DEMAND MANAGEMENT IN GLOBAL SUPPLY CHAINS

A Thesis
Presented to
The Academic Faculty

by

Evren Özkaya

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2008

DEMAND MANAGEMENT IN GLOBAL SUPPLY CHAINS

Approved by:

Professor Pınar Keskinocak, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor John H. Vande Vate, Co-advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Paul Griffin
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Mark Ferguson
College of Management
Georgia Institute of Technology

Professor Julie L. Swann
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 18 September 2008

To my beloved parents,

Semra and Mustafa Özkaya,

and my brother and best friend,

Engin Özkaya,

for their eternal love and support.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Pinar Keskinocak for her tremendous support since the first days of my Ph.D. study and for her continuous motivation and trust at every step of my research. I would like to give my special thanks to my co-advisor Dr. John Vande Vate for creating an extraordinarily beautiful and supportive environment for research. I am honored to be a part of the H. Milton Stewart School of Industrial and Systems Engineering doctoral program and I want to thank each an every one of its renowned faculty.

I would like to take this chance to thank my proposal and thesis committee. I appreciate their invaluable time and effort in giving me guidance and support. I would like to thank Dr. Roshan Joseph Vengazhiyil for his great contributions to the second chapter of this thesis. For the same study, I would like to thank Ryan Weight and Schneider Logistics, Inc. for their support and for letting me access their industry data. For making my Ph.D. study a truly exceptional one, I would like to extend my special thanks to Michael B. Waithe and Global Supply Chain Modeling and Solutions group at Intel Corporation; I really enjoyed the opportunity of working with great colleagues.

Finally, I am deeply grateful to my mom, dad and brother for always being by my side even though they are thousands of miles apart from me. My heartfelt gratefulness goes to my dear wife Banu, my companion and my true love. I would have never made this far without her unconditional love, support and friendship. Thank you for everything you endured with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In this thesis, we investigate the potential of improving demand management activities in the global supply chains. In the increasingly global world, commerce is becoming more complex with an incredible amount of internal and external information available for businesses to select, analyze, understand and react. We identify opportunities for companies to convert data and business information into actionable intelligence.

We first study the logistics industry with real data. In the Less-than-Truckload (LTL) market, we analyze an extensive historical shipment database to identify important factors to estimate LTL market rates. Quantifying critical expert knowledge, we develop a price estimation model to help shippers reduce their logistics cost and carriers to better manage their demand. In our second study, we analyze a global supply chain in the high tech industry. Using the demand dependency structure of certain products, we identify collaboration opportunities in the ordering practices that results in increased forecast accuracy. In our third study, we focus on using historical product adoption patterns for developing good pre-launch forecasts for new product introductions. Through a normalization approach and algebraic estimation procedures that use intuitive parameters, our models provide opportunities to significantly improve pre-launch forecast accuracy. Finally, in our fourth study, we develop novel approaches for modeling and mitigating the impact of demand seasonality in new product diffusion context. Focusing mainly on practical applications, our research shows that companies can find innovative ways for turning raw data into valuable insights leading to better demand management activities.

CHAPTER I

INTRODUCTION

The Supply Chain and Logistics world is in the middle of a big transformation. This transformation is like Darwin's natural selection theory, which eliminates companies that are lagging behind the use of technology, real time information and data-based decision support systems to give operational, tactical and strategic decisions. In the internet era, what is lacking is not the data or information, but is the ability to convert this abundant data and information to meaningful knowledge that helps foster good decision making. The following research topics are the attempts of establishing such data-based decision making tools into supply chain and logistics functions of the global companies.

1.1 Less than Truckload (LTL) Market

With the globalization of supply chains in the manufacturing and retail industries, there is an increasing need for faster delivery of smaller shipments at lower cost. Less-than-Truckload (LTL) is a mode of transportation that serves this need by handling shipments smaller than full truckload and larger than small package. LTL is a $34 billion industry in the U.S. and LTL freight is priced significantly higher per unit weight than truckload (TR) freight. Given the trade off between higher service levels and higher cost of LTL shipments, LTL purchasing managers increasingly focus on getting better rates from the carriers. Lack of transparency and complex discount practices make the less-than-truckload (LTL) market rates a challenging piece of business information both for carriers and shippers. In Chapter 2 of this thesis, we present a regression-based methodology that can estimate the Less-than-Truckload (LTL) market rates with high reliability using an extensive database of historical shipments from the continental United States. Our model successfully combines the quantitative data with qualitative market knowledge to produce better LTL market rate estimations that can be used to produce benchmark studies allowing carriers and shippers to identify cost saving opportunities. Model results also outline and rank the important factors that affect LTL pricing.

1.2 Platform Products Supply Chains

In Chapter 3, we investigate a form of inter-enterprise supply chain collaboration by exploring the value of demand information for platform products and the use of this information in an intra-enterprise vertical collaboration setting. In the semiconductor industry, suppliers deliver multiple components to the OEMs at different times that are then assembled as a single platform (i.e. Personal Computer) by OEMs. We analyzed scenarios where we can use this component demand dependency on quantity and timing as advance demand information to improve forecasting accuracy. Our approach investigates the benefits of this advance demand information (given from customer to supplier) under stochastic demand scenarios using a Monte Carlo approach. We developed an integrated supply chain simulator to quantify the potential benefits of the resultant forecast improvement on Intel Corporation's global supply chain, where the forecast improvements are used vertically within the company to improve supply chain efficiencies.

1.3 Pre-launch Forecasting using New Product Diffusion Models

In global economy, growing demand combined with increasing rates of innovation in both product and process technologies drive shorter product life cycles. Forecasting product life cycle behavior with the limited information available prior to the product launch remains one of the most important and difficult challenges faced by many businesses as they introduce new products. Diffusion models are good candidates for providing comparisons of potential market penetration scenarios using historical product adoption patterns.

In Chapter 4, we present a practical framework for the analysis of historical product diffusion patterns and propose several methodologies for algebraically estimating new product diffusion parameters. We introduce user-friendly versions of the classic Bass Diffusion Model with new sets of parameters that are more intuitive and have natural interpretations in terms of more easily estimated market characteristics. We test our models on high tech industry data sets and report significant forecast improvement opportunities.

1.4 Seasonality Considerations for Diffusion Models

In forecasting new product diffusions with short life cycles, seasonality plays a significant role. Seasonal data series have not been widely used in the diffusion modeling context as the majority of the studies focus on macro-level diffusion models that use annual data. Increasingly, managers need to forecast new product diffusions at more granular level both in product and time dimensions. Product-level diffusion models that aim to produce monthly or quarterly demand forecasts, therefore require the proper treatment of seasonality factors.

In Chapter 5, we analyze the impact of seasonality on new product diffusions and propose models to improve forecast accuracy through better estimation of seasonality factors. We propose two novel approaches for better identifying and removing seasonality from the data series. Under both simulated data and real data, we show that we can significantly improve seasonality factor estimates, especially for short data series with nonlinear trend and high random error variance, resulting in improved potential for higher forecast accuracy.

CHAPTER II

ESTIMATING AND BENCHMARKING LESS-THAN-TRUCKLOAD (LTL) MARKET RATES: A TOOL FOR LTL MARKET VISIBILITY

2.1 Introduction

Today's competitive marketplace requires companies to operate at low-cost, which increases both the importance of the market knowledge and the price companies are willing to pay for acquiring such knowledge. According to the 18th Annual State of Logistics report by Council of Supply Chain Management Professionals, the logistics costs add up to $1.31 trillion in the U.S. in 2006, which constitutes 10% of the US GDP of the same year, following an increasing trend since 2003 when the logistics costs were $1.01 trillion with 8.6% of the US GDP (23). In order to reduce logistics cost, shippers are trying to gain a better understanding of the market rates offered by the carriers or other logistics providers for their services. Negotiations become an important part of cost savings in this Business-to-Business (B2B) market environment.

With the globalization of supply chains in the manufacturing and retail industries, there is an increasing need for faster delivery of smaller shipments at lower cost. Less-than-Truckload (LTL) is a mode of transportation that serves this need by handling shipments smaller than full truckload and larger than small package. LTL is a $34 billion industry in the U.S. and LTL freight is priced significantly higher per unit weight than truckload (TR) freight (Shultz (2007) (74)). Given the trade off between higher service levels and higher cost of LTL shipments, LTL purchasing managers increasingly focus on getting better rates from the carriers. However, often times neither a customer nor an LTL carrier knows how the offered rates compare to the other rates for similar shipments in the industry. Since every shipper-carrier pair contract their own rates based on many parameters, the knowledge of "market rates" requires historical shipment data from a variety of shippers and carriers and a systematic process for analyzing the data.

The LTL mode differentiates itself from the other modes, because the shippers do not pay for the entire truck/container cost based on "rate per mile", but they pay only a portion based on their own freight. The LTL carriers are therefore interchangeably called "common carriers" in the transportation industry. In LTL, since shipments belonging to different shippers are carried in one truck, the pricing structure is much more complex compared to truckload (TL) shipments. For carriers, it is a very challenging task to estimate what the real costs are for different loads. LTL carriers use a transportation network with break-bulk facilities and consolidate LTL freight to a full-truck-load or break a full-truck-load into local deliveries. These facilities incur extensive handling and planning costs, which are hard to track down to ration to each shipper. In order to simplify the pricing structure, the carriers use industry standards called "tariffs". Based on these tariffs (such as "Yellow500" and "Czarlite") the freight is priced based on its origin-destination (O-D) zip codes, its freight class (i.e., freight class ranges from 50 to 500) and its weight (150 lbs to 12000 lbs). However, these tariffs are often used as a starting point for negotiations and the carriers usually offer steep discounts (generally between 50-75%) from the tariffs.

The main goal of this study is to develop an analytical decision-support tool to estimate LTL market rates. To the best of our knowledge, such a tool currently does not exist in the industry. Having market rate estimates which consider various factors such as geographic area, freight characteristics and relative market power of the shipper (or carrier) will help shippers better understand how much they currently pay with respect to the market and why, and whether there are opportunities for cost savings. Shippers can also use these estimates in their network design studies as a source of reliable LTL prices for the proposed new lanes. On the other hand, carriers would benefit from market rate estimates in pricing their services. A similar-purpose analytical benchmarking model, Chainalytics Model-Based Benchmarking (MBB), has been developed by Chainalytics, a transportation consulting firm, for the long-haul Truckload and Intermodal moves. MBB analyzes the cost drivers for the realized market rates of shippers that form a consortium and share shipment data. "MBB only shares information regarding the drivers of transportation costs - not the actual rates themselves. ... [It] quantifies the cost impact of operational characteristics." (18). In our analytical model we not only quantify the impact of captured tangible factors, but also analyze non-captured market information and provide the full view of the cost drivers of LTL market rates. Our results show that qualitative expert inputs - which can be crucial in LTL shipments - can be quantified and used in the econometric models to further improve the market rate estimations. Suggested methodology can also be applied to other Business-to-Business (B2B) markets for improving price estimations with qualitative market information.

2.2 Literature Review

In the Less-Than-Truckload business, rates continue to rise and with the sharp increase in fuel charges in recent years, LTL purchasing managers are looking for different ways to reduce cost (Hanon 2006a (33)). A recent poll conducted by Purchasing Magazine among LTL buyers reveals some suggestions to reduce costs such as using standardized base rates, always asking for discounts and using online bidding tools to level LTL rates with competitive market rates. The deregulation of LTL industry with the Motor Carrier Act of 1980 brought today's complex pricing structure. Leaving carriers free for setting any discount levels, shipment rates started to be called with their percent discounts off of the carrier set base prices. Later, fuel charges sky rocketed when added as an additional surcharge on top of the LTL rate. Fuel surcharge is stated to be a major problem across the industry, which originally emerged to protect carriers from sudden increases of fuel costs. However, the LTL industry lacks a standard fuel surcharge program. FedEx CEO, Douglas Duncan, states that there is "inconsistency in pricing in the LTL market since deregulation in 1980. Every customer has a different idea of pricing in their base rates and surcharges" (Hanon 2006b (34)). Grant and Kent (2006) ((29)) survey the methods used by LTL carriers to calculate fuel surcharges. Extra services provided by LTL carriers such as pallet handling are charged separately under accessorial charges. Barett (2007) ((7)) explains the evolution of free market into a very complex pricing structure: "Things soon got out of control in the newly invigorated competitive marketplace. Carriers bulked up their base rates outrageously, to support more and more increases in those customer-attracting discounts, and the process became self-perpetuating. Thus it is that discounts in the 70th, even the 80th percentile have become the order of the day now." Recently, there are further attempts to remove remaining regulations on the LTL industry. Surface Transportation Board (STB) (formerly, Interstate Commerce Commission) decided on May 7, 2007 (Ex Parte No. 656) to remove anti-trust immunity previously enjoyed by LTL carriers who met at the National Motor Freight Committee (NMFC) meetings to set classification of goods or at rating bureaus. The impact could potentially eliminate the NMFC or freight classification of commodities. Future studies could replace freight class with freight density, which are highly correlated. However, there is no final decision on this major change yet (Bohman 2007 (14)).

While there has been little research done to analyze industry practices in pricing the LTL services, researchers investigated other interesting aspects of the LTL industry. On the operational side, Barnhart and Kim (1995) ((6)) analyze routing models for regional LTL carriers. Chu (2005) ((22)) develops a heuristic algorithm to optimize the decision of mode selection between truckload and LTL in a cost effective manner. Katayama and Yurimoto (2002) ((45)) present a solution algorithm and corresponding literature review for LTL load planning problem for reducing LTL carrier operating costs. Hall and Zhong (2002) ((32)) investigate the equipment management policies of the long haul LTL shipments. Murphy and Corsi (1989) ((65)) model sales force turnover among LTL carriers. Chiang and Roberts (1980) ((20)) build an empirical model to predict transit time and reliability of the LTL shipments. An operations model constructed by Keaton (1993) ((46)) analyzes and reports significant cost saving opportunities in economies of traffic density. Many mergers and acquisitions in the LTL industry can be explained by this potential cost savings opportunity. On the pricing side, research shows that auctions or bidding projects are important ways of reducing transportation costs. Elmaghraby and Keskinocak (2003) ((26)) analyze combinatorial auctions focusing on an application by The Home Depot in transportation procurement and reported significant cost savings. Smith et al. (2007) ((77)) analyze a US LTL carrier's shipments with statistical models to estimate revenues from different customers at different lanes. They compare the regression estimated expected revenues with actual revenues to identify opportunities for re-negotiations when there is a systematic difference in estimated and actual revenues. However, their models do not estimate market rates at individual shipment level, and their analysis is limited to the single carrier's dataset. Market rate estimation requires a diverse set of carriers and shippers with diverse freight characteristics. Although many articles such as Baker (1991) ((4)) and Centa (2007) ((17)) reveal that there are many factors considered in the LTL pricing and it is a complex mechanism of convoluted relationships and considerations, we find no study so far that attempts to analytically model the LTL pricing structure and estimate individual LTL shipment market rates. Our research is aimed to fill this gap by analyzing LTL industry data with statistical methods to provide market rate estimates for the US LTL shipments. Our paper focuses on estimating the total line haul cost of transportation that excludes the fuel surcharges and additional accessorial charges.

2.3 Problem Definition

The LTL Market is fragmented among hundreds of carriers, which are generally grouped into 3 major categories based on the area they serve: regional, super-regional and national. The pricing structure of the LTL market is mostly based on contracts signed by these carriers and the shippers. Unlike the small package (parcel) carriers, one cannot check the prices online from the web sites of major carriers (i.e., UPS, FedEx, USPS) and find out the best price for a specific shipment. Even for the major LTL carriers, prices are negotiated and contracted for at least 1 or 2 years. Hence, the LTL prices are mostly hidden between the corresponding shipper and carrier. The same LTL carrier most likely has different negotiated prices for different shippers based on the desirability of the freight, suitability of the freight to the carrier's current network, negotiation power of the carrier relative to the shipper and many other factors. Under this complex pricing structure, our objective is to create a robust model that can reliably estimate LTL market rates for all possible continental US shipments at any given time, freight class, weight and other factors.

The first challenge to achieve this objective is to obtain enough market data that has sufficient diversity in freight class, in Origin-Destination pairs (lanes), as well as in carrier-shipper pairs, in order to represent the current market dynamics. Also the market data has to have some important LTL shipment information such as origin-destination, freight class, weight, shipper-carrier information, etc. that has direct or indirect effect on the final price.

Using Schneider Logistics' extensive LTL market database, we obtain detailed information about each LTL shipment that can be used as potential predictors. Our second challenge is to analyze the pricing structure and find consistent relationships of final price with the limited dependent variables under the sparse data reality.

Data: We use a dataset of shipments from February to April 2005 containing information about origin and destination zip codes, cities and states, the carrier and shipper names, the weight, the class of the freight, the total line haul price paid to the carrier, the unique bill of lading number, and some other operational information such as date of shipment.

The cleaned dataset contains $90 million worth of LTL transactions incurred by 485 thousand shipments during these 3 months, spanning the freight of 43 shippers moved by 128 carriers that covers 2126 state-to-state lanes (92% of all possible state-to-state combinations inside the continental USA excluding Washington DC). Cleaning procedure involves removing shipments that are missing or have erroneous crucial information such as Zip code, freight class and weight. Also, the dataset is filtered to include shipments that are within reasonable LTL market bounds such as the weight to be between 100-12000 lbs (higher then 12000lbs is generally more cost effective with Truckload shipments) and minimum discount is set to 40%. The discount levels range from 50-75% for the majority of the shipments. With regards to lane coverage, diversity of transactions among freight class and diversity of carrier-shipper combinations, we used one of the most extensive data set available in the industry.

Seasonality: LTL prices are negotiated for long term (i.e., one or two years) and contracted. Price contracts cover the entire contract period with the same negotiated prices without including possible seasonal changes of the logistics costs. Therefore, the seasonality is generally not part of the LTL pricing (excluding the fuel surcharges that are seasonally affected by constantly changing fuel costs). Seasonality might be present in the spot market for last minute services such as expedited shipments during the holiday season. However, for our study we are not considering the seasonality effect of the LTL market rates. Since we also have only 3 months of data, we may not observe any seasonal effect in other parts of the year.

Descriptive Statistics: More detailed statistics are available in Appendix C.1. The majority of the LTL shipments in the dataset are priced between \$65 and \$1100; however, there are also higher priced LTL shipments up to \$3000-4000 level. The average distance of the shipments is 933 miles with an average weight of 1713 lbs. Freight class changes from 50 to 150 for different types of freight including general merchandise, industrial lubricants, cereal, automotive spare parts and other goods. For some of the US regions (i.e., New York City, South Florida, Rocky Mountains), the cost of LTL shipments might be higher or lower due to special reasons such as congestion (New York City), supply-demand mismatch (South Florida) and level of urbanism (Rocky Mountains).

Sparse data: Although we have an extensive database of LTL market transactions with half a million records, almost no two shipments are the same in details. For example, there exists only one LTL shipment from Atlanta, GA to Chicago, IL of class 85 that weighs between 1000 lbs and 2000 lbs. Hence, estimating the LTL market rate for this level of detail is not reasonable with this one shipment.

Table 1 and Table 2 show example summaries of the shipments from Atlanta, GA indicating sparse data even at a higher detail level. In Table 2, there are no shipments from Atlanta to New York City; and only 2 shipments to the State of New York. There are 18 different freight classes and 6 different weight-brackets (weight interval where the unit LTL base price is the same) and 2304 state-to-state origin-destination couples (not even going into Zip code level detail). The combination of these three major characteristics creates 250,000 different shipment types, leaving less than 2 shipments on average per combination. This excludes the carrier-shipper details, city and zip code level details, and the geographical area of the shipment.

Next, we present our statistical model that uncovers some important LTL pricing characteristics.

Table 1: All shipments from Atlanta to Chicago of class 85.

O City	**O St**	**D City**	**D St**	**Class**	**Wt Bracket**	**Shipments**	*Avg.* **c/cwt**
Atlanta	GA	Chicago	IL	85	100 to 300	2	$ 24.96
Atlanta	GA	Chicago	IL	85	300 to 500	5	$ 17.09
Atlanta	GA	Chicago	IL	85	500 to 1000	3	$ 13.67
Atlanta	GA	Chicago	IL	85	1000 to 2000	1	$ 12.25
Atlanta	GA	Chicago	IL	85	2000 to 5000	1	$ 10.07

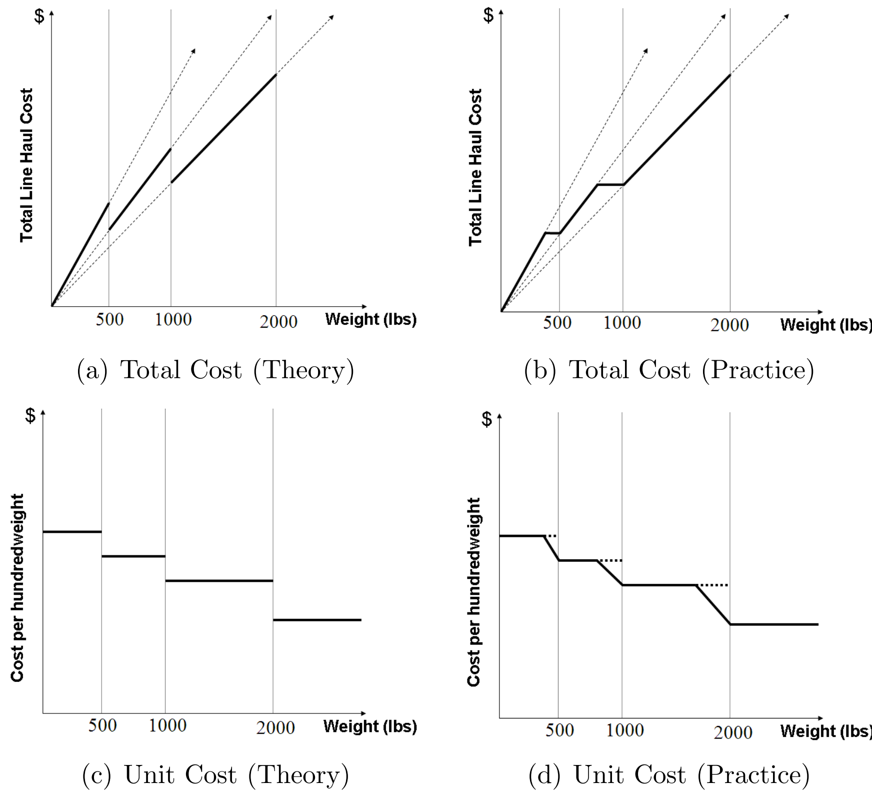
Table 2: All shipments from Atlanta to State of New York.

O City	**O St**	**D City**	**D St**	**Class**	**Wt Bracket**	**Shipments**	**Avg. c/cwt**
Atlanta	GA	New York	NY	All	All	0	N/A
Atlanta	GA	All Cities	NY	60	All	2	$ 17.50

2.4 Modeling Approach

The currently realized LTL shipment rates are based on the contracts between the shippers and the carriers. In contract negotiations many different factors are taken into account, some of which are captured and/or easily calculated such as Freight Class, origin and destination zip codes, weight and mile. Some others also affect the price significantly but generally are not captured in the data; for example negotiation power (i.e., if the company is using a third party logistics company with combined purchasing power), freight desirability (i.e., whether the freight is stackable or palletized, whether the drivers need to wait long times to get the freight, etc.) or the economic value that the shipper receives from this LTL service.

Our approach breaks down the above factors in two categories, namely, tangible and intangible; and then formulates a multiple regression model using both type of factors to estimate the total LTL service price for the specific shipments. For the intangible factors, using expert knowledge we develop a scorecard methodology that captures the information in a score that impacts the final price. LTL experts could be able to evaluate (i.e., score) the majority of the shippers in the dataset that corresponds to 75% of the total shipments. Therefore for our model, we use the scored dataset that contains 363 thousand shipments. More details will be given on



(a) Total Cost (Theory)
(b) Total Cost (Practice)
(c) Unit Cost (Theory)
(d) Unit Cost (Practice)

Figure 1: LTL Linehaul cost pricing and discount schedule. All unit discount scheme based on weight for a given Origin-Destination zip code and freight class.

this market scorecard methodology in the following sections.

2.4.1 LTL Pricing: Discounts and Minimum Charge

In the LTL Market, in order to simplify the pricing and contracting process, the carriers are using industry standard tariffs. These tariffs are basically tabulated market rates that give the rate according to freight's origin-destination (O-D) zip codes, its freight class and its weight. There are few industry-wide tariffs that are the most commonly used; however there are 292 internal tariffs that are being used today according to Material Handling Management online newsletter (58).

Percent Discount: is a discount offered by the carriers off of the posted tariff base price.

Weight Discount: is an all-unit discount scheme with breakpoints currently set at 500 lbs, 1000 lbs, 2000 lbs, 5000 lbs and 10000 lbs. Figure 1 presents how the weight discounts affect the total and unit base prices posted by the tariffs.

Minimum Charge: is the base price set by the carrier (usually between $40 and $80) for a specific O-D pair and freight class, such that any shipment that is rated (discounted) under the minimum charge is raised to the minimum charge. Hence, there is no shipment that costs less than the designated (in the contract) minimum charge. In practice, today's LTL prices are given as a percentage discount and a minimum charge, on which the carriers and shippers contract based on the selected tariff, O-D pair and freight class.

In this research we focus on estimating the total price of the LTL shipments. Therefore our market rate estimates are independent of any tariff. Our aim is to create a model that will allow us to predict the market rate for any type of LTL shipment with high confidence given the origin-destination, freight class and weight. Desired minimum charge level can then be applied if the market rate estimates are less than certain minimum charge level. Next, we present this holistic approach with multiple regression modeling.

2.4.2 Regression Model

We propose a process and a model that estimates LTL market rates. Our prediction process has the following three steps: (1) Regionalization, (2) Multiple regression model, and (3) Post-regression analysis (Optional).

Geographical regions impact the pricing of LTL services depending on the characteristics of the carriers operating within those regions and their different pricing policies. In our estimation process we first propose specific regions. Then we run our multiple regression model that consists of both tangible and intangible predictors together with Origin and Destination Region information. Finally we allow the users to bring their expertise into the analysis by considering other factors that may not be captured by the general trend in the dataset.

2.4.2.1 Regionalization

With LTL expert knowledge on the current industry practice and the distribution of regional LTL carriers' service areas we group the US states into five non-overlapping regions, namely, West, Mid West, South Central, South East and North East. See Figure 2 for the regionalized US map. To add some flexibility to our process we create a regionalization assignment table in a database. This table contains the assignment information of all Zip3 regions (i.e., an area that is the collection of Zip5 regions with common first three digits) in the continental US that are assigned to specific regions. Each Zip3 belongs to a state, and each state belongs to a region as assigned.

If the user prefers to conduct analysis in much smaller regions (therefore more number of regions), it is possible by only altering the zip3-region table. A Zip3 area cannot be in two regions at the same time. However, by changing (squeezing) the regions, there is a trade-off between getting more specific results and decreasing the reliability of the estimates. Smaller regions mean fewer shipments, which translates into less variance being captured in the region. Altering the Zip3 assignment table option can be considered at a later stage to get more specific/precise results if we have more historical shipments or if we increase the time-span of the data to a longer horizon.

Regionalization of the data splits the entire database into 5 regions. One LTL shipment can only originate from one Origin Region and can go to one Destination Region. Looking at the distribution of shipment origin and destination regions in Figure 3, we can say that our database of shipments is distributed almost evenly among the regions and fairly represents the US LTL market. Mid West is the most industrialized region, therefore it contains the most number of inbound and outbound

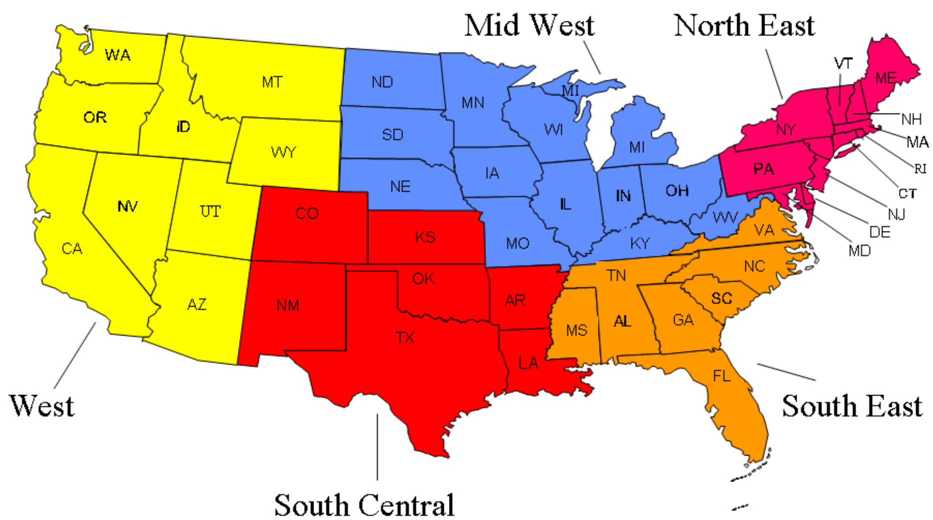


Figure 2: LTL Regionalization map based on pricing structures.

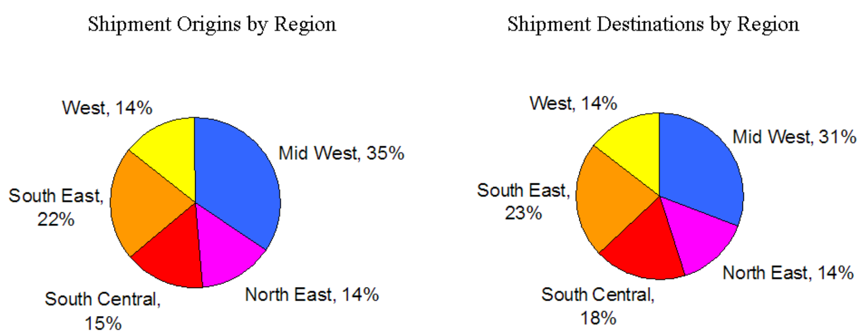


Figure 3: Distribution of Shipments by Origin and Destination Region.

shipments.

2.4.2.2 Multiple Regression Model

We analyzed the initial list of important tangible and intangible factors that impact the LTL pricing, and then we selected the most important ones to use in our regression equation.

Tangible factors in the model are Weight (W), Mile (M), Freight Class (FC), Origin (O) and Destination (D) Region and Carrier Type (CT). Instead of using Origin-Destination Zip codes, we quantified total distance in miles. Freight Class is the contracted type of freight being carried by the LTL service provider. Carrier Type is classified under Regional, Super-regional or National by Schneider Logistics based on the number of states served by the carrier.

Freight class is found to be less relevant although it is directly included in the base price tariff calculations. The reason is that the contracted freight class might differ from the actual freight class that carriers see, so the carriers are not willing to give big discounts for freight that is contracted with less than its actual freight class. This phenomenon is observed in several instances. For example, a major retailer found contracting its freight at a freight class of 50 (lowest freight class), although corresponding general merchandize has a freight class of 100 or more. In comparison with the other freight class 50 shippers, this retailer was paying significantly more for its LTL shipments. We consider and address this issue by creating a freight class index (Freight Index) that is modeled by expert input to be used as part of Intangible factors.

Intangible factors in the model: Following intangible factors affect the LTL pricing but they are not captured in the data set. However, with expert knowledge it is possible to quantify these characteristics with a survey methodology.

- **Freight Desirability:** is what makes freight appealing to the carriers. The reasons why a particular freight type is more desirable to a carrier vary; we focus on whether or not the freight is stackable, palletized, high density and whether driver delays occur while handling this freight.

- **Negotiation Power of the Shipper:** is how much influence the shipper has with carrier. We measure negotiation power based on whether the shipper bid its freight within the last year, whether the shipper has high freight spend (usually above $20M/year), whether shipper uses a consulting company and uses carrier tariffs for base price calculations.

- **Economic Value Estimate:** includes additional shipper factors that directly influence the pricing structure of the shipment. Whether the shipper is low-cost oriented, requires time windows or guaranteed delivery and whether it prefers national carriers are all part of this measure.

- **Perceived Freight Class:** is the freight class as carrier sees it based on true product density and not as stated by the shipper.

Tangible factors such as mile and weight are easy to incorporate into the regression model. However, for intangible factors it is hard to quantify the values of the each or it is subject to judgment. For example, we may not know how to express the negotiation power of the shipper or the desirability of a particular type of freight to a carrier. To overcome this problem, we propose a **market scorecard** methodology that considers all the intangible factors and weighs them to get a score for a shipper. We create Shipper Index to relatively score each shipper based on their characteristics, and reflect the shipper's relative position in the market. Similarly, we create Freight Index, which relatively defines the actual (perceived by the carrier) freight class.

Shipper index (SI): corresponds to a score between 0% and 100% and is calculated by answering the survey questions under three major categories for each shipper, namely, freight desirability, negotiation power, and economic value estimate. The higher a shipper's score, the higher is the LTL price that is likely to be charged for a similar shipment. These survey questions are designed to be yes/no questions for simplicity, and they are answered by LTL experts for each shipper in the database. Some of the questions have a positive impact on the price for the shipper, meaning that they "decrease" the LTL price, versus others have negative impact. Table 3 shows the 12 yes/no questions (4 for each category) presenting their relative weight (importance level) within their category, their direction of impact (positive/negative) and the relative contribution of each category to the Shipper Index.

The Shipper Index is the weighted average of each category scores, and shown in

Table 3: 12 Yes/No questions within 3 categories used to calculate Shipper Index for each shipper. Weights of questions represents their importance within each category.

Freight Desirability (60%)	(+) **Stackable Freight?**	**20%**
	(+) **Palletized?**	**30%**
	(+) **High Density?**	**20%**
	(-) Driver Delays?	30%
Negotiation Power (30%)	(+) **Bid within last year?**	**40%**
	(+) **High Freight Spend?**	**10%**
	(+) **Using Consulting Company or 3PL?**	**20%**
	(-) Using Carrier Tariffs?	30%
Economic Value Estimate (10%)	(+) **Low Cost?**	**60%**
	(-) Time Windows?	10%
	(-) Guaranteed delivery?	20%
	(-) Prefers Nat'l Carriers over regional?	10%

Equation (1).

$$SI = \sum_i c_i S_i \tag{1}$$

where i is the category index, S_i is the Category Score of category i, c_i is the weight of category i as listed in Table 3. With the current category weights, Shipper Index is calculated as follows:

$$\begin{aligned} SI = \; & 0.6[Freight\_Desirability\_Score] \\ & +0.3[Negotiation\_Power\_Score] \\ & +0.1[Economic\_Value\_Estimate\_Score] \end{aligned}$$

Each category score is calculated starting with its baseline and adding and subtracting the question weights that are answered "yes." The positive question weights are subtracted, while the negative question weights are added to the baseline. The resulting category score is anywhere between 0% and 100%, but is a discrete number as determined by the question weights. Equation (2) shows the category score calculations.

$$S_i = B_i - \sum_{j \in Q^{i+}} w_j y_j + \sum_{j \in Q^{i-}} w_j y_j \tag{2}$$

where

$$
\begin{aligned}
B_i \quad &= \textstyle\sum_{j \in Q^{i+}} w_j,\ \textit{is the Baseline of category } i\\
i \quad &= \textit{Category index}\\
j \quad &= \textit{Question index}\\
w_j \quad &= \textit{Weight of Question } j\\
y_j \quad &= \begin{cases} 0, & \textit{if the answer to question } j \textit{ is no}\\ 1, & \textit{if the answer to question } j \textit{ is yes.} \end{cases}\\
Q^{i+} \quad &= \textit{Positive Questions in category } i\\
Q^{i-} \quad &= \textit{Negative Questions in category } i
\end{aligned}
$$

With this type of formulation we allow relative scoring at two levels. The first one is at the category level. Each shipper has a score that determines their category position from 0% to 100%. Then these category scores are weighted again at the higher level, which constitutes the Shipper Index. Shipper Index quantifies the currently non-captured information on each of the three categories.

Freight Index (FI): is the second component of the relative scoring mechanism. With data analysis and industry expert knowledge we know that some firms do not contract with their actual freight class. The dataset captures only contracted freight class values, not the actual freight class as perceived by the carrier. We incorporate the Actual (Perceived) freight class into the intangible factors, specifically in the Freight Index that we calculate. An interesting observation about the freight class is that the freight class value, which is generally referred as the product category in the industry, has almost a perfect correlation with the base price if considered as a value. Testing a widely used industry tariff, we identify the linear relationship of freight class value with the tariff's base prices. Figure 4(a) illustrates this relationship with a linear regression line fitted with 99.7% coefficient of determination (R-Square). Figure 4(b) illustrates the relationship if the freight class values were taken as categories instead of numbers. Freight class (FC) values in use today are listed in Table 4.

Extensive testing of this linear relationship showed that the average increase in the base price with each incremental freight class value is changing between 1.4% and 1.9% of the starting base price (i.e., freight class 50). This value varies based on origin and destination regions, but stays practically constant for any given lane. However it points out a very important relationship of Base Price (P) and the underlying tariff formula. Fitted regression line has the following general form:

$$\hat{P}(FC) = \alpha FC + \beta$$

where FC represents the freight class and the slope α is dependent on the intercept β, as explained by this relationship. Without loss of generality we can assume that the y-axis starts at freight class 50 (i.e.,$\hat{P}(50) = \beta$). Then the freight class and Base Price relationship takes the following form:

$$\hat{P}(FC) = \alpha(FC - 50) + \beta \tag{3}$$

where $\alpha/\beta \in [0.014, 0.019]$.

We propose a method to calculate Freight Index (FI) to be used as one of the LTL Market Rate predictors. First we define Freight Index (FI) using the regression equation (3), such that:

$FI = \hat{P}(FC^{actual}) = \alpha(FC^{actual} - 50) + \beta$

Our dataset has a maximum freight class of 150. Therefore, we take freight class 150 as the 100% baseline. For Freight Index (FI) to take a maximum of 100% score (like the Shipper Index score takes 100% as the maximum score), we need to scale this regression formula. Therefore, we set $FI^{\max} = 100$.

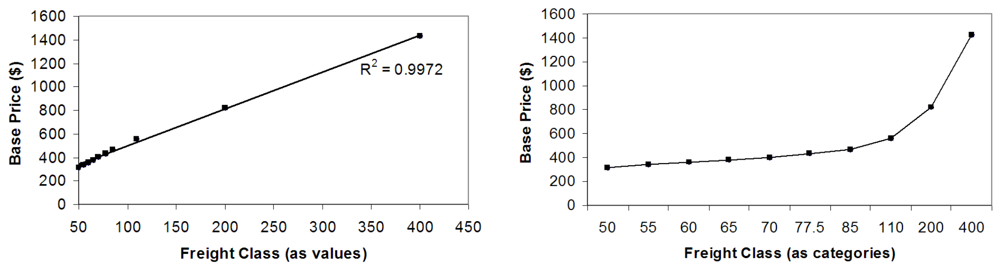
Let $FC^{\max}$ be the maximum freight class available in the dataset.

Then the following should hold:

$FI^{\max} = \hat{P}(FC^{\max}) = \alpha(FC^{\max} - 50) + \beta = 100$

Table 4: Currently used freight class categories. 18 different classes of freight. Higher the class, more expensive the LTL shipment is with all the other variables held constant.

Freight Class Categories								
50	55	60	65	70	77.5	85	92.5	100
110	125	150	175	200	250	300	400	500



(a) Freight Class as values (b) Freight Class as categories

Figure 4: Illustration of the linear and non-linear relationship of Freight Class as “values” and “categories” with the generally accepted LTL tariff Base Price for a given OD pair and Weight.

Since we do not know the actual distribution of α/β, we pick Mid West Region (that has the most number of inbound and outbound shipments) as the representative lane and find that $\alpha/\beta = 0.017$ is consistent for Mid West to Mid West shipments. Substituting it to the above equation we find:

$\beta^* = \frac{100}{(0.017)(FC^{\max}-50)+1}$ and $\alpha^* = \frac{1.7}{(0.017)(FC^{\max}-50)+1}$.

Here scaled slope α^* ensures that regression equation (3) gives a maximum value of 100. Using the linear relationship, we can re-write the Freight Index formula as in equation (4). Therefore, for a given Actual (or perceived) freight class value, the Freight Index is calculated as follows:

$$FI = 100 - \alpha^*(FC^{\max} - FC^{actual}) \tag{4}$$

Statistical results show that both Shipper Index (SI) and Freight Index (FI) we calculated for all the shippers are significant contributors of the final LTL price prediction. In fact, Freight Index (FI) is found to be much more significant than the original contracted Freight Class (FC).

Multiple Regression Model: Shipper Index and Freight Index are both designed to relatively quantify each shipper's market position. Therefore, we create the following main predictors: Mile, Weight, Freight Class, Shipper Index, Freight Index, Carrier Type, Origin region, and Destination region. To estimate the market rates of the specific LTL shipment we propose the following general model:

$$y = f(M, W, FC, SI, FI, CT, O, D) + \varepsilon \tag{5}$$

where ε is the random error due to the other unobservable factors.

After extensive model building and testing steps, it is found that miles (M) and weight (W) are the most important factors. In addition, their interaction effect ($M * W$) and the quadratic effect of weight (W^2) found to be significant. The M and W are also found to interact with the other factors. Therefore, we propose the following regression model:

$$y = \beta_0 + \beta_1 M + \beta_2 W + \beta_3 M^* W + \beta_4 W^2 + \varepsilon,$$

where $\beta_0, \beta_1, \beta_2$, and β_3 are functions of the other predictors FC, SI, FI, CT, O and D. However, β_4 is taken as a constant because W^2 has no interaction with the other factors. We use a linear predictor for each parameter β_j. Thus, for each $j = 0, 1, 2, 3$:

$$\beta_j = \alpha_{j0} + \alpha_{j1} FC + \alpha_{j2} SI + \alpha_{j3} CT[N] + \alpha_{j4} CT[R] + \alpha_{j5} O[MW] + \alpha_{j6} O[NE] + \alpha_{j7} O[SC]$$

$$+ \alpha_{j8} O[SE] + \alpha_{j9} D[MW] + \alpha_{j10} D[NE] + \alpha_{j11} D[SC] + \alpha_{j12} D[SE]$$

Note that the carrier type predictor is replaced with two 0-1 dummy variables $CT[N]$ and $CT[R]$, where $CT[N] = 1$ when the carrier type is National and 0 otherwise and $CT[R] = 1$ when the carrier type is Regional and 0 otherwise. Similarly, four 0-1 dummy variables are introduced for Origin and Destination regions, where MW, NE, SC, and SE stand for Mid West, North East, South Central, and South East regions, respectively. Thus, there are a total of $13 \times 4 + 1 = 53$ parameters in our regression model.

Table 5: Mean and standard deviation of numerical predictors.

Predictor	Mean	St.Dev
W	1492.4	1801.7
M	1013.3	738.0
SI	27.4	24.3
FI	66.5	11.9
FC	65.2	18.5

Standardizing the Variables: Because the scales of the variables are quite different, we standardize the numerical predictors such as Weight, Miles, Shipper Index, Freight Index and Freight Class. This makes the relative comparison of the coefficients meaningful.

As a general rule, the predictor X is standardized as follows:

$$x = \frac{X - \bar{X}}{s_x}$$

where $\bar{X}$ is the mean of the predictor X, and s_x is the standard deviation of X, which is calculated from the data. Small x is the standardized predictor. Since our predictors are given as standardized predictors, the corresponding mean and standard deviations are needed to use the regression model properly. Table 5 provides the necessary means and standard deviations for convenience.

One needs to standardize any shipment with the details listed in Table 5 to be able to use it in the regression model. Weight is given in pounds. Any interaction can be achieved by multiplying the corresponding standardized predictors.

Model Selection: Since the initially proposed model contains 52 variables (and thus 53 parameters including the intercept), it is difficult to interpret and use in practice. Moreover, some of these variables may have practically insignificant effects on the line haul price. Therefore, removing some of them will not adversely affect the prediction, but can help in simplifying the model. We employ a backward elimination strategy to select the best model. We start with the full model containing 52 variables and then we eliminate the least important variable (with the least F-ratio) obeying

Table 6: Multiple Regression Model Performance. R^2 and $RMSE$

RSquare	0.9374
RSquare Adjusted	0.9374
Root Mean Square Error (RMSE)	37.964
Mean of Response	175.632
Observations	356,425

the hierarchical principles. This step by step procedure reduce the model size down to 29 variables without significantly affecting the model performance.

2.4.2.3 Model Results and Interpretation

After the backward elimination steps we select the best model that gives high performance with minimum number of variables. We observe that the selected model has very high predictive power that explains 93.7% of the LTL market rate variability. With such a high multiple coefficient of determination performance and also relatively narrow confidence intervals, we can help the LTL market analyst to identify a small interval that the given shipment should be priced within, according to the current market dynamics. When building these models with real data, we know that our dataset might have some outliers due to different reasons such as expedited shipments (much higher price) and data errors (wrong weight, wrong price, wrong zip code, etc.) due to non-standardized transaction recording methods such as typing. After building our model, we remove only the most extreme outliers with the studentized residual method, if the prediction error is at least 4 σ (i.e., if the absolute value of studentized residual is greater than or equal to 4). We repeat this procedure twice and on the average remove only 1.8% of the total data. Table 6 summarizes the results and Appendix C.1 presents all the regression parameter estimates for each predictor.

Interpretation of the Model Results and Predictor Importance: One of the major contributions of our paper is to capture market expertise with market score indices to improve market rate estimations. We find both Shipper Index (SI)

and Freight Index (FI) scores to be statistically significant predictors of the model. Furthermore, Freight Index (FI) turned out to be much more significant than the original contracted Freight Class (FC). While the t-ratio of FI is 301.9, it is only 30.4 for the original FC and the corresponding FI coefficient in the final model is 22.4 while FC coefficient is 2.4, which suggests that the Freight Index (FI) we calculated using our derived formula is almost 10 times more important than the original contracted Freight Class (FC). These results strongly support the expert observation that the freight classification practice is manipulated by the shippers to get a better rate. However, carriers respond this by not giving steep discounts to those shippers that lower their actual freight class in the contract. Therefore, actual freight class (captured by FI) turned out to me much more significant than the contracted freight class (captured with FC). Furthermore, FI and SI made it into the regression model as the 5^{th} and 10^{th} most important variables according to the absolute effect sizes presented in Appendix C.1. FC is the 8^{th} most important one. Weight (W) is the most important predictor. $W * M$ was the second most important, while Miles (M) is the third most important predictor. Both FI and SI contributed to the model further with their interaction variables. $W * FI$ was the 6^{th} most important variable in the model, while $W * SI$ is 7^{th}.

Carrier Type (CT) also contributed to the model performance by being the 9th most important variable. As expected, National carriers charge more than Super-regional carriers, which is more than regional carriers. The regression equation reveals this effect by having parameter estimates of CT as β(National)> β(Super-regional)> β(Regional). In the final model, Super-regional type is taken as the base line, so the coefficient for National Carriers is found to be 10.5, while Regional Carriers coefficient was -4.4. This shows that there is almost a \$15 difference on the average between the prices of National vs. Regional carriers. This corresponds to 8.5% difference between National and Regional Carrier pricing considering \$175.6 mean of market rates. Since CT also has an interaction variable with Weight (W) and $CT * W$ variable also presents a similar coefficient structure, we can say that this gap gets larger for shipments with higher than average weights. This additional difference is calculated as extra \$11.25 for additional 1000lbs weight increase in the shipment from the mean weight of approximately 1500lbs.

One important LTL market rate characteristic is the all-unit-discount schedule implemented on the weight brackets. Having discounts at certain weight threshold levels raises the question of W^2 impact. As expected, this phenomenon is also observed by a significant W^2 parameter (which negatively contributes to the price) being the 4^{th} most important variable.

Regionalization presents itself by having significant interactions with other predictors. During our backward elimination procedure, Destination Region (D) variables are eliminated from the model and three groups of Origin Region (O) variables stayed. This suggests that shipment origin is more important for an LTL shipment than the shipment destination in pricing. Statistically significant interactions of O with M and $W * M$ are included in the final model. However, we can interpret from the regression results that region related variables are in fact less important than the remaining variables in the model.

Following section identifies some of the other factors that can be used to fine-tune estimations.

2.4.2.4 Post-Regression Analysis

The final part of the LTL Market Rate estimation procedure is the post-regression analysis. There may be some further improvement opportunities at this stage especially for special situations. Certain cases may need special analysis such as focusing on specific cities or states that are known to have higher or lower LTL market rates.

Below we identify some areas to consider to further refine the market rate estimates:

1. **Origin-Destination States:** If certain states or cities are known to have higher/lower rates (i.e., it is cheaper to ship freight out of South Florida than into)

2. **Minimum charge analysis:** If there are many "low weight," "low price" shipments, then different minimum charge levels will affect the LTL prices significantly. It is advised to investigate and use accurate minimum charge levels. Regression model does not provide minimum charge considerations. However it can also be easily automated if minimum charge levels are known with enough confidence.

3. **Complementary Network:** Certain freight might be really desirable for certain carriers because it may fit into their network structure and increase their utilization, while decreasing dead miles. In these cases, carrier can offer steeper discounts.

The above list of special analyses is a limited catalog. However, there might be other cases that require further investigation. If the post-regression analysis is being done because of a benchmarking study, then the analyst should understand the nature of the shipments (including shipper and freight) as well as the nature of the business that the shipper practices. Some topics to keep in mind during benchmarking studies are covered in the next section.

2.5 Benchmarking LTL Shipments

One of the most valuable uses of price predictive models in the LTL industry is the ability to provide market rate benchmarking studies. For most of the companies that use LTL as part of their logistics activities, LTL constitutes one of the most costly items in their expense list. The road for shippers to reduce cost in their LTL network passes through self-awareness within their market, which is achieved by industry benchmarking studies. For shippers with high LTL spend such as large retailers or big auto makers, a few percentage point reductions in their cost might translate into millions of dollars of annual savings.

Today, benchmarking studies are mostly conducted by experts with limited market knowledge. If certain 3PL companies or freight payment companies have market knowledge on certain lanes, they can provide benchmarking services to their clients for these specific lanes. Shippers may even be able to create coalitions to leverage each other's knowledge. The prediction model we present in this paper provide a significant value both from quality and coverage perspective (analytical model that considers tangible and intangible factors to estimate market rates for entire US) and resource perspective (automation of benchmarking process saves very valuable expert time).

The following simple benchmarking algorithm illustrates the automation of the benchmarking analysis.

Simple Benchmarking Algorithm:

1: Prepare the data set to be benchmarked against LTL market

- 1.a. Calculate "mile" column for each shipment using O-D Zip codes
- 1.b. Calculate Shipper Index and Freight Index for the shipper and add to dataset

2: For each shipment in the benchmark dataset find the market rate estimate

- 2.a. Standardize the numerical predictors with given mean and st.dev.
- 2.b. Use standardized predictor columns to generate market rate estimate

3: Compare the actual LTL payment with the market estimates

3.a. Compare the mean

3.b. Find the position in the confidence interval

4: Combine the benchmark dataset with market estimates and report

At Schneider Logistics, benchmarking algorithms are automated with user interface using MS Access software package. This application enables the analysts to benchmark thousands of client shipments with the current market rate estimates within seconds on a shipment-by-shipment basis and results can be easily classified and reported from different perspectives. For example, it is possible to report the benchmark results for different regions, showing the shipper's performance with respect to market for similar freight. A shipper therefore can see which regions to focus on for cost savings. Another reporting can be done across the same industry. If the market data includes other companies within the same industry, then clients can see their position within this market. Benchmarks can also show the purchasing effectiveness of 3PL companies by comparing their current customers with other companies, therefore showing the extra potential cost savings for 3PL's customers.

Shipment level benchmark studies enable analysts to provide tailored and specific benchmarks of interest. Combined with estimation models, benchmarking algorithms are clearly a market visibility tool. It can be used to research market dynamics/trends over time, to find market areas for developing successful new product offerings and to enable greater negotiation leverage for both shippers and carriers.

2.6 Conclusion

The main contribution of this study is the development of a decision-support tool and a procedure for estimating LTL market rates with minimum manual intervention. We propose a powerful estimation model that explains a significant majority of the market rate variability. We consider both tangible and intangible factors important to LTL pricing and provide a scorecard method to quantify intangible factors to be used in the multiple regression model. Finally, we present additional opportunities for prediction improvement for special cases as part of the post-regression analysis step.

One of the most beneficial areas to apply our predictive model is in benchmarking studies. Our research identify best ways to exploit the predictive power of regression models. Benchmarking algorithms provide additional value to the prediction models. It enables analysts to benchmark any company's historical LTL shipments with the same-dated market data. This will allow the shipper to realize its market performance and position from many different perspectives, potentially leading the way to extensive cost savings. Benchmarking algorithms offer the flexibility for analyzing the market performance from multiple points of view. Algorithms compare the shipper performance with the market rate estimates on a shipment-by-shipment basis instead of giving an overall financial benchmark, which allows the generation of many performance reports based on any factor or parameter. Automation of these models and algorithms makes the process of market research much faster. This creates an opportunity to reduce the cycle time of benchmarking studies, decreasing the time required from experts and making it possible to serve many potential clients with a consulting type service offering.

Overall, the ability to easily estimate and benchmark LTL market rates is a very beneficial market visibility tool that allows analysts to do extensive research on the LTL market. For 3PL companies, it can be a revenue source for consulting service offerings in terms of benchmarking studies. Also it can be used as the source for accurate market rates for doing network design studies. This research can also serve as the baseline for potential market rate analytics projects for other B2B markets such as truckload, small package, etc. For shipper companies, this tool can bring much better market visibility. Even without owning the extensive LTL market data, most up-to-date regression models can serve the shipper needs. For carriers, the market knowledge brings a serious competitive advantage. Knowing the pricing of US LTL market, carriers might benefit in pricing their own services or offering new services that can be successful in the market.

We consult with LTL market experts in order to understand the most important factors in LTL pricing. The market score mechanism proposed as part of our approach can further be revised if the conditions on the market change. Weights in the suggested formulas can be updated to give better correlations with the LTL prices, or more survey questions can be added if necessary. The market scorecard method is by no means the absolute way of quantifying the intangible factors. For our research we find that the market scores (i.e., SI and FI) are highly correlated with the LTL prices, which in turn improves the predictions. Future research can look into different ways for quantifying the non-captured factors and compare with currently proposed methodology. Similar market scorecard methods can also be tested in different B2B markets to validate the improvement potential in prediction power.

Based on the size of the data set and distribution of shipments on different lanes, future LTL pricing studies may consider adjustments to the regionalization map. Regionalization step is proposed to distinguish regional pricing differences. For larger data sets, altering the regionalization assignment table (to get more, but smaller regions) might be considered given that smaller regions still contain enough diversity for number of carriers and shippers as well as different freight classes involved. Smaller regions may help to reduce errors of predictions, but this may cause the loss of important information otherwise captured with larger regions.

CHAPTER III

VALUE OF DEMAND INFORMATION IN PLATFORM PRODUCTS SUPPLY CHAINS

3.1 Introduction

In this research, we focus on using advance demand information as a tool to improve forecast accuracy and gain insights on production and inventory planning of platform products. We define "platform product" as a group of components (or products) that are separately available to the market, but designed to perform better when assembled and/or used together. Our motivation is the latest trend in the semiconductor industry, that is projected to have over \$250B in sales in 2006, to provide platform products that consist of components (i.e. CPU, Chipset, Wireless card) that are validated to work together more efficiently in terms of higher performance with lower energy consumption.

"Platforms" first emerged as a marketing concept in Semiconductor Industry. The component suppliers were already supplying different components to the Original Equipment Manufacturers (OEMs) that produce the final product, which is a personal computer (PC). Suppliers wants to sell the whole set of components, whereas the customers have the option of buying different components from different suppliers. Later, this marketing concept started to show itself in the design of these components, such that now the several components that go into a PC were carefully configured to work together more efficiently, namely with less energy and more performance.

Platform products have two dimensions. The complementariness dimension is determined by the components that complement each other to form a platform product (i.e. processor, chipset and wireless card). On the other hand, the substitutability dimension is determined when the substitutable components such as different variants of a particular component (i.e. processors with different speeds or energy consumption levels) can be used in a platform. For simplicity, we will analyze the complementariness dimension, and leave the other dimension as a future research direction. The customers (i.e. OEMs) can buy all the complementary components from the same supplier to form a platform, or they can purchase some of the components from different suppliers. For a platform product supplier (i.e., Intel), this means an ever changing market environment with different demand levels of different components, since the market is mostly driven by continuously changing/decreasing prices and fierce competition for market share. Some of the demand volume for individual products/components comes from customers who eventually purchase all the components to build a platform. However, there are also customers who purchase only some of the components that go into a platform, but not all, as they may purchase the other components from other suppliers (For instance, Dell buys Intel processor and chipset but uses its own brand wireless card for some of its mobile platforms). Hence, the demands of the components are partially dependent on each other. Generally suppliers have no demand information at the platform level, i.e., they do not capture whether a customer will purchase all the components in a platform, or only some, and they manage their component supply chains independently by forecasting each component demand separately. The lack of demand information at the platform level has two primary causes: (i) Customers prefer to purchase the components at different times, depending on when they are needed in the production process. They also want to benefit from price reductions over time. Hence, they do not want to commit to purchasing all the components in the platform at the beginning, even if they intend to do so over time. (ii) The ordering and planning systems in many companies can handle orders at the component level, but not as a "batch" order which includes multiple components.

In Semiconductor Industry the manufacturing processes are very complex and consist of several steps with total manufacturing lead times as long as 3 to 4 months. Long lead times combined with high demand uncertainty make the impact of forecast error more detrimental and the coordination of component supply decisions is therefore harder. At each manufacturing step, the demand information can only be used to a minimal extend to affect the output of the process, because the processes do not allow postponement of product differentiation to later stages. Therefore, planners need to make production planning decisions long before they have reliable demand signals from the market.

Our goal in this paper is to answer the following research questions:

- If supplier can get demand information at the platform level (i.e., customers "order as a kit" versus "order as components"), how does this information impact the forecasting and production/inventory management process, the forecast accuracy and inventory levels?

- How can we quantify the costs/benefits of the availability of advance demand information to the supplier?

We look at this information at two levels: A customer orders the platform as a kit, specifying (1) the platform quantity and (2) the timing of when she wants to receive each component. In this setting, when the customer places a platform order, the supplier will know the delivery times and quantities of the corresponding components with some confidence, taking into account the possibility of order changes and/or cancelations. Our model will assess the value of this advance demand information for the whole supply chain, considering the anticipated forecast improvements, and the supply chain inventory impact of this improvement.

The rest of the paper is organized as follows. In Section 3.2, we talk about related literature specific to platform products and platform strategy. We also mention advance demand information and nonstationary demand modeling literature. In Section 3.3, we present our Monte Carlo Model for analyzing the forecast improvements with extra demand information using platform ordering. We also present our numerical study on Intel Corporation's global supply chain, where we quantify the potential forecast improvement scenario. We give the results in Section 4.5.2.1, and discuss the outcomes and future research in Section 3.5.

3.2 Literature Review

There are different streams of literature related to our research. We are focusing on quantifying the value of a certain type of advance demand information. The advanced demand information literature offers important insights and directions for our research. Since we are analyzing the value of this extra information on the platform products, we are particularly interested in the platform products supply chain research. Although many previously published papers are related to our research from different perspectives, they are only similar in certain parts. Below we are presenting the different literature as they relate to our research.

Platform products and their supply chain: Platform products are typically investigated under product variety management, and Platform is defined as the shared common components of a product line that is used as a strategy to reduce the cost of manufacturing, lowering safety stock and having more accurate forecasts (Huang et al. 2005 ((36)), Meyer and Lehnerd 1997 ((62))). Different focus areas are explored under the platform products including more systematic approaches like optimizing supply chain configuration and design (Huang et al. 2005 ((36)), Kim et al. 2002 ((47)), Salvador et al. ((73)) and Park 2001 ((68))). Other more focused areas are lower safety stock (Baker 1985 ((5)), Dogramaci 1979 ((25))) and simplified planning and scheduling (Berry et al. ((13))).

Generally speaking, enabling the platform products and the product commonality, suppliers can deal with the "sku proliferation" or "product variety" problem by pooling the common components, which improves the forecast accuracy and decreases safety stock. Our approach for platform products is focusing on forecast improvements using the demand dependency of the components that forms the platform. In other words, we are focusing on the common parts of the platform that becomes different products with the addition of modular parts. The example from the computer industry is that the PCs that are based on Intel Centrino Platforms. Intel Centrino Platform has 3 components (i.e. CPU, Chipset and Wireless Card) and a PC that is based on this platform can have multiple variety of hard-disk sizes, memory options and others modular options. Most of the literature focuses on the platforms as it relates to the interaction of the platforms with the other (modular) components that creates the product variety. For instance the PC Platform that Huang et al. (2005) ((36)) analyze is consisting of a platform subassembly and modular option of having either DVD Drive or CD-RW. Our research is focusing solely on the demand dependency of components within the platform.

Advance Demand Information and Forecasting: This area of research contains numerous papers on the benefit of advance demand information in the inventory systems. Literature review of this field can be found in Gallego and Ozer (2002) ((28)) and Benjaafar et al. (2007) ((12)). The general focus of this growing literature is to find optimal inventory/production policies under various setting of having advance demand information. Although having an insight about optimal policies under different settings is important, application of these optimal policies to the real life supply chains are very limited due to the assumptions being involved with the mathematical models and quantifying the real benefits to the supply chain may not be representative. Our research is only focusing on forecast error improvements under certain advance demand information setting (that comes with customer platform orders) and tries to quantify the impact of this setting on a real supply chain using a representative supply chain simulation model. We analyzed a forecasting system where the orders are arriving periodically and give information about both the immediate orders (first delivery of platform components) to be filled from stock and some information about the future orders (remaining components of the platform order). The information about future orders has two vectors. First, the timing of the future orders, is known and constant , and it depends on the manufacturing process of the customers (i.e. customers require Chipset first to install onto the motherboard, they will later require the wireless card and CPU as the last steps of their manufacturing process). Second, the quantity of the future orders, which may change over time due to order updates or cancelations. In our paper, the benefits of having this information about future orders is used in the forecasting process and the forecast accuracy improvements are found using a Monte Carlo Simulation approach with demand being a nonstationary process that change over time with the product life cycle phase of the product and the seasonality of the demand. Advance demand information literature focusing only on the forecasting accuracy improvements is limited. One approach of getting extra information about the future demand is using leading indicator products. Time-lagged correlations of certain products might give information about the future demand structures of other products. A study on this approach can be found in Wu et al. (2003) ((86)).

Modeling nonstationary demand: Nonstationary demand is a part of real life systems. Compared to the stationary demand systems, considerably fewer researchers use nonstationary demand in their supply chain and inventory models (Graves and Willems 2005 ((31))). This type of demand is especially applicable to high tech industry where the products have short life cycles and we generally do not observe steady demand. The techniques used to model nonstationary demand vary. Some papers use integrated moving average techniques (Graves 1999 ((30))), some others employ

Markov-modulated (state-dependent) Poisson demand process (Chen and Song ((19)), Abhyankar and Graves 2001 ((1))). Johnson and Anderson (2000) ((40)) employs nonstationary demand that has product life cycle pattern. One of the techniques for generating nonstationary arrivals is by White (1999) ((84)), where bivariate thinning approach is used to model customer arrivals to an electronic store that changes by the day of the week and the hour of the day. In our paper, we modified this final methodology to fit into supply chain context, which allowed us to model the product life cycle pattern as well as the seasonality of the demand.

3.3 Model

In order to analyze platform ordering scenario, "Kits vs. Components", and quantify the benefits on the supply chain we developed two separate models. The first model is a Monte Carlo Simulation approach, which analyzes the forecasting process in the as-is scenario (Components ordering) and the to-be scenario (Platform ordering), then compares the forecast errors and calculates the savings. For a numerical case study we developed the second model, which is an end-to-end supply chain simulation of Intel's global supply chain. This model is integrated with optimization models to dynamically give production planning decisions during the simulation run. This way it represents the real system "planning" and "execution" cycles. We use this model to quantify the supply chain wide impact of forecast error savings achieved with the platform ordering scenario. Section 3.3.1 proposes a Monte Carlo model. In Section 3.3.2, we describe the Platform Supply Chain Simulator, which is an integrated supply chain simulation model we used to quantify the forecast accuracy improvements we calculated from the Monte Carlo Model.

3.3.1 Monte Carlo Simulation Model

Monte Carlo Simulation is used to test the forecasting system under stochastic demand. To be able to conclude that a certain way of forecasting is better than the other, we cannot rely on historical demand information, since the historical demand is only one realization of the underlying stochastic demand process. We need multiple replications to test two different forecasting scenarios under consideration (i.e., with or without platform orders). This enables us to find whether one method is consistently better than the other. Since the difference between two forecasting methods is coming from the usage of advance demand information, we will use a simple forecasting technique (i.e., single exponential smoothing) and enable the use of this extra information in the to-be scenario forecasting.

3.3.1.1 Model structure and assumptions

A Monte Carlo Simulation model is constructed for the numerical study of a platform product under consideration. This platform consists of three components: Chipset, Wireless Card and Processor (CPU). There is one supplier of these components and aggregated demand of many customers. We are given one year of historical demand data for each of these components. We assume that in the as-is scenario, each component demand is forecasted independently on a week to week basis with the reveal of new demand information at each demand period. For customers ordering these components to build a platform, we assume that they need to order the Chipset first. Then they order the wireless card, and finally CPU order is placed, which is the current industry process. However, the timing information is not known to the supplier. On the other hand, not all the components from this supplier are assembled to be a platform on the customer site, but for instance CPU components can be assembled with a competitor's chipset. So overall, there is no exact one-to-one relationship in component quantities.

To-be Scenario: In the to-be scenario, currently not-captured platform orders will be captured. Therefore, the customers place a platform order by specifying the quantity and the time when they need each of the components. For simplicity in the numerical study, we assume that the lead times between Chipset and Wireless Card; and between Wireless Card and CPU shipments are 2 weeks. Based on the historical data, we understand that the total shipment quantities show this pattern: Wireless Card < Chipset < CPU. We assume that the weekly platform orders should follow the time-lagged minimum rule, such that the platform order quantity cannot exceed the minimum quantity of the ingredient components that corresponds to the same platform order time (i.e., $Platform(t) \leq min\{Chipset(t), Wireless(t+2), CPU(t+4)\}$)

In this setting, we only get advance demand information for Wireless Card and CPU orders, because the platform order's first delivery is the chipset and chipset gives the advanced demand information about the Wireless card and CPU. We assume that platform orders are uniformly distributed between 80-90% of the minimum quantity, which is later relaxed for sensitivity analysis. Our final assumption is that a platform order cannot constitute perfect advance demand information for the future component orders. In other words, we allow for randomness even for the remaining parts of platform order, such that after the delivery of the Chipset part, we allow a quantity change in the remaining component demands by ±10% per week until its delivery time. This corresponds to a uniform demand of 80-120% for Wireless Card and 60-140% uniform demand for CPU part (i.e., $P_{wireless}(t) \sim Uniform(0.8 \times Platform(t-2), 1.2 \times Platform(t-2))$. This randomness is due to order cancellations and order updates that the customers are allowed until last minute, or it can be attributed to the randomness in manufacturing lead times (i.e., timing information), and it is a conservative demand assumption.

As the forecasting model, we selected single exponential smoothing for its simplicity and used it with a selected parameter of $\alpha = 0.7$. This parameter selection is preferred based on the performance of the method on the historical demand data that minimized the mean absolute percentage error (MAPE) on CPU families for all

$\alpha = \{0.1, 0.2, ..., 0.9\}$.

3.3.1.2 Stochastic Demand Generation with Bivariate Thinning

The idea of Monte Carlo Simulation is to test two forecasting methods under random demand. Therefore we need to characterize the randomness of the demand. In semiconductor industry, the demand is almost never stationary. It is observed in many instances that the new technology products follow a short life cycle pattern with quick ramp-up, high volume and end of life periods that is mostly single modal. This type of demand is nonstationary as it is a time dependent process.

We modeled our demand arrivals based on the approach by White (1999) ((84)) that is nonstationary demand process with bivariate thinning. In his paper, White shows that nonstationary Poisson process with bivariate thinning is a good approach in modeling customer arrivals to an electronic store. Time dependent arrivals are generated independently by day and by hour with a maximum arrival rate and the thinning factors. This methodology was also shown recently to be appropriate for the simulation of arrivals in traffic by Woensel (2006) ((85)). We employ the same base method with some modifications to fit into the supply chain context for modeling the demand arrivals to the system over a product life cycle. Therefore, we have leveraged the thinning factors from day-hour pairs to the quarter-week pairs.

Modifications: There are two sets of modifications on the random arrival generations with bivariate thinning. The first set is due to the nature of demand in Semiconductor industry. We do not assume that the arrivals (i.e. weekly demand arrivals) follow a Poisson process as in White (1999) ((84)). Our nonstationary demand process still follows the piecewise-constant rate, but we generate our random demand with normally distributed noise term around the piecewise-constant mean demand that changes by the quarter and by the week. The randomness is therefore symmetric around the mean, which is coming from $Normal(0, \sigma^2(t))$, where $\sigma^2(t)$ is the time dependent variance of the demand and calculated from the data for each quarter (i.e. $\sigma(1)$ is the standard deviation of the first quarter demand). To be able to prevent extreme/unlikely demand quantities, we truncate the demand at $\pm\sigma$.

Under this definition, the demand is nonstationary $Normal(\mu(t_{ij}), \sigma(t)^2)$, where

- t_{ij} represents the i^{th} week within j^{th} quarter and $i = \{1, ..., 13\}$, $j = \{1, 2, ..., Q\}$
- $\mu(t_{ij})$ is the nonstationary demand mean that is piecewise constant at each $\{i, j\}$ pair
- $\sigma(t_{ij}) = \sigma(t_j)$ for all $i = \{1, ...13\}$, so the demand variance is constant within the quarter
- Truncation happens at $[\mu(t_{ij}) - \sigma(t_i)]^+$ and $[\mu(t_{ij}) + \sigma(t_i)]$ for all i, j where $[x]^+ = max\{x, 0\}$.
- $\mu(t_{ij})$ is calculated with bivariate thinning factors η_i (weekly) and δ_j (quarterly) with the following equation:

$$\mu(t_{ij}) = \mu^{max}\eta_i\delta_j \tag{6}$$

where

$$\eta_i = \frac{\bar{x}_i}{max(\bar{x}_i)} \quad \forall i \in 1, ..., 13 \tag{7}$$

$$\delta_i = \frac{\bar{x}_j}{max(\bar{x}_j)} \quad \forall i \in 1, ..., Q \tag{8}$$

$$\mu^{max} = m(t_{ij}) \quad for(i, j) \ with \ \eta_i = \delta_j = 1 \tag{9}$$

Here $\bar{x}_i$ is the mean weekly demand for each week $i = \{1, ..., 13\}$ and $\bar{x}_j$ is the mean quarterly demand for each quarter $j = \{1, ..., Q\}$. As proposed by White (1999) ((84)) these are calculated as:

$$\bar{x}_i = \frac{1}{Q}\sum_{j=1}^{Q}\bar{x}_{ij} \quad \forall i \in 1, ..., 13 \tag{10}$$

$$\bar{x}_j = \frac{1}{13}\sum_{i=1}^{13}\bar{x}_{ij} \quad \forall j \in 1, ..., Q \tag{11}$$

Where $\bar{x}_{ij}$ is the mean arrivals for given period $\{i, j\}$ and Q is the number of quarters that we have data.

Our second set of modifications to the method emerges due to this technique being used in a supply chain context. For a given product life cycle demand, we do not have a second chance to observe the same demand process. It is not like the electronics store, where we can observe multiple Monday morning 10-11 a.m. periods each week. Therefore we cannot see the same life cycle phase of the product (quarter and week) more than once. So we replace $\bar{x}_{ij}$ with x_{ij}, which is the only observed demand of a given product in quarter j and week i. However, these observations are still smoothed with the bivariate thinning factor calculations. One way to get multiple observations is to use multiple products, but in general every product has its own demand and product life cycle characteristics that may differ from each other. So this analysis is ad hoc to the product for which we have demand data, but still gives insights about the general short life cycle semiconductor products and the value of advance demand information.

Having a short life cycle impacts the weekly thinning factor calculations. The weekly patterns within the quarter may be substantially different at each stage in the product life cycle, namely in ramp-up, high volume and end of life. Like the separation of weekday data from weekend data in the analysis of White (1999) ((84)) and Woensel (2006) ((85)) due to distinct pattern differences, we also divide the data into three life cycle phases: (1) Ramp-up, (2) High Volume (HV) and (3) End Of Life (EOL). Each phase provide data for calculations of three separate sets of weekly thinning factors $\{\eta^{Ramp}, \eta^{HV}, \eta^{EOL}\}$ and these weekly thinning factors are only used for the respective quarters that are assigned to that life cycle phase. For our numerical study, one of the products has total of $Q = 11$ quarters in the life cycle, first 3 of them are assigned to ramp-up, following 6 is high-volume and the remaining 2 are assigned to end-of-life phase with expert judgment. This modification improved the fit of the model to the actual demand data by reducing the mean absolute error by 45% while providing much better fits on both the ramp-up and end-of-life tails of the demand curve.

The final modification is for the quarterly thinning factors. In his analysis, White (1999) ((84)) used bivariate thinning for staffing decisions in an electronic store. Therefore, the ad hoc formulation for thinning factors helped to get a smoothed mean arrival pattern and it is aimed to staff the store with the right number of sales associates at the right time. Looking at "mean arrivals" for the day they can optimize the trade off between service level and number of sales associates. In the supply chain context, service level is more related to the peak demand instead of mean demand. Both the capacity decisions and demand fulfillment have to consider the demand upsides to be able to provide high service levels. Also in an electronic store, or in any kind of other retailer, customer will only experience a longer wait time in the cashier due to low staffing, however in the demand fulfillment case if there is not enough products on the shelf the customer may be lost to competitor and the cost of this shortage is much more detrimental. Therefore, in order to capture the demand peaks better we modified quarterly thinning factors δ_j by changing the "mean" arrivals with the "peak" arrivals. And we observed that the model gave a better fit to the actual data than using the mean arrivals in the thinning factor calculations. In our numerical study, the mean absolute error is further reduced by 22% with this final modification resulting in a total of 57% reduction of the mean absolute error over the original method and better capturing of the demand peaks.

With all the modifications applied, the nonstationary random demand for week i of quarter j (D_{ij}) is generated with the following formulas from top to bottom:

$$D_{ij} \sim Normal(\mu(t_{ij}), \sigma^2(t_j)) \tag{12}$$

where

$$\mu(t_{ij}) = \mu^{max}\eta'_{ij}\delta'_j \tag{13}$$

$$\eta'_{ij} = \begin{cases} \eta_i^{Ramp} = \frac{\bar{x}_i^{Ramp}}{max_i(\bar{x}_i^{Ramp})}, & i \in \{1,...,13\} \quad j \in Ramp; \\ \\ \eta_i^{HV} = \frac{\bar{x}_i^{HV}}{max_i(\bar{x}_i^{HV})}, & i \in \{1,...,13\} \quad j \in HV; \\ \\ \eta_i^{EOL} = \frac{\bar{x}_i^{EOL}}{max_i(\bar{x}_i^{EOL})}, & i \in \{1,...,13\} \quad j \in EOL. \end{cases} \tag{14}$$

$$\delta'_j = \frac{x_j^{Peak}}{max(x_j^{Peak})}, \quad \forall j \in \{1,...,Q\} \tag{15}$$

$$\mu^{max} = \mu(t_{ij}), \quad for \quad \eta'_{ij} = \delta_j = 1 \tag{16}$$

$$\bar{x}_i^{Ramp} = \frac{1}{|Ramp|}\sum_{j \in Ramp} x_{ij}, \quad \forall i \in 1,...,13 \tag{17}$$

$$\bar{x}_i^{HV} = \frac{1}{|HV|}\sum_{j \in HV} x_{ij}, \quad \forall i \in 1,...,13 \tag{18}$$

$$\bar{x}_i^{EOL} = \frac{1}{|EOL|}\sum_{j \in EOL} x_{ij}, \quad \forall i \in 1,...,13 \tag{19}$$

$$x_j^{Peak} = max_i\{x_{ij}\} \quad \forall j \in 1,...,Q \tag{20}$$

The Monte Carlo Simulation Model is designed to run for both “as-is” and “to-be” scenarios to use the same random numbers. Hence the use of random numbers is synchronized to give a more precise comparison of the two alternative systems. The model is run for 1000 replications to test the systems under almost all possible demand scenarios. The model algorithm logic is as follows:

Monte Carlo Simulation Algorithm:

1. Load the component demand data for 3 components

2. Construct the piecewise constant demand mean for all time periods

3. Construct the upper and lower truncation limits based on demand variance

4. Loop until replication number = 1000

    (a) Generate random demand for planning horizon based on the nonstationary demand model with bivariate thinning probabilities

    (b) Generate random platform orders based on the minimum rule assumptions

    (c) Component Scenario Forecasting

        i. Forecast the component demand stream with as-is model

        ii. Calculate forecast error

    (d) Platform Scenario Forecasting

        i. Forecast the platform and independent component demand stream with to-be model using Advance Demand Information

        ii. Combine the forecasts into component forecast

        iii. Calculate forecast error

    (e) Compare forecast errors and store results by replication number

5. end Loop

6. Publish 95% Confidence Intervals on forecast error savings

3.3.2 Platform Supply Chain Simulator

Our objective is to test different forecasting scenarios (with and without advance demand information) and quantify its benefits on a real supply chain. Therefore we build a simulation model that is automated to interact with optimization models for dynamic production planning decisions in order to mimic Intel's Global Supply Chain realistically. For Intel's global supply chain, we model flow of materials, order fulfillment, production processes and planning system. We call this model the "Platform Supply Chain Simulator". It is validated that the model gives very similar results as real Intel supply chain gives in real life. The validation results can be found at the end of this section.

The model is designed for the product family level. Although it is capable of simulating SKU level details, due to the limitations on data availability and the research scope being at a strategic level, we keep the model at family level. For this case study, real data is used in the Platform Supply Chain Simulator. Total of three supply chains are simulated for each of three product lines (CPU, Chipset and Wireless Card). Data collected and used in the simulation covers one year horizon from July 2005 to June 2006.

3.3.2.1 Modeling Approach - Why Simulation?

We quantify the forecast benefits on the Intel supply chain with a separate simulation model. In order to realistically assess these benefits on a complex supply chain, we need to build some of the critical complexities into the model and simplify the rest. Simulation modeling can help us define these complexities when the closed form analytic solutions are not available or desirable.

In the global supply chain arena, even a basic product can have a very long supply chain from the raw material suppliers to the end consumers, including steps like supplier's supplier, supplier, manufacturer, third part logistics (3PL) companies, carriers and even for some cases outsourced marketing and sales forces. Each entity along this supply chain has its own decision making system either systematic or qualitative or generally both, and the interaction of these entities along the supply chain is very complex. Early inventory management literature assumed a centrally managed supply chain to optimize the inventory levels at each echelon. However, in today's "era of outsourcing", companies are outsourcing most of the services and concentrating on the core competencies that creates a highly decentralized supply chain systems that is run by many different companies in different industries with different business objectives. Even within the same company - especially for multinational companies - each group or function has its own goals and objectives in the supply chain, which are usually not perfectly aligned with the overall corporate objectives. Under this reality setting, most of the assumptions that help us build analytical models fail to hold, which makes the results of these models less viable. Simulation models help us define most of these complicated interactions and control rules that manage supply chain operations.

Model Structure: Platform Supply Chain Simulator is a simulation model developed in Rockwell Automation's Arena 10.0 Software. This simulation model is integrated with production planning optimization models through Visual Basic for Applications (VBA) codes embedded inside the Arena model. Arena Model automatically calls ILOG OPL and Excel Optimization models for production planning decisions. This automated system works with a network of Excel files that serves as the intermediary input/output data storage environment.

In this structure, we have three major components besides the Excel input/output network. Those are:

1. Arena Simulation Model that contains the supply chain logic and serves as the mastermind of the Platform Supply Chain Simulator by calling the other decision models.

2. ILOG OPL Optimization Model that is called each simulation month for weekly production (wafer start) decisions for the Fab/sort Manufacturing for the next month. It imitates the current production optimization models at Intel.

3. Excel ATM Production Model that is called each simulation month for the daily

ATM production decisions for the upcoming month in the simulation.

At a high level, the system works the following way. At the beginning of the simulation time (week zero), simulation starts by calling ILOG Model to get the wafer start plan for the Fab/Sort Manufacturing. ILOG model takes the inputs like forecasts, yields, inventory targets and previously scheduled production that are stored in an Excel file (week zero input) and produces the output excel file (week zero output) that contains the weekly wafer start production decisions. At the same time Arena also calls Excel ATM Production Planning Solver to get daily ATM productions for one month. All these production plans are loaded into Arena simulation, and the simulation runs for one month like the real Intel supply chain runs. In this month the orders (from historical data) arrive daily and fulfilled from component warehouse if there is enough inventory. Inventory levels over time are tracked and written into Excel files. One simulation-month later, the procedure repeats itself by calling the production planning models again, but this time with the updated forecasts, inventory levels, inventory targets and yield over time values. The model can be run as long as we have historical data, generally as the multiples of months. Simulation length determines how many times the following simulation cycle turns.

Platform Supply Chain Simulator - Monthly Simulation Cycles:

1. Update forecast for 3 quarters out (Read from Excel - staggered forecast file)
2. Optimize Fab/Sort and ATM Production Plans (by calling optimization models)
3. Simulate (execute) 1 month with optimized production plans
    (a) Production
    (b) Material flow
    (c) Demand generation
    (d) Demand fulfillment

(e) Inventory check/update

4. Update ADI and CW inventory levels (continuously)

5. Update optimization model inputs

   (a) Yield over time

   (b) Inventory targets

   (c) Penalty parameters

Model components are explained with more details as follows:

3.3.2.2 ARENA Simulation Model

We used Rockwell's ARENA Simulation Software to model Intel's entire global supply chain at a high level. Arena provides connectivity options with many other applications such as optimization software and Excel. This enables the simulation model built in Arena to exchange data at any time during the simulation run and dynamically adjust to this new input stream. Arena modeling structure consists of very basic building blocks that are required in most of the simulation models. These elementary building blocks such as "create", "process", "decide", "assign", "hold", "separate" and "dispose" help us define the flow and control logic of the supply chain we are modeling. By carefully aligning these blocks and providing them with the correct parameters we can imitate the material flow within Intel's supply network. Once the flow structure is set, we use the connectivity options to imitate the decision making in the supply chain.

Network Flow and Decision Making: For Platform Supply Chain Simulator, after we setup the network flow with the elementary blocks we defined above, we used ILOG OPL Studio to design a mathematical optimization model to give wafer start production decisions and we designed an Excel model to give ATM production decisions. These production decisions are dynamically integrated with the Arena model by VBA codes. Other decisions like inventory target decisions and forecasts are deterministically given to the model, since there is not much analytics going into these decisions and therefore it is harder to imitate these kinds of decision making with analytical algorithms and models.

3.3.2.3 ILOG OPL Optimization Model

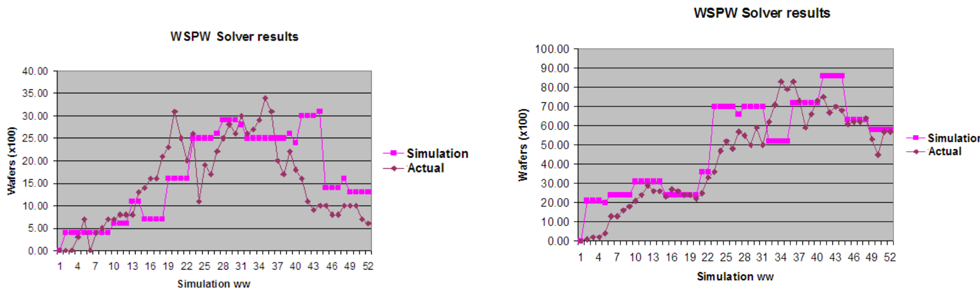
There are two optimization models developed in ILOG OPL Development Studio 4.2. They are both used for FSM production plans, which are called for wafer start plans, for CPU and Chipset supply chains. Advanced mathematical models named Build Plan Solvers (BPS) are used at Intel that optimize weekly wafer start decisions based on the penalty structure defined within these models.

We reviewed these solvers and simplified them for 2 product family case by re-writing our own optimization models in ILOG OPL Studio. We changed the penalty structure a little bit to relax the capacity considerations. We assumed no capacity constraint and defined no penalty for capacity utilizations. However, we maintained the inventory target penalties as well as production smoothing penalties in the models. The models we developed are end-to-end supply chain models that consider FSM build plans, involve yield over time numbers, calculate ATM build plans and ADI (Assembly Die Inventory) inventory levels, and finally CW (Component Warehouse - Finished Goods) inventory levels. The models are dynamically solved every month and they also exchange data within themselves because the previous month's production decisions are inputs as scheduled/fixed plans to the next month's solver runs. In order to accommodate this information exchange, we used Excel spreadsheets that are connected to each other. ILOG OPL reads the inputs from the input Excel files, and writes it to the output Excel files. But input Excel files read some of their values from the previous months' output Excel files and so on.

After setting up the models this way, integration of them with Arena is handled by VBA codes. In Arena, every month (considering 4+4+5 week split of a 13-week quarter) a VBA script is called. This script runs the ILOG OPL application; pulls out the correct model file of the month, and then runs this model until the solution is found. The model automatically writes the production decisions into the output Excel files. Script closes the ILOG OPL application and resumes the Arena model. Arena model at this time starts with a fresh new production plan by reading the values of wafer starts from the specified Excel file. This loop continues for 12 months with 12 ILOG calls.

This model is an linear relaxation of a mixed integer program (MIP) with 620 constraints, 1019 variables and it considers end-to-end supply chain, mimics the current solver. Objective function includes penalties for inventory targets and production smoothing. Model is called at the beginning of each simulation month, and it solves 9 months out for weekly production quantities, each call taking less than 1 second of CPU time.

ILOG Model Validation: In order to validate that our model works similarly like the actual BPS models, we can check the realized wafer start per week (WSPW) values versus the ones that are given by our models. However this is not a perfect validation, since the real BPS outputs are judged and changed before realization by the various management levels. We plot the actual vs. optimized WSPW graphs to see how close plans our models give with respect to the realized production levels. On the other hand, we also employ expert validation by consulting to the creators of the original solvers. After analyzing the actual vs. solver results in Figure 17, the solver team endorsed our work by acknowledging that the model is successful and the results are close enough to what the real solver might give.



(a) CPU Product (b) Chipset Product

Figure 5: Actual versus Solver Model wafer start per week values for two example products

3.3.2.4 Excel Model for ATM Production Planning

This model is designed to give daily ATM production decisions both for CPU and Chipset supply chains. Two different models for two different supply chains operate with the same logic. First of all, the daily production schedules are based on the monthly production quantities calculated by the model. This monthly numbers are then converted to daily numbers based on the 4+4+5 week split of a quarter. Therefore the first month in a quarter is assumed to have 4 weeks, and this monthly production quantity is divided by $4 \times 7 = 28$ days. This is same for the second month. The third month of the quarter is however divided by $5 \times 7 = 35$ days. These daily numbers are then read by Arena simulation model.

Like the BPS solvers, Excel ATM models are also called monthly. The inputs, such as ADI and CW inventory levels and scheduled FSM production information are updated and come from the Arena simulation model at the beginning of each simulation month. The decisions for monthly ATM productions are given by a single variable optimization model, in which the objective function is to minimize the difference between the demand and supply levels and the only decision variable is the ATM production (supply) quantity. The constraint is the available ADI inventory for production based on ADI inventory targets.

The decision components are illustrated in Figure 6, although the production

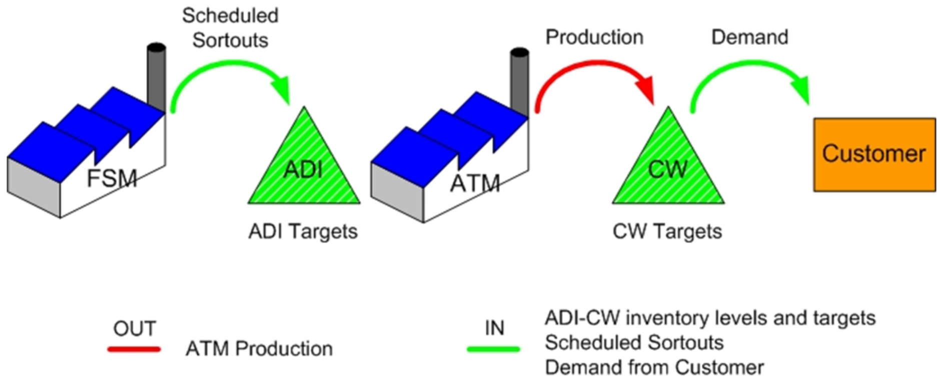


Figure 6: ATM production decision components: inputs and the output

strategy changes little bit among the simulated product families with respect to the product's life cycle curve. For instance, new products' build strategy is to "produce if the demand is present" versus a mature product's build strategy might be to "push as many products to CW". The reason is that for a mature product, Intel wants to push the supply chain pipeline inventory to the CW, where it can be immediately sold to customers before a newer product fully transitions to its place. As these production strategies might change over time and from product to product, we try to find out the best strategy for each product family that gives the closest estimate of the supply chain operations for that product. We give more details on this topic in the Model Validation at the end of this Section. The general optimization model for a product is give below:

Minimize

$$|Sup(t) - [JD(t) - CW(t-1) + CWtar(t)]| \tag{21}$$

Subject to:

$$Sup(t) \leq ADI(t-1) + Sortout(t) - ADItar(t) \tag{22}$$

$$Sup(t) \geq 0 \tag{23}$$

where,

- t: months from 1 to 12
- $Sup(t)$: Decision Variable, ATM Production quantity for month t
- $ADI(t)$: ADI inventory level at the end of month t
- $CW(t)$: CW inventory level at the end of month t
- $Sortout(t)$: Scheduled arrivals to ADI inventory from FSM at the beginning of month t
- $ADItar(t)$: Target ADI inventory at the end of month t
- $CWtar(t)$: Target CW inventory at the end of month t
- $JD(t)$: Judged demand (forecasted demand) for month t

The above optimization model can be specified by a single formula in Excel, such that:

$$Sup(t) = max\{0, min\{JD(t) - CW(t-1) + CWtar(t), ADI(t-1) + Sortout(t) - ADItar(t)\}\} \tag{24}$$

Where all the inputs are known and taken from the Arena model, and $[X]^+ = max\{X, 0\}$.

Although this supply formula (and therefore strategy) works for the newer products better, we can manipulate this formula for some more mature products to accommodate for their production strategies. For example, after some point in time, we may switch from this strategy to a "push" strategy by changing the above formula to the following:

$Sup(t) = max\{0, JD(t) - CW(t-1) + CWtar(t)\}$

In reality, knowing the build plan strategies for each product over time is very hard to find out for historical build plans. We try several possible strategies for a product and check the CW inventory levels over time for this product in the simulation. If there is a close match between these values versus the actual reported CW inventory levels for the same time period, then we conclude that the final strategy we tried is the best match for that specific product.

3.3.2.5 Model Validation

We built Platform Supply Chain Simulator by integrating Arena Simulation Model, ILOG OPL Optimization Model and Excel ATM Production Planning Model with the use of VBA codes. It is also automated so that it runs for one year (Q3'05 to Q2'06) without any user intervention. Now we expect the model on computer to run like actual Intel's Global Supply Chain under the same demand stream. In order to validate our model and ensure that it represents Intel's supply chain, we compare the model outputs with the actual system outputs.

The most readily available supply chain output data is the inventory levels over time. Especially for CW inventories, we can retrieve weekly inventory level changes for any Intel product. Therefore we use this information and compare these values against the simulation generated CW levels over time. In these comparison graphs we look at the point-wise correlation for the 52 pairs of values, one pair for each simulated week. We also check on the average how close we get to the real inventory levels. Figure 7 presents two validation graphs belonging to a CPU and a Chipset family.

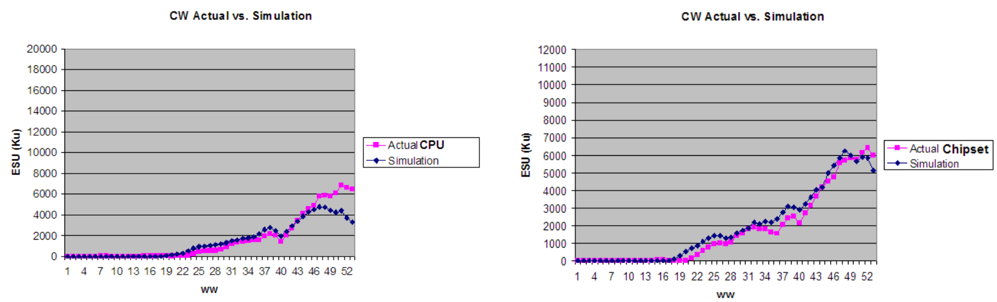
In these graphs we observe highly correlated results. We also notice immediately that the inventory level over time patterns, such as the ones that show seasonality within the quarter almost perfectly match. Quarter-end phenomenon, which happens in many industries, shows itself here too. These quarter-end sudden decreases in inventory levels are tracked each time by Platform Supply Chain Simulator.

In generation of these simulation results, all the production decision mechanisms are modeled, automated to give the best production plan for the available supply chain information at the time. These mechanisms employ very similar techniques as real planners use in today's Intel. So the very close match between actual inventories versus simulation inventory levels show that FSM and ATM production planning models are estimating their actual counterparts pretty well. They react very similarly to the changes in the demand, forecast and inventory values; therefore the final outputs match nicely.

There is one important fact however that the product families in Figure 7 were new products in mid 2005. This means we have zero pipeline inventories, and we perfectly know the initial state of the supply chain before a single CPU or Chipset was produced in Q4 of 2005. This is obviously an advantage for our model, specifically for these products. It is clear that for the other product families, which were in their maturity in mid 2005, having a very good validation like this is harder. One of the reasons is that we cannot perfectly know the distribution of inventories in the supply chain at the time of the simulation start. Another but more important reason is that for this type of products, inventory and production strategies change over time, and they lose their inventory, capacity and sales priority to the newer products. A lot of human interventions to the previously set strategies occur, and as a result it gets much harder to analytically estimate production planning decisions with the same algorithms. For these products, as explained under Excel Model for ATM Production Planning section, we try several strategies and decide on the best match occurred so far.

3.4 Main Results

With the Monte Carlo approach we developed in Section 3.3.1, we analyzed the forecast improvements of the proposed system. In the proposed system, the orders are placed as platform orders, therefore the actual dependency between the platform components is known to a certain extend. For instance when a customer places an



(a) CPU Family

(b) Chipset Family

Figure 7: Comparison of Simulation Model Output versus Actual inventory levels over time

order of 100 platforms, we would know that we have to supply 100 chipsets first, then 100 wireless cards after two weeks, and finally 100 CPU's two weeks after the wireless card delivery. As explained in Section 3.3.1, we also consider the possibility that some parts of platform orders might change. So a customer actually might cancel the rest of the order after the chipset part is delivered. In order to account for order cancelations and order changes we assume that any remaining part of the platform order can be altered by 10% in quantity per every week until its delivery. So for our case study, wireless card order quantity can be changed from 80% to 120% of the original order quantity. And the demand distribution of this change is assumed to be uniform. Similarly, CPU part of the platform order can change between 60% to 140%.

Under this setting, we run the simulation for 52 weeks. We generate the random demand from the stochastic demand generation algorithm we developed using the modified bivariate thinning method. The demand is generated for all three components separately. For the first scenario (i.e. without platform demand information), each stream of demand is forecasted separately for the upcoming week using the single exponential smoothing technique. Every week when the new demand is revealed, it is added to the list of historical demand data set, and used to forecast the upcoming week of demand. After 52 weeks, the actual (generated) demand is compared with the forecast and forecast error is found in mean absolute percentage error.

Generating the Platform Demand: For the second scenario, we actually try to capture the platform level demand, which is currently not captured. We know that some parts of component demand are actually coming from platform orders. But the platform orders are not disclosed by the customers. Since we do not have historical data on the platform demand, we assume that most of the components are actually going to be platforms. So first, we use minimum rule to find the most number of platforms possible to build for each time period. This is like in any manufacturing process, where you have multiple components to build the final product. Materials requirement planning (MRP) uses Bill of Materials (BOM) structure to find the maximum number of finished goods that can be produced with the current component inventory. In our case it is easier to calculate this, because the BOM structure has one components of each type to build the platform. But we have to consider the time-lag that each component is ordered. In fact, the chipset ordered at time "t", corresponds to the CPU ordered at week "t+4". So we used a time-lagged minimum rule to find the maximum possible platform quantity that could be ordered in week "t" for all "t". After finding the maximum order quantity, we assume that platform orders change anywhere between 80% to 90% of this maximum quantity. Therefore the model is generating the platform orders by: (1) Finding the time-lagged minimum of component order quantities, (2) Calculating the [80%, 90%] range of this quantity to generate the uniformly distributed platform orders. This way, we make sure that ordering that many platforms is feasible.

Using advance demand information: Since now we capture the platform orders in the second scenario, we can use this extra information to tell something about the future component deliveries, i.e. wireless card and CPU's. Although the remaining components may change in quantity, we still would know the average units of wireless cards and CPU's needed for that platform order. So we will in fact use these averages as the unbiased estimators of the known part of the demand, and we will only forecast the remaining non-platform component orders, therefore reducing the uncertainty. The value is coming from knowing the majority of the future wireless and CPU orders with some confidence. The more we know, the more we will improve the forecast accuracy. Running the model for 1000 replications, each with different random demand stream, we concluded that CPU forecast error can be reduced by 4%, whereas wireless card forecast error can be reduced by 8% in MAPE. When we convert these improvements in relative terms, it is a [9.9%, 10.7%] reduction in CPU MAPE at 95% confidence level; and [25.5%, 26.7%] reduction for wireless card at 95% confidence level. The reason that the Wireless card forecast improvement is more than the CPU products is mostly from the fact that wireless cards are less in order quantity and platforms orders constitute a bigger portion of the overall wireless card orders. This means we know more about the wireless card future demand than we know CPU products. Another reason to this is that the uncertainty of the future demand associated with wireless card is less than the CPU as it only waits 2 weeks versus 4 weeks for CPU, where every week of wait adds another 10% demand variability.

Above results assume that majority (i.e. 80% to 90%) of the individual components are part of platform orders. Iterating this assumption to understand the impact of the proportion of platform orders on the overall order quantity, we re-run the model by generating the platform orders as 70-80%, 60-70% and so on until the extreme case of having no platform orders. Figure 8 shows the corresponding forecast error reduction for both Wireless cards and CPU products. Starting with no platform order case, where we have no additional information, therefore no savings, as we increase the percentage of platform orders we observe diminishing returns.

Supply Chain Impact of Forecast Error Savings due to Platform Demand Information: As we mentioned previously, we build "Platform Supply Chain Simulator" to be able to quantify the supply chain impact of the forecast figures.

Now that we have the forecast improvement results from the Monte Carlo model, we can run our platform supply chain simulator with the improved set of forecasts. To be able to measure the inventory impact on the supply chain, we assume that all of the forecast error savings are decreasing the actual forecast figures. In other words, we assume the current forecasting system is over-estimating the demand; therefore we use the savings to reduce the forecast numbers. After calculating forecast savings over the 52 week horizon with the Monte Carlo Model, we recalculate the forecast figures and re-run the platform supply chain simulator.

Platform supply chain simulator is capable of tracking inventory levels continuously at any given supply chain node. We tracked and recorded the inventory levels at the most critical nodes at Intel supply chain, which is the work-in-process inventory in the FSM, Assembly Die Inventory (semi-finished goods) and Component Warehouse (finished goods). Results show that the simulated CPU products can have significant inventory reduction without any loss in the service level at the product family level. Overall supply chain inventories are estimated to be reduced by 6.7% and 5.1% for the two CPU product families tested.

The advantage of creating the platform supply chain simulator is many-fold. First of all this supply chain model is validated to behave like real Intel supply chain. So instead of forecasting scenarios, any major scenarios can be run to quantify the performance difference of the supply chain as it relates to inventory levels. Some of the supply chain scenarios that can be run with this model are as follows:

1. Forecasting scenarios (scope of this paper)
2. Production planning scenarios (different optimization models, different strategies)
3. Lead time scenarios (transit lead times, production lead timesetc.)
4. Manufacturing Yield scenarios (yield improvement over time)

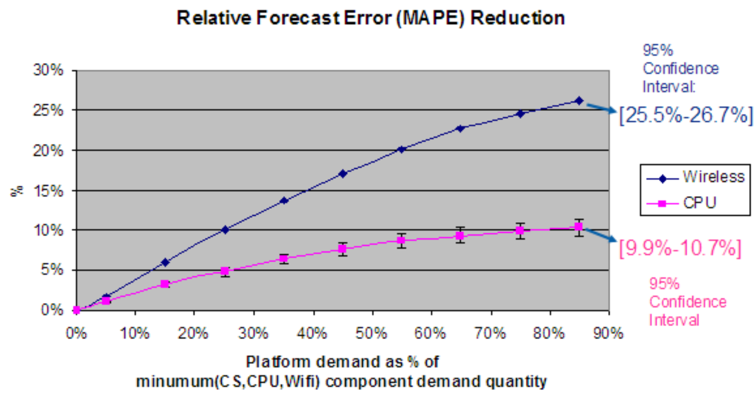


Figure 8: Effect of Platform demand as percentage of component demand on Forecast Error Reduction

5. Demand Scenarios (minor ordering pattern changes, major demand scenarios require automated forecasting model embedded in the supply chain simulation).

Considering a relatively short runtime of the simulation (i.e., 20 minutes), all of the above strategic supply chain scenarios are feasible to try on as necessary.

3.5 Conclusion

In this paper, we analyzed the value of advance demand information on platform supply chains. The advance demand information is coming from the ordering system of the customers. We assume that the customers of a supplier are placing platform orders (i.e. CPU, Chipset and Wireless Card) instead of just giving independent component orders. Knowing the rest of the platform order (i.e. future deliveries) gives valuable information from the forecasting perspective. We quantify the forecast error savings of such a system with a Monte Carlo Simulation approach using a modified Bivariate Thinning methodology to generate nonstationary demand structure. In order to quantify the forecast error improvements on the supply chain level, we build a supply chain simulation model using Rockwell's Arena software and production planning optimization models using ILOG OPL Studio and Excel and connect them together using Visual Basic for Applications (VBA) codes. Resultant simulation model called

Platform Supply Chain Simulator is validated with Intel's global supply chain and it offers many other supply chain strategic scenario analyses. Having a multi-purpose supply chain simulator may allow managers to test their strategic decisions on the computer and get insights on the supply chain impact of the potential changes.

Focusing on the forecasting area, we identified a source for advance demand information, i.e. using customer orders. The relationship between the individual components' demand can be also used without asking customers to place platform orders. One of the future directions of this research is to understand the leading indicators of CPU products from the historical time-lagged demand correlation of Chipset to CPU products and Wireless Cards to CPU products. Another research direction is to look at the product life cycle behavior of the dependent product families. For instance, if a certain Chipset product family is compatible/supporting a certain CPU product family, then how does this demand dependency shows itself in the product life cycle stages?

The implementation of this research at Intel Corporation is continuing to focus on the demand side of the supply chain. Therefore, future related research is on the predictive demand models and understanding the impact of strategic decisions such as new product introductions, price cuts and capacity creation and allocations.

CHAPTER IV

PRE-LAUNCH FORECASTING OF NEW PRODUCT DIFFUSIONS: DIFFUSION PATTERN ANALYSIS AND ALGEBRAIC ESTIMATION PROCEDURES

4.1 Introduction

Shorter product life cycles and increased product variations mean managers face the challenge of forecasting demand for new products more frequently than ever. The success of these new products in the marketplace critically depends on having accurate demand forecasts well before the product launch, since these forecasts drive important pre-launch decisions such as capital equipment purchases and capacity allocations that determine future product availability. In industries with long manufacturing lead times, such as the semiconductor industry, supply decisions must be made months before the product launch. Overestimating demand can lead to excess inventories and millions of dollars in inventory writedowns. On the other hand, underestimating demand can result in significant stock outs, and reduced market share. A survey of 168 firms by Kahn (2002) ((41)) on new product forecasting shows that informal methods based on managerial judgments and look-alike analysis are still heavily preferred by managers over more formal structured methods such as regression or diffusion models. Kahn (2002) ((41)) found that informal forecasting methods based on analogies is negatively correlated with the accuracy. In fact, the survey respondents reported that only 40% to 65% of new product forecasts are reported as accurate by the respondents; pointing out a necessity for more systematic approaches that allow managerial judgments to provide higher forecast accuracy and ease of use.

Diffusion theory, since its introduction to management science in the early 1960s, has been widely used to address many forecasting related business problems as noted by Putsis and Srinivasan (2000) ((70)) and can be a good candidate to bridge the gap between systematic analysis and managerial judgment. With its roots in epidemiology, diffusion theory attempts to model the adoption of a new product or technology over time based on certain assumptions about the target population. The most studied and widely-used diffusion models are those based on the work of Frank Bass (8). Putsis and Srinivasan (2000) ((70)) note that most diffusion modeling studies center around the work of Bass (1969) ((8)). Since its introduction in 1969, almost 750 publications based on the Bass diffusion model have explored extensions and applications and analyzed more than 2200 empirical cases (see the diffusion research database in (37)). The Bass Model forecasts the diffusion of a new product into a market over time based on an estimate of the potential market size m and two diffusion parameters p (parameter of innovation) and q (parameter of imitation), which is described in Section 4.2 in more detail. The model is simple, yet captures the essential social dynamics in terms of how much internal and external influence (through parameter of innovation and imitation, respectively) individuals have on their decisions to purchase or adopt the new product. Although there have been many extensions, according to Bass et al. (1994) ((10)) and Meade and Islam (1995) ((59)), the original Bass diffusion model tends to perform as well or nearly as well as its many extensions and is much easier to work with. Most of the extensions (discussed in Section 4.2) incorporate additional information about the market. In the case of pre-launch forecasting, however, managers often have only very limited information about the market, and so the simpler models that do not require many inputs are preferred. We, therefore, base our investigation of pre-launch forecasting on the original Bass diffusion model.

Bass et al. (2001) ((9)) acknowledge that "the most critical forecast is the forecast prior to product launch"[1]. Putsis and Srinivasan ((70)) state that "the estimates

[1](9) p.S87

of relevant diffusion model parameters early on in the diffusion process can be extremely valuable managerially and can serve a variety of purposes depending on the strategic objectives of the firm"[2]. Nevertheless, pre-launch forecasting remains the most challenging and only a limited number of methodologies and empirical studies address this important topic. Many researchers (see, for example, Thomas (1985) ((80)), Bayus (1993) ((11)), Bass et al. (2001) ((9))) have proposed adapting diffusion model parameters from similar, previously launched products to generate pre-launch forecasts for new products, a method generally referred to as "guessing by analogy". This method requires managers to assess the extent to which the new product will exhibit a sales pattern similar to those of its predecessors. The parameters for the new product are estimated via a weighted average of historical parameters.

Implementations based on "guessing by analogy" generally suffer two principal shortcomings: (1) The diffusion parameters p and q often vary significantly from product to product and (2) Managers find it difficult to relate the model parameters to tangible product or market characteristics and so to identify appropriate adjustments for the new product. Sultan et al. (1990) (79) note that only 30-50% of the variance in model parameters can be explained with meta-analytical studies that attempt to quantify how parameter values change based on the nature of the product and the market.

In this paper, we present a practical framework for analyzing historical product diffusion patterns based on a normalization approach that facilitates easy comparison and enhances intuition about the relevant parameters. Commonalities among diffusion curves from different generations can be obscured by differences in the market sizes or product life times. Normalization standardizes these aspects of the process and thereby improves our ability to identify commonalities and differences across

[2](70) p. 280

product generations. We propose several models based on the normalization approach to algebraically estimate new product diffusion parameters. We enhance the ability of our models by incorporating parameters to replace scalar parameters p and q that are relatively less intuitive and harder to judge. Like other algebraic estimation methods such as those presented in Mahajan and Sharma (1986) ((55)) and Lawrence and Lawton (1981) ((49)), some of our models allow managers to include judgmental information about tangible product characteristics such as peak time, product life time or first period shipments to improve forecast accuracy. We also provide means of fine-tuning these estimates in the context of multi-generational products. We test our models with six different multigenerational data sets. Four of these data sets include microchip product diffusion data that are provided by a major semiconductor manufacturer[3]. The other two data sets are also industry data from IBM Mainframe Computers and DRAM generations.

The rest of the paper is organized as follows. In Section 4.2, we briefly discuss the previous diffusion research with a specific focus on work related to early life cycle and pre-launch forecasting. In Section 4.3, we introduce a normalization approach and a pre-launch forecasting tool. In Section 4.4, we introduce several models that enhance model intuitiveness and improve pre-launch forecast accuracy. In Section 4.5, we compare the performance of the proposed models with the "guessing by analogy" approach on the industry data sets. We compare our pre-launch forecasts for the microchip data with those published by the company and report significant improvements. Under mild assumptions, our models reduce the median pre-launch forecast error from 30% to as low as 22%, and the average error from 46% to as low as 27% MAPE. Better estimating time-based parameters such as peak time or product life time can further improve the forecast accuracy. We discuss managerial insights based on these results in Section 4.5.3. In Section 4.6, we summarize our contributions and

[3]Due to confidentiality, we will refer to this company as X throughout the text.

discuss future research.

4.2 The Bass Diffusion Model & Pre-launch Forecasting

Meade and Islam (2006) ((61)) reviewed the last 25 years of diffusion literature from single innovation diffusion models to multigenerational models, from single market analyses to multiple country studies. Evident from the many papers they reviewed, the classical Bass Diffusion Model (8) remains the central model. The Bass model offers important insights and constitutes the nucleus of several more complex forecasting models. We review the assumptions behind the Bass model and its key extensions in the literature.

In his original paper, Bass (1969) ((8)) assumes a market population of m individuals, whose purchasing decisions are influenced by diffusion parameters p and q through the Bass Model Principle that reads "the portion of the potential market that adopts at time t given that they have not yet adopted is equal to a linear function of previous adopters", which is expressed as $f(t) = (p + qF(t))(1 - F(t))$ (see, (37)). Here, $f(t)$ is the portion of the potential market that adopts at time t, while $F(t)$ is the cumulative fraction of the market that adopts by time t. Cumulative number of adoptions by time t can be expressed as $N(t) = mF(t)$; and given the three parameters p, q and m, the Bass model forecasts cumulative sales as a function of time using equation (25).

$$N(t) = m\frac{[1 - e^{-(p+q)t}]}{[1 + (q/p)e^{-(p+q)t}]} \tag{25}$$

The derivative of equation (25) gives the rate of sales $n(t)$ as a function of time shown in equation (26).

$$n(t) = m\frac{(p+q)^2}{p}\frac{e^{-(p+q)t}}{[1 + (q/p)e^{-(p+q)t}]^2} \tag{26}$$

The Bass Model assumes that the market potential m and the diffusion parameters

p and q remain constant over time, the buyers of the new product purchase only one unit, and there is no interaction with previous or succeeding generations of the product (i.e., no cannibalization effect). The literature since the original paper Bass (1969) ((8)) has attempted to relax or eliminate many of these assumptions. Norton and Bass (1987) ((67)) extended the original model to consider successive generations of products and their substitution effects. Kalish (1985) ((42)) allowed the market potential m to be a function of price over time. Sharif and Ramanathan (1981) ((75)) modeled the market potential as a function of population growth. Other works including Kamakura and Balasubramanian (1988) ((44)) and Jain and Rao (1990) ((38)) analyzed the effect of price on the adoption rate. Kalish and Lilien (1986) ((43)) defined the parameter of imitation q as a function of changing product characteristics over time. Bass et al. (1994) ((10)) presented the Generalized Bass Model (GBM), allowing the inclusion of decision variables such as price or advertising as a function of time in the original Bass model.

Mahajan et al. (1990) ((54)) state nine assumptions of the original Bass model and review several papers that relax them. Although extensions that include decision variables such as price and advertising are found to fit some data sets better than the original Bass model, in most comparisons the original model provides an equally good fit with less model complexity ((10), (54)). Mahajan et al. (1990) ((54)) concluded that the analytical elegance surpasses the empirical validation of the derived results for the more complex models. Moreover, complex models require more data such as price, advertising and competition that may be hard to obtain, especially in pre-launch contexts.

Although many researchers and practitioners share the belief that more accuracy is needed in pre-launch forecasting, little research has been done in this area. Most studies estimate diffusion parameters p and q, while separately estimating m from market research studies. For example, Bass et al. (2001) ((9)), Thomas (1985)

((80)) and Bayus (1993) ((11)) use “guessing by analogy” together with product grouping procedures to estimate p and q. Meta-analytical studies such as Sultan et al. (1990) ((79)) and Van den Bulte and Stremersch (2004) ((82)) draw conclusions about how external information such as income heterogeneity, cultural index or product characteristics impact the parameter values. Lenk and Rao (1990) ((52)) and Xie et al. (1997) ((87)) employ Hierarchical Bayesian procedures and Kalman Filters, respectively, to dynamically update the prior distributions for p and q.

Since the diffusion parameters p and q, do not correspond directly to intuitive and readily measured market characteristics, some researchers attempt to estimate these parameters from managerial estimates of other more tangible information. The model of Lawrence and Lawton (1981) ((49)) (hereinafter referred to as the LL model) requires three pieces of information: $p+q$, first period sales $n(1)$ and the market potential m. As Putsis and Srinivasan (2000) ((70)) observe, the drawback of this method is that estimating $p+q$ is difficult. Although, Lawrence and Lawton (1981) ((49)) offer general guidelines such as $p+q$ is 0.5 for consumer goods and 0.66 for commercial goods, such generalizations fail to represent individual characteristics of a particular product. On the other hand, the model of Mahajan and Sharma (1986) ((55)) (hereinafter referred to as the MS model) rely only on managerial estimates of intuitive and measurable market characteristics including the market potential m, the time of peak sales and the peak rate of sales. However, estimating both the peak time and the peak rate of sales is difficult. While managers may have certain expectations about the peak time, producing a forecast for peak rate of sales with high accuracy is more difficult.

Only a small number of studies in the literature report on actual commercial pre-launch forecasting efforts. Choffray and Lilien (1986) ((21)) report significant improvements in forecast accuracy at a leading paper producer and a leading zinc producer after implementing a diffusion-based forecasting system. The paper producer reported projections for four-year cumulative sales with less than 30% error, while the zinc producer achieved a pre-launch forecast error of less than 15% for the first five years of cumulative sales (also see, Virrolleaud (1983) (83)). Lee et al. (2003) ((50)) describe a case of pre-launch forecasting in the music industry. They report a reduction in the MAPE of pre-launch forecasts from 69% to 52%, with further reductions to 30% as sales data become available. Both Choffray and Lilien (1986) ((21)) and Lee et al. (2003) ((50)) employ a "guessing by analogy" approach together with a database of relevant exogenous variables to enhance the parameter estimates. For many instances, however, obtaining additional data beyond historical sales is very difficult and without a systematic approach, the "guessing by analogy" method can provide poor forecasts.

Our experience in implementing diffusion based forecasting models at a major semiconductor company also suggests that managers have insights about the new products, but are unable to translate those insights into terms compatible with the diffusion parameters p and q. One immediate consequence is the frequent discrepancies between products that managers deem similar and those that actually enjoy similar diffusion parameter values. A little more structured guessing by analogy approach is implemented in a software program introduced by Liliean et al. (2000) ((53)), where available historical products are divided into four categories: (low p, low q), (high p, low q), (high p, high q) and (low p, high q) to aid in managerial judgment. However, no systematic approach is offered for selecting an analogous product, leaving this decision to the managers, who cannot intuitively select any of these categories.

In Bass et al. (2001) ((9)), the authors recognize that "the [art] of forecasting the diffusion of new products prior to product launch is still in development. Algorithms for choosing analogies are being developed."[4]. We contribute to the art and science of pre-launch forecasting by introducing a normalization framework for analyzing historical diffusion patterns. This framework helps managers better understand various diffusion patterns and select the best representative pattern. In Section 4.3, we introduce volume and time normalization procedures that help isolate the effects of the diffusion parameters p and q on the forecast. We then introduce a pre-launch forecasting tool that allows managers to produce a pre-launch forecast by just estimating the product life time and the market potential of a new product. In Section 4.4, we introduce several algebraic estimation procedures to estimate the Bass diffusion parameters from other parameters that are related to tangible market characteristics and so are more intuitive to estimate. Normalization procedure is used in the multigenerational context to help estimate these more intuitive parameters, where management judgment can also be employed in further fine-tuning. We test proposed models with real world data in Section 4.5 and show significant pre-launch forecast accuracy improvement opportunities.

4.3 Normalization and Blueprint Approach

Differences in market potential and product life can obscure similarities in the diffusion curves of similar products or between generations of a single product line. In Section 4.3.1, we introduce a normalization approach to help managers identify the similarities in historical product diffusions by scaling both volume and time. In Section 4.3.2, we propose a simple pre-launch forecasting model, called the Blueprint approach, where managers only need to estimate the new product's market potential and life time to forecast sales for the next generation product.

[4](9) p. S91.

4.3.1 Normalization Approach

Differences in market potential and product life times can obscure similarities in diffusion curves of similar products and even in different generations of the same product. The influence of market potential is immediate, intuitive and simple to identify. The influence of product life time, however, is more subtle and is buried in the unintuitive parameters p and q. For example, simply scaling the parameters p and q of a diffusion curve to pT and qT with some constant $T > 0$ maintains the essential features of the diffusion, but changes the time scale on which those features play out. Note that as t goes to infinity, $N(t)$ increases to the market potential m, i.e., if the company continues to market the product in perpetuity, its cumulative sales will approach the market potential and the product will saturate the market. In this sense, a Bass diffusion curve assumes the product life time is infinite. In practice, every product has a finite, though often indistinct product life time and organizations carefully pace the introduction of new generations and manage product roadmaps based on these life times. To normalize out time scale effects we need to establish a finite product life time T that is both consistent with managerial practice and has the appropriate analytical properties for forecasting. Among the several possible definitions, we consider the following three:

Definition 1: Given a fraction $0 < \alpha < 1$, define the product life time to be the time at which cumulative sales reach the fraction $(1 - \alpha)$ of the market potential, i.e., the product life time T_1 is defined so that

$$N(T_1) = (1 - \alpha)m \tag{27}$$

It is not difficult to verify that

$$T_1 = \frac{ln(\frac{p+q-\alpha q}{\alpha p})}{p + q} \tag{28}$$

satisfies equation 28. This definition is simple, intuitive and is consistent with much of industry practice. in those industries where management decides to end the active management of a product when the remaining market size is small and moves resources to new products instead of continuing to invest in a vanishing market. The appropriate choice of the specific fraction α can and should change from industry to industry, company to company and even product segment to product segment. For high-volume, low-margin "commodity" products α should be large. For high-margin, low-volume "customized" products, the appropriate choice of α will likely be small.

Definition 2: Given a fraction $0 < \beta < 1$, we can define the product life time to be the time T_2, after which the rate of sales never exceeds the fraction β of the peak sales rate, i.e.,

$$T_2 = min\{t \geq 0 : n(\tau) \leq \beta n(t^*), \forall \tau \geq t\} \tag{29}$$

It is not difficult to verify that

$$T_2 = \frac{ln(\frac{q(2-\beta)+2q\sqrt{1-\beta}}{\beta p})}{p+q} \tag{30}$$

satisfies equation 30. Note that if $p > q$, then $t^* < 0$ meaning that the sales rate will begin below $n(t^*)$ and decrease. This definition can nevertheless be applied if we adopt the convention that the maximum rate of sales is taken over all values of t, including values less than 0. Definition 2 is appropriate in capital intensive industries where capacity utilization is a priority. When change over costs are high or economies of scale are great, the fraction β should be larger. For flexible manufacturing systems, where the capacity can be easily transferred to the newer generations at low cost, β should be lower.

Definition 3: When $q > p$, the peak rate of sales occurs at a time $t^* > 0$ and, given a scalar $k > 0$, we define the product life time T_3 to be $(1 + k)t^*$. Thus, under this definition, the product life time is given by:

$$T_3 = (1+k)\frac{ln(q/p)}{(p+q)} \tag{31}$$

This definition is preferred for situations when the end of life decisions depend on the duration of declining sales. Since after the peak sales, the sales decline will continue for a (relatively) long time, if the management do not redeploy assets, they run the risk of falling behind the technology curve. For multigenerational products, the timing of new product introductions may be seasonal, such that after the peak season for the old generation product, there is a constant time period to introduce the new generation on time, so that the new generation can also has its maximum sales potential during the next peak season. For example, few months before the holiday season, PC manufacturers are releasing their newest products so that it can reach its maximum potential during the holiday season. However, after the peak season, there is only a constant time until the next holiday season, therefore the active management of each product generation ends after a constant multiple of the peak time. Usually, the number of generations being actively managed is limited, therefore it is also possible to end a product life cycle at the time of the newest product introduction.

The planning and operational implications of product life time can be quite varied ranging from the cessation of sales to moving the product into secondary markets or shifting its production to secondary capacity, etc. Whichever definition of the life time T we choose, we can use that value to normalize a diffusion curve so as to isolate the effects of time from those of the diffusion parameters p and q. In particular, given a Bass diffusion curve $N(t)$ with market potential m, diffusion parameters p and q and life time T, we define the normalized cumulative diffusion curve:

$$N(t;T) = \frac{[1 - e^{-(p+q)Tt}]}{[1 + (q/p)e^{-(p+q)Tt}]} \tag{32}$$

Note that $N(t;T)$ is a Bass diffusion curve with market potential 1 and diffusion parameters pT and qT so that the product life time is scaled to 1 as well. Scaling time by a factor of T scales the sales rate by the same factor, in particular, the normalized rate of sales as a function of time is:

$$n(t;T) = T\frac{(p+q)^2}{p}\frac{e^{-(p+q)Tt}}{[1+\frac{q}{p}e^{-(p+q)Tt}]^2} \tag{33}$$

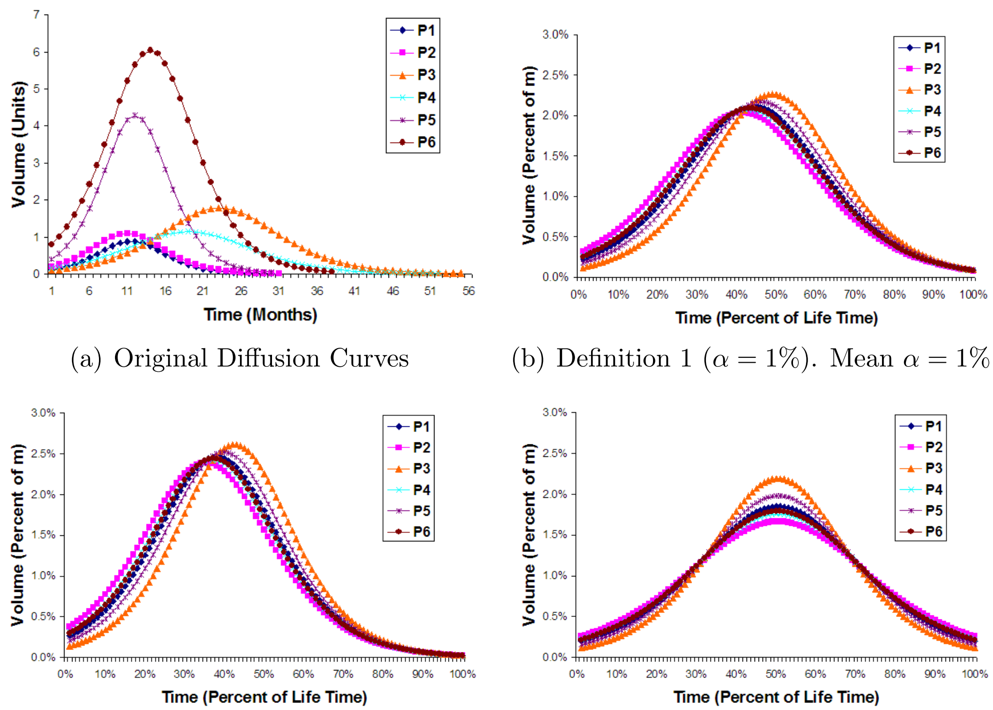
i.e., the normalized rate of sales is given by $n(t;T) = Tn(Tt)$. Each of our three proposed definitions is consistent with this scaling in that if we define the product life time T for the original diffusion curve $N(t)$ according to one of those definitions, then according to the same definition the product life time of the scaled curve $N(t;T)$ will be 1. This can be shown by observing that $N(1;T_1) = (1-\alpha)$ for Definition 1, $n(1;T_2) = \beta n(t^*;T_2)$ for Definition 2; and $t^* = 1/(1+k)$ for Definition 3, where t^* is the peak time for the normalized curves for the last two observations.

Figure 9 shows the historical and normalized diffusion curves for 6 product generations from a microchip product line using each of our three definitions for product life. As seen from this figure, eliminating the differences attributable to product life and market potential reveals the similarities among the diffusion patterns within the product line. In the rest of the paper, we will use the Definition 2 that relates the end of life to the level of peak sales, a definition consistent with the semiconductor industry.

Managers can use the normalized curves to estimate the diffusion parameters p and q of a normalized curve for the new product, separately estimate the new product's life time T and market potential m and reverse the normalization process to obtain a diffusion curve forecasting sales of the new product. In particular, the cumulative Bass diffusion curve for the new product is:

$$N(t) = m\frac{1-e^{\frac{p+q}{T}t}}{1+\frac{q}{p}e^{\frac{p+q}{T}t}} \tag{34}$$

and its derivative



(a) Original Diffusion Curves

(b) Definition 1 ($\alpha = 1\%$). Mean $\alpha = 1\%$

(c) Definition 2 ($\beta = 1\%$). Mean $\alpha_\beta = 0.3\%$

(d) Definition 3 ($k = 1$). Mean $\alpha_k = 2.8\%$

Figure 9: Original and Normalized Diffusion Curves based on Different Life Time Definitions.

$$n(t) = \frac{1}{T}\frac{(p+q)^2}{p}\frac{e^{-\frac{p+q}{T}t}}{1+\frac{q}{p}e^{-\frac{p+q}{T}t})^2} \tag{35}$$

gives a forecast of the rate of sales. We refer to this process of obtaining a forecast from a normalized diffusion curve as *de-normalizing*. It is instructive to observe the effects of moving from a historical Bass diffusion curve with market potential m, diffusion parameters p and q and product life time T to a diffusion curve for a new product with market potential m' and life time T' via this process of normalizing and de-normalizing. The resulting forecast is a Bass diffusion curve with market potential m' and diffusion parameters $p\frac{T}{T'}$ and $q\frac{T}{T'}$.

Section 4.3.2 introduces an approach, called Blueprint approach, that helps managers select the most likely diffusion pattern for the upcoming new product, and produce an easy pre-launch forecast using the de-normalizing.

4.3.2 Blueprint Approach

The main purpose of the normalization approach is to better identify how to select the representative (i.e., analogous) curves. Using normalization, managers can choose relevant curves from the pool of normalized curves for historical products that they believe represent the upcoming product. Averaging the selected normalized curves based on certain criteria would give a baseline expectation for the new product. We call this curve the “Blueprint” curve, because it gives managers a blueprint or a template to construct their future expectations.

There are multiple alternative approaches to produce a Blueprint curve from a set of normalized curves.

1. Pick individual normalized curve and de-normalize.

2. Average the diffusion parameters of normalized curves.

3. Average the normalized diffusion curves themselves.

4. Find a normalized curve that best fits the given normalized curves and de-normalize that.

We now investigate each of these options based on their technical advantages and disadvantages and their easiness for implementation.

First option is to pick individual normalized curves as representative scenarios for the new product, and de-normalize them for forecasting purposes as explained in Section 4.3.1. This approach is the simplest of all and it is easy to implement. However, selecting the representative curves would require managers to analyze the reasons of their choice and justification of why they choose a specific curve. In the situations where the normalized curves are clearly identified and associated with certain business context, this approach can produce good forecasts. The success rate of this method would depend on which normalized curves are chosen and whether those curves would represent the upcoming product accurately, besides the estimates of m and T for the new product.

Second option distributes the risk of choosing the wrong representative curve by averaging the normalized parameters p_iT_i and q_iT_i of all the historical products $i \in I$, where I represents the set of historical products that are selected as possible candidates. Instead of simple averages, weighted averaging can also be applied, so that the estimated parameters p and q would equal to:

$$p = \sum_{i \in I} w_i p_i T_i, \qquad q = \sum_{i \in I} w_i q_i T_i \tag{36}$$

where the w_i's are the weights for the averaging and so are non-negative and sum to 1. The technical disadvantage of this formulation is that the averaged parameters p and q may not correspond to a normalized curve. In other words, their product life time may not equal to 1, due to differences in original life times (i.e., T_i) of the normalized curves. But since a closed-form solution for the product life time is available, we can scale the averaged parameters to achieve a life time of 1, such that:

$$p' = pT \qquad q' = qT \tag{37}$$

where T is the life time of diffusion curve with parameters p and q, based on the selected life time definition. Having fixed the normalization issue, this approach is also relatively easy to implement.

On the other hand, if we average the normalized curves themselves rather than averaging the parameters, then we obtain a curve that is generally not a Bass diffusion curve. We cannot talk about formal normalization of this curve as it is not bound to any specific functional form. However, the resultant curve is averaged over the unit life cycle, therefore it will have a life cycle of 1. And since the normalized curves captured a market volume close to 1, the averaged curve will also contain almost the same amount until time 1. We can de-normalize this curve to obtain a forecast for the new product. But the issue with this approach is that a resultant forecast will not reflect a true diffusion pattern. It will correspond to average market penetration across all historical products for a given phase in the product life cycle.

Finally, in order to find the best representative Blueprint curve, in the form of a Bass curve, we can fit a curve to find the parameters p and q that minimizes the (weighted) sum of the square errors between the representative curve and the normalized historical curves, such that:

Minimize

$$SSE(p,q) = \sum_i w_i \int_0^\infty (bass(t,p,q) - n_i(t;T_i))^2 dt \tag{38}$$

Subject to:

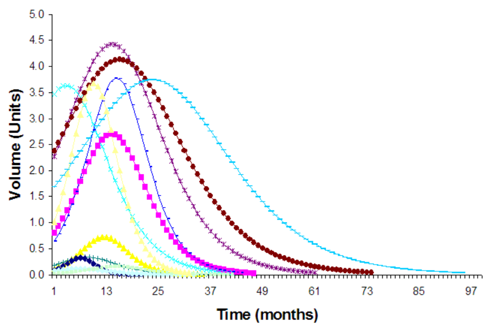
$$p + q = ln(\frac{p + q - \alpha q}{\alpha p}) \tag{39}$$

$$p > 0, \quad q > 0 \tag{40}$$

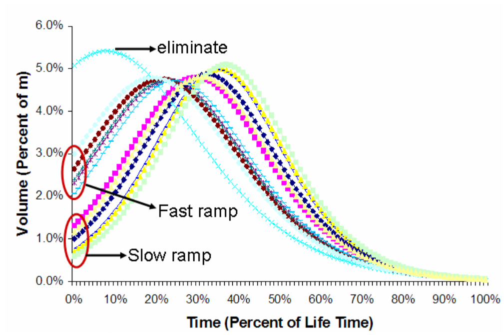
where $bass(t,p,q)$ is the best fitted Bass curve and $n_i(t;T_i)$ is the normalized diffusion curve for product i.

This approach is the most complicated one among other alternatives, and it is not practical for implementation. Although, this best fitted normalized curve may represent the historical normalized curves with minimal sum of square error, managers would still want to consider multiple scenarios in pre-launch forecasting and they will want to understand a range of cases.

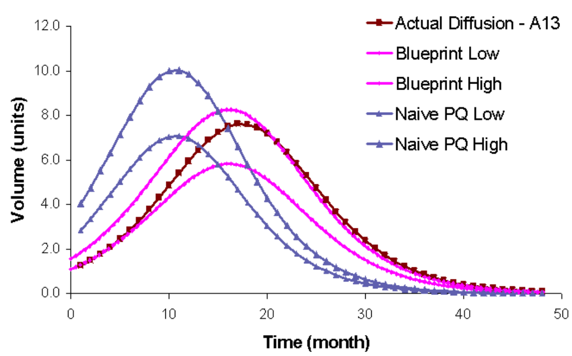
In order to illustrate the benefits of using the Blueprint approach and the normalized curves as a visual tool, we use a product line that consists of 12 real-life product generations. Figures 10(a) and 10(b) present the actual best fitted diffusion curves, and their normalized versions, respectively. Using the Blueprint approach, we want to construct a blueprint curve for the upcoming new product launch (product A13), which is expected to have a four years of life cycle ($T = 48$ months) and an uncertain market potential of 120 to 170 units. Analyzing the normalized curves, two groups of diffusion types are identified as fast vs. slow ramp, containing 5 and 6 product generations respectively. One of the diffusion pattern is observed to be unique, and it is eliminated from consideration. Product A13 is assumed to belong to the slow ramp family of products and the corresponding blueprint curve is constructed by averaging the normalized curves of slow ramp group and de-scaling with the worst and best case scenarios for market potential (i.e., $m = 120$ and $m = 170$) and product life time of four years (i.e., $T = 48$). For comparison purposes, we assume that a naive manager would like to use historical diffusion curves, but instead of normalizing the curves she directly averages the p and q parameters (let us call this method "Naive PQ"). Then she also uses the same market potential estimates as the manager using the Blueprint approach (i.e., 120 and 170 units). Figures 10(c) shows the results of both methods compared with the actual diffusion of product A13. Without reporting the obvious forecast error difference, we want to point out the importance of (1) analyzing historical products on the same scale and understanding different diffusion types, (2) using the estimate of product life in pre-launch forecasts.



(a) Original Diffusion Curves



(b) Normalized Curves (Def.2 with $\beta = 1\%$)



(c) Comparison of the Blueprint and the Naive PQ method

Figure 10: Pre-launch forecasting scenario using Blueprint and Naive PQ method.

Normalization and Blueprint approaches are heuristic solutions and handy tools for forecasters. Managerial inputs and judgments on product life time T and total market potential m are necessary to use these tools for forecasting purposes. These methods provide an intuitive big picture analysis by setting a standard for characterization of the historical product diffusion patterns and a benchmark for easy comparison.

One advantage of normalization is that simple features of the normalized curve completely define the entire curve: Specifically, if we know the time t^* of the peak of the normalized curve and the value $n(t^*)$ of the curve at the peak, then we know the entire normalized curve. This suggests that rather than finding a representative curve for the next product diffusion, we might instead find a representative for the set of points (i.e., $(t_i^*, n_i(t_i^*))$: $i \in I$) that define the peak positions and therefore represent the curves. Since normalization approach scales the peak position for the historical products, the managerial judgment of these values would be relatively simple to do on the same scale than judging the peak positions of the original curves with different scales.

With this motivation, Section 4.4 introduces formal algebraic models to estimate diffusion curve parameters from other sources of information such as percentage-based, time-based and unit-based parameters that can be used to define the diffusion peak position.

4.4 Algebraic Estimation Procedures for Diffusion Parameters

Commonly used "guessing by analogy" is designed to make it simpler for managers to easily judge the similarity of product being launched to that of historical products. However, one needs to recognize that all products are different and it is hard to standardize the basis of the managerial judgments. For example, one can think that DirecTV is a subscription based product, so it should be similar to another subscription based TV product: CableTV; others might consider the similarity of the product price. Yet no one knows which product characteristic or which external variable would be the most influential on the diffusion pattern. Instead of focusing on analogous products and indirect similarities such as product characteristics to estimate original Bass model parameters, we visually analyze historical diffusion patterns. We propose percentage-based parameters and other intuitive pieces of information such as peak time, product life time and first period sales to be judged by managers. We then use these as inputs to calculate implied diffusion parameters.

In Table 7 we present an extensive list of potential parameters that can be used to estimate a diffusion curve, which are either scalar, percentage-based, unit-based or time-based. Table 8 summarizes and compares the current algebraic estimation procedures in the literature with our models based on their information requirements. Our main focus is to move away from less intuitive scalar parameters p and q towards more intuitive percentage-based abc parameters and other inputs that can help us formally define diffusion curves easily. Most of the definitions in Table 7 can be applied to any diffusion model. For the Bass model, percentage based parameters a, b and c can be calculated using the following equations:

$$a = \frac{ln(q/p)}{(p+q)T} \qquad b = \frac{(p+q)^2}{4q} \qquad c = \frac{q-p}{2q} \tag{41}$$

All of our models share the same assumption that we can estimate market potential m from marketing research studies together with managerial judgment. Literature also supports this assumption by focusing mainly on p and q estimations. Bass (1969) ((8)) states that “the parameter for which one has the strongest intuitive feeling is m”. However, one cannot find the same intuitiveness for parameters p and q. Hence we focus on estimating these parameters from other (more intuitive) sources of information. For convenience, model names are constructed from the required input parameters (i.e., the $b-c$ model requires b and c parameters to estimate p and q). Some model inputs can be more (or less) intuitive for different businesses or industries.

Table 7: List of various diffusion parameters by category that are required by different models

Parameter	Description
p	Parameter of innovation
q	Parameter of imitation
$p+q$	Sum of parameter p and q
q/p	Ratio of parameter q to p
a	Normalized Peak Time (t^*/T)
b	Normalized Peak Height ($f^* = n^*/m$)
c	Fraction of adopters at peak ($F^* = N^*/m$)
n^*	Noncumulative sales at peak
N^*	Cumulative sales at peak
$n(1)$	First period Sales
m	Market potential
t^*	Peak Time
T	Product Life Time

To help managers select the best model, in Section 4.5 we employ sensitivity analysis to compare the forecasting performances of these models under varying levels of input accuracy.

4.4.1 b-c Model

As an alternative to Mahajan and Sharma (1986) ((55)) (MS1 and MS2) models (see table 8), the $b-c$ model does not require the time of the peak t^* as an input. Algebraic rearrangement of equation (41) gives equation (42), which can be used to calculate p and q directly from b and c, while m is estimated separately.

$$p = \frac{b(1-2c)}{(1-c)^2} \qquad q = \frac{b}{(1-c)^2} \tag{42}$$

The advantage of this model over MS models is that the managers can estimate percentage-based b and c parameters instead of unit-based n^* and N^*, by comparing the normalized curves of the historical products on the same scale. Since m is estimated separately, this method does not require the estimation of t^*, which can be calculated from p & q or directly from b & c as follows:

Table 8: Comparison of Algebraic Parameter Estimation Procedures by their input requirements.

	Information Required to Estimate a Diffusion Curve			
Model	**Scalar**	**Percent**	**Unit**	**Time**
Bass (1969)	p q		m	
Mahajan & Sharma (MS1) (1986)			n^* m	t^*
Mahajan & Sharma (MS2) (1986)			n^* N^*	t^*
Lawrence & Lawton (LL) (1981)	$p+q$		$n(1)$ m	
Blueprint Method			m	T
$b-c$ Model		b c	m	
$b-t^*$ Model		b	m	t^*
$c-t^*$ Model		c	m	t^*
$n(1)-t^*$ Model			$n(1)$ m	t^*
Time Controlled Bass Model			m	t^* or T
$a-T$ Model		a	m	T
$a-t^*$ Model		a	m	t^*
$T-t^*$ Model			m	t^* T

*TCB is an adjustment procedure. It also requires initial estimates of p and q.
**Blueprint Method requires historical diffusion pattern analysis.

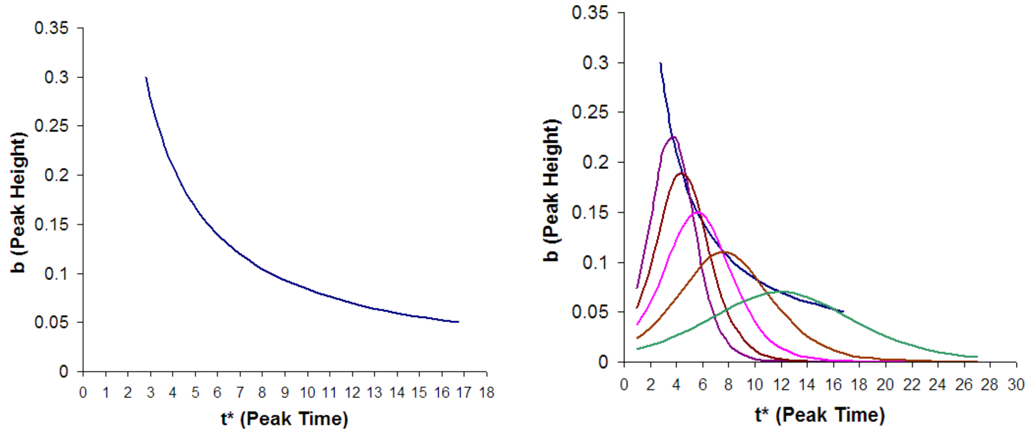
$$t^* = \frac{ln(q/p)}{(p+q)} = \frac{ln(1-2c)^{(c-1)}}{2b} \tag{43}$$

Using equation (43) for the peak time, one can derive two more models: $b-t^*$ and $c-t^*$. In both of these models, market potential m is estimated separately, while the other inputs are used to estimate p and q. The $b-t^*$ Model corresponds to MS2 Model, since MS2 model requires m, n^* and t^* as the inputs and b is implied by n^* and m (i.e., $b = n^*/m$).

4.4.2 $b-t^*$ and $c-t^*$ Models

Peak time can be calculated by inputs b and c. If we know the peak time t^* and one of either b or c, we can calculate the other. Then we can convert b & c into p & q as in the $b-c$ model. Equation (43) gives us the opportunity to estimate any two of the three unknowns that determine the peak position. Following peak position charts in Figure 11 help managers see the tradeoff between b, c and t^*. It visualizes the diffusion curve peak position to aid in judgment process.

Based on the confidence in estimating these three inputs, managers can choose any



(a) Peak Position Path (b) Example Curves

Figure 11: Peak position Chart and Visualization of different peak height (b) and peak time (t^*) combinations (c is fixed at 0.48)

two out of three parameters that they have more confidence. To help assess different levels of confidence we can construct the feasibility region for the peak position as in Figure 12, which defines the area for the expected peak position for the new product based on historical peak positions. Upper and lower bounds of each parameter can be calculated from the span of historical diffusion pattern analysis with normalization, while the most likely case is given by the Blueprint average of b and c. Managers can then visually adjust these bounds with their judgments.

4.4.3 $n(1) - t^*$ Model

This model can be regarded as the hybrid model of Mahajan and Sharma (1986) ((55)) and Lawrence and Lawton (1981) ((49)). We pick the most intuitive pieces of information in both of these models (i.e., t^* from MS and $n(1)$ from LL) to estimate the original parameters p and q. With m estimated separately, we solve the following two equations simultaneously:

$$N(1) = n(1) = m\frac{(1 - e^{-(p+q)})}{1 + \frac{q}{p}e^{-(p+q)}} \tag{44}$$

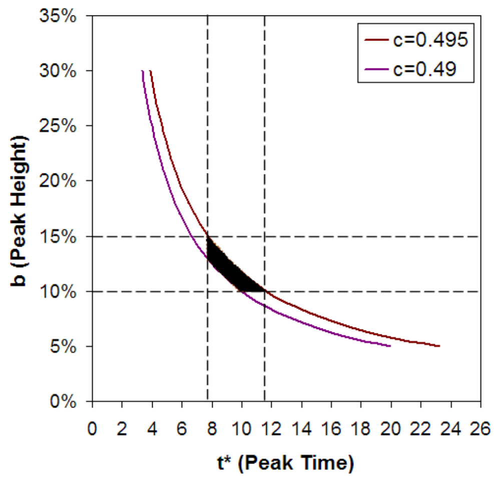


Figure 12: Feasible region for peak position. Upper and lower bounds for b, c and t^* determined by the diffusion pattern analysis define the area for peak position.

and

$$t^* = \frac{ln(q/p)}{(p+q)} \tag{45}$$

Since no closed-form solution exists for p and q, we first numerically find the set of (p,q) pairs that give the estimated t^* value using equation 45. Figure 13 presents a chart called Iso-peak time PQ chart showing these (p,q) pairs.

Then we use the formula in equation (44) to pick the (p,q) combination that gives the desired level of first period sales $n(1)$. This method eliminates both p & q and b & c estimations and focuses on estimating m, $n(1)$ and t^*. As explained by Lawrence and Lawton (1981) ((49)), both m and $n(1)$ can be obtained from market research techniques. In order to reduce some uncertainty, one can also wait to observe the first period sales and fix the value of $n(1)$. Estimation of t^* can come from various sources of information such as product roadmaps and seasonality information such as time of peak selling season (i.e., November-December for high tech products). We believe that estimating t^* instead of $p+q$ is more intuitive for managers, since it is closely related to business practices. Performance of this model is separately analyzed in Section 4.5.2.1 with Color TV adoption data, and found that it can match the performance of MS1 model, while requiring easier to guess information ($n(1)$ vs. $n(t^*)$).

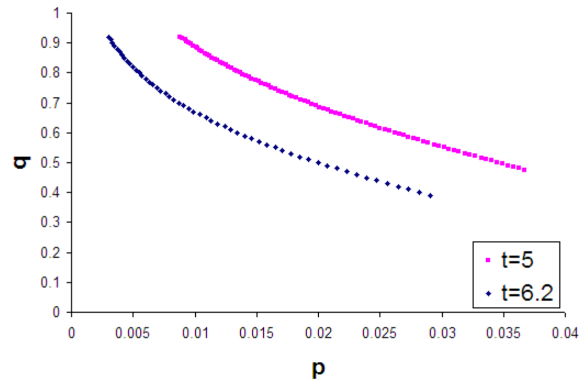


Figure 13: Iso-peak time (p,q) combinations that give the same peak time t^*

4.4.4 Time Controlled Bass (TCB) Model

In Section 4.3.1 we introduced the time scaling method. Time Controlled Bass Model employs time scaling method to use the initial estimates of the p and q parameters, and adjust them with equation (46) to account for any known or estimated timing information.

$$\hat{p} = p_0 \frac{T_0}{\hat{T}} \qquad \hat{q} = q_0 \frac{T_0}{\hat{T}} \tag{46}$$

where p_0 and q_0 are initial parameter estimates (that may be calculated from Naive PQ method, or with analogous products chosen by diffusion pattern analysis of the normalized historical products) and T_0 is the product life calculated from these initial parameters.

In this regard, it is the same model suggested in the Blueprint approach. We include this approach in our algebraic models for comparison purposes, but we also extend its capability to use any critical timing information such as peak time t^*, instead of just product life time. Time controlled Bass Model is an adjustment procedure to account for timing information. If the initial parameter estimates create peak time or product life time that is not parallel to the managerial expectations, above formulas can be directly used to adjust for the known or estimated time. Other than

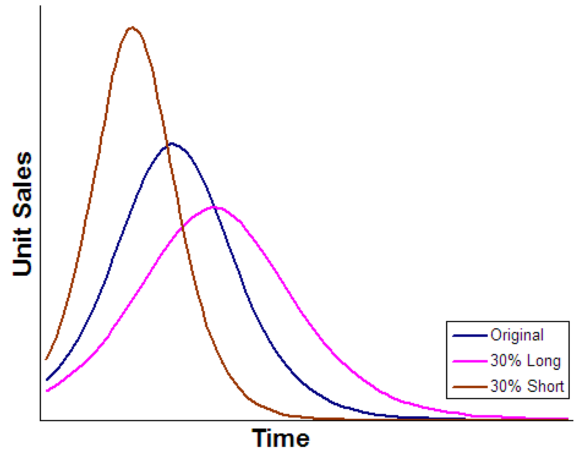


Figure 14: 30% extended vs. 30% shrunk life cycle

market potential m, it is sufficient to estimate t^* or T to run this model, however, the model performance will depend on the ratio of the initial parameters q_0/p_0, which stays unchanged after the adjustment, while the sum of the parameters decreases (increases) for longer (shorter) product life or peak time.

Without employing any definitions for product life T, one can still use the Time Controlled Bass Model by scaling the time axis with a scaling factor θ, where $\theta < 1$ extends the time axis to a longer horizon, and $\theta > 1$ shrinks the time axis to a shorter horizon by "$|1-\theta|$" percent. Equation (47) presents required formulas, while Figure 14 illustrates an example with $\theta = 1.3$ and $\theta = 0.7$.

$$\hat{p} = p\theta \qquad \hat{q} = q\theta \tag{47}$$

4.4.5 $a - T$ Model

Although b and c parameters are related to a and T parameters through the calculation of peak time t^*, there is no closed-form solution that only uses a and T to estimate diffusion shape parameters of p and q. Therefore, the $b-c$ Model cannot be algebraically transformed into the $a-T$ Model. However, empirical study of 38 Microchip product families (under 4 product lines) shows a strong negative correlation between b and T (between -0.883 and -0.975), and positive correlation between c and

Table 9: Estimates and fit statistics of linear regression models suggested for parameters b and c.

Model for b		**Model for c**	
Term	Estimate (α_i)	Term	Estimate (β_i)
Intercept	0.4745	Intercept	1.1686
T	0.0013	a	-0.8155
ln(T)	-0.1251	ln(a)	0.3814
RSquare	0.951	RSquare	0.998
RSquare Adjusted	0.948	RSquare Adjusted	0.998
RMSE	0.00518	RMSE	0.00129
Mean of Response	0.0667	Mean of Response	0.4729
Observations	38	Observations	37[1]

[1]One data point is excluded from regression model for c with studentized residual of 5.527

a (between 0.945 and 0.987) within each of the four product lines. Performing linear regression analysis on this data set of parameters, following simplified regression models, for b and c respectively, are suggested after employing stepwise elimination of less important variables.

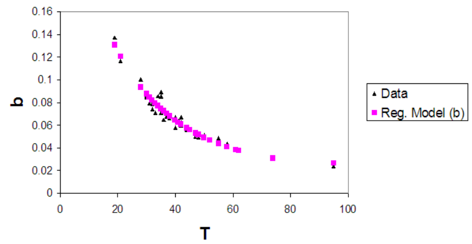
$$y_b = f_b(T) + \epsilon = \alpha_0 + \alpha_1 T + \alpha_2 ln(T) + \epsilon \tag{48}$$

$$y_c = f_c(a) + \epsilon = \beta_0 + \beta_1 a + \beta_2 ln(a) + \epsilon \tag{49}$$

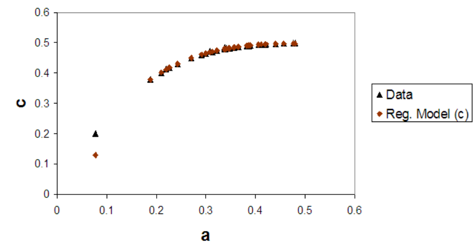
Both models have high levels of R-square values ($R_b^2 = 0.951$ and $R_c^2 = 0.998$), suggesting that they can be used to estimate parameters b and c with high confidence. Since the $b - c$ model can directly estimate diffusion shape parameters, using these regression equations, one can derive the $a - T$ model by substituting regression functions $f_b(T)$ and $f_c(a)$ into the places of b and c. Equation (50) defines the $a - T$ model formulas, Table 9 presents the coefficient estimates of the regression functions, and Figure 15 presents the Actual vs. Estimated b and c parameters.

$$p = \frac{f_b(T)(1 - 2f_c(a))}{(1 - f_c(a))^2} \qquad q = \frac{f_b(T)}{(1 - f_c(a))^2} \tag{50}$$

One should be careful in using the suggested regression equations. Not every product line or group may lend itself for high confidence and simple regression models.



(a) Regression fit for b



(b) Regression fit for c

Figure 15: Data vs. Regression Fit to the parameters of 38 Microchip product families

In those cases, information on parameter a and T can be used to calculate t^* (i.e., $t^* = aT$), and consequently Time Controlled Bass Model (equation (46)) can be employed with initial parameter estimates (i.e., p_0 and q_0) coming from Naive PQ or Selective PQ method.

4.4.6 $a - t^*$ and $T - t^*$ Models

Since parameters a, t^* and T are related directly (i.e., $t^* = aT$), the $a - t^*$ and $T - t^*$ models can be derived from the $a - T$ Model, such that for the $a - t^*$ model, T is replaced with t^*/a, for $T - t^*$ Model a is replaced with t^*/T in the original model equation (50), to arrive at corresponding model equations (51) and (52).

$$p = \frac{f_b(t^*/a)(1 - 2f_c(a))}{(1 - f_c(a))^2} \qquad q = \frac{f_b(t^*/a)}{(1 - f_c(a))^2} \tag{51}$$

$$p = \frac{f_b(T)(1 - 2f_c(t^*/T))}{(1 - f_c(t^*/T))^2} \qquad q = \frac{f_b(T)}{(1 - f_c(t^*/T))^2} \tag{52}$$

Similar to the relationship between b, c and t^*, we can plot upper and lower bound expectations for parameters a, t^* and T to help managers in judgment process. Special advantage of $T - t^*$ Model is that by just estimating time related information (the peak time and product life time), one can estimate a diffusion curve. In many organizations, product roadmaps and long-range plans provide important guidelines in estimating these values by considering next product launch times or planned end-of-life times.

In summary, in Section 4.3 we developed managerially intuitive methods such as Normalization and Blueprint Approaches to analyze historical product diffusion patterns. Benefiting from percentage-based parameters and important timing information, in Section 4.4, we developed several algebraic procedures to estimate new product diffusions parameters. We investigated interrelationships of the proposed parameters to provide alternative formulations of these models so that managers can have different options based on the their confidence level of the required inputs. Section 4.5 provides forecast performance results of these models in the real data setting. Model sensitivity to the various inputs are also analyzed.

4.5 Numerical Results

All of the nine proposed models require market potential m to be estimated separately, which is a common practice in the literature. Market potential estimates can come from marketing reports, survey of purchasing intensions and managerial judgement. Our models estimate the diffusion shape parameters p and q from several alternative formulations of other parameters. Blueprint Method and the Time Controlled Bass (TCB) Model are adjustment procedures, which adjust the initial diffusion shape estimate with one time-based parameter (either t^* or T) in addition to m. All the remaining models (total of seven) are estimation procedures that need two parameters to estimate the shape of the diffusion curve. Since first period sales $n(1)$ can be expressed in percentage (i.e., $n(1)/m$), these seven models require either percentage-based or time-based parameters or both. Five models out of seven require one percentage-based and one time-based parameters $(a-t^*, b-t^*, c-t^*, a-T, n(1)-t^*)$, while the $b-c$ model needs two percentage-based parameters and the $T-t^*$ model needs two time-based parameters.

We provide empirical results by testing our models with six multigenerational real world diffusion data sets. Table 10 summarizes the data set properties. Section 4.5.1 describes parameter estimation procedures. Section 4.5.2 presents the test procedures

	DRAM	**IBM**	**M1**	**M2**	**M3**	**M4**
Data frequency	yearly	yearly	monthly	monthly	monthly	monthly
Number of Generations	8	4	10	8	13	7
Start Date (yr-mo)	1974	1955	2001-01	2001-06	2000-03	2001-05
End Date (yr-mo)	1998	1978	2008-04	2008-04	2008-06	2008-06
Market	Global	USA	Global	Global	Global	Global

Table 10: Multigenerational data sets used for testing.

and empirical results. In Section 4.5.3 we provide managerial insights and discuss implementation strategies.

4.5.1 Estimating Model Parameters

For all the 50 product generations from six multigenerational data sets, we used the standard Nonlinear Least Square (NLS) procedure proposed by Srinivasan and Mason (1986) ((78)) applied to the data series[5] to estimate the three parameters of the Bass Model. In order to provide good starting points for parameters, we followed initialization suggested by Van den Bulte and Lilien (1997) ((81)), and used a grid search to find initial values for p, q and m[6]. Parameters p and q are exactly used as found in the grid search, and m is inflated by 20% to be consistent with Van den Bulte and Lilien (1997) ((81)), who used population size M (i.e., $m \leq M$) as the initial value for m. All of the 50 products but one[7] converged in the NLS procedure. Using equation (30) with $\beta = 0.01$, we find product life time T and then using equation (41) we find the resultant a, b and c parameters for all products. Peak time t^* and first period sales $n(1)$ are found using the equations (45) and (44), respectively.

[5] Monthly frequency data sets are not seasonally adjusted because our tests show negligible practical significance of the seasonal adjustment procedure on diffusion curve parameter estimates.

[6] We evaluated all $p-q$ combinations in the interval of (0, 0.9] with increments of 0.0025 at every incremental m value changing in the interval of $[80\% - 120\%]N(T)$, where $N(T)$ is the cumulative number of adaptors in the last data observation.

[7] Only 8^{th} DRAM generation with only 5 data points did not converge. Since this generation is the last generation in the product line, it's parameters are not used for any pre-launch testing. Grid search results are reported for this product.

4.5.2 Impact on Forecast Accuracy

After constructing the database of all the parameters, we proceed with our pre-launch forecasting tests to assess their forecast accuracy. We conduct these tests within each product line and report the results as the averages across product lines. We first rank and order the products within each product line based on their launch times. We simulate each product launch scenario starting from the second product launch (making the first product within each product line as the seed product, therefore simulating 44 product launch scenarios in total) by assuming that all the previously launched product parameters are known. For the product launch scenario of product $i + 1$, all the historical products $1...i$ in the same product line are used. Market potential m is assumed to be known. Effect of biased estimates of m is explored in Appendix B.3. If a model requires one of the percentage-based parameters (a, b or c), it is calculated from the historical averages. The time-based parameters (t^* or T) for these models are first assumed to be known, then systematic bias is introduced with increments of +/-10% to test the model sensitivity to time-based parameter estimates.

We propose a simple algorithm to estimate peak time t^* from product roadmap[8] information and algorithm estimated peak time values are also tested. According to this algorithm, the product being launched will have its peak realized at the time of the next major product introduction. Since product roadmaps plan for product launches far ahead of time, this algorithm is very practical and the assumption is realistic. For Microchip data sets, we were able to estimate 32 out of 38 families peak time from the product roadmap information. The remaining 6 products are assumed to have 20% positive bias in estimating t^*, which produced an overall t^* estimation

[8]Product roadmaps are long term plans for product development and introduction times. It is common in the semiconductor industry to have roadmaps that cover the next 5 year of product development and introduction activity. For this algorithm, we assumed that future major product introduction times are known at the time of current product introduction.

error of 16.1% (MAPE).

Forecast performance comparisons for the Microchip datasets are based on: (1) ratio of model generated forecast's mean absolute deviation (MAD) to the optimal (fitted) diffusion curve's MAD evaluated over four forecast horizons: 3, 6, 9 and 12 months; (2) mean absolute percentage error (MAPE) for the cumulative demand forecasted by the model for the first 12-month horizon. For yearly data (DRAM and IBM), we used first 6-year horizon as the forecast period, which constitutes the majority of the early life cycle sales for most of the products.

The MAD of a particular method is calculated over three period horizon as:

$$MAD_3 = \frac{|e_1/X_1| + |e_2/X_2| + |e_2/X_2|}{3} \tag{53}$$

where $e_t = F_t - X_t$ is the error, F_t is the forecast and X_t is the actual data for period t.

The MAPE of the cumulative demand for the first 12 months period horizon is calculated as:

$$MAPE_{12} = \frac{|(F_1 + ... + F_{12}) - (X_1 + ... + X_{12})|}{(X_1 + ... + X_{12})} \tag{54}$$

The MAD ratio measures closeness of pre-launch model forecasts to the optimal diffusion curve tested at each period. This metric is more sensitive to monthly or annually point forecasts, than the cumulative demand MAPE metric, which is designed to compare our model performances to some of the published results in the literature.

4.5.2.1 Results

The $n(1) - t^*$ model is illustrated separately using the Color TV diffusion data and test procedures used in Mahajan and Sharma (1986) ((55)). Main advantage of the hybrid model $n(1) - t^*$ observed in Figure 16 is that it almost achieves the same forecast performance (slightly worse MAD=0.603 vs. 0.602, slightly better MAPE=12.9% vs. 14%) as in Mahajan and Sharma (1986) ((55)) MS1 model, but instead of using peak

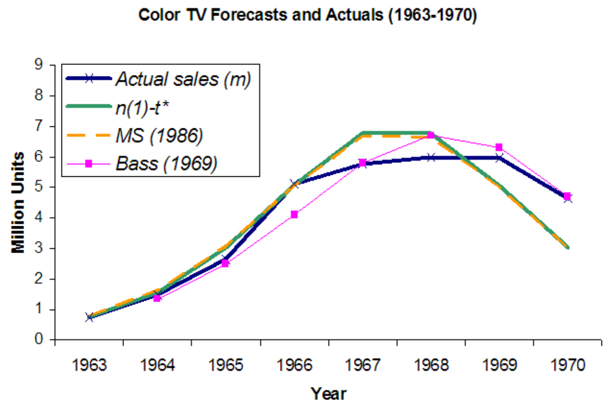


Figure 16: Comparison of the $n(1) - t^*$ model with (8) and (55) in the Colort TV Diffusion Case.

height information we used first year sales information. Same peak time (5 years) and market potential (35 million) is assumed. Based on simple forecasting principles, 1-step ahead forecast $n(1)$ is managerially easier and generally more accurate than 5-step ahead $n(t^* = 5)$ forecast. This is only one example of how the Bass model parameters can be estimated with more intuitive parameters without sacrificing the forecast accuracy. In fact, the superior performance of the $n(1) - t^*$ model becomes apparent on the early life cycle forecasts. When we compare only the first 4-year forecast accuracy of the Bass (1969) ((8)) and Mahajan and Sharma (1986) ((55)) MS1, the $n(1) - t^*$ model clearly outperforms the others with 5.6% MAPE versus 8.7% in the MS1 and 11.4% in the Bass Model.

Table 11 presents the Microchip data performance results of the models that use only one time-based information, which is subject to managerial judgment. Varying levels of accuracy on the time-based parameters are observed to impact forecast accuracy. The other parameter in these models is a percentage-based parameter, which is obtained by averaging historical observations. The $b - c$ and $T - t^*$ models are not tested as they can be directly converted to the $a - t^*$ or $a - T$; and the $b - t^*$ or $c - t^*$ models, for which we already provide test results. As it can be seen from Table 11, none of the proposed models consistently dominates the others. Model $b - t^*$ performs relatively better in both MAD and MAPE metrics, when t^* is negatively biased, but it is outperformed when time-based information is positively biased. One important observation is the comparison of the contribution of t^* and T from the paired models $a-t^*$ and $a-T$, and the Time Controlled Bass (TCB) Models with t^* and T. Models with t^* outperform the models with T in both cases, implying that the peak time is more important than the product life time information in estimating early life cycle demand.

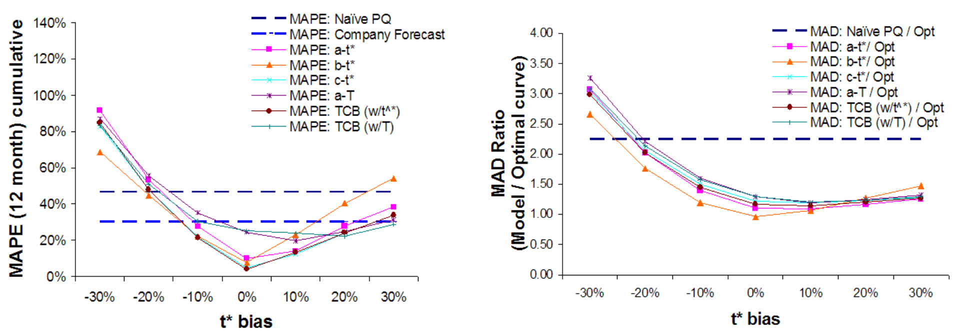
Figure 17(a) shows that naively averaging historical p and q parameters -even within the same product line- can result in very high forecast errors. In this instance, company generated forecast numbers are much better than the Naive PQ method (i.e., cumulative pre-launch forecast error over 12 month horizon is 47% with Naive PQ method vs. 30% with company forecasts). However, by including the timing information with the proposed methodologies, one can significantly improve pre-launch forecast performances. Using the peak time information, the best performances of our models can estimate first year cumulative demand within 4-10% of actual demand. Even if not perfectly known, peak time t^* has some room for error. Within the interval [-10%, 20%] for t^* bias, total of six models are evaluated at 4 different time levels creating 24 combinations of forecasts, our models outperform Company X forecasts in 22 of those 24 scenarios. Another observation from both Figures 17(a) and 17(b) is that a negative bias in time-based parameters creates more forecast error than the same amount positive bias. This error gap widens increasingly as the bias increases.

We also test our models on multigenerational data sets from the literature such as four IBM Mainframe Computer generations and 8 DRAM generations. In both of these data sets, we obtain very similar results and significant forecast accuracy improvement opportunities. Since these data series are at annual frequency, we compare their first 6-year cumulative demand forecast accuracy (MAPE). In Appendix C.1, Figure 26 presents the performances of the t^*-based models that do not require

	Time-based Parameter (t* or T) Estimation Bias							
	-30%	**-20%**	**-10%**	**0%**	**10%**	**20%**	**30%**	$t^*_{roadmap}$ [1]
MAD: Naive PQ / Opt	2.25	2.25	2.25	2.25	2.25	2.25	2.25	-
MAD: $a - t^*$ */ Opt*	3.07	2.02	1.40	1.11	1.09	1.17	1.26	1.41
MAD: $b - t^*$ */ Opt*	2.66	1.76	1.19	0.96	1.07	1.27	1.47	1.23
MAD: $c - t^*$ */ Opt*	3.01	2.07	1.50	1.22	1.19	1.23	1.28	1.47
MAD: $a - T$ */ Opt*	3.26	2.21	1.60	1.29	1.19	1.24	1.32	-
MAD: TCB (w/t) / Opt*	2.98	2.02	1.45	1.17	1.14	1.20	1.27	1.43
MAD: TCB (w/T) / Opt	3.08	2.13	1.58	1.29	1.20	1.24	1.30	-
MAPE: Naive PQ	47%	47%	47%	47%	47%	47%	47%	-
MAPE: $a - t^*$	92%	53%	28%	10%	14%	28%	38%	24%
MAPE: $b - t^*$	69%	45%	22%	8%	23%	40%	54%	30%
MAPE: $c - t^*$	83%	48%	22%	5%	13%	24%	34%	23%
MAPE: $a - T$	87%	56%	35%	25%	20%	25%	31%	-
MAPE: TCB (w/t)*	85%	48%	21%	4%	13%	24%	34%	22%
MAPE: TCB (w/T)	83%	51%	30%	25%	24%	22%	29%	-
MAPE: Company X Forecast[2]	30%	30%	30%	30%	30%	30%	30%	

Table 11: Median values for MAD ratio and MAPE (average of the four Microchip product lines, containing total of 34 pre-launch forecast scenarios).

[1] $t^*_{roadmap}$ is estimated from roadmap data, from next major product introduction time.
[2] Company X's published forecast figures. Most recent forecast figures just before each product launch are evaluated for the first 12-month forecast horizon.



(a) MAPE for 12 month cumulative demand

(b) MAD Ratio (avg. of 3, 6, 9 and 12 month forecasts)

Figure 17: Comparison of Proposed Models and Company Published Pre-Launch Forecasts

regression modeling to calculate parameters (therefore, models with parameter a are excluded). In DRAM data set, where parameters p and q vary significantly across generations (coefficient of variations 125% (p) and 31% (q)), the Naive PQ method produces a median forecast error of 185% MAPE in calculating the first 6-year cumulative demand. Our models can bring this error down to as low as 15%, while leaving considerable room for error in estimating t^*. For IBM data set, where parameters p and q are much more stable (37% (p) and 15% (q) coef. of variation), the Naive PQ method gives a median MAPE value of 25% for the first 6-year cumulative demand. Our models can decrease this error down to as low as 5% MAPE. For this case, the Naive PQ method barely outperforms the average performance of our models only when t^* is underestimated as much as 10% or overestimated as much as 15%. In addition to what we learned from Microchip data, we see from DRAM and IBM data sets that the variance of the original Bass model parameters p and q across generations affects the performance of the Naive PQ method, and more stable parameters produce higher accuracy pre-launch forecasts. Moreover, even in the best case scenario of stable p and q parameters, our models present an opportunity to significantly improve pre-launch forecast accuracy by estimating peak time t^*. In Appendix B.2, we illustrate this statement by forecasting 4^{th} Generation IBM Mainframe Computer diffusion, pre-launch. While the Naive PQ method produces a pretty good pre-launch forecast performance of 25% MAPE for the first 6-year cumulative demand, the $b-t^*$ model gives 4%, the $c-t^*$ model gives 5.1% and the $TCB(t^*)$ model gives 5% MAPE for the same horizon. Peak time for these models are assumed to come from historical averages of year-to-peak values.

4.5.3 Managerial Insights

"[Diffusion Models] are not meant to replace management judgment; rather they should be used to aid that judgment, to help run sensitivity analyses, and to compare the attractiveness of alternative market scenarios" (53). Regardless of the method being used to estimate new product diffusions, it is crucial to recognize that it involves high amounts of managerial judgment and this is best achieved by working with managers and making these estimation procedures as easy and intuitive as possible for them to understand. Any effort to easily combine managerial judgments with analytical models will bring a competitive advantage to companies. In this context, we learned from our modeling effort and numerical results some important lessons that can help managers in their forecasting practice:

- *It is important to analyze historical product diffusions on the same scale:*

  We showed that the analogy of a diffusion curve and normalization procedures improve the ability of managers to see product diffusion pattern similarities. These analyses can be used to standardize the definitions of product life cycle phases, and establish a common platform to study new product introduction scenarios. Based on the diffusion shapes, blueprint curves can be constructed to help managers put their expectations into visual scenarios.

- *Original Diffusion parameters p and q are not managerially intuitive. Diffusion models can be reformulated with more intuitive parameters to provide improved judgment and accuracy:* Table 11 presents the averages of four Microchip product line Naive PQ performances to be 47% MAPE over the first year forecasting horizon, and 125% more error (i.e., 2.25 MAD ratio) from the optimal curve when evaluated over 3, 6, 9 and 12 month horizons. Naive PQ uses scalar p and q parameters, and it is hard for managers to judge and update these numbers based on their expectations. We propose percentage-based and time-based parameters to be used in the pre-launch new product diffusion forecasts. Test results show that our models can significantly improve forecast accuracy, especially when accurate time-based parameter inputs are used.

- *Peak time t^* is more valuable in pre-launch forecasting than the product life T, and product roadmaps can be used to estimate the peak time with high confidence:*

  Comparable models that can either use t^* or T showed that models with t^* can be significantly more accurate than the same model with T. And we show that we can estimate a good baseline for t^* from the roadmap data using a simple algorithm. Application of this method to the microchip data set created t^* estimates within 16.1% of the actuals on the average. Last column in Table 11 presents the results of using the output of this procedure in various methods. In all of the forecast cases, our models that use t^* estimates from this simple algorithm and use the historical averages of the percentage-based parameters still outperform Naive PQ method and the company forecasts. This is a very strong result, because with the use of roadmap data, this forecast method can even be automated and run by just including a market potential m estimate. With the inclusion of managerial judgement to fine-tune these parameters, performance of our models can further increase.

- *Over-estimating time-based parameters produces less forecast error than under-estimating:*

  All of the models show higher sensitivity to the negative bias of time-based parameters and they increase in error much steeper on the negative bias direction. This can be explained by the type of shape change in diffusion curve when a shorter peak time or product life time is used. Earlier peak times require much faster ramp up, taking the product demand much higher levels early in the life cycle. This creates a significant upward bias. Using the same amount of shrinkage, extended vs. shrunk life cycles visualize this effect (see Figure 14). Over versus underestimation of peak time or product life time should also consider the inventory holding cost versus stock-out penalties. Higher t^* estimates may show less absolute forecast error, but higher t^* generally underestimates the demand and it can increase the risk of stock-outs. Therefore, t^* needs to be analyzed carefully considering stock-out penalties versus inventory holding costs and other effects.

4.6 Conclusion

In this paper, we present alternative formulations of the Bass diffusion model that rely on a set of intuitive parameters with natural business interpretations. These percentage-based and time-based parameters are investigated and their interrelationships are explored to provide alternative models. Using data from various multi-generational product lines we show that our version of the Bass diffusion models significantly improves the quality of pre-launch forecasts. Our models emphasize the importance of time-based parameters such as peak time and product life time. Under mild assumptions our models outperformed the forecasts of a major semiconductor company even when time-based parameters were off by -10% to 20%. Using the product roadmap information, we provide a simple procedure to estimate peak times from the next planned major product launch times. Using these peak time estimates and historical averages of percentage-based parameters, all of our models improved the current company forecast accuracy. First 12-months cumulative demand MAPE is improved from 30% down to as low as 22% using this simple procedure. Although unlikely in actual practice, if used with perfect peak-time estimates, our models can improve this forecast metric down to 7% MAPE on the average. These results suggest that by just focusing on estimating peak time or product life time of the product being launched (and relying on the stable historical averages of percentage-based parameters), significant improvements can be achieved in the pre-launch forecast accuracy.

To provide an easy-to-use tool to managers for analyzing the historical product diffusion patterns, we develop Normalization and Blueprint approaches. Normalizing the diffusion curves help us standardize the way we look at different product diffusions. Normalized diffusion curves can improve managers' capability in observing the similarity of historical diffusion patterns. Grouping similar diffusion patterns based on expected product life of the upcoming product and creating a representative blueprint curve, managers can have a baseline estimate for the upcoming new product's life cycle, and they can covert this baseline to a pre-launch forecast easily by just estimating product life time and market potential.

The majority of the data sets analyzed by the diffusion literature are in annual frequency. Past applications of diffusion models are generally applied at macro level representing entire industry and product families. Individual level diffusion modeling is required increasingly to address the new product introduction and forecasting problems at a more granular level. Today's managers need to forecast demand in monthly, sometimes weekly levels. However, increased data frequency comes with an important challenge of dealing with seasonality. Appropriate modeling should address understanding and mitigating the impact of seasonality in diffusion based forecasts. Next chapter focuses on modeling and reducing the affect of seasonality on new product diffusion models, and combining them with pre-launch forecasting models we developed here to further improve early life cycle forecast accuracy.

CHAPTER V

SEASONALITY MODELS FOR NEW PRODUCT DIFFUSIONS: SHRINKING SEASONAL SPLIT AND PRODUCT MIX APPROACHES

5.1 Introduction

"Virtually every product in every industry in every country is seasonal" (Radas and Shugan 1998 ((72))). From weather to holidays, from tax returns to major sport events, there are many reasons behind seasonality for different products. While certain industries are subject to more intense seasonal demand patterns than others, being aware of seasonal variations in demand presents significant forecast accuracy improvement opportunities. One such opportunity is in forecasting the demand of new products that have short life cycles, where the availability of demand information before the product launch is often very limited. Pre-launch estimation of demand of a new product gives valuable insights to the managers in capacity planning, marketing and pricing policies.

In this chapter, we focus on using diffusion models for forecasting the demand of new products with short life cycles and seasonal demand patterns. Shortening product life cycles increase the frequency of new product introductions, therefore managers face the pre-launch forecasting scenarios more often than ever. Already difficult task of estimating new product diffusions becomes more complicated with the addition of strong seasonal variations. The importance of seasonality and its impact on forecast accuracy is recognized and potential models are compared in several time series studies. However, to the best of our knowledge, no such comparative study exists in the diffusion literature, where the majority of the studies focus on macro level diffusion models that use annual sales data. Putsis (1996) ((69)) notes that the traditional statistics and advertising-sales response literature suggests using data series with higher frequency, since it improves parameter estimates of the models. Testing extensive data from diffusion literature, Putsis (1996) ((69)) finds that using deseasonalized monthly or quarterly data results in better fit and higher forecast accuracy than the same model using annual data. For products with short life cycles, diffusion models cannot produce stable parameter estimates using few aggregated annual data points. Heeler and Hustad (1980) ((35)) recommend using at least 10 years of input data, including the data on peak sales, which is much longer than some of today's product life cycles. Therefore, higher frequency data is required to estimate good diffusion parameters.

When higher frequency data is used, certain demand patterns that were disguised in the annual data become more visible. Especially the seasonal demand variations within the year can be significant, and it needs to be properly modeled to estimate the true underlying trend. On the other hand, modeling seasonality can improve the quality of parameter estimates, therefore lead to higher forecast accuracy. Better understanding of the seasonal variations in the diffusion context, therefore, can improve both the monthly/quarterly forecast accuracy and the parameter estimates of the underlying diffusion model trend. Figure 18 shows a real life microchip product, where actual sales show significant deviations from the fitted diffusion curve due to seasonal variations.

In this chapter, we address the following research questions:

- Can classical seasonality methods (i.e., ratio-to-moving-averages) estimate seasonality factors accurately under nonlinear trends, which is generally observed in short life cycle products?

- How to improve the seasonality factor estimates under nonlinear trends and use

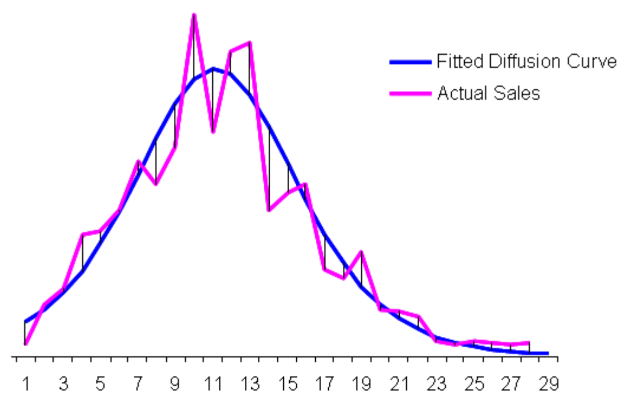


Figure 18: Actual sales of a microchip product versus the fitted diffusion curve.

them to improve monthly forecast accuracy?

- How to model seasonality for short life cycle products when we do not have enough data to run classical seasonality procedures?

This chapter is organized as follows. In Section 5.2, we give an overview of the literature on estimating seasonality factors and estimating diffusion parameters where seasonality can play significant role. In Section 5.3, we extend the findings of Miller and Williams (2003) ((63)) on seasonality models under linear type trends to diffusion type trends. In Section 5.4, we propose two novel approaches, namely, Shrinking Seasonal Split (SSS) and Product Mix Forecasting (PMF), for improving forecast accuracy through seasonality models for short life cycle products. SSS relates seasonality factors not only to the level of the trend as in the classical seasonality models, but also to the slope. Exploiting this relationship, we can better estimate seasonality factors and, therefore, improve forecast accuracy. PMF explores the relationship of seasonality factors for products that belong to the same product line (for example, multigenerational products such as microchips). PMF method avoids the calculation of seasonality factors by estimating the shares (or ratios) of individual product demands within the product line. In the multigenerational diffusion context, we show that PMF can mitigate seasonal variations and improve forecast accuracy. We illustrate forecast improvement potential of both methods with real-world data from a major semiconductor manufacturer. We discuss the conclusions and future research in Section 5.5.

5.2 Literature Review

Two streams of literature are related to our research: (i) the time series forecasting literature that focuses on economic time series data with seasonality, (ii) the diffusion literature that uses higher frequency data such as monthly or quarterly.

Seasonality has been studied extensively in the time series forecasting literature. De Gooijer and Hyndman (2006) ((24)) review methods to estimate seasonality factors including standard decomposition procedures such as X-11, X-12-Arima and their variants (Findley et al. 1998 ((27)), Quenneville et al. 2003 ((71))), shrinkage estimators (Miller and Williams 2003, 2004 ((63)), ((64))) and group seasonality factors (Bunn and Vassilopoulos 1993, 1999 ((15)), ((16))). Many of these studies use simulation experiments, or data from M-Competition (Makridakis et al. 1982 ((56))) and/or M3-Competition (Makridakis and Hibon 2000 ((57))). Some also report results on real life data series, however, as noted by De Gooijer and Hyndman (2006) ((24)), "the best performing model varies across the studies, depending on which models were tried and the nature of the data. There appears to be no consensus yet as to the conditions under which each model is preferred". Widely used databases of monthly data series, namely, M-Competition and M3-Competition, do not include diffusion type trend and we found no study that considers seasonality in conjunction with diffusion models. Our aim is to fill this gap by highlighting important studies of seasonality and diffusion models and combining them where appropriate to develop insights on seasonality for new product diffusion forecasting.

In the limited diffusion literature that uses higher frequency data (i.e., monthly, quarterly), Putsis (1996) ((69)) highlights the theoretical and empirical relationships of the parameter estimates for diffusion models that use data with different frequencies. This paper shows that the use of seasonally-adjusted monthly or quarterly data outperforms the same model that uses annual data. The authors suggest the use of classical X-11 procedure of the U.S. Bureau of the Census (Shishkin 1958 ((76))) for deseasonalizing the data and show that higher frequency data produce better parameter estimates, which in turn produce more accurate "annual" forecasts. However, this study does not provide insights on how better seasonal factors can be found to improve predictions of monthly or quarterly sales. In order to address this issue, our first analysis is to extend the study of Miller and Williams (2003) ((63)), who present a simulation experiment design to estimate seasonality factors with different methods by randomly generating monthly data series representative of the data series in the M-competition database. However, only linear trend cases (1% per month or no trend at all) is included in this simulation design. Authors report that Empirical Bayes methods (by James and Stein (1961) ((39)) (JS)) and Lemon and Krutchkoff (1969) ((51)) (LK)), also known as shrinkage estimators, are superior to the Classical-Decomposition (CD) method for a variety of seasonality patterns simulated. We first replicate their results and then extend their analysis to include nonlinear (diffusion-type) trend. We show that the suggested shrinkage estimators are no longer robust under diffusion-type trend, and on average they can perform significantly worse than the classical decomposition method.

Few studies such as Kurawarwala and Matsuo (1996) ((48)) and Radas and Shugan (1998) ((72)) combine the seasonality terms with the underlying diffusion model (or any model for the latter case). Both of these studies employ the concept of "rescaled or transformed time", where time passes quickly during the high season letting the model accumulate more sales, and then it slows during the low season relatively decreasing the sales volume. Both papers suggest using historical information on seasonality patterns to adjust the current sales model. Kurawarwala and Matsuo (1996) ((48)) analyze four personal computer products and the seasonal model they propose provides a good fit to the data ($R^2 = 0.867$ or higher). Radas and Shugan (1998) ((72)) investigate the seasonality of the movie industry, and using only few data points (first five weeks of data) they show that the transformed-time diffusion model fits better to the data and provides higher forecast accuracy than the observed-time diffusion model for the two films' box office performances tested.

Comparative studies in the diffusion literature generally use annual data, therefore, seasonality is not part of the performance comparisons ((59), (60)). One major conclusion of these studies is that the analytically simpler models perform as well as more complex ones (Armstrong 2006 ((3))). Given similar observations from the M-Competition literature ((56), (57)) and the diffusion literature ((81)), we use the simple Bass Diffusion Model (Bass 1969 ((8))) as the underlying nonlinear trend to represent the demand for certain types of short life cycle products such as high tech microchips. However, the models we propose in this paper are generally applicable to other types of linear and nonlinear trends. The Bass model formula for the cumulative sales at time t is given in equation (55).

$$N(t) = m\frac{[1 - e^{-(p+q)t}]}{[1 + (q/p)e^{-(p+q)t}]} \tag{55}$$

where m is the total market potential, p is the parameter of innovation and q is the parameter of imitation. The derivative of the cumulative sales is $n(t) = dN(t)/dt$, represents the sales at time t. For many real life scenarios such that $p < q$, $n(t)$ is unimodal, i.e., it increases until its peak at $t^* = ln(q/p)/(p + q)$, then decreases. We use $n(t)$ as the basis for the nonlinear trends and introduce several versions of these trends by controlling the signal using the ratio q/p and the speed using $p + q$ (Van den Bulte and Lilien 1997 ((81))).

In Section 5.3, we present comparative test results of several key seasonality models and analyze their performances in estimating seasonality factors under diffusion type trends.

5.3 Methods for Estimating Seasonality Factors under Diffusion Trend

In this Section, we extend the simulation study of Miller and Williams (2003) ((63)) to the diffusion-trend case. Miller and Williams (2003) ((63)) consider two different trend types: no trend or 1% per month trend. Table 12 shows the simulation design setting of (63) that is used to test four seasonality models, namely, Classical Decomposition (CD), Armstrong (2) (A), James-Stein (39) (JS) and Lemon-Krutchkoff (51) (LK). The authors observe that the performances of the seasonality models are almost indistinguishable. We first replicate their analysis to validate their results, then include diffusion-type trends with varying levels of signal (q/p) and speed ($p+q$) as suggested by (81) to the simulation and provide the same performance metrics for different diffusion-trends. We show that performances of shrinkage estimators suggested by the authors are not nearly as superior to classical decomposition (CD) method under diffusion-type trends as they were under no trend or 1% trend.

In simulation design of (63), there are two levels of data length, four levels of variation for the random component, four levels of variation and four levels of skewness for the seasonality factors, and two types of trend, i.e., a full factorial design of $2 \times 4 \times 4 \times 4 \times 2 = 256$ combinations. Since zero variance in seasonality factors can only correspond to zero skewness, this produces 13 different seasonality types (down from 4×4=16), reducing the total number of combinations to 208. We replicate half of these combinations, only focusing on 3 year long data series to represent shorter life cycles better. In addition, we include 6 diffusion-type trends (3 levels of signal (q/p) times 2 levels of speed (normal vs. 20% reduced)), and test them under the same combinations of seasonality factors and random error levels. Every combination is tested with 500 independent replications as suggested by (63).

Table 12: Simulation design factor levels used in Miller and Williams (2003).

No years	SD (E)	SD (S)	Skewness (S)	Trend
3	0.025	0	0	0
6	0.05	0.05	0.6154	1% per month
	0.125	0.15	1.4035	Diffusion ($q/p = 14$)[1]
	0.25	0.35	2.8868	Diffusion ($q/p = 27$)[1]
				Diffusion ($q/p = 50$)[1]

[1]Diffusion-type trend factors added to the simulation design.

Table 13 shows the relative performance of each model in estimating the seasonality factors using the ratios of mean squared error (MSE). For example, rMSE: JS/CD performance represents the ratio of the MSE for the James Stein (JS) method to the MSE of the Classical Decomposition (CD) method. The first two lines of each comparison show the performances reported in (63), and the results we obtained using the same simulation setting. We found that the difference between the results published in (63) and the replicated results are not statistically significant at significance levels of 1%, 5% and 10%. The next three lines of each model comparison present the results for different diffusion-trend cases with different signal (q/p) values[1].

There are three main observations from Table 13: (1) The Armstrong (A) method performs significantly better under certain diffusion trends than under the linear trend, but its average performance is still very poor compared to the (CD) method; (2) Shrinkage estimators (JS and LK) perform worse than the CD method under certain diffusion trends and have very poor worst case performances under the diffusion trends than under the linear trends; and (3) The relative performances of A, JS and LK deteriorate with an increasing q/p value. Table 13 also shows the statistical significance of each difference in performance.

Although not shown, we also tested the 20% slower diffusion trends (achieved by

[1]$q/p = 27$ is the base signal that corresponds to the median values of $p = 0.009$ and $q = 0.24$ for a sample of 38 real-life microchip product families we obtained from a major semiconductor manufacturer. $q/p = 14$ and $q/p = 50$ corresponds to the first and third quartile values for q/p distribution of these 38 products. We change p, while holding q constant to obtain different signal values.

Table 13: Comparison of MSE ratios of seasonality models in estimating seasonality factors.

	Trend	**Mean**	**min**	**Q1**[3]	**Median**	**Q3**[3]	**Max**
rMSE: A/CD	Linear - M&W[1] (2003)	17.523	0.202	0.707	2.404	18.526	150.881
	Linear	20.953	0.210	0.861	2.609	13.196	168.789
	Diffusion $q/p = 14$	1.774***	0.118	0.267	0.933	1.922	10.503
	Diffusion $q/p = 27$	5.865***	0.202	0.469	1.429	5.390	59.425
	Diffusion $q/p = 50$	13.109	0.206	0.775	2.183	10.578	103.068
rMSE: JS/CD	Linear - M&W (2003)	0.826	0.219	0.775	0.951	0.992	1.001
	Linear	0.819	0.213	0.772	0.952	0.995	1.014
	Diffusion $q/p = 14$	0.952	0.086	0.275	0.799	1.194	3.179
	Diffusion $q/p = 27$	2.868***	0.109	0.723	1.331	2.235	30.692
	Diffusion $q/p = 50$	5.109***	0.124	0.921	1.624	5.173	35.943
rMSE: LK/CD	Linear - M&W (2003)	0.677	0.407	0.486	0.660	0.784	1.214
	Linear	0.677	0.394	0.493	0.676	0.802	1.222
	Diffusion $q/p = 14$	0.483***	0.066	0.231	0.363	0.622	1.620
	Diffusion $q/p = 27$	1.173**	0.082	0.273	0.792	1.107	8.980
	Diffusion $q/p = 50$	1.961***	0.162	0.315	0.899	2.228	13.085
rMSE: LK/ JS	Linear - M&W (2003)	0.962	0.422	0.669	0.815	1.190	2.395
	Linear	0.990	0.451	0.689	0.910	1.202	2.267
	Diffusion $q/p = 14$	0.964	0.033	0.511	0.859	1.256	3.365
	Diffusion $q/p = 27$	0.910	0.018	0.413	0.943	1.111	2.901
	Diffusion $q/p = 50$	0.915	0.009	0.503	0.931	1.097	2.816

[1] M&W represents the (63) paper and their results in testing linear trends for the simulation design setting explained in Section5.3. [2] Only seasonal series are shown. [3] Q1 and Q3 denote first and third quartiles. [4] Asterisks indicate P values for two-tailed t-tests: *$P < 0.10$; **$P < 0.05$; ***$P < 0.01$ on the hypothesis that the group mean is equal to the values reported by Miller and Wiliams(2003).

reducing p and q by 20%) and observed that the relative performances of JS and LK methods are more robust (improved worst case performance) in slower speeds ($q/p = 14$ being the exception). For six types of diffusion trends tested, the average performance of JS was better than CD in only one case, while LK was better in three out of six cases. The main insight from this study is that shrinkage estimation methods (JS or LK) are less robust for the diffusion trends than the CD method and their average performance is not nearly as superior as the performance of the CD method in nonlinear trend case. Considering that shrinkage estimators perform significantly better than CD under linear trends, we can conclude that trend plays an important role in seasonal factor estimates.

5.4 Two Novel Approaches for Modeling Seasonality under Diffusion Trend

In Section 5.3, we found that trend plays an important role in seasonality factor estimation. Moreover, for the methods we tested, we assume that there are enough data points required by seasonality procedures to calculate seasonality factors, which may not be the case in today's competitive landscape with short product life cycles. In fact, many products have less than 36 months of life cycle, which is the minimum required data length for classic seasonality models such as classical decomposition (CD).

We identify two major problems to be addressed in modeling seasonality in the diffusion context:

- How to develop seasonality models to address both linear and nonlinear trend types?
- How to find seasonality factors when we do not have enough data points to run classical seasonality models?

In Section 5.4.1, we propose a novel approach, called "Shrinking Seasonal Split" (SSS), to understand the seasonality of monthly data series with different trend types. Observing the interaction of the data series with trend and seasonality, we introduce the seasonal split (SS) factor that is a function of (1) the seasonality factors and (2) the slope of the trend relative to its level, and employ this relationship for linear and nonlinear trend types to find better seasonality factors. We observe that the SS factors show shrinking behavior over the course of the product life time. Differentiating the contributions of seasonality factors and underlying trend to this shrinking behavior, we can obtain better seasonality factors, which translate into higher forecast accuracy. SS factors also allow managers to use their judgment for aggregate level quarterly demand, and easily split the quarterly forecasts into individual months.

In Section 5.4.2, we propose a model, called “Product Mix Forecasting” (PMF), that avoids the calculation of the seasonality factors, which is advantageous in scarce data situations. PMF forecasts the demand for products in “percent mix” instead of in units, by estimating the shares of individual product demands within the total product line demand. Using product mix forecasts, one can divide the aggregate level demand forecasts that tend to be more accurate, into individual product demands. Seasonality can be considered at the aggregate product line level, where longer data series are available. This way the demand forecasts for individual products are obtained with a top-down approach without estimating seasonality factors for each product.

5.4.1 Shrinking Seasonal Split - SSS

In the monthly data series, we observe a specific pattern in the way the quarterly volume is split into monthly volume due to the interaction of the seasonality and the trend. We introduce a novel approach that uses this concept, called “Shrinking Seasonal Split” or “SSS”, that splits the quarterly level demand into monthly level using appropriate seasonal split (SS) factors. We define the seasonal split factors as follows.

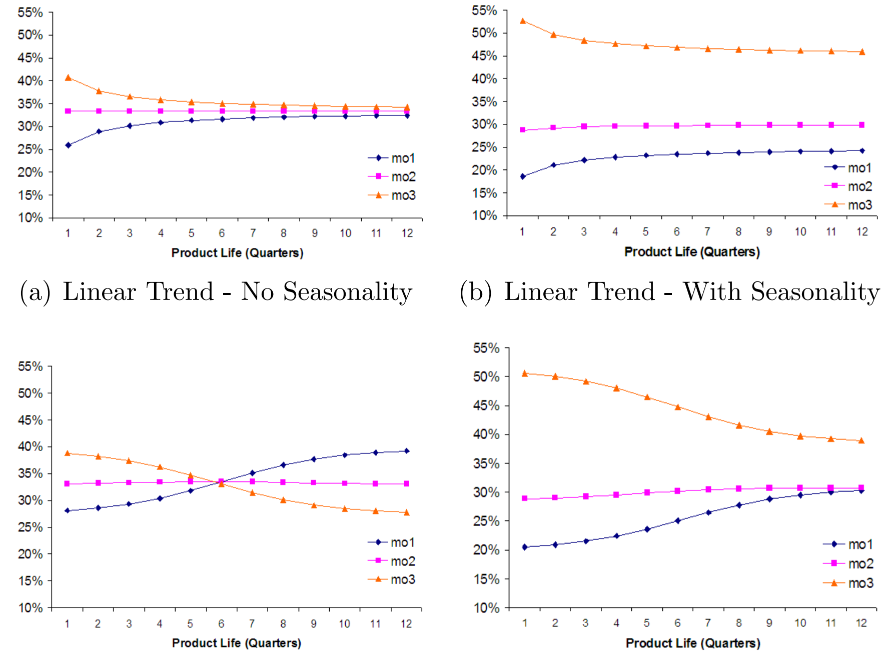
$SS(t+i-1)$ is the seasonal split factor of the i^{th} month ($i \in \{1,2,3\}$) for the quarter starting with month t, such that $SS(t) + SS(t+1) + SS(t+2) = 1$ for any quarter.

A classic multiplicative model decomposes the data series into the trend component T_t and the seasonality component S_t such that $X_t = T_t \times S_t$ (we exclude the random error component for simplicity). Our proposed model replaces the multiplicative seasonal model by $X_t = Q_t \times SS_t$, where Q_t is the quarterly demand and SS_t is the seasonal split component which splits this quarterly demand into individual months. There are two important advantages in using this model. First, the trend component T_t is replaced by a more aggregate level demand component Q_t, which is easier to forecast and judge by managers. Second, the seasonality factors and the underlying trend's relative slope can be used to estimate better seasonality factors, and therefore produce higher forecast accuracy.

5.4.1.1 SSS with Linear Trend

Seasonal split factors depend on both the trend and the seasonality of the data series. For the simplest case when there is no seasonality and no trend, seasonal splits for a quarter starting with month t are equal to each other, such that $SS(t) = SS(t+1) = SS(t+2) = 1/3$. Assuming no seasonality, linear trend and diffusion trend produce the seasonal split factors over time as shown in Figures 19(a) and 19(c). We observe that positive sloping trend curves produce the shrinking behavior of seasonal split factors, such that the percentage split of the first month and the third month sales volume has the biggest gap at the beginning of the life cycle, and this gap shrinks (approaches 33.3%) over time, while the seasonal split factor for the second month of each quarter is already close to 33.3% due to symmetry. This is because the slope of the trend stays constant (for the linear model), while the level (or magnitude) increases over time, shrinking the percentage gap between monthly shares. The opposite behavior is observed with downward sloping trends, where the gap between the SS factors widens over time with the third month SS decreasing and the first month SS increasing (i.e., after the peak of a diffusion curve).

When we add within-the-quarter seasonality, the gap between first and third month SS factors widens. The reason is that the third month seasonality factors are usually greater than 1, while the first month seasonality factors are smaller than 1. Figures 19(b) and 19(d) show the SS factors, when seasonality factors are applied to the underlying trend. We notice that while the trend (slope and intercept) creates the shrinking behavior, seasonality introduces an additional static gap (i.e., not



(a) Linear Trend - No Seasonality (b) Linear Trend - With Seasonality

(c) Diffusion Trend - No Seasonality (d) Diffusion Trend - With Seasonality

Figure 19: Shrinking Seasonal Split - SS factors under linear and diffusion trend.
[1] $y = 50 + 20t$ is selected arbitrarily as the linear trend. [2] A Bass diffusion curve with parameters $p = 0.009$ and $q = 0.176$ is selected arbitrarily to represent the nonlinear trend. [3] Seasonality factors for each month within the quarter is selected as 0.75, 0.9 and 1.35 for the first, second and the third month respectively.

changing over time), since seasonality factors that are greater (less) than 1 increase (decrease) SS factors.

In summary, the SS factor is found to be a function of (1) the slope of the trend relative to the level and (2) seasonality factors. For a monthly linear trend with intercept a and slope b (i.e., $y(t) = a + bt$, where t is the month), equation (56) gives the formula for SS factors for a given quarter (starting with month t) when there is no seasonality.

$$SS(t) = \frac{1}{3} - \frac{g(t)}{3}, \qquad SS(t+1) = \frac{1}{3}, \qquad SS(t+2) = \frac{1}{3} + \frac{g(t)}{3} \tag{56}$$

for $t \in 1, 4, 7...$ where $g(t)$ is the slope of the trend relative to level (measured at its mid-point) for the quarter starting with month t, such that $g(t) = b/y(t+1)$ for a linear trend curve $y(t) = a + bt$.

When we add seasonality factors (s_1, s_2, s_3), the SS factors for the i^{th} month of a quarter starting with month t can be easily updated by equation (57):

$$SS'(t+i-1) = \frac{s_i SS(t+i-1)}{\sum_{i=1}^{3} s_i SS(t+i-1)}, \quad i = 1, 2, 3 \tag{57}$$

Equation (56) explains the shrinking behavior of the SS factors. When we have a trend with positive constant slope b, the level of the trend increases over time decreasing the relative slope $g(t)$. Therefore, the first month SS increases over time, while the third month SS decreases both approaching 33.3%. Equation (57) changes $SS(t+i-1)$ proportional to the seasonality factor s_i such that $SS'(t+i-1) = SS(t+i-1) \times s_i$, however, the resultant sum of the new SS factors within the quarter may not equal to 1. Therefore, we simply scale the new SS factors by the new sum to adjust for this effect. Proposition 1 gives the exact closed-form relationship of the SS factors under linear trend.

Proposition 1 *For a linear trend $y(t) = a + bt$, relative slope for a quarter starting with month t is defined as $g(t) = b/y(t+1)$. Given the within the quarter seasonality factors s_1, s_2 and s_3, such that $(s_1 + s_2 + s_3)/3 = 1$, seasonal split (SS) factors for a given quarter starting with month t are calculated as:*

$$SS(t) = \frac{s_1(1-g(t))}{3+g(t)(s_3-s_1)}, \; SS(t+1) = \frac{s_2}{3+g(t)(s_3-s_1)}, \; SS(t+2) = \frac{s_3(1+g(t))}{3+g(t)(s_3-s_1)} \tag{58}$$

for $t \in 1, 4, 7, \ldots$.

See Appendix C.1 for the proof of Proposition 1. If there exists no seasonality (i.e., $s_1 = s_2 = s_3 = 1$), equation (58) reduces to equation (56), which gives the effect of trend on the SS factors. On the other hand, Proposition 2 identifies the effect on the SS factors that is purely attributable to seasonality.

Proposition 2 *The contribution of trend to the SS factors diminishes over time and the SS factors converge to the values only attributable to the effect of seasonality.*

The pure-seasonality effect is found by taking the limit of equation (58) as t goes to infinity, giving:

$$lim_{t\to\infty} SS(t) = lim_{t\to\infty} \frac{s_1(1-g(t))}{3+g(t)(s_3-s_1)} = \frac{s_1}{3} \tag{59}$$

$$lim_{t\to\infty} SS(t+1) = lim_{t\to\infty} \frac{s_2}{3+g(t)(s_3-s_1)} = \frac{s_2}{3} \tag{60}$$

$$lim_{t\to\infty} SS(t+2) = lim_{t\to\infty} \frac{s_3(1+g(t))}{3+g(t)(s_3-s_1)} = \frac{s_3}{3} \tag{61}$$

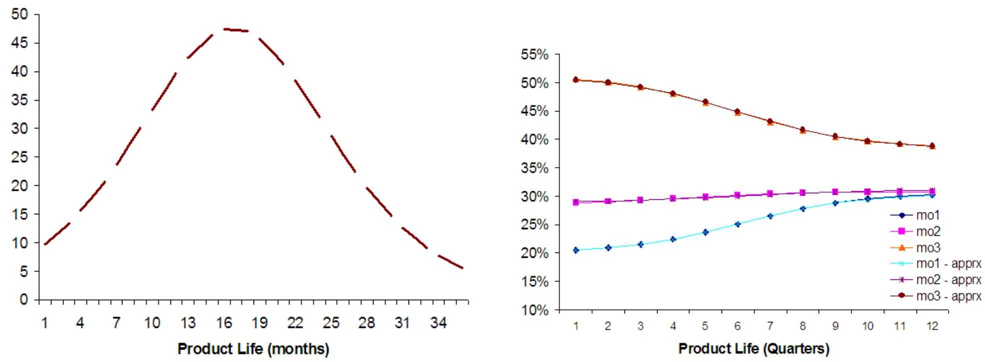
for $t \in 1, 4, 7, \ldots$*.*

From Proposition 2, when there is no seasonality (i.e., $s_1 = s_2 = s_3 = 1$), all SS factors converge to 33.3%, which is in line with equation (56).

5.4.1.2 SSS with Nonlinear Trend

In this section, we extend our analysis on SS factors to the diffusion trend case, and point out why the SSS method can be useful in the diffusion context. Under diffusion trend, the slope of the trend curve changes constantly over time with a positive slope until the peak, and a negative slope thereafter. Given the closed form solution of a diffusion curve such as the Bass model, one can easily calculate SS factors over time. Noting that any general trend curve, i.e., $T(t)$, behaves almost linearly in short intervals, we estimate the curve with a piece-wise linear function such as $y_t(x) = a(t) + b(t)(x-t)$ for each quarter, where y_t represents the linear approximation of the trend curve for the interval $[t, t+2]$. It is assumed that the approximated piecewise linear trend curve y_t passes through the first and the third month levels of the original trend curve for a given quarter, such that $T(t) = y_t(t)$ and $T(t+2) = y_t(t+2)$. Therefore, the slope of y_t can be approximated by $b(t) = (T(t+2) - T(t))/2$.

Based on these definitions and linear approximations of trend, we can extend Proposition 1 for any trend curve by replacing $g(t)$ by approximated values, such that



(a) Approximated Piecewise Linear Diffusion Curve

(b) Actual vs. Approximated SS factors

Figure 20: Piecewise Linear Approximation of a diffusion curve and Approximated SS factors.

$g(t) = b(t)/(y_t(t+1))$. Figure 20(a) shows the linear approximation of a Bass diffusion curve, while Figure 20(b) compares the actual SS factors with the approximated factors. The mean absolute deviation (MAD) of approximated SS factors are found to be less than 0.1% across the 36 individual SS factors estimated. This result shows that our assumption of estimating a nonlinear trend curve with a piecewise linear trend function is a reasonable assumption, and that the approximation functions for SS factors produce values that are very close to the actual.

If the underlying trend type is not known, symmetric moving averages (which is also employed by the classical decomposition method) can be used to estimate trend. The estimate of the relative slope of the trend at any given quarter can then be approximated the same way by a piecewise linear function.

5.4.1.3 Estimating Seasonality Factors using SSS

So far we assumed that seasonality factors s_1, s_2, s_3 are constant across quarters and their average is equal to 1. This assumption ignores intra-year seasonality and only focuses on intra-quarter seasonality. Proposition 3 generalizes Proposition 1 to allow seasonality factors to change across quarters within the year by defining 12 distinct seasonality factors that repeat every year. We then use Proposition 3 to estimate seasonality factors from a given data series. We compare these estimates to the shrinkage estimators proposed by (63) using a similar simulation design and find that SSS can significantly improve seasonality factor estimates, especially for the diffusion trend cases when there are high levels of random error. Under some conditions, SSS reduces the CD method's estimation error by as much as 78% and on average by 50% for the highest random variation setting across all trend types (both linear and nonlinear).

Proposition 3 *For a piecewise linear trend function $y_t(x) = a(t) + b(t)(x - t)$, and seasonality factors s_t, $t \in 1, ..., 12$; seasonal split (SS) factors for a given quarter starting with month t are calculated as:*

$$SS(t) = \frac{s_t(1 - g(t))}{(s_t + s_{t+1} + s_{t+2}) + g(t)(s_{t+2} - s_t)} \tag{62}$$

$$SS(t+1) = \frac{s_{t+1}}{(s_t + s_{t+1} + s_{t+2}) + g(t)(s_{t+2} - s_t)} \tag{63}$$

$$SS(t+2) = \frac{s_{t+2}(1 + g(t))}{(s_t + s_{t+1} + s_{t+2}) + g(t)(s_{t+2} - s_t)} \tag{64}$$

for $t \in 1, 4, ..., T - 2$ where $g(t)$ is the relative slope for the quarter starting with month t such as $g(t) = b(t)/y_t(t + 1)$.

The proof of Proposition 3 follows from the proof of Proposition 1 by keeping $s_1 + s_2 + s_3$ instead of 3, since their sum may not be equal to 3 with the relaxed assumption.

Equations (62), (63) and (64) find the allocation of quarterly demand into monthly for a given set of seasonality factors and the relative slope of the trend curve within the quarter. These are general equations that can be used for any linear trend curve within the quarter, or they can be used as an approximation to nonlinear trend curves using the piecewise linear procedure described in Section 5.4.1.2. One advantage of the SS factors over classical seasonality factors is that for a variety of trend curves SS

factors can be represented via a closed form function, which can be used to select the best possible seasonality factors. We propose a procedure to employ the SSS concept, which improves monthly forecast accuracy through better estimation of seasonality using this property.

A Simulation Study: Performance of the SSS Method:

We conduct a simulation study similar to the one in Section 5.3 with N=50 replications for a known seasonality pattern under 4 different levels of random error. We use 3 types of diffusion signal ($q/p = 14, 27, 50$) and 2 levels of speed (normal vs. 20% slower). SSS approach employs the following steps to find seasonality factors:

- Fit a Bass diffusion curve to the data series
- Estimate the fitted diffusion curve with piecewise linear functions for each quarter and find relative slope ($g(t)$) for each quarter.
- Solve the following optimization model to find the best seasonality factors that fit the data.

Minimize

$$SSE = \sum_{t=1}^{T}(SS(t) - SS_X(t))^2 \qquad t \in 1, 2, ..., T \tag{65}$$

Subject to:

$$SS(t) = \frac{s_t(1 - g(t))}{(s_t + s_{t+1} + s_{t+2}) + g(t)(s_{t+2} - s_t)} \qquad t \in 1, 4, ..., (T-2) \tag{66}$$

$$SS(t+1) = \frac{s_{t+1}}{(s_t + s_{t+1} + s_{t+2}) + g(t)(s_{t+2} - s_t)} \qquad t \in 1, 4, ..., (T-2) \tag{67}$$

$$SS(t+2) = \frac{s_{t+2}(1 + g(t))}{(s_t + s_{t+1} + s_{t+2}) + g(t)(s_{t+2} - s_t)} \qquad t \in 1, 4, ..., (T-2) \tag{68}$$

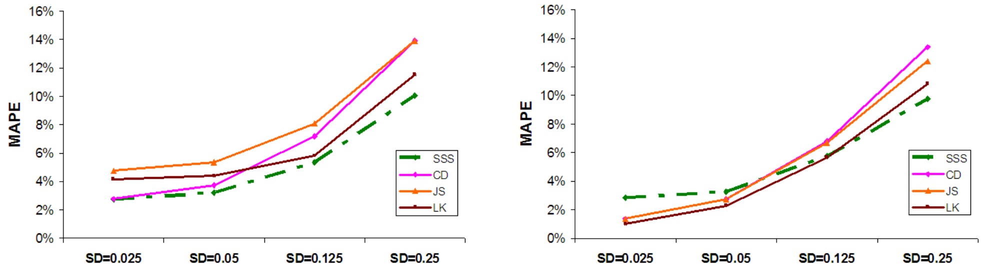
$$s_t = s_{t-12} \qquad t \in 13, ..., T \tag{69}$$

$$\frac{1}{12}\sum_{t=1}^{12} s_t = 1, \qquad s_t \geq 0, \; t \in 1, ..., 12 \tag{70}$$

where s_t are decision variables for seasonality factors, T is the length of the monthly data series consisting of $T/3$ full quarters, $SS(t)$ is the seasonal split factor for month t, and $SS_X(t)$ is the actual split value such that $SS_X(t) = X(t)/(X(t)+X(t+1)+X(t+2))$ for a quarter starting with month t. Equation (69) makes the seasonality factors periodic, while equation (70) provides the nonnegativity constraints and adjusts the average of seasonality factors to 1 for a time series data $X(1)...X(T)$.

We apply the Classical Decomposition (CD), James-Stein (JS) and Lemon-Krutchkoff (LK) methods to the same data series using the procedures described in Miller and Williams (2003).

Table 14 compares the ratios of mean square errors (MSE) of the SSS, JS and LK methods to the CD method for a given set of known seasonality factors demonstrating an end-of-quarter effect. SSS approach results in significantly better seasonal factor estimates, especially when the random error has high variance. For this specific seasonality factor set, LK consistently outperforms JS due to the asymmetry of true seasonality factors, which is favored by LK by design. For the highest two levels of random error across all trend types "including" the linear trend and no trend, the SSS approach outperforms the CD, JS and LK methods in 14 out of 16 cases, reducing the MSE by 45% on average. For the remaining two cases (the Diffusion $q/p = 50$ trends), only the LK method outperforms SSS. Under two out of six diffusion trends (i.e., $q/p = 14$, $q/p = 27$), SSS consistently outperforms CD, LK and JS in all of the four error levels. Over all diffusion type trends and all error level combinations (i.e., total of 24), SSS outperforms in 15 of them, CD outperforms in 6, and LK outperforms in 3 of these combinations. With slower diffusion trends and longer life cycles (i.e., $q/p = 27$-slow and $q/p = 50$-slow) which have flatter slope, and for the linear trend types, results are mixed. SSS outperforms for higher random error cases, while it performs worse for low random errors. These results can be summarized as follows:



(a) Nonlinear Trends (b) Linear Trends

Figure 21: Mean Absolute Percentage Error (MAPE) Performance of Seasonality Estimators under Nonlinear and Linear Trends.

- SSS performs better under high uncertainty regardless of trend.
- SSS performs better under nonlinear trends than linear trends.
- SSS performance improves under diffusion trends with shorter life cycles (low q/p) and faster speeds (high $p+q$), which is generally the hardest case for seasonality models.

in comparison to the CD, JS and LK methods. These insights are summarized from a different perspective in Figure 21, which shows the mean absolute percentage error (MAPE) values calculated against the true values of seasonality factors for each estimator under the same nonlinear and linear trend types. For this particular metric, SSS outperforms all the other methods for all the error levels for the nonlinear trends, while it only performs better than the other methods in the highest error level for the linear trends.

In the simulation tests, SSS approach assumed that there is no model misspecification. In other words, data is generated from the Bass diffusion curves, and again the Bass diffusion curves are used to estimate the trend. This may bring an advantage to the SSS method over the CD method, which uses symmetric moving averages to estimate the trend of a given data series. In order to test the hypothesis that the superior performance of SSS is not due to the correct model specification only, we also

Table 14: Ratios of Mean Square Errors (rMSE) for *SSS*, *JS* and *LK* methods against *CD* method across different diffusion trends, 1 linear trend and no trend cases under 4 different levels of random variance.

Trend	rMSE	**SD=0.025**	**SD=0.05**	**SD=0.125**	**SD=0.25**
Diffusion ($q/p = 14$)	SSS/CD	0.22	0.27	0.36	0.50
	JS/CD	2.25	2.16	1.63	1.26
	LK/CD	0.60	0.60	0.76	0.92
Diffusion ($q/p = 27$)	SSS/CD	0.68	0.64	0.57	0.53
	JS/CD	3.31	2.72	1.66	1.11
	LK/CD	1.10	0.86	0.63	0.92
Diffusion ($q/p = 50$)	SSS/CD	3.02	1.41	0.70	0.47
	JS/CD	6.72	2.73	1.40	1.11
	LK/CD	4.52	1.62	0.62	0.91
Diffusion ($q/p = 14$ slow)	SSS/CD	1.16	0.85	0.59	0.56
	JS/CD	1.77	1.50	1.18	0.98
	LK/CD	1.36	0.92	0.60	0.83
Diffusion ($q/p = 27$ slow)	SSS/CD	3.48	1.46	0.55	0.44
	JS/CD	1.88	1.30	1.11	0.95
	LK/CD	2.65	1.11	0.61	0.83
Diffusion ($q/p = 50$ slow)	SSS/CD	3.01	1.36	0.62	0.49
	JS/CD	1.17	1.19	0.96	0.92
	LK/CD	1.44	0.92	0.54	0.80
No Trend	SSS/CD	4.89	1.57	0.70	0.47
	JS/CD	0.99	0.99	1.02	0.88
	LK/CD	0.61	0.72	0.71	0.80
1% Linear Trend	SSS/CD	5.35	1.61	0.67	0.50
	JS/CD	1.00	1.00	0.97	0.92
	LK/CD	0.57	0.68	0.69	0.84

used the Bass diffusion curves to calculate CD seasonality factors. As expected, the relative performance of SSS compared to CD is somewhat degraded, however, the insights we reported earlier remained the same. For the two highest error cases, SSS still outperforms CD in all of the 16 cases, however the error reduction potential for SSS over CD fell from 45% to 21%. Given many practical situations with shortening product life cycles, where the random error component is a significant contributor to noise, we can claim that SSS has an advantage over classical seasonality methods in providing improved seasonality factors. Better seasonality factors can, in turn, improve the potential for higher forecast accuracy.

5.4.1.4 Impact on Forecast Accuracy

In the previous section, we tested how alternative seasonality models perform in estimating the true seasonality factors, but we have not assessed the impact of better seasonality factors on forecast accuracy. To test the performance of SSS within a forecasting framework, we introduce a simulation design. We generate random data series for a horizon of 54 months both with nonlinear and linear trends. We use the first 36 months of data to calculate CD and SSS seasonality factors, while withholding the remaining 18 months for forecasting purposes. We then separately deseasonalize first 36 months of data using both methods and we fit a Bass diffusion curve to the deseasonalized series for nonlinear trend, and we fit a straight line to the linear trend case. Then we reseasonalize the data using the CD and SSS seasonality factors. We measure forecast accuracy over five different horizons: 1, 3, 6, 12 and 18 months using mean absolute percentage error (MAPE). We use the same random error variance levels as described in Section 5.4.1.3 with the same known seasonality factors. To represent the nonlinear trends, we tested the diffusion curve with the longest life cycle (i.e., $q/p = 50$ slow). For linear curves, we used a 1% trend.

Table 15 presents the relative average performances of SSS and CD methods for

Table 15: Forecast performance of the SSS and CD methods [MAPE: CD/SSS (%)].

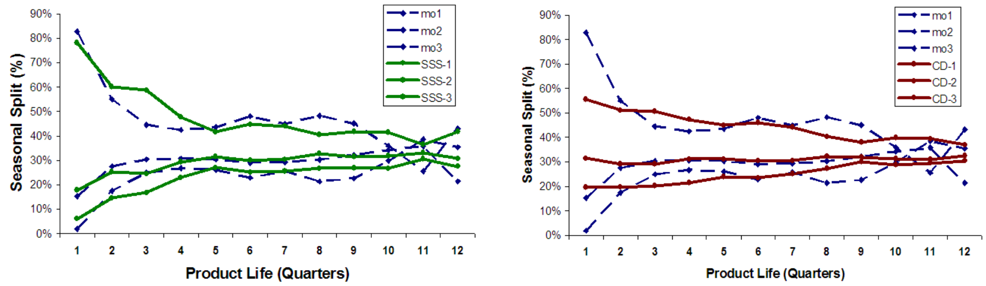
	Horizon	Error Level (SD=St.Dev.) **SD=0.025**	**SD=0.05**	**SD=0.125**	**SD=0.25**
Nonlinear Trend:	1	2.3 / 3.0	6.1 / 6.6	11.9 / 11.1 *	28.1 / 24.6 *
	3	2.6 / 3.0	5.8 / 6.0	12.9 / 12.1 *	30.0 / 25.0 *
	6	2.7 / 4.0	5.9 / 6.4	14.2 / 13.3 *	31.5 / 28.8 *
	12	3.1 / 5.0	6.2 / 6.7	15.3 / 15.1 *	33.2 / 33.2 *
	18	3.2 / 5.0	6.6 / 7.1	16.3 / 16.2 *	36.3 / 35.5 *
Linear Trend	1	2.5 / 6.2	4.5 / 7.0	11.5 / 12.8	29.0 / 26.3 *
	3	2.5 / 6.7	4.8 / 7.2	11.9 / 12.2	28.9 / 24.4 *
	6	2.6 / 4.6	4.7 / 5.9	12.0 / 11.9 *	27.6 / 24.8 *
	12	2.6 / 3.5	4.6 / 5.1	12.2 / 11.8 *	27.6 / 25.6 *
	18	2.7 / 4.1	4.7 / 5.4	12.3 / 11.8 *	28.1 / 25.5 *

* Indicates that SSS outperforms CD.

this simulation for each forecast horizon and each random error level tested. Similar to our findings in Section 5.4.1.3, SSS performs better than CD at higher error levels regardless of trend, and overall performance of SSS is better in nonlinear trends than in linear trends. CD outperforms SSS in the lower error levels.

In summary, we have shown that SSS produces improved seasonality factor estimates, especially when the random error variance is high and data series is short and nonlinear. These improved seasonality factors can in turn improve the forecast accuracy.

Another forecast accuracy improvement opportunity with the SSS approach is the use of seasonal split (SS) factors. In practice, managers prefer forecasting demand at more aggregate levels, such as quarterly versus monthly, since it is easier to achieve higher forecast accuracy. Then the quarterly demand forecasts are split into monthly volume based on a mixture of factors such as historical seasonality, customer backlog and business judgment. To illustrate the potential use of SSS method in a real life setting, we analyze a sample microchip product using data from a major semiconductor manufacturer. We estimate the seasonal split factors for this product by averaging the historical seasonal splits of the group of similar products over the



(a) SSS vs. Actual (b) CD vs. Actual

Figure 22: Estimates of the Seasonal Split values for a sample microchip product.

period of first 12 quarters in their product life cycles. We assume that this average gives a good estimate for the seasonal split of quarterly volume into monthly volume. Figure 22(a) compares the actual seasonal split versus the SSS estimates. Another way of finding seasonal split factors is to use the seasonality factors and the underlying trend curve by Proposition 1. In order to compare the performance of the classical decomposition method, we employed Proposition 1 together with the CD seasonality factors. Estimates of the CD method versus the actual seasonal splits are presented in Figure 22(b). Finally we compared these seasonal split factor estimates to that of company published numbers (derived from monthly and quarterly forecast figures). The results are presented in Table 16. Accuracy of seasonal split estimates for SSS is better than CD, and together they outperform the company seasonal splits. Moreover, the company forecasts are published only one quarter before the actual sales, while SSS and CD estimates are calculated from similar products before the product launch. Considering this performance improvement, we can conclude that SSS can help managers split the quarterly demand forecasts into monthly volume more accurately, while managers continue to rely on their judgments at the quarterly level forecasts.

The next section proposes an alternative approach for modeling seasonality in the context of multigenerational product diffusion, called Product Mix Forecating (PMF).

Table 16: Mean Absolute Deviation (MAD) of the *SSS* and *CD* methods in estimating the seasonal split values.

	Seasonal Split Error		
Horizon	SSS	CD	Company
1-year	5.0%	7.7%	8.6%
2-year	3.7%	4.9%	5.7%
3-year	4.1%	5.3%	5.3%

Note: Company seasonal split errors are found from monthly and quarterly forecast figures published one quarter ahead of the actual sales. *SSS* and *CD* estimates are prior to product launch.

5.4.2 Product Mix Forecasting - PMF

For calculating the monthly seasonality factors, classical seasonality methods require a minimum of 36 data points. Although the first and last 6 data points are used as inputs, no seasonality factor estimates can be generated for these points, due to the 12-month symmetric moving averages used in trend estimation. This is referred as the end-point-problem in the time series literature. Asymmetric moving averages methods such as Musgrave method (Musgrave 1964 (66)) can be used to eliminate the end-point problem, however minimum required points cannot be reduced to less than 24, since at least two full cycles are necessary to estimate preliminary seasonality factors. Assessing the usefulness of the classical seasonality methods in the context of short life cycle products, minimum data requirement presents an important challenge. Many products in today's marketplace may not have long enough life cycles that would allow for the estimation of seasonality factors. Even if we have 36 months of data for a given product, the remaining few months of this product's life cycle would be much less important from the demand management perspective than the early life cycle. In this section, we propose a model, called Product Mix Forecasting (PMF), to address this problem. PMF assumes that short life cycle products are generally members of a larger product line that has multiple generations of successive products, which are introduced into the market one after another. At any given time, the transitions from older to newer generations are actively managed by companies. PMF attempts to estimate the shares of individual products within their product line, while eliminating the common seasonality effects on the demands of these products.

In the multigenerational context, the product mix idea exploits the observation that different products within a product line might be subject to similar seasonality. Multiple products that are in transition are considered together to calculate each product's monthly demand with a top-down approach from aggregate level to product level. Product Demand Mix (or Product Mix) is crucial information for planning of resources. Ability to understand the product transition from old products to new products allows managers to accurately allocate production capacity. Furthermore, right product mix will have an impact on the capital equipment purchases to make sure the new equipment is being purchased on time to enable the manufacturing of new products, while minimizing the excess capacity.

One of the basic rules of forecasting is that the aggregate forecasts tend to be more accurate. Therefore, estimating the aggregate demand for the entire product line, then using the "product mix" forecast to calculate individual product demands offers opportunities for higher forecast accuracy. The variance of the demand for the product line that is composed of multiple generations of products is generally much less than the variance of the demand for each individual products. Moreover, managers have the advantage of analyzing the product lines over a longer horizon, since the life cycle of product lines are much longer than individual products. This makes it easier for managers to forecast the underlying trend and seasonality of the product line. If one has good product mix forecasts, then the aggregate level demand forecasts can be used with a top-down approach to produce product level demand forecasts.

When forecasting demand for short life cycle products using diffusion models at monthly granularity, actual demand can differ significantly from the underlying diffusion trend due to seasonality. We propose the PMF method for calculating product mix from the estimated diffusion curves that mitigates this seasonality effect. Following equation calculates the product mix from estimated diffusion curves:

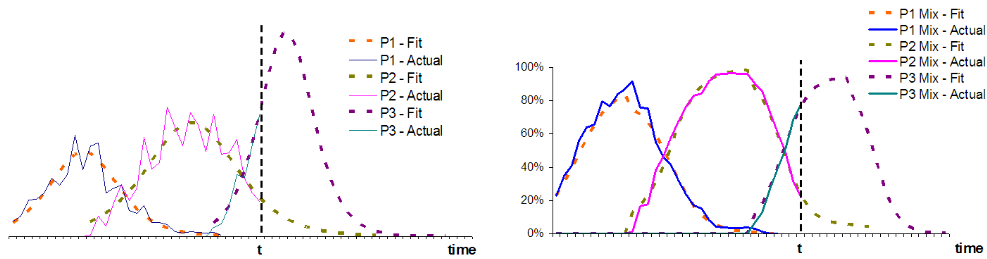
$$mix_i(t) = \frac{n_i(t)}{\sum_{j \in I} n_j(t)} \tag{71}$$

where $n_i(t)$ is the estimated diffusion curve for product i evaluated at time t, I is the set of active products within the product line and $\sum_{j \in I} mix_j(t) = 100\%$.

This transformation gives a simple ratio of a particular product diffusion curve to all the active products' diffusion curves. The basic idea is that two (or more) products are experiencing similar seasonality at the same time (i.e., the end-of-quarter effect); and when we find shares of the actual sales of these products and observe the underlying product transition curve in percentage mix, we mitigate the common seasonality effect on these products.

We assume a multiplicative model such that $X_i(t) = A_i(t)S_i(t)$, where $X_i(t)$ is the actual demand at time t, $S_i(t)$ is the seasonality factor at time t and $A_i(t)$ represents the deseasonalized data series for product i. For products experiencing a diffusion type trend, $A(t)$ can be modeled by a diffusion curve $n(t)$. Therefore, when $S_i(t) \approx S_j(t)$ for any product $i, j \in I$, the actual product mix for product i at time t is given by $X_i(t)/\sum_{j \in I} X_j(t) = A_i(t)/\sum_{j \in I} A_j(t)$, and it can be estimated by $n_i(t)/\sum_{j \in I} n_j(t)$.

With this transformation, estimated diffusion curves can forecast product mix over time, which has an improved model fit. In order to illustrate this claim, we analyze a sample product line using data from a major semiconductor manufacturer. Figure 23 presents three products within this product line. Part (a) shows the actual sales versus fitted diffusion curves. In Part (b), we calculate percent mix graph using PMF method, which significantly reduces the effect of seasonality and improve the model fit with R-square values equal to 0.984, 0.979, 0.975 for products P1, P2 and P3, respectively. Time t represents the current time marked with a vertical dashed line, until which we use all the data to fit diffusion curves. We compare actual product



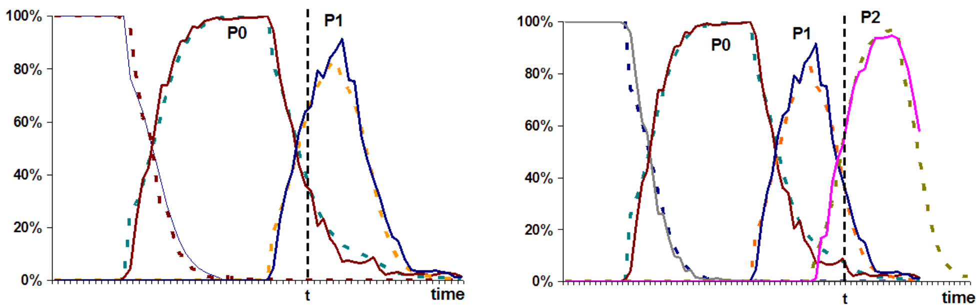
(a) Unit Graph (b) Percent Mix Graph

Figure 23: Illustration of Percent Mix Model fit versus Regular Diffusion fit. Preceding product P0 and succeeding product P4 are not shown.

mix vs. fitted product mix until time t and extrapolate the product mix beyond time t for forecasting purposes.

We illustrate the forecast performance of the proposed model by simulating two real-life product launch scenarios from the same sample product line, namely P1 and P2. For each scenario, we use the actual sales data that captures up to the first six months sales of the product being launched. We then forecast the future product mix of these products until the next product launch time using diffusion models and the PMF method. First, in the launch scenario of P1, we use available data to fit diffusion curves to products P0 and P1. We transform this fit to product mix using equation (71). We forecast the product mix until the product launch time of P2 and compare with the actual product mix using mean absolute deviation (MAD). Figure 24(a) illustrates this scenario. Similarly, we execute the product launch scenario of P2 and forecast the product mix for this product until the launch time of the next product, i.e., P3. This scenario is shown in Figure 24(b). Monthly and quarterly forecast performances of both scenarios are presented in Table 17 and compared with the performance of the actual company published forecasts.

In the first scenario, the PMF method achieves a forecast performance of 3.0% MAD at the monthly level and 2.9% MAD in the quarterly level. For the same horizon, company published forecasts achieved a performance of 5.7% and 4.4% MAD, for monthly and quarterly levels, respectively. For the second product launch scenario,



(a) P1 Launch Scenario

(b) P2 Launch Scenario

Figure 24: Product Mix graphs of two product launch scenarios tested.

Table 17: Product mix forecast performance of the PMF method and the corresponding company published forecasts in mean absolute deviation (MAD).

		Monthly	Quarterly
Launch of P1	PMF	3.0%	2.9%
	Company	5.7%	4.4%
Launch of P2	PMF	2.2%	2.0%
	Company	5.2%	5.1%

[1] A simple seasonal split method (i.e., 30-30-40%) is used to convert quarterly company forecasts into monthly levels.

PMF achieves 2.2% and 2.0% product mix forecast performance against the company forecast performance of 5.2% and 5.1%, when measured in MAD at monthly and quarterly levels.

In both of these analyses, we assume that parameter m (market potential) is known for the product being launched (therefore, we only fit parameters p and q, while fixing m). However, sensitivity analysis shows that our model is quite robust in m, such that relaxing this assumption by $\pm 25\%$ still provides better results than the company forecasts for both products.

When high volumes of sales are considered, calculating the lost revenues and excess inventory costs, one can easily understand that even a 1% product mix forecast deviation can be a significant cost for businesses. In practice, forecasted product mix can significantly deviate from the true demand mix. In these cases, managers generally engage in price moves to make sure that old product inventories are consumed (by lowering price) and new product demand is aligned with the available supply. However, profit margins are often sacrificed to meet the supply determined by the inaccurate forecast figures, causing significant losses. Accurate product mix forecasts, therefore, can be very useful for businesses to provide the right mix at the right time without sacrificing profits to clear the market.

Using the PMF method, in order to obtain the demand forecasts for individual products in units, managers need to multiply the aggregate level product line demand with the product mix forecasts. Therefore, the forecast accuracy depends on the accuracy of both the aggregate level demand and the product mix demand. Calculating the aggregate demand from time series analysis is beyond the scope for our paper. However, it is expected that the aggregate level product line demand forecasts will be more accurate than the individual product level forecasts. Moreover, seasonality is easier to determine at this level of aggregation, since more data is available for longer period of times.

5.5 Conclusion

Seasonality is part of almost any demand management activity, and it needs to be treated carefully to understand the real trends and improve the forecast accuracy. In forecasting time series data with seasonality, the common practice is to first deseasonalize the data to obtain a trend curve estimate, then forecast the trend and reseasonalize it to obtain final forecasts numbers. Therefore, having good seasonality factor estimates has 2-fold benefits by: (1) improving the trend estimate, (2) improving the final forecasts. The majority of the time series forecasting literature investigates the seasonality models under linear trends when there are enough data points to calculate seasonality factors. In this study, we focused on combining seasonality models with diffusion models to provide improved forecast accuracy. In order to understand the performances of current seasonality models under diffusion modeling context, we first extended the simulation analysis of Miller and Williams (2003) ((63)) to the nonlinear trend cases, and found that the performances of seasonality models depend on the underlying trend. We found that suggested shrinkage estimators (James-Stein and Lemon-Krutchkoff) can perform much worse than the classical decomposition method under certain diffusion type trends.

We proposed two novel approaches to model seasonality by exploiting the certain relationships of seasonality factors. The Shrinking Seasonality Split (SSS) approach identifies the relationship between seasonality factors, the slope of the trend relative to its level and the seasonal split factors, which we define as the percentage split of each month within its quarter. Employing this relationship, we showed that we can improve seasonality factor estimates and therefore forecast accuracy, especially for short data series under nonlinear trends with high random error. Under high levels of random error, SSS improves both seasonality estimates and forecast accuracy, regardless of the underlying trend. The second approach we proposed, called Product Mix Forecasting (PMF), addresses the challenge of not having long enough data series for calculating seasonality factors. PMF mitigates the impact of seasonality by using simple ratios of the trend estimates in the multigenerational diffusion context. Using real data from microchip products, we showed that PMF can improve the model fit and the product mix forecast accuracy. Moreover, both the SSS and the PMF models allow managers forecast at an aggregate level, which generally leads to higher forecast accuracy. SSS can be used to split aggregate level (i.e., quarterly) forecasts into monthly using the appropriate seasonal split factors. PMF lets the managers forecast the demand at the aggregate level for a group of products, then product mix forecasts can be used to split these forecasts into individual products. These top-down forecasting approaches are both easier for managers to employ their judgments at higher levels and they provide increased potential for higher forecast accuracy.

Classical decomposition methods only use one type of information in estimating seasonality factors, i.e., the level of the trend. Seasonality factors are simply the ratios of actual data to the estimated trend levels. The Shrinking Seasonal Split method utilizes additional information, which is the slope of the trend, and provides an insight on how seasonality factors contribute to the the percent split of quarterly volume into monthly volume. This relationship can be further investigated, and further benefits can be researched. One such area of future study is to understand the nature of seasonality factors in different parts of the life cycle phases. External variables such as price, marketing effort or sales incentives can be analyzed for their effect on seasonality factors and therefore on seasonal split factors. Providing insights on how certain managerial decisions affect the split of quarterly volume into each month can be very valuable for businesses. Especially in the new product launch scenarios, decisions made prior to launch time are very critical to the success of the products in the marketplace. Studies analyzing product launch scenarios together with seasonality models would certainly contribute further to the insights we obtained with this study.

CHAPTER VI

CONCLUSION

In this thesis, we focused on demand management activities of global firms. Through innovative thinking coupled with data intensive analyses, we identified opportunities for companies to improve their demand management and forecasting activities.

In Chapter 2, we analyzed an extensive pricing data set of the U.S. Less-than-Truckload (LTL) market, and explored opportunities to quantify expert knowledge and improve market rate estimates. Beneficial to both shippers and carriers, proposed regression based methodology can be used as a market visibility tool to improve understanding of the realized market rates, therefore reduce cost and improve service through better negotiations of these market rates.

In Chapter 3, we conducted a simulation study on a major semiconductor manufacturer's global supply chain. Identifying demand dependencies of products related to the ordering mechanisms, we tested a strategic collaboration scenario that provides advanced demand information on future orders. Quantifying the forecast improvement potential of such a scenario, we measured the inventory impact on the supply chain on sample products using a supply chain simulation model. This study provided the insight to the managers that thinking across business units and exploiting demand dependencies can improve supply chain efficiency.

In Chapter 4, we attempted to provide user-friendly approaches for estimating new product diffusions prior to product launch. We first introduced a normalization approach, a tool to visually analyze historical diffusion patterns on the same unit scale. Using the insights from normalization, managers can select representative adoption patterns for their new products, then use them easily to construct pre-launch forecasts. In the second part of this chapter, we introduced several versions of the Bass diffusion model that use more intuitive and easier to estimate parameters. Testing industry and company data, we showed that our models can significantly improve the pre-launch forecast accuracy.

In Chapter 5, we extended the diffusion modeling research to include seasonality considerations. We first analyzed the classical seasonality models under diffusion-type nonlinear trend and found that methods that are shown by other researchers to perform well under linear trends may not perform well under nonlinear diffusion trends. We then proposed two novel approaches to model seasonality designed for diffusion trends. Our models have important advantages over classical methods, such that they estimate seasonality factors more accurately especially under diffusion trends with high levels of random error, require less data points to do so, and improve forecast accuracy.

Intersection of Chapter 4 and 5 open up new interesting research directions. Further analysis of seasonality models within the diffusion modeling literature would be a fruitful area of research. The proposed Shrinking Seasonal Split (SSS) approach should be tested and compared against other standard seasonality methodologies under both real and simulated data to better assess its strengths and weaknesses. Special attention to the models that require less data points to provide at least the same amount of accuracy is especially valuable from the short life cycle products perspective. There is always a need for higher accuracy in pre-launch forecasts. Future research can focus on improving the abilities of managers to engage more with analytical techniques in estimating new product launch scenarios, rather than relying only on their judgments. Provided models in Chapter 4 are the first steps to involve hesitant managers in diffusion modeling, however, more extensive analysis of these methods under specific conditions is necessary.

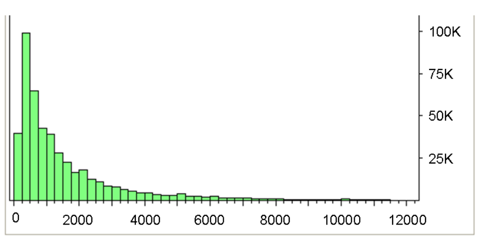
APPENDIX A

ADDENDUM FOR CHAPTER 2

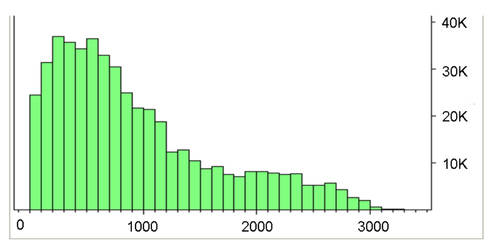
A.1 Descriptive Statistics

Table 18: Descriptive Statistics of the dataset.

Average Shipment Price	$185.80
Average Mileage	933.6 miles
Average Weight	1713.4 lbs
Minimum (Freight Class)	50
Maximum (Freight Class)	150
Number of shippers	43
Number of carriers	128
Number of State-to-State lanes covered	2126 (excluding Washington DC)
Most number of shipments (Origin State)	California
Most number of shipments (Destination State)	Texas
Least number of shipments (Origin State)	Wyoming (excluding Washington DC)
Least number of shipments (Destination State)	Vermont (excluding Washington DC)
Total number of shipments	484,612 (75% of which are scored)



(a) Weight Distribution (lbs)



(b) Mileage Distribution

Figure 25: Weight and Mileage Distribution of LTL Shipments

A.2 Regression Model

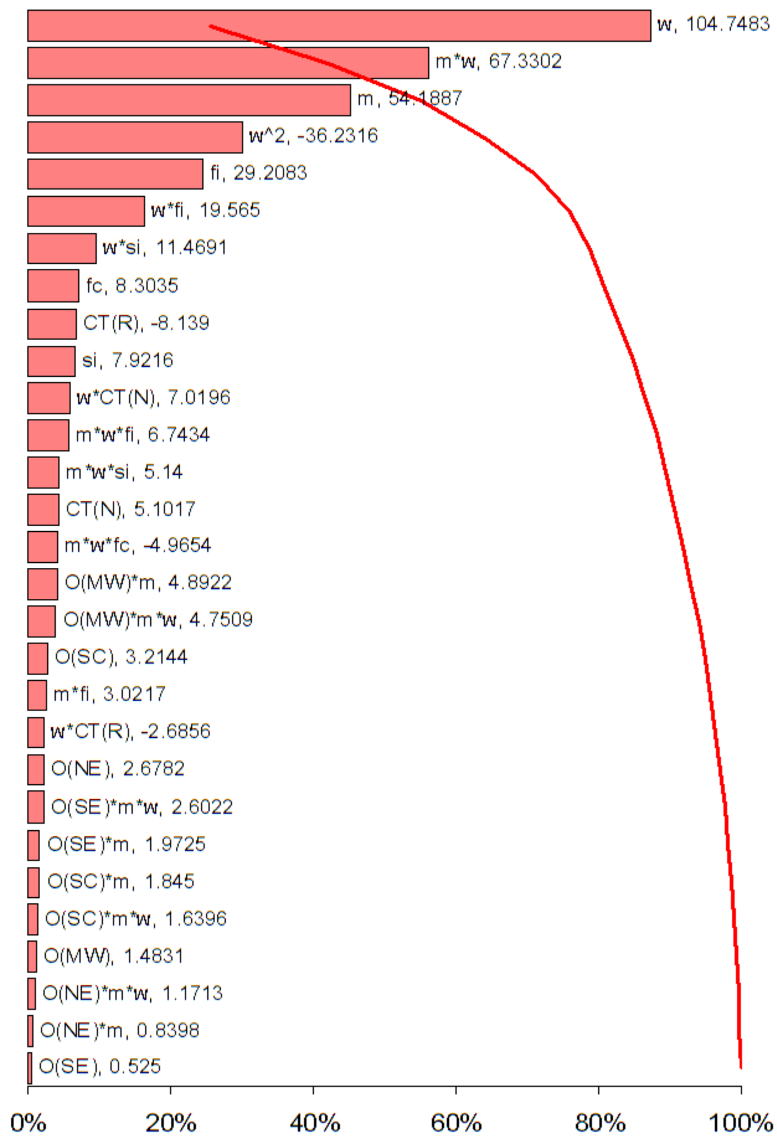
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**	
Model	29	7696500700	265396576	184146.1	
Error	356395	513646549	1441.2283		Prob > F
C. Total	356424	8210147249		0	

Parameter Estimates				
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>** $\|t\|$
Intercept	207.594	0.09903	2096.2	0
w	186.677	0.13764	1356.2	0
m	67.900	0.09428	720.2	0
si	10.014	0.08553	117.1	0
fi	22.272	0.07376	301.9	0
fc	2.391	0.07862	30.4	<.0001
$CT(N)$	10.532	0.10964	96.1	0
$CT(R)$	-4.392	0.14857	-29.6	<.0001
w^2	-7.565	0.04033	-187.6	0
$m * w$	78.630	0.10449	752.5	0
$m * fi$	9.401	0.08363	112.4	0
$w * si$	12.773	0.10134	126.0	0
$w * fi$	22.731	0.07302	311.3	0
$w * CT(N)$	12.429	0.10862	114.4	0
$w * CT(R)$	-7.850	0.12759	-61.5	0
$m * w * si$	6.940	0.10763	64.5	0
$m * w * fi$	9.175	0.07979	115.0	0
$m * w * fc$	-6.366	0.08388	-75.9	0
$O(MW)$	-0.088	0.12513	-0.7	0.4825
$O(NE)$	-0.429	0.18042	-2.4	0.0174
$O(SC)$	5.620	0.15078	37.3	<.0001
$O(SE)$	0.891	0.13586	6.6	<.0001
$O(MW) * m$	4.733	0.13899	34.1	<.0001
$O(NE) * m$	-1.698	0.17072	-10.0	<.0001
$O(SC) * m$	3.604	0.19897	18.1	<.0001
$O(SE) * m$	5.159	0.14213	36.3	<.0001
$O(MW) * m * w$	4.893	0.13210	37.0	<.0001
$O(NE) * m * w$	-2.275	0.20682	-11.0	<.0001
$O(SC) * m * w$	2.604	0.21337	12.2	<.0001
$O(SE) * m * w$	6.286	0.15361	40.9	0

All numerical predictors are standardized using the mean and standard deviation values reported in Table 5. (i.e., "w" is the standardized version of "W", which represents the weight of the shipment).

A.3 Pareto Plot of Transformed Estimates

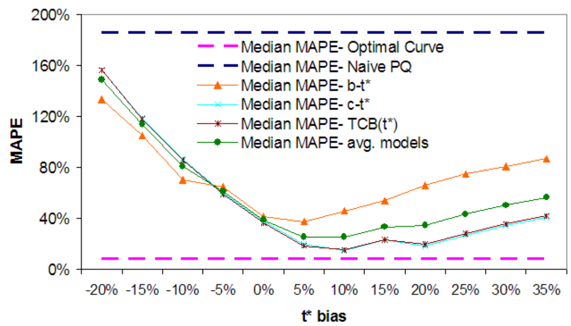
Absolute effect sizes from high to low, and how they add up.



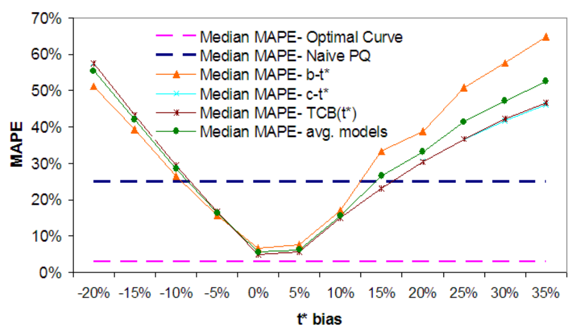
APPENDIX B

ADDENDUM FOR CHAPTER 4

B.1 Performances of Proposed Models for DRAM and IBM data sets



(a) DRAM Generations



(b) IBM Mainframe Computer Generations

Figure 26: Median MAPE for 6-year cumulative demand. 7 pre-launch scenarios tested for DRAM data. 3 pre-launch scenarios tested for IBM data. Median performances are plotted with varying levels of t^* bias.

B.2 Illustration of the IBM Gen-4 Case

For this case, m is assumed to be known. t^* for the proposed models are calculated from the historical averages of peak times. Historical averages of b, c, p and q are used when required.

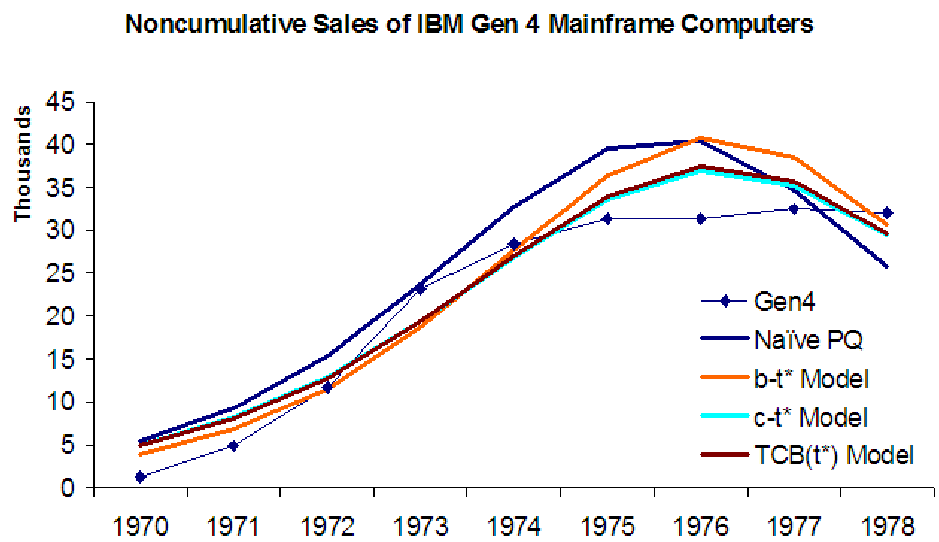


Figure 27: Noncumulative Demand vs. Noncumulative Pre-launch Forecasts

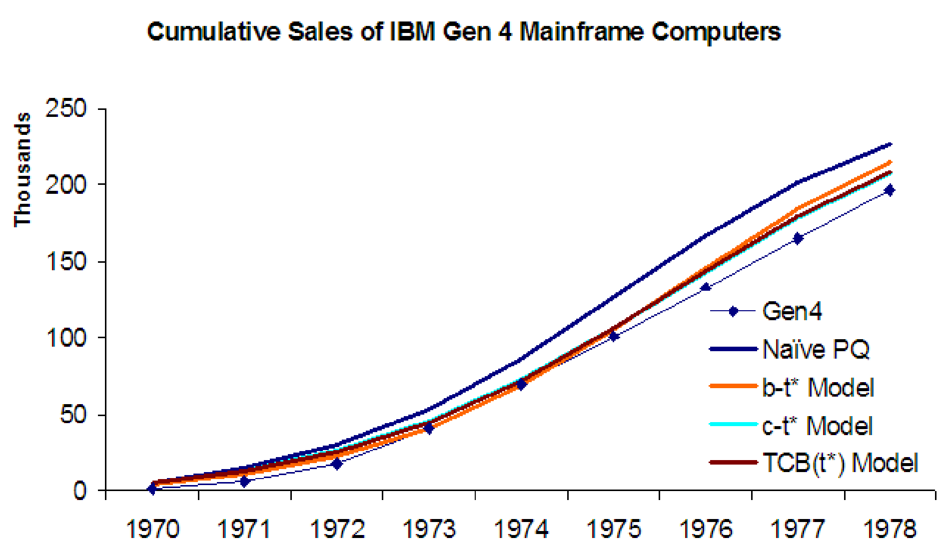


Figure 28: Cumulative Demand vs. Cumulative Pre-launch Forecasts

B.3 How biased estimates of market potential m affect forecast accuracy?

Through out the paper we assumed that market potential m is known. However, in real life situations this is hardly possible. Potential bias in estimating m can be incorporated into the forecast accuracy calculations easily, and in the situations where estimating m is really difficult, forecast accuracy scenarios can be generated with various bias amounts.

Let $N(T)$ be the forecasted cumulative demand for the first T periods using the optimal market potential m and $A(T) = X(1)+...+X(T)$ is the actual cumulative demand until time T. The forecast accuracy of first T-period cumulative demand is given by $MAPE(T) = [N(T) - A(T)]/A(T)$ when there is no bias in m estimate. Let b represent the bias in estimating market potential m. Then the equation (72) gives the forecast with biased market potential and equation (73) gives the resultant forecast accuracy in MAPE.

$$N_b(T) = (1+b)m\frac{[1 - e^{-(p+q)T}]}{[1-(q/p)e^{-(p+q)T}]} = (1+b)N(T) \tag{72}$$

$$MAPE_b(T) = \frac{|N_b(T) - A(T)|}{A(T)} = |bN(T)/A(T) + N(T)/A(T) - 1| \tag{73}$$

Following cases explain the direction and the magnitude of change in MAPE calculations:

- Case 1. $N(T)/A(T) > 1$ (overestimation case)

  if $b > 0$, then MAPE is increased by $bN(T)/A(T)$ (positive bias)

  if $b < 0$, then MAPE is decreased by $bN(T)/A(T)$ (negative bias)

- Case 2. $N(T)/A(T) < 1$ (underestimation case)

  if $b > 0$, then MAPE is decreased by $bN(T)/A(T)$ (positive bias)

  if $b < 0$, then MAPE is increased by $bN(T)/A(T)$ (negative bias)

Since reported MAPE values are calculated from $N(T)$ and $A(T)$ such that $N(T)/A(T) = 1 \pm MAPE$, MAPE value for a given bias b is calculated by equation (74):

$$MAPE_b = MAPE \pm b(1 \pm MAPE) \tag{74}$$

Where $\pm$'s are replaced by (+) signs when $N(T)/A(T) > 1$, and by (-) signs when $N(T)/A(T) < 1$. For example, with 5% positive bias in m, originally reported 20% MAPE can increase to 26% MAPE. On the other hand, negative 5% bias in m can reduce 20% MAPE down to 14%. If this was an underestimation case (i.e., $N(T)/A(T) < 1$), then +5% bias would result in 16% MAPE, while -5% bias would give 24% MAPE. Average MAPE with +5% bias is (26+16)/2=21%, while with -5% bias, the average MAPE is (24+14)/2=19%.

Equation (74) therefore shows that overestimating m is more risky than underestimating. Because, in the situations where positive bias in m increases MAPE, it does more than the same amount negative bias; and where positive bias decreases MAPE, it does less than the same amount negative bias. This result is opposite of peak time t^* bias, where underestimation is more risky in terms of MAPE calculations. As in the peak time bias case, one needs to be careful in assessing the impact of under vs. overestimation of any parameter in the context of inventory holding costs versus stock out penalties.

APPENDIX C

ADDENDUM FOR CHAPTER 5

C.1 Proof of Propositions

C.1.1 Proof of Proposition 1

For a linear trend curve $y(t) = a_0 + bt$ and monthly seasonality factors s_1, s_2 and s_3, we would like to find seasonal split factors $SS(t+i-1)$, for the i^{th} month within the quarter that that consists of months $t, t+1$ and $t+2$.

When the linear trend curve is subject to multiplicative seasonality, then the value of the time series for month $(t+i-1)$ is given by $X(t+i-1) = y(t+i-1)s_i$. Seasonal split (SS) factors are defined as $SS(t+i-1) = X(t+i-1)/(\sum_{i=1}^{3} X(t+i-1))$. Plugging the formula for the time series into this equation, we obtain:

$$SS(t) = \frac{s_1(1 - b/(a+b))}{3 + b(s_3 - s_1)/(a+b)} \tag{75}$$

$$SS(t+1) = \frac{s_2}{3 + b(s_3 - s_1)/(a+b)} \tag{76}$$

$$SS(t+2) = \frac{s_3(1 + b/(a+b))}{3 + b(s_3 - s_1)/(a+b)} \tag{77}$$

where $a = a_o + bt$.

Mid point (or average) level of the linear trend curve is given by $(a+b)$. We define $g(t)$ as the slope of the trend curve relative to its average level for a quarter starting with month t. Therefore, for linear trend cases $g(t) = b/(a+b)$. Therefore, substituting this definition into equations (75), (76) and (77), we obtain the formal relationship of seasonal split factors with seasonality factors and relative slope of the trend.

References

[1] Abhyankar, H. S. and Graves, S. C., "Creating an inventory hedge for markov-modulated poisson demand: An application and a model," *Manufacturing & Service Operations Management*, vol. 3, pp. 306–320, 2001.

[2] Armstrong, J. S., *Long-range forecasting*. New York: Wiley, 1985.

[3] Armstrong, J. S., "Findings from evidence-based forecasting: Methods for reducing forecast error," *International Journal of Forecasting*, vol. 22, pp. 583–598, 2006.

[4] Baker, J., "Emergent pricing structures in ltl transportation," *Journal of Business Logistics*, vol. 12, no. 1, pp. 191–202, 1991.

[5] Baker, K., "Safety stocks and commonality," *Journal of Operations Management*, vol. 6, no. 1, pp. 13–22, 1985.

[6] Barnhart, C. and Kim, D., "Routing models and solution procedures for regional less-than-truckload operations," *Annals of Operations Research*, vol. 61, pp. 67–90, 1995.

[7] Barrett, C., "Discounts are legitimate, but beware," *Traffic World*, p. 38, 2007. July 2.

[8] Bass, F. M., "A new product growth model for consumer durables," *Management Science*, vol. 15, no. 5, pp. 215–227, 1969.

[9] Bass, F.M., G. K. F. T. and Githens, M., "Directv: Forecasting diffusion of a new technology prior to product launch," *Interfaces*, vol. 31, pp. S82–S93, 2001.

[10] Bass, F. M., K. T. and Jain, D., "Why the bass model fits without decision variables," *Marketing Science*, vol. 13, no. 3, p. 203223, 1994.

[11] Bayus, B. L., "High-definition television: assessing demand forecasts for a next generation consumer durable," *Management Science*, vol. 39, no. 11, pp. 1569–1585, 1993.

[12] Benjaafar, S., C. W. L. and Mardan, S., "Production-inventory systems with imperfect advance demand information and due date updates." Working Paper, University of Minnesota, 2007.

[13] Berry, W.L., T. W. and Boe, W., "Product structure analysis for the master scheduling of assemble-to-order products," *International Journal of Operations and Production Management*, vol. 12, no. 11, pp. 24–41, 1992.

[14] Bohman, R., "Stb hands down major trucking deregulation decision," *Logistics Management*, 2007. June 1, http://www.logisticsmgmt.com/article/CA6451283.html.

[15] Bunn, D. W. and Vassilopoulos, A. I., "Using group seasonal indices in multi-item short-term forecasting," *International Journal of Forecasting*, vol. 9, pp. 517–526, 1993.

[16] Bunn, D. W. and Vassilopoulos, A. I., "Comparison of seasonal estimation methods in multi-item short-term forecasting," *International Journal of Forecasting*, vol. 15, pp. 431–443, 1999.

[17] Centa, H., "Facts on freight - what determines ltl discount percentages," *Production Machining*, pp. 20–21, 2007. December 1.

[18] Chainalytics, L., "Model-based benchmarking consortium." http://www.chainalytics.com/services/transport_mbbc.asp, 2008.

[19] Chen, F. and Song., J.-S., "Optimal policies for multiechelon inventory problems with markov-modulated demand," *Operations Research*, vol. 49, pp. 226–234, 2001.

[20] Chiang, Y. S. and Roberts, P. O., "A note on transit time and reliability for regular-route trucking," *Transportation Research Part B: Methodological*, vol. 14, no. 1-2, pp. 59–65, 1980. March-June.

[21] Choffray J. M., Lilien, G., "A decision support system for evaluating sales prospects and launch strategies for new products," *Industrial Marketing Management*, vol. 15, p. 7585, 1986.

[22] Chu, C. W., "A heuristic algorithm for the truckload and less-than-truckload problem," *European Journal of Operational Research*, vol. 165, no. 3, pp. 657–667, 2005.

[23] CSCMP, "18th annual state of logistics report," *CSCMP.org*, 2007.

[24] De Gooijer, J. G. and Hyndman, R. J., "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, pp. 443–473, 2006.

[25] Dogramaci, A., "Design of common components considering implications of inventory costs and forecasting," *AIIE Transactions*, pp. 129–135, 1979.

[26] Elmaghraby, W. and Keskinocak, P., *Combinatorial Auctions in Procurement*. C. Billington, T. Harrison, H. Lee, J. Neale (editors), Kluwer Academic Publishers, 2003. The Practice of Supply Chain Management.

[27] Findley, D. F., M. B. C. B. W. R. O. M. C. and Chen, B. C., "New capabilities and methods of the x-12-arima seasonal adjustment program," *Journal of Business and Economic Statistics*, vol. 16, pp. 127–152, 1998.

[28] Gallego, G. and Ozer, O., *Optimal use of demand information in supply chain management*. J. Song and D. Yao, editors, Kluwer Academic Publishers, 2002. Supply Chain Structures: Coordination, Information and Optimization.

[29] Grant, K. B. and Kent, J. L., "Investigation of methodologies used by less-than-truckload (ltl) motor carriers to determine fuel surcharges." Technical Report by Midwest Transportation Consortium, MTC Project 2006-03, 2006.

[30] Graves, S. C., "A single-item inventory model for a nonstationary demand process," *Manufacturing & Service Operations Management*, vol. 1, no. 1, 1999.

[31] Graves, S. C. and Willems., S. P., "Strategic inventory placement in supply chains: Non-stationary demand." Working Paper, MIT, available from http://web.mit.edu/sgraves/www/papers/, 2005.

[32] Hall, R. W. and Zhong, H., "Decentralized inventory control policies for equipment management in a many-to-many network," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 10, pp. 849–865, 2002. December.

[33] Hannon, D., "Best-kept secrets for reducing ltl costs," *Purchasing*, pp. 33–36, 2006. May 18.

[34] Hannon, D., "Ltl buyers face an uphill battle," *Purchasing*, p. 42, 2006. April 6.

[35] Heeler, R. M. and Hustad, T. P., "Problems in predicting new product growth for consumer durables," *Management Science*, vol. 26, pp. 1007–1020, 1980.

[36] Huang, G.Q., Z. X. Y. and Lo, V. H., "Optimal supply chain configuration for platform products: impacts of commonality, demand variability and quantity discount," *International Journal of Mass Customisation*, vol. 1, no. 1, 2005.

[37] Institute, B. B. R., "New-product forecasting research database: Diffusion modeling." http://www.frankmbass.org/NPF/Default.aspx, 2008.

[38] Jain, D. C. and Rao, R., "Effect of price on the demand for durables: Modeling, estimation, and findings," *Journal of Business and Economic Statistics*, vol. 8, no. 2, pp. 163–170, 1990.

[39] James, W. and Stein, C., "Estimation with quadratic loss," *In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, vol. 1. Berkeley, CA: University of California Press*, p. 361379, 1961.

[40] Johnson, M. E. and Anderson, E., "Postponement strategies for channel derivatives," *The International Journal of Logistics Management*, vol. 11, pp. 19–35, 2000.

[41] Kahn, K. B., "An exploratory investigation of new product forecasting practices," *The Journal of Product Innovation Management*, vol. 19, p. 133143, 2002.

[42] Kalish, S., "A new product adoption model with price, advertising and uncertainty," *Management Science*, vol. 31, pp. 1569–1585, 1985.

[43] Kalish, S. and Lilien, G. L., *Applications of Innovation Diffusion Models in Marketing*. Mahajan V. and Wind Y. (editors), Cambridge, MA: Ballinger, 1986. Innovation Diffusion of New Product Acceptance.

[44] Kamakura, W. A. and Balasubramanian, S. K., "Long-term view of the diffusion of durables," *International Journal of Research in Marketing*, vol. 5, pp. 1–13, 1988.

[45] Katayama, N. and Yurimoto, S., "The load planning problem for less-than-truckload motor carriers and solution approach." Working Paper, Ryutsu Keizai University, Ibaraki, Japan, 2002.

[46] Keaton, M. H., "Are there economies of traffic density in the less-than-truckload motor carrier industry? an operations planning analysis," *Transportation Research Part A: Policy and Practice*, vol. 27, no. 5, pp. 343–358, 1993. September.

[47] Kim, B., L. J. P. K. Z. G. and Lee, S., "Configuring a manufacturing firm's supply network with multiple suppliers," *IIE Transactions*, vol. 34, no. 8, pp. 663–677, 2002.

[48] Kurawarwala, A. A. and Matsuo, H., "Forecasting and inventory management of short life cycle products," *Operations Research*, vol. 44, no. 1, pp. 131–150, 1996.

[49] Lawrence, K. D., L. W. H., *Applications of diffusion models: some empirical results.* Wind, Yoram, Mahajan, Vijay, Cardozo, R.C., (editors), Lexington Books, Lexington, MA, 1981. New Product Forecasting.

[50] Lee, J., B. P. and Kamakura, W., "A bayesian model for pre-launch sales forecasting of recorded music," *Management Science*, vol. 49, p. 179196, 2003.

[51] Lemon, G. H. and Krutchkoff, R. G., "Smooth empirical bayes estimators: with results for the binomial and normal situations," *In Proceedings of the symposium on empirical Bayes estimation and computing in statistics, Texas Tech University mathematics series no. 6*, p. 110140, 1969.

[52] Lenk, P. and Rao, A., "New products from old: Forecasting product adoption by hierarchical bayes procedures," *Marketing Science*, vol. 9, pp. 42–53, 1990.

[53] Lilien, G. L., R. A. V. d. B. C., *Diffusion Models: Managerial Applications and Software.* Mahajan V. and Wind Y. (editors), Cambridge, MA: Ballinger, 2000. Innovation Diffusion of New Product Acceptance.

[54] Mahajan, V., M. E. and Bass, F., "New-product diffusion models: A review and directions for research," *Journal of Marketing*, vol. 54, pp. 1–26, 1990.

[55] Mahajan, V. and Sharma, S., "Simple algebraic estimation procedure for innovation diffusion models of new product acceptance," *Technological Forecasting and Social Change*, vol. 30, pp. 331–346, 1986.

[56] Makridakis, S., A. A. C. R. F. R. H. M. L. R. N. J. P. E. and Winkler, R., "The accuracy of extrapolation (time series) methods: results of a forecasting competition," *Journal of Forecasting*, vol. 1, pp. 111–153, 1982.

[57] Makridakis, S. and Hibon, M., "The m3-competition: results, conclusions and implications," *International Journal of Forecasting*, vol. 16, pp. 451–476, 2000.

[58] Management, M. H., "Tariff practices add complexity for carriers." Material Handling Management Online, http://www.mhmonline.com/nID/2252/MHM/viewStory.asp, 2008.

[59] Meade, N. and Islam, T., "Forecasting with growth curves: An empirical comparison," *International Journal of Forecasting*, vol. 11, pp. 199–215, 1995.

[60] Meade, N. and Islam, T., *Forecasting the diffusion of innovations: Implications for time series extrapolation.* J. S. Armstrong (editor), Boston: Kluwer Academic Publishers, 2001. Principles of Forecasting.

[61] Meade, N. and Islam, T., "Modelling and forecasting the diffusion of innovation - a 25-year review," *International Journal of Forecasting*, vol. 22, pp. 519–545, 2006.

[62] Meyer, M. and Lehnerd, A., *The Power of Product Platforms: Building Value and Cost Leadership.* New York: Free Press, 1997.

[63] Miller, D. M. and Williams, D., "Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy," *International Journal of Forecasting*, vol. 19, pp. 669–684, 2003.

[64] Miller, D. M. and Williams, D., "Damping seasonal factors: Shrinkage estimators for the x-12-arima program," *International Journal of Forecasting*, vol. 20, pp. 529–549, 2004.

[65] Murphy, P. and Corsi, T., "Modelling sales force turnover among ltl motor carriers: A management perspective," *Transportation Journal*, vol. 19, pp. 25–37, 1989. September 1.

[66] Musgrave, J., "A set of end weights to end all end weights." Unpublished internal note, US Bureau of the Census, Washington, 1964.

[67] Norton, J. and Bass, F., "A diffusion theory model of adoption and substitution for successive generations of high-technology products," *Management Science*, vol. 33, pp. 1069–1086, 1987.

[68] Park, B., "A framework for integrating product platform development with global supply chain configuration," *GIT PhD Dissertation, Georgia*, 2001.

[69] Putsis, W. P., "Temporal aggregation in diffusion models of first-time purchase: Does choice of frequency matter?," *Technological Forecasting and Social Change*, vol. 51, pp. 265–279, 1996.

[70] Putsis, W. P. and Srinivasan, V., *Estimation techniques for macro diffusion models.* Mahajan V. and Wind Y. (editors), Cambridge, MA: Ballinger, 2000. Innovation Diffusion of New Product Acceptance.

[71] Quenneville, B., L. D. and Lefrancois, B., "A note on musgrave asymmetrical trend-cycle filters," *International Journal of Forecasting*, vol. 19, pp. 727–734, 2003.

[72] Radas, S. and Shugan, S. M., "Seasonal marketing and timing new product introductions," *Journal of Marketing Research*, vol. 35, no. 3, pp. 296–315, 1998.

[73] Salvador, F., F. C. and Rungtusanatham, M., "Modularity, product variety, production volume, and component sourcing: theorizing beyond generic prescriptions," *Journal of Operations Management*, vol. 20, pp. 549–575, 2002.

[74] Schulz, J. D., "Ltl market update: Getting the balance right," *Logistics Management*, 2007. June 1, http://www.logisticsmgmt.com/article/CA6451284.html.

[75] Sharif, M. N., R. K., "Binominal innovation diffusion models with dynamic potential adopter population," *Technological Forecasting and Social Change*, vol. 20, pp. 63–87, 1981.

[76] Shiskin, J., "Decomposition of economic time series," *Science*, vol. 128, no. 3338, 1958.

[77] Smith, L.D., C. J. and Mundy, R., "Modeling net rates for expedited freight services," *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, p. 192207, 2007.

[78] Srinivasan, V. and Mason, C. H., "Non-linear least squares estimation of new product diffusion models," *Marketing Science*, vol. 5, no. 2, pp. 169–178, 1986.

[79] Sultan F., F. J. U. and R., L. D., "A meta-analysis of applications of diffusion models," *Journal of Marketing Research*, vol. 27, pp. 70–77, 1990.

[80] Thomas, R. J., "Estimating market growth for new products: An analogical diffusion models approach," *Journal of Product Innovation Management*, vol. 2, pp. 45–55, 1985.

[81] Van den Bulte, C., L. G. L., "Bias and systematic change in the parameter estimates of macro-level diffusion models," *Journal of Product Innovation Management*, vol. 16, no. 4, pp. 338–353, 1997.

[82] Van den Bulte, C., S. S., "Social contagion and income heterogeneity in new product diffusion: A meta-analytic test," *Marketing Science*, vol. 23, no. 4, pp. 530–544, 2004.

[83] Virrolleaud, P., "Les ventes des trois premières années prévisibles à 30 percent," *Usine Nouvelle*, 1983.

[84] White, P. K., "Simulating a non-stationary process using bivariate thinning: the case of typical weekday arrivals at a consumer electronics store," *Proceedings of the 1999 Winter Simulation Conference*, pp. 458–461, 1999.

[85] Woensel, T. V., W. B. and Vandaele, N., "Validating state-dependent queueing models for uninterrupted traffic flows using simulation," *4OR: A Quarterly Journal of Operations Research*, vol. 4, p. 159174, 2006.

[86] Wu, S. D., A. B. B. R. T. and Armbruster, C. A., "Predicting high-tech market demands using leading indicators," *Tech. report, Lehigh University, Department of Industrial and Systems Engineering, Bethlehem, PA,*, 2003.

[87] Xie, J., S. M. S. M. and Wang, Q., "Kalman filter estimation of new product diffusion models," *Journal of Marketing Research*, vol. 34, pp. 378–393, 1997.

VITA

Evren Özkaya (born in Izmir, Turkey in 1982) graduated summa cum laude with B.Sc. in Industrial Engineering from Bilkent University, Ankara, Turkey in 2004. In 2005, he received his M.S. in Industrial Engineering from Georgia Tech, while pursuing his Ph.D. in the same department with a minor in Management of Technology. Özkaya is currently working for McKinsey and Company as an Associate Consultant in the Operations Practice and based in the Atlanta Office. Previously, he has worked for companies like Intel Corporation, Schneider Logistics, Case New Holland, Coca-Cola and Aselsan Electronics in areas like Demand Management and Forecasting, Supply Chain Modeling, Logistics Pricing, Manufacturing, Sales and IT.

Özkaya served as the President of Operational Research Club at Bilkent University. In his senior year at college, he was the Founder and Project Leader for Career Project of European Students of Industrial Engineering and Management (ESTIEM), an organization comprised of 65 universities in 24 countries in Europe. He is the former President of Turkish Student Organization and the former Vice President of Management Consulting Club at Georgia Tech. He also coordinated and instructed Georgia Tech's Turkish Folk Dance Group for two years during his graduate studies. In his last year at Georgia Tech, Özkaya served in the Student Health Insurance Program (SHIP) Committee of the University System of Georgia representing international students of Georgia.

Özkaya is the first recipient of distinguished John L. Imhoff Scholarship of Institute of Industrial Engineers (2006) for his international contributions to the industrial engineering profession. He has been selected as the Global Logistics Scholar of the Supply Chain and Logistics Institute at Georgia Tech in 2007. Özkaya's current research focuses on Demand Management and Forecasting, Supply-Demand Alignment and Supply Chain Information Systems.