Improving the Consumer Demand Forecast to Generate More Accurate Suggested Orders at the Store-Item Level

By

Susan D. Bankston

Bachelor of Science, Mechanical Engineering, Texas A&M University, 2001

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

Master of Business Administration

AND

Master of Science in Mechanical Engineering

In conjunction with the Leaders for Manufacturing Program at the

Massachusetts Institute of Technology

June 2008

Signature of Author______________________________
Department of Mechanical Engineering and
MIT Sloan School of Management
May 9, 2008

Certified by____________________
David Simchi-Levi, Thesis Supervisor
Professor of Civil & Environmental Engineering and ESD

Certified by____________________
Roy Welsch, Thesis Supervisor
Professor of Statistics & Management Science, Director CCREMS

Certified by____________________
David E. Hardt, Thesis Reader
Professor of Mechanical Engineering and ESD

Accepted by____________________
Lallit Anand, Professor of Mechanical Engineering
Chair, Department Committee on Graduate Students

Accepted by____________________
Debbie Berechman
Executive Director of MBA Program, MIT Sloan School of Management

This page has been intentionally left blank

Improving the Consumer Demand Forecast to Generate More Accurate Suggested Orders at the Store-Item Level

By
Susan D. Bankston

Submitted to the MIT Sloan School of Management and the
Mechanical Engineering Department on May 9, 2008 in Partial Fulfillment of the
Requirements for the Degrees of Master of Business Administration and
Master of Science in Mechanical Engineering

ABSTRACT

One of the biggest opportunities for this consumer goods company today is reducing retail stock-outs at its Direct Store Delivery (DSD) customers via pre-selling, which represents approximately 70% of the company's total sales volume. But reducing retail stock-outs is becoming constantly more challenging with an ever-burgeoning number of SKUs due to new product introductions and packaging innovations. The main tool this consumer goods company uses to combat retail stock-outs is the pre-sell handheld, which the company provides to all field sales reps. The handheld runs proprietary software developed by this consumer goods company that creates suggested orders based on a number of factors including:

- Baseline forecast (specific to store-item combination)
- Seasonality effects (i.e., higher demand for products during particular seasons)
- Promotional effects (i.e., lift created from sale prices)
- Presence of in-store displays (i.e., more space for product than just shelf space)
- Weekday effects (i.e., selling more on weekends when most people shop)
- Holiday effects (i.e., higher demand for products at holidays)
- Inventory levels on the shelves and in the back room
- In-transit orders (i.e., orders that may already be on their way to the customer)

The more accurate that the suggested orders are, the fewer retail stock-outs will occur. This project seeks to increase the accuracy of the consumer demand forecast, and ultimately the suggested orders, by improving the baseline forecast and accounting for the effect of cannibalization on demand.

Thesis Supervisor: David Simchi-Levi
Title: Professor of Civil & Environmental Engineering and ESD

Thesis Supervisor: Roy Welsch
Title: Professor of Statistics & Management Science, Director CCREMS

Thesis Reader: David E. Hardt
Title: Professor of Mechanical Engineering and ESD

This page has been intentionally left blank

Acknowledgments

I would like to thank the following people:

The Leaders for Manufacturing Program
for the opportunity to be part of this incredible experience.

My thesis supervisors, David Simchi-Levi and Roy Welsch,
for their guidance and support throughout this project.

The Supply Chain Selling Systems group at the consumer goods company
for sponsoring this project in conjunction with an LFM internship.

My fellow LFM and MIT Sloan classmates
for making this two of the most fun and memorable years of my life.

Todd Cooper, LFM '98,
for his mentorship and for introducing me to the LFM Program.

My extended family that lives in the greater Boston area, especially my grandparents,
for providing me with a local support network that included many delicious home-cooked meals.

Finally, a special thanks to my immediate family
for the constant love, support, and guidance they have provided throughout my life.

This page has been intentionally left blank

Table of Contents

This page has been intentionally left blank

Table of Figures

Table of Tables

This page has been intentionally left blank

1. Introduction

This global consumer goods company sells high-velocity products and competes on the basis of sales, customer service, merchandising, and operations. The company's Supply Chain Selling Systems group supports the tenet of providing superior customer service through operations. The group is responsible for servicing the Direct Store Delivery (DSD) customers via pre-selling, which represents approximately 70% of the company's total sales volume. Most consumer goods companies ship their products from their manufacturing warehouses to their customers' warehouses, which then distribute the product to the customers' retail outlets. However, in the DSD model that comprises the majority of this consumer goods company's sales, it ships products from its manufacturing warehouses directly to its approximately 500,000 customer endpoints. Besides delivering the product to the customer's back room, the company also employs merchandisers who are responsible for transferring product from the back room to the customer's shelves and for building displays in conjunction with manufacturer and customer promotions.

The motivation behind providing such excellent customer service is to ensure that the company continues to maintain control over writing customer orders rather than customers writing orders themselves. The company believes that its ability to write accurate orders is better than any of its customers would be able to accomplish independently. This proved to be the case for at least one customer who attempted to write its own orders for a trial period and reversed course due to poor results. Merchandising is a natural extension of this control, as it enables the company to maintain control over how its products are presented in its customers' stores.

The combination of accurate orders and excellent merchandising should prevent stock-outs, which will ensure that this consumer goods company maximizes not only its own sales but also its customers' sales. Each time the product is not on the shelf when the consumer wants to buy it, the company risks losing a sale. Although the consumer may substitute with an alternative from within the brand's family (e.g., different flavors or package types), the danger is that the consumer might substitute with a competitor's product. Consequently, retail stock-outs make the company vulnerable to losing a consumer's loyalty to its brand.

Writing an accurate order means minimizing the customer's back room inventory while simultaneously keeping enough inventory on the shelves for every product at every customer store (approximately 12 million combinations!) to avoid stock-outs. The company estimates that its retail stock-out rate is currently about eight percent, and some of its customers are complaining about stock-outs. Furthermore, some of the company's customers are actually monitoring stock-outs on the shelves using their own handheld scanners as frequently as twice a day.

The main tool this consumer goods company uses to combat retail stock-outs is the pre-sell handheld, which the company provides to all field sales representatives. Based on inventory levels (on the shelves and in the back room) and promotional information input by the sales representative, as well as other factors such as holidays and seasonality, the handhelds run code that calculates a suggested quantity for each store-item combination. Requirements were developed by the company and coded by consultants retained in 2003, and the company has made incremental improvements to the code since then.

All calculations start with a number known as the baseline forecast, which is the quantity that would sell if there were no other factors involved. Other factors are subsequently layered onto the baseline forecast including:

- Seasonality effects (i.e., higher demand for products during particular seasons)
- Promotional effects (i.e., lift created from sale prices)
- Presence of in-store displays (i.e., more space for product than just shelf space)
- Weekday effects (i.e., selling more on weekends when most people shop)
- Holiday effects (i.e., higher demand for products at holidays)
- In-transit orders (i.e., orders that may already be on their way to the customer)

The more accurate that the suggested orders are, the fewer retail stock-outs will occur. The goal of this project was to help the company mitigate its stock-out rate by improving the consumer demand forecast, and ultimately the suggested orders, at the store-item level. This was accomplished by analyzing historical delivery and Point-of-Sale (POS) data. The delivery data is based on company records, while the POS data is customer-provided transaction data based on bar code scans made by the POS systems used at checkouts. The consumer demand forecast was analyzed in two phases during the course of this project. The first phase consisted of improving the baseline demand forecast at the store-item level. The second phase of the project involved modeling the effect of cannibalization on demand for high volume products during promotions and new product introductions.

2. Literature Review

Much has been written on the subject of demand forecasting. Typically historical sales data is readily accessible and can be used to predict future sales, particularly for manufacturing companies who also sell the product that they produce. In the case of this consumer goods company, it sells its products through customers to the end consumer, so its visibility of sales data is limited by its customers' willingness to share information. In general, customers often fear that sharing too much information with their suppliers might expose them to competitive threats, but in reality such collaboration usually generates mutually beneficial results. In the case of this consumer goods company, the sharing of POS data is in the best interest of the both the company and its customers as both stand to gain increased sales volume through reduced stock-outs and potentially reduced inventory levels throughout the system (e.g., the company's warehouses and the back rooms of its customers).

The lack of sales data has required this consumer goods company to be innovative relative to other companies in its approach to forecasting. The company receives POS data from a limited number of customers – approximately 0.82% of its customers. Fortunately, these five customers are large national chains that comprise approximately 14% of the company's sales volume, so their data provides a representative sample of customer data. The company's approach is to use the POS data that it does have to develop forecasting methodologies which it can then apply to the customers for which it does not have POS data. In addition, the company leverages the POS data to create customized forecasts for those customers who provide it, so they benefit from increasingly accurate forecasts.

This project was part of this consumer goods company's innovative approach to forecasting, and the baseline work described in this document provided a methodology that will be likely prove to be instrumental to the company's baseline forecasting. The company's dilemma was that it knew the rate at which product entered the back of *all* stores (via its direct store deliveries to customer stores), and it knew the rate at which product left the front of *some* stores (via customer-provided POS data). These rates are much different due to variability in supply and demand. Up until this project began, the company was creating its baseline forecast mostly on supply information (historical delivery data) because it did not have a way to relate supply to demand. (The company had created custom thresholds based on POS data to eliminate some of the outlying delivery data that it was regressing to create its forecast.)

The baseline work documented here provided a missing link for the company by relating supply to demand. A methodology was developed using historical delivery and POS data to create POS-like data from delivery data. The company can apply this methodology to create POS-like data for customers for which it does not have POS data using only its own internal delivery data. Although this methodology can and should continue to be refined, it hopefully has provided a breakthrough for companies who are trying to forecast demand in the absence of sales data.

This page has been intentionally left blank

3. Data Integrity

The importance of data integrity, particularly as it related to this project, merits an upfront discussion in this document. Theoretically, what the consumer goods company reports that it delivered to the customer's back room should match what the customer's POS data indicates was sold and went out the front door. However, this is not always the case. One potential cause of this mismatch is scanner error. The typical scenario under which this occurs is when the consumer has several different flavors of the same package type in his or her cart, and the clerk simply scans one and multiplies by the quantity being purchased. Obviously the integrity of the data on flavor is compromised in this situation. Some experimentation of aggregating a popular package type for a particular customer was done in an attempt to validate this theory, but the results did not substantiate the hypothesis.

Another potential cause of error that the consumer goods company acknowledged as a primary source of mismatch arises because of the asymmetry of internal and external information. This manifests itself at this company in two fashions, which are likely common at other companies whose products are sold by their customer to the end consumer. The first manifestation of this asymmetry is that the company has a unique customer ID for each store that must be matched to the customer's own internal store number. The company can not use the customer's internal store number because it would not be unique among all of the company's customers. The second manifestation of this asymmetry is that the company tracks deliveries using item IDs while the customer's POS data is collected based on UPC codes. Therefore, the company has to match its item IDs to the appropriate UPC codes. The obsolescence of products, the presence of limited time offer products, and the introduction of new products make this mapping very dynamic and challenging.

There are other general causes of mismatch between company and customer data. An interruption of the flow of data from the customer to the company would be classified as data feed error. If POS data must be sent manually by the customer to the company, this would introduce room for human error on either end. For example, the customer might neglect to send the data, and it might be irretrievable if the customer continually overwrites its files. Or if the company must manually upload the data to its own internal systems, it might accidentally miss an upload, thereby omitting data. Companies can minimize the risk of data feed error by automating the process of uploading customer data, or better yet, by having electronically connected databases (e.g., electronic data interchange or EDI).

In order to ensure that the data being analyzed was robust, a historical delivery data and POS data comparison was performed several times throughout the course of the project with incremental improvements made along the way. The most comprehensive set of results are shown in Table 1. The results are shown for three of the five customers that were providing the consumer goods company with POS data at the time of this work. The company had specified these as the focal customers for this project. A rating of "good" matches within 10%, "so-so" matches within 20%, and "bad" matches beyond 20%. Only Customer A's data was analyzed due to data integrity issues, and the majority of all analyses considered only those store-item combinations that matched within 10% of each other, as referenced later in this document.

OVERALL STORE COMPARISON PER YEAR SUMMARY

CUSTOMER A	2005		2006		2007 THROUGH 07NOV07	
	percent	count	percent	count	percent	count
good	17.6%	211	10.8%	130	71.9%	868
so-so	76.2%	914	69.0%	834	22.5%	272
bad	6.3%	75	20.2%	244	5.6%	68

CUSTOMER B	2005		2006		2007 THROUGH 31OCT07 (SKIP MAY - JULY)	
	percent	count	percent	count	percent	count
good	11.3%	124	0.3%	3	34.6%	330
so-so	65.2%	715	0.5%	5	49.0%	467
bad	23.4%	257	99.2%	967	16.5%	157

CUSTOMER C	2005		2006		2007 THROUGH 31MAY07	
	percent	count	percent	count	percent	count
good	0.9%	1	86.5%	96	85.2%	92
so-so	60.4%	67	11.7%	13	5.6%	6
bad	38.7%	43	1.8%	2	9.3%	10

STORE-UPC COMPARISON PER YEAR SUMMARY

CUSTOMER A	2005		2006		2007 THROUGH 07NOV07	
	percent	count	percent	count	percent	count
good	49.6%	97,285	51.0%	111,403	26.5%	58,740
so-so	15.8%	31,036	14.0%	30,538	29.7%	65,851
bad	34.6%	67,905	35.0%	76,583	43.8%	97,110

CUSTOMER B	2005		2006		2007 THROUGH 31OCT07 (SKIP MAY - JULY)	
	percent	count	percent	count	percent	count
good	28.9%	30,115	3.9%	5,390	28.4%	44,219
so-so	23.6%	24,656	6.0%	8,265	22.3%	34,649
bad	47.5%	49,517	90.1%	124,846	49.3%	76,734

CUSTOMER C	2005		2006		2007 THROUGH 31MAY07	
	percent	count	percent	count	percent	count
good	11.7%	1,784	53.0%	6,734	50.8%	5,255
so-so	29.2%	4,466	16.0%	2,030	16.1%	1,662
bad	59.2%	9,058	31.1%	3,949	33.2%	3,435

Table 1: Results of Historical Delivery Data and POS Data Comparison

4. Baseline Demand Forecast

4.1. Overview

At the time that this work began, the consumer goods company was using its historical delivery data in isolation to calculate the baseline forecast. This methodology was developed concurrently with the original handheld code to calculate suggested orders. In the interim, the company received POS data from some of its customers in order to improve forecasting and thereby enhance order quality. The company had used POS data to make improvements to handheld logic since then, and it had recently undertaken an initiative to improve the baseline forecast calculation dramatically using this data. This involved a three part approach:

1. Converting delivery data into POS-like data
2. Eliminating promotional spikes
3. Calculating the baseline forecast using non-promotional POS-like data

The company was already piloting a new method for eliminating promotional spikes and had recently begun using a one year trend line instead of a three month moving average to calculate the baseline forecast. What the company still needed was a way to convert delivery data into POS-like data.

The consumer goods company's own delivery data is intermittent and irregular – deliveries are not made every day, and they are not made at regular intervals. Therefore, it is difficult for the company to get an accurate picture of consumer demand since it is not able to measure how the inventory gets depleted between deliveries. However, a fairly accurate picture of consumer demand is captured by its customers' POS data on a daily basis.

The logical next step would be for this consumer goods company to use POS data to calculate the baseline forecast for customers. However, the company only had data from 0.82% of its customers. Fortunately, these five customers were national chains that comprised 14% of the company's sales volume. Since the company did not have POS data from every customer, it wanted to develop a methodology to estimate customer POS data using its own historical delivery data. This way, the company could generate estimated POS data for those customers for which it did not have POS data. Furthermore, the estimated POS data would give the company a more accurate picture of consumer demand, which it could then use to calculate the baseline forecast.

4.2. Development of Methodology

The boat chart shown in Figure 1 shows the development of the methodology to convert delivery data into estimated POS data. This occurred in four stages, beginning with understanding the original method the company was using to calculate baseline. The second stage was the most time-intensive because it involved having to devise a new and different methodology to convert the company's historical delivery data into estimated POS data by utilizing customer-provided POS data. The third stage resulted in a technique that was a slight variation of the second, while the fourth stage resulted in a methodology that was a marked improvement over the third.

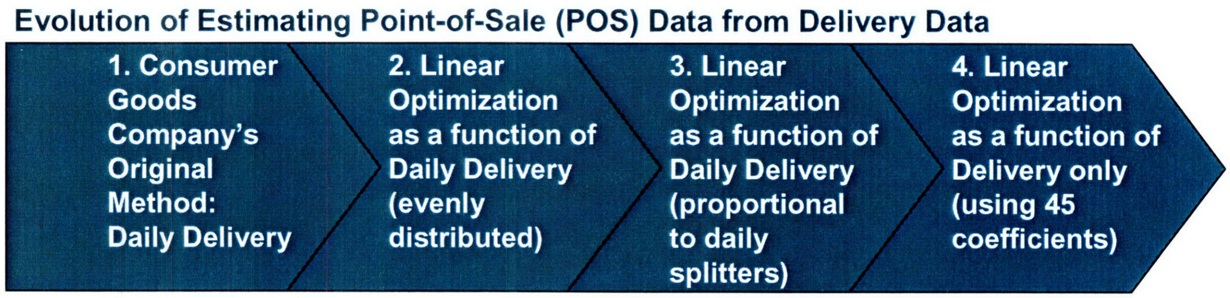

Figure 1: Evolution of Estimating POS Data from Delivery Data

4.2.1. Method 1 – Daily Delivery (Company's Original Method)

As previously mentioned, this consumer goods company's original method of calculating the baseline forecast used historical delivery data in isolation. Since delivery data is intermittent and irregular, the company came up with a metric known as the daily delivery, where:

$$\text{Daily Delivery} = \frac{\text{Delivery Quantity}}{\text{Days between Deliveries}}$$

Delivery data during one quarter for a given store-item combination is shown in Figure 2, while the corresponding daily delivery data is shown in Figure 3. For example, a delivery of about 60 cases was made on 5/03/2006 and the next delivery was made six days later on 5/09/2006, the daily delivery was about 10 cases per day.

The consumer goods company was basically using daily delivery as a proxy for daily sales. The company took the average of the daily delivery points over a one year period and then calculated 1.5 times that average, represented by the red line in Figure 3. The company assumed that any points above 1.5 times the average daily delivery were promotional points, so all points above that line were thrown out. The remaining points were presumed to be the baseline points, and a linear regression trend line was fit to those points to yield the baseline. The problem with this method is that if deliveries were made on two successive days, then the first delivery quantity all gets allocated to a single daily delivery, creating an artificially high spike on the graph. Two instances of this are shown in the graph in Figure 4.

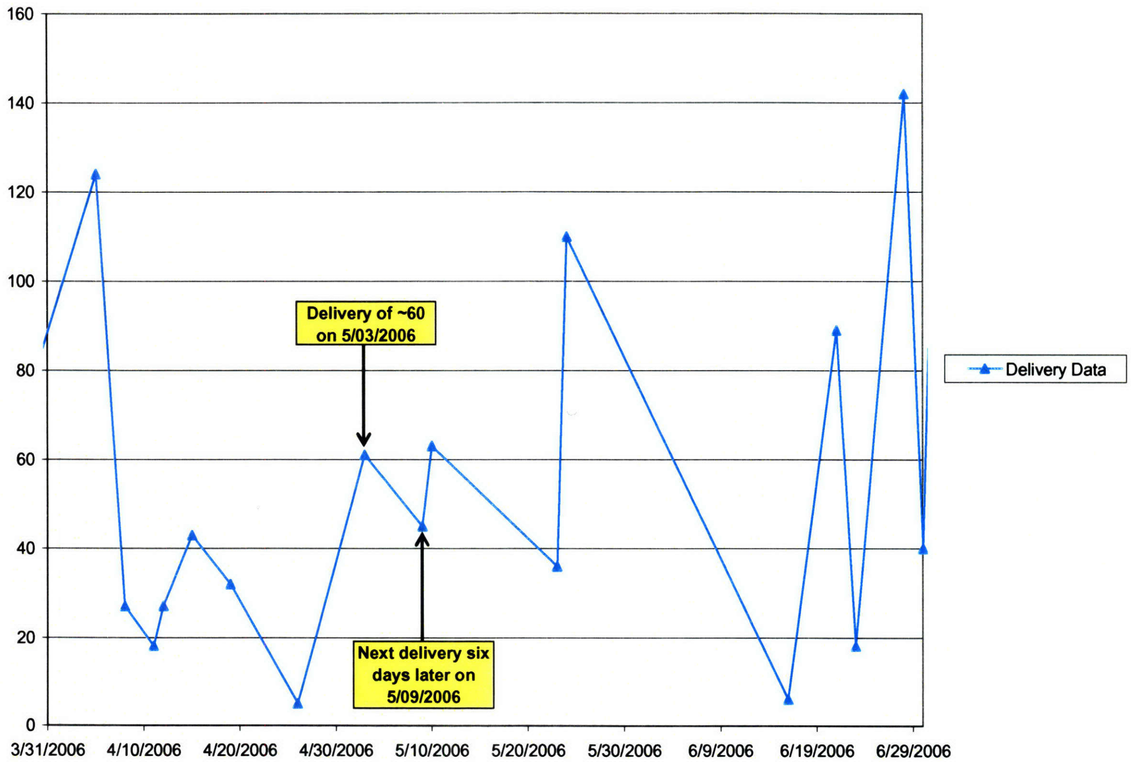

Figure 2: Delivery Data

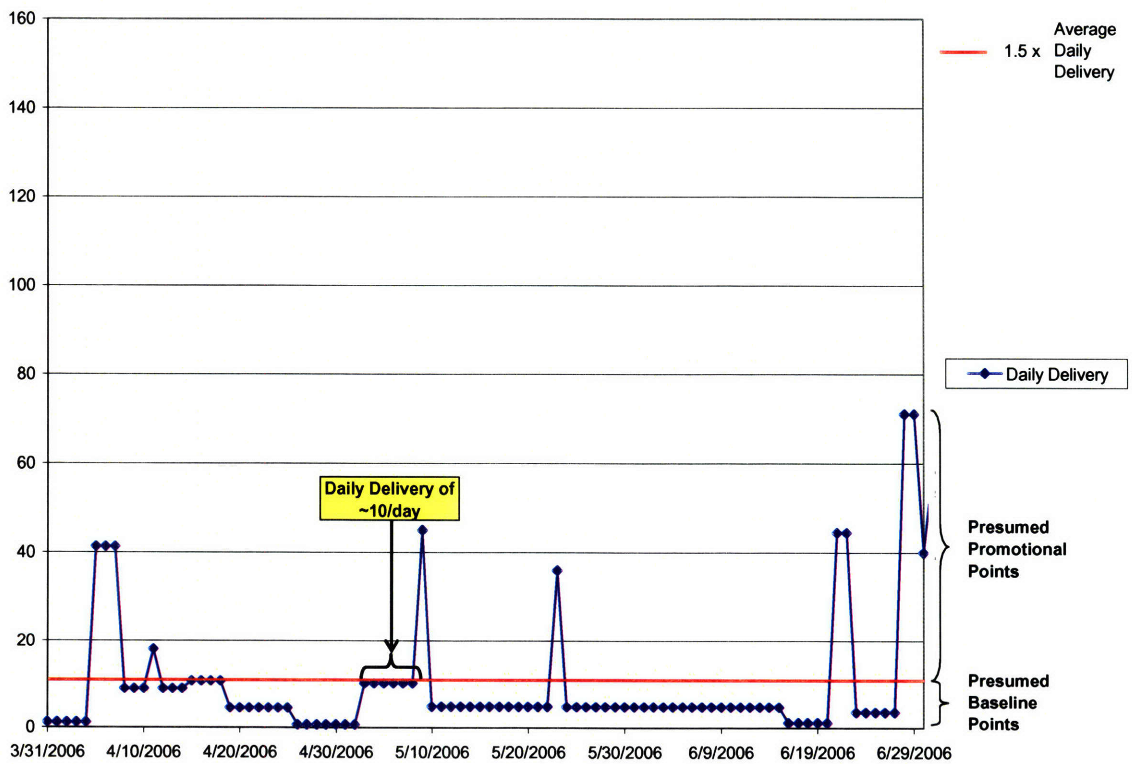

Figure 3: Daily Delivery (Calculated as Delivery Quantity/Days between Deliveries)

A promotional point by definition would be one where a decrease in price caused an increase in quantity sold. The consumer goods company's method of identifying promotional points assumes that adjacent deliveries are being made because a promotion is driving higher sales volume. There are two problems with this premise. The first problem is that since the high spikes in the graph appear in instances where deliveries were made on two successive days, the delivery quantity that gets thrown out is somewhat arbitrary, as is the delivery quantity that gets included by being spread across several days. This relationship between delivery and daily delivery is illustrated visually as shown in Figure 4. Those deliveries whose daily deliveries were included in the baseline regression are circled in green, while those deliveries whose daily deliveries were excluded from the baseline regression are crossed out in red.

What is actually happening is that the second delivery quantity is generating arbitrarily low daily deliveries that the company has to try to identify and exclude from the linear regression of the baseline points, while valuable data is being lost by eliminating the first delivery quantity. In reality, the quantity of the *combined* deliveries should be spread out across those days in keeping with the assertion that the daily delivery is a proxy for daily sales. The second problem is that this method fails to account for the time lag between when the product is delivered to the customer's back room and when the inventory is actually moved onto the customer's shelves by the merchandiser. In the case of adjacent deliveries, allocating the entire delivery quantity to a single day effectively implies that everything that was delivered to the customer's back room was sold the same day, which does not accurately capture the flow of inventory through the store.

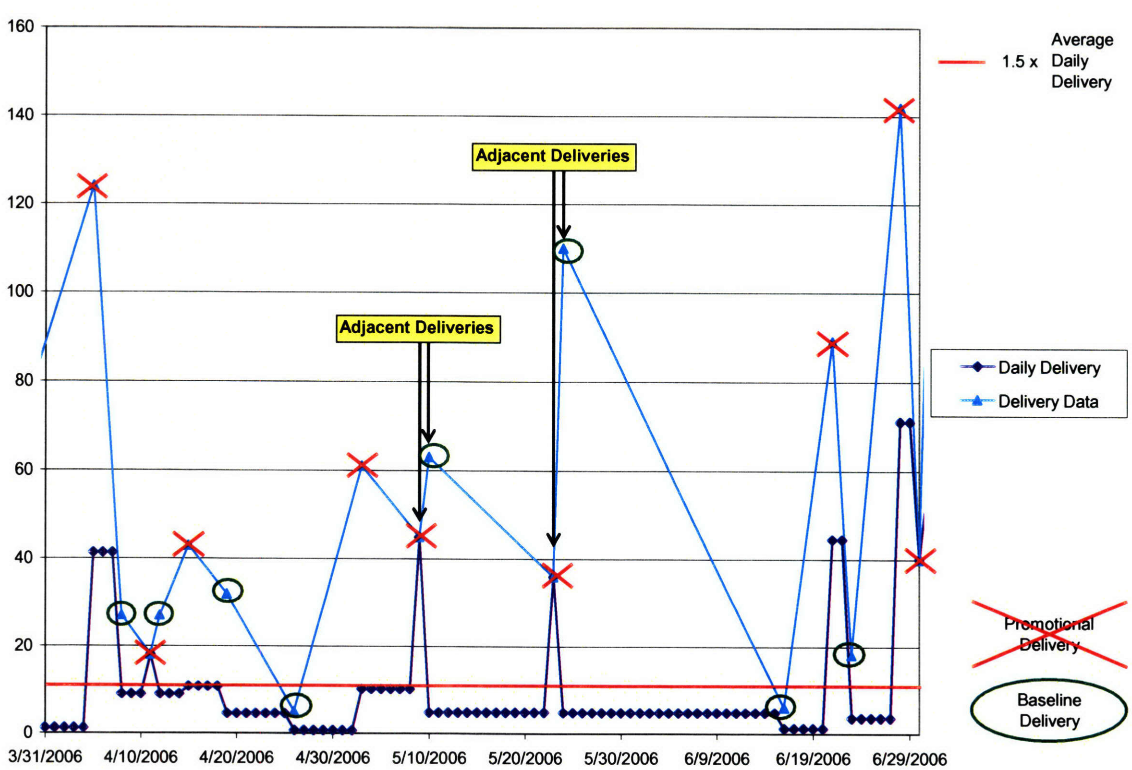

Figure 4: Relationship between Delivery Data and Daily Delivery

These issues imply that the daily delivery is not always an accurate proxy for sales. This was substantiated by comparing the daily delivery data to actual POS data: a sample quarter of data is shown in Figure 5. The estimated POS data (daily delivery in this case) appears as a navy blue line labeled "POS New Est Cases" while the actual POS data is represented by a pink line labeled "POS Vol Cases". The arbitrarily high spikes created by adjacent deliveries are obvious places where the daily delivery does a poor job of approximating actual POS data.

Establishing a metric for the accuracy of the consumer goods company's original method would facilitate comparison with the development of subsequent methods. Since daily delivery data was a proxy for daily sales, logic implied that the difference between the estimated POS data (in this case daily delivery) and the actual POS data should be measured at each point. This is the traditional definition of forecast error where:

$$Error(\%) = \frac{|(Actual - Forecast)|}{Actual}$$

This error was averaged across all of the points where POS data was available (since an actual value of zero in the denominator would generate an infinite error). This results in a traditional metric used for these types of measurements known as the Average Absolute Relative Forecast Error (AARFE):

$$AARFE = \frac{\sum_{1}^{n} \frac{|(Actual - Forecast)|}{Actual}}{n}$$

The company's original method yielded an AARFE of 1.52, which means that daily delivery as a proxy for daily sales (actual POS) yields an error with an average magnitude of 152% (for this particular store-item combination).

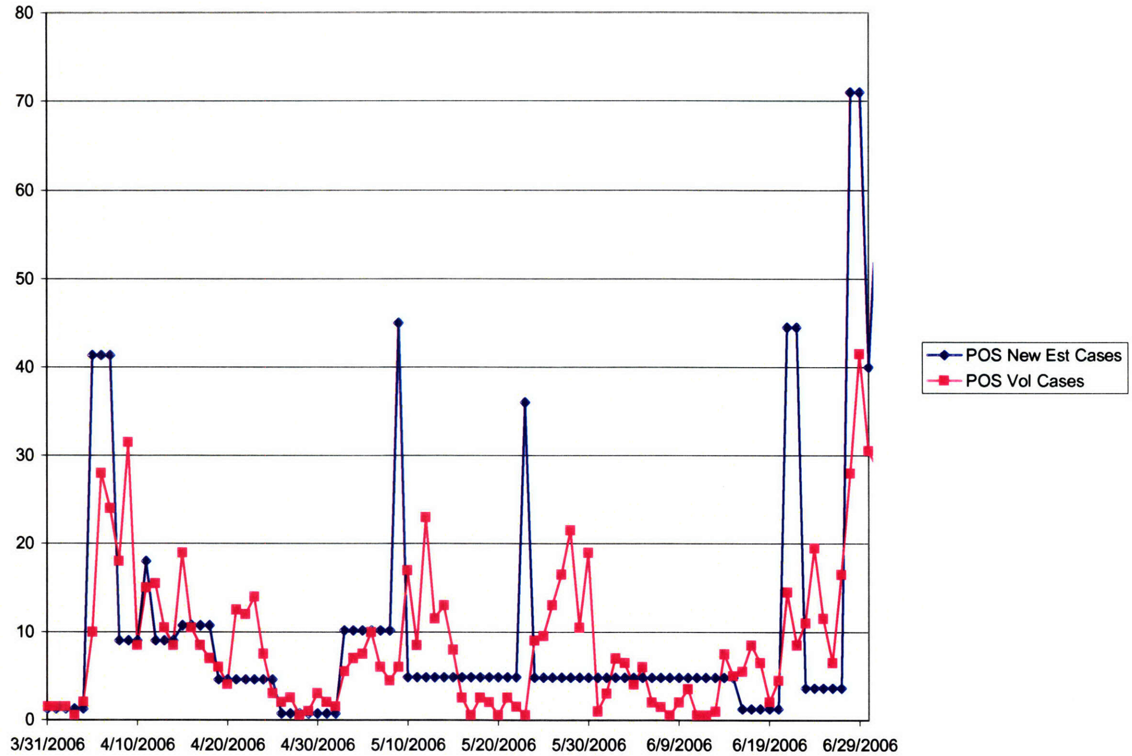

Figure 5: Method 1 – Daily Delivery (Consumer Goods Company's Original Method)

4.2.2. Method 2 – POS as a Function of Daily Delivery (Evenly Distributed)

Reflecting on the first problem with the premise of the consumer goods company's original method, it seemed to make sense to consider nearby deliveries rather than a single delivery in isolation. Contemplating the second problem, an approach that could account for the time lag was needed. Consultation with several sources, including faculty advisors whose specialties were statistics and supply chain, indicated that there were no mathematical functions that could describe the relationship between delivery and POS data. In fact, there was skepticism among these specialists that *any* methodology could be developed that would improve upon the current one. The circumstances called for a heuristic approach, and linear optimization was selected. The estimate of POS would be a function of the historical delivery data in the optimization. The objective function would be to minimize the AARFE while constraining the coefficients to be greater than zero (to prevent prediction of negative sales) and to sum to 1 (to enable the total delivery volume to match the total predicted sales volume).

Analysis of historical delivery data from a specific fast moving store-item combination indicated an average 4.5 days between deliveries. Based on this information, the original linear optimization considered daily deliveries for 4.5 days on either side of the current delivery, with the intention of accounting for the delivery before and the delivery after the current delivery (on average). Relatively good results were achieved, but since the store can not sell inventory that has not yet been delivered, the method did not reflect the reality of what was happening.

In order to capture the actual flow of inventory, the optimization was modified so that the current days sales were a function of the daily delivery over the past ten days (including the current day). The graphical results of this methodology are shown in Figure 6, over the same time period and using the same historical delivery data and actual POS data used to generate the plot in Figure 5. This second methodology yielded an AARFE of 1.16: an absolute improvement of 36.4% and a relative improvement of 23.9% over the company's original method.

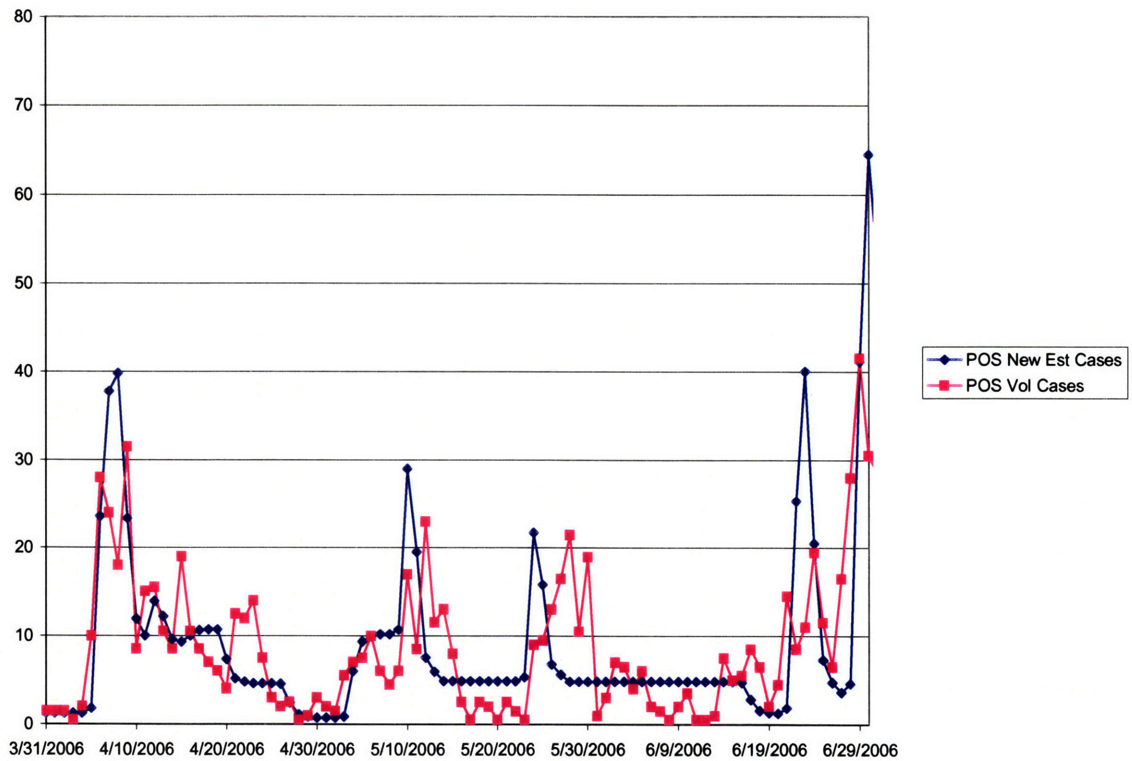

Figure 6: Method 2 – POS as a Function of Daily Delivery (Evenly Distributed)

4.2.3. Method 3 – POS as a Function of Daily Delivery (Using Splitters)

Based on the consumer goods company's input, the second methodology was slightly modified to include information that the company was already using in forecasting known as the daily splitter. The company had developed the daily splitter through analysis of historical data, and it indicated what percentage of the week's sales typically took place on a given day for a particular customer store. This is described in the introduction as weekday effects. For example, the company tended to sell more on weekends when most people shop.

Since the daily splitter summed to 100% for a given one week period, the daily splitter had to be applied proportionally to the delivery quantity given the number of days that it was being spread across (i.e., the number of days between deliveries). Applying the daily splitter to generate the daily delivery meant that the daily delivery quantity was no longer uniform – it varied slightly depending on the day of the week. As such, the quantity being optimized was referred to as the *proportional* daily delivery. The graphical results of this methodology are shown in Figure 7, over the same time period and using the same historical delivery data and actual POS data used to generate the plots in Figure 5 and Figure 6. This third methodology yielded an AARFE of 1.16: an absolute improvement of 37.4% and a relative improvement of 24.6% over the company's original methodology.

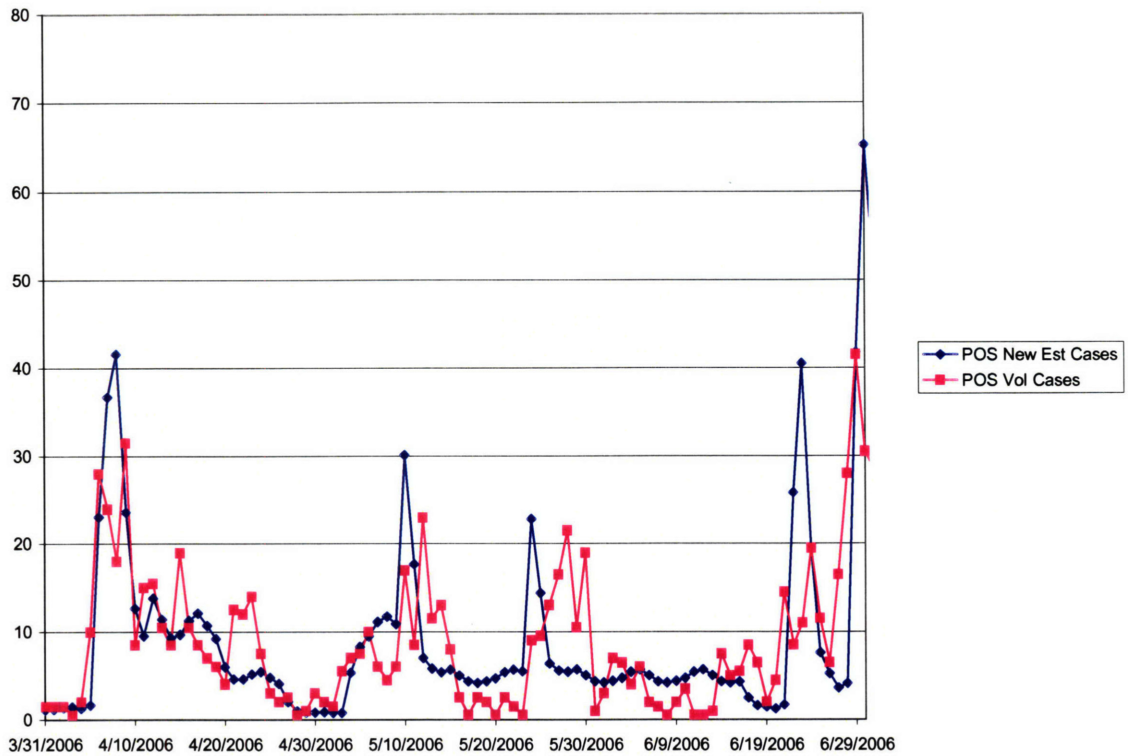

Figure 7: Method 3 – POS as a Function of Daily Delivery (Using Splitters)

4.2.4. Method 4 – POS as a Function of Past 45 Days of Delivery

Several modifications to the third methodology were experimented with in an attempt to generate improved results. One variation was to eliminate the absolute value and optimize based on the Average Relative Forecast Error (ARFE):

$$ARFE = \frac{\sum_{1}^{n} \frac{(Actual - Forecast)}{Actual}}{n}$$

However, it did not appear to produce significantly better results. The decision was made to keep AARFE, which measured the error like ARFE but used the absolute value, since the magnitude of the error was more important than the direction of the error.

Another modification that was attempted was to include an intercept term in the optimization. Although this produced slightly improved results, the intercept was typically negative and generally accounted for the difference between the total delivery volume and the total actual sales volume (since they did not exactly match). A non-zero term was created even when the intercept was constrained to be positive, but the consumer goods company rejected the idea of using an intercept since the function could theoretically predict sales in the absence of deliveries.

A faculty advisor whose expertise is in statistics suggested optimizing on the original delivery data itself in lieu of the daily delivery data or proportional daily delivery. Experimentation with this proposal yielded incrementally improved numerical results but substantially improved visual results. The optimization using the original delivery data produced estimated POS data that was a much better fit to the actual POS data, especially in comparison with the company's original method. The graphical results of this methodology are shown in Figure 8, over the same time period and using the same historical delivery data and actual POS data used to generate the plots in Figure 5, Figure 6, and Figure 7. This fourth methodology yielded an AARFE of 1.06: an absolute improvement of 45.9% and a relative improvement of 30.2% over the company's original methodology.

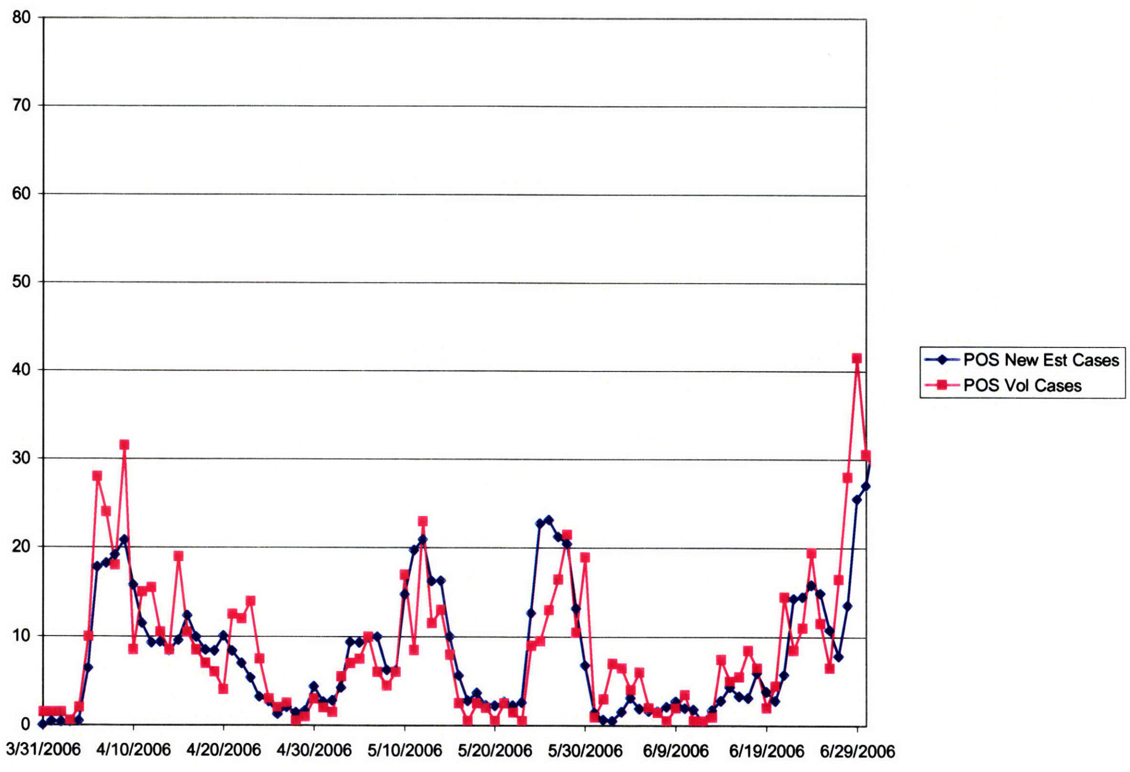

Figure 8: Method 4 – POS as a Function of Past 45 Days of Delivery

4.3. *Identification of Cluster Attributes*

Once the methodology was finalized, the next step was to determine how to generalize it for application to various flavors and package types. This process was referred to as clustering. Delivery frequency seemed like a logical cluster attribute since the coefficients would vary substantially for a store-item combination that was fast-moving versus one that was slower. Analysis was conducted to determine deliveries per year for all store-item combinations in the pre-sell system, and these were plotted on a histogram, which is shown in Figure 9. Clear inflection points in the histogram seemed to imply natural break points for groupings at 52 deliveries per year (approximately once per week) and 104 deliveries per year (approximately twice per week).

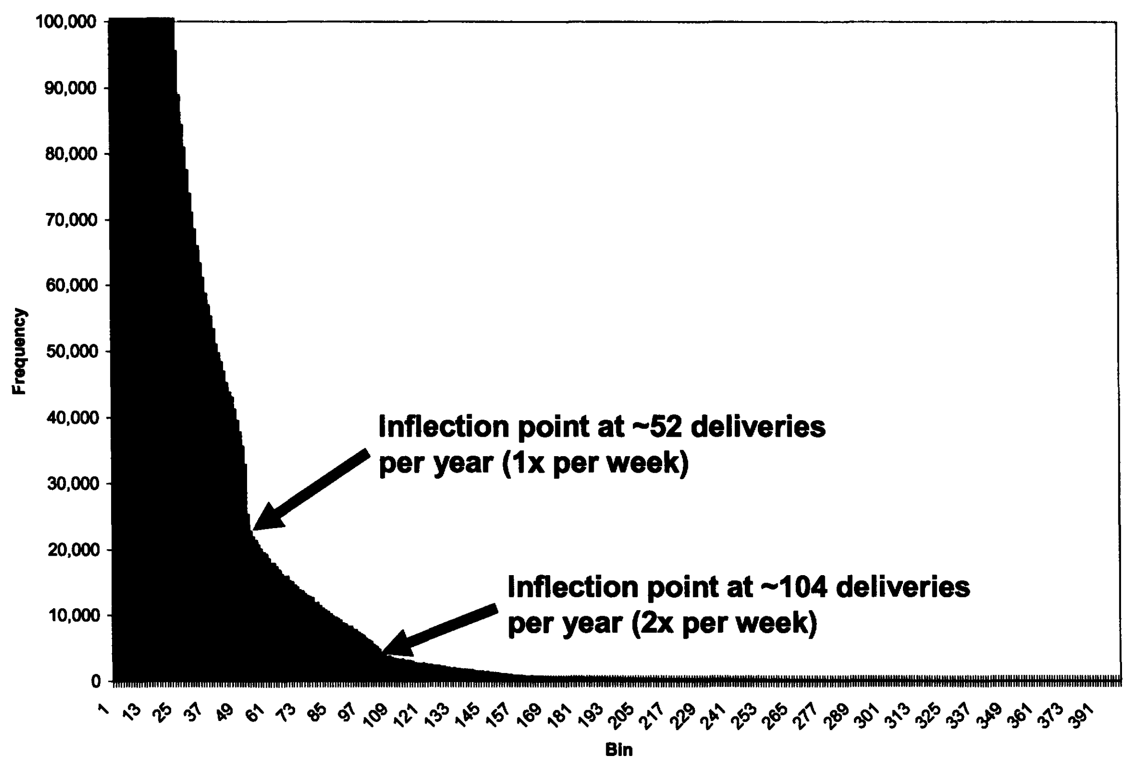

Figure 9: Frequency of Deliveries per Year

Examination of sample delivery and POS data seemed to indicate that package type would also be a natural cluster attribute. Products are ordered in cases (delivery data), whereas they are purchased in units (actual POS data), and the conversion from cases to units varies from one to as many as 30 for different package types. Ordering patterns differ depending on package type because sales reps must order in cases and try to order in pallet "layers". This is illustrated by comparison of the sample delivery data plots for Flavor A Package A and Flavor A Package B in Figure 10 and Figure 11. For example, a case of Package B with 8 units suffers from a greater bullwhip effect than a case of Package A with only 2 units. Sales patterns differ depending on package type due to the way they are promoted and consumer buying behavior. This is illustrated by comparison of the sample POS data plots for Flavor A Package A and

Flavor A Package B in Figure 12 and Figure 13. For example, Package B sales volume is more dependent on price than Package A sales volume, so Figure 13 exhibits much higher volume spikes due to promotions than Figure 12.

Note the presence of returns in the delivery data in Figure 10 and Figure 11. It should be pointed out that these returns were added back to previous deliveries, effectively smoothing the data, during the development of the methodology described above. The problem with including returns in the optimization is that they would appear as negative deliveries, artificially altering the coefficients and potentially allowing the function to predict negative sales.

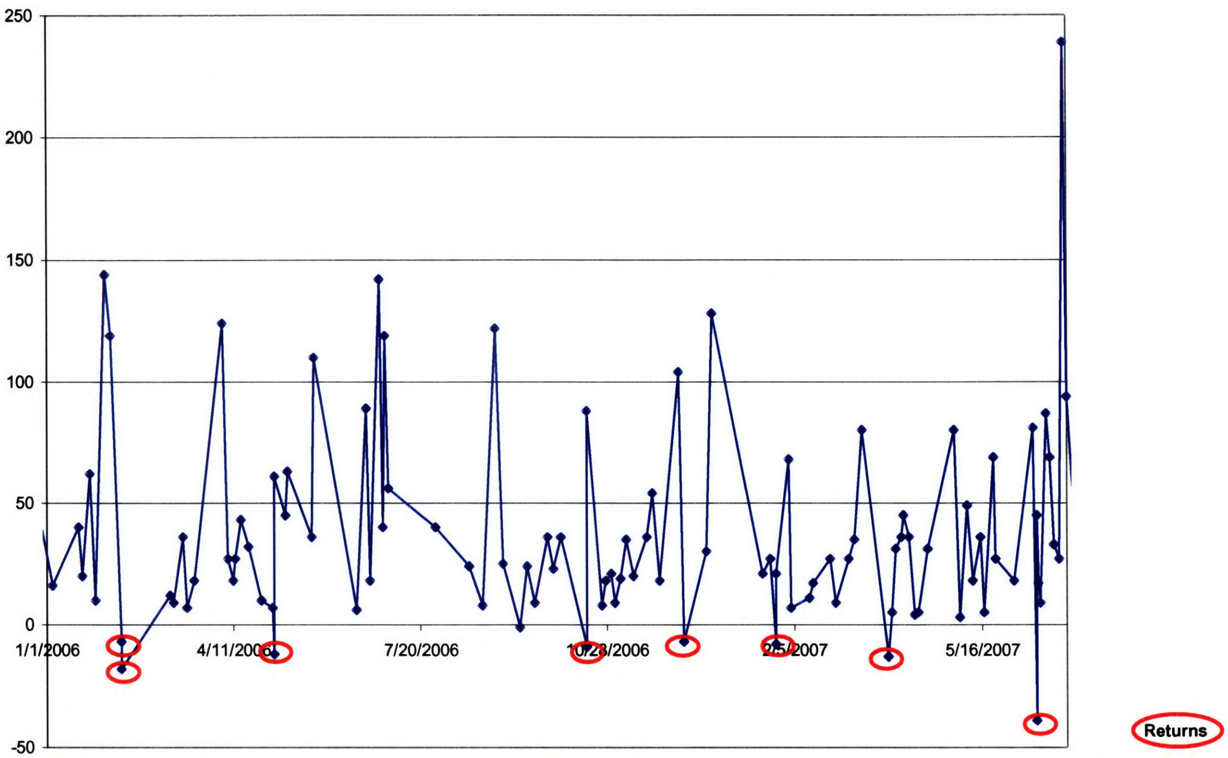

Figure 10: Sample Delivery Data for Flavor A Package A in Customer A Store 1880

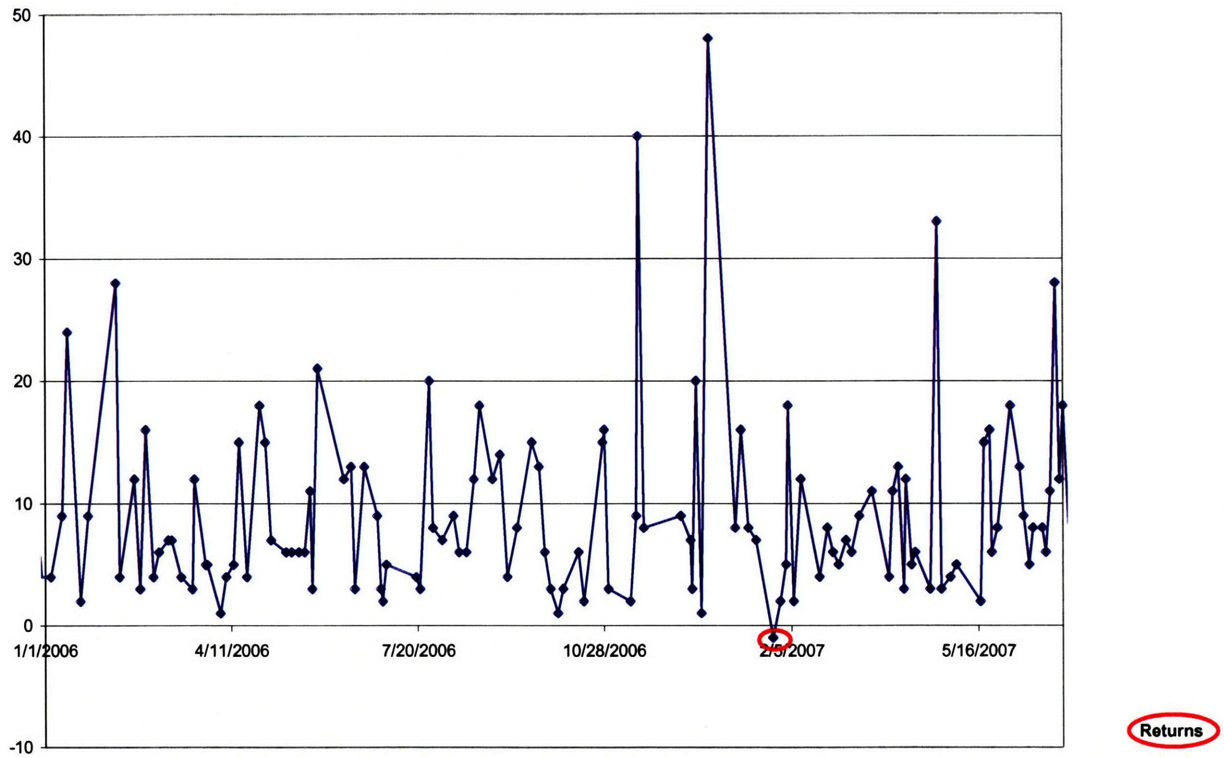

Figure 11: Sample Delivery Data for Flavor A Package B in Customer A Store 1880

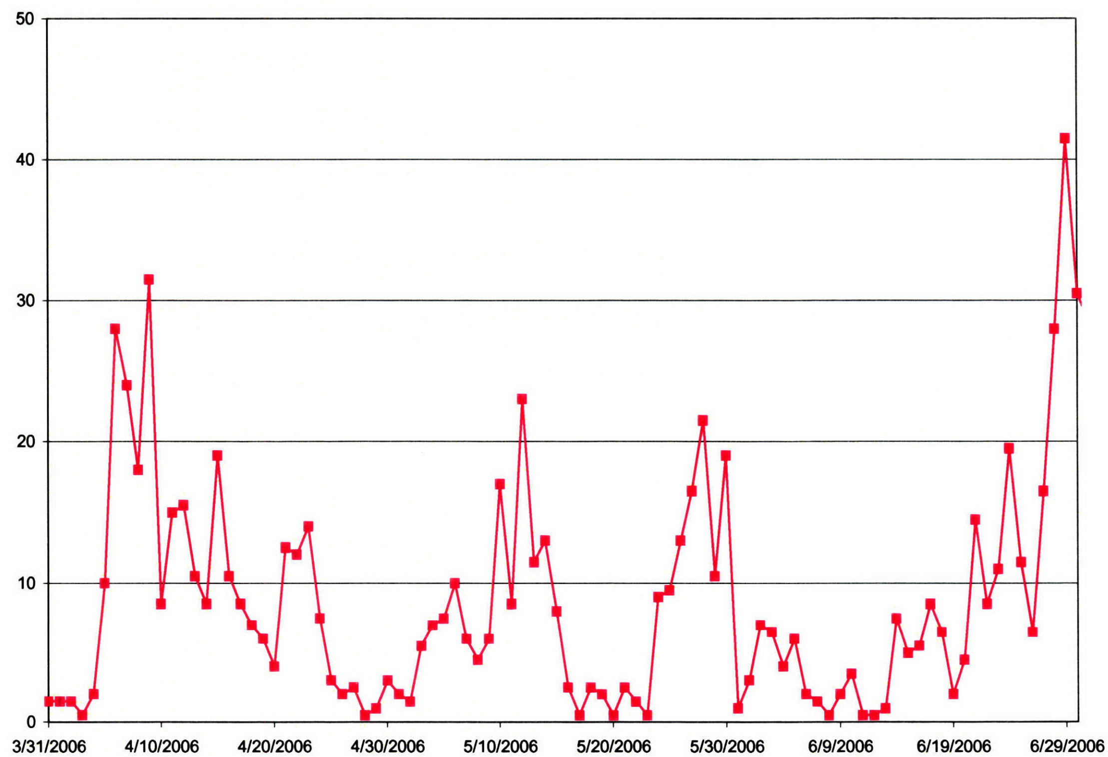

Figure 12: Sample POS Data for Flavor A Package A in Customer A Store 1880

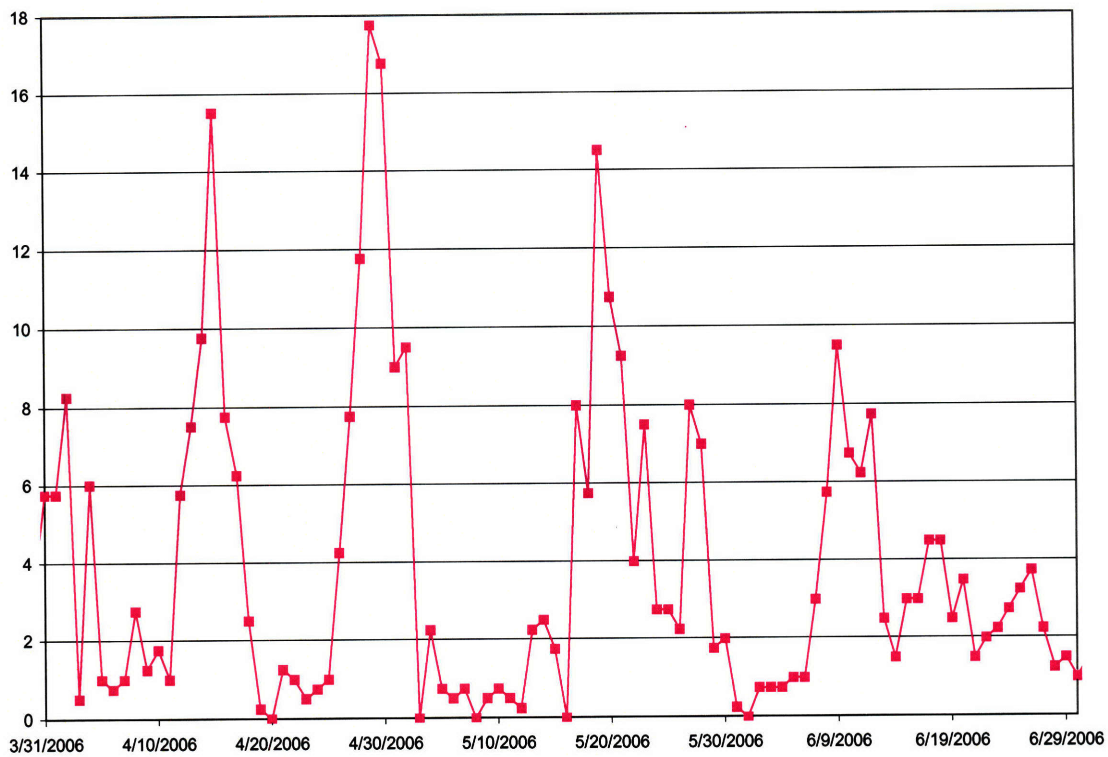

Figure 13: Sample POS Data for Flavor A Package B in Customer A Store 1880

To further refine package type as a cluster attribute, some experimentation with a small number of store-item combinations was performed. Other fast-moving products that were sold in the same store as the particular store-item combination used to develop the methodology (Package A) were identified, and their historical delivery and POS data were obtained. Optimization of coefficients for each of these unique package types was then conducted, the results of which are represented by the blue bars in Figure 14. Then the original coefficients generated for the particular store-item combination used to develop the methodology (Package A) were applied to these other package types, the results of which are represented by the purple bars in Figure 14. Comparison of these results clearly confirmed that package type was indeed a cluster attribute. Experimenting by applying different coefficients to different package types determined what patterns might exist. Based on insights gained during this testing, coefficients of package types that were similar (e.g., only varied in flavor) were blended and then applied back to the package types used to develop them, the results of which are represented by the beige bars in Figure 14. Fairly good results were obtained using this method, as shown by comparison of the blue bars and the beige bars in Figure 14. Armed with two strong cluster attributes, generalization of the methodology to all store-item combinations could begin through development of a comprehensive set of coefficients.

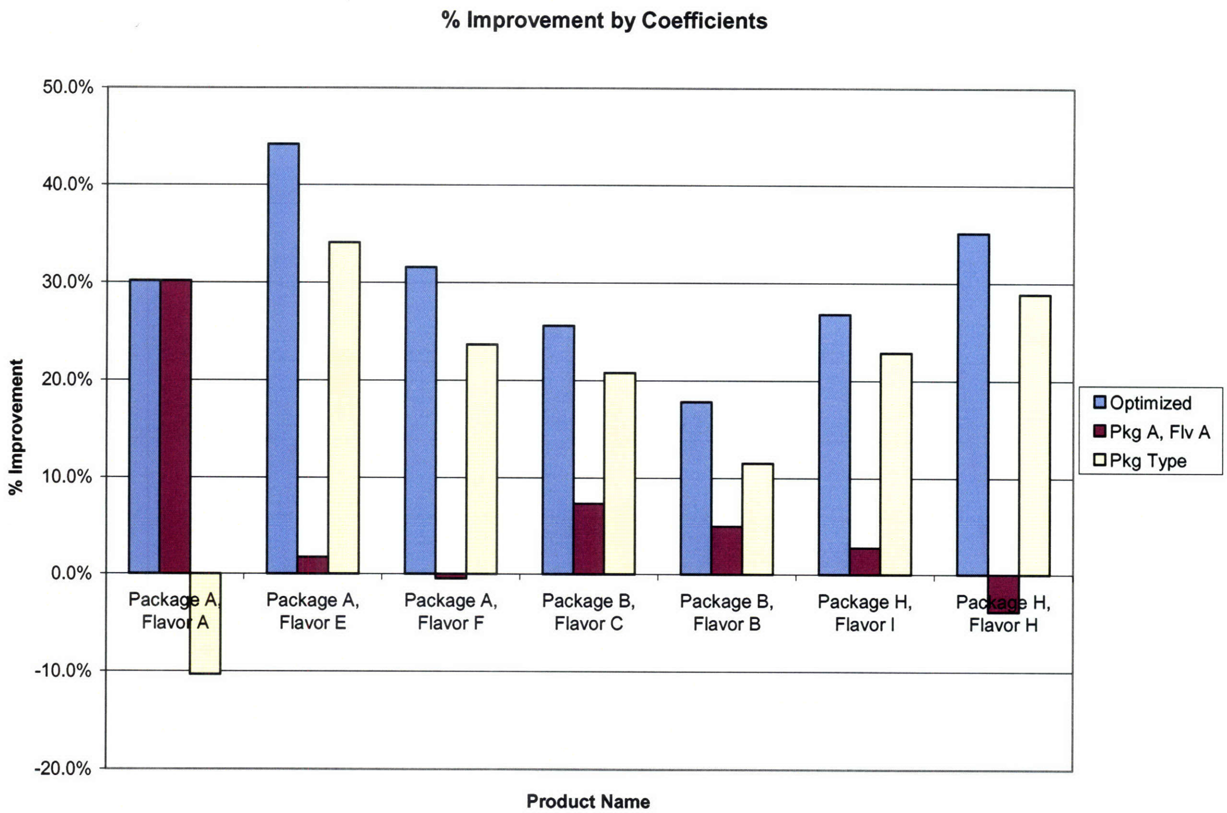

Figure 14: Results of Coefficient Experimentation

4.4. Creation of Coefficients

While coefficients theoretically could be developed for every single package type sold, doing so was impractical given the sheer variety of package types that existed in the consumer goods company's system, numbering close to 300. Only Customer A's 2005 – 2006 daily POS data was robust enough to use for creation of these coefficients. A volume analysis by package type performed on this data revealed that 11 package types accounted for 95% of the customer's overall sales volume. Now that the package type cluster attribute had been addressed, the delivery frequency cluster attribute needed to be considered. Further analysis was done to determine which flavor was the highest-selling by volume within each of the 11 package types identified. The logic behind this analysis was that the highest-selling volume flavor would yield the broadest range of delivery frequencies within a given package type. The results of the package type and delivery frequency cluster attribute analysis are shown in Table 2.

Package Type	Highest Volume Flavor[1]	Percent of Customer A Sales Volume[2]
Package A	Flavor A	54.6%
Package B	Flavor D	12.3%
Package C	Flavor A	7.2%
Package D	Flavor D	6.8%
Package E	Flavor G	3.9%
Package F	Flavor G	2.8%
Package G	Flavor B	1.7%
Package H	Flavor I	1.5%
Package I	Flavor J	1.5%
Package J	Flavor I	1.2%
Package K	Flavor A	1.0%

[1] Where delivery and POS data matched within 10% of each other; this was usually the same as overall volume except for Package B

[2] Based on 2005 – 2006 delivery data; numbers are approximate

Table 2: Package Type and Delivery Frequency Cluster Attribute Analysis

Since there were 11 package types and three delivery frequencies within each of those package types, a total of 33 sets of coefficients were developed. The process used to determine these coefficients was to subset the data for a particular package type in order to identify the highest-selling flavor by volume. It should be noted that the match between the historical delivery and actual POS data for the flavor were confirmed to be within 10% of each other. If the data did not meet this requirement, the next highest-selling flavor by volume that satisfied this criterion was selected. The package type data was then further subset to isolate the particular highest-selling flavor. The mean within each delivery frequency was determined, and then the stores around the mean were identified. A visual depiction of this method is shown in Figure 15.

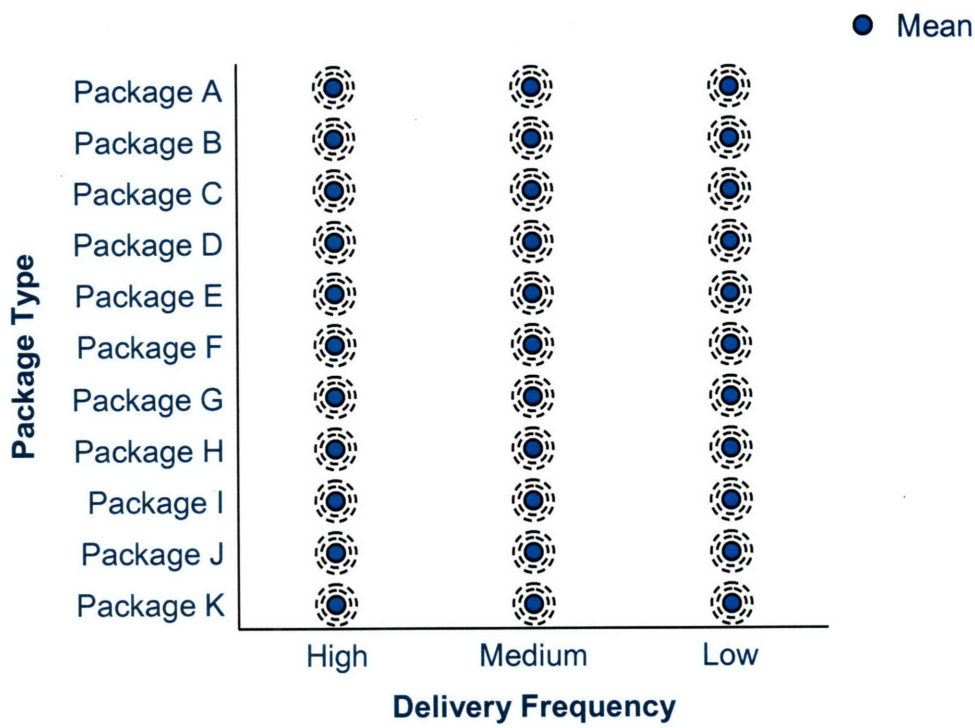

Figure 15: Visual Depiction of Clustering Methodology

An example of the application of this methodology for a specific package type is shown in Table 3.

Package: Package A		Flavor: Flavor A		UPC: UPC A	
High (> 104 / yr)		Medium (52 – 104 / yr)		Low (≤52 / yr)	
Store Code	Deliveries/ Year[1]	Store Code	Deliveries/ Year[1]	Store Code	Deliveries/ Year[1]
1928	129.5	1795	71.5	945	43.5
1681	129	2116	71.5	1076	43.5
1791	128.5	2120	71.5	1369	43.5
1938	127.5	2121	71.5	2134	43.5
2913	126.5	2197	71.5	2266	43.5
				3208	43.5
Cluster Mean	**128.1**	**Cluster Mean**	**71.6**	**Cluster Mean**	**43.3**

[1] Average based on 2005 – 2006 delivery data

Table 3: Stores Selected for Three Delivery Frequencies within a Package Type

4.5. Measurement of Results

A SAS program was created to measure the estimated POS accuracy of the new methodology versus the consumer goods company's original method. This involved applying the 33 sets of coefficients that were developed to the 11 package types to which they applied across the entire Customer A system. Only stores where delivery and POS data were within 10% of each other were considered in this analysis. The rationale is that since error is being measured in the context of POS data, if the historical delivery data and POS data are wildly off, there will already be a significant error built into the data. The results of this analysis are shown in Table 4.

Package Type	Delivery Frequency			Overall
	High (> 104 / yr)	Medium (52 – 104 / yr)	Low (≤52 / yr)	
Package A	12.8%	12.7%	8.8%	11.6%
Package B	6.2%	13.2%	9.3%	11.2%
Package C	8.9%	11.7%	11.7%	11.6%
Package D	12.8%	3.2%	10.9%	7.7%
Package E	10.2%	10.9%	10.7%	10.6%
Package F	17.2%	16.2%	9.7%	12.9%
Package G	10.7%	14.4%	12.8%	13.7%
Package H	13.9%	18.5%	16.8%	17.6%
Package I	17.3%	15.4%	11.5%	12.6%
Package J	14.5%	14.0%	14.2%	14.1%
Package K	15.6%	15.4%	13.4%	14.4%
Overall (including other)	**11.1%**	**9.6%**	**10.1%**	**9.9%**

Note: Improvements shown are average differences in AARFEs (old – new methodology)

Table 4: Measurement of Results

An example of what the estimated POS data look like when the optimized coefficients (for that particular store-item combination) are applied versus when the clustered coefficients are applied are shown in Figure 16 and Figure 17. The optimized coefficients yield a 23.5% improvement, whereas the clustered coefficients give a 16.0% improvement over the consumer goods company's original method.

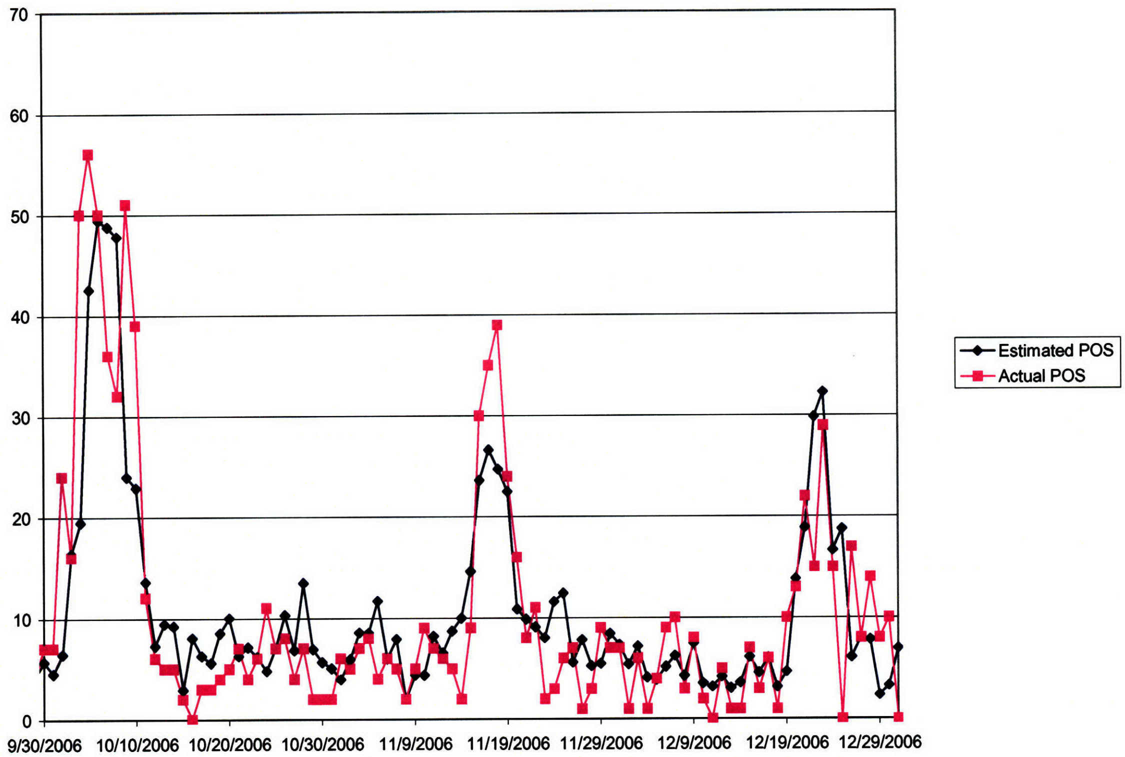

Figure 16: Estimated POS When Optimized Coefficients Applied to Package E (Flavor G)

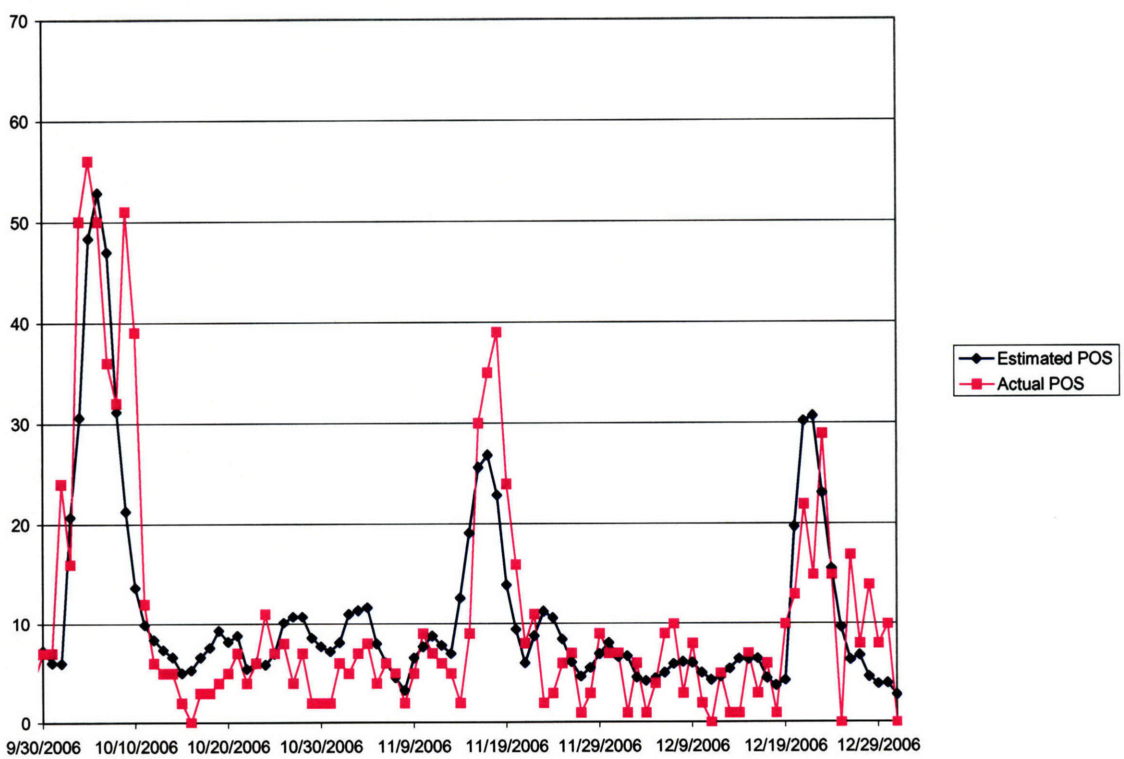

Figure 17: Estimated POS When Clustered Coefficients Applied to Package E (Flavor G)

4.6. Implementation

The consumer goods company plans to implement the baseline forecast methodology developed during the course of this work. This section briefly outlines a set of suggested next steps that the company could pursue.

The first step would be to change the baseline code in the consumer goods company's delivery database, which is known as Data Warehouse. The coefficients developed would be applied to generate estimated POS data for Customer A. Custom thresholds, developed just prior to the commencement of this project by another member of the Supply Chain Selling Systems group, would be used to identify promotional points. The new baseline forecast would then be calculated excluding promotional points.

The next step would be to conduct a new baseline validation. This could be accomplished by comparing the new baseline to the old baseline and the POS baseline to determine the improvement. During the course of this project, SAS programs were developed to compare the delivery baseline and the POS baseline for Customer A and Customer B. These SAS programs can easily be extended to Customer C, Customer D, Customer E by the consumer goods company, and any other customers that may provide POS data in the future.

The final step would be to develop these coefficients for other key accounts where the consumer goods company has POS data (Customer B, Customer C, Customer D, Customer E). Alternatively, the company could try applying the coefficients developed for Customer A to these other key accounts. If the measurable improvement (determined as detailed in this document) were significant, it would indicate that the coefficients are effective when applied to other key accounts.

This page has been intentionally left blank

5. Cannibalization Effect on Forecast

5.1. Overview

At the time this work began, the consumer goods company suspected the existence of cannibalization among its own products due to promotions. That is, when the company promoted a product that competed with another one of its products, it reduced the sales of the latter. Anecdotes from the company's field sales reps, who had relayed their observations to the Supply Chain Selling Systems group, substantiated the existence of cannibalization and determined it to be a priority. Furthermore, the field sales reps believed that this was occurring among package types (as opposed to flavors) and had identified specific package types that they believed were cannibalizing each other.

The presumed existence of cannibalization motivated the second phase of the project – modeling the effect of cannibalization on demand for high volume products during promotions and new product introductions. There were three major package types whose cannibalization interactions were of significant interest to the company. First, since Package A is its highest-selling package type, the company wanted to understand the effect that promoting other package types had on Package A's sales. These other package types fell into two categories: promotional products and new products. The primary promotional product under consideration was the Package C, known as an "in-and-out" product within the company because the customer does not always carry it in stores. The main new product introduction being investigated was the Package M, which was essentially the same as the Package A except its unit packaging had a different form factor. Therefore, the second phase of the project would consist of modeling the effect of Package C on Package A and the effect of Package M on Package A using customer-provided POS data.

5.2. Regression Analyses

In order to model the effect of Package C on Package A sales, a SAS program was created to run regression on Customer A POS data at the weekly level. This was consistent with the consumer goods company's current practice of aggregating at the weekly level. Daily POS data would be too granular to create a predictive model for several reasons. Modeling based on data aggregated at the weekly level would help reduce the variation in sales, effectively smoothing consumer demand. For example, days where no sales occurred would be inconsequential when rolled up at the weekly level. Days where few products were sold would not confuse the model or create noise since they were aggregated at the weekly level. Finally, since the manufacturer and customer promotions are run on a weekly basis, running regression on data aggregated at the weekly level that is consistent with promotion start and end dates makes sense.

Initially, the program was run for a single Customer A store and multiple flavors and was later generalized across all of Customer A's stores. During this process, it was determined that the regression should be run on the logarithm of volume rather than the volume itself. Using the logarithm had a leveling effect that enabled the SAS regression to better interpret the relationships among variables, thereby yielding models with better predictive power. A plot was created to visually interpret the predictive power of the cannibalization models and is shown for a specific store-item combination in Figure 18.

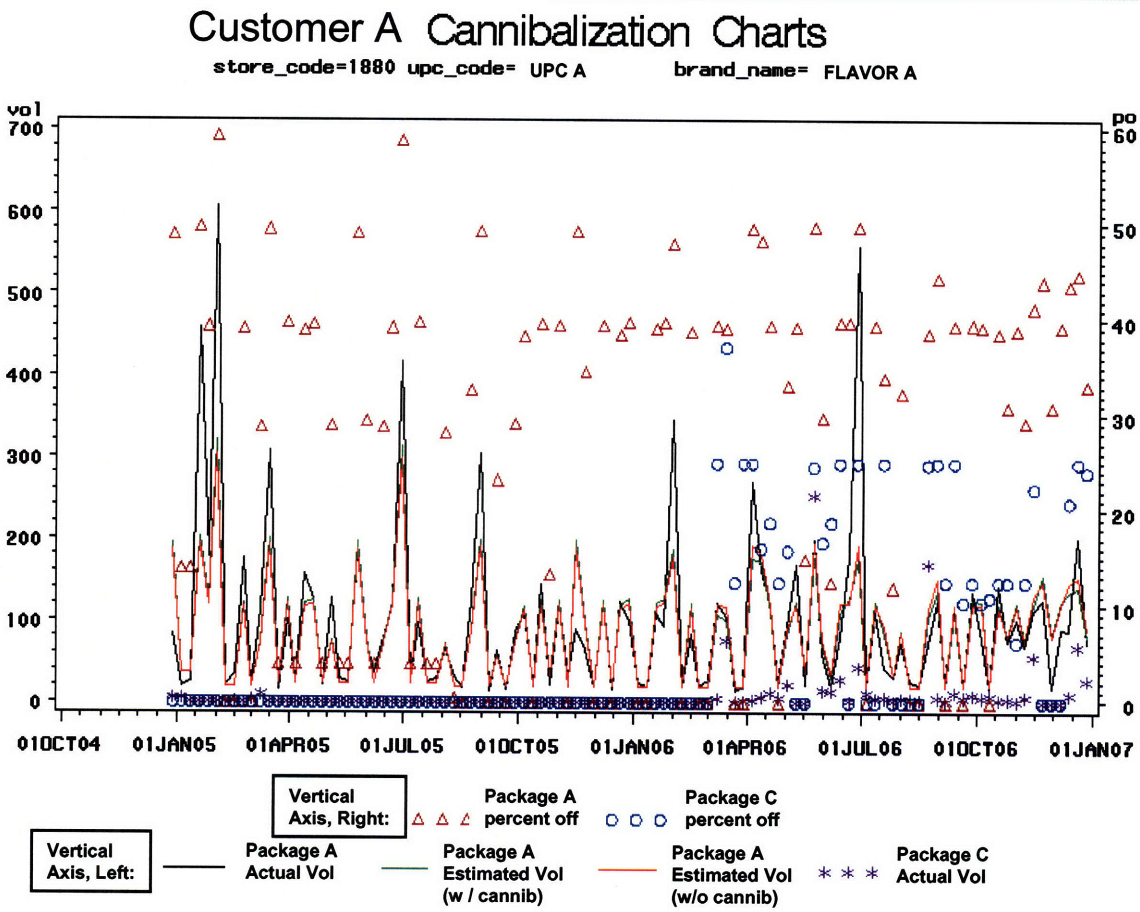

Figure 18: Sample Plot of Package A Volume Predicted by Regression Models

One of the main variables considered was a binary flag that indicated whether or not Package C was being promoted based on a particular Package C percent off (PO) level. This metric measured the magnitude of the promotion by comparing the shelf price to the retail (promotional) price. The other main variable was the Package C percent off itself. Various combinations of these variables were experimented with: each in isolation, together (independently), together (effectively the interaction term), and the logarithm of Package C percent off (in isolation). This analysis indicated that the interaction term was the best variable because it generated the greatest percentage of valid models with Package A coefficients and cannibalization, as shown in Table 5.

	1. Models Generated whose POS-Delivery Matches w/in 10%	2. Positive Package A Coefficient Generated	3. Negative Package C Coefficient Generated	Percent of All Models with Cannibalization (3/1)	Percent of Models with Package A Coefficients and Cannibalization (3/2)
Binary Flag	61,844	9,566	6,872	11.1%	71.8%
Log ("Package C" PO)	61,844	10,747	7,027	11.4%	65.4%
"Package C" PO	61,844	10,747	7,803	12.6%	72.6%
Binary Flag & "Package C" PO	61,844	10,747	9,670	15.6%	90.0%
Binary Flag x "Package C" PO	61,844	9,566	9,566	15.5%	100.0%

Table 5: Cannibalization of Package A Sales by Package C Promotions across all Customer A stores

A faculty advisor whose expertise is in statistics suggested that using the interaction term along with the independent variables themselves could potentially generate improved models. Regression analyses were performed manually in Excel for sample store-item combinations to test this hypothesis. However, using the other terms only marginally improved the predictive power of the models. It was decided that this slight improvement did not justify the additional complexity these terms layered onto the cannibalization models, particularly with regard to the implications this complexity would have for the eventual implementation of these models.

The equation for the models was therefore:

$$\log(volume_{PackageA}) = \alpha + PO_{PackageA} * coeff_{PackageA} + interaction\ term_{PackageC} * coeff_{PackageC}$$

where α is the intercept term and PO stands for "percent off". Taking the inverse of the logarithmic function on both sides yields:

$$volume_{PackageA} = e^{\alpha} * e^{PO_{PackageA} * coeff_{PackageA}} * e^{interaction\ term_{PackageC} * coeff_{PackageC}}$$

The demand forecast for Package A could be modified to account for cannibalization by reducing the volume of Package A using this equation, which will be discussed later in the implementation section of this document.

Once the interaction term had been identified as the key variable, the cannibalization models were further refined. Some experimentation was conducted to determine the optimum levels of Package C percent off for the binary flag. Originally, a Package C percent off level of 15% was being used to flag Package C promotions. Inspection of sample data revealed that a Package C percent off of 10% was the minimum level where a correlation to decreased Package A sales was observed.

There were a few general criteria that the cannibalization models had to meet in order to be considered valid. First, since the consumer goods company believed that its delivery data was more accurate than POS data, cannibalization models were only considered valid for those store-item combinations where the historical delivery data and POS data matched within 10% of each other. The second criterion was that there had to be a positive correlation between percent off and sales volume, which was indicated by the generation of a positive coefficient for the Package A term. This follows the basic principle of microeconomics that price and quantity are inversely related – that is, when price decreases (percent off increases), quantity increases. A third criterion was the presence of cannibalization, which was indicated by the generation of a negative coefficient for the Package C interaction term (which would offset the positive coefficient for the Package A term). These are the three criteria used in Table 5.

Since the cannibalization models were going to impact the suggested order, they should be statistically significant. To test this, minimum thresholds were set that resulting models had to meet in order to be considered in implementation. These were developed with guidance from a member of the consumer goods company's Supply Chain Selling Systems group whose expertise is in statistics, along with some data experimentation for what levels were realistically achievable. The final minimum thresholds used were as follows:

- EDOF (error degrees of freedom) > 10
- R-squared > 0.4
- F-statistic: p-value < 0.1
- t-statistic: p-value < 0.1

Note that the F-statistic tests whether the model is valid for *all* variables, while the t-statistic tests whether a particular variable is relevant or not. The results and implications of this analysis will be described in later sections of this document.

A similar SAS program was developed to model the cannibalization effect of Package M on Package A. However, when the program ran, SAS indicated that there was not enough data to generate statistically significant results. Examination of the data indicated that most flavors of Package M were not introduced until late May 2007 at Customer A. Since the consumer goods company's SAS database only contained data through mid-July 2007, only about seven weeks of data were available. Although the SAS database was eventually updated to contain data through the first week of September 2007, it turned out that Customer A's 2007 data was not suitable for regression. Given these circumstances, an alternative method of evaluation would have to be developed to analyze Package M's cannibalization effect on Package A.

5.3. *Visual Analyses*

A member of the consumer goods company's Supply Chain Selling Systems group recommended conducting a visual analysis of POS data for products where sales history was not long enough to generate a statistically significant regression. It was hoped that aggregating volume for multiple package types at the weekly level (starting when the new product was introduced) would indicate volume trends for those package types. The hypothesis was that the sales volume of existing package types would be decreased by the sales volume from the new product introduction. Volumes for each package of the same flavor (Flavor A) were converted to eight ounce servings in order to facilitate a consistent comparison across package types.

Due to several orders of magnitude difference between the Package A volume and the Package M volume, the aggregation method was deemed ineffective for visualizing volume trends. A member of the consumer goods company's Supply Chain Selling Systems group suggested using a simplified line chart instead that (1) rolled up data at the periodic level[1], (2) only included Package A and Package C sales (excluding Package K, for example), (3) included the previous year's Package A sales to facilitate comparison, and (4) used two different vertical axes for the different package types. The chart for all Customer A stores is shown in Figure

[1] The consumer goods company divided the year into 13 sales periods.

axes for the different package types. The chart for all Customer A stores is shown in Figure 19, while the chart for the Customer A store that sells the highest volume of Package A is shown in Figure 20. Note that the left vertical axis is the scale for Package A sales while the right vertical axis is the scale for Package M sales, both in terms of eight ounce servings. Unfortunately, this analysis was inconclusive regarding whether the new package type was affecting sales volumes of existing package types. The spike in Package M sales volumes during the fifth period does not appear to correlate with any significant reduction in Package A sales volumes around that period.

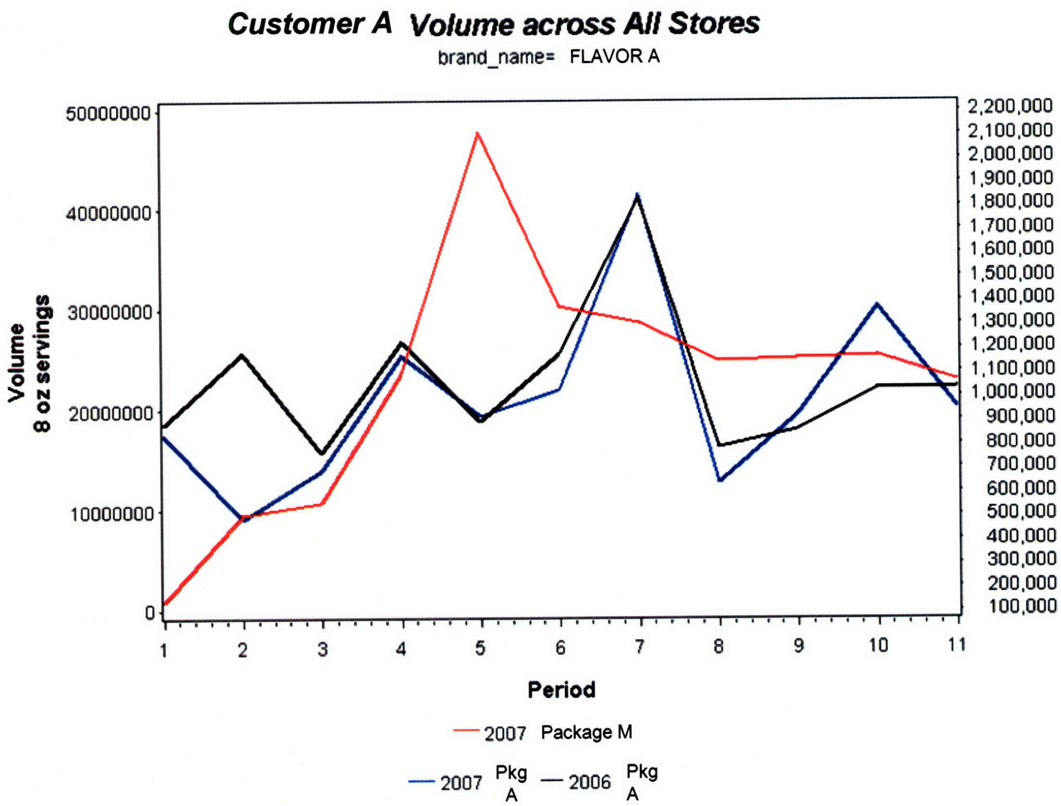

Figure 19: Effect of Package M on Package A Volume across All Customer A Stores

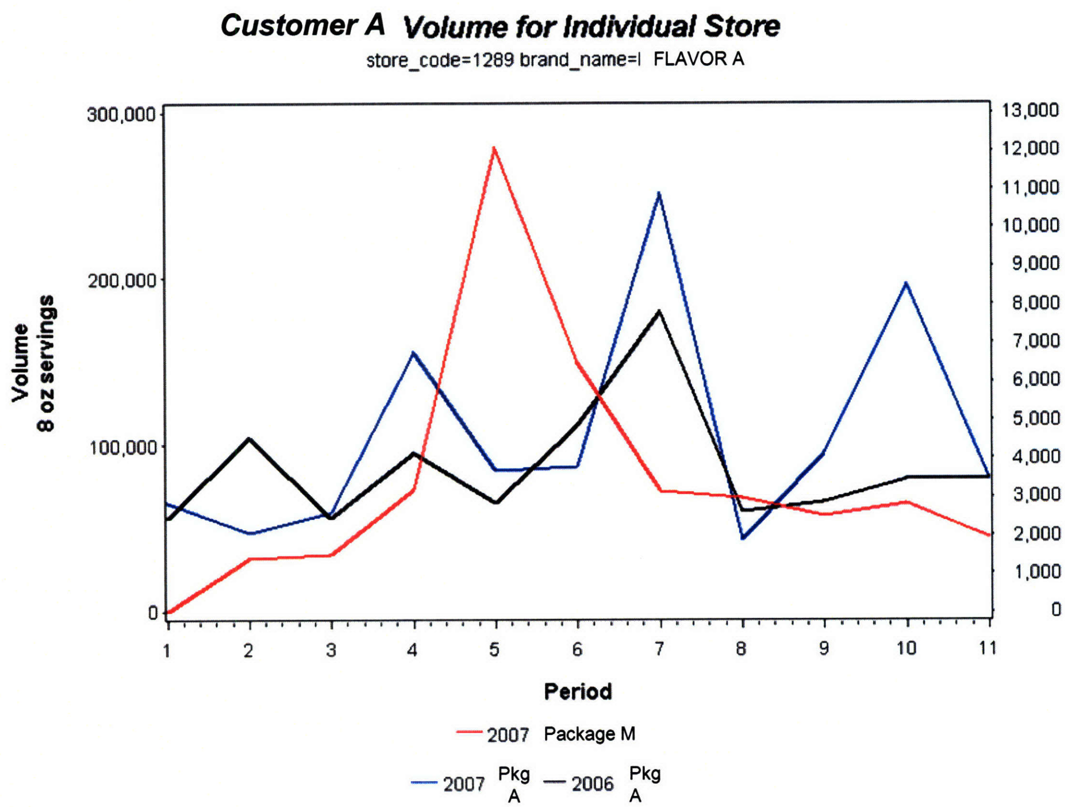

Figure 20: Effect of Package M on Package A Volume at Individual Customer A Store

It was postulated that perhaps the Package M was cannibalizing other package types, whose unit packaging had a similar form factor, as opposed to Package A, whose unit packaging had a different form factor. A discussion with a former field sales representative suggested that the most likely packages being cannibalized were the Packages E/G/I/N/O/P/Q/R/S. Therefore, a visual analysis of the effect of Package M on Packages E/G/I/N/O/P/Q/R/S (combined) was conducted. The chart for all Customer A stores is shown in Figure 21, while the chart for the Customer A store that sells the highest volume of Packages E/G/I/N/O/P/Q/R/S is shown in Figure 22. Note that the left vertical axis is the scale for Packages E/G/I/N/O/P/Q/R/S sales while the right vertical axis is the scale for Package M sales, both in terms of eight ounce servings. Unfortunately, this analysis was also inconclusive regarding whether the new package type was affecting sales volumes of existing package types. Again, the spike in Package M sales during the fifth period does not appear to correlate with any significant reduction in Packages E/G/I/N/O/P/Q/R/S sales volumes.

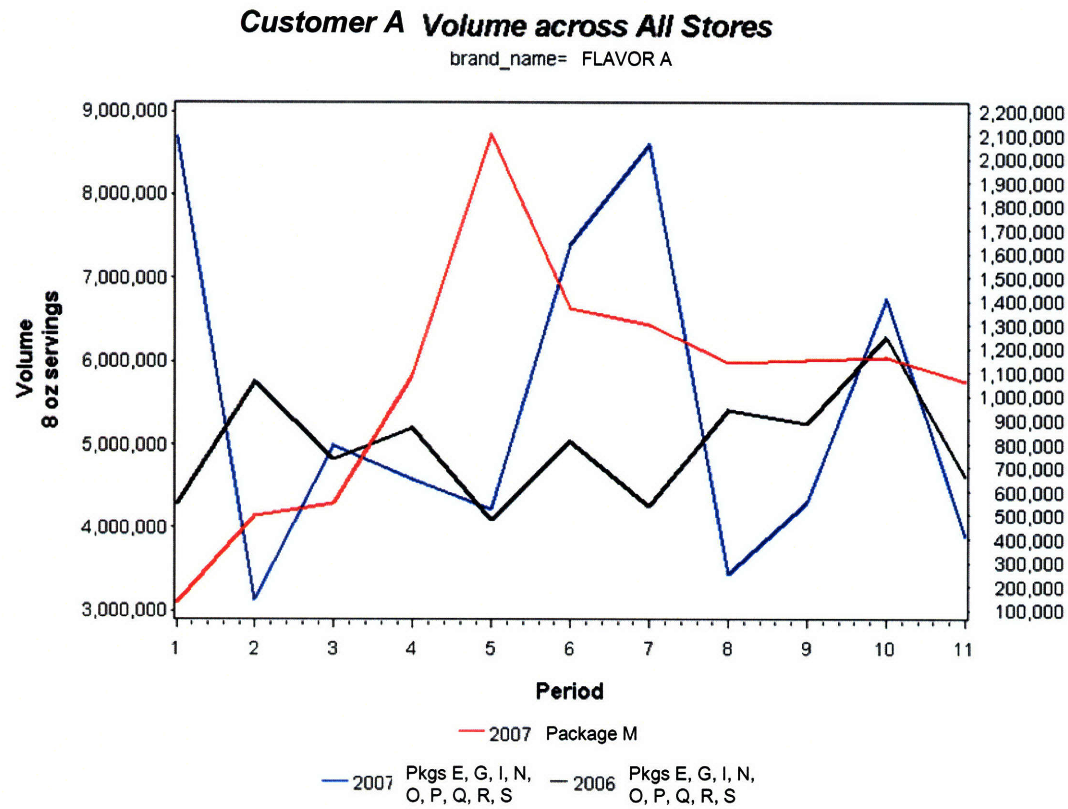

Figure 21: Effect of Package M on Packages E/G/I/ N/O/P/Q/R/S Volume across All Customer A Stores

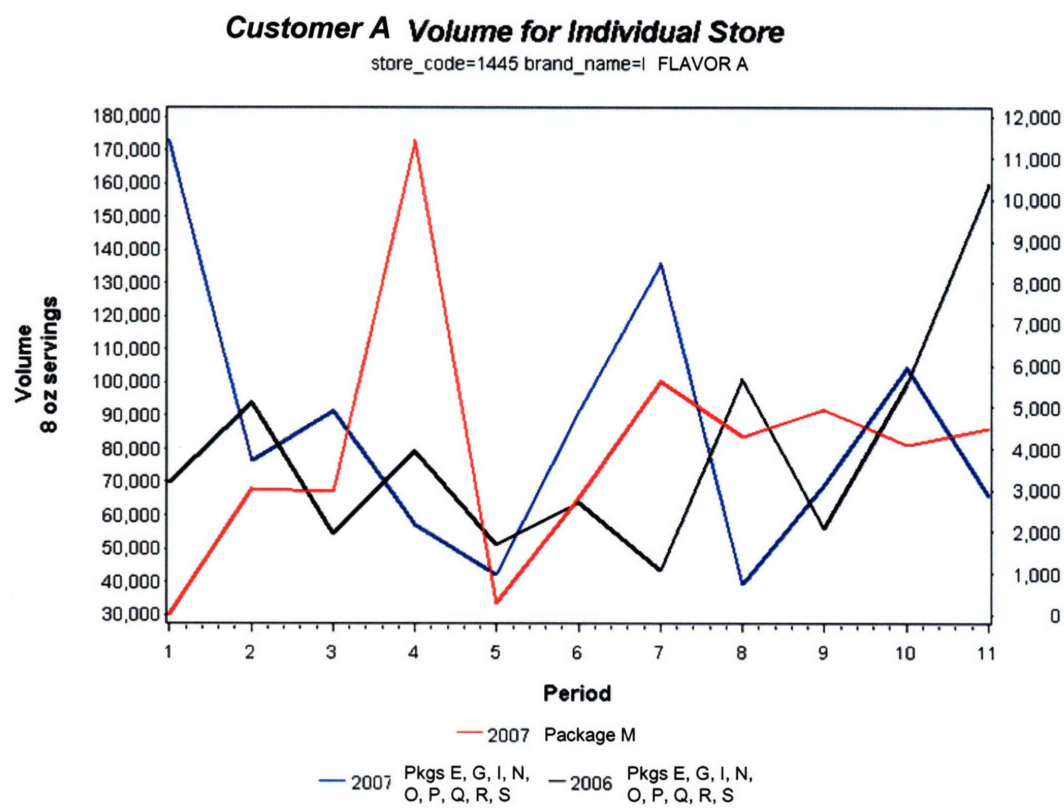

Figure 22: Effect of Package M on Packages E/G/I/ N/O/P/Q/R/S Volume at Individual Customer A Store

All of the visual analyses documented above were also run using the consumer goods company's delivery data but yielded similar results that were equally inconclusive. In an effort to understand why the cannibalization trends could not be detected visually, system-wide warehouse production data was obtained. Plotting the production volumes of Package A and Package M revealed that Package A volume was quite volatile and several orders of magnitude larger than Package M, as shown in Figure 23. A key trend that was discovered through analysis of this data was that although sales of packages with unit packaging having a similar form factor to Package A's unit packaging had been contracting overall, Package A sales were actually increasing while sales of other packages (such as Package T) were decreasing. Package A volume had increased 1.2% from 2005 to 2006 and was nearly flat (-0.1%) from 2006 to 2007, as shown graphically in Figure 24. This further substantiated that the introduction of the Package M did not appear to affect Package A sales, although it is plausible that the Package M might have displaced any growth that Package A volume would have experienced otherwise from 2006 to 2007.

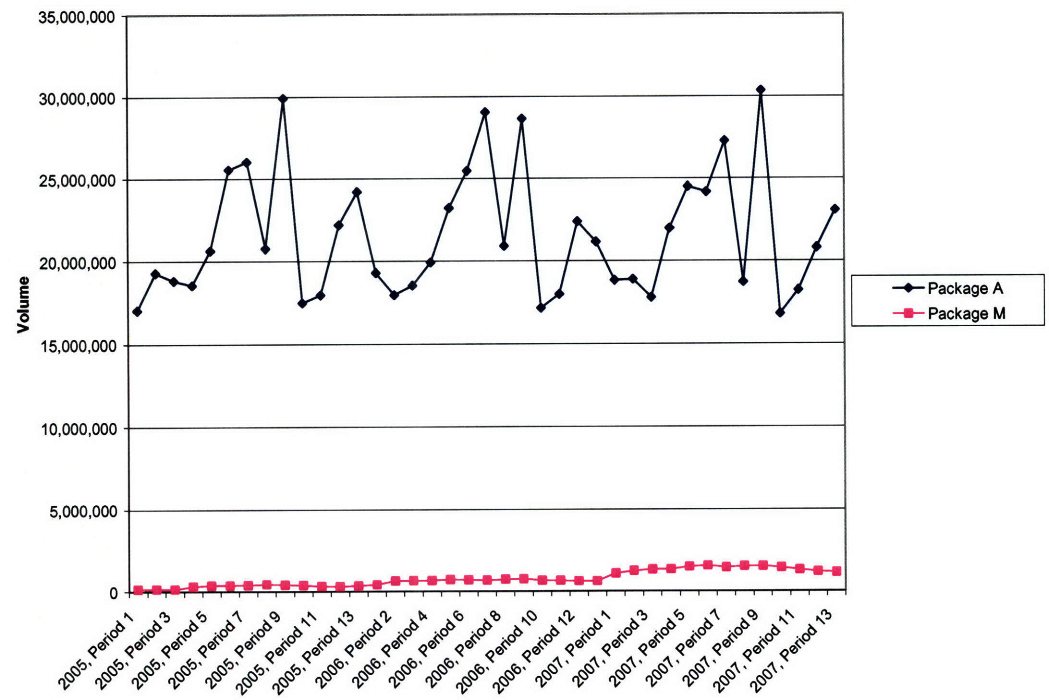

Figure 23: Company's 2007 System-wide Production of Package A and Package M

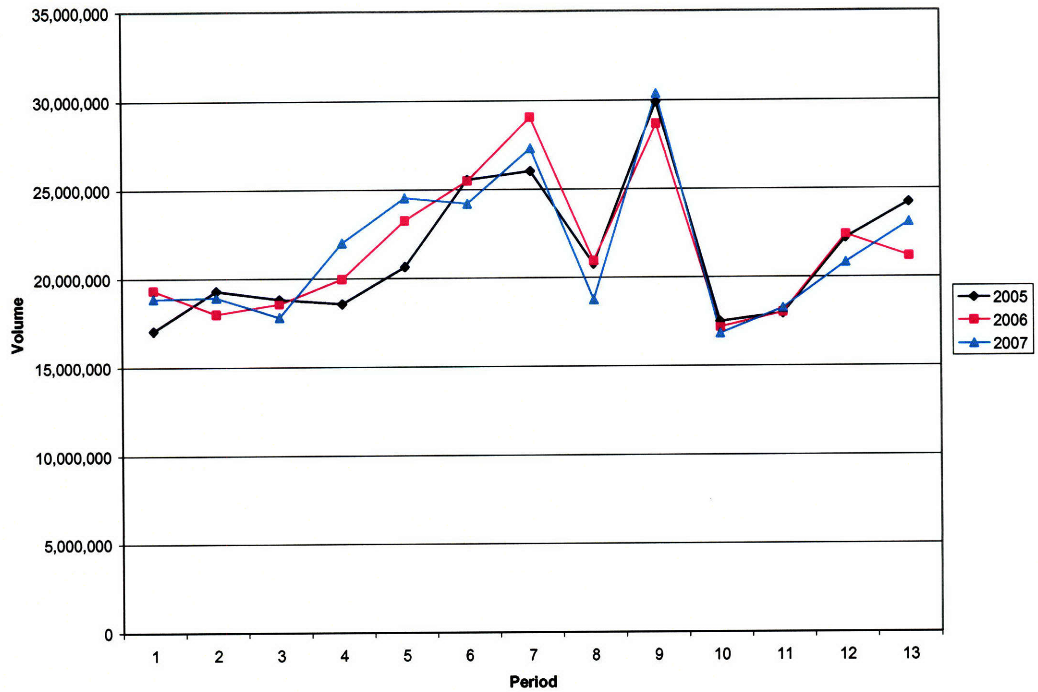

Figure 24: Company's Year-over-Year System-wide Production of Package A

5.4. *Measurement of Results*

The results of the regression analysis can be summarized as follows. From the store-item combinations whose delivery and POS data matched within 10%, 21.8% valid models were generated. Of the valid models, 19.2% of them were valid cannibalization models. The models generated indicated that a 30% off promotion on Package C typically decreases Package A sales by 40.6% (on average, weighted by volume). The next step in modeling the effect of cannibalization on demand for high volume products during promotions was to determine how to implement the findings.

The visual analyses done for new product introductions were summarized numerically, as shown in Table 6, to quantify the presence of cannibalization. Again, the results proved to be inconclusive regarding whether the new package type was affecting sales volumes of existing package types. Since there was no evidence of cannibalization from the visual analyses, no implementation would be considered for the effect of cannibalization on demand for high volume products during new product introductions.

Theory 1: Package M is cannibalizing Package A

231,088,428	Package A 2007 volume (Flavor A)
11,563,740	Package M 2007 volume (Flavor A)
242,652,168	**Total**
	versus
250,502,076	**Package A 2006 volume (Flavor A)**

Theory 2: Package M is cannibalizing Packages E, G, I, N, O, P, Q, R, S

59,967,288	Packages E, G, I, N, O, P, Q, R, S 2007 volume (Flavor A)
11,563,740	Package M 2007 volume (Flavor A)
71,531,028	**Total**
	versus
54,966,024	**Packages E, G, I, N, O, P, Q, R, S 2006 volume (Flavor A)**

Table 6: Numerical Cannibalization Analysis (All Volumes in 8 oz Servings, through 12-04-07)

5.5. Implementation

The consumer goods company plans to adjust the forecast for the cannibalization effect based on the findings of this work. The main implementation question is how to implement this on customer-item combinations where the company does not have POS data. This section briefly explores how such an adjustment might be implemented. It was postulated that the cannibalization effect might be correlated to the ratio of Package C to Package A (per store, on an annual basis), where the cannibalization effect is the percent reduction in sales on the cannibalized package (Package A in this case). Hence, the cannibalization effect and corresponding ratio were plotted for each of the store-item models developed as shown in Figure 25. One would have expected a clustering of points around a straight line if that were the case, but instead the chart exhibited scattered points, which did not support the hypothesis.

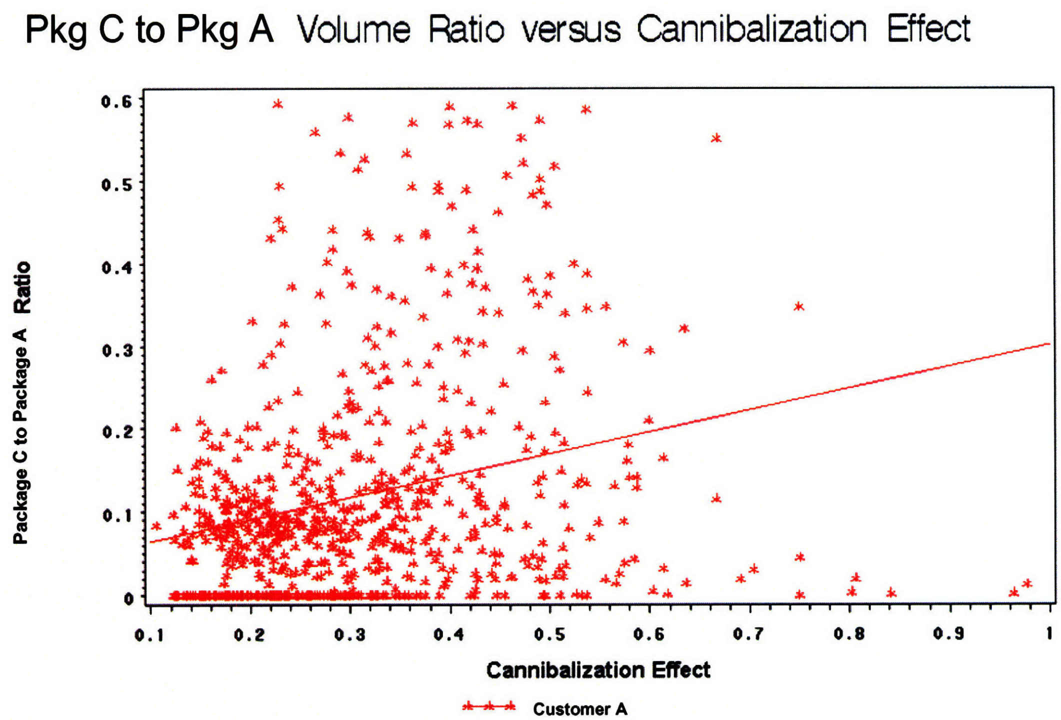

Figure 25: Package C to Package A Volume Ratio versus Cannibalization Effect

The next idea was to examine the distribution of the cannibalization coefficient to see if it might suggest an implementation strategy. The distribution is shown in Figure 26. The diagonally striped column contains the range of the most commonly occurring cannibalization coefficients, which suggests applying the mean of that range when Package C is present. The distribution of the cannibalization effect produced by these cannibalization coefficients was also considered and is shown in Figure 27.

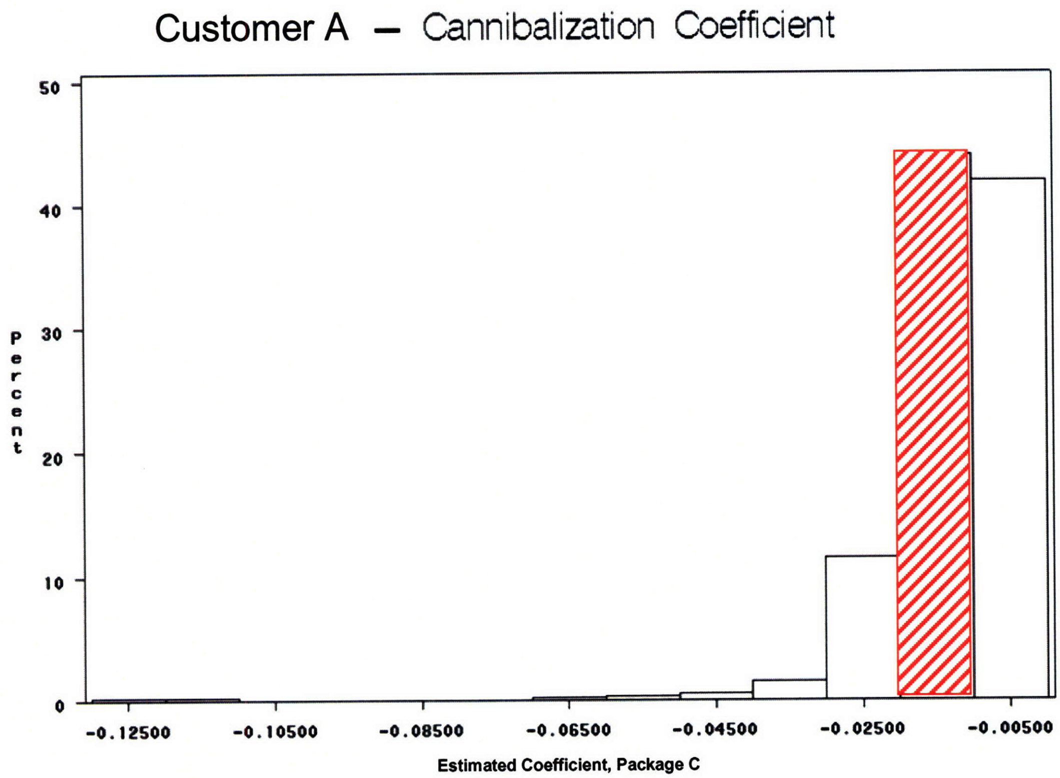

Figure 26: Distribution of Cannibalization Coefficient for Customer A

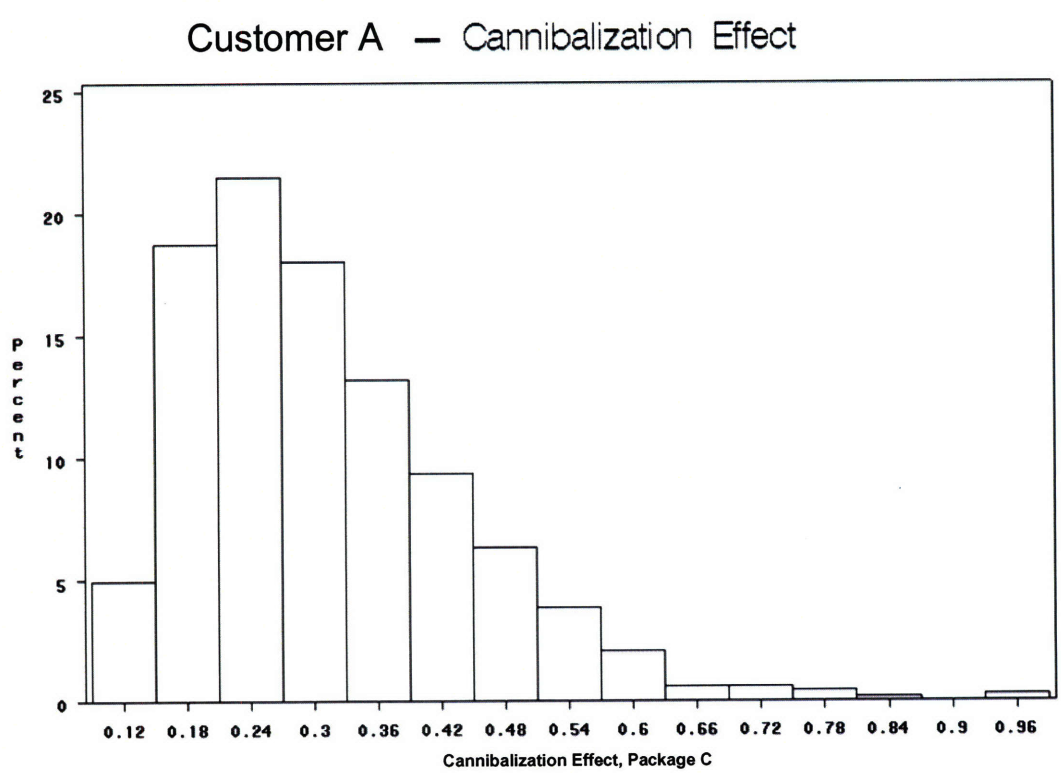

Figure 27: Distribution of Cannibalization Effect for Customer A

As previously mentioned, the equation for the cannibalization models was of the form:

$$\log(volume_{PackageA}) = \alpha + PO_{PackageA} * coeff_{PackageA} + interaction\ term_{PackageC} * coeff_{PackageC}$$

where α is the intercept term and PO stands for "percent off". Taking the inverse of the logarithmic function on both sides yields:

$$volume_{PackageA} = e^{\alpha} * e^{PO_{PackageA} * coeff_{PackageA}} * e^{interaction\ term_{PackageC} * coeff_{PackageC}}$$

One concern regarding modifying the demand forecast for Package A was that it would not be sufficient to multiply the volume by a reduction factor consisting of $e^{interaction\ term_{PackageC} * coeff_{PackageC}}$ because the cannibalization model also transformed the intercept and coefficient for Package A. Therefore, it was necessary to compare the intercept and coefficient for Package A from the cannibalization model to the intercept and coefficient for Package A from the model without cannibalization. (Regression was run twice in the SAS program to generate two models for each store-item combination.) The form of the latter model was:

$$\log(volume_{PackageA}) = \alpha' + PO_{PackageA} * coeff_{PackageA}'$$

Taking the inverse of the logarithmic function on both sides yields:

$$volume_{PackageA} = e^{\alpha'} * e^{PO_{PackageA} * coeff_{PackageA}'}$$

Note that the intercept term α' in this model differs from the intercept term α in the cannibalization model. Analogously, the coefficient term for Package A ($coeff_{PackageA}'$) in this model differs from the coefficient term for Package A ($coeff_{PackageA}$) in the cannibalization model.

A final chart was produced to investigate the relationship between the intercept and the Package A coefficient in order to understand whether and how either would have to be modified (either by inflating or by deflating) when the cannibalization coefficient for Package C was used to decrease the sales of Package A. This is shown in Figure 28 but was largely inconclusive given the scattered nature of the points on the chart.

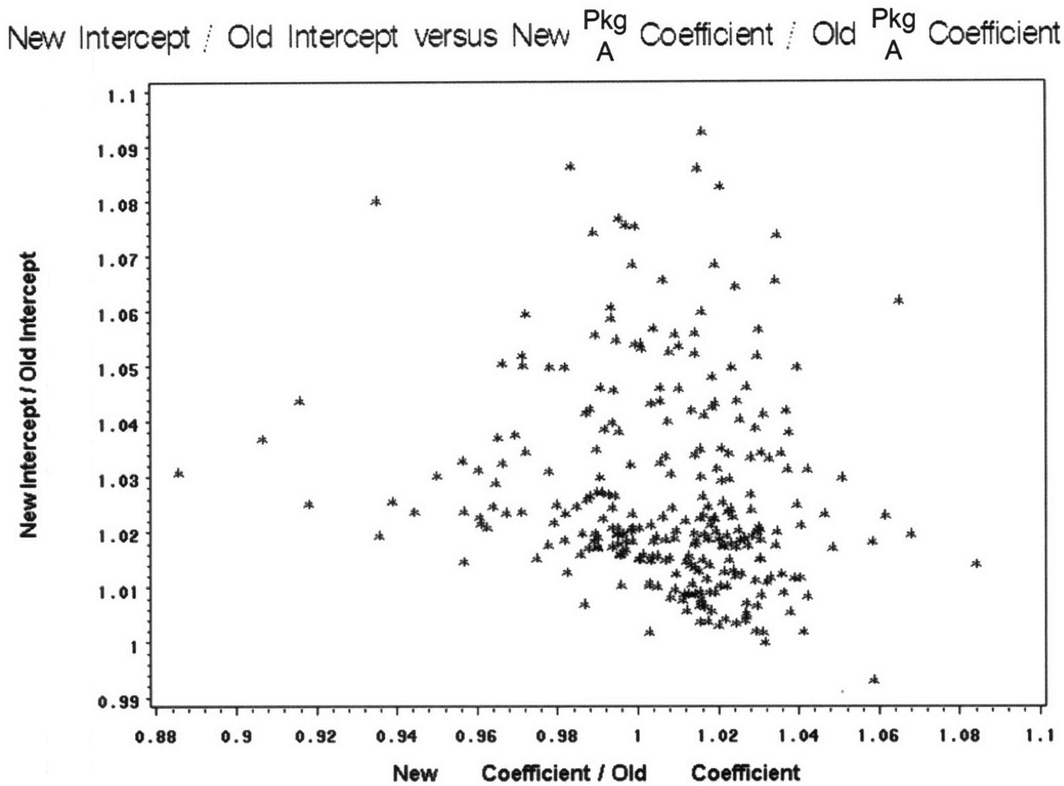

Figure 28: New Intercept / Old Intercept vs. New Pkg A Coefficient / Old Pkg A Coefficient

Given that the extensive analyses performed to determine an implementation strategy for cannibalization were inconclusive, the consumer goods company has a couple of options. The first option is to choose a conservative cannibalization coefficient to use in reducing the projected sales volume for Package A. This value could be sent down to the handheld to adjust the demand forecast for Package A, which would effectively modify the amount of Package A inventory being shipped to that particular store in anticipation of decreased sales due to cannibalization by Package C. Alternatively, it was discovered late in this project that a parallel effort was being conducted by a team in a partner company using Bayesian Regression to predict demand for multiple products, including competitors' products, with prices and other factors input by the user. This work could potentially be incorporated to modify the demand forecast of Package A when Package C is present in the store.

6. Observations and Recommendations

This section contains observations and recommendations treated in four areas: strengths, weaknesses, opportunities, and threats. Evaluation using these four criteria is commonly known as a "SWOT" analysis. In terms of strengths, this consumer goods company is constantly improving its handheld logic to more accurately reflect the reality of inventory. For example, the company was improving inventory handling by considering factors that were not accounted for previously. An additional strength of the company is that their approach is based on solid supply chain and statistical principles.

The biggest weakness of the consumer goods company is its data integrity, which was discussed in detail following the introduction of this document. Although the company strives to exclude mismatched historical delivery data and customer-provided POS data when analyzing data or building models, the sources of mismatch merit further exploration. Interpretation of data analysis would benefit from a better understanding of the nature and frequency of these errors, particularly scanner error and data feed error. This investigation would necessitate involvement and cooperation on the part of the company's customers. Scanner error could be measured by monitoring clerks in a customer's store for a specified period of time. The number of clerks should be chosen to provide a representative sample that would generate statistically significant results. An appreciation of data feed error could be accomplished by sitting with the customer employees responsible for maintaining the data.

Regular and ongoing data validation of data sets via statistical sampling should be performed, especially to ensure consistency between daily and weekly data sets. This data validation should include checking for duplicates and the presence of all dates. During the course of this project, SAS programs were developed to conduct data validation around duplicates and dates for Customer A and Customer B. These SAS programs can easily be extended to Customer C, Customer D, Customer E, and any other customers that may provide POS data in the future. Other external data sources that the company uses (e.g., those provided by industry organizations) should also be validated on an ongoing basis.

This consumer goods company has a strong heritage as a sales culture, but the behavioral incentives of sales and operations may at times be at odds with one another. For example, the sales reps are measured on their compliance with suggested orders, which are calculated to provide adequate inventory coverage over the demand period between deliveries. The company has improved sales reps' compliance from 40% to 70% (although this means that any improvements the Supply Chain Selling Systems group makes to the handheld logic, including this work, will only have a 70% impact). However, sales reps might choose to override suggested orders for legitimate reasons such as new product introductions, sales volume targets, customer store manager requests, and warehouse productivity (e.g., ordering full pallets or layers to facilitate pallet building at the warehouse). It is difficult to interpret the compliance measure given that the sales rep must balance so many factors simultaneously.

This consumer goods company has several opportunities besides the ones that emerge from the data integrity issue already addressed. Given the thin margins of its business, the company should aspire to create behavioral incentives that are consistent with the objectives of both sales and operations. Such a transformation will require getting buy-in from the rest of the organization (corporate, field sales reps, etc.) to redesign them from scratch in a way that is consistent the company's overall strategy. In doing so, the company also could strive to stabilize delivery quantities. The benefit of stabilizing delivery quantities would be to generate more accurate coefficients, which would lead to more accurate baseline numbers and ultimately, more accurate suggested orders. In addition, the company stands to save money from reduced fuel, handling, and carrying costs.

A final opportunity for this consumer goods company is that many groups are doing forecasting within the organization, including integrated planning, key accounts (national customers), finance, and possibly others. The company should explore whether it is possible to leverage the work being done by these other groups. At the very least, efforts could be coordinated so that forecasting is consistent. An added benefit would be that best practices could be shared among the groups.

7. Conclusion

This project laid the groundwork for some substantial improvements to this consumer goods company's demand forecasting at the store-item level. This section briefly outlines potential future work that the company could pursue in the areas of the baseline demand forecast and the effect of cannibalization on the forecast.

In terms of the baseline demand forecast, the consumer goods company should eventually develop coefficients for other key accounts (e.g., Customer B, Customer C, Customer D, and Customer E). Such coefficients would be more specific to these customers rather than applying Customer A's coefficients. The company also could create more granular clusters by using more delivery frequencies (e.g., five instead of three) or by identifying another cluster attribute (e.g., store size – large versus small). Another direction that the company could take this work is to use more sophisticated modeling to estimate POS data, for example by incorporating other factors besides delivery quantity (e.g., price, presence of an in-store display, etc.). Additionally, the company could use a more rigorous solver to develop coefficients, which it would probably need were it to pursue more sophisticated modeling.

On the cannibalization front, the consumer goods company should continue to seek evidence that Package M is cannibalizing Package A, which the multiple visual analyses performed during this project were not able to confirm. This is particularly relevant since the company will have nearly a year of sales history for Package M by the time this work is published. If the company is able to find such evidence, it should incorporate Package M into the regression analysis done to capture the cannibalization effect of Package C on Package A. This way, Package A's forecast can be adjusted for the presence of either or both of these package types (Package C and/or Package M). Furthermore, the Supply Chain Selling Systems group should continue to seek out anecdotes from field sales reps that will indicate what type of cannibalization they are seeing in customer stores. There is at least one other suspected cannibalization effect among several package types that merits investigation, for example Package E and others of that type (such as Package I, Flavor J) on Package A.

In conclusion, this work has gone a long way toward advancing this consumer goods company's demand forecasting at the store-item level. The development of a methodology that can convert delivery data into estimated POS data is a notable progression that is likely to provide vastly better baseline forecasts than the company's original daily delivery method. While the company will have to determine how to best generalize adjusting forecasts for cannibalization, this project did confirm that Package C is cannibalizing Package A and developed models at the store-item level to predict sales volume accordingly. This work also established a method of regression and a method of visual analyses that can be replicated to explore the existence of cannibalization among other package types. Finally, this document has left the company with a suggested set of next steps in each area, as well as ideas for potential future work that the company could pursue. This work has not only given the company improvements that can be implemented immediately but also established a foundation upon which the company can continue to build and expand upon to garner further enhancements in demand forecasting at the store-item level.

This page has been intentionally left blank

Bibliography

Bertsimas, Dimitris and Robert M. Freund. *Data, Models, and Decisions: The Fundamentals of Management Science.* Belmont, MA: Dynamic Ideas, 2004.

Datta, S., Granger, CWJ, Barari, M., and T. Gibbs. "Management of supply chain: an alternative modeling technique for forecasting." *Journal of the Operational Research Society* 2007: 1-11.

Devore, Jay L. *Probability and Statistics for Engineering and the Sciences*. Fifth ed. Pacific Grove, CA: Thomson Brooks/Cole, 2000.

Hubert, Mia, Rousseeuw, Peter J., and Stefan Van Aelst. "High-Breakdown Robust Multivariate Methods." Academic Paper. August 9, 2006.

Klein, Janice A. *True Change: How Outsiders on the Inside Get Things Done in Organizations*. San Francisco, CA: Jossey-Bass, 2004.

Kowalski, Scott M., and Geoffrey Vining. *Statistical Methods for Engineers*. Second ed. Boston: Thomson Brooks/Cole, 2006.

Simchi-Levi, David, Philip Kaminsky, and Edith Simchi-Levi. *Designing & Managing the Supply Chain: Concepts, Strategies, & Case Studies*. Second ed. New York, NY: McGraw-Hill/Irwin, 2003.

Welsch, Roy E. "Lecture Notes Part 8 – Time Series" *15.064J/ESD.751J Engineering Probability and Statistics* Summer 2007.