Analyzing the Accountability, Systems and Efficiency of Demand Planning Processes in a Consumer Products Environment
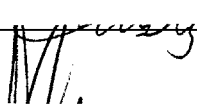
By

Christy A. Prilutski

B.ChE. Chemical Engineering, University of Delaware (1998)

Submitted to the Sloan School of Management and the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration

and

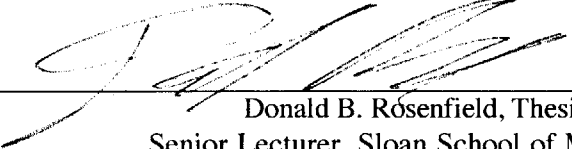Master of Science in Civil and Environmental Engineering

In Conjunction with the Leaders for Manufacturing Program at the Massachusetts Institute of Technology
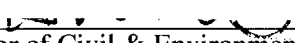
June 2006
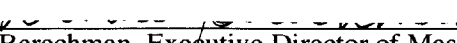
Signature of Author _____

Sloan School of Management
Department of Civil and Environmental Engineering

Certified by _____

Jonathan Byrnes, Thesis Supervisor
Senior Lecturer, MIT Engineering Systems Division

Certified by _____

Yossi Sheffi, Thesis Reader
Professor, MIT Department of Civil and Environmental Engineering

Certified by _____

Donald B. Rosenfield, Thesis Supervisor
Senior Lecturer, Sloan School of Management

Accepted by _____

Andrew J. Whittle, Professor of Civil & Environmental Engineering
Chairman, Committee for Graduate Students

Accepted by _____

Debbie Berechman, Executive Director of Masters Program
MIT Sloan School of Management
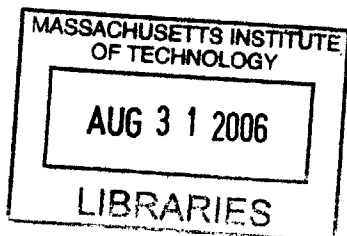
Analyzing the Accountability, Systems and Efficiency of Demand Planning Processes in a Consumer Products Environment

By
Christy A. Prilutski

Submitted to the Sloan School of Management and Department of Civil
Engineering on May 6, 2006 in partial fulfillment of the Requirements
for the Degrees of Master of Business Administration and
Master of Science in Civil and Environmental Engineering

As consumer products companies like P&G strive to achieve a consumer driven supply network, the value of forecast accuracy comes into question. Many companies push for faster cycle times and shorter supply chains, driving towards make-to-order production. These trends may appear to reduce the importance of forecasts. However, a closer look into P&G and their business reveals that sales forecasts are still very important and have a far reaching impact stretching from the supply network through to Wall Street. This thesis evaluates the forecasting process in a company like P&G. The thesis delves into the accountability around sales forecasts and proposes a top-down, statistical process for creating and tracking forecast accuracy which was implemented across P&G's global organization. Another analysis is conducted on the evaluation of a new demand planning system which provides more granular input data for generating forecasts, and the implications from this trial on the demand planning process. Finally, an assessment on the efficiency of current systems is also detailed.

Thesis Supervisor: Don Rosenfield
Title: Senior Lecturer, Sloan School of Management

Thesis Supervisor: Jonathan Byrnes
Title: Senior Lecturer, MIT Engineering Systems Division

Thesis Reader: Yossi Sheffi
Title: Professor, MIT Department of Civil and Environmental Engineering

This page is intentionally left blank.

This page is intentionally left blank.

Table of Contents

1 Introduction

The objective of this thesis is to evaluate the processes surrounding forecast accuracy within the consumer products organization of Procter and Gamble, P&G.

1.1 Thesis Motivation

Significant excitement exists around shortening and simplifying supply chains, and creating make-to-order systems. This emphasis of having responsive and flexible supply chains calls into question the need for forecasts. Whether in discussions with students studying supply chain and operations, or even global process owners within P&G's organization, a consistent questioning of the importance of forecasting was evident. The following quote from Sam Ouliaris exemplifies a common message, "If we could always adjust instantaneously and costlessly to new conditions there would be no need for forecasts."[1] My research at P&G indicated that even as the organization makes major strides in moving towards a more responsive supply chain, forecast accuracy is still critically important to the business. This thesis will provide sufficient background to allow readers to obtain a basic understanding of P&G's business and then will present ideas surrounding the importance of forecast accuracy. The thesis will back up these ideas with a review of an initiative to drive forecast accuracy accountability, an evaluation of an investment to improve forecast accuracy, and an analysis of the efficiency of the current demand planning process. The objective of the thesis is to evaluate the importance of forecast accuracy to P&G, and provide enough detail to allow readers to apply the lessons to their own organizations.

1.2 Thesis Overview

Procter and Gamble has a successful history of building and maintaining brand leadership in the consumer product marketplace. Their success lies not only from excellent management, innovation and marketing, but from leveraging their size to capture economies of scale. By driving innovation in finding the most efficient business processes, P&G lowers the cost of bringing products into the hands of consumers, allowing them to spend more money on activities such as marketing and product development. In order to continue along this model of success, integrating supply chains between P&G and the retailers will be necessary to reach the next breakthrough in efficient supply chains. Additionally, P&G looks inside the organization for cost savings and process improvement. This thesis introduces P&G, its organization, and supply chain in Chapter 2. Next, Chapter 3 seeks to assert the importance of forecast accuracy, and the effects poor forecasting has in the organization. The effects include the cost of inventory, lost sales from not having product on the shelf, cost of expedited procurement, cost of lost resources when deploying people and capital assets to projected business growth, and the hit to the stock

---

[1] Ouliaris, Sam. "The Art of Forecasting" NUS Business School & Monetary Authority of Singapore, http://www.ess.org.sg/The%20Art%20of%20Forecasting.ppt#256,1,The Art of Forecasting, accessed 3/1/2006.

price from poor forecasting. These issues are rising in importance, and more attention is being drawn towards improving forecast accuracy.

Having set the stage for forecast accuracy being very important to business, the remainder of the thesis evaluates forecast accuracy within the P&G Organization. Chapter 4 introduces the team at P&G responsible for demand planning, the vocabulary and processes employed, and the relationship to the rest of the organization. A relatively large organization at P&G, the demand planning group is well organized and follows standard procedures worldwide to forecast weekly demand at the SKU level for all of P&G products. Chapter 5 describes the technical details of the measures P&G employs to assess forecast accuracy.

Understanding the P&G approach to forecasting and measuring accuracy, the next three chapters outline two internal and one integrated approach to improving forecast accountability, accuracy, and efficiency. Chapter 6 presents a process for deploying metrics top-down using statistics to cascade tolerance limits depending on the size, complexity, and capability of organizations. The development of the process and implementation is discussed. Chapter 7 evaluates the pilot of an integrated system connecting P&G's customer service system with the customer's distribution center. Predicted accuracy savings were not realized. The most promising hypothesis seem to be that the opportunity for applying this system was not as large as originally believed and that organizational challenges hindered implementation and adoption of new systems in a timely and cost-effective manner. The system tested in Chapter 7 did show improvements, but at a huge financial and organizational cost. This introduced the idea of efficiency, and the cost of improvement, leading into an efficiency study in Chapter 8. Following on a similar study at Radio Shack[2], the efficiency of the P&G forecast output was compared against simple statistical models to evaluate areas for efficiency improvement.

Lastly, an organizational review is presented to demonstrate the importance of P&G's strategy, politics and culture on corporate efforts. Each of these areas is analyzed and the particular impact of these areas on the projects is discussed.

---

[2] Harwell, Jack and Michael Gilliland. "Applying Forecast Value Added at Radio Shack." Conference Presentation at the Institute of Business Forecasting Supply Chain Forecasting Conference, Phoenix, AZ, March 1, 2005.

2 Procter and Gamble: The world's greatest consumer product company

This chapter provides an introduction to Procter and Gamble, setting the background for the thesis. Clearly stated in the 2005 Annual Report, the vision of P&G is:[3]

> "Be, and be recognized as, the best consumer products and services company in the world.
>
> Our Promise: Two billion times a day, P&G brands touch the lives of people around the world. P&G people work to make sure those brands live up to their promise to make everyday life just a little bit better.
>
> Touching lives, improving life. P&G."

While leaving a heart-warming feeling, these words do not speak to the fundamental statistics and strategy which allow for P&G to fulfill the vision expressed.

2.1 History

Formed as a soap and candle business by two brothers-in-law in 1837, P&G had a humble beginning. Surviving difficult market conditions and the civil war, the company grew to a multi-million dollar corporation by 1890. At the same time, P&G built one of the first product research labs in America. This investment into research led to the development of new products including soap for clothes and dishes, and Crisco, the first all-vegetable shortening. In 1920, P&G chose to sell directly to retailers avoiding the uneven production needs from dealing with wholesalers, changing the way the grocery trade would operate.[4]

Throughout the 1930s and 40s, Procter and Gamble expanded into beauty products, shampoo, and toilet goods. At the same time, P&G developed a system for brand management and began advertising on both radio, and eventually TV. In fact, the sponsorship of radio programs by P&G's soap powder led to the term, "soap opera" to be used for these popular radio programs. Even in these early days, P&G advertising and sponsorship was invading US households. P&G also started early with international expansion, and in 1935 P&G acquired Philippine Manufacturing Company, the first operations in the Far East.[3]

In 1946, the introduction of Tide laundry detergent proved so successful that it funded new growth into other product lines and other markets. Between 1946 and 1980, P&G would expand into new products including Crest toothpaste, toilet tissue and paper towel, Pampers disposable diapers, Pringles potato chips, Folger's coffee and Downy fabric softener. P&G also expanded

---

[3] P&G Annual Report 2005.
[4] "A Company History" P&G, 2005

geographically starting on-the-ground operations in Mexico, Europe and Japan. By 1980 P&G was doing business in 23 countries around the world with sales of over $10 billion. [5]

Throughout the last two and a half decades, P&G has continued to expand into new product markets including health care and cosmetics, as well as geographic markets as it now sells its products in over 160 countries worldwide.

2.2 Key Competitive Advantages

P&G publicizes that its core strengths lie in four main areas: branding, innovation, go-to-market, and scale. [6] These core strengths lead to a competitive advantage in the market place.

2.2.1 Branding

Known for marketing expertise, especially their frequently copied methodology of "brand management," P&G has over three hundred brands. P&G prides itself on building large brands which are hugely popular globally. P&G introduced the concept of "billion dollar brands" referring to brands which independently generates over $1 billion dollars in sales each year. P&G has 17 brands in this category in 2005, and with the acquisition of Gillette, will add another 5 brands anticipating 22 brands in 2006. Some of these brands are quite huge. Pampers alone generates over $6 billion in sales each year. Another favorite, Tide, generates $3 billion.

Figure 2.1 Billion-Dollar Brands[6]

Having twenty-two brands each selling over a billion dollars in sales each year may seem impressive. But the pipeline is also exceptional, and P&G has another thirteen brands with sales of over $500 million in the pipeline. Each of these is capable of crossing the billion-dollar mark in the next several years.

---

[5] "A Company History" P&G, 2005
[6] P&G Annual Report 2005.

2.2.2 Innovation

P&G's size, resources and history allows P&G to have massive repositories of consumer and market knowledge spanning the globe and decades of transactions. P&G leverages this knowledge, along with science and engineering expertise to continually innovate both products and processes. P&G has collected consumer data from more than 100 million consumers across 30 countries and 20 categories of business. Additionally, P&G invests nearly $2 billion a year in research and development, more than most of their direct competitors combined. This knowledge and commitment to investment drives P&G innovation and brand building and strengthens the value provided to both retail partners and consumers. [7]

2.2.3 Go-to-market

P&G can leverage its knowledge, size and expertise to developing strong relationships to bring products to consumers through the retail channel. P&G provides retailers with consumer and shopper research, supply chain solutions, branding and marketing expertise, and more. In a recent industry survey of U.S. retailers, P&G was ranked #1 in six of eight categories: clearest strategy, most innovative, most helpful consumer and shopper information, best supply chain management, best category management, and best consumer marketing. [7]

2.2.4 Scale

With $56 billion dollars in annual sales, not including a recent acquisition of Gillette which adds another $10 billion in annual sales, P&G is one of the world's largest consumer products organizations. With over 100,000 employees operating in over 80 countries, the large organization can function with significant economies of scale P&G both creates and capture value from its massive size in areas such as purchasing, distribution, and business services at the company level.[7]

2.3 Moments of Truth

P&G's CEO, AG Lafley, has created a vision of two moments of truth in which P&G needs to win to be successful. This vision has energized the organization and focused people on two core ideas. The first moment of truth occurs in the store, when the marketing, pricing, and availability of a product entice a consumer to purchase it. The second moment of truth occurs when the consumer uses the product, and is pleased with the quality enough to become a repeat user.

2.4 Organization

P&G also implemented an organizational change to setup a structure around these two moments of truth, which would provide the ability to integrate new businesses while continuing to build

---

[7] P&G Annual Report 2005.

existing businesses. P&G is divided into three major areas: Global Business Units (GBU), Market Development Organizations (MDO), and Global Business Services (GBS). The main idea is to capture the benefits of focused smaller companies through dedicated GBUs while capturing the go-to-market strengths and capabilities of a $50 billion company through local market development organizations. Lastly, shared business services organization and lean corporate functions groups ensure P&G's functional disciplines will continue to lead the industry.[8]

Global Business Units manage the profit and loss for brands globally. They develop and implement long-term strategies for P&G brands. Operating globally, they can leverage deep consumer understanding to determine common consumer needs and quickly expand brands and product innovations to different markets around the world. GBUs are measured on shareholder return and their development of strong market leadership positions in their individual industries. The GBUs include functions such as global marketing, product development, and product supply. P&G divides its businesses into three main GBUs: Beauty Care, Household, and Health, Baby and Family Care. The GBU is focused on winning the "second moment of truth" – when the consumer uses the product and evaluates how well the product meets their expectations – by focusing on delivering superior products, packaging and marketing.

Market Development Organizations are responsible for understanding local needs and leveraging their scale in distributing to customers in local regions. MDOs are responsible for the entire portfolio of brands in a particular geography, and this breadth provides flexibility to serve the needs for local retailers. These regional MDOs can focus on the customers, retailers, supply chains and governments associated with local markets. Essentially functioning like a sales and customer service organization, they are aligned behind top-line growth, market share, cash, cost, and value-creation objectives. The key advantage of this structure is that MDOs can focus 100% of their resources on local consumers and customers without duplicating product innovation, product sourcing, brand advertising or other activities that are led by the Global Business Units. This eliminates inefficient overlaps and frees up resources to collaborate better with customers and focus exclusively on winning in local markets. Thus, the MDO is focused on winning the "first moment of truth" – when a consumer stands in front of the shelf and chooses a product from among many competitive offerings.

Each of the Global Business Units is headed by a vice-chairman reporting to the chief executive officer of the company. All of the MDOs also report to the Vice-Chairman of Operations. Thus employees within the GBU or MDO do not officially connect in the organization until reaching the CEO level.

Global Business Services, the third major organization, operates as the "back office" for the GBU and MDO organizations, providing world-class technology, processes and standard data tools to understand the business and serve consumers and customers better. This organization includes functions such as accounting, IT and human resources.

Two other smaller corporate organizations also exist at P&G: Global MDO and Corporate Functions. Corporate functions operate across the GBUs for functions typically conducted

---

[8] P&G 2004 Annual Report.

within the GBU. These functional organizations include marketing, finance, and product supply. They support the GBUs ensuring functional discipline and installing industry best practices. Likewise, the Global MDO supports functions which occur across the different regional MDOs, such as sales, physical distribution, and demand planning.

2.5 Customer versus Consumer

There is an important distinction between the term customer and consumer. P&G uses 'customer' to refer to the retailers, primarily mass merchandisers, grocery stores, membership club stores and drug stores that purchase products from P&G. The term 'consumer' refers to the person who purchases the product and uses it in the household. So while P&G works to improve relationships with their customers to reduce costs and improve service, the purpose statement of P&G stresses the importance of the consumer:

> "We will provide branded products and services of superior quality and value that improve the lives of the world's consumers."

In order to provide superior quality and value to today's demanding, diverse and changing marketplace, each brand maintains a significant amount of variety and is constantly refreshing the product portfolio and to add new products. This adds complexity to the relationships with customers, a critical part of the supply chain.

2.6 Supply Chain: Scope, complexity and magnitude

The supply chain refers to the network of suppliers, manufacturing plants, distributors, and retailers that participate in the sale, delivery and production of a particular product. The emphasis on the first moment of truth indicates the importance of P&G's supply chain in ensuring the product is on the shelf at the retailer and able to be sold.

P&G has a large supply organization, operating over 100 manufacturing plants, prior to the Gillette acquisition, and engaging in numerous contract manufacturing relationships.[9] P&G also has its own distribution network, in addition to that of its customers. Selling over 50,000 products to mass merchandisers, grocery stores, and drug stores worldwide, P&G products can be purchased from a huge number of retailers worldwide. This supply chain is depicted below.

---

[9] P&G leaders, including Richard Clark, Mark Kremblewski and Jim Yuhas, provided much of the background about the company and its supply chain.
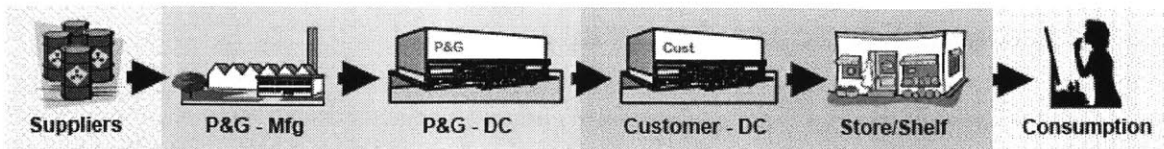
Figure 2.2 P&G Supply Chain[10]

This picture shows the many components in the channel between manufacturing the product and finally getting the product into hands of the consumer. The breadth and size of P&G further complicate the system. Strong differences occur between businesses and regions. For example, the relative importance of customers varies significantly by region. Over 30% of the sales volume passes through Wal-Mart in North America. However, in Western Europe there are more relatively equivalent retailers responsible for selling to consumers. One example can be used to highlight both the importance of Wal-Mart, and supply chain policies such as inventory management. Inventory reductions at Wal-Mart are cited by the Dow Jones Newswires as triggering the forecast update at P&G, which caused the stock price to fall 2.3%.[11]

The size and scope of P&G's product variety also complicates the supply chain. First, from a size perspective, P&G operates sixty Sales and Operations Planning (S&OP) processes around the globe each month. S&OP refers to process where various departments within P&G come together to convert the business plan into an operating plan, balancing and determining the appropriate demand and supply plans. There is also significant variety with the supply processes for different P&G businesses. For example, some businesses are quite responsive and have the flexibility to change the schedule of what they plan to make a few hours before it is actually produced. These same businesses execute the planning and scheduling processes multiple times a day. On the other hand, some businesses plan production weeks in advance, and fix the schedule a month out, leaving little flexibility for last minute changes. In this wide variety of business environments, the importance of forecast accuracy begins to unfold.

---

[10] Cheung, Christine. "A Short-range Forecasting and Inventory Strategy for New Product Launches." MIT Thesis. 2005.
[11] Lloyd, Mary Ellen. "Procter & Gamble Down After Tempering E Europe, Asia Outlook" Dow Jones Newswire. March 14, 2006.

3 Importance of Forecast Accuracy

This chapter will analyze the importance of forecast accuracy and the direct impacts it has on the organization. The forecast enables the first moment of truth – having the product available on the shelf.

3.1 Inventory Levels

Many supply chain and operations textbooks and journal articles discuss the direct relationship between forecast error and inventory levels. With perfect accuracy, and instantaneous capacity, no inventory would be required in the system as companies could plan to make the right quantity of equipment exactly at the right time. However, it is not feasible that companies would have infinite instantaneous capacity. Thus, even if the forecast accuracy is perfect, inventory will be required to ensure that demand can be met. This is referred to as process tock, pipeline stock, or WIP. Once forecast error is introduced, safety stock inventory rises to manage the variability in demand. The ownership of inventory ties corporate dollars which could have been invested otherwise. Thus, over-predicting sales can lead to excess costly inventory levels. It can also lead to costs of inventorying the material, warehouse, handling and management costs. Additionally, as this unpredictability of demand amplifies finished product inventory, it can pass through to raw material inventory safety stock, requiring higher levels in this category as well. Finally, excess inventory created from poor forecasting may not be able to be sold at all, with the rapid product life cycle occurring for many consumer goods. P&G refers to this inventory as remnants, excess inventory without any plans to sell.

3.2 Customer Service

The relationship between forecast accuracy, inventory, and instantaneous capacity demonstrates how having additional inventory or capacity protects against forecast error. However, if the inventory is carried in the wrong product, a consumer may find the shelf empty of his or her favorite variety. This refers to another factor impacted by forecast accuracy, customer service. Customer service refers to the availability of goods for sale on the shelf. Under-predicting sales can lead to supply shortages resulting in empty shelves in the store. This would result in a failure at the first moment of truth as a customer can not find the particular product that they desire. Thus, this can result in lost sales to the customer. This is not always a major issue. Sometimes a customer will wait until the material is in stock, purchase a similar SKU still from P&G, or travel to a different store that does have the SKU in stock. However, there is a risk of devastating long term effects on profitability if the customer were to purchase a competitors product and cease being a repeat user of P&G's product, and switch to the competitor's product. This lost revenue can extend for long periods of time and be exceptionally costly to the business. Therefore P&G strives to have a high level of customer service, keeping product on the shelf. Other mechanisms can be used to minimize out-of-stock occurrences on the shelf, for example, rush shipping of product. But this premium freight is also expensive and reduces the overall profits of the enterprise.

3.3 Procurement Costs

The impact of poor forecasting on inventory levels stretches beyond finished product to raw materials. It can also be felt in many other procurement areas, including the expedition costs of rushing material, in the demurrage cost of not being able to offload excess material, in negotiating future vendor relationships and in lining up adequate supply in the future. For example, one plant listed detention and demurrage costs in excess of $10,000 per week. Under-forecasting raw material requirements can lead to a shortage of material, forcing a shipment to a customer to go with fewer cases than requested. This would not be reflected in forecast error, as the forecast is compared with actual shipments, and not customer demand. Additionally, raw material shortages can lead to unplanned downtime as equipment is wasted without being able to make product. To provide an example of the magnitude of potential savings in this area, one of the P&G manufacturing plants spends in excess of one billion dollars per year on raw material cost. If the forecast accuracy could be improved, and thus the variability reduced, a significant percentage of raw materials, and thus cash, could be freed up for better ventures.

3.4 Business Planning

Not only does forecast error have a direct impact on the daily inventory and costs, and product sales, it also impacts longer term decision making. Forecast accuracy is important to driving big business decisions and can impact pricing decisions, timing of new product introductions, and investment decisions for the allocation of resources between R&D, sales, and marketing. For example, some businesses carry extra capital assets to be able to deliver instantaneous capacity, where these investments could be avoided if better forecasting were available. Even when business budgets are allocated to new product development and the introduction of a new product, P&G customer service managers give examples of how development production has been required to shift to actual production for immediate sale, delaying initiative timing of new products. Capital investment, R&D investment, and marketing budgets decisions hinge on appropriate forecasts.

3.5 Sales Projections to Financial Analysts

Lastly, the course of work at P&G highlighted the importance of accurate forecasts in areas beyond sales and demand. Forecast accuracy is important for building credibility on Wall Street and protecting the company's stock price. Many journal articles discuss the relationship between earnings forecasts and stock performance. Copeland writes:

> "Changes in noise, measured by changes in the standard deviation of analyst forecasts of earnings, have a significant negative correlation with total return to shareholders. If the

company can decrease "noise"—that is, the dispersion of expectations across analysts and investor forecasts — then its share price should increase."[12]

This essentially asserts that forecast error reduces the stock price, and improving accuracy should improve the stock price. One article expresses how superior forecasting has a great impact on stock prices.[13] Another interesting article studied how since investors reward firms that meet or exceed earnings expectations, executives of firms with sizable option components in their compensation plans have increased incentives to report earnings that meet or exceed analysts' forecasts.[14] It further went on to investigate the correlation between options granting and a history of exceeded expectations. This last article touches upon the management motivation for introducing forecast bias to ensure that forecasts are always met. This bias can be difficult to impossible to remove from the system under the current policies. However, in light of the fact that error has a strong effect on stock price, it is evident that improvements in forecast accuracy are better for the shareholders.

---

[12] Copeland, Tom, Aaron Dolgoff, and Alberto Moel. "The Role of Expectations in Explaining the Cross-Section of Stock Returns" Review of Accounting Studies. 9, 149–188, 2004

[13] Park, Chul W. and Earl K. Stice. "Analyst Forecasting Ability and the Stock Price Reaction to Forecast Revisions" Review of Accounting Studies. 5, 259–272 (2000)

[14] Bauman, Mark P and Kenneth W. Shaw. "Stock option compensation and the likelihood of meeting analysts' quarterly earnings targets." Review of Quantitative Finance and Accounting. (2006) 26: 301–319

4 Demand Planning

Forecast accuracy hinges on the demand planning organization, whose objective is to generate volume forecasts which are timely, unbiased, and reasonable. The organization operates at an intersection of two of the major organizations within P&G.

4.1 Organizational Structure

As described earlier, the majority of P&G employees work in one of the three major organizations: GBUs, MDOs, or GBS. Global Business Services (GBS), which includes groups such as information technology, are not relevant to most of this thesis. The demand planners operate at an interesting bridge in the organization between the other two major groups. The three Global Business Units have responsibility for product supply – which includes every phase from product development, procurement, manufacturing plants and their processes. Therefore the GBU essentially owns creating and maintaining the inventory for the corporation. The Market Development Organizations are regional groups which operate across the business units and align with customers, thus coordinating sales, promotions, shipments, and customer service across the different business units. The demand planners fall within the MDO organizations and are responsible for collecting the input from the respective parties and generating the sales forecast which is then used by the business to determine how much product to make and when to make it.

As this is a fairly involved role for a large consumer products company like P&G, demand planning is quite a sizeable organization. There are over five hundred demand planners throughout the P&G organization. This is believed to be one of the largest demand planning organizations. While demand planners reside within the MDO and report through that organization, they are aligned with GBUs. For example, although the MDO is not divided by brands, the demand planners within the MDOs are. With the large breadth of P&G businesses and the number of general managers representing various countries, brands and profit centers, these planners are required handle a large volume and breadth of situations.

Turnover is another interesting consideration. There are approximately 200 new planners turning over in the role each year. Many people leave this role for other positions within P&G, thus training is a high priority of the organization, and consistent processes are necessary for bringing new people up to speed quickly. To retain knowledge, enable standard process maintenance and improvement, a Global Demand Planning organization exists. This organization has regional leads which support the planners within their regions, as well as global experts focusing on improving the processes. This work was conducted from the global demand planning organization, though funded by the product supply corporate function group.

4.2 Demand Plan Terminology

When discussing forecast accuracy, certain terms will be used repeatedly. Below, the common terms used by P&G will be presented and used subsequently in this thesis.

4.2.1 Horizon, weeks-out

The horizon refers to the time period between when the forecast is generated, and when the shipments are expected to occur. Horizon is synonymous with weeks-out. A one-week-out horizon indicates that the forecast was generated one-week before the shipment was expected to occur. The horizon is often expressed as month-n, where n is the horizon. When creating a forecast for the month of September in July, the P&G convention is to call this horizon month-1, indicating that the forecast was generated a full month ahead. A forecast for sales in September generated in August, would be referred to as having a horizon of month-0. This can be a source of confusion, as different groups within P&G use different rules for the zero period.

4.2.2 Time bucket

The phrase, "time bucket" is used to refer to the period of time for which the forecast and shipments are aggregated. For example, within P&G the most common time buckets are weeks and months. A weekly time bucket indicates that the forecast is generated for a seven-day period, thus alleviating any concern on which day within a week the shipment will occur. This weekly forecast will be compared with the shipment total for that week. Monthly time buckets are often used for evaluating business forecast accuracy. Essentially the forecast generated for a month will be compared with the month's shipments.

4.2.3 Aggregation

Aggregation refers to combining together different units to represent a broader perspective. For example, demand for different fragrances of deodorant may be aggregated together to represent an aggregated demand for deodorant. Aggregation may occur across products, across businesses, even across time. Many logistics and supply chain experts indicate that aggregate forecasts are more accurate than individual forecasts. Yossi Sheffi gives the example of how Cadillac aggregated demand across its Florida dealers, instead of forecasting each dealer individually. This resulted in much higher accuracy and improved customer service.[15]

4.2.4 Product hierarchy

Different levels of the organization will refer to different levels of products. For example, the stock keeping unit, or SKU, is the lowest level of the hierarchy and refers to a unit that is sold of a product. Sometimes P&G will replace one SKU number with a new SKU number when the image on the packaging changes, although it is the same product in the same quantity. In this case, to ease the transition between the two numbers, though different SKUs, the term, demand forecast unit, or dfu, is used to refer to the level above SKU referring to the size for which a product is forecast. Above this, products are aggregated according to similarities. The hierarchy varies based on the product, but an example is shown in the figure below:

---

[15] Sheffi,,Yossi. "A Demand for Steady Supply" Financial Times. August 22, 2005.

PRODUCT HIERARCHY

Category: Fabric Care

Brand: Tide

Packaging: Bottle

Fragrance: Clean Breeze

Sku: Tide Clean Breeze Liquid, 100 oz

Figure 4.1 Sample of Product Hierarchy

As shown in Figure 4.1, Tide, Era, Gain, Cheer, and Ivory Snow are all P&G brands of laundry detergent. Combining the concepts of aggregation, predicting the monthly sales of the laundry category is easier than predicting a particular brand. Once inside the liquid form of the Tide brand, it is more difficult to predict whether a customer will stick with Mountain Spring Tide, or switch to April Fresh. Lastly, the largest amount of forecast error would occur for the SKU, which defines the unit sold. Thus, as you move up the product hierarchy, products are aggregated together according to common features, and the forecast also tends to become more accurate.

4.2.5 Events

The term "event" refers to different instances which represent a deviation from normal day-to-day business operations. Events can have a large impact on demand, and are thus critical to the forecast. Different types of events include merchandising, price changes, new item introductions, planogram or distribution changes. Merchandising refers to special efforts to encourage sales of an existing product, such as a special advertising campaign, end-of-aisle display or other shelf arrangement, brochures, or perhaps coupons. When a new price is rolled out for an existing product line, the purchasing behavior from the customer can be impacted. For example, customers may over-order in advance of a price increase. New item introductions are difficult to

forecast as explained by the thesis of the previous P&G intern.[16] A 'planogram' is a graphic diagram showing how and where specific retail products should be placed on retail shelves or displays. Changes in the customer stores can have a large impact on product sales.

4.3 Planning Processes

This section will explain the process used by demand planning, building upon the terminology discussed previously. The over five hundred demand planners execute a weekly process where they generate a SKU-level forecast, in weekly buckets, for every week for a period of years. This process employs historical data, user input, and the use of a statistical forecasting tool which incorporates many of the important inputs to determining a forecast.

4.3.1 Factors Influencing Demand

The factors coming together in the statistical forecasting tool are demonstrated in the following diagram excerpted from P&G's Demand Planning Global Process Owner, Richard Clark.[17]
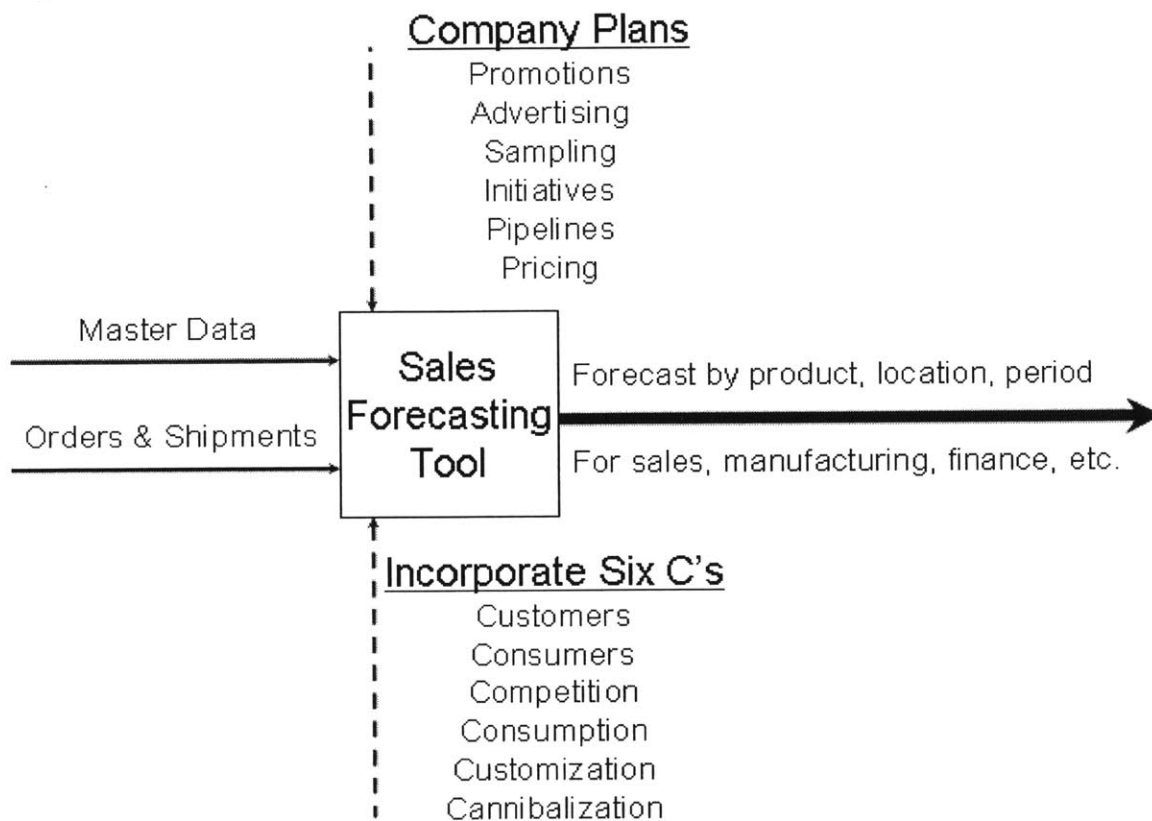
Figure 4.2 Inputs to the Sales Forecasting Tool [17]

[16] Cheung, Christine. "A Short-range Forecasting and Inventory Strategy for New Product Launches." MIT Thesis. 2005.
[17] Clark, Richard A. "Demand Planning Overview: Key Elements of P&G's Global Demand Planning Process." Internal P&G Documentation. December 9, 2004.

Figure 4.2 depicts the demand planner's use of the sales forecasting tool to generate forecasts. The sales forecast tool takes historical data, including years of shipment history, along with future orders as the base data. Additionally, master data is recorded around the size of shipping quantities, choice of forecast models, and the parameters used. The list of company plans are also important to consider. Promotions typically refer to discounts passed through to customers to encourage buying within a certain time window due to a reduced cost. Advertising can be accomplished through multiple media, and is intended to have an impact on sales. Many studies evaluate consumers and competitors response to changes in advertising and promotion.[18] Distributing of product samples, launching of new initiatives and the time and quantity to fill a new product pipeline, along with fundamental pricing changes all reflect P&G decisions which have an impact on the forecast. Additionally, the external "C's" have an impact on the sales forecast. Customers who control the channel through which the products are sold make independent decisions with regard to their advertising, promotions, store placement which impact sales. Consumer tastes and preferences continually change and impact demand. Direct competition can be beneficial by increasing the awareness of a market, or can steal demand by attracting customers away whether based on quality or cost. Previous history of consumption is important, as with consumer goods, the inventory that the average consumer has will impact buying patterns. For example, when consumers run out of toilet paper at home, they are likely to purchase another roll, regardless of price. However, if they have recently purchased a mass merchandiser case of 100 rolls, they are unlikely to purchase even more regardless of the promotion offered. Lastly, P&G can cannibalize their own products as they have multiple brands which serve the same marketplace.

4.3.2 Statistical Forecast

The first step in the demand planning process is to develop the statistical forecast. The demand planner is responsible for evaluating, selecting, maintaining and updating the forecast model which predicts sales based on historical data and model parameters. The model parameters are set based on a review of historical data, assumptions about market behavior, and an evaluation of how well different models perform at predicting future demand. The statistical forecast is the output of the computerized sales forecasting tool which predicts demand by demand forecast unit, in most cases the SKU, by week for an extended period of time.

4.3.3 Independent Net Forecast

With the statistical forecast, the demand planner classifies the SKUs according to their historical demand profile. For each supply chain, focus is placed on those which have the most impact on overall shipment variation and forecast error for the category. The demand planner meets with organizations such as the marketing and sales functions to collect intelligence on special causes which might impact demand. The planner uses historical data to determine what manual forecast adjustments are warranted from the information. This generates the "independent net forecast," the final forecast which is communicated to the organizations within P&G.

---

[18] Ailawadi, Kusum L., Donald R. Lehmann, and Scott A. Neslin. "Market Response to a Major Policy Change in the Marketing Mix: Learning from Procter & Gamble's Value Pricing Strategy" Journal of Marketing, 65, 44-61, 2001.

4.4 Relationship to Organization

Demand planning requires significant interaction with the other organizations at P&G. Organizational motivations can have an impact on the forecast. In some markets, regional sales input might be inclined to over-forecast demand to ensure that product is available and no shortages occur. Since inventory is owned by the GBU, not the MDO, there is no penalty for the MDO if excess material is not able to be sold. In a softening business, market representatives may encourage planners to keep demand high to justify additional spending. Thus much of the work of the demand planner is to drive out bias introduced by organizational motivations. Much of the success of the demand planner comes from managing by influence.

The output of the demand planning process is a forecast by SKU, for a group of customers, for a weekly bucket by region. In order to transform this forecast into more detailed demand plan which can be used for site integrated planning, split tables are maintained which maintain fixed percentages for materials and the manufacturing plants. For example, if two plants, A and B, both make a certain SKU of Tide, the table maintains a data record for Tide indicating that 30% will be manufactured in plant A and 70% will be manufactured in plant B. These tables are also maintained for the days of the week, keeping fixed percentages for each day, based on material.

The demand planning organization essentially presents a volume forecast to a particular product, split according to groups of customers. However, it is the product supply organization, which lies within the GBUs, that takes the forecast and translates it into a plan detailing exactly when and what each plant will make. This group also maintains the inventory levels and is responsible for procuring raw materials. Thus the impacts of the good or bad forecast are typically not felt by the planning organization, but by another branch of P&G.

5 Measuring Forecast Accuracy

It is clear that demand planning plays an important, yet complicated role in the organization. Thus, measuring forecast accuracy is critical not only for analyzing the performance of the forecast process, but also for determining how to properly design the rest of the organizational processes to deal with this uncertainty. For example, inventory levels and customer service are directly impacted by the forecast error. This section will introduce the main areas of forecast accuracy used at P&G.

5.1 Forecast Error Index

A simple index is commonly used to review forecast error at P&G. The index is simply a ratio of the actual sales, as measured by shipment volume to the sales forecast for that period. This is shown as follows:

$$\text{Forecast Error Index: Actual Sales/Forecast} * 100$$

This simple ratio is represented as a percent, so that a value of 100 indicates that the sales matched the forecast, and a number over 100, indicates that the sales exceeded the expectations. Whereas a number below 100 indicates that sales did not meet the forecast. This not only gives an indication of the size of the error (the difference from 100), but also a visible measure of the bias, which is the tendency to over, or under-forecast demand.

5.2 Bias

Silver, Pyke and Peterson indicate that the term bias is used to indicate that on average a forecast is substantially above or below actual demand.[19] Bias is commonly indicated by evaluating a period of forecast error indices to determine if over an extended period of time, the errors lie heavier on one side of 100. It can be also evaluated using probabilities, to be explained later. Common perception is that bias can be removed by making a model correction, or by management intervention.

5.3 WAPE: Weighted-Absolute Percent Error

The global standard definition that P&G uses for forecast error is WAPE, weighted-absolute percent error.[20] P&G defines a process for calculating this factor, aggregating it up through the

---

[19] Silver, Edward A., David F Pyke, and Rein Peterson. Inventory Management and Production Planning and Scheduling. 3rd Ed., John Wiley & Sons, 1998.
[20] Kremblewski, Mark. "Supply Chain Forecast Error: Global Standard Definition" P&G Internal Documentation. March 2005

organization and handling exception cases. This is detailed below based on training provided by Mark Kremblewski, Global Demand Planning Business Expert.

Supply Chain Error is based on a weekly measure of Absolute Percent Error (APE) by Demand Forecasting Unit (DFU) by Market:

APE = Absolute (Shipment-Forecast)/Forecast x 100%

When determining the forecast for a period of time, the appropriate 'lead-time' or 'time fence' between when the forecast was created and when the shipment occurred is left to the discretion of the demand planner but should in general reflect the 'supply chain cumulative lead time' for that DFU. For simplicity, entire brands, categories or even countries may use the same average lead-time.

Results are reported in a monthly aggregate. Aggregation is volume weighted based on shipping volume. DFU APE's are volume weighed, based on shipped, not forecast volume, horizontally across the month: this is the WAPE for a DFU for that month DFU. Additionally, APE's are volume weighed (based on shipped volume) vertically to the Brand: this is the WAPE for a Brand for that month. All further aggregations (to country, category, GBU, MDO, Global etc.) are all done on a shipped volume weighted basis.

$$WAPE = \frac{\left[\frac{1}{n} \sum_{t=1}^{n} shipment_t \times \frac{\left| shipment_t - forecast_{t-1} \right|}{forecast_{t-1}} \right]}{\sum_{t=1}^{n} shipment_t} \times 100$$

Standards are applied to other circumstances to ensure uniformity across P&G. For a DFU/week with a positive forecast but 0 (zero) shipments, APE should indicate 100%. This will not affect the Brand total WAPE since the volume weight of that data point is null (zero). For a DFU/week with a positive Forecast but <0 (negative) shipments, the APE will show a real calculated value. The weighting calculation will use the absolute value of the negative shipment. This may occur if product returns in a given period exceed shipments. Thus the error will affect the WAPE result in the correct direction. (If the negative were to be used in the weighing, this DFU/Week data point would serve to actually reduce the overall WAPE result). For a DFU/week with a 0 (zero) forecast and positive shipments, the APE calculation fails due to the divide by zero term, therefore these instances need to be excluded from the calculation. As these data points are excluded completely from aggregate calculations, removing them also essentially detracts from the actual volume shipped term of the equation. If the actual shipment is zero, even if there is a forecast, the data point will essentially be excluded as the weighting applied will be zero. This is summarized in the following table:

Forecast	Actual shipment	Data point shows	Weighted Roll Up
Real	Real	XX%	Included
Real	0	100%	Not included due to 0 weighting
Real	< 0	XX%	Absolute value of the Actual is used

			for weighting (increases WAPE)
0	Real, 0, <0	Div/0 or NA	Excluded

Additionally, in some cases if the forecast is close to zero, but a large shipment occurs, the error may become huge. Therefore P&G specifies a cap of 1000% for individual APE data points. This occurs when a very large shipment occurs (as in the shipment of a remnant) for a very small forecast. In severe situations, these data points can make an entire country or region's WAPE result meaningless.

5.4 MAPE, Mean-Absolute Percent Error

The mean absolute percent error, MAPE follows identical rules to WAPE, however, it is not weighted based on the shipment volume, and is instead averaged across the data set, giving equal weight to each data point. Silver, Pyke and Peterson present MAPE as another intuitive measure of forecast error.[21]   Essentially, it measures the absolute percent error for each of the data points, and takes an un-weighted average across the data points, and then represents it as a percent error as shown below:

$$MAPE = \left[\frac{1}{n} \sum_{t=1}^{n} \frac{\left| shipment_t - forecast_{t-1} \right|}{forecast_{t-1}} \right] \times 100$$

It is important to note that MAPE is not appropriate if demand values are very low. For example, a forecast of one unit of demand matched with an actual value of two units shows an error of 100 percent.

5.5 Probability

Another measure that has been commonly adopted at P&G is a probability. Managers easily understand a percent of how likely a behavior is to happen. Therefore probability is frequently used to evaluate bias, and indicate how likely an organization is to have a forecast that is over 100% versus a forecast that is under 100% - when measured by the forecast error index. The probability that a forecast will be under 100 is determined by evaluating a normally distributed curve with a mean and standard deviation determined from the forecast error index data. The probability that a forecast will be over 100 can be determined by determining the difference from one. P&G strives for a forecast which is 50% likely to be over and under 100.

---

[21] Silver, Edward A., David F Pyke, and Rein Peterson. Inventory Management and Production Planning and Scheduling. 3rd Ed., John Wiley & Sons, 1998.

6 Enabling Accountability for Forecast Accuracy

Previous chapters demonstrated the importance of forecast accuracy, and also discussed the general organizational structure at P&G. While there are a number of resources dedicated to demand planning, this organization is not functionally close with the supply organization which uses the forecast generated from demand planning. While the supply process and manufacturing resides in the GBUs (Global Business Units), demand planning resides with the MDOs (Market Development Organizations). So while many functions may point to the demand planning process as contributing to issues, they do not have responsibility over the department. This becomes quite interesting when the MDO is responsible for budgeting the resources dedicated to forecasting, while it is the GBUs that are most impacted by the quality of the forecasts generated.

This chapter details a process P&G employed to create accountability for forecast accuracy across the organization. Using a top-down metric driven approach, targets have been set for the company as a whole, and a process has been developed to disaggregate a metric into appropriate targets for sub-components, allowing managers to further demand accountability from within their organizations.

6.1 Motivation for Statistical Accountability Process

Forecast accuracy is critically important to P&G, as inaccuracy can have a direct impact on the stock price. P&G publishes quarterly statements, which indicate the sales revenue (in dollars) for the quarter. If these revenues exceed expectations, the stock price typically rises. If these revenues fall short of expectations, the stock price falls. Therefore the ability to accurately predict and manage future sales is critically important to the evaluation of the company.

When talking about the forecast generated for P&G as a whole, typically this is discussed in terms of dollars sold, instead of units sold. However, predicting quarterly profit is difficult as the contributing factors fluctuate with time. These factors include the mix of products sold, the cost to manufacture, distribute and sell these products, and the price achieved for each of these products. One of the challenges discussed is whether businesses generate the volume forecast of the units that will be sold, and then use this to drive the financial forecast (dollars of revenue), or whether businesses set the revenue (dollar) target, and then back calculate the volume forecast (units to sell).

In a move to create more accountability, accuracy, and clarity in the system, P&G has recently moved to use the S&OP Process (Sales and Operations Planning) to have all of the relevant parties represented at a routine meeting to generate the appropriate volume forecasts for businesses. This S&OP process should generate the volume forecast, which should be the basis for the financial forecast. As Finance has a seat in the S&OP process, their opinions should be represented at this earlier point in the process, and the most realistic and accurate forecast should be generated.

With this new emphasis on generating accurate forecasts through the S&OP Process, a project was created to generate corporate accountability for accurate forecasts. Historically, P&G has set a certain tolerance range for each of the business regions and units, and the same range was used for each entity, regardless of the size or complexity or the business region or unit. With this seemingly arbitrary metric setting policy, these metrics were frequently missed, and managers did not buy into their importance. The main objective was to set a global metric for forecast accuracy for P&G as a whole entity, then develop a process for disaggregating the metric as aligned with business units.

6.2 Setting a Global Metric

Historically, P&G had measured its accuracy as a ratio of the shipments over the forecast, expressed as a percentage. Thus if the percent was over 100, then there were more shipments than forecast. A single percentage point deviation was associated with millions of dollars of sales either over or under forecast. To encourage high forecast accuracy, P&G had historically operated by setting a global tolerance where the shipment/forecast ratio needed to be within a set tolerance. However, their ability to deliver within this tolerance was questionable, and the first area to be discussed.

6.2.1 Historical evaluation of P&G performance

The basis of this proposal for creating accountability for forecast accuracy hinged on selecting a global metric which was both realistic, yet a stretch target. Senior leadership wanted a tight tolerance range, which would demand better forecasts. However, I worked with P&G's Global Analytics departments to investigate what would be an appropriate target. Oscar Rosen, from the Global Analytics department had access to the historical forecast and sales data. We presented the following graph to senior leadership:
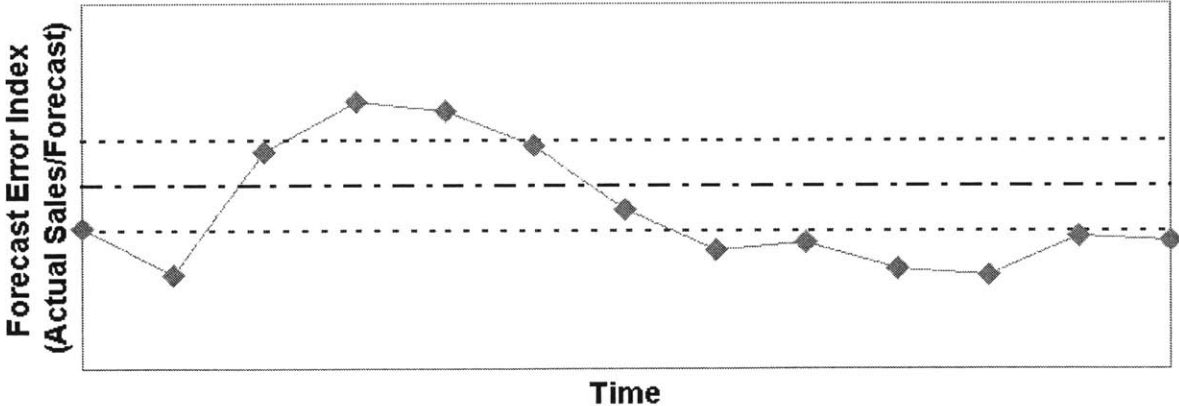
Figure 6.1: Forecast Error Index (Actual Sales/Forecast) for P&G as a whole

Figure 6.1 demonstrates the tolerance ranges that had been in place and the historical company performance against these metrics. Each data point represents a three month aggregate, though these are re-measured each month. In the year of data displayed, it is evident that only four out of thirteen months did the sales fall within the set tolerance range of the forecast.

The executives driving the project were concerned about accurate reporting to Wall Street and the financial impact of forecast error. Thus, an important consideration for the global tolerance was the financial impact from different ranges. The question was posed whether the finance organization would prefer a tight tolerance, with a low chance of being met, or a wider tolerance with a higher chance of being met, even if this meant higher forecast error. As financial reserves can be maintained to balance certain swings, the financial organization indicated that there was a definite cost associated with higher error, but better certainty of the error range could enable more efficient planning. To enable better decision making, a probability estimate was warranted.

6.2.2 Using historical data to predict future performance

In choosing an appropriate metric, there appeared to be a pretty clear trade off between setting a tighter tolerance, and the likelihood that P&G would be able to deliver within the tolerance range set. To enable decision making and provide clarity to this tradeoff, a numerical assessment of the likelihood that P&G could perform within certain tolerance ranges was conducted. To perform this analysis, an assumption was made that the forecast error would follow a normal distribution, as recommended by Silver, Pyke and Peterson.[22] This assumes that bias is not present. However, since P&G strives for a balanced forecast, management directives are in place to remove bias from the forecasts. Thus, bias should be limited in P&G forecasts.

Probability (within tolerance range) = NORMDIST (upper limit, mean, standard deviation)
  − NORMDIST (lower limit, mean, standard deviation)

Using this relationship, different tolerance ranges were tested to determine the likelihood that P&G's forecast error would fall within the range specified if P&G repeated historical performance. To protect the confidentiality of data, the four data tolerances were replaced with the phrases, "very low," "low," "high," and "very high." The following table was generated to show the likelihood that P&G could meet each of the tolerance ranges:

Global P&G Corporate Tolerance (+/-)	Very Low	Low	High	Very High
Probability of meeting tolerance with today's process and no improvements	26%	50%	69%	83%
Probability of meeting tolerance if all regions remove all forecast bias	29%	54%	73%	86%

Figure 6.2: Probability of Meeting New Global Tolerance Limits

---

[22] Silver, Edward A., David F Pyke, and Rein Peterson. Inventory Management and Production Planning and Scheduling. 3rd Ed., John Wiley & Sons, 1998.

Figure 6.2 shows that the current process of setting a tolerance range based on the financial organizations need for accuracy, results in a high probability that the tolerance will be missed as the rest of the P&G organization is not capable of delivering forecasts within this tolerance range on a repeatable basis. With this data in hand, the management team was able to weigh the different factors impacting the decision to choose what the global tolerance range should be. Factors included were the motivational impact of choosing metrics which supported continuous improvement, but didn't discourage organizations by seeming unachievable. Additionally, the organization was about to start a massive effort to integrate the Gillette Corporation, one of the larger acquisitions P&G has undertaken, so caution was exercised in choosing a metric that would create buy-in and motivate people to achieve. By presenting the data in Figure 6.2, and discussing the factors above, we enabled senior leadership to choose a global tolerance range which they felt would be best for P&G as a company. By choosing this range for which the company would be expected to perform at a global level, top-down direction would follow to ensure that the business would be the basis in further statistical analysis.

6.3 Choosing Appropriate Factors for Statistical Process (Size, Complexity, Capability)

Having chosen a global tolerance range, the next step was to choose appropriate factors to enable the statistical process of taking a global tolerance range, and enabling it to be converted into tolerance ranges for the different business regions and units. As recognized by the company's senior leadership, in order to have a meaningful metric, people had to buy-into its relevance and feel accountable. Therefore, the process for allocating the metric needed to be relatively straightforward, transparent, and clear enough to create buy-in.

6.3.1 Availability of data

One key decision concerned the data which would be used for the process. P&G has multiple methods for tracking and reporting forecast error. Discussion with involved parties indicated that the forecast error index was the desired measure as it clearly indicated bias. Another objective of the S&OP Intervention effort was to drive bias out of the system, therefore the generation of a new metric system needed to clearly demonstrate bias. Additionally, the index had been used historically, so was familiar both to demand planners, and the organization which used the demand plan or was responsible for its accuracy. To create the index, simple forecast and shipment data obtained from the company information system, and aggregated over a rolling-three month basis was deemed straightforward, and the organization readily accepted this as the fundamental source of forecast accuracy information.

6.3.2 Popular opinion

Having determined that the model would be based on simple forecast and actual sales data, the next question was how to use this data. Discussions with various demand planning process owners, experienced planners, and managers accountable for the forecast accuracy indicated that there was a fairly standard opinion regarding what impacted forecast accuracy. If the process appealed to people's intuition, it would be easier to generate buy-in for the process. Therefore,

popular opinion was considered heavily when selecting factors to use in the statistics. For example, some regions are experiencing rapid growth and popular opinion was that it is easier to forecast a static market, compared with a rapidly changing dynamic market. Additionally, P&G breaks the world into different segments, with one of these segments including over one hundred different countries stretching from Africa, across the Middle East, Eastern Europe and even the countries formed from old Soviet Republics. The cultural and geographical diversity across this region, along with the communication challenges was commonly believed to make it harder to forecast. Essentially, popular opinion seemed to indicate that some segments were intrinsically harder to forecast than other segments, just by the nature of the segment.

6.3.3 Statistical relationships

To select factors, popular opinion indicated that some measure of the intrinsic difficulty to forecast a segment would be important. To also appeal for the need for simplicity, the variability within sales was tested to see how strongly it correlated with forecast error. Oscar Rosen, the Global Analytical statistical expert who devised the process chose to represent variability using a "volatility factor". This fact was the coefficient of variation, or the standard deviation divided by the average.[23]

$$\text{Volatility factor, } v_i = \frac{\sigma_i}{avg\ sales}$$

While using this volatility factor to represent the intrinsic difficulty in forecasting a segment was simple, there were many other ideas about other factors other than volatility which would impact forecast error. Thus we felt it was important to statistically justify that a strong relationship existed between this volatility factor and forecast error.
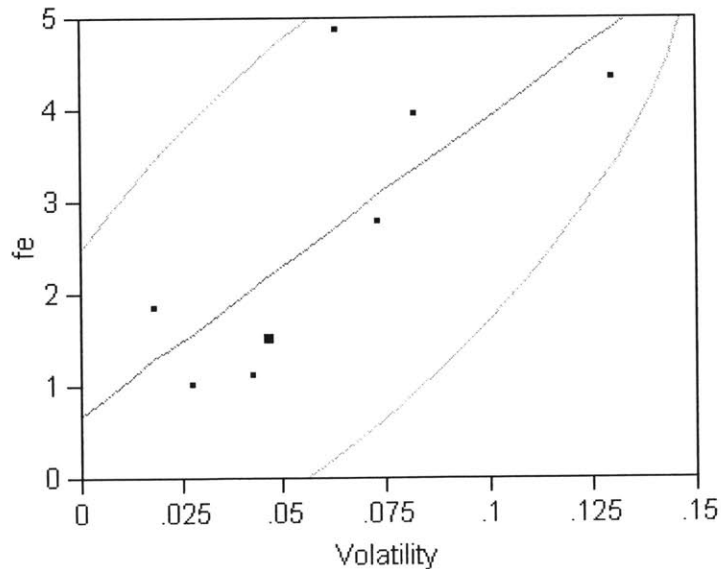
Figure 6.3: Tested relationship between shipment volatility and forecast error (fe)[21]

---

[23]Rosen, Oscar. "Forecast Error Allocation and Tracking" P&G Internal Documentation. 6/13/2005.

Figure 6.3 indicates a relatively good relationship between volatility and forecast error. Further iterations were performed, and eventually it was determined that the forecast error was correlated most strongly with the square root of the volatility. Thus this volatility factor was used in the subsequent statistical model.

Additionally, one point that came across from discussions was that the regions that contributed larger portions to total P&G sales, would have a larger impact on the overall forecast error. For example, 5% forecast error in North America, which accounted for almost half of P&G's sales volume, would have a much larger impact on the total company accuracy, compared with a 5% forecast error in Australia or India. Thus another component of the statistical process was chosen to be the total sales volume of the segment.

P&G's Global Analytical department was able to use these two factors, and standard statistical theory around the aggregation of data to estimate correlations and break out the global tolerance into regional targets. To enable buy-in, the probability analysis discussed in 6.2.2 Evaluating probabilities and the likelihood to perform in the future at certain levels was repeated for the segments. If our assumptions were valid, we would expect historical performance to indicate that the probabilities of meeting these tolerances would be relatively equivalent across segments. However, it was surprising to note that the probabilities were quite different. For example, it was noted that Latin America had an exceptionally high probability of meeting its target allowance (greater than 99%), whereas other regions of the world were less than 60% likely to perform within their given tolerance. This could create questioning of the assumptions, and questioning whether the metrics were fair, and whether they would be accepted and incorporated into the organizations.

6.3.4 Fairness but flexibility

Discussions with the Demand Planning team of experts indicated that the use of simple sales volatility and volume was not enough to predict how well a region would perform at forecasting. There are other factors leading to intrinsic difficulty in forecasting a region, one in particular being the people in the organization. The average lifespan of a demand planner is 3 years in the role, so that 1/3 of the organization is new in the role each year. Thus the capability and experience of the demand planner can impact accuracy. Therefore room was left for another factor, the current forecast error, to represent the capability of a region and account for influences not accounted for by the other factors. Thus the three factors to be used would be size, as approximated by volume, difficulty as approximated by volatility, and capability, as approximated by forecast error.

6.4 Technical Decisions

The forecast and shipment data was used to determine the factors outlined; however, on running the statistical model, certain questions arose. Although this was the first time the statistical process would be used to take the global metric and split it out for sub-units, the desire was to create a robust and repeatable process that would not require significant rework in the future. This required evaluating and making appropriate decisions related to certain details.

Additionally, the model was built with certain flexibility and expandability so that it could be easily adjusted in future years as the business continued to change, and other requirements were added.

6.4.1 Scope

When applying a statistical process to take a global metric and translate it into appropriate metrics for smaller categories, one question arose as to whether all of the components of the total had to be included. For example, P&G divides the company into seven regional Market Development Organizations (MDOs) based on geography. However, when allocating the company sales volume, not every sale falls into one of these seven MDOs. For example, when a new company is acquired, or when a business is reorganized, sales might temporarily fall into categories such as "Global Organization" or "Corporate Adjustments". These two extraneous categories do not fall in the mainstream corporate hierarchy, and more importantly, do not have business authorities presiding over them. Within P&G, each of the regional MDOs has a President, who is accountable for the regional business. However, accountability is harder to trace for these non-regional groups. In the case of the global P&G data, these groups together accounted for a small fraction of the volume. However, upon review with the P&G analytical department, the decision was made that all components of the total must be included in the statistical process. For example, if these groups were excluded from the analysis, but had a large error associated with them, it might be possible for the components which were included in the analysis to meet their targets, but the overall organization to miss due to the excluded segments. Therefore, the process was defined that all components for a total must be included when splitting out the data.

6.4.2 Duration of history

Another debate ensued surrounding the amount of historical data to use in generating the metrics. The company stores three years of statistical forecast accuracy history. There are many factors which would indicate that a long history is better for the statistical process. First of all, longer time horizon would more accurately show the different environments in which the business performed. For a consumer products company like P&G, business can be significantly impacted by the overall health of the economy, something that moves and shifts relatively slowly. Additionally, certain events such as a hurricane, tsunami or terrorist event can have an impact on forecast accuracy. Thus the longer time horizon followed, the more these rare events are diluted in a broader set of data. However, arguments also existed for maintaining a shorter time horizon. The consumer products business changes quickly with multiple company acquisitions and ever changing competition. Thus, the most accurate snapshot of the marketplace is a current one. Additionally, many of the regions P&G services are undergoing rapid changes, and using older data may not represent the capability moving forward. Additionally, the turnover of demand planners leads us to believe that a shorter time horizon more accurately represents regional planning capability. Balancing these factors, and synchronizing with the fiscal year led to the selection of using twelve months of three month rolling data.

6.4.3 Update frequency

The frequency at which the metric should be updated was also questioned. For example, should the integration of a new business trigger a re-run of the process to generate new metrics? Should the target be allowed to move based on trending business conditions? Or, once the tolerances are generated, should they be held constant indefinitely so that people grow accustomed to them and don't believe they are trying to meet a moving target? Discussion with business leaders and the statistical process owners led to the decision to update the metric every fiscal year. This is fairly consistent with many metric setting and evaluations processes and is also aligned with the length of history used to generate the metric.

6.4.4 Leveling Process

The last factor discussed in 6.3.4 Fairness but flexibility also presented difficulties in incorporating this factor into the statistics, and in allocating appropriate weight it. Essentially the statistical process used the shipment and forecast history to develop two key factors – shipment volume and volatility – and from these roll out the target for the aggregate into the sections based on the shipment volume and volatility in each section. After this was completed the probability of each section meeting the assigned tolerance was determined based on a year of historical forecast error. A leveling process was then applied to account for this forecast error, and essentially re-allocates tolerance range percentage points from the section most likely to meet the tolerance to the section least likely. An algorithm was determined to essentially level the playing field by giving each section an equal probability of meeting the tolerance range. This procedure had a setting allowing it to level to only a certain percentage of what the total difference in tolerance ranges would be between the range assigned based on volatility and volume only, versus the range accounting for forecast error. For example, the first step of the process resulted in Region A having a 70% probability of meeting a tolerance of 3.0%, and Region B had a 99% probability of meeting a tolerance of 1.5% the tolerance range for Region A would be relaxed and the range for Region B would be tightened. Supposed both regions reached an equal tolerance of 88% if the range for Region A was increased to 4.0%, and the range for Region B was tightened to 1.0%. Instead of using the ranges that ensured an equal probability for all regions, scaling was applied. Thus the process would result in Region A having a range between 3 and 4% and Region B having a range between 1.0 and 1.5%. Discussions with management ensued as to whether each region should be given a metric that they were equally likely to meet. Another option would be to leave some margin for difference in capability. This would reward regions who had dedicated significant effort in the past to getting to the high level of capability they had today. Regions with a more difficult metric might be motivated to invest more effort into improving the forecast accuracy. For example, in the case of Latin America, significant resources and effort had been dedicated to improving forecast accuracy, and a substantial improvement had been made in recent history. The decision was made to set the leveling to 70% of the total change that would have occurred to put the regions on equal probability. This was to account for differential effort between regions.

6.4.5 Correlations

A slightly more complicated decision involved the statistical process treatment of correlations between the data. Essentially, statistical routines exist to account for the fact that all of the

components may not behave completely randomly, but that there may be a link in how they move. With forecast error between regions, if there is a general worldwide economic slump, then forecast errors will probably increase around the world. Thus world issues will impact all regions, thereby correlating the data. On the other hand, most of the contributors to error are smaller and associated with promotions and events, and these are typically random across regions. Even seasonality can be diluted across regions. From that standpoint, for a product consumed more heavily in the summer, there may be negative correlation between northern hemisphere and southern hemisphere regions that experience summer during different months of the year.

6.4.5.1 What Do They Mean: Positive, Negative, None

Positive correlations imply data is not completely random, some trends occur across regions in the same direction at the same time. If positive correlations are found, the statistical process to account for this will tighten the tolerance ranges necessary to deliver the main tolerance as it is likely that error would move in the same direction across groups at a given time, and not benefit from aggregation and leveling of variability by combining units. Negative correlations imply data for different groups will trend in the opposite direction. If this exists, the statistical process to account for this will loosen the tolerance ranges necessary to deliver the set tolerance as it is likely that error would move in the opposite direction across groups at a given time. No correlation implies data is completely random.

6.4.5.2 Implications of Correlation Assumptions

A decision had to be made on how to account for correlations in the data. Since only a year of data is used, it is possible that the algorithm to determine if correlations exist might account for a phenomenon which was purely random. Many discussions were held between the global demand planning experts to assess their opinion of whether or not correlations exist in the real world between groups of products or forecast errors. Essentially, if we chose to allow positive correlations, the tolerance ranges might be set tighter than necessary, challenging the business more than required, but leaving a higher likelihood that if the segments meet their tolerances, that the aggregated tolerance would be met as well. On the other hand, if we allow negative correlations, and the data indicates this type of correlation exists where it was only an anomaly, the tolerance ranges for the subgroups might be set too broadly, allowing for each of the subgroups to make their tolerance goal, but the overarching organization to miss.

6.4.5.3 What Did We Choose and Why?

After reflection, the conservative approach was chosen to allow for positive correlations. The model was built flexible so that this approach could be modified for future years. The choice to use positive correlations will result in more conservative (narrow) tolerances, although it risks being too conservative due to correlations from chance alone. The model does not give credit if the overall covariance is negative because it means that credit is being given for groups offsetting each other, and we did not want to count on that behavior being repeatable.

6.4.6 Standard Deviation

Options for different calculation methods for standard deviation were presented and debated.

6.4.6.1 Options

"Short-Term" Standard Deviation: Our statistical expert found reference to a non-standard method for calculating the standard deviation term in the probability formula which is argued to be better at predicting future behavior based on history. It is used in the JMP statistical software tool for capability studies and is computed using the method for control charting and is based on the moving range.

$$\sigma = \frac{\sum_{i=1}^{n-1}|x_{i+1} - x_i|}{n-1} \; [24]$$

Additionally, the conventional standard deviation algorithm could be used.

$$\sigma = \sqrt{\sum_{i=1}^{n}\frac{(x_i - \bar{x})^2}{n-1}}$$

In the above equations, $\bar{x}$ represents the average of the data, x_i represents an individual data point, n represents the total number of data points.

6.4.6.2 Evaluation

The different options for standard deviation provided different probabilities of meeting a given tolerance. These differences hinged on a behavior called "chasing." Chasing is a phenomenon where data points trend in the same direction for a period of time. The following figure demonstrates considerable chasing:

---

[24] JMP Stat Graph Guide. Version 5.1
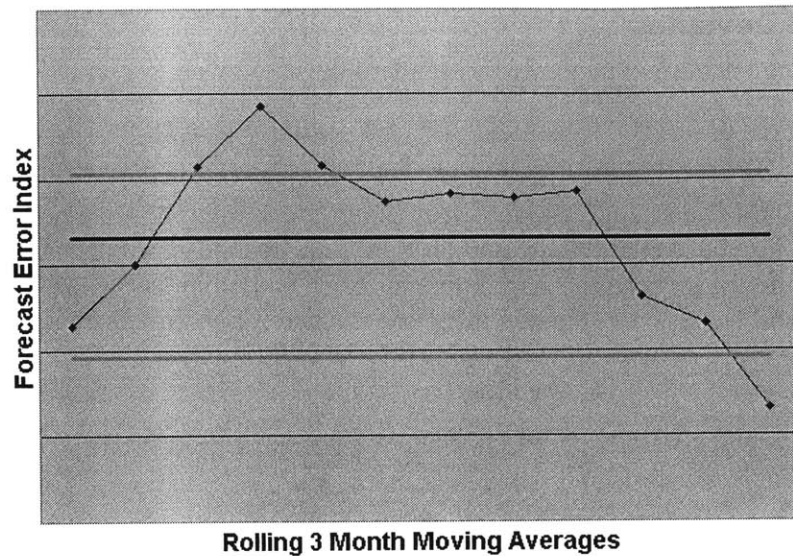
Figure 6.4 "Chasing" phenomenon where data points tend to trend in a single direction

For units with chasing, as shown in Figure 6.4, the probability of meeting the tolerance is much higher using the 'short-term' standard deviation as it assumes improvements will occur and the data will be more randomly up and down. For the data displayed in Figure 6.4, the conventional standard deviation indicated a 71% probability that the future data points would fall within range, whereas the predictive standard deviation indicated a 96% probability that future data points would fall within range. On the other hand, some regions had data which tended to swing back and forth without consistently moving in one direction as shown in the following figure:
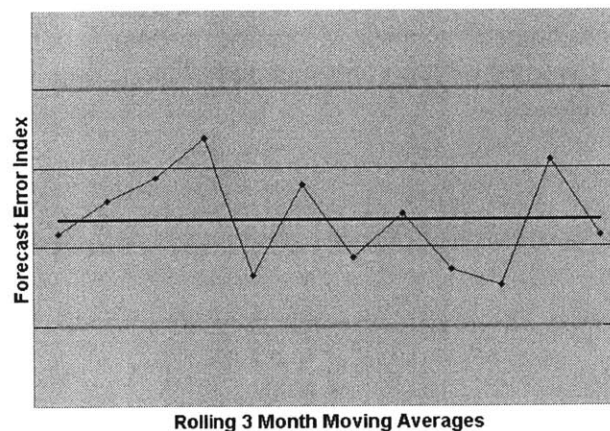
Figure 6.5 No "chasing" phenomenon where data points tend to fluctuate both up and down.

For figure 6.5 the conventional standard deviation indicates a 99% probability that future data would fall within the specification range; whereas the predictive standard deviation indicates a 96% probability that data would be within specification. Reviewing these behaviors and the data, we chose to use conventional standard deviation, as it appeared to give a more appropriate relationship for our rolling three month data – where between two successive data points, two out of the three months included in each of these data points are the same.

6.5 Implementation

The previous sections detailed the development of the statistical process for taking a broad tolerance, and determining what the appropriate tolerances would be for the subcomponents of the total. However, the new accountability process had to be implemented to be successful.

6.5.1 Implementation Strategy

During development and testing of the statistical model, the global process experts for demand planning, along with the managers responsible for accuracy were frequently consulted for their advice on process development. Their input and suggestion also helped generate feelings of commitment, ownership and buy-in from this respected and well-experienced group of individuals. Senior management would also need to be sold on the idea and to drive the top-down metrics. Finally, the managers accountable for the new metrics had to buy into their importance, and understand a tool developed to allow managers to take their tolerance range and break it out for their own sub-units. This is demonstrated in the following figure:

Global Target	x%

2.0%
Total NA MDO

USA/PR	2.0%

Canada	4.0%

3.0%	2.0%
US HABA	US HHN

4.0%	5.0%
CN HABA	CN HHN

Fem	6.0%
PHC	10.0%
Oral	8.0%
Hair Care	5.0%
Hair Color	9.0%
Prof Hair	
Cosmetics	9.0%
PCC	8.0%
DOOS	5.0%
Skin	9.0%
Pharma	8.0%

11

Fabric	4.0%
Home	6.0%
Family	4.0%
Baby	5.0%
Coffee	8.0%
Snacks	8.0%
IAMS	4.0%
CPG	4.0%

8

Deployed by Global

GM Level

Fcst Team Level
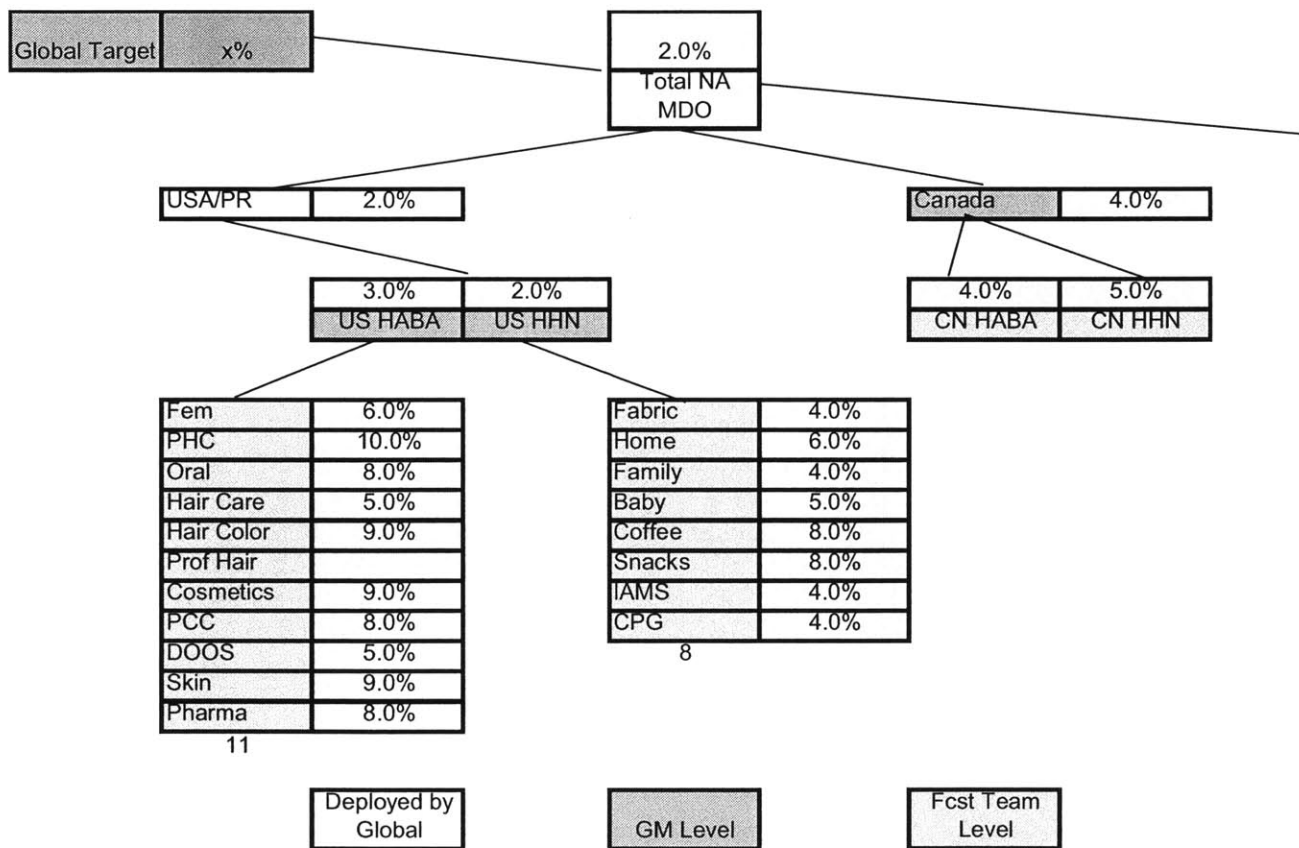
Figure 6.6 Tolerance deployment process (data was adjusted to protect confidentiality)[25]

---

[25] Yuhas, Jim. Internal P&G Documentation. North American Demand Planning. Modified for confidentiality.

Figure 6.6 demonstrates how a regional manager uses the process to first take the tolerance from upper management and deploy it throughout the organization. This manager first splits North America into the US and Canada, then splits the tolerance by business unit, and then into business segment. The figure also depicts the general trend of tolerance ranges increasing the deeper the level of the organization. This is due to the disaggregating and smaller volumes at smaller levels. Essentially, as long as the tolerances at these small levels are met, the tighter tolerances at the higher levels are met as errors balance out when combined.

6.5.2 Issues and Execution

The plan for implementation was to ensure that representatives who carried weight within the organizations to be impacted were part of the process at each step. The statistical work was conducted by an expert from the Global Analytical department. Demand planning input was obtained by frequent reviews with the Global Demand Planning process owner, and an expert with over twenty years of demand planning experience. Additionally, the person from the sales organization responsible for reporting on forecast accuracy for the company was included in each of the team meetings. Throughout the implementation process, these people were included in the team reviews and consulted for their opinions. This ensured buy-in from the organizations in general. However, the process needed support both up and down the organization.

The team met with corporate executives representing global Finance, Consumer Market Knowledge, Customer Service and Logistics, and Sales organizations. This was the set of executives who had mandated the intervention to create accountability for forecast accuracy. The group reconvened and the team presented an overview of the statistical process and received management approval. Additionally, as this was a top-down process, the set of executives deployed the global tolerance discussed in 6.2 Setting a Global Metric and accepted the output of the statistical process to deploy forecast accuracy tolerances to the regional business presidents.

Meanwhile, the regional business presidents knew that new metrics were going to be deployed soon, and were asking their demand planning process owners for any sort of early information about tolerances or capability. This created a bit of excitement throughout the organization, and we used this to start obtaining buy-in from the organization to test the tool throughout the development process. For example, the North American Demand Planning Manager found several issues with the statistical model which we were able to improve upon before rolling it out. Also, obtaining his buy-in early in the process smoothed the rollout to the rest of the organizations worldwide. To deploy it to the rest of the world, a teleconference was held with regions around the world to discuss the new metrics, the statistical process, to deploy the tolerances and give them a tool to be able to sub-divide them throughout their organization. For example, Western Europe first subdivided into countries, then into businesses within the country.

Although behind schedule, the process was deployed within the second month of the new fiscal year and was met with acceptance. The number of questions and feedback generated is a good indication that it is capturing attention throughout the organization.

6.6 Summary and Recommendations

In review, this project was successful in its set goal of creating accountability for forecast accuracy throughout the organization. The previous policy of setting arbitrary and identical targets for accuracy across all levels, regardless of size or complexity, had not been supported, and the metrics were generally ignored. The statistical process was developed with significant input from the representative parties, and implemented in an approach which generated internal buy-in to the new metrics. The statistical analysis and attention at a high level to accuracy produced not only higher accountability, but highlighted areas of opportunity for the organization.

6.6.1 Bias – Using the Process to Emphasizing its Effect

One area for improvement clearly visible from the study surrounded the area of forecast bias. The term bias is used to describe the tendency for errors to occur in one direction – either consistently over- or under- forecasting demand. Where the statistical analysis indicated a strong history of bias, management was questioned to determine what policies were in place encouraging this bias. For example, numerous publications reference how earnings estimates are optimistically biased[26]. At the executive level, the drive for a 50/50 forecast began to be pushed throughout the organization. This indicated that the probability of forecasting under/over should be 50/50 indicating that it is a true reflection of believed market behavior, and not slanted due to internal policies. Management felt strongly that capability and tolerances should be set assuming that businesses can operate without bias. For example, in the leveling process which adjusts the tolerances for the capability of regions, bias is estimated and removed from the historical data. Essentially, the average index over the history is determined, and each data point is adjusted up or down as necessary to set the average index to 100, indicating no bias is present. With this adjusted data, a new capability is determined, which is used to adjust the final tolerances. For example, some groups had a history of poor accuracy, driven by poor bias, but were fairly consistent with their errors. While historical data exhibits poor capability, once adjusting the bias out, the capability would look much better and the group would be given a tighter specification assuming they could remove the management bias immediately. The following figure shows an example of this:

---

[26] Han, Bong H, David Manry, Wayne Shaw. "Improving the Precision of Analysts' Earnings Forecasts by Adjusting for Predictable Bias" Review of Quantitative Finance and Accounting, 17: 81–98, 2001.
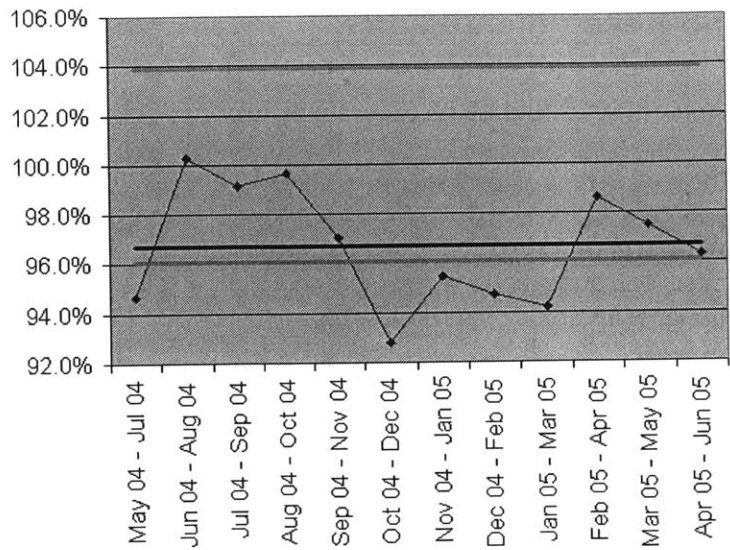
Figure 6.7 Historical Accuracy plotted against tolerance limits

Figure 6.7 shows the impact of bias. The last year's worth of data for this business indicates that only once did a data point exceed 100% (having more shipments then forecast). Essentially, this business consistently forecasts sales to be higher than they actually prove to be. The solid lines at the outer range indicate the tolerance limits set by the process. According to these limits, five of the historical data points would fall outside of the limit. This is due to bias, shown by the black line at 96.7% indicates the mean of the data. However, if bias was removed and all of the data points were shifted upwards adjusting the mean to 100% as shown below:
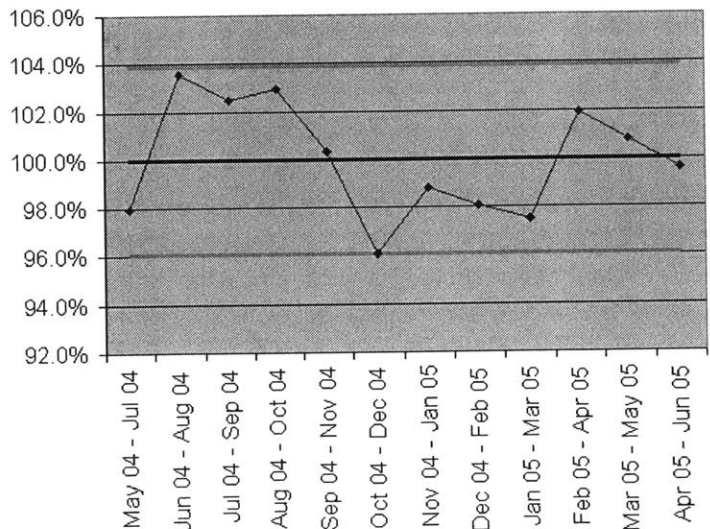
Figure 6.8 Historical Accuracy adjusted to remove bias, against tolerance limits

Figure 6.8 shows the impact of removing bias. Now, with all of the data points adjusted, we can see that each of the historical data points would fall within the tolerance range. The impact that

this powerful graphic has conveys the importance of the management directive to attain unbiased forecasts.

6.6.2 Organizational impact

Overall, accountability for forecast accuracy and sponsorship for improvement is difficult in P&G's organizational design. The Global Business Units essentially run the branded business for P&G, and include marketing, finance, and the product supply organizations. The product supply organization controls the cost of manufacturing, making sure the right amount is available at the specified time. However, product supply is not responsible for generating the demand forecast. The demand planners are part of the Market Development Organization. The MDO includes the sales, customer service and logistics organizations across all the brands for a region. Thus, the organization that controls the process for determining the sales forecast does not actually use the forecast. Therefore, the MDO may decide to reorganize, reduce headcount, or make other changes which may impact forecast accuracy. The main impact would be felt in the GBU, which might be asked for funding to support additional headcount. While processes are in place to smooth the collaboration between the units, the S&OP Intervention is taking a broader step, by setting metrics for forecast accuracy in the scorecard of managers in the GBU. Essentially they are now responsible for meeting targets which are set by people not in their organization. This is yet another example of how work at P&G is conducted through relationships as much as it is accomplished through formal structures and authority.

7 Improving Forecast Accuracy

This thesis has discussed the importance of forecast accuracy and the impact poor forecasts can have on many different levels of the organization. While the global process organization does an excellent job of evaluating and improving the demand planning process, real funding for improvement needs to be justified by the organizations affected by the forecast improvement. Thus, establishing metrics and accountability for forecast accuracy lends transparency to the expectations and the results, and also aids in justifying further improvement efforts. The effects of forecast accuracy are so widespread, that the impact they have on the customer service organization can be enough to justify investment into improvements. This chapter will evaluate the trial of a new system for customer accounts where P&G manages the customer's inventory. A similar sharing of information across customer relationships has been discussed regarding P&G and Wal-Mart. Michael Green, the Director of Information Technology for the Wal-Mart Sales Team at P&G writes about the effects of the experience.[27]

> "P&G replenished Wal-Mart's inventory based on inventory data received from Wal-Mart's distribution center (DC). This data allowed P&G to manage the inventory levels to ensure that P&G products were in stock at all times. P&G reduced the order cycle time (amount of time from the order generation to delivery) by 3-4 days. This process also dramatically increased inventory turns which resulted in a reduction in the inventory of the entire system."

Thus the P&G and Wal-Mart experience shows that shared information can have a direct benefit to the business. Thus the new system discussed in this chapter evaluated for P&G and one of its grocery customers was intended to provide more granular input data for generating forecasts, and thus expected to improve accuracy.

7.1 Current System: Customer Replenishment Process, SAP FIRST

Vendor Managed Inventory, VMI, refers to the relationship between P&G and a customer where P&G tracks the inventory at the customer distribution center, and replenishes stock according to the customer's needs. The customer, for example, a mass-merchandiser or grocery store chain, shares its shipment and inventory data with P&G, so that P&G can generate and fulfill purchase orders to maintain an agreed quantity in the customer's distribution centers.

P&G's customer service organization is responsible for working with the customer to track the inventory level and place the shipment orders so that the customer's inventory will be properly maintained. The process that the customer service organization utilizes to monitor and replenish material is referred to as the Customer Replenishment Process, or CRP. The consumer service representative is evaluated based on the inventory levels maintained at the customer, the service level (preventing out-of-stocks) and the customer feedback. There is a perception that the more

---

[27] Grean Michael and Michael Shaw. "Supply-Chain Integration through Information Sharing: Channel Partnership between Wal-Mart and Procter & Gamble"

personal attention, and the more "touches" a representative makes to the process for the customer, the better the service.

The demand planning process is the same regardless of whether or not P&G has visibility to the customer inventory level. The demand planner uses a system co-developed by P&G and SAP, called FIRST, to maintain a demand model and generate a demand plan. The planner also consults with account reps to learn when an event such as a promotion or introduction of a new product is planned, and accounts for these instances. While the demand planner generates a forecast for each individual SKU, the forecasts are aggregated for weekly buckets, and for customer groups. For example, all of the grocery chains may be aggregated into one customer group.

The level that the customer service representative works is much more detailed than the level that the demand planner functions at. Thus, in the current system, there is little connection between the demand planners' weekly process and the customer service reps' process for generating orders for the customer's distribution centers.

7.2 Proposed System: SAP-Responsive Replenishment

P&G has partnered with SAP to develop a new system solution, Responsive Replenishment, which improves short term demand visibility for cases specifically where a VMI relationship exists with a customer. The new system, RR, includes an interface between the customer's information system and P&G providing visibility to inventory levels and shipments from customer's distribution centers on an automated basis. The new system also integrates the inclusion of special event information. The new system automatically generates a short-term forecast from the recent history, thus the demand planning process is not necessary for short-term forecasting.

The new system is only applicable for very short-term forecasting. The P&G demand planning process develops a forecast for two years into the future. The timely visibility to customer inventory and sales data will provide little benefit to long-term forecasting. Thus RR is intended to only replace the demand plan for only a couple of weeks. Additionally, the RR system is only applicable for customers with a VMI relationship with P&G. The following figure depicts how RR is only applicable for a fraction of the planning process.
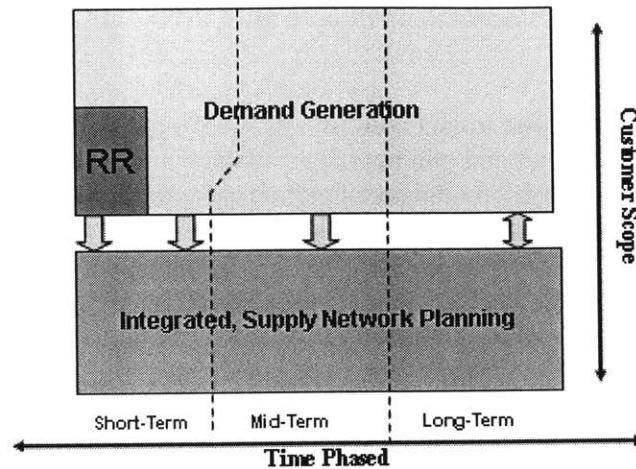
Figure 7.1 Fraction of the Planning Process that can utilize Responsive Replenishment

Figure 7.1 depicts that the RR system is only applicable for part of the short-term forecast, and will have no effect on mid to long-term planning. This can be seen as the RR box only stretches a short distance in the x-direction. Additionally, RR is limited in the y-direction as the system can only be used where P&G manages the inventory at the customer's location. Thus, this block represents the small fraction of P&G's demand planning which RR can support.

The business process also changes using RR. The system now automatically conducts forecasting, DRP (demand requirements planning), TLB (truck load building) and order creation. The customer service representative monitors the output and can intervene if necessary. The RR system uses the history and model parameters to forecast the demand requirements for a few weeks in the future. RR executes DRP which assesses how much is needed to keep the customer in-stock and then rounds the quantities into orders based on P&G order constraints. An example of a constraint is that full cases or pallets must be shipped as the distribution center should only order a minimum quantity of a case of deodorant, not a single stick. The customer service representative then uses RR to run the TLB process which optimizes the order to add or remove products to minimize customer out of stocks while balancing a low inventory level. Finally, the purchase orders are created and the customer service representative reviews them and releases the orders for processing.

7.3 System Differences

Although RR can only be applied to a fraction of P&G's business, the impact was expected to be large and justify the substantial development effort. The following table summarizes the key differences between the systems and processes with the current and proposed systems.

CRP + SAP-FIRST	RR
• Existing system and process • Not currently linked to live	• New system and process • Linked to customer information

customer data	system
• Produces weekly forecast	• Produces daily forecast
• Updates forecast weekly	• Updates forecast daily
• Forecasts for aggregated customer group (e.g. All Mass Grocers)	• Forecasts at customer distribution center level (e.g. Boston Warehouse for Shaw's Grocery)
• Event information collected for major customers and aggregate scaled	• Event information visible and input at a customer DC level,
• Active management required	• Exception driven process

Figure 7.2 Comparison of Current versus New System

Figure 7.2 depicts how more granular input data will be available for generating forecasts. The current system takes snapshots from the customer data, and then updates the long range plan, generating forecasts for weekly buckets at a customer-group, regional level. It only updates when it is executed each week, thus if new information is received in the middle of the week, it will not be reflected until the plan is run the subsequent week. The new system is linked to the customer information system and thus it can be set to update the forecast daily. Additionally, as it sees the daily transactions at the customer distribution centers, it can be configured to predict daily sales from particular P&G plants or distribution centers, as it knows the exact planned location of the shipment, and is the same software that rounds the trucks out to create full-truck loads. Hypothetical examples of forecasts from the two systems highlight these differences:

Current Demand Planning Example Forecast: For week 26, Wal-Mart in North America will purchase 1000 units of Cold Water Tide in 100oz bottles. All-mass-grocers will purchase 800 units of Cold Water Tide in 100 oz bottles.

RR Demand Planning Example Forecast: For week 26, Monday, Wal-Mart Distribution Center in Arkansas will purchase 32 units of Cold Water Tide in 100oz bottles from P&G's Cincinnati Distribution Center. Tuesday, the Shaw's Market Distribution Center in Cambridge will purchase 20 units of Cold Water Tide in 100 oz bottles from P&G's Boston Distribution Center.

These highlighted differences between the proposed system and the existing system would clearly provide more granular data, which should theoretically produce a better forecast. To quantify the savings potential from the implementation of the new system, a pilot test was conducted.

7.4 Pilot Scope

An integrated pilot was commissioned to investigate the impact from implanting the new developed RR (Responsive Replenishment) system. This section will provide context and scope for the pilot trial.

7.4.1 Motivation and Broad Scope

Turn forecasts are relatively predictable with the current process. Opportunities for improvement lie in better event visibility and in incorporating more detailed information to the forecast. This additional level of detail should be more accurate than the current logic of maintaining average relationships of how a product volume is split between days and sites, commonly called the "split tables". Thus, this new system appeared promising as it would incorporate granular promotion forecasts and shipment data into the demand planning process. It would allow for actual event/promotion performance feedback to be incorporated into the process. All of this information would now be seen and able to be analyzed and improved at the producing site level, improving the supply planning operation at P&G. With this new system providing these clear advantages, quantifying the value of these savings became important as the costs were non-trivial. Not only was the cost of development of the system expensive, but the cost of the implementation, along with the organizational changes, were worrisome, especially since this large fixed cost system solution was only applicable to a fraction of P&G's customer relationship. Thus, the pilot test was commissioned to get the new system up and running for a test scope to prove its functionality, and to quantify the savings possible. These savings were then to be extrapolated to the rest of P&G to determine whether a full implementation was justified.

7.4.2 Expected Impact

The differences from the new system implied many changes which were expected to have a significant impact in the following areas:

- Improved Sales: Specifically, the closer connection between the customer distribution center and the forecast was expected to reduce the out-of-stock (OOS) occurrences. Out-of-stocks can lead to lost sales, which can be very costly for a branded product and can translate into a lost customer. These savings were supposed to have a larger impact for merchandising events and instances when supply disruption type activities occur.
- Reduced operating cost: This improved visibility would reduce costs associated with emergency orders to respond to product gaps, avoiding the out of stocks in the above example.
- Reduced inventory: True visibility to the customer was expected to reduce variability and therefore reduce safety stock.
- The implementation and organizational cost would be measured for the infrastructure and supporting organizational changes.

A larger effort was underway at P&G to invest in numerous projects leading towards a "Consumer Driven Supply Network." This RR project was intended to serve as an example of the type of time, cash and costs savings from implementing an end-to-end integrated solution. It would demonstrate the ability to leverage results in a core P&G business – big category, big country, and big customer. It would also quantitatively demonstrate that demand information for events can flow from the customer to the producing site to enable visibility in the supply network.

7.4.3 Detailed Scope

To accomplish the goals of the project, a narrow scope was defined using one of P&G's core business areas:

- Product: Essentially all of the brands and products associated with a major category for P&G were chosen to be in scope.
- Customer: One major grocery chain was selected and 11 of its distribution centers were included in the scope.
- Duration: All volume over 3 month period was to be measured.
- Information: The systems would be implemented as a final design and implementation costs, supply network benefits, and organizational/work process lessons would be captured

7.5 Evaluation

The major cost savings was intended to come from inventory savings from improved forecast accuracy. Thus, to evaluate the savings potential, the accuracy improvement had to be measured and translated into a dollar value. P&G has a department, "Global Analytics" which set the statistical process for determining safety stock.

7.5.1 P&G Safety Stock Model

Wim Van de Velde provided training and background on the P&G process, which will be summarized in this section.[28] P&G utilizes one model for all P&G regions, all business units. For some complex businesses a 'custom model' is built to address the need, or the model is upgraded to include the scenario. Four main factors influence safety stock:

- Supply chain uncertainty/variability, σ. This includes forecast error, production schedule unreliability, and lead time uncertainty. Most effort should go towards estimating the biggest source of uncertainty. Generally, P&G has managed to control production and transportation relatively well. The big 'unreliability' struggled with is forecast error. Thus most attention is paid to demand uncertainty for determining safety stock.
- Total Reaction Time, TRT, is the time between the identification of a production need and the finished product being available at the P&G DC to be picked for shipment to the P&G customer. This time includes the lead time to get into a production schedule, supply lead time for materials not held in inventory, quality release time, transportation time.
- Target availability, Avle, is the desired service level, which represents the percentage that the product is available on the shelf.
- Cycle Time/Batch coverage, represents the batch size. For example, frequent production of small batches will result in lower inventory.

Safety stock is determined by the following formula:

$$SS_{Days} = k \times \sigma_{daily} \times \sqrt{TRT}$$

The variability term is determined by taking the standard deviation of relative forecast error: (Forecast-Actual)/Forecast. However, as this is not readily available, most P&G businesses

---

[28] Van de Velde, Wim. "Finished Product Inventory Model Training." P&G Internal Documentation. March 2006

approximate the term using the following relationship: $\sigma_{FE} \approx 1.25$ MAPE, where MAPE is the mean-average-percent-error discussed in 5.4 MAPE, Mean-Absolute Percent Error . However, for the purposes of setting safety stock, MAPE should not be aggregated across products, and needs to be determined across a time period. With this information determined, it was clear that the appropriate measure for evaluating an opportunity in inventory savings would require forecast accuracy to be measured using MAPE.

7.5.2 Measuring Forecast Accuracy Improvement

To measure the improvement in forecast accuracy, data was extracted and analyzed each week regarding the forecasts and shipments for both the RR and FIRST systems.

7.5.2.1 MAPE by SKU → Method

Each week, the forecast and shipment data was pulled for each SKU in scope for the previous week's shipments. The shipments were compared to ensure that the scope was identical, and that no data issues were present with data collection. An issue in the master data delayed collection of accuracy data for four weeks. Forecast horizons for 1, 2, 3 and 4 week intervals were evaluated. Thus for each SKU, and for each forecast horizon, the absolute percent error was determined. Simultaneously, the process checked to make sure that a forecast and shipment were recorded in both system, FIRST and RR. If either system did not indicate a shipment or a forecast, the data point was not ignored for that SKU. This was done so that the same group of SKUs and same amount of data would be used in the comparison. The discrepancies were fed back to the respective groups to be investigated. The direct correlation to inventory savings would come from the MAPE calculated by averaging across a given SKU for the duration of the pilot. The MAPEs for the different SKUs could then be averaged across the set of products to give an aggregated accuracy measure for the pilot. While this was the final method to be related to inventory savings, each week the MAPE was aggregated across the SKUs to give an indication of how accurate the system was performing.

7.5.2.2 MAPE by SKU, and by manufacturing site → Method

One of the expected benefits from the RR system was the detailed level of information, providing forecasts for the producing site. The RR system had visibility to the exact customer distribution center to which the shipment would be sent, and thus could provide P&G site specific forecasts. The current system, SAP-FIRST, generates forecasts at the regional level, and then a split table divides the demand between the plants. For example, one product line with two manufacturing plants had a 30%/70% split maintained. After the demand plan was generated, the split table divided the total demand into plants for the supply planners to fulfill. A calculation was used to determine the accuracy improvement of RR over the SAP-FIRST system at the site level. To perform this analysis, the component absolute percent errors (APEs) were determined from the RR forecast by site, and averaged to determine the MAPE across product lines by site, for a given forecast horizon. The regional forecasts were extracted from SAP-FIRST, and split according to the split table logic to determine what the site forecasts would have been. The MAPE accuracy was then calculated for these sites from SAP-FIRST. This analysis should be more applicable to the business cost as it directly indicates how the forecast is

consumed by the downstream DRP (demand requirements planning) operation. However, the smaller volumes at each site will have larger error as they have lost the impact of aggregation from combining two demand streams.

7.5.2.3 Daily MAPE

While most of the accuracy calculations were based on weekly buckets, RR has the ability to forecast on a daily basis, as it evaluates the daily shipments from the customer distribution center. A calculation was also desired to determine the improvement of RR over SAP-FIRST. A MAPE was calculated for RR for daily buckets. For SAP-FIRST, the daily split tables were used to convert weekly forecasts into daily forecasts, and then the actual shipment data was obtained on a daily basis. However, the shipment volume to the customer in the pilot was too small for this analysis to be meaningful. Typically, there would be one day a week where a product was shipped, and since the forecast was spread across the entire week, the scale of the pilot could not support daily analysis. This same approach was also attempted for rolling three day consecutive shipments to attempt to alleviate the problem above, however, it did not yield much information.

7.5.2.4 MAPE by Customer Distribution Center

Additionally, the MAPE was determined for each of the customer distribution centers. Each week we attempted to analyze if a particular distribution center had higher error than another. However, this analysis did not provide any insight and only assisted in investigating the root cause of forecast errors.

7.5.3 Extrapolating to P&G

Wim Van de Velde, from Global Analytics championed a process for extrapolating the savings potential to all of P&G.[29] This process was important as it would be critical to a multi-million dollar investment decision. To extrapolate savings, several assumptions were required. Each of these was supported by the Global Analytics department. First, it was assumed that the percentage-point improvement in forecast accuracy would be achievable for the rest of P&G's businesses. Next, this improvement would only occur for the fraction of the business operating under VMI. So while that fraction of the total volume will have an improvement in forecast accuracy, the effect on the accuracy of the remainder would is unknown. It is possible that when VMI inventory is removed, the forecast accuracy would worsen for the remaining volume as disaggregating increases error. However, it was assumed that error would be approximately the same for the total volume, and for the non-VMI volume, as VMI represents a small fraction of the total volume. This could be verified if forecast history could be differentiated based on VMI and non-VMI relationships. However, forecast history is not stored by customer, therefore this is not possible using the information system, and would have to be built manually by collecting data points.

---

[29] Van de Velde, Wim. "Finished Product Inventory Model Training." P&G Internal Documentation. March 2006

7.6 Findings

Surprisingly, the accuracy achieved by RR was only marginally better than FIRST, and not sufficient to justify further investment. This section will review the results of the analysis and other findings uncovered which help to explain why the RR project was canceled.

7.6.1 MAPE by SKU → Results

The key driver for cost savings was expected to come from improved MAPE which would translate into inventory savings from reduced safety stock. The calculation was performed as described above with a few modifications. As mentioned before, approximately four weeks of data were not usable due to an initial master data issue. Then subsequently, communication problems at the end of the pilot led to confusion regarding when the transition back to CRP was to occur. This resulted in the last week of data having to be discarded as the RR data was not complete. Other issues related to the work process employed by the end users produced points for a particular SKU for a particular week of exceptionally high error. The largest contributors to forecast error were evaluated each week, and if attributed to poor training for the new system startup, and not expected to recur, the forecast was adjusted to compensate for this. The concluding results of the forecast accuracy comparison are shown in the table below for the smallest time horizon (1-week).

	14-Aug	21-Aug	28-Aug	4-Sep	11-Sep	18-Sep	25-Sep	2-Oct	10-Oct	Vert Avg	Horz Avg
DP MAPE	64%	56%	55%	44%	49%	56%	53%	64%		**58%**	55%
RR MAPE	51%	31%	71%	38%	56%	63%	37%	70%		**60%**	53%
Point Improvement	13%	25%	-16%	6%	-7%	-7%	16%	-6%		-2%	2%
Percent Improvement	21%	44%	-28%	13%	-14%	-13%	30%	-9%		-4%	4%

Figure 7.3 Forecast Error Comparison for 1-week forecast horizon

Reviewing the MAPE for all of the SKUs, aggregated across time in Figure 7.3 shows small to no improvement in most weeks, and essentially no improvement overall. The last column is blank to emphasize the week of missed data, and how one organizational miscommunication can result in a loss of over 10% of the data possible. Each of the weeks was included to highlight that the error in RR seemed to be much more variable than SAP-FIRST. While SAP-FIRST had minimal attention from the Demand Planner as this customer represented a very tiny fraction of the overall demand volume, the forecast was relatively steady and error occurred from variability in shipments, varying between 44 and 64% error. On the other hand, RR varied from 31 to 71% error. RR ranged from 25% points better than FIRST, to 16% points worse that FIRST. This behavior was unexpected and provoked further analysis.

Products were classified into New (newly introduced SKU with little history), Promo (highly promoted products), and Turn (the majority of the SKUs without any exceptional behavior. The error measure (MAPE) was aggregated for products along these lines in the following tables:

Forecast Accuracy by Material Type: 8 Wk Average by Sku												
	RR DATA						FIRST					
Weeks Back	PROMO		NEW		TURN		PROMO		NEW		TURN	
	MAPE	# skus	MAPE	# skus	MAPE	# skus	MAPE	# skus	MAPE	# skus	MAPE	# skus
wk 01	77%	1	137%	10	51%	80	76%	1	93%	10	53%	80
wk 02	N/A	0	129%	9	47%	79	N/A	0	161%	9	51%	79
wk 03	N/A	0	463%	9	66%	82	N/A	0	255%	9	58%	82

Figure 7.4 Forecast Error Comparison broken out by New, Promoted or Turn Products

Evaluating the differences between the turn, new and promoted products showed that for 1, and 2 week forecast horizons, RR was nominally better than FIRST. For 1 week forecast horizons, the difference between RR and FIRST came from the new product data. Additionally, for the 2, week forecast horizon, RR was better than FIRST in both categories. This prompted evaluation of multiple weeks back as shown in the following table:

	RR DATA	FIRST
Weeks Back	MAPE	MAPE
wk 01	60%	58%
wk 02	55%	63%
wk 03	105%	78%

Figure 7.5 Summary of RR versus FIRST Forecast Error for different horizons.

It was surprising to see the accuracy improve when forecasts were generated further from the actual shipment date. Standard supply chain management textbooks emphasize that long-range forecasts are less accurate than short-range forecasts.[30] A potential reason for this, to be discussed in further detail later, is that as the actual shipment date approached, the customer service reps were paying more attention to the order and making more user interventions intended to improve service to the customer. But before discussing this, a little more analysis was conducted to analyze the behavior of RR versus FIRST for different forecast horizons. The following figure shows data points for each of the forecast horizons over time. If RR was better than FIRST, the data point appears above the bold line on the top half of the graph. If FIRST had less error than RR then the data point appears below the bottom line on the bottom half of the graph. The magnitude represents the difference between the two systems.

---

[30] Anupindi, Ravi, Sunil Chopra, Sudhakar Deshmukh, Jan A. Van Mieghem, Eitan Zemel. Managing Business Flows. Prentice Hall, 1999.

RR Better than FIRST ??
(removing worst 5 data points)
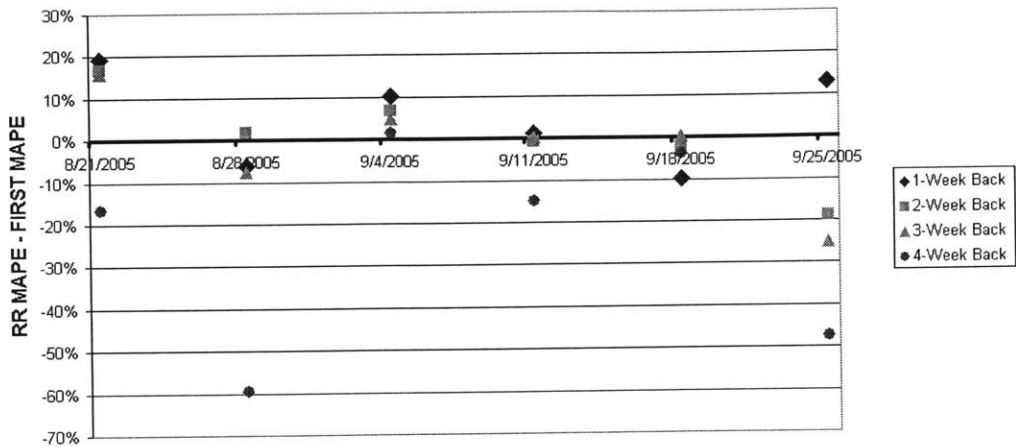
Figure 7.6 Difference between RR and FIRST forecast errors

Evaluating the figure above, FIRST clearly does a better job of forecast 4 weeks out. But after that, the data points get much closer together. While there is fluctuation from week to week, this graph shows that there are many weeks where RR does outperform FIRST, or the difference is negligible. To see this more clearly, one more graph was generated for the one and two week back horizons.

Forecast Error Over Time - System Comparison, 1 and 2 Weeks Back
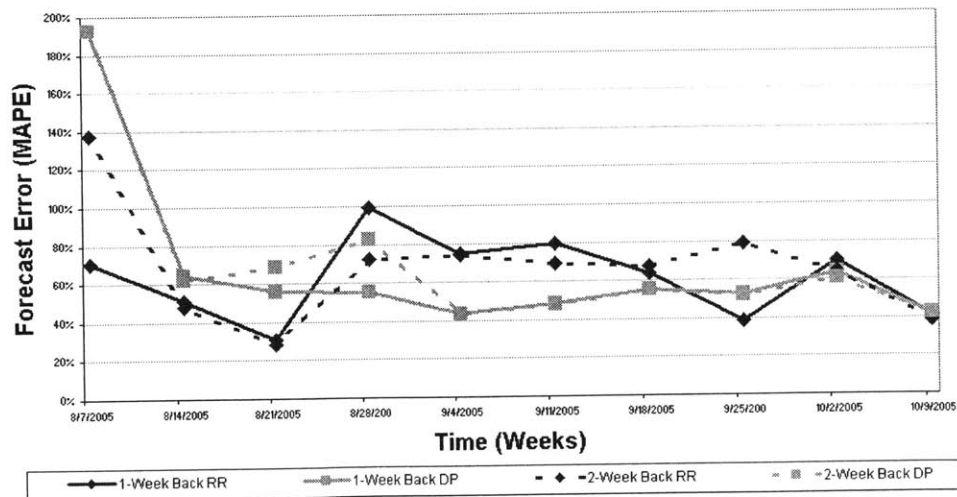
Figure 7.7 Comparing RR and DP for 1 and 2 week horizons

The solid lines indicate the one week horizon, and the dotted lines represent the two week horizon. The lower the line is, the lower the error. Thus the RR lines (darker) seem to start

below the FIRST lines, then swap above, and then come quite close to the FIRST error. Another hypothesis as that at the beginning of the implementation, users were still being trained in the system and were not making overrides or adjustments. But as they became more comfortable, the number of interventions increased, decreasing the optimality of the system.

7.6.2 MAPE by SKU, and by manufacturing site → Results

Disappointed in the improvement at the aggregated level for the shipments in total, an analysis was conducted evaluating the accuracy at the plant levels for the producing plants at P&G. The results are shown in the two figures below.

1 Wk Back MAPE
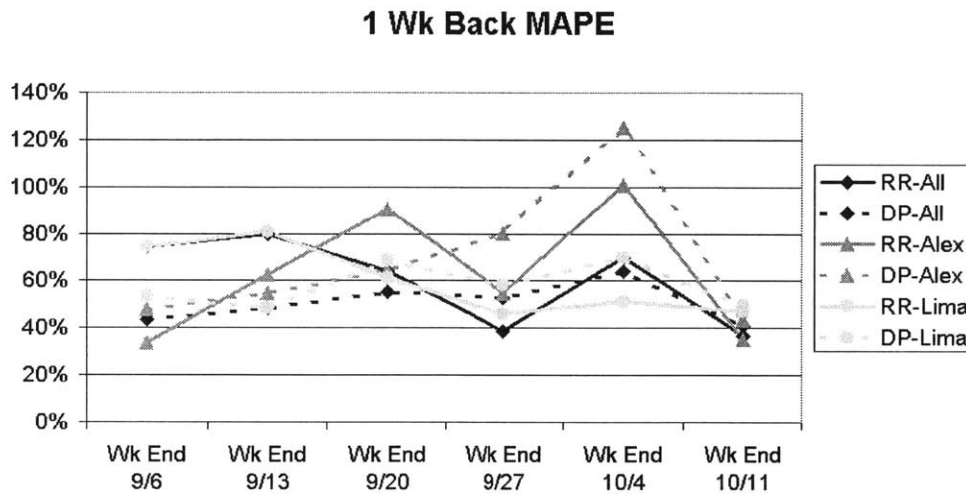
Figure 7.8 Comparing RR and DP at the plant level for a 1 week horizon

In Figure 7.8, the solid lines represent RR and the dotted lines represent SAP-FIRST. Standard supply chain textbooks indicate that aggregated forecasts should be more accurate than individual forecasts.[31] Therefore the darkest line which represents the total volume should be expected to have the lowest error. As the demand is disaggregated for the producing plants, the error is expected to increase. RR was expected to have much lower error, as the forecast for the producing plant was developed based on the exact shipments for that plant. DP was expected to have higher error as a fixed ratio was used to divide demand between plans without any time-based intelligence. These charts do not show any definite improvement of RR over DP for the manufacturing site. There are some instances where RR is an improvement over DP; however, we did not see the overwhelming improvement expected. This demonstrates that the split tables used to split customer demand into the manufacturing sites, while seemingly crude, are fairly accurate.

---

[31] Anupindi, Ravi, Sunil Chopra, Sudhakar Deshmukh, Jan A. Van Mieghem, Eitan Zemel. Managing Business Flows. Prentice Hall, 1999.

2 Wk Back MAPE
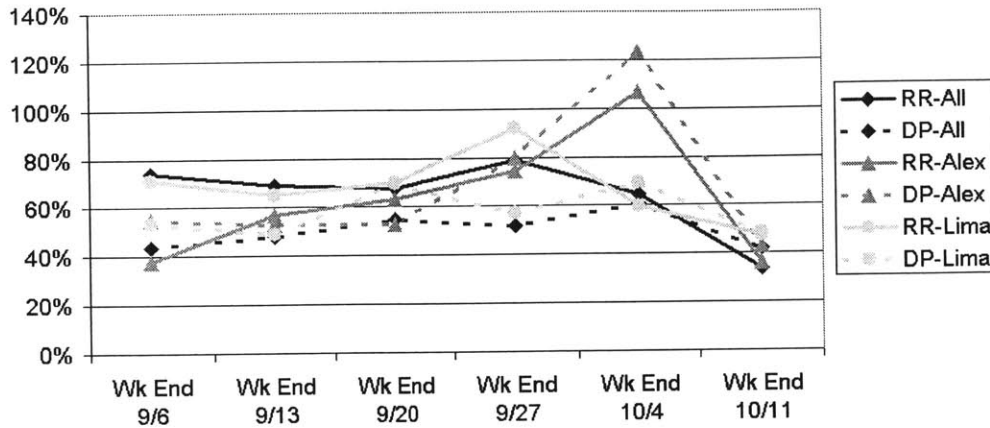
Figure 7.9 Comparing RR and DP at the plant level for a 2 week horizon

In Figure 7.9, again, the solid lines (RR) appear to generally have larger error than their dotted (FIRST) counterparts indicating that even for the two week forecast horizon, FIRST is just about as good as the complicated RR system.

7.6.3 SAP-FIRST Forecast better than anticipated

The above findings that RR provided minimal to no improvement over the FIRST forecast generated the idea that maybe the SAP-FIRST forecast did not have significant room for improvement. Thus, the following analysis was conducted using the legacy system, SAP-FIRST, to evaluate the different customer groups and their accuracy for the business analyzed.
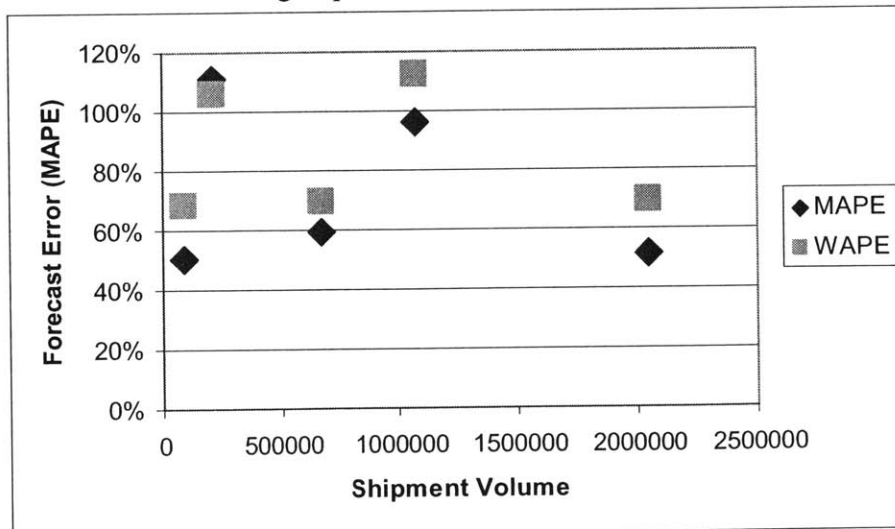
Figure 7.10 Forecast errors from the legacy system (SAP-FIRST) for different customer groups for the piloted business category

The expected relationship between shipment volume and error for the same products in the same region, would be that the larger volume customers would have more predictable behavior as the different stores comprising the large volume would balance out each others abnormalities. However, Figure 7.10 does not show a strong relationship between shipment volume and forecast error for the customer groupings and business displayed. Instead, the data point associated with the pilot has the smallest volume on the graph, as it represented only 11 distribution centers for one of the grocery store chains. However, it was surprising to see that this very small volume had one of the lowest errors.

Following this, a series of studies were conducted across all of P&G's businesses using the legacy system (SAP-FIRST) to determine if a strong relationship could be determined between forecast error and some measure of magnitude – whether absolute shipment volume, shipment % of business, or the number of SKUs.

MAPE vs Shipment as a % of Business

MAPE vs Shipment Volume
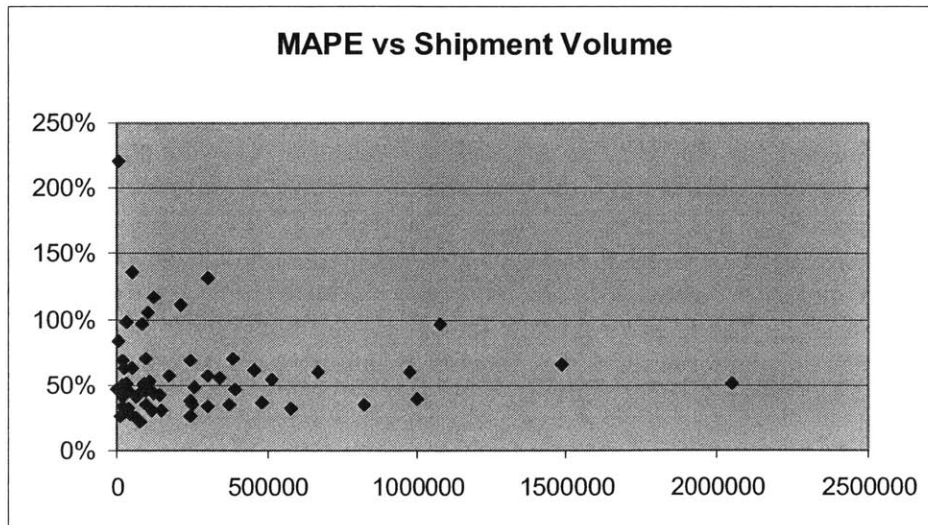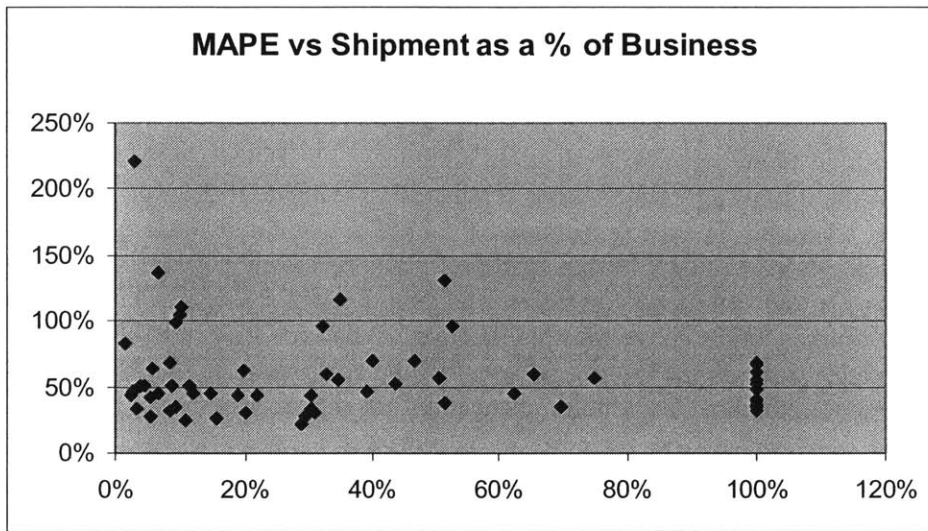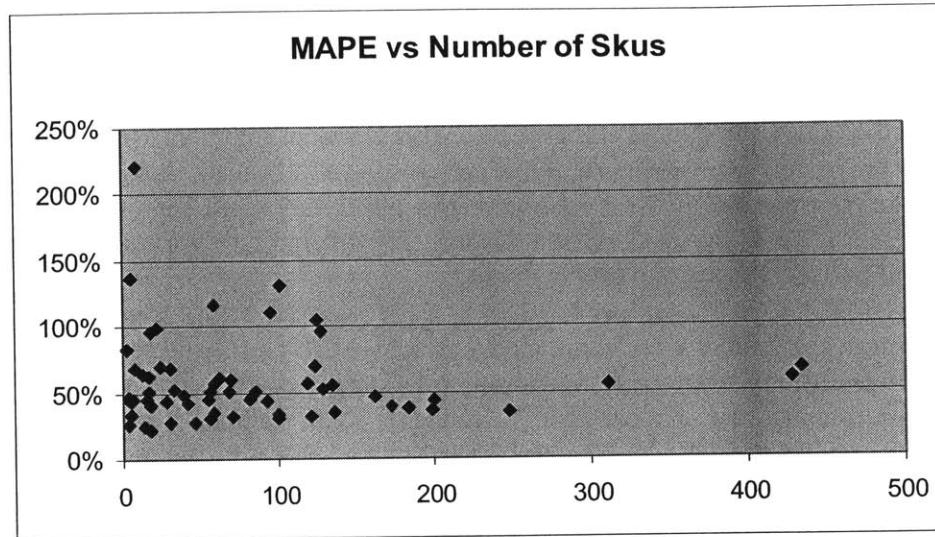
Figure 7.11 Error versus absolute shipment volume, shipment % of business, and the number of SKUs

Attempts were made to fit an equation to the data in figure 7.11. Linear, logarithmic, exponential, and polynomial regressions were conducted to evaluate if the data would fit any of these relationships. Across all of the regressions, the highest r-squared value achieved was 0.03. This indicates a very poor fit of data to the equation. Thus, no strong correlations were indicated by this data. However, the abnormally high error tends to occur for the smaller volumes and the abnormally high volumes tend to have reasonable error. What this may indicate, is that there are other important effects, such as the buying behavior of the customer, the relationship with P&G, or the products and plan themselves which may account for the error more so than the volume and aggregation.

7.6.4 Event Management

Another hypothesis was that the major savings from RR would come from closer integration to events occurring. Simchi-Levi, Kaminsky and Simchi-Levi emphasize how collaborative forecasting processes with customers produce excellent supply chain improvements by providing clarity to market demand, the impact of promotions, pricing events, and advertising.[32] However, during the period tested for the RR pilot, the magnitude of the events was small in comparison. This section will investigate this hypothesis. Prior to start of the full scale pilot test, one distribution center was converted to RR and evaluated over a period of time. The volume was too small to draw conclusions which could be applied to all of P&G, but some lessons were captured that were useful. The following figure indicates a different color line for each SKU in scope for this distribution center across a few months of shipments. The y-axis shows the MAPE – mean average percent error for these SKUs across time.

---

[32] Simchi-Levi, David, Philip Kaminsky, and Edith Simchi-Levi. Designing & Managing the Supply Chain. 2nd Ed. McGraw-Hill, 2003.

Weekly MAPE for RR Skus
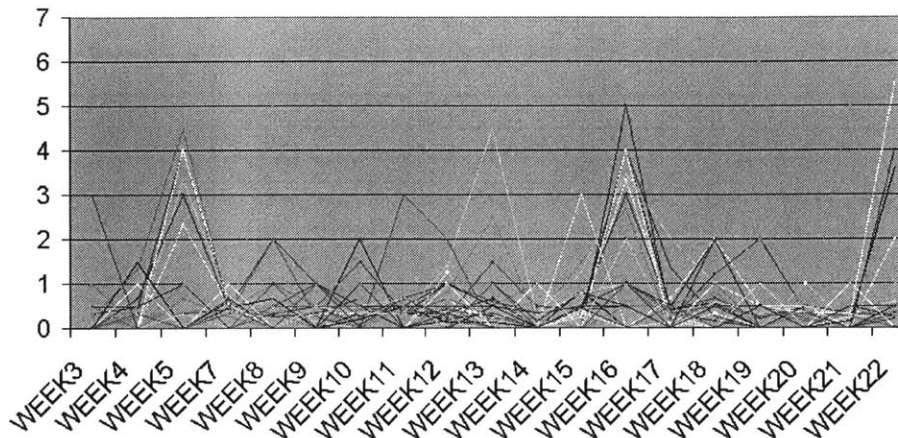
Figure 7.12 Weekly MAPE for set of SKUs, one distribution center

Figure 7.12 shows how for the majority of SKUs, the majority of the time, the error is less than 100% and the bulk of the lines falls under 1. However, certain periods occur where the accuracy is quite high and it is significant to notice that these peaks tend to occur at the same time across different products. Notice the peaks around weeks 6, 13, and week 17. Multiple products have large errors around the same time, indicating that a promotion or event was likely. These errors contribute more to the overall variability in demand than the weekly amounts. It is also interesting to see this data weighted for shipment volume as shown in the following figure.

Weekly WAPE for RR Skus
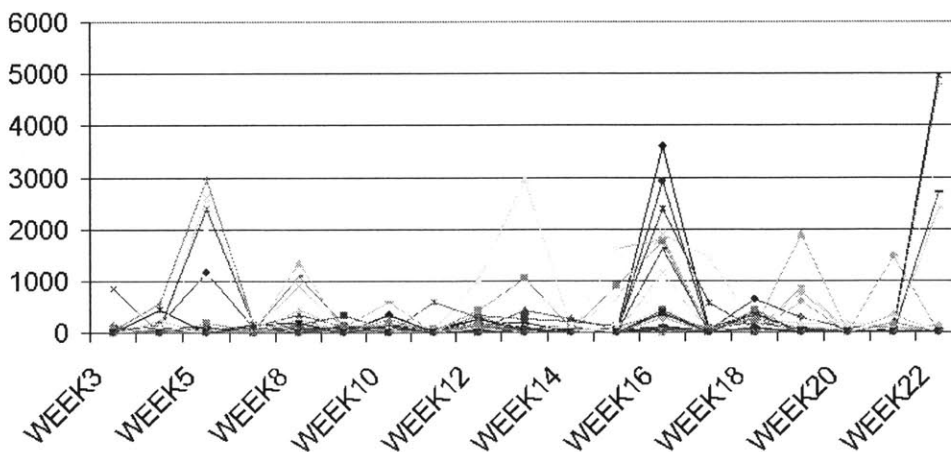
Figure 7.13 Weekly WAPE for set of SKUs, one distribution center

Figure 7.13 accentuates the behavior seen in Figure 7.3, as with the shipment weights we can see that the weeks that looked particularly bad in the non-volume weighted error measure are

exceptionally worse in the volume-weighted measure, indicating that these are also high volume shipment weeks, thus supporting the hypothesis that an event was occurring in this time frame.

The following figure shows the raw data for this preliminary analysis.

	RR DATA						FIRST - DP DATA					
	WAPE	MAPE	Data Points	No RR Forecast	No Shipment	No DP Forecast	WAPE	MAPE	Data Points	No DP Forecast	No Shipment	No RR Forecast
WEEK3	15%	27%	27	57	1	24	26%	40%	27	60	20	2
WEEK4	46%	29%	33	71	3	2	35%	37%	33	29	38	9
WEEK5	103%	73%	39	66	2	2	138%	63%	39	29	28	13
WEEK7	25%	29%	54	40	8	7	48%	39%	54	29	25	1
WEEK8	36%	64%	33	68	1	7	37%	56%	33	29	25	22
WEEK9	30%	34%	46	49	7	7	137%	80%	46	29	30	4
WEEK10	51%	45%	42	58	3	6	45%	41%	42	29	29	9
WEEK11	44%	37%	35	56	16	2	57%	54%	35	23	44	7
WEEK12	35%	40%	61	42	3	3	281%	118%	61	23	14	11
WEEK13	53%	44%	46	62	1	0	73%	46%	46	23	35	5
WEEK14	26%	20%	36	67	3	3	81%	49%	36	22	41	10
WEEK15	75%	40%	41	57	9	2	81%	47%	41	22	45	1
WEEK16	137%	151%	40	65	1	3	187%	125%	40	22	34	13
WEEK17	31%	18%	47	57	3	2	65%	42%	47	22	37	3
WEEK18	53%	45%	49	53	6	1	160%	74%	49	22	33	5
WEEK19	90%	66%	36	55	15	3	57%	64%	36	23	44	6
WEEK20	20%	18%	47	57	2	3	40%	32%	47	23	33	6
WEEK21	41%	33%	57	46	3	3	47%	41%	57	23	19	10
WEEK22	60%	54%	54	49	4	2	162%	61%	54	23	29	3
	51%	46%	43	57	5	4	93%	58%	43	27	32	7

Figure 7.14 Accuracy metrics for SKUs for one distribution center

Figure 7.14 indicates that a relatively small amount of SKUs (27-61) were contained in the analysis. The accuracy calculation was conducted for both RR and SAP-FIRST to test the algorithms. The highlighted sets of data are the only instances where the RR data point had a higher error than the FIRST data point. These also tend to coincide with the events. As evaluated in the previous section, larger volumes are expected to balance out the variability and result in smaller percent error. This was tested in the following chart for a small number of SKUs for 22 weeks prior to the pilot:
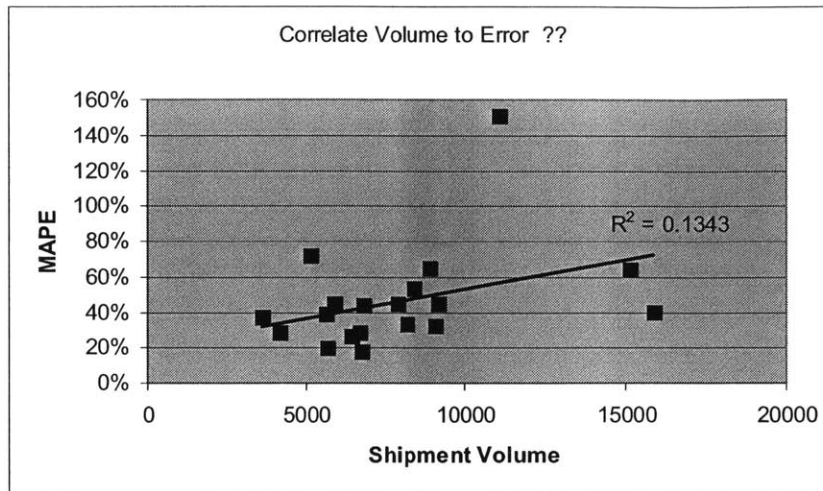
Figure 7.15 Relationship between MAPE and Shipment Volume using SAP-FIRST for a small subset of pilot data

Figure 7.15 shows little correlation between shipment volume and MAPE. If anything, the trend would be for error to increase as volume increases. While it may seem unusual for larger volumes to have higher error, larger volumes are more likely associated with events and promotions. This graph illustrates how the three points with exceptionally high volume have average to high forecast error. This perhaps could be driven by promotions.

This assertion can be used to highlight why negligible improvement was measured in forecast accuracy. The following figure demonstrates the shipment volume for all of the RR SKUs in scope for the pilot for an extended time period, and draws a box around the period during which the RR pilot was conducted.
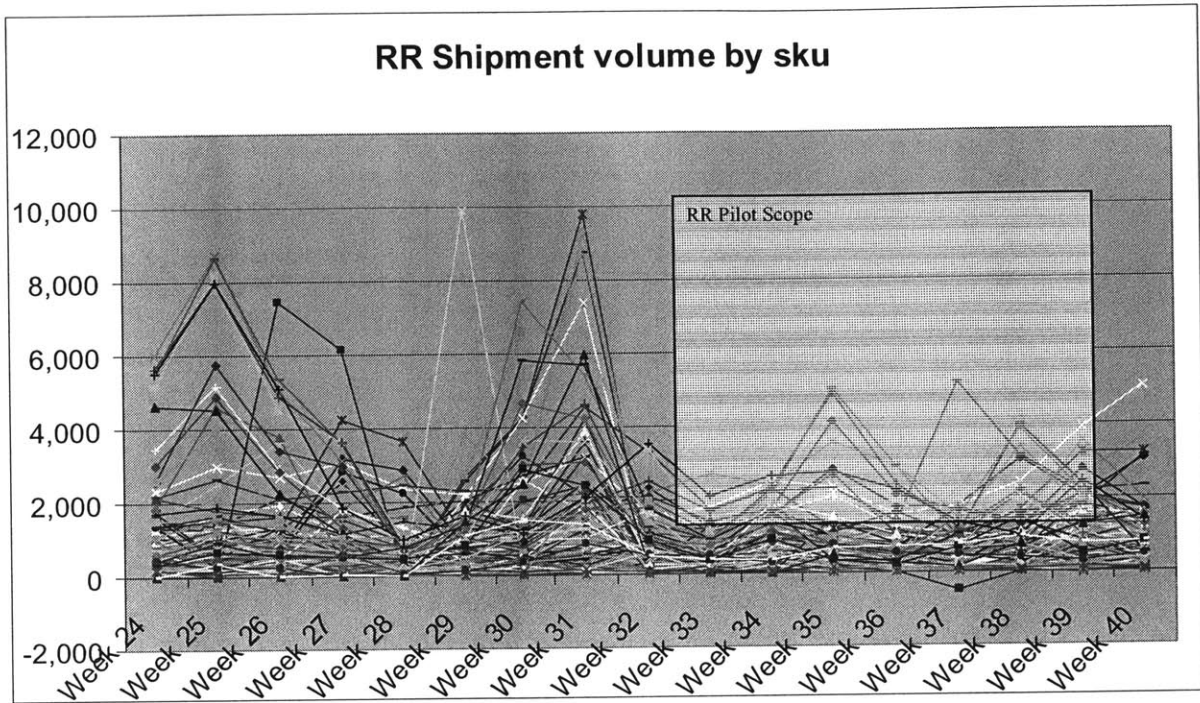
Figure 7.16 Weekly volume across SKUs in scope

Figure 7.16 highlights how the volume fluctuates from week to week. In the P&G business market, it is common for promotions to occur around the end of the fiscal year in June associated with the "Kids around the world" campaign and desires to meet year end targets. The pilot was conducted in August through October. The graph highlights how the volume during the period tested was much more stable with smaller promotional activity in the Pilot Scope period.

7.6.5 Organizational Changes

Another possible reason for the poor performance of RR is the challenge of implementing a new system which changes the business processes for the organization. Alejandro Lopezmonjardin studied the three customer service reps using the RR system and the new processes. The metrics displayed in Figure 7.17 show in general the performance of three customer service representatives and additionally, the amount of interventions, titled "fine-tuning," each of them tended to perform. These are easier to see graphically, as shown below.

Summary	User 1 (15% orders finetuned)	User 2 (1% orders finetuned)	User 3 (44% orders finetuned)	Total	Goal
Service	98.1	98.9	98.8	98.5	97.8
Inventory	18.2	13.9	22.6	17.2	17.0
Productivity index	1.26	0.99	1.45	1.17	1.00

Figure 7.17 Customer Service Representative Metrics During the RR Pilot[33]

Productivity Index (RR vs CRP)
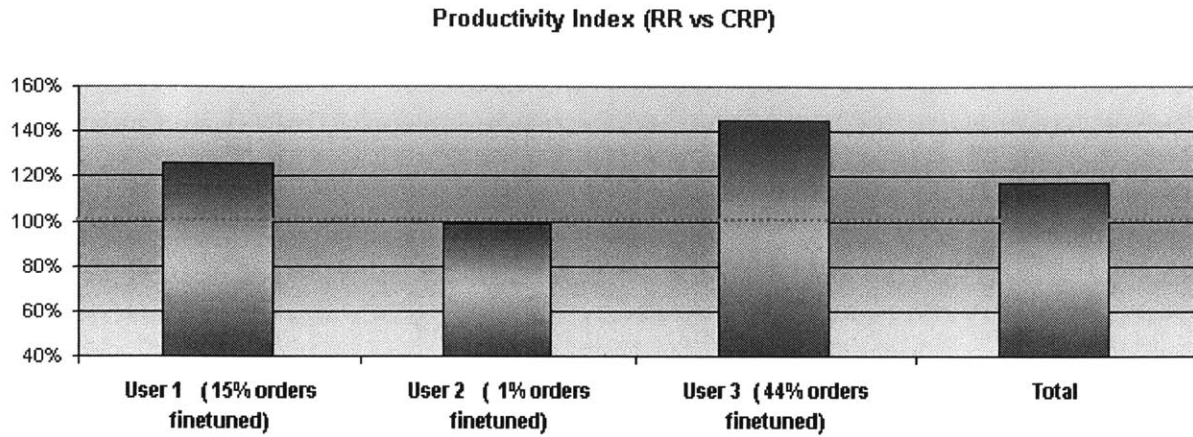
Figure 7.18 Productivity Index across Different Customer Service Representatives

The Productivity Index shown in Figure 7.18 is the ratio of the hours required for the customer service representative to perform the job using the new system (RR) versus the legacy system (CRP). Thus, User 2 spends the same amount of time with the new system as with the legacy system. However, Users 1 and 3 both spend more time with the new system. Thus the users who make the most interventions to the system, also take longer to perform their roles.

Service Results (% Service)
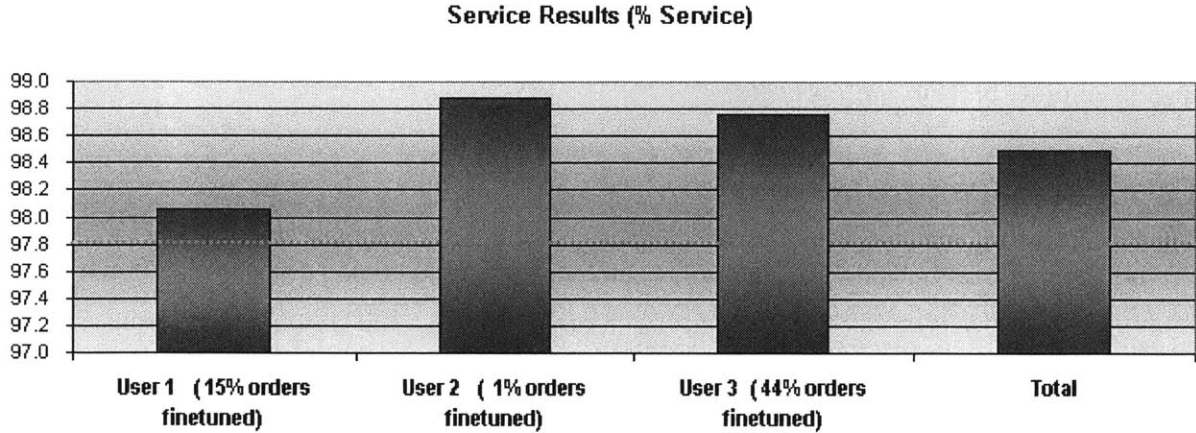
Figure 7.19 Service Results across Different Customer Service Representatives

Service measures the amount of inventory coverage at the customer's distribution center. The higher the coverage is, the smaller the chance is for costly out-of-stock occurrences. Here, it is interesting to note that despite the larger intervention and time spent from Users 1 and 3, it was User 2 who had the highest customer service.

---

[33] Lopezmonjardin, Alejandro. "RR Leadsite Results: Productivity Tracking" P&G Internal Documentation. 10/10/2005.

Inventory Results (DOH)
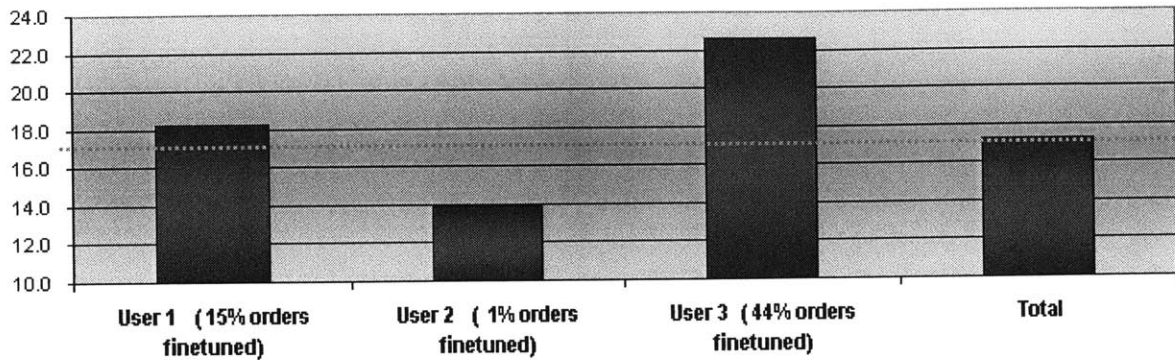
Figure 7.20 Inventory Results across Different Customer Service Representatives

Inventory Results measures the inventory measured as days on hand (DOH) at the customer's distribution center. The lower the inventory is, the better the performance is. Again, User 2 has the highest performance.

The conclusion here is that despite the training efforts and coaching provided, the historical mindset of the customer service representatives is that they are valued for their interventions. Historically, customers value this added 'touch' and results are better. However, the sophisticated RR system was designed to minimize the need for user interface and was intended to be "exception driven." In the duration of the pilot, it was difficult to impossible to change this culture and thus it has had an impact on the results. The key learning is that implementing a system is difficult if people are strongly reluctant to change. A careful plan is necessary to change the culture and the work process.

7.6.6 Analyzing Errors in RR

Saddled with the organizational data provided above, significant effort was spent trying to determine the potential behind RR if the organizational challenges could be overcome. Each week a root cause analysis was conducted reviewing the forecast errors, evaluating the contributing factors and assigning one of the following reasons:

- New Product – a new work process and training was needed for the VMI analysis, root cause identified the source of many errors was that the work process was not followed by customer service representatives, and that they were still adjusting to a new learning curve.
- Promotion – a new process needed to be trained and utilized between P&G and the customer to account for event data, this process was slow getting up and running, a new learning curve adjustment took time.
- Variation in movement – unusual demand, unpredicted promotion.
- Shipment Execution – truck not leaving when specified.
- TLB Sizing – rounding to truck size not as system predicted, might be due from strong promotional activity.
- Customer Data Issue – possible error in customer data.
- Source One Issue – some double counting or data appearing from unknown origin.

- User Adjustment – manual inputs or change to RR output.

Root Cause for RR Forecast Errors
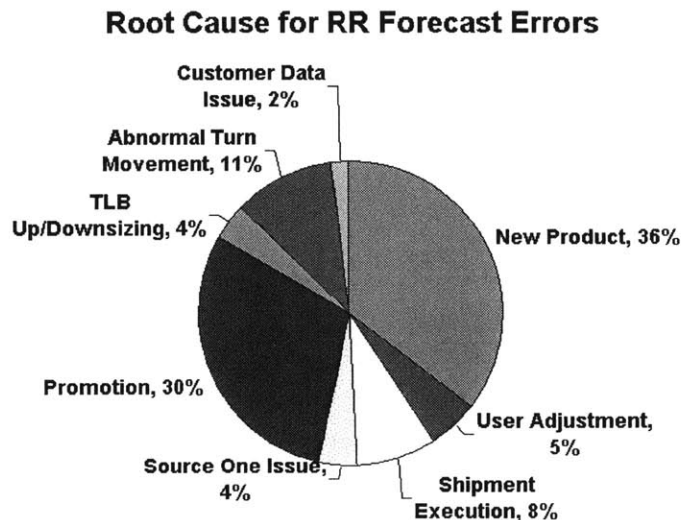
Figure 7.21 Distribution of Root Causes for RR Forecast Error

Analyzing the distribution of the types of forecast errors displayed in Figure 7.21 shows that the majority of the errors occurred from New Product and Promotion categories, which team leads feel could be improved in the future as work processes are stabilized. This indicates that the duration of the pilot was not long enough to really test if the RR system could operate more optimally if the organization were given time to shift.

7.6.7 Plant Interview

Much of the evidence above indicates that the optimal conditions for RR were not met in the trial, and that by looking at the snapshot, perhaps a large savings might still be possible. However, discussions with the plant led to the conclusion that the potential was small. The business chosen to conduct the system trial has a very short cycle time and plans production multiple times a day. With this system flexibility, their inventory level is not set by forecast error, but instead by emergency inventory storage in case of a catastrophic event. So the only potential savings would occur from raw material savings. However, discussions with the plant indicated that the magnitude of these savings was small in comparison to the cost of implementation at a single facility. Thus any savings from this site would be overwhelmed by the cost. Additionally, even if another business could be used, in general the fraction of customers using VMI accounts for such a small total volume, that even if the accuracy improvement was huge, it might not result in significant savings overall. For example, if 20% of the volume was able to be perfectly predicted, but the other 80% was completely unpredictable and subject to wide variations, aggregating the two might not make much difference to the amount of inventory required to compensate for the volatility in 80% of the total volume.

Additionally, discussions with the other businesses at P&G with longer cycle times indicated that the very short 1 to 2 week forecast horizons would not be beneficial to the business due to the

manufacturing and transit lead time for certain items. Therefore magnitude of the opportunity to roll out such a system, even if the forecast accuracy improvement had been large, is limited by the application of VMI, within the businesses with an appropriate cycle time.

7.6.8 Gillette Acquisition

One additional consideration is that simultaneous with the RR Pilot, P&G's acquisition of Gillette was approved and the organization suddenly had a new priority drawing attention. Where adequate resources and organizational attention may have brought the cultural change, and development of a system capable of delivering forecast accuracy improvements at a reasonable cost, it was not the company's priority at the current period.

7.7 Summary

The development and testing of this new SAP system provided a few valuable lessons.

7.7.1 Improvements Too Costly

The analysis detailed above indicates that RR was directionally better than the current system for certain horizons and would make a good customer replenishment tool, should the CRP tool need to be replaced. It provided better turn accuracy, lower inventory levels and higher service levels for the distribution centers. It also proved that it could integrate event information at a low level visible through the supply network.

However, the organizational challenges to replacing the current customer service, 'hands-on' organization with a much more automated tool would require more time and careful handling. Additionally, RR does not deliver significant enough benefits when compared to the large development and implementation costs. The implementation cost was much greater than originally expected due to complex, unique customer processes.

7.7.2 Scope Determination Key

Considerable resources, both financial and personal, were invested over a span of five years at bringing this system through development and through the test cycle. Some recommendations for future scope determination could help alleviate the cost of future trials.

- Savings potential analysis – Analysis of the DP forecast accuracy of the pilot scope conducted before the scope decision was made would have provided an assessment of the savings potential. In this case, were it known that the business and customer combination tested had little margin for improvement, another business and customer combination might have been chosen.
- Business Selection – A business should have been selected which depends on the forecast. The pilot business does not use the forecast unless it is a significant event. It is only used for raw material planning. Sites that utilize the forecast more actively as part of production planning could have provided a better representation of the impact of forecast accuracy improvements.

- Timing – A longer trial would allow time to see more representative business swings and opportunity to collect data and monitor the impact of work process improvements during the pilot.
- Volume – Improving Visibility will not impact the site until "bigger" customer spikes are removed. Later reviews determined that the full VMI Customer base would still not have big enough spikes.
- Opportunity – More detailed analysis into the potential magnitude of improvement. If RR only improves accuracy up to two weeks out, it will not help sites with a longer than two week supply chain.

7.7.3 Opportunities for Other Improvement

The study did present some interesting findings which open the door for more questions about improving forecast accuracy:
- The finding that this particular business and customer have such a high accuracy for such a relatively small volume indicates that there may be opportunities to learn how this partnership is so successful and to carry it to other businesses.
- While improvements in accuracy seem capable, the cost of these appears quite high. Are there cheaper alternatives?
- Opportunity for forecast improvements seems to lie with event forecasting and tying in information to the system to provide better updates.
- Granular data seems to present its own challenges in how easily a shipment sliding one day can strongly impact a daily measure. For the work involved at collecting and storing the additional level of detail, is it worth the additional granularity?
- Manual adjustments to sophisticated computer systems do not seem efficient. This topic will carry over to the next chapter.

8   Evaluating the Efficiency of Current Systems

This thesis has demonstrated how sales forecasts are critical to the success of P&G. It has proposed a method for creating accountability for forecast accuracy within the organization and has evaluated a proposed system for improving forecast accuracy. While the system proposal did not prove to be cost effective, the question remains, "How else can we work on improving forecast accuracy?"

This chapter evaluates a hypothesis from the Global Demand Planning Process Owner, Richard Clark, that the demand planning process has become so large, and the planners themselves are so over-worked, that opportunities for efficiency improvements may lay in the historical forecast data. This idea was partially motivated by the Radio Shack/SAS presentation at the Supply Chain Forecasting Conference in March 2005 which stressed that "MAPE is an appropriate metric for evaluating the magnitude of your forecast error......but MAPE is not an appropriate metric for evaluating the effectiveness of your forecasting process".[34]   This chapter will discuss the motivation behind such a study, the data available and the approach taken to evaluate efficiencies, the results and subsequent analysis, and finally some conclusions and recommendations for further improvements.

8.1   Motivation For Efficiency Study

With over 500 demand planners, P&G has one of the largest demand planning organizations known. Every demand planner interviewed is incredibly busy with their responsibilities and has little time to be pursuing interests, ideas or analysis that is not required for the organization to run. These planners work diligently to consolidate information from many different parties which impact the forecasts on different time horizons. They are responsible for predicting sales three years out on a SKU basis for different groups of regional customers. They need to keep up to date through discussions with the account managers to know of changes occurring in the business and promotions coming, and make sure they are providing the right input to the organizations which depend on the forecast. So while they have a tremendous amount of historical data available to them, they have little time to pursue new ideas or mine through historical data. Additionally, the organization has sufficient turnover with approximately 1/3 of the demand planners turning over to other jobs within P&G each year. Thus, in many cases by the time the demand planners are up-to-speed enough to recognize potential areas for improvement and data analysis that would be warranted, they may be moved on to another position.

Not only are intense personnel resources invested in forecast accuracy, the information systems that store and organize the data, and maintain models to predict future behavior are costly. Additionally, an entire support organization for these systems exists.

---

[34] Harwell, Jack and Michael Gilliland. "Applying Forecast Value Added at Radio Shack." Conference Presentation at the Institute of Business Forecasting Supply Chain Forecasting Conference, Phoenix, AZ, March 1, 2005.

When evaluating the magnitude of the resources invested in demand planning, questions arise about the benefit when weighed against the cost. Other organizations with much leaner demand planning organizations have been acquired by P&G, so people question if this department is a competitive advantage enabling P&G to function more efficiently, or if demand planning has reached the point of diminishing returns where the value of increased forecasting effort is not worth the associated cost.

8.2 Analytical Approach

To evaluate efficiency, it seems important to have reference data to how efficient other processes or organizations are with their demand planning. Benchmarking was considered as an option to comparing demand planning processes and organizations across different companies and determining some relationship between accuracy and effort and seeing how efficient different companies compared. However, it quickly became clear that each company has very different work processes, organizational structures, and methods of measuring and tracking forecast accuracy. To simplify the analysis and make it as objective as possible, the first approach undertaken was to compare the demand forecast generated for different businesses within P&G with each other. Anecdotal discussion with the demand planners indicated that a more significant amount of time and effort was dedicated to shorter term (weeks to months out) forecasts relative to longer term (years out) forecasts. Thus efficiency improvements would be more meaningful if they could be realized for the shorter-term forecast. Further analysis was conducted by comparing forecast accuracy from P&G business processes with statistical moving average.

8.2.1 Data Available

P&G information systems store three years of forecast and sales history, by SKU. A SKU is referred to as a finished product code, or FPC. Sometimes FPCs are combined to form a demand forecast unit, or DFU. For example, two different SKUs may represent the same size, same fragrance, of a bottle of hair shampoo, but when the packaging was updated to reflect the most recent marketing development, the new packaging receives a new SKU or FPC number, although they fall under the same DFU. A hierarchy exists within the product line and DFUs roll up based on certain product characteristics (perhaps fragrance, or size) until they roll up typically into a brand, (for example, Pantene). A few brands may roll up together to make a category (i.e., hair care). The first decision was to choose the appropriate level of analysis. As we evaluated accountability in a previous section, the first analysis was conducted at the category level for all P&G businesses, to evaluate how different categories compared in their forecast accuracy. For one business in particular, one brand was investigated for the different types of products sold under that brand. Eventually, the analysis would reach the DFU level.

Additionally, another key factor in evaluating forecast accuracy is the amount of time in advance that the forecast is generated. Historical forecast data existed for forecasts generated 1-wk, 2-wk, 3-wk, 4-wk, 6-wk, 8-wk, and 12-wks before the actual sale. Analysis was conducted for each of these forecast horizons, and eventually as the analysis went into more depth, the horizon was chosen to match that of the supply chain. In each case, the duration of the forecast was one week.

Also available in the historical system were forecasts generated at two different points in the P&G demand planning process. The statistical forecast refers to the forecast generated from the computer model which evaluates three years of sales history against certain parameters at a SKU level and predicts future sales. The demand planners are responsible for selecting the appropriate forecast model used and for updating the parameters. The output from the computer model, the *statistical forecast*, is subsequently updated by the demand planners to include timely information about merchandising, promotions, special events, or key information from the account executives. This final forecast is called the *independent net forecast* and is what is used by the organizations to make decisions. Both of these forecasts were used in accuracy calculations.

8.2.2 First Approach: Category Accuracy

The first statistical analysis extracted 52 weeks of forecasts and shipments for over thirty categories of P&G business. Sample categories included: Antiperspirant/Deodorant, Baby-wipes, Color Cosmetics, Coffee, Diapers, Feminine Care, Fabric Enhancers, Filtration, Healthcare, Hair care, Hair color, Laundry, Men's Grooming, Oral Care, Skincare, and Snacks. The forecast accuracy was determined for both the statistical and independent net forecasts for each week, for each set of forecast horizons. Forecast accuracy was measured as an index (actual/forecast) for each of the forecasts. The standard deviation was also determined for the forecasts across the year. The same analysis was repeated, but measuring forecast accuracy by WAPE across the 52 weeks.

8.2.3 Second Approach: Accuracy of product forms within a brand

The second statistical analysis took one category and delved into one brand in particular. One brand may be sold in different forms, and the same analysis as depicted in 8.2.2 First Approach: Category Accuracy was conducted, but for a lower level in the product hierarchy.

8.2.4 Third Approach: DFU level accuracy within a category

Lastly, 52 weeks of forecast and shipment history were extracted for each DFU within a category. This was only conducted for one forecast horizon, and an appropriate horizon was chosen to match the average length of the supply chain for the associated category. For example, laundry has a short supply chain and a 1-week forecast horizon was used, where-as a longer horizon was used for baby-wipes. For each DFU, the WAPE was calculated across the 52-weeks for both the statistical and independent-net forecasts. In addition, a simple moving average was calculated for each week based on four weeks of data. To match the horizon, the four weeks of historical data used, were taken from the number of weeks back to match the corresponding P&G process forecast horizon. So if the P&G process accuracy analysis used the forecast from four weeks out, the moving average was calculated based on the average of data points from eight to four weeks out.

8.3 Analytical Results and Interpretation

The analytical approach discussed above generated a large amount of data. A few tables and extracts are included below to illustrate the findings and how they led to subsequent analysis.

8.3.1 Category Level : Forecast Index

The category analysis first gave a measure of forecast accuracy for different time horizons. Figure 8.1 indicates an example of an analysis. Each category is disguised for confidentiality. The index is presented where the closer to 100% the more accurate the forecast. If the percentage is over 100%, the shipments exceeded the forecast. If the percentage is less than 100%, the forecast exceeded the shipment. The greater the standard deviation is, the higher the variability is in the forecast accuracy index. Thus a low standard deviation and a high percentage would indicate that the category was consistently over-shipped.

| | Independent Net | | Statistical | |
	Average	Stdev	Average	Stdev
Category 1	104%	32%	149%	43%
Category 2	97%	18%	113%	20%
Category 3	77%	27%	128%	41%
Category 4	95%	23%	102%	20%
Category 5	95%	24%	128%	31%
Category 6	98%	20%	112%	24%
Category 7	86%	48%	100%	49%
Category 8	98%	18%	105%	19%
Category 9	97%	18%	102%	19%
Category 10	96%	16%	103%	17%
Category 11	98%	19%	102%	21%
Category 12	110%	56%	108%	45%
Category 13	96%	21%	198%	17%
Category 14	100%	24%	112%	20%
Category 15	103%	86%	103%	86%
Category 16	110%	22%	117%	23%
Category 17	96%	15%	101%	16%
Category 18	104%	45%	118%	40%
Category 19	92%	28%	96%	28%
Category 20	102%	26%	102%	25%
Category 21	91%	21%	122%	30%
Category 22	95%	19%	105%	21%
Category 23	98%	27%	99%	22%
Category 24	102%	21%	108%	20%
Category 25	57%	54%	60%	58%
Category 26	93%	17%	102%	17%
Category 27	105%	32%	123%	42%
Category 28	92%	21%	131%	29%
Category 29	85%	22%	89%	25%
Category 30	103%	22%	123%	25%
Category 31	100%	18%	109%	23%
Category 32	101%	17%	104%	17%

Figure 8.1 Forecast accuracy index across 52-weeks of data for a particular forecast horizon.

From this analysis, we saw that in almost every category, for every set of forecast horizon (number of weeks out the forecast was generated) the statistical indices were larger than the independent net. This indicates that the statistical forecast generated by the computer model predicted fewer shipments than the independent net; and that the manual process was overall increasing the weekly forecasts. Also noticed was that standard deviation tended to be larger in the statistical model. Looking at the confidence intervals represented by the outlier box plots shown below, the confidence interval for standard deviation is lower and tighter for the independent net forecast.
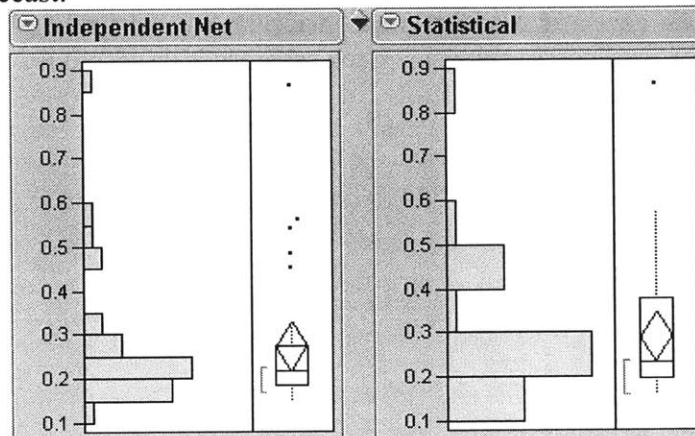
This indicates that the manual adjustments by demand planners were overall producing less variable error.

These trends can more easily be seen by averaging the data across all of the categories as shown in the following figure:

Forecast Accuracy Index
Averaged Across Categories
Weekly Time Buckets - Year Avg

| | Independent Net | | Statistical | |
Fcst Date	Average	Stdev	Average	Stdev
1-week	97%	27%	112%	29%
2-week	97%	27%	112%	29%
3-week	94%	27%	109%	29%
4-week	93%	27%	108%	28%
6-week	93%	26%	108%	28%
8-week	93%	27%	108%	28%
12-week	94%	27%	108%	28%

Figure 8.2: Forecast Accuracy Index for Independent Net versus Statistical Forecasting

This analysis was quite interesting, and clearly shows that the statistical averages were consistently higher than the independent net averages. Again, this shows that the statistical forecasts generated by the computer model tend to under-predict sales; while the adjustments made to create the independent net forecast tend to over-predict sales. This indicates that the statistical forecast has a bias that it always under-predicts sales, perhaps this could be accounted for by the lack of last minute promotion or event data in the statistical forecast. Another possibility is that the planners adjust parameters to force the forecast to under-predict sales

leaving room to add market intelligence. The independent net forecast has an apparent bias to over-predict sales, indicating that the adjustments made tend to be larger than necessary. Reviewing the standard deviation between the two forecasts indicates only a slight improvement in variability in the independent net versus the statistical forecast.

The other interesting finding by this analysis is shown when reviewing the accuracy and standard deviation across the forecast horizons. Standard theory would predict that as the actual date of shipment is approached, the forecast should become more accurate. While we can see this in the independent net forecast as the indices reach closer to 100% as the forecast horizon shortens, looking at the statistical forecast we see different behavior where the forecast accuracy decreases as we get closer to the actual ship date.

These findings impact business planning decisions, but provided limited insight possible efficiency improvements. Another analysis was conducted by drilling into a brand in particular.

8.3.2 Product Level Within Brand : Forecast Index

The first approach was repeated for a particular brand within the laundry business of P&G. Laundry can come in different forms such as liquid, powder, or tablets and therefore different levels of this brand were analyzed to examine the forecast accuracy behavior. Most of the behaviors at the aggregate level were repeated at this smaller increment. The following figure represents a snapshot of this analysis:

1 week Fcst				2 week Fcst				3 week Fcst			
Independent Net		Statistical		Independent Net		Statistical		Independent Net		Statistical	
Average	Stdev	Average	Stdev	Average	Stdev	Average	Stdev	Average	Stdev	Average	Stdev
97%	15%	105%	19%	98%	15%	105%	19%	98%	16%	104%	19%
97%	17%	108%	22%	98%	17%	108%	22%	99%	18%	108%	22%
107%	35%	118%	32%	110%	36%	119%	33%	111%	36%	121%	32%
71%	78%			67%	75%			73%	73%		
83%	48%	110%	59%	95%	63%	112%	60%	105%	58%	130%	70%
14%	33%	15%	37%	14%	34%	16%	37%	15%	35%	16%	37%
15%	37%	15%	36%	15%	36%	15%	35%	15%	36%	15%	35%

Diff Avg	Diff Stdev	Diff Avg	Diff Stdev	Diff Avg	Diff Stdev
8%	4%	7%	4%	6%	3%
12%	5%	10%	5%	9%	4%
11%	-2%	9%	-3%	10%	-4%
27%	11%	17%	-3%	24%	12%
2%	4%	1%	2%	1%	2%
-1%	-1%	0%	-1%	0%	-1%

*** Red cells indicate that the statistical forecast is better than the independent net forecast

Figure 8.3: Forecast Accuracy for Statistical and Independent Net forecasts generated for different product types within a brand

Figure 8.3 demonstrates the same behavior that was displayed at the category level, across the different forecast horizon. This shows again that statistical forecasts tend to under-predict sales, with a couple exceptions; whereas the independent net forecasts tend to over-predict sales, or at least predict higher sales than the statistical forecast generated.

With this smaller subset of data, we can see that there is a large error associated with the last two product lines analyzed. Both of these have indices less than 20%, which indicates that actual sales were much more than anticipated. Perhaps something unusual was occurring in the brand for those product lines. With a few anomalous data points aggregated into the total picture as these are, the summary for the brand may be significantly impacted. This is shown in the following figure.

Forecast Accuracy Index
Determined for LND-Tide & Averaged
Weekly Time Buckets - Year Avg

	Independent Net		Statistical	
Fcst Date	Average	Stdev	Average	Stdev
1-week	69%	38%	79%	34%
2-week	71%	40%	79%	34%
3-week	74%	39%	82%	36%
4-week	74%	39%	85%	34%
6-week	70%	35%	86%	31%
8-week	65%	34%	86%	31%
12-week	70%	38%	86%	30%

Figure 8.4 : Forecast Accuracy aggregated at the product line with a brand level

Figure 8.4 demonstrates that while the majority of the product lines within the brand may be under-forecast (indices above 100%), once the two product lines are included, the entire brand appears to have a gross over-forecast bias, which is created from a small segment of the brand. To better account for the relative size of the contributing products, another analysis was performed using a forecast accuracy metric popular at P&G, weighted-average percent error, or WAPE.

8.3.3 Products Level Within Brand : Forecast WAPE

P&G commonly uses the forecast error metric, WAPE, when evaluating performance. WAPE reviews the absolute difference between the forecast and the actual shipments, and calculates the error as a percent of total volume. Thus the higher the WAPE is, the higher the percent error is and the worse the forecast performance is. The year's worth of statistical and independent net forecast history for the same brand outlined in section 8.3.2 Product Level Within Brand : Forecast Index was evaluated against WAPE.

1 week Fcst		2 week Fcst		3 week Fcst		4 week Fcst	
Indp Net Average	Stat Average	Indp Net Average	Stat Average	Indp Net Average	Stat Average	Indp Net Average	Stat Average
8%	14%	8%	14%	8%	13%	9%	13%
10%	18%	11%	17%	11%	17%	11%	17%
30%	29%	30%	30%	30%	31%	31%	32%
15%	10%	13%	11%	13%	10%	11%	11%
11%	11%	12%	12%	11%	11%	12%	12%

Diff Avg	Diff Avg	Diff Avg	Diff Avg
6%	6%	5%	4%
8%	7%	6%	5%
-1%	0%	1%	1%
-5%	-2%	-3%	-1%
0%	0%	0%	0%

*** Shaded cells indicate that the statistical forecast is better than the independent net forecast

Figure 8.5: Weighted-Average Percent Error for a set of product lines within a brand for P&G

This time, the larger the number is, the worse the error is. WAPE does not indicate bias and whether or not a forecast is above or below the actual shipments. Instead, the conclusion we can draw from this analysis is that the independent net forecasts on a whole appear to be very close or slightly better than the statistical forecasts, with the exception of the product line which is second from the bottom. Here, it appears that the manual adjustments being made to the statistical forecast are creating more error as opposed to if the forecast had just been left alone.

Averaging across the product lines for the brand, the following summary chart was obtained:

Forecast Accuracy WAPE
Determined for LND-TIDE & Averaged
Weekly Time Buckets - Averaged over a year

Fcst Date	Independent Net Average	Statistical Average
1-week	22%	89%
2-week	25%	90%
3-week	24%	94%
4-week	25%	94%
6-week	17%	32%
8-week	20%	32%
12-week	19%	32%

Figure 8.6: Weighted-Average Percent Error averaged across the product lines for a particular brand.

Figure 8.6 demonstrates that the forecasts are generally much better following the adjustments made to the statistical output. The statistical errors appear rather large. Another unusual finding is that the error tends to increase closer to the actual shipment date. Again, it is hard to determine what is contributing to this effect, as perhaps new event only products are added to the forecast close to the ship date.

These analyses provide interesting insights into forecast accuracy. However, global demand planning experts suggested using the metric WAPE aggregated to a business category. This is examined in the next section.

8.3.4 Category : Forecast WAPE

The weighted-average percent error was calculated for each week across an entire category, for each possible forecast horizon. This was repeated for each category. Averaging across the categories gives the following data:

Forecast Accuracy WAPE
Determined for Category & Averaged
Weekly Time Buckets - Averaged over a year

Fcst Date	Independent Net Average	Statistical Average
1-week	20%	38%
2-week	20%	38%
3-week	20%	38%
4-week	21%	38%
6-week	21%	38%
8-week	21%	38%
12-week	24%	38%

Figure 8.7: Weighted-Average Percent Error for different forecast horizons averaged across P&G categories.

Figure 8.7 demonstrates that the manual process executed by the demand planners generates a forecast that has much lower percent error than the statistical output. However, it might also indicate an opportunity to update the statistical models to generate more accurate output.

The WAPE was calculated for each category type of forecast across the period of a year. This also prompted the idea to compare the statistical and independent net forecasts with a simple statistical representation, the moving average. This analysis was conducted for the one-week forecast horizon, as this is where both the independent net and statistical forecasts have the greatest accuracy. The moving average was calculated by averaging four weeks of data leading to the forecast period. The results are depicted in the following figure:
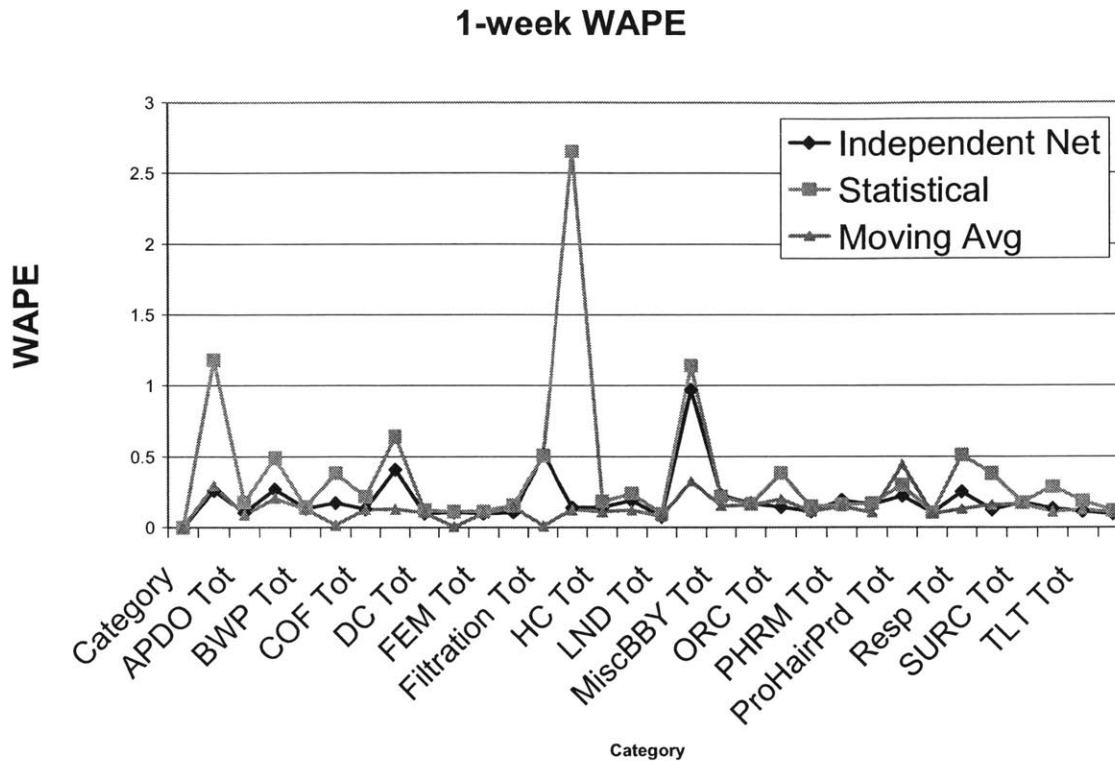
1-week WAPE

Figure 8.8: Weighted-Average Percent Error for three different types of forecasts generated for different business categories. These are determined for a year worth of history, and a one-week forecast horizon.

Figure 8.8 shows an interesting finding. For most categories, the statistical forecast clearly generates the highest error. In fact, there only appears to be one category where the statistical forecast does not generate the most error. The independent net forecast appears to be significantly better than the statistical forecast for many categories, indicating the value of the personal step in the demand planning process. The most interesting finding is the accuracy of the moving average forecast compared with the other two forecasts generated from the P&G demand planning process. Here, we can see that in a number of cases the moving average is better than the independent net forecast, although it does not have the business intelligence that the P&G process supplies. It might appear that replacing the P&G demand planning process with a simple moving average for very short forecast horizons would reduce the effort required to generate these forecasts. However, category level planning at this short interval is not useful to many people in the business. At the short time interval, forecasts are more useful and drive the business at a SKU level. Additionally, the one-week forecast horizon may be too short for some businesses which have longer processes. Therefore, the next analysis seeks to understand better if this trend continues to a lower level, by choosing a set of categories, and evaluating the accuracy at a SKU level for a time horizon in line with the average cycle time for the business.

8.3.5 DFU level accuracy within a category

To delve into the possibility that a moving average forecast might be very accurate relative to current processes, a year's worth of forecast and shipment history for a particular category was extracted at the demand forecast unit, DFU, level. Chapter 4 explained that a DFU represents a SKU, or a set of nearly identical SKUs that are forecast as a group. This is the smallest forecast unit at P&G. For the category, an appropriate forecast horizon was chosen which would be the appropriate length to provide a forecast in the right time frame to be able to have an impact on the supply chain. For each SKU, the WAPE was calculated for a year's worth of data for the statistical, independent net, and moving average forecasts. Below are a series of charts showing the distribution of SKUs based on how the independent net forecast compares to the statistical forecast. These are evaluated for three different P&G categories.
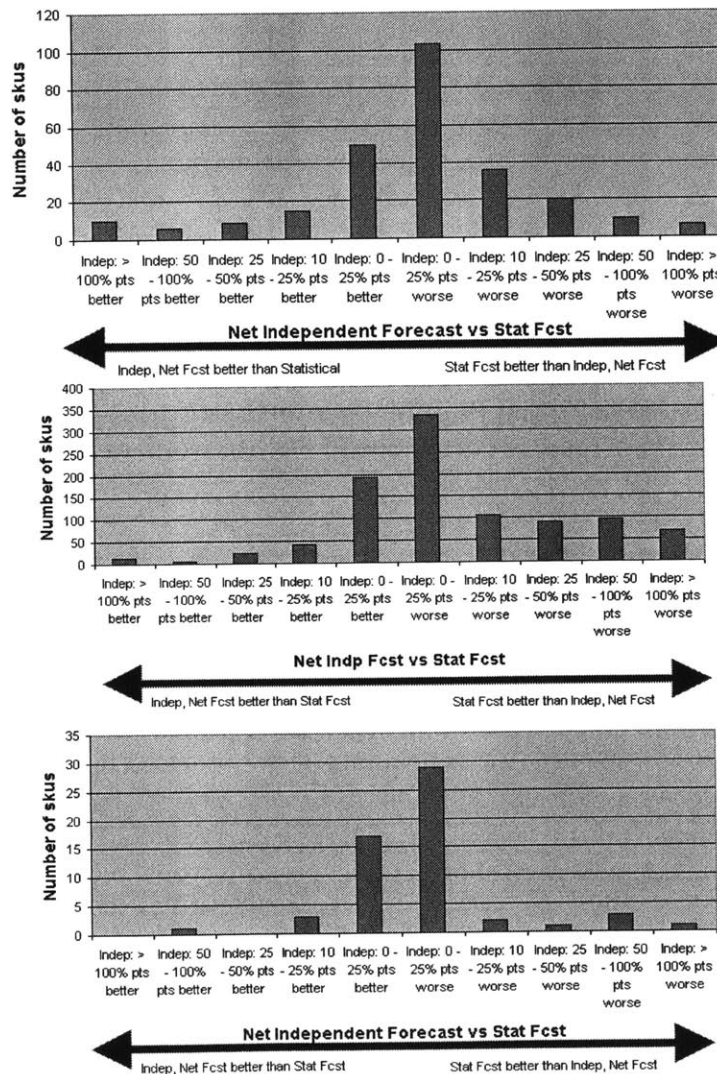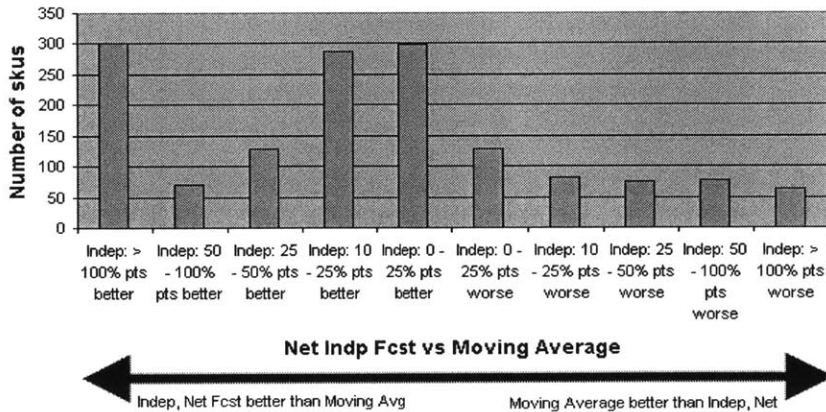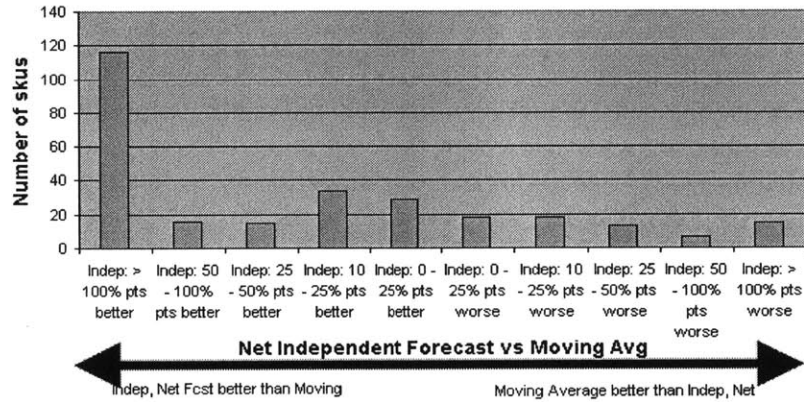

Figure 8.9: Distribution of SKU performance for the statistical versus the independent net forecast for three categories.

Figure 8.9 determines the number of WAPE percentage point improvement between the independent net and statistical forecasts. A certain pattern emerges that the majority of products fall in the range where the independent net and statistical WAPEs are within 25 percentage points of each other. The appearance of SKUs at the left of the graphs indicates that the net independent forecast does much better than the statistical forecast. The numbers of SKUs on the right of the chart show that there are products out there where the adjustments made to the statistical forecast are not improving the situation.

Also of interest, was comparing the independent net forecast with a moving average forecast. The moving average was calculated by averaging four weeks worth of shipments going back from the time to match the forecast horizon. Then the moving average forecast error, measured by WAPE, was then compared to the forecast error for the statistical and independent-net forecasts. The analysis is demonstrated for three categories below:
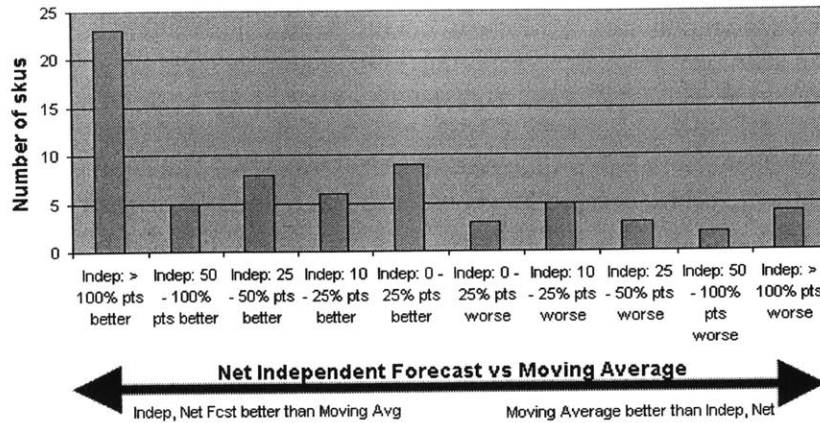
Figure 8.10: Distribution of SKU performance for the independent net versus the moving average forecast.

Figure 8.10 indicates that the independent net forecast significantly outperforms the moving average for a large number of SKUs. However, there are still a relatively large number of SKUs that fall into the category where the moving average outperforms the independent net forecast. To understand this better, a few theories were tested to break the analysis down by different product types. One differentiating factor is a category referred to a CSU, which is a complex ship unit typically produced for a one-time promotion, and not a lasting product. Removing the CSU from the analysis showed the same general trend as shown in the graphs above. The only difference was noticed when looking at the CSUs alone as shown in the figure below:
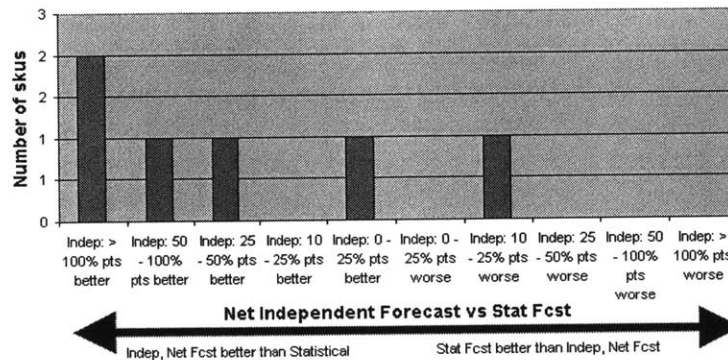
Figure 8.11: CSU comparison of Net Independent Forecast versus Statistical Forecast

Figure 8.11 shows how the CSUs tend to have independent net forecasts which are more accurate than the statistical forecast. More individual attention and planning are required for these one-time use SKUs, which may only be sold for a short period of time.
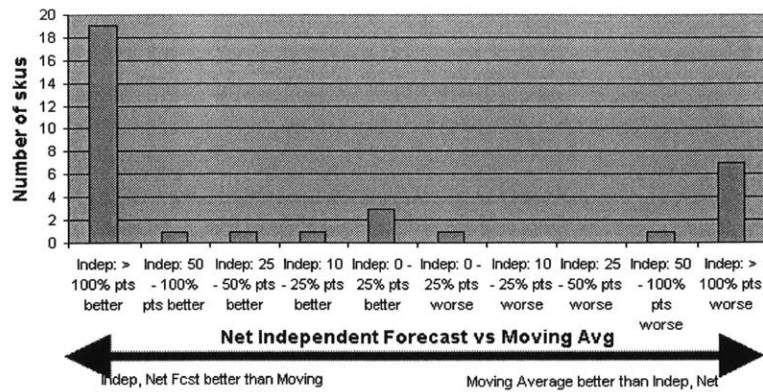
Figure 8.12: CSU comparison of Net Independent Forecast versus Moving Average Forecast

Figure 8.12 shows how the CSUs also tend to have less error independent net forecast, when compared to the moving average forecast. However, there are still a significant number of SKUs that have a moving average forecast which is much better than the independent net. Another differentiating factor considered was the volume associated with each SKU.

To further investigate a reason why different SKUs fall into different categories, each SKU was assigned an "A", "B," or "C" depending on the annual volume of shipments. The "A" SKUs represent approximately 80% of the shipping volume, though are only a smaller fraction of the SKUs. The "B" SKUs are the next 15% of the shipping volume, while the "C" SKUs are the remaining 5% of the shipping volume. This further segregation was applied to the net independent forecast versus the statistical forecast. This can be seen in the figure below for two of the categories.
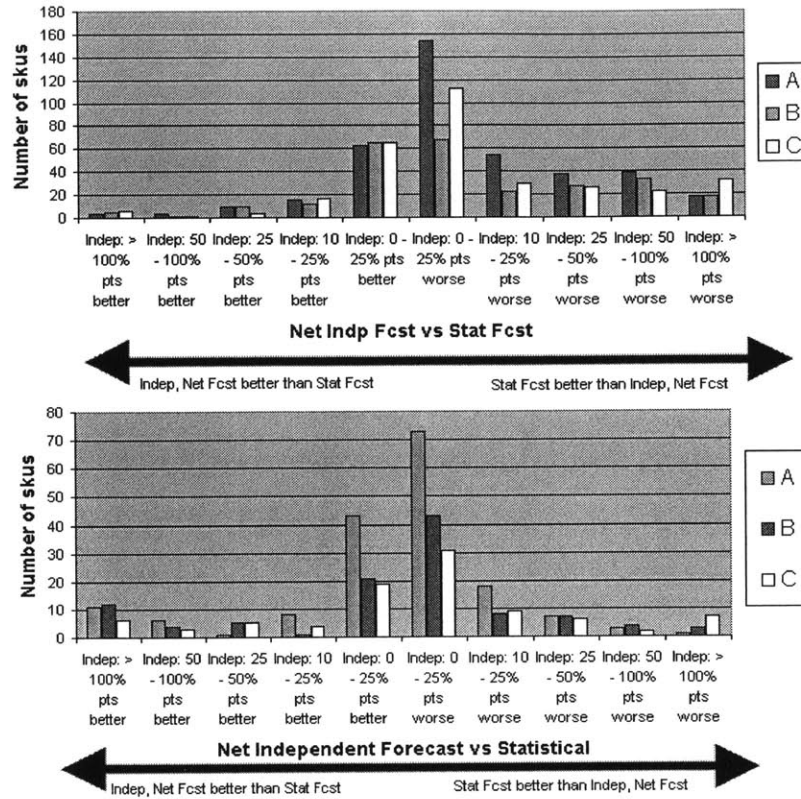
Figure 8.13: ABC analysis for comparing independent net forecast error with statistical forecast error for two categories.

Figure 8.13 shows essentially the same trend as before, without any obvious trend changes due to the category. The "A", "B", and "C" SKUs seem to be dispersed across the charts and no generalizations were able to be made.

The same analysis was conducted on the net independent forecast versus the moving average comparison. This is shown in the following figure:
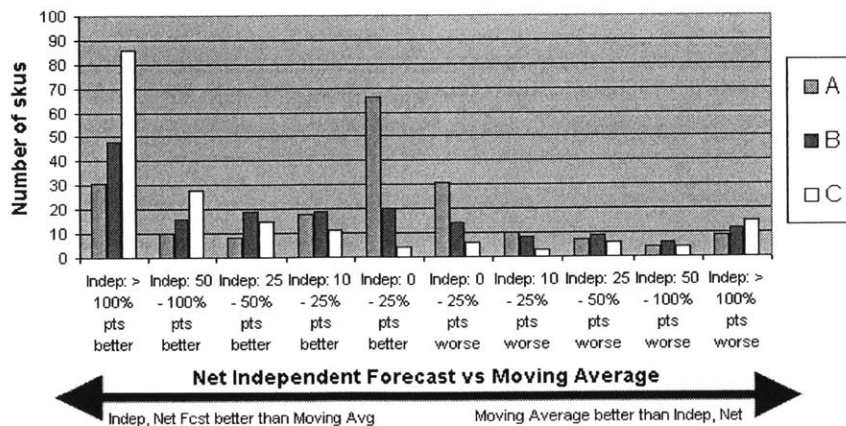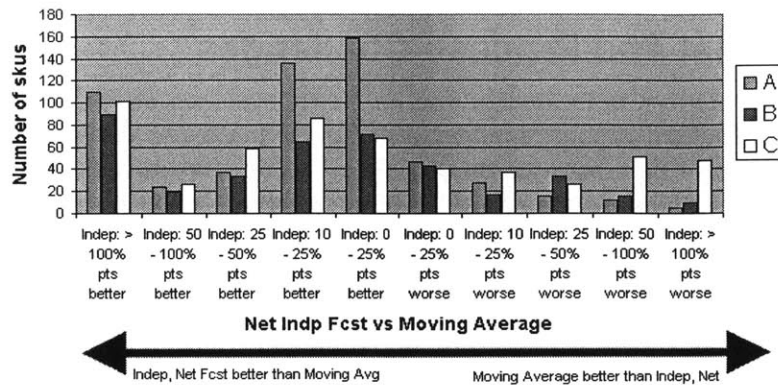
Figure 8.14: Examples of ABC comparison of Net Independent Forecast versus Moving Average Forecast for two different P&G categories

In the top graph, Figure 8.14 shows that the "C" SKUs, which represent a smaller volume, represent a large number of the cases where the independent net forecast is much more accurate than the moving average forecast. Some "C" SKUs have the opposite behavior where the moving average is much better than the independent net. Thus no overall conclusions are possible and the SKUs need to be evaluated on a case-by-case basis. Additionally the "A" SKUs still seem to appear on the side where the moving average forecast has less error than the independent net.

8.4 Summary and Recommendations

Review of this analysis indicates some key conclusions with respect to evaluating forecast efficiency. The level of product hierarchy at which the analysis is conducted can lead to different conclusions. For example, at the category level the independent net forecast appeared to have consistently higher error than the statistical forecast. However, when evaluating forecasts at a SKU level, many SKUs had statistical forecasts that was as good, or even better, than the independent net forecasts. Thus, it is important to understand the objective and potential impacts of an evaluation and choose the appropriate level for the analysis.

Aggregated analysis shows that independent net forecasts provide value and are capable of reducing error over the statistical forecasts. SKU-level analysis identifies opportunities for improvement. For example, simple moving average forecasts seem relatively accurate. I left spreadsheet tools and directions to repeat this analysis as necessary for other categories. These tools generate lists of the SKUs which can be reviewed by people closer to the business operation. The next step would be for a business expert to analyze what efficient SKUs have in common, and whether lessons can be captured and extended from one business to another.

This chapter demonstrates the challenges in performing an efficiency study, yet indicates the opportunities that exist for forecasts generated using less personal interaction to be relatively accurate. So while a previous chapter found that a potential method for improving forecast was not cost effective, redistributing effort between the SKUs may be able to provide efficient improvements.

9 Organizational and Process Challenges

While the thesis has presented an interesting analysis, it is important to consider the context of the study, and the impact of P&G's organization on the findings and ability to impact change. This chapter will review the organization through three different perspectives, or lenses: strategic, political and cultural. Lastly, the chapter will evaluate the work discussed in the thesis within this organizational context.

9.1 Strategic Lens

A review through the strategic lens begins with an assessment of organizational goals and the organizational design to accomplish this strategy. Subsequently, the impact of the strategy on the work conducted for this thesis will be reviewed.

9.1.1 P&G Corporate Strategy

The 2005 annual report explains the company's strategy in the words of the CEO[35]:

> "P&G's strategies remain unchanged. We are continuing to focus on core businesses and on P&G's leading brands, countries, and customers. We are continuing to build P&G leadership in the faster-growing, higher-margin beauty and health care businesses. We are continuing to invest in growth with lower-income consumers in developing markets. I'm confident these remain the right choices for sustained growth."

In essence, these comments point to a focus on building size, maintaining strong brand leadership positions, growing into new markets and building higher margin businesses. These comments do not point to the overarching strategy, but instead are really methods of accomplishing their main goal, which is dominance of the consumer market place.

P&G's short-term operational strategies are aligned around meeting this long term goal of dominating the consumer market globally. Their focus on growing size and dominance leads to increasing market share in the primary categories. Increasing size leads to economies of scale as their business can spread the cost of fixed assets and advertising across more products, and use size to leverage efficiency in relationships with other parties in the channel. These economies of scale lead to lower unit costs, resulting in greater profitability and thus larger budgets. This allows P&G to invest heavily into research to understand consumer needs, to develop and bring new products to market, and to improve marketing to increase awareness and consumer demand. This theory is substantiated by the identification of P&G's core strengths in the annual report as being branding, innovation, go-to-market, and scale. Essentially the first two strengths represent the marketing and product development focus, and the last two represent the efficiencies from

---

[35] P&G Annual Report 2005.

scale, especially in their go-to-market capabilities. This strategy creates a virtuous cycle[36], whereby increasing market share leads to lower costs, and thus higher profits, allowing for investment into research and development and improved marketing, driving further increases in demand. Applying system dynamics to represent this behavior, Vensim ® demonstrates this cycle is shown in the figure below:

```
                    ➤ Market Share ➤

       Consumer                         Economies of
       Demand                              Scale

         ⬆
          \
           \
     Improved Marketing and
        New Product                  Profitability
        Development ➤
```
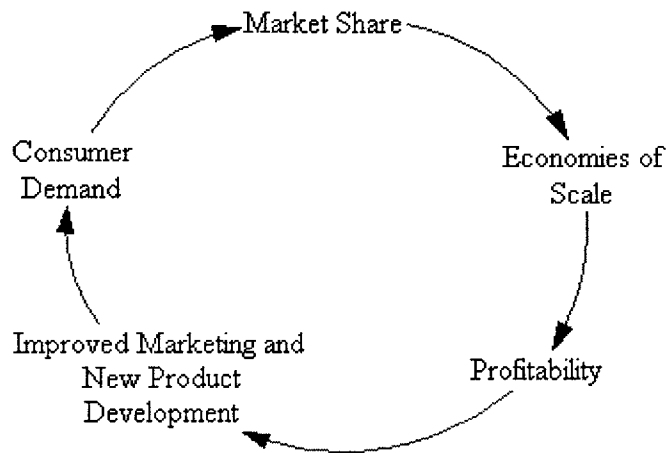
Figure 9.1 Virtuous cycle of dominating the consumer product marketplace

Thus, driven to dominate the consumer product market, Figure 9.1 demonstrates how P&G has become so successful and where it focuses its efforts. An excerpt from the annual report highlights this importance.[37]:

> "[We are focusing on] industry leadership that enables growth. Leadership is important because it creates opportunities to grow, to sustain growth over the long term, and to earn superior returns from consistent growth. P&G aspires to be the leading consumer products company in sales, profitability, market capitalization, shareholder return, and – particularly important – in each of our core strengths. We are creating sustainable leadership advantages in branding, innovation, go-to-market capability, and scale – and we aspire to be the industry best in each area."

P&G thus places strategic importance on marketing and new product development, while also enabling sales and lowering costs from capturing economies of scale, thus funneling money back into increasing demand through marketing and new product development.

Following these strategies, P&G has reorganized to most efficiently operate along these principles. The organizational structure outlined in Chapter 2 is designed to synergize the core activities of a brand – product development, marketing and product supply – globally in GBUs, while capturing the economies of scale across product lines locally in MDOs. See figure below:

---

[36] Sterman, John. Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill, 2000.
[37] P&G Annual Report 2005.
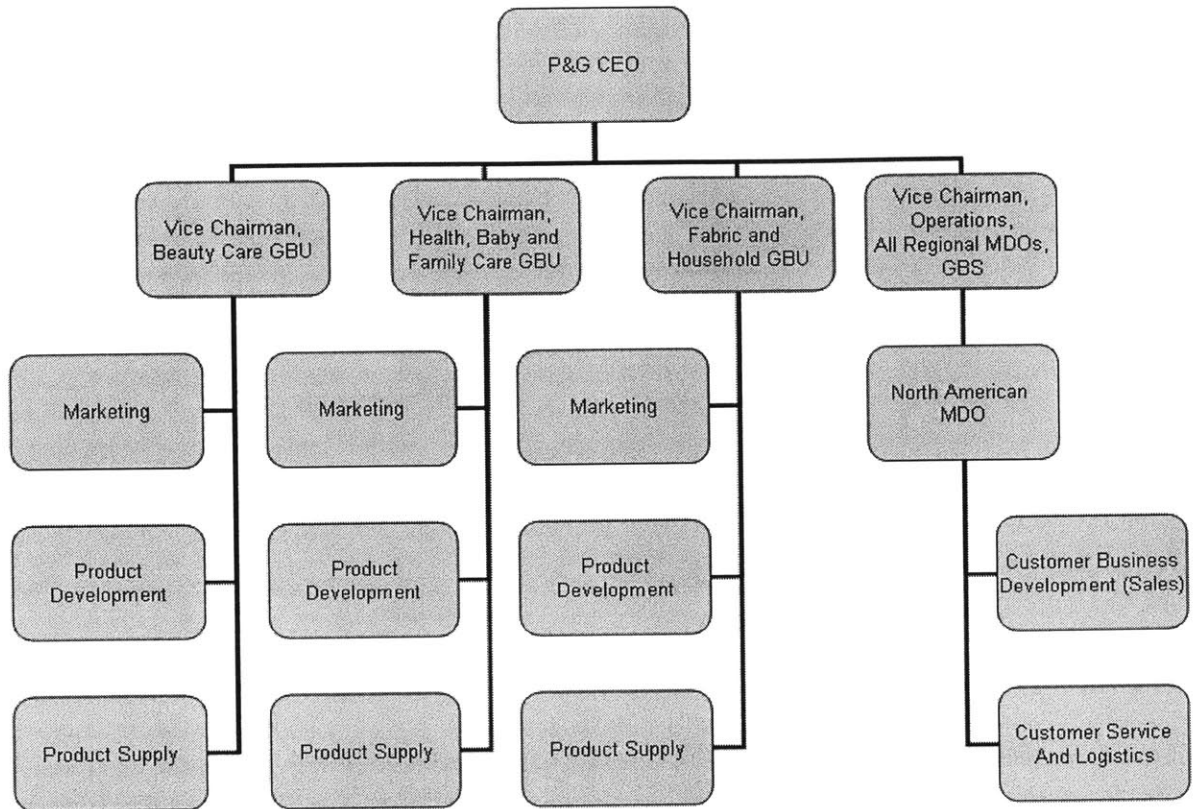
Figure 9.2 P&G Organizational Design

For example, product improvements and brand marketing can be applied globally, thus global brand organizations are more efficient. Local market development organizations are efficient to allow P&G to leverage local market knowledge and relationships with customers on a regional basis. This segregation of responsibility and budgets all the way up to the CEO level allows organizations to focus on these goals with minimal conflict. Essentially, this explains how P&G's corporate strategy has influenced the organizational design.

Another benefit from the organizational design is the ability to quickly integrate other companies into the organization. P&G's growth strategy is operationally achieved not only through organic growth, but also through numerous acquisitions of varying sizes. The organizational design allows for other organizations to be assimilated into P&G easily by either rolling up similar businesses into business units, or by creating new business units for new business ventures. The MDOs are already accustomed to supplying a diverse set of products to customers regionally, and can add the new products to the portfolio efficiently.

9.1.2 Impact of P&G Strategy on Projects

The strategies discussed above provide a context for the environment in which this work was conducted. For example, P&G's strategy of growth through acquisition leads to the importance of developing flexible processes. Chapter 6 discussed the design of a statistical process to set forecast accuracy metrics and create accountability throughout the organization. At the same

time that the design work was underway, the acquisition of Gillette was approved. Thus throughout the design process, flexibility was built into the model to allow for changing numbers of business units, setting metrics with limited history, and updating easily as large changes were anticipated. This also emphasizes the importance for projects to be simple and easy to implement. The forecast trial discussed in Chapter 7 was conducted while the planning for the Gillette acquisition. This could serve as a resource drain for the MDOs which are responsible for sales and distribution of product to the customers regionally, and would be taking on responsibility for Gillette products soon.

At a higher level, the importance of the primary strategy highlighted in Figure 9.1 is also reflected in the project work. Attention is drawn to forecast accuracy as it is cited as the reason for many costs in the business. For example, work by another P&G intern evaluated the financial value of P&G's inventory, and then looked at benchmarking this to other companies. A major cause for inventory is safety stock to account for variability in demand. To improve efficiency and reduce costs, forecast accuracy needs to be improved. Therefore, attention to forecast accuracy stems from P&G's core strategy as forecast accuracy is viewed an opportunity for cost savings.

Another part of P&G's core strategy is leveraging size to capture economies of scale. Hence, P&G creates horizontal process networks and global process organizations which facilitate the standardization of processes globally and the rapid sharing of important lessons and system improvements. While Chapter 7 discusses a trial for one customer, one category, in North America, it was important to keep the perspective of how this would extrapolate to the rest of the company. This necessitated having discussions with global experts, people from different regions and businesses to understand unique situations and how applicable project lessons would be to the rest of the business. For example, while the RR system in theory seemed like a great opportunity, the cycle times and needs of other businesses in P&G decreased the savings potential on total rollout to the company. This importance on leveraging lessons across regions as efficiently as possible led to a time pressure for the forecast accuracy metric rollout discussed in Chapter 6. When a cost-saving initiative was planned, rapid deployment was considered a strategic strength of P&G and demanded from projects.

P&G's size is not only leveraged in organizational efficiencies, but also through improving service to other parties in the supply chain. For example, P&G supplies such a large quantity of products (estimated to be 5% of Wal-Mart's volume) that investments in strengthening these relationships to simplify the supply chain pay for themselves in reduced cost, increased efficiency and reduced out of stocks. This was a direct motivator for the project discussed in Chapter 7. Here, the information sharing between P&G and its customer using Vendor Managed Inventory was used to reduce costs and improve systems for both P&G and the customer distribution center. While this project did not prove successful, it was aligned with the overall strategy of P&G and other ventures in similar areas are likely to be pursued in the future.

Lastly, the organizational design created to enable P&G's goals formally separated the demand and supply organizations. As seen previously in Figure 9.2, the people responsible for the demand forecast and the people who are impacted by it in the supply organization do not meet until the highest levels of the organization. While this design is efficient for the overall

operations of P&G, it leads to projects such as that discussed in Chapter 6 to drive accountability through the system. Thus General Managers who reside in the GBUs will be held accountable for their forecast accuracy, although the forecast is generated by personnel within the regional MDOs.

9.2 Political Lens

As the previous discussion highlighted the overall strategy of P&G and how this impacted the project work, the organizational design and strategy have created a political system which also affects project work. This section will investigate the interests and power of different stakeholders within the organization, and how this impacted the project.

9.2.1 P&G Corporate Politics

Interpersonal relationships are key to successful operation of P&G businesses. This is driven by a number of factors including P&G's "promote from within" mantra and the organizational design. P&G has a strict policy of only hiring at the entry level. As people reach higher levels of the organization, they unanimously have had extensive experience operating within the company. This policy leads to the average employee having a significant history at P&G. With the average years of working at P&G being quite high, it is likely that people will work together at multiple points along their career. Additionally, as relationships build over time, an informal network connects employees across organizations. These relationships are seen as a strategic advantage and even written about in the annual report:

> "In addition, there is an intangible but important benefit that comes from P&G's "promote from within" culture. The men and women working in the GBUs and MDOs often know each other because they've spent their entire careers at P&G. They're focused on the same purpose. They have similar values. They've had similar career experiences. This strengthens their ability to debate issues, make decisions and execute with excellence.[38]"

This excerpt indicates that even though employees may be separated by function within the hierarchy of the organization, these relationships enable collaboration and grease the processes in place. Understanding the importance of relationships, next the relative sources of power will be discussed.

One source of power within the organization is tenure. The "promote from within" policy creates an interesting dynamic within the organization, and that's the importance placed on the number of years an employee has worked for the company. The more years of service an employee has, the more respect they command from their peers. Respect can also be obtained by people with less experience, but this is harder to earn. When teams form, or groups come together for a meeting, introductions always include the 'years of service' at P&G. In many of the meetings attended, the years of service were in the twenty-thirty range.

---

[38] P&G Annual Report 2004.

Another source of power within the organization is the department represented. Anecdotally, many employees commented that P&G is really a "marketing company" and that decisions are primarily driven by marketing. Another comment indicated that "the organization is aligned around marketing and consumer knowledge." This dynamic also plays into the functional organizational structure. Business general managers are groomed through the marketing organization. Additionally, Marketing is the fastest promotion path within the organization, and is referred to as an "up or out" function. Thus, having a marketing person on a team could have an interesting impact on the dynamics. This person has the potential to rise quickly through the organization to a leadership position, thus their relative power on a team might be higher. On the other hand, the rapid turnover in these positions presents another challenge. While Marketing has been used as an example of how power is derived from the function of the organization, this also extends to other functions within P&G. I would hypothesize the GBU and MDO organizations have higher relative power by association than the GBS unit. It appears that business really drives GBS services such as information technology, as opposed to having the information technology drive business change.

Lastly, two other factors appear to contribute to power within the organization: position, and closeness to the customer. Positional authority is deemed from the relative place of people in the corporate structure. While recommendations may be carefully framed, important decisions were left to higher levels in the organization, and respect was always shown for authority. However, people did not seem to wield structural power. Information and respect flowed in both directions. Less importance seemed to be shown to people's relative place in the hierarchy, as opposed to their years of service. The other factor which could be drawn on for power, and input to decision making, was proximity to the customer. With the strategy of leadership brands and high service, people in roles with a direct impact on the final product or sales demanded attention.

The "promote from within" strategy may appear challenged by the strategy of growth by acquisition. For example, in 2004 Gillette had approximately 29,000 employees. Folding this organization into P&G will prove a cultural challenge, as Gillette employees tend to have more diverse backgrounds. Thus the sources of power may be changing with the latest acquisition.

Having reviewed power and relationships within the organization, the impact of these on the project work will now be discussed.

9.2.2 Impact of P&G Politics on Projects

The importance of relationships in the organization, coupled with the demand planning function operating in the MDO, but affecting the GBUs, heightened the importance of working with other people for this research. Fortunately, my supervisor commanded organizational respect due to his long tenure, functional excellence in demand planning, and global leadership role. Thus, some credibility was possible as any analysis would pass through him. Additionally, he identified the key organizational players for the project work, and provided an introduction to enable relationship-building. For example, critical to the success of the forecast accuracy accountability project outlined in chapter 5 was top-down leadership from the executive level, and buy-in from the MDO and GBU organizations. Having no experience with the company, and essentially no organizational power, my ability to drive the project's success hinged on the

shoulders of the key players involved. Throughout the process, input was sought from the global demand planning experts, and especially the North American Demand Planning Manager. Being a relationship-based enterprise, the MDO regional managers had a tight network and communication flowed easily. North America represents half of the business of P&G, and its size and proximity to headquarters leaves this a powerful position within the regions. Therefore, obtaining the input, buy-in and support from the North American lead gave the project credibility earlier with the remote regions.

P&G's organizational politics also deserves a brief discussion on its relevance to the demand planning process. As indicated in Chapter 4, demand planners turn over on average every three years, leaving a high percentage of the demand planners new to the role at any given time. Additionally, input to the demand plan comes from multiple parts of the organization. Marketing, and Consumer Market Knowledge provide input based on the expected effects of advertising and new product design on consumer demand. Meanwhile, account executives from the sales organization work closely with the customers and have foresight to planned promotions and events at a customer level. To justify advertising spending, Marketing needs to point to a high expected impact on demand. Additionally, sales representatives want to ensure that product is available on the shelf for the consumers to purchase. Thus, both of these organizations have an incentive to predict sales which will be higher than actual, although the cost of excess inventory at P&G, a main result from over-predicting sales, would not impact either of their organizations. As these branches have relatively higher power in the organizational framework, it takes a demand planner who has excellent interpersonal skills and the ability to hold their ground in the demand planning process, analyzing the input from different parties in light of their incentives. Incentives also exist to under-predict demand so that the business is seen as exceeding expectation.

9.3 Cultural Lens

While the strategy and politics set an important part of the organizational context for the project, the analysis would not be complete with a review of the strong P&G corporate culture.

9.3.1 P&G Corporate Culture

P&G has a very strong corporate culture, cultivated by its "promote from within" policies and selective hiring process. Carefully controlling the hiring of people into the organization to be a good fit for P&G's culture, and then only promoting people from these entry level positions, serves multiple purposes. One, the hiring process includes a personality test and ensures that people selected are likely to succeed and fit within the P&G culture. Second, the "promote from within" organization starts all new hires at entry level positions; therefore, if someone makes it through the hiring process who happens to be a bad fit for the organization, they are held to the lowest levels and will not likely progress through the organization. Additionally, the "promote from within" culture gives P&G years to mold and shape each new employee, so that by the time people reach managerial level, they have been indoctrinated into the P&G culture. These policies support a strong corporate culture.

While I can not speak to exactly what the P&G personality test is looking for in new employees, a certain consistency was evident in the employees. In my work at P&G I met one of the most outwardly diverse groups of people. P&G managed to attract employees from all different diversity categories – race, ethnicity, citizenship, gender, religious affiliation. I never expected to hear so many accents at the P&G corporate headquarters in Cincinnati, Ohio. However, while so many of the employees came from such different backgrounds, there was a uniformity of personality amongst the people. In general, I found employees to be intelligent, hard-working, loyal, outgoing, personable, friendly, ambitious, and relatively conservative. Across the board, all of the employees I worked with were very driven, putting in long hours to keep ahead of their busy schedules, and outwardly motivated to ensure their project's success. While everyone was quite driven, it was also an environment where all of the employees were very friendly, easy to get along with, and very personable. In essence, P&G employees all tend to have the salesman quality of being able to engage people, tell a good story, and keep the interactions fun. Even when interacting with career statisticians or production workers, who might have the stereotype as being less extroverted, these people were incredibly warm and engaging. Creating a staff of hard-working yet personable employees enables P&G to accomplish significant goals through their relationships, as this type of person is likely to be motivated to push the company ahead, yet be pleasant enough to build positive networks and relationships across the organization.

The value of loyalty deserves an entire section to itself. P&G encourages this loyalty through their "promotion from within" policy, and their excellent benefits, especially their profit sharing plan. Under P&G's profit-sharing plan, originally set up in 1887, US employees receive 5% to 25% of their salary in P&G stock, as their annual retirement savings. Fostering employee loyalty and a sense of ownership in the company, the plan doesn't even require employee contributions.[39] What's more astounding, is that media reports between 87 and 95% of employees' retirement plans are invested in P&G stock.[40] [41] The numbers demonstrate how employees back up their faith in P&G to continue to prosper and deliver significant return with significant investment. This loyalty even survived a tumultuous period under CEO Dirk Jaeger, when, on March 7, 2000, P&G stock fell 30 percent. This crash is evident in the figure below:

---

[39] Nelson, Emily. "Could A Bear Market Hurt Your Retirement?" Wall Street Journal Online. Accessed 3/15/2006. http://www.careerjournal.com/myc/retirement/20000414-nelson.html
[40] Zall, Milton. "Too much stock in one basket?" Today's Chemist February 2002. Vol 11, No. 2, pp 77-78.
[41] Fox News. "Employees Beware: Company-Stock 401(k) Plans Not Wise Investments" Friday January 18, 2002. Accessed 3/15/2006. http://www.foxnews.com/story/0,2933,43353,00.html
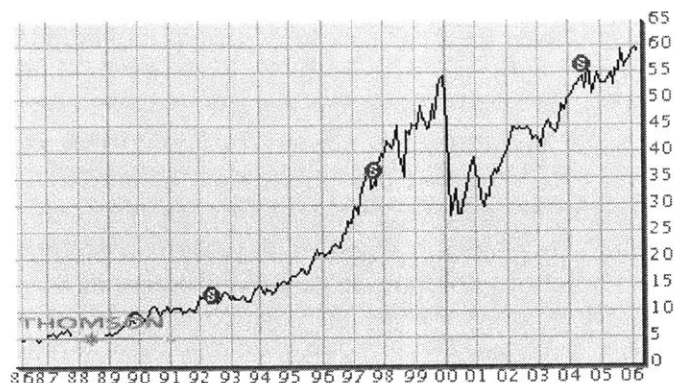
Figure 9.3 P&G Stock History for the past 20 years (the S indicates a stock split) [42]

Former P&G employee, George Fay, cashed in about half his holdings after he left the company in 1995, but still held about 10,000 shares. However, his loyalty is evident in the following quote "P&G got me here and I'm sticking with them," he said.[43]

Employees feel loyal to the company but expect strong leadership to drive success. They tend to develop legends around their CEOs. It is common to walk through the halls of P&G and hear the names "AG" and "Pepper" discussed. John Pepper was a beloved CEO and Chairman of the Board of P&G in the late 1990s. He was replaced by Durk Jager upon retirement. However, employees talk about how cold and unfriendly Mr. Jager appeared, and how he rode P&G stock down through the crash mentioned above. Following this crash, Durk Jager resigned and the current CEO, AG Lafley, was promoted to the job, before some people thought he was ready. Mr. Pepper came out of retirement to chair the board again, and allow Lafley to focus on the CEO job while Pepper ran interference for him with the board. This cemented John Pepper as a hero in employees' eyes. The public certainly approved as the stock has recovered nicely since then. After Pepper retired for the second time, Lafley instilled his own vision of building core brands, designing the current organization, and focusing the organization on winning at the two "moments of truth." This vision resonated with the employees and the "moments of truth" are frequently referenced by employees.

A discussion of P&G culture would not be complete without mentioning the norms of behavior. When it comes to meeting other employees, P&G uses the word, "join-up" to refer to the routine of two colleagues spending an hour just getting to know each other. This is standard practice at P&G, and one employee will send another an email, indicating that work may have them cross paths, and request this informal 'join-up' before beginning any formal working partnership. Norms stretch beyond introductions to communication at P&G. Templates exist for power-point presentations, standard colors are laid out and rules about how the circles and colors can be combined and used. P&G has a language all its own. So many acronyms and phrases are in use, that P&G maintains a web dictionary of P&G terminology.

Lastly, the external community provided many clues to the P&G culture. Employing more than 13,000 employees in Ohio, P&G is a formidable part of the local community. Throughout my

---

[42] P&G website © Thomson Financial
[43] Coolidge, Alexander. "P&G stock floats near century mark" The Cincinnati Post. January 10, 2004.

experience in Cincinnati, I learned the impression outsiders had of the organization. Locals saw employees as very conservative and traditional, and nicknamed them Proctoids. At the same time, there was significant respect for the company, and while some extremists believed some of the behaviors mirrored a cult, overall people felt it was a successful company, a good neighbor and an enviable place to work.

9.3.2 Impact of P&G Culture on Projects

Overall, the strong culture of the organization facilitated my internship. I have had a longstanding, deep respect for P&G and was excited to experiment with becoming part of the organization. My excitement was so visibly evident that people cautioned me about maintaining my objectivity for the quality of research. However, the strong culture set me up as a natural, "outsider on the inside."[44] This enabled my objectivity despite my natural proclivity to fall-in with the norms of the organization.

The existing norms did facilitate operating in the environment. With so many standards in place, there were many presentation templates to copy from, clear expectations about the length, number of bullets/slide that management would expect, etc. While I had to get used to P&G vocabulary, once I was up to speed, it was very easy to be productive.

The join–up process facilitated relationship building early in my internship, which was the key to the projects' success. My first day of work at P&G, I was given a list of over twenty names of appropriate people that I should 'join-up' with. The personality of P&G ensured that these people were all very open, eager to talk, and very helpful throughout the course of the project. The skills and personality of people involved made it very easy to get work done.

9.4 Evaluation of Project Outcomes

In summary, the context of P&G provided factors which enabled project success. The strong relationships between employees, and philosophy of hiring people who are friendly and intelligent facilitated work. People were willing to support and help out where needed. While some challenges from the organizational complexity made the statistical process more complex for rolling out metrics, they were offset by the other factors.

These also highlighted a few organizational areas of importance to keep in consideration when evaluating forecast accuracy at P&G. News travels fast through these organizations. Once a prototype was shown to one of the regional managers, demand rose from the other regions. In rolling out the new metric system, it was important to have the process tested and ready to deploy early, and to keep key stakeholders updated. This also facilitated getting the new process up and running quickly. While the project findings seemed readily accepted, this was the case due to the relationships and support of key organizational players. Lastly, projects such as the accountability metric setting will continue after the internship, and their long term success is supported by the key organizational players who supported and drove the change.

---

[44] Klein, Jan. True Change: How Outsiders on the Inside Get Things Done in Organizations Jossey Bass. 2004.

10 Leadership Lessons

Having the context of the organization highlighted in Chapter 9, reflection on the organizational context also revealed some personal leadership lessons gathered during the analysis.

10.1 Metric Setting

Creating metrics for which organizations feel accountable can be challenging. Several factors appeared to be quite important in the implementation of the projects. Although the statistics for generating appropriate metrics are important, maintaining a soft perspective and obtaining support behind a deployment are also critical. It was clear that metrics perceived as arbitrary, though numerically equal for all regions, were not supported or adhered to. The organization believed that other factors intrinsically impacted accuracy and should be considered when evaluating performance or setting metrics. Therefore, it is important to work within, or at least understand, the organization's mental model to create functional metrics.

Another important consideration was balancing the priorities of people responsible for delivering the objective. For example, discussions with various demand planners indicated that while they were responsible for generating the forecast, they would have strong pressure to keep a forecast high even as business began to slip, since adjusting the forecast downward would indicate that the business was slipping and would draw unwanted attention to the business manager, and might limit the resources available, often justified on growth forecasts.

Lastly, reiterating points made earlier in the paper, an effective goal needs to be stretch to create motivation yet reasonable, so not to de-motivate. The following table demonstrates some of the impacts on choosing a metric:

	Global Error Tolerance (+/-)		
	Low (tight) ← --- → High (loose)		
Finance (preference for smaller variability)	+		-
Product Supply (inventory savings)	+		-
CBD (effort required to improve forecasting)	-		+
Cost (additional cost required to meet tolerance)	-		+
Motivation and Business Buy-In (feasible but stretch)	-	+	-
Realistic Target (probable to meet business requirement)	-		+

Figure 10.1 Impact on choosing a forecast error tolerance.

Figure 10.1 shows directionally how different business areas and factors relate to potential different tolerances. Finance prefers smaller forecast error, and thus a tighter tolerance, as long

as this can be met. Product supply also favors a tighter forecast tolerance as this reduces the inventory carried. CBD (consumer business development, a multi-functioning organization which interfaces with the customer) prefers looser tolerances for meeting sales commitments. If the company were to choose a tighter tolerance, then expectations would be set to invest in programs to improve capability. Thus a tighter tolerance would have an additional cost to P&G in terms of the investment required to increase capability. Motivation and business buy-in require a feasible but stretch target. If the target is too easy the organization will not work to meet it. If the target is too difficult, it will be ignored as it will be viewed as unachievable. Thus the range needs to strike a balance between feasible, yet stretch to motivate personnel. Lastly, business management requests a metric that they can expect to be delivered. This summarizes considerations that need to be balanced in metric setting.

10.2 Consistency

When making change within an organization, consistency is critically important as people are inclined to ignore requests or changes if they do not feel that they be lasting, or if they are in a state of flux and are not yet settled. It is very common for people to wait to see how something 'settles out' before responding. Therefore, were we to change the target ranges for forecast accuracy too frequently, people would cease to pay attention to them as they could argue that they were always changing. We ran into a few issues with this as the initial set of data used to perform the statistical analysis was based on June-May data, which is one month off from the fiscal year cycle of July-June. This occurred since the analysis was begun before June data was obtained. Expectations were for the statistical process to be complete before June data was obtained, however this did not prove to be the case. When the question arose whether the global metrics and tolerances should be re-generated with data that was one month more recent, this concept of consistency was raised. Review of the changes to data that this would produce indicated that it would be in the decimal places for the small percentage numbers, and the team decided not to update the tolerances that had already been presented to senior management.

Consistency came into question again when developing the statistical process to determine tolerance ranges, and the use of probabilities to determine capability. As discussed in section 6.4.6 Standard Deviation there were two options for standard deviation, and initially the "short-term" standard deviation was chosen as this appeared to be more relevant for our predictive needs. However, in reviewing the output and process with the North American Demand Planning Manager, he noticed that the probability calculations did not appear to correlate with the historical data displayed and none of the statistical explanations resonated with the data. It became evident that the process should change and use the common definition of standard deviation. So although the statistical process changed, the original tolerances calculated were not updated as they had already been deployed throughout the organization. This produced some questions later surrounding the probability number discrepancies between the original rollout, and subsequent analysis, but these were handled as simply as possible.

10.3 Balancing Stakeholder Needs

The accountability project was a practical lesson in understanding and balancing stakeholder needs. Each of the parties involved had different priorities, and recognizing and accounting for these issues allowed the project to proceed without significant delay. For example, some regions were more concerned with monthly tolerance ranges as opposed to the corporate focus on rolling quarterly tolerances. Some regional leads wanted to know the conversion from quarterly tolerances to monthly tolerances. Others took the top-down metric at face value and were not interested in understanding a monthly conversion. When these issues arose at the global teleconference, they were discussed long enough to be understood. Subsequently, the interested parties were identified and separate meetings were setup to be conducted offline from the global group. This accounted for everyone's needs without slowing down the group's progress.

Another challenge in the accountability project was striking the balance between having a statistically perfect model that would not present future issues and having the model rolled out in a timely manner for the fiscal year with a consistent process that was easily bought into. The business was persistent that they needed the metric immediately. Applying a metric after the year had begun was far from optimal. Meanwhile, the statistical drive for perfection was causing model development to take much longer than anticipated. Finally, the needs of the business for speed and consistency overcame the desire for perfection and a less than perfect was deployed.

11 Conclusions

The final chapter will summarize the key lessons and applications stemming from this thesis research. It presents the author's opinion on the P&G demand planning process and key lessons which can be applied to other areas, or companies based on this research.

11.1 P&G Demand Planning Process Comments

P&G has implemented a world class business process in their demand planning organization. Many factors point to the strengths of the demand planning organization and display evidence of its success. Consistency across businesses and regions allows for speed in introducing process improvements globally. A team of global experts constantly innovates to ensure that the best processes for P&G are in place and to plan new advances for the future. The most common discussed impacts of forecast accuracy include inventory and customer service. Benchmarking analysis conducted by another intern indicated that P&G performs very well in these areas, indicating that it would be logical for their forecast process to additionally perform well relative to competition.

The downside of having a common global process for demand planning is that it may not be optimized for local requirements or business needs. Analysis at the SKU level indicated that using a moving average as a benchmark forecast, some SKUs were much better than the moving average, while others were much worse. This seems to indicate that different levels of attention would produce a more level result.

Additionally, another concern is that with such a developed demand planning process the opportunities for further improvement may be costly. In fact, improvements achieved may not justify the additional cost to improve the processes. This was demonstrated in the RR project analysis.

11.2 Key Applications

My thesis research led to a few key lessons which can be applied to other areas. Review of the importance of forecast accuracy for P&G indicated that in general, forecasts can have a major impact on many areas of the business. Thus, forecasts should still be regarded as critically important to other organizations. Even companies which produce to order need to have an idea of forecast demand to be able to operate effectively. These companies will still need to maintain relationships with their suppliers, ensure they have adequate capital assets to provide sufficient capacity, and provide accurate financial projections to Wall Street to be viewed successful.

While this research demonstrated the importance of forecast accuracy, it also revealed how challenging demand planning may be in an organization as the inputs and outputs from the process come from many different parts of the company. The motivations behind all the parties providing input data to the forecast decision can have a major impact on accuracy. Additionally,

the costs of forecast error are widespread throughout the organization. However, the cost of demand planning may not be aligned with the cost of the output, thus organizational challenges can arise quickly.

A number of general lessons applying to evaluating forecast accuracy were also evident. Evaluating forecast accuracy can be challenging as many companies and businesses have different processes for measuring accuracy. Even P&G and Gillette had very different metrics. Therefore, it is important to consider what is appropriate when determining how to measure accuracy. For example, choosing appropriate time buckets, forecast horizons, and deciding how to incorporate bias into the error measurement are all important decisions. In the course of this research, measurement techniques were matched up with the largest expected impact of forecast accuracy. As the RR project was expected to provide inventory savings, the forecast accuracy term in the algorithm to calculate inventory was the appropriate metric for accuracy.

When implementing metrics, or even performing an analysis on an organization, the analysis itself can drive behavior. Therefore, considering the interactions in the system is also important when designing a forecast accuracy analysis. For example, the three-month rolling data for one of P&G's geographic regions showed consistent fluctuation from over-predicting to under-predicting demand. As these are three-month rolling averages, it was clear that the one month averages must be largely over- or under- the forecast to be able to swing the data points so widely each month. This indicates that the one-month forecasts were highly likely to be much more erroneous than the three-month forecast to ensure that bias is eliminated. So while this region may perform well according to the three-month forecasts, the one-month forecasts are abnormally poor.

A few themes seemed to run through multiple projects. Manual adjustments to statistical or computer generated forecasts are not always efficient. In both the RR project and efficiency evaluation, evidence was shown that cases exist where forecasts were better without manual adjustment. Additionally, these projects indicated that improvements in forecast accuracy may not be easy and the point of diminishing returns may have been met as it becomes prohibitively costly to achieve significant improvement. The use of much more granular and detailed data did not produce significant accuracy improvements. Additionally, events seem to be the major contributor to forecast error, and can distort standard principles. For example, larger shipments do not show a strong correlation to lower error as one would expect from aggregation. Thus with the concern about reaching diminishing returns, focus on improving event forecasts is the area demanding attention as improving turn forecasts will have negligible impact until event forecasts are improved.

Lastly, the concept of taking a global or corporate target and breaking it down to component targets for individual units based on statistics can be applied to areas other than demand planning. This process seems novel and could be extrapolated beyond forecasting.

12 References

Ailawadi, Kusum L., Donald R. Lehmann, and Scott A. Neslin. "Market Response to a Major Policy Change in the Marketing Mix: Learning from Procter & Gamble's Value Pricing Strategy" Journal of Marketing, 65, 44-61, 2001.

Anupindi, Ravi, Sunil Chopra, Sudhakar Deshmukh, Jan A. Van Mieghem, Eitan Zemel. Managing Business Flows. Prentice Hall, 1999.

Bauman, Mark P and Kenneth W. Shaw. "Stock option compensation and the likelihood of meeting analysts' quarterly earnings targets." Review of Quantitative Finance and Accounting. (2006) 26: 301–319

Cattini, Kyle and Warren Hausman. "Why Are Forecast Updates Often Disappointing?" Manufacturing & Service Operations Management Vol 2, No. 2. Spring 2000 pp 119-127.

Cheung, Christine. "A Short-range Forecasting and Inventory Strategy for New Product Launches." MIT Thesis. 2005.

Clark, Richard A. "Demand Planning Overview: Key Elements of P&G's Global Demand Planning Process." Internal P&G Documentation. December 9, 2004.

Coolidge, Alexander. "P&G stock floats near century mark" The Cincinnati Post. January 10, 2004.

Copeland, Tom, Aaron Dolgoff, and Alberto Moel. "The Role of Expectations in Explaining the Cross-Section of Stock Returns" Review of Accounting Studies. 9, 149–188, 2004

Crandall, Richard. "Demanding Disciple" APICS May 2005 p. 16 – 18

Fox News. "Employees Beware: Company-Stock 401(k) Plans Not Wise Investments" Friday January 18, 2002. Accessed 3/15/2006. http://www.foxnews.com/story/0,2933,43353,00.html

Gilliland, Michael, "Is Forecasting a Waste of Time?," Supply Chain Management Review. July/Aug. 2002, Vol. 21, No. 2.

Gleason, C. and C. Lee. Analyst Forecast Revisions and Market Price Discovery. The Accounting Review 2003: 78, 193–225.

Grean Michael and Michael Shaw. "Supply-Chain Integration through Information Sharing: Channel Partnership between Wal-Mart and Procter & Gamble"

Han, Bong H, David Manry, Wayne Shaw. "Improving the Precision of Analysts' Earnings Forecasts by Adjusting for Predictable Bias" Review of Quantitative Finance and Accounting, 17: 81–98, 2001.

Harwell, Jack and Michael Gilliland. "Applying Forecast Value Added at Radio Shack." Conference Presentation at the Institute of Business Forecasting Supply Chain Forecasting Conference, Phoenix, AZ March 1, 2005.

JMP Stat Graph Guide. Version 5.1

Klein, Jan. True Change: How Outsiders on the Inside Get Things Done in Organizations Jossey Bass. 2004

Kremblewski, Mark. "Supply Chain Forecast Error: Global Standard Definition" P&G Internal Documentation. March 2005.

Lloyd, Mary Ellen. "Procter & Gamble Down After Tempering E Europe, Asia Outlook" Dow Jones Newswire. March 14, 2006.

Lim, T. Rationality and Analyst Forecast Bias. Journal of Finance 2000: 55, 369–385.

Lopezmonjardin, Alejandro. "RR Leadsite Results: Productivity Tracking" P&G Internal Documentation. 10/10/2005.

Matsumoto, D. "Management's Incentives to Avoid Negative Earnings Surprises." Accounting Review 2002: 77, 483–514.

Murphy, Jean V. "More Accurate Forecasts Give Sales, Operations Planning New Life at Sara Lee Unit" Global Logistics & Supply Chain Strategies. Keller International Publishing. March 2003.

Nelson, Emily. "Could A Bear Market Hurt Your Retirement?" Wall Street Journal Online. Accessed 3/15/2006. http://www.careerjournal.com/myc/retirement/20000414-nelson.html

Ouliaris, Sam. "The Art of Forecasting" NUS Business School & Monetary Authority of Singapore, Accessed 3/1/2006
http://www.ess.org.sg/The%20Art%20of%20Forecasting.ppt#256,1,The Art of Forecasting,.

P&G Annual Report 2004.

P&G Annual Report 2005.

P&G, "A Company History", 2005.

P&G website, Investor Information, Stock Quotes and Chart, © Thomson Financial

Park, Chul W. and Earl K. Stice. "Analyst Forecasting Ability and the Stock Price Reaction to Forecast Revisions" Review of Accounting Studies. 5, 259–272 (2000)

Rosen, Oscar. "Forecast Error Allocation and Tracking" P&G Internal Documentation. 6/13/2005.

Sheffi,,Yossi. "A Demand for Steady Supply" Financial Times. August 22, 2005.

Silver, Edward A., David F Pyke, and Rein Peterson. Inventory Managemnet and Production Planning and Scheduling. 3rd Ed., John Wiley & Sons, 1998.

Simchi-Levi, David, Philip Kaminsky, and Edith Simchi-Levi. Designing & Managing the Supply Chain. 2nd Ed. McGraw-Hill, 2003.

Sterman, John. Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill, 2000.

Van de Velde, Wim. "Finished Product Inventory Model Training." P&G Internal Documentation. March 2006

Yuhas, Jim. Internal P&G Documentation. North American Demand Planning. Modified for confidentiality.

Zall, Milton. "Too much stock in one basket?" Today's Chemist February 2002. Vol 11, No. 2, pp 77-78.