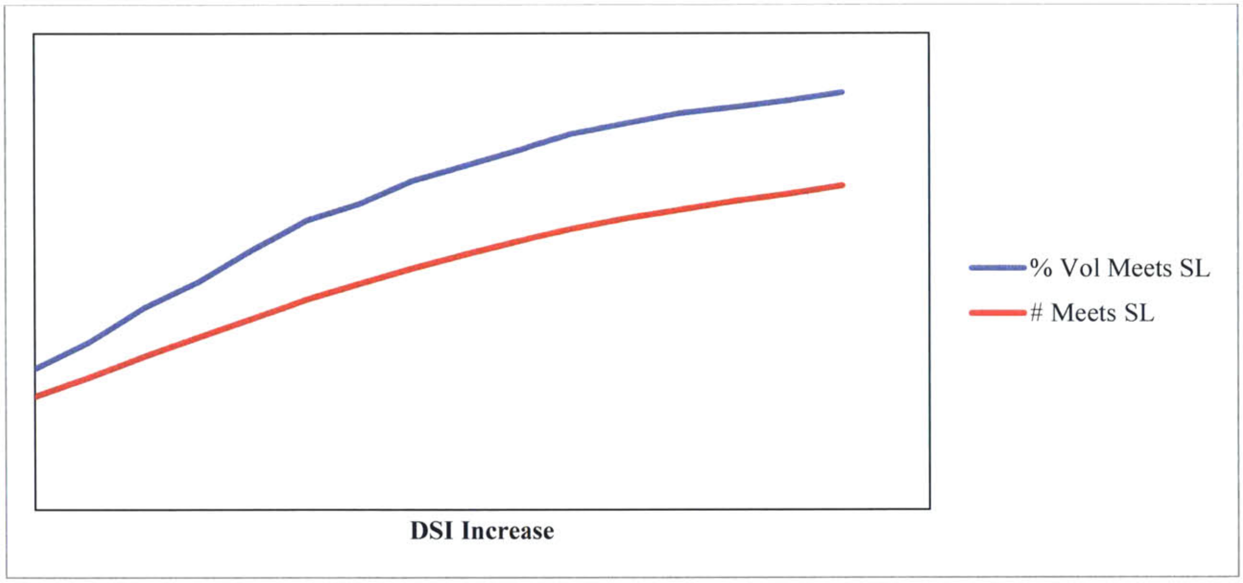

Figure 18: Apparel Service Level with DSI Adjustment

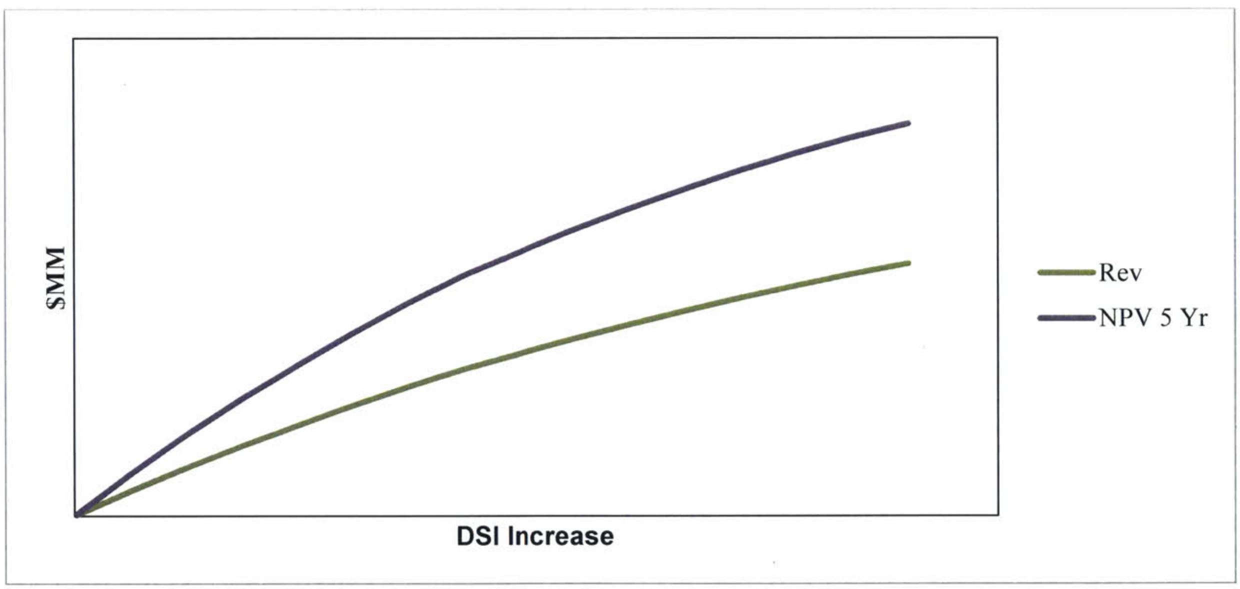

Figure 19: Monetary Benefits with DSI Adjustment for Apparel

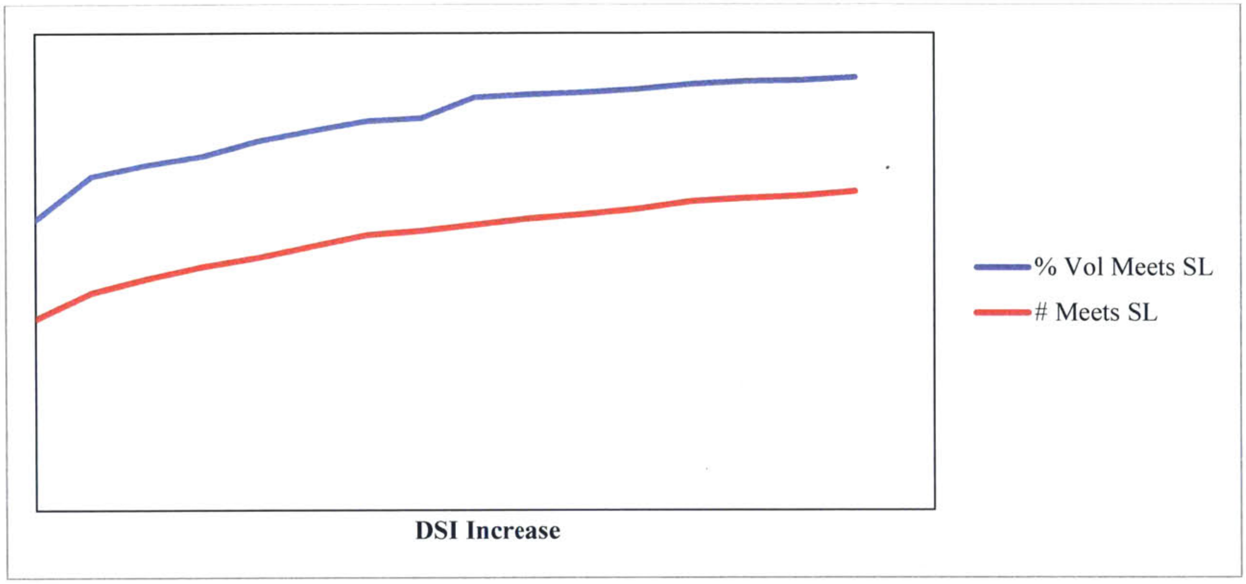

Figure 20: Equipment Service Level with DSI Adjustment

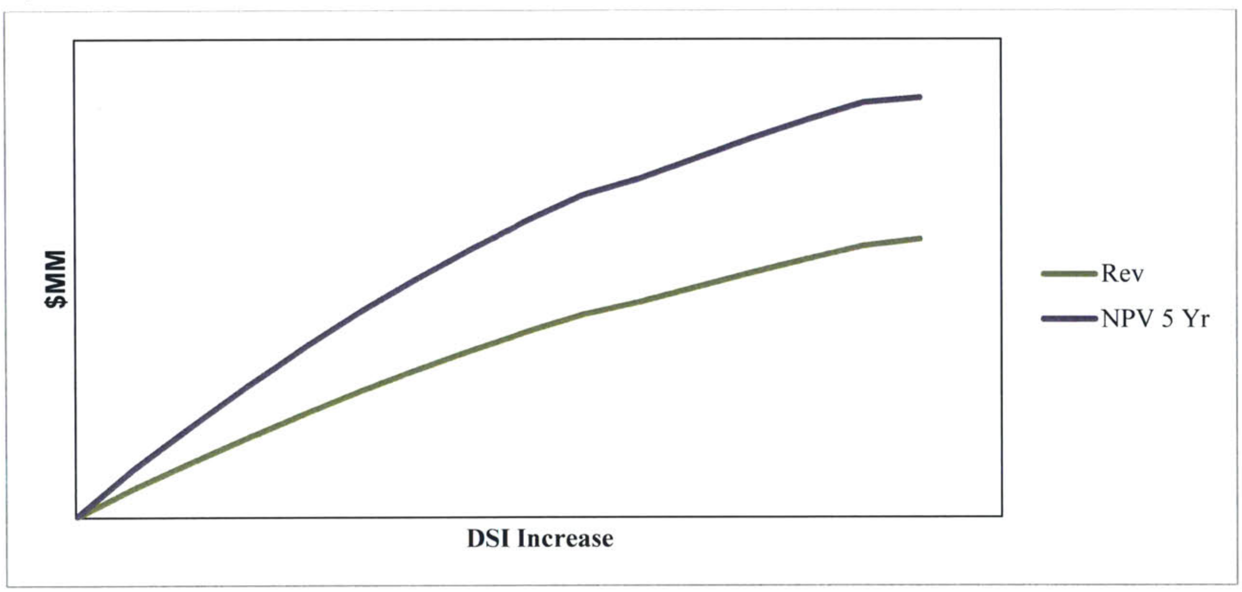

Figure 21: Monetary Benefits with DSI Adjustment for Equipment

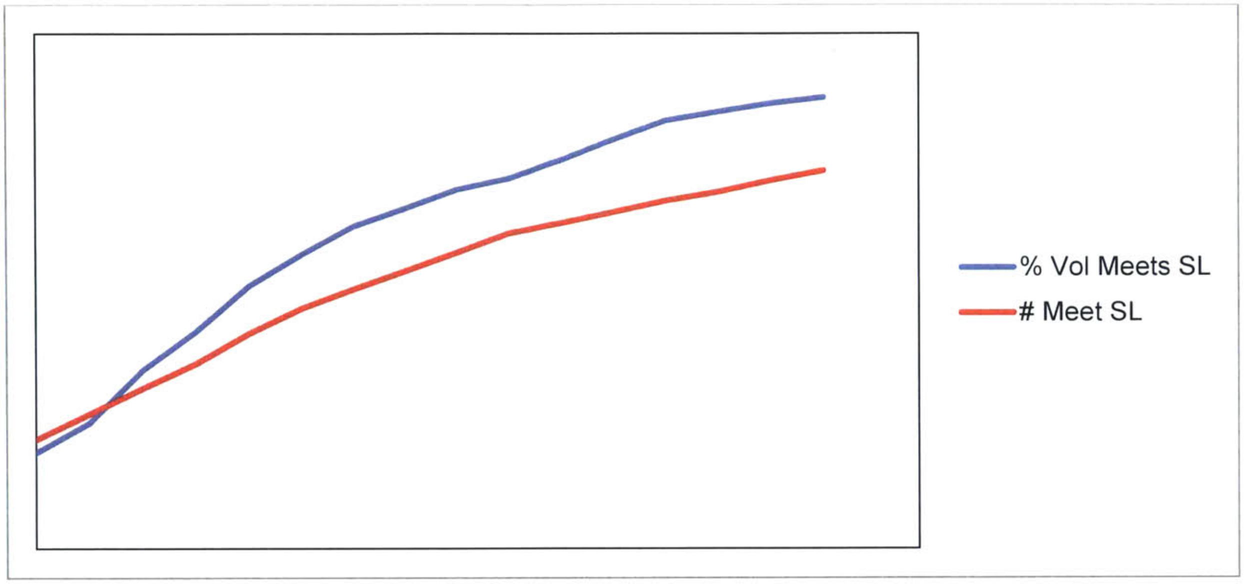

Figure 22: Footwear Service Level with DSI Adjustment

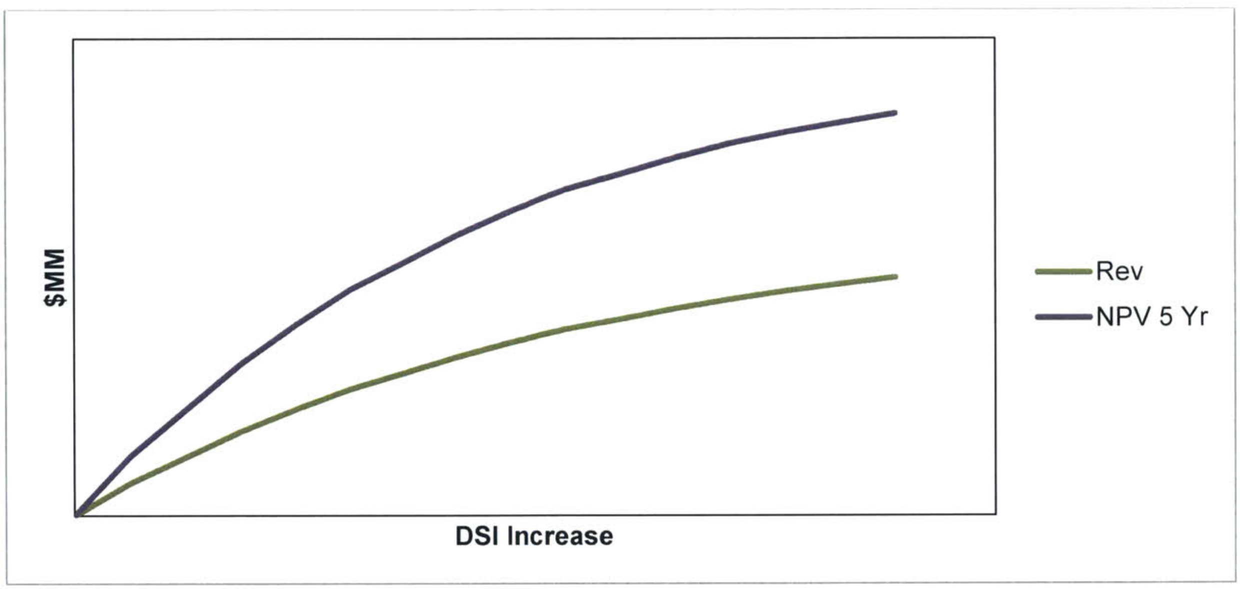

Figure 23: Monetary Benefits with DSI Adjustment for Footwear