PRICING AND ORDER FULFILLMENT FOR ONLINE MARKETPLACES

by

Yasin Yıldız

B.S., Industrial Engineering, Boğaziçi University, 2016

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering
Boğaziçi University

2019

PRICING AND ORDER FULFILLMENT FOR ONLINE MARKETPLACES

APPROVED BY:

Assist. Prof. Mustafa Gökçe Baydoğan ....
(Thesis Supervisor)

Prof. Taner Bilgiç ....

Prof. Gürhan Kök ....

DATE OF APPROVAL: 02.08.2019

ACKNOWLEDGEMENTS

Firstly, I would like to express my thanks to all of my professors in Boğaziçi University. I offer special thanks to Asst. Prof. Mustafa Baydoğan and Prof. Taner Bilgiç for their guidance and support during this study. Also we thank Prof. Mehmet Gümüş for providing the data used in this thesis.

I am indebted to my friends and family for their endless support during this thesis. I am thankful to Çağdaş Aktepe, Mehmet Yekta Tabanlı and Emre Haliloğlu for their technical and academic support throughout this study.

This research was partially supported by Boğaziçi University Research Fund Grant Number 13501.

ABSTRACT

PRICING AND ORDER FULFILLMENT FOR ONLINE MARKETPLACES

The main purpose of this thesis is to provide a wholistic model for both pricing and order fulfillment problems of an online retailer. Real sales and shipment data of a company are used for both forecasting and fulfillment decisions. A tree-based ensemble model is offered for the demand forecasting process by considering pricing and promotion effects. The generated sales forecasts are added to the fulfillment model as future orders. These orders can be fulfilled by any FCs by considering the corresponding fulfillment costs. Therefore, the offered data driven model tries to optimize total profit of the company while minimizing these operational costs. These results are compared across different cases for price and capacity levels. Due to the randomness of the generated demand forecasts, a prescriptiveness coefficient is used to evaluate the reliability of the offered results. As a result of this study, an optimal inventory allocation, fulfillment and pricing strategy are provided to the company.

ÖZET

ÇEVRİM İÇİ PAZARLARDA FİYATLAMA VE SİPARİŞ KARŞILAMA

Bu tezin temel amacı, bir çevrim içi perakendecinin hem fiyatlandırma hem de sipariş yerine getirme problemleri için bütünsel bir model sağlamaktır. Şirketin gerçek satış ve sevkiyat verileri hem satış tahminleme hem de sipariş karşılama modelleri için kullanıldı. Talep tahmin süreci için fiyatlandırma ve promosyon etkileri göz önünde bulundurularak ağaç tabanlı bir topluluk modeli oluşturuldu. Üretilen satış tahminleri sipariş karşılama modeline, gelecek siparişler olarak eklendi. Bu siparişler, ilgili karşılama maliyetleri göz önünde bulundurularak herhangi bir sipariş karşılama merkezi tarafından karşılanacak şekilde modellendi. Bu nedenle, sunulan modelin, işletme maliyetlerini minimize ederken, şirketin toplam kârını da optimize etmesi amaçlandı. Genel problemi değişik açılardan değerlendirebilmek için, fiyat ve kapasite temelli farklı senaryolar yaratıldı. Üretilen talep tahminlerinin stokastikliği nedeniyle, hesaplanan sonuçların güvenilirliğini değerlendirmek için bir güvenilirlik katsayısından faydalanıldı. Bu çalışma sonucunda, şirkete optimum stok tahsisi, sipariş karşılama ve fiyatlandırma stratejisi sağlanmış oldu.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

β	Coefficient vector for regression
β_i	i^{th} element of coefficient vector β
x	Feature vector
x_i	i^{th} element of feature vector x
y	Response vector
y_i	i^{th} element of response vector y
A	Actual data for regression
A_i	i^{th} element of vector A
F	Fitted value vector for regression
F_i	i^{th} element of fitted value vector F

LIST OF ACRONYMS/ABBREVIATIONS

FC	Fulfillment Center
RF	Random Forest regressor
SST	Total sum of squares
SSE	Sum of squared errors
$MAPE$	Mean absolute percentage error
SKU	Stock keeping unit

1. INTRODUCTION

Online marketplaces are the platforms that make possible to buy or sell products/services online without need for a physical communication between the seller and the buyer. All of the transactions are made online, until the buyer receives the service or the product. It saves huge amount of unnecessary investments and lots of variable costs for both customers and suppliers.

People tend to keep shopping time as short as possible by using e-commerce platforms therefore the security systems, prices and fulfillment performances are very crucial for all online sellers. All of the online suppliers, try to build an optimal shopping experience for possible customers. By using the current potential of the information technologies, nearly all of the trading actions (clicks, followed path on the website etc.) of customers are being tracked and stored in database systems. There are some restrictions that regulate and protect the usage of personal data. These regulations are not same in all countries but in most, keeping and using basic transaction data like buying price, shipping cost etc. is allowed. Not only the data usage is regulated, but the personalized campaigns, pricing and different offers are also being regulated by laws in most countries.

The other side of that transaction is the customers and their actions. Customers always look for a better option while performing a trade. The factors that describe what a better option is not same for all customers. Customers want to compare all of the suppliers according to their fulfillment performances and proposed prices for a specific product. The fastest way to perform that mini-research is just checking different platforms, sites or suppliers online. This situation helps customers to find the best product-seller pair that perfectly fits their needs and willingness to pay levels. Most of the time, that challenge for the online market owners turns out to be a well-known problem 'Prisoner's Dilemma'. Once a service or product provider cuts the price of a product/service, all of the suppliers try to mimic that action to keep customers at their side. Therefore, the online markets have an invisible equilibrium and even a simple action for one action taker may change all the rules on the game.

For an online retailer, fulfillment cost consists of two main parts: Shipping cost and store/warehouse operation cost. As the product assortment of a retailer enlarges, these fulfillment costs get higher. The main reason is, not all the products have same size and characteristics to apply same fulfillment strategy which causes a decrease in fulfillment efficiency. Figure 1.1 shows Amazon's shipping and operational costs as a percentage of net sales [1]. Fulfillment cost ratio of Amazon has a smooth increase pattern in last 12 years. It is caused by the increase in the service levels of the online retailers such as shorter expected delivery times for sale orders.

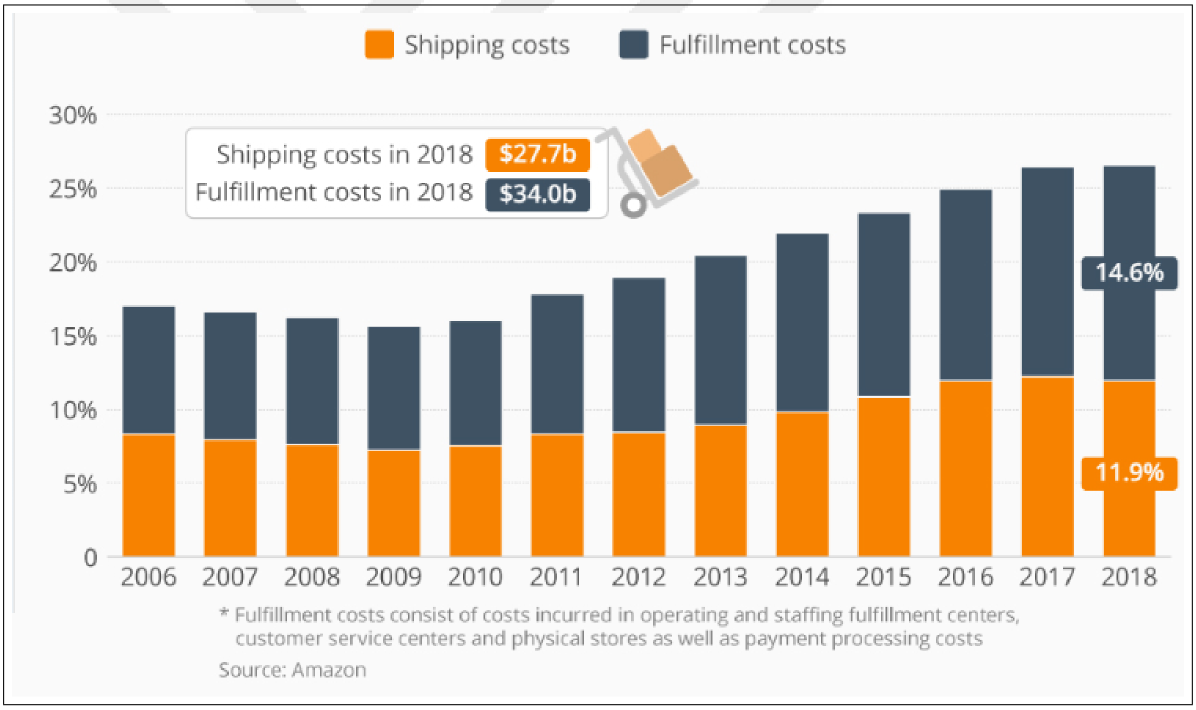

Figure 1.1. Shipping and operational costs of Amazon as a percantege of net sales

In today's retail industry, brick and mortar stores lose their importance on markets with the huge improvements on e-commerce markets and their technologies. Figure 1.2 shows the steady growth on the share of e-retail industry in the United States [2]. Both the growth pattern in Figure 1.2 and the increase in cost ratios in Figure 1.1 explains why fulfillment efficiency is that crucial in online retailing industry.

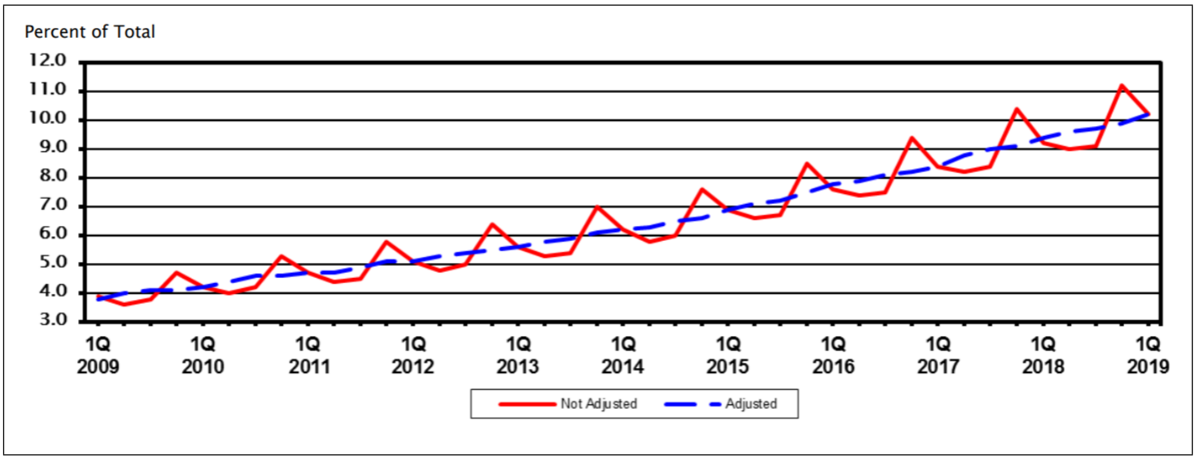

Figure 1.2. Estimated Quarterly U.S. Retail E-commerce Sales as a Percent of Total Quarterly Retail Sales

In this thesis, a data-driven joint forecasting and fulfillment model is developed and applied to data obtained from a US company which sells technological devices and computer parts. Almost all of the products they sell are imported from abroad therefore they have respectively longer lead times. That situation forces the company to optimize its inventory levels not only at the purchasing stage but during the in-sale period. The main instrument to manage that inventory level is proposed prices for the products. In this study, different inventory and price levels are used to build a model to get an optimal fulfillment process.

The main objective of the stated problem is to maximize company's total profit while considering the fulfillment costs. The optimal results offer about 5% to 25% improvement on total profit for different price and capacity levels. Optimal results suggest that, pricing and capacity levels can be used to manage service levels and inventory capacities of a company. By using such a model, a company can manage its resources more effectively to maximize its total profit.

2. LITERATURE REVIEW

Order fulfillment and pricing issues gain importance as the online marketplaces grow. There has been extensive research on the literature especially on the fulfillment issues. The pricing part of the problem is being researched mostly in the last 20 years with the help of information technologies. Most of the research tried to construct models for industry to maximize the expected profit for specific problems. In this thesis, we proposed a model that solves the fulfillment problem by using the forecasts generated by a data-driven model which calculates and uses the effects of the external factors.

The literature on this area can be divided into three main branches: Demand forecasting & pricing, order fulfillment and the joint models for these two problems.

2.1. Demand Forecasting & Pricing

Demand forecasting has a crucial importance on most of the supply chain problems. Forecasts are used as input values for the optimization problems to simulate real life in a specified time horizon. In literature, there is numerous research related to forecasting with different approaches.

Demand models either start from individual choice behavior of consumers and work upwards or start at an aggregate level to describe the market level demand. Huang *et al.* studied these demand functions on their survey research [3]. Their research proposed an overview for the deterministic and stochastic demand functions which have different parameters such as price, competition effect etc. These functions were used in models by assuming that the demand function of a product or a service is known and can be used in the supply chain operations. In the last 20 years, with the huge development on data collection and handling technology, different forecasting methods are evaluated and implemented in industry. In literature, researchers focused on different methods to find and analyze the historical sales data patterns. In the last decade, researchers focused on alternative tailor-made forecasting methods to analyze and use the effects of external factors such as promotions. Kuo *et al.* analyzed a fuzzy neural network method to generate sales forecasts [4]. Their research compares the traditional forecasting methods such as simple regression models with more complicated neural networks. Their aim was to find the effects of unusual events such as promotions and incorporate these effects in the forecasting model. In their research, they proposed that time-series based methods are successful only on data-sets which have clear seasonality or time dependency. The proposed model used an artificial neural network to find and use the effects of the nonlinear relations of different inputs. The model was constructed in a way the base forecast was generated by ANN and the external effects were predicted using different methods. After external factors were evaluated, an ensemble model was constructed to get the final forecasts.

The other branch of the pricing and forecasting studies is based on historical/statistical learning methods. The literature on the learning of the price effect on the demand forecasts compares different methodologies on different problems. Chinthalapati *et al.* studied reinforcement learning for a dynamic pricing model by including the competition effects in their pricing model with different scenarios [5]. The information shared with the competitors was different in different scenarios. The model they built offered to solve the dynamic pricing problem and inventory replenishment in a competitive and stochastic environment. Their forecast methodology was similar to the method in this study but general problem to be solved is different than the proposed problem in this thesis. They modeled the problem by considering the stochasticity of the sales behaviour of the customers and they used stochastic optimization techniques to get optimal solution. However, in this thesis, we modeled the system in a stochastic way by using a coefficient which can be defined as a transformation coefficient, and an equivalent LP is utilized to get final results. The details of that coefficient will be described on Section 4.2.4.

On the online marketplaces, the demand comes from different geographical zones. Each city and region has its own characteristic pruchasing behaviour. In literature, researchers focus on location-based forecasting especially for energy markets which shares many features with online markets. Both have a couple of distribution centers and all demand is met from there. Based on the source of the demand, resources are distributed to minimize the fulfillment costs. Filizadeh and Omran studied a location-based forecasting model for vehicle charging stations [6]. They investigated possible external factors and built a fuzzy system to estimate the demand. The number of charging stations in an area and distance from city centers were crucial inputs for their location-based forecasting model. The proposed model tried to estimate the demand for a charging station based on its location by considering the external factors such as climate conditions. The external factors for forecasting methods differ based on the industry and the customer behaviour for a specific product. For online marketplaces, prices are very important for a customers' decision to buy a product from a vendor. Therefore, the demand of a specific product is highly dependent on its price in the online markets. Promotions, special days and other events that may affect the price of the product are also important inputs for demand forecasting models. In their study, Ramanathan and Muyldermans [7] used a soft drink company's data to find and analyze the effects of these special events on the sales pattern of the products. Their study proposed that the effects of these special events were different for different product categories. They analyzed three different product families and generated input parameters using structural equation modeling (SEM). Unlike other researches, their model did not propose a forecasting model but discussed the effects of special events on the sales of the products. In industry, prices of the products mostly stay stable until a campaign or discount offer is launched. There are always some price fluctuations during the normal lifetime of a product, but a discount process may change all the characteristics of the sales pattern. In literature, most of the research focused on the scenario where the price effect of a product is very similar on different time intervals. The special day effects, promotion effects etc. were ignored on most papers. Ferreira *et al.* proposed a solution to the pricing and inventory management problem of an online retailer which offers short-time promotions frequently [8]. Obviously, finding the best price at a given time was very crucial to maximize the revenue. Their main approach was the demand curve of a product during a short-time discount could be learned by an algorithm. The company had some products which were never sold before, therefore similar product features were used to predict the future sales for these new products. As an initial solution to the pricing problem, price of the competing style was used. As a result, the model offered optimal prices for products on a daily basis. However, that study did not take the fulfillment processes into consideration while solving the specified supply chain problem.

Another important point for a service supplier company is its service level, therefore the action takers on online market places try to take this into account while constructing a supply chain system which determines the order fulfillment strategies. Xia *et al.* offered a model [9] that include both pricing and the service level as a factor so that their joint effects on uncertain demand is more realistic. They proposed a revenue management model that highly depends on the demand forecast. Their demand estimation model used a hybrid model that consists of Bayesian learning and simulated annealing algorithms. By considering the service level offered by them, the firm needed to manage its resources to meet the promised service level. Therefore, the proposed model was able to show the joint effects of the pricing and service level strategies.

Because of the publicly shared price data of the products in different e-retail platforms, the competition effect on the pricing problems is a fundamental fact for the online markets. These competition effects on dynamic pricing strategies is analyzed by Fisher *et al.* in their study [10]. It was based on the fact that the responses of the e-tailers to their competitors should be taken into consideration while modeling a pricing problem on online platforms. They tried to find the best response to the price changes of the competitors by adding the customer choices and stock out effects into the model. As a result, they could give the best-response prices which maximize the total revenue.

In this research, the transactional sales data of an online retailer company are used to forecast future sales. Machine learning algorithms are used to evaluate special day, price and product family effects. Unlike the other studies, time series property of the sales data is added to the model by generating new features. Because of the nature of the online retailing, the location-based sales data are also added to the forecast model to increase the accuracy of the proposed model. All of the products have different forecasts for different cities which is very crucial for the fulfillment part of our research.

2.2. Order Fulfillment

In online retailing industry, companies fulfill the customer demands with different fulfillment centers by considering the corresponding costs and capacities. To construct a model that maximizes the company's total profit, the relationship between the fulfillment process and other supply chain operations should be established in a systematic way. Lin and Shaw [11] described the business flow of the order fulfillment process of an enterprise by adding the effects of the operational elements. They proposed an information system approach that connects the supply chain networks and order fulfillment processes to increase the efficiency of the fulfillment process by improving the information share strategies, capacity management approaches and resource allocation strategies. Today, most of the e-commerce companies still struggle with that issue. It is because the supply chain operations of the enterprises should be integrated in a way that all parts of the system consider the global efficiency. The other valuable research area in the literature is optimizing the fulfillment processes by better operational decisions. There are two main constraints in these kind of problems: inventory level of the products and the fulfillment capacities of the companies. In their study [12], Acimovic *et al.* built a forward-looking fulfillment model for the online retailing industry. There was more than one fulfillment center and the fulfillment cost of an order differed by the fulfillment centers. The proposed linear model took orders into consideration one by one. By doing this, they could analyze the effects of split-or-not on the multi-item orders in their optimization model. Their objective was to reduce the fulfillment costs by deciding on fulfillment strategies for the given orders. Unlike Acimovic's study, Torabi *et al.* allowed fulfillment centers to perform transshipment between these centers [13]. This difference allowed the proposed model to use the total inventory in a more effective way. Their study aimed to assign customers to the fulfillment centers by considering the inventory levels and fulfillment costs. The proposed solution offered transshipment if it is necessary which also had an operational cost. Similar to the Acimovic's paper, they proposed an optimal solution for whole fulfillment process.

All of the mentioned research assumed that an order can be fulfilled by the fulfillment centers. However, an incoming order for a product may be fulfilled not only with fulfillment centers but also by using the existing brick and mortar stores and shipping from that location. Ishfaq *et al.* studied these alternative ways to fulfill the orders [14]. The proposed model gave the company different fulfillment options to the company by adding the effects of the corresponding costs. They used a linear price-demand function by assuming that the revenue function is concave and that the demand decreases as the price increases. As a result, they offered the optimal solution under these assumptions for an online retailer.

In this thesis, the main fulfillment problem has similarities with the aforementioned researches. There is more than one fulfillment option and fulfillment centers has a finite inventory capacity. But in our problem, we offered the optimal initial inventory levels for fulfillment centers that minimize the stock keeping costs and lost sales. Transshipment operation is not allowed, and the unfulfilled orders are charged with a high cost. Real shipment costs are used by considering the weights of the products, therefore more realistic solution is achieved.

2.3. Joint Pricing and Fulfillment Process

As already mentioned, pricing and order fulfillment processes are highly related with each other and in order to get an global optimal solution to the general problem, both pricing and fulfillment parts should be solved together. The main reason behind that is if a company changes the price of a product, total demand for that product also changes and all of the fulfillment operations might be affected by this price adjustment. In literature, these two problems are mostly solved separately. However, Lei *et al.* proposed an alternative model that solves the pricing and fullfilment problems jointly [15]. The proposed model determines the optimal prices and the fulfillment strategy that maximizes the total profit. They studied the behaviour of the customers with the change on the prices. The main reason behind is that the customers tend to choose best option considering both the service quality of the supplier and the price of the product. There were multiple fulfillment centers and the company could change the prices of the products in each period. However, the prices of a particular product could not differ in different cities/districts so there was always one price for a product on a particular time. The objective of the proposed model was to maximize the profit by considering the fulfillment and price constraints.

Similar to joint fulfillment model, airline pricing and revenue management systems try to charge customers with optimal prices to meet customer demands with their highest willingness to pay values. Otero *et al.* established a dynamic stochastic model to decide prices for the tickets in different time intervals [16]. Their main assumption was there is no strong competition on the market and all price changes affect the customers directly. They first created fare families and then calculated ticketing and buying probabilities of these families. The dynamic model offered a solution to maximize total revenue until the flight. The proposed solution is very similar to the model in this thesis. Both have a finite capacity to meet customers' demand and both tries to maximize revenue with pricing actions. There is a fundamental difference between these two problem. In airline industry, the tickets are perishable and cannot be used in next periods. However, our model is able to carry inventory to next period with a defined cost.

Another research area of pricing and inventory management problems is the Service & Inventory systems. This kind of problems describe the fulfillment process as a service system which serves the customers with the given service and arrival rate distributions. A well-known example for these researches [17] has been published by Marand *et al.*. In their research, they assumed that arrival rate of the customers is price dependent and that the service time, which is the fulfillment process in this problem, is exponentially distributed. Most of the cost terms of their model were similar to the other researches in this area: inventory holding cost, fulfillment cost and lost sales. However, there was another cost term which was called waiting cost. This cost was used to minimize the total waiting time of the customers on the service queue.

The system was modeled in a way that allowed the inventory to be replenished during the given time horizon. The traditional (r, Q) replenishment strategy was used with different parameter values and the performances were compared in a statistical way. The final model offered insights for the relationship between pricing and fulfillment operations with different strategies.

Most of the research about the pricing problems focused on the retail industry. However, pricing is a key concept for almost all supply chain operations. Ahn *et al.* built a pricing model jointly with a production problem [18]. The first model they proposed was deterministic and had demand for each period as a function of the price on that particular period. However, in the second model demand did not depend only on the price but also on other factors such as previous orders. For that reason, the model became non-concave and difficult to solve. They solved that problem by adding price ordering constraints called fixed-ordering constraints. After adding these constraints to the model, they offered an algorithm which could find the optimal solution for the given pricing and production problem. The production problem is different than the main problem in this thesis, but production process has similarities with the order fulfillment of an online retailer. In both of these papers, there are customers to be satisfied and revenue is made by performing fulfillment operations.

All research on these problems tried to come up with alternative approaches and to offer some applicable strategies. However, converting the statistical insights to a decision is another difficult research area in the literature. Most of the proposed solutions for the specific problems may not be perfectly applicable for real systems.However, Bertsimas *et al.* worked on creating good and applicable decisions using predictive methods of both machine learning and operations research literature [19]. Their study aimed to find a better use of auxiliary data to make standard predictions become better decisions. The main assumption was that taking only the mean of the sample is not an acceptable method on most cases in the real world. Therefore, they offered a metric named coefficient of prescriptiveness to get decisions from a sample data which has an unknown distribution. For different methods, their study offered different weighting formulations which make the model converge to an optimal decision. Their approach also helped to convert stochastic problems to a deterministic one by using the coefficient of prescriptiveness.

In this study, the fulfillment problem is modeled considering the pricing effect by establishing a forecasting method that uses both the time series property of the sales data and the effects of the external factors such as price etc. Unlike the other studies on this area, the pricing/fulfillment model is solved by using the coefficient of predictiveness which is mentioned in the Bertsimas and Kallus' study. As in their study, different input cases are created and evaluated using the dataset. The sales forecasts are added to the model directly and different prices are evaluated in different scenarios. Therefore, this coefficient add the stochasticity to the studied fulfillment problem. The proposed model offers the optimal initial inventory levels for the products by considering the fulfillment costs, the total sales revenue and the inventory keeping costs for all products.

Table 2.1 summarizes the literature review. Forecasting model specifications, external factor on their forecasting methods and used fulfillment/replenishment procedure stated on the summary table.

Table 2.1. Summary of Literature Review

Reference	**Forecasting Model**	**External Factors for Forecasting**	**Fulfillment and Replenishment Procedure**
Huang *et al.* [3]	Surver Research, Multiple Models	Price, Rebate, Advertising, Allocated Space, Lead Time, Quality	-
Kuo *et al.* [4]	Artificial Neural Network	Promotion, Price	-
Chinthalapati *et al.* [5]	Reinforcement Learning	Price, Competition	Replenishment
Ramanathan *et al.* [7]	Structural Equation Modeling	Promotion, Price	-
Ferreira *et al.* [8]	Hybrid Learning Algorithm, Time Series Methods	Promotion, Price, Substitution	Fulfillment
Xia *et al.* [9]	Bayesian Learning, Simulated Annealing	Price	Fulfillment
Fisher *et al.* [10]	Choice model, triple-difference estimator.	Competition, Stock-Out Factor, Price	-
Lin *et al.* [11]	-	-	Fulfillment
Acimovic *et al.* [12]	-	-	Fulfillment
Torabi *et al.* [13]	-	-	Fulfillment Transshipment between FCs
Ishfaq *et al.* [14]	Linear Price-Demand Model	Price	Fulfillment
Lei *et al.* [15]	Location Based, Linear Price-Demand Model	Price	Fulfillment
Otero *et al.* [16]	Dynamic Stochastic Model	Price, Cluster Effect(Fare Families)	Fulfillment
Marand *et al.* [17]	Service&Inventory System	Price	Fulfillment Inventory Holding
Ahn *et al.* [18]	Bi-level Optimization Model	Price, Last Order Effect	Production
Bertsimas *et al.* [19]	No specific model, offers approach	-	-
Our Study	Tree-Based Ensemble Model with Time-Series and Location-Based properties	Price, Promotion, Location Effect	Fulfillment Inventory Holding

3. BACKGROUND

3.1. Forecasting Methods

In this section, forecasting methods which are used and evaluated in this thesis are summarized.

3.1.1. Linear Regression

Linear regression methods offer a mathematical formulation that represents the relationship between predictors and output values with a linear function. This formulation process uses statistical methods to get best response values for the input dataset. The most basic application of linear regression method is forecasting the response by using only 1 predictor (Simple Linear Regression). However, most of the models in real life utilizes more than 1 predictor to estimate response value (Multi-Variable Linear Regression). The general formula for multi-variable linear regression model is the following:

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_i p + \epsilon_i \tag{3.1}$$

where Y_i is the outcome (or dependent variable), X_{ij} are independent variables, and ϵ_i is the error term.

Linear models generally fitted by minimizing the sum of squared residuals. However, different scoring methods are available such as Lasso (L-1 Norm) and Ridge (L-2 Norm) regressions. In this thesis, least squares approach is used to fit forecast models.

Figure 3.1 exhibits the fitted values for a simple linear regression on a two-dimensional graph where the straight line stands for fitted values and the dots indicate real data points.

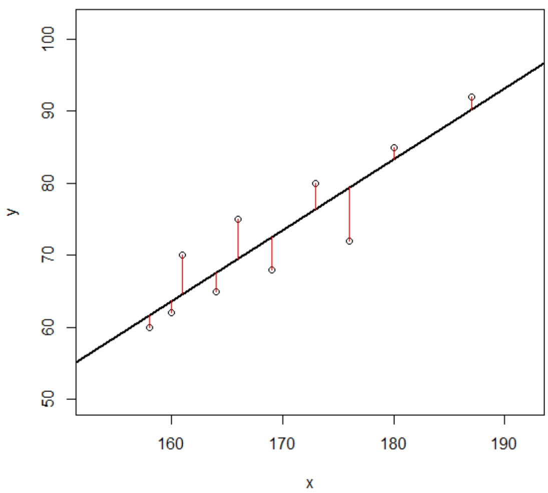

Figure 3.1. Linear Regression example in a 2-dimensinal graph

3.1.2. Log-Linear Regression

Log-linear methods aim to find a linear relationship between input and response variables in a logarithmic scale. For some problems (elasticity estimation, marginal effect formulation etc.), log-linear models provide better prediction models according to the determined score functions.

	X	
Y	X	$\log X$
Y	*linear* $\hat{Y}_i = \alpha + \beta X_i$	*linear-log* $\hat{Y}_i = \alpha + \beta \log X_i$
$\log Y$	*log-linear* $\log \hat{Y}_i = \alpha + \beta X_i$	*log-log* $\log \hat{Y}_i = \alpha + \beta \log X_i$

Figure 3.2. Logarithmic transformation models

To establish a log-linear regression model, transformation process is a must. Therefore, dependent and independent variables which will be transformed to the log scale should be positive. Figure 3.2 summarizes 4 different transformation approaches for log-linear regression models.

3.1.3. Random Forest

Random forest method (RF) is an advance application of simple decision trees. RF is an ensemble method that uses bagging procedure while creating new decision trees. For each split, random number of estimators is chosen and based on the performance metric, a new split is performed by using these features [20]. These random feature selection prevents the model from overfitting.

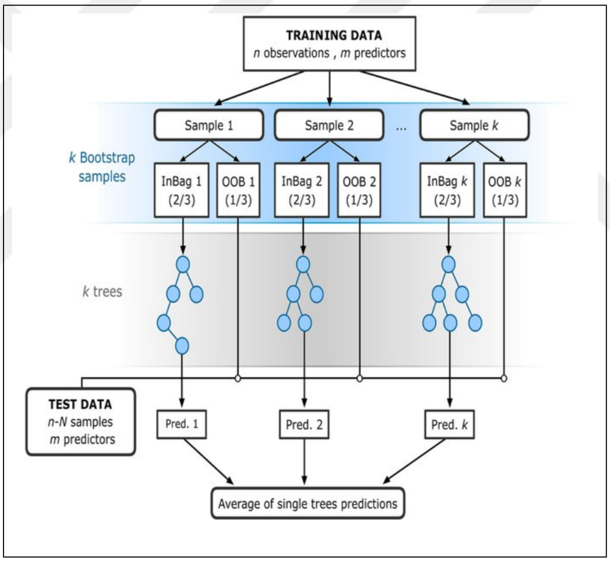

Figure 3.3. Random Forest Regression Flow-Chart

After generating a specific number of trees (predictors), a voting procedure is performed to get final decision for classification problems. The vote of the majority is chosen as the final result. For the regression problems, two possible result generation approaches are used: mean of the predictions, constructing a new model using the predictions. Figure 3.3 summarizes the prediction process of random forest models.

3.1.4. Extremely Randomized Trees

Extremely Randomized Trees (ExtRa-Trees) method is very similar to the Random Forest method in most ways. There is a major difference on the split selection procedure. Random forest algorithm selects k features and performs best split according to the information gain metric (Gini, entropy etc.). However, in Extra-Trees algorithm, splits are performed randomly which may reduce the variance on some cases [21].

Both Random Forest and ExtraTrees regressors have capability to generate more than one response values using a single input feature vector. By using single vector for multiple outputs, joint effects of the features can be evaluated in a simple and effective way.

Figure 3.4 visualizes multi-response regression models in general.

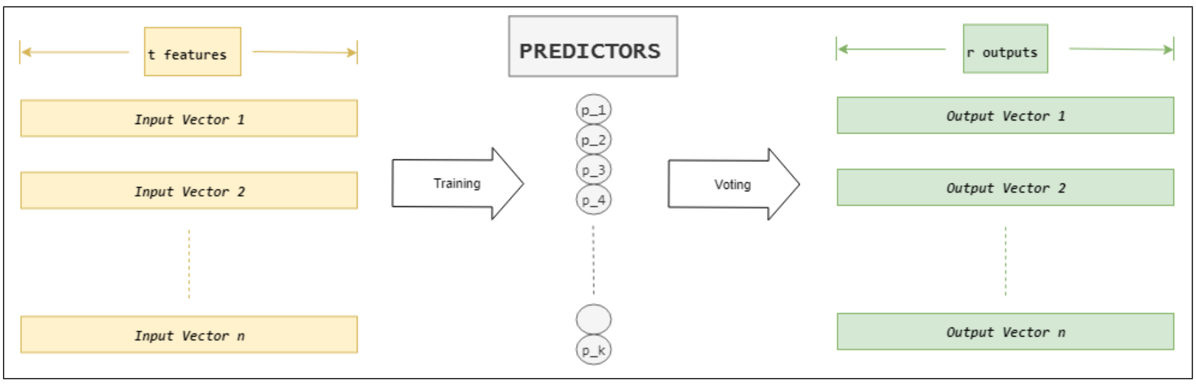

Figure 3.4. Multi Response Regression Models

3.2. Performance Metrics for Forecasting Methods

During the evaluation process for forecasting methods, some performance metrics are used. The most popular ones are examined in this section.

3.2.1. R-Square

R-Squared metric is used to represent the strength of the relationship between input and output vectors and forecast quality. Statistically, R-Squared value shows how much variation can be explained by the input features.

$$R^2 = 1 - \frac{SSE}{SST} \tag{3.2}$$

Equation 3.1 formulates the R-Squared value in a mathematical way. SSE stands for sum of squared errors where SST is the sum of all squared values. Therefore, as the errors which cannot be explained by the model decrease, the quality of the regression model increases.

3.2.2. Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) is a metric that represents the average absolute accuracy of the fitted values. Unlike the R2 formulation, MAPE penalizes deviations with absolute values. Equation 3.3 shows the mathematical formula for MAPE metric.

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\frac{|A_k - F_k|}{A_k} \tag{3.3}$$

MAPE formulation gives same weight for each data point. Therefore, a very small value with a respectively biased forecast, may lead a very high loss in MAPE metric. It makes models not only focus on the large valued data points but tries to maximize accuracy on each data points.

4. PROPOSED APPROACH

4.1. Forecasting Model

To get a more applicable and consistent solution for the general problem, a data-driven forecasting method is established. Because of the limitations of traditional forecasting methods, machine learning based approaches are used to establish a more accurate model. By using these methods, external factors such as promotions are added into the model in an easy and effective way. Different forecasting methods are tested on the data-set and one of them is chosen as the main method. The performance criteria, the sampling methodology, the cross validation algorithm and the feature selection procedures which are used in this study are described in the following sections.

4.1.1. The Data Set

We used sales and shipment data of an e-retailer company which mostly sells computer parts and technological devices. Table 4.1 outlines the sales dataset in general.

Table 4.1. Dataset Details

# of Products	17
# of Shipped-From States	22
# of Shipped-To States	52
Min. Sales Date	2012-07-01
Max. Sales Date	2012-12-31

There are 17 distinct products used in the forecasting model. All of them are computer hardwares which mostly directly imported from abroad. Therefore there is a respectively long lead time for the purchasing process. Table 4.2 summarizes the sales and shipment data for all products. These products are highly selling products throughout the period over all locations sampled from hundreds of SKUs.

Table 4.2. Summary for Product Data

No	ProductID	Item Description	Total Sales	Weighted Avg. Unit Price	Max. Unit Price	Min. Unit Price
1	33	CPU INTEL—CORE I5 2500K 3.3G 6M R	211782	218.87	219.99	197.99
2	179	CPU INTEL—CORE I7 3770K 3.5G 8M R	496651	327.03	352.49	290.16
3	187	CASE ROSEWILL—CHALLENGER RT	280742	47.71	59.99	39.99
4	219	CPU INTEL—CORE I5 3570K 3.4G 6M R	741950	225.64	233.32	175.99
5	227	CPU COOLER CM— RR-212E-20PK-R2 R	316468	33.27	37.99	29.99
6	287	CASE ANTEC—THREE HUNDRED BK RT	343014	53.3	69.99	38.99
7	305	BLU-RAY BURNER LG— WH14NS40 %	209288	64.14	79.99	54.99
8	322	SSD 128G—CRUCIAL CT128M4SSD2 R	205900	110.37	160.71	89.99
9	416	CPU AMD—4-CORE FX-4100 3.6G 8M R	222524	107.13	109.99	79.99
10	496	CPU COOLER CM— RR-B10-212P-G1 RT	335509	28.12	34.99	19.99
11	528	AMD GIFT DIRT3 GAME COUPON	226113	49.99	49.99	49.99
12	570	CPU AMD—PH II X4 965 3.4G AM3 R	256449	104.26	119.99	84.99
13	600	ACC HDD BYTECC— BRACKET-35225 RT	241470	5.04	6.89	1.99
14	1008	CPU INTEL—CORE I3 2100 3.1G 3M R	221203	119.74	119.99	103.99
15	1049	CPU INTEL—CORE I3 2120 3.3G 3M R	150536	123.24	124.99	104.49
16	2932	KB MICROSOFT — KEYBOARD 200 R	106027	9.66	11.99	4.99
17	3785	CASE ROSEWILL—R363-M-BK RETAIL	133476	51.02	54.99	49.99

The product summary Table 4.2 shows that, products used in this study have high prices so fulfillment costs have lower effect on the fulfillment process. The shipment costs of the products are calculated using real shipment data. As a percentage of total sales, the cost ratio changes between 1% to 24% for different SKUs. It highly depends on volume and sales price of a product. Products like computer case, have high volume and low sales prices so the shipment cost rate is respectively high. However, SSDs and CPU units are the product with lower ratios because of the higher price/volume rate.

The company keeps their sales data with transactional details, but we aggregated the data to daily level. The other dimension of the sales data is the location. All of the realized sales have ship-to information with zip-code detail on the data set. The sales are aggregated to the corresponding states. After the aggregation process the columns of the base data used for the forecasting model are listed in Table 4.3.

Table 4.3. Details of sales data

Data Field	Data Type	Description
ShippingDate	date	Date of the shipment
ProductID	int	Unique ID for Products
StateID	int	Unique ID for States
TotalUnitsShipped	int	Total Realized Sale Quantity

Table 4.3. Details of sales data (cont.)

Data Field	Data Type	Feature Description
AverageUnitPrice	float	Weighted Average of the Sales Price

The sales transaction data has geographical information of both sender and receiver. The sales are generally cumulated on specific states which can be observed on Figure 4.1. Even the company only sells technological devices, the location information of the buyers are very effective on the purchasing behaviour. It does not need to be a geographical effect but socioeconomic levels of the customers highly related with the state/province they live in. Therefore, location based model is used on the proposed forecasting model in this thesis.

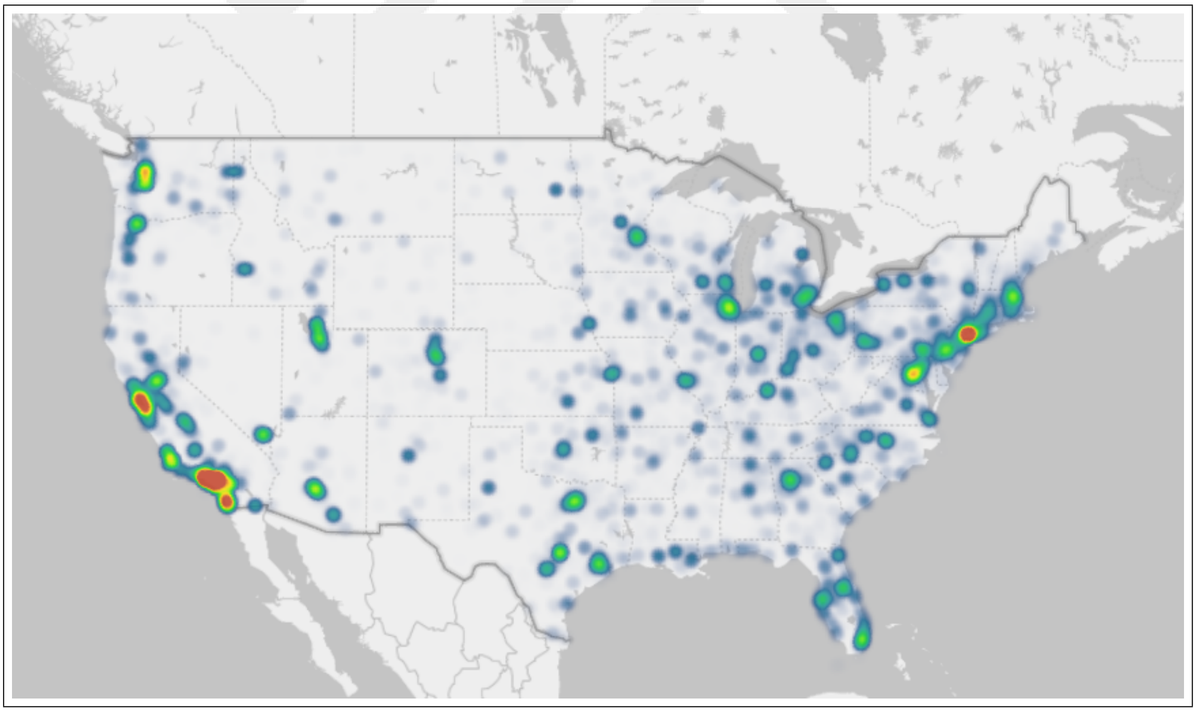

Figure 4.1. Heat-Map for Sales Data by Ship-To Location

4.1.2. Feature Selection and Generation

Before fitting the model, features which are used in the forecasting method are selected. Because of the nature of the retail industry, the time series effect should be kept in the model. Therefore, sales performance of the products in the last week are added to the input data set as features. This period is kept short because of the pricing process is highly correlated with the sales frequency of the product in last periods.

The other critical issue in this problem is the non-numeric features such as demand zones. 52 locations are used in the model which correspond to the 52 states of the USA in the sales data. In order to use these data for forecasting, 52 different binary variables should be added to the model which might cause a lack of robustness for the general solution. To get rid of this problem, city sales are used as the key indicator for each state. However, sales performances of the locations are different for each product. Therefore, sales performances of the products in a state is added as a feature to increase the accuracy of the location-based forecasting model (Table 4.4). Sales performance factor is added by using two different features :

(i) Total sales across all periods and all products in a state.
(ii) Total sales of specific product across all periods in a state.

Table 4.4. Forecast Input Features

Feature Number	**Feature Name**	**Description**
1	SalesLW	Sales Quantity, Last Week
2	AverageUnitPrice	Average Sales Price of the Product in that Period
3	BlackFridayFlag	Black Friday Indicator
4	stateSales	Total Sales Quantity for All Products in that State
5	stateProductSales	Total Sales Quantity of the Product in that State

The price factor is added to the model directly for each product. Realized sales might show different prices for different transactions because of the intraday price changes. At this point, average sales price is added to the model for each time stamp to get a single price for each data instance.

The special days and the other events have a significant importance on the sales patterns of the products. However, adding all of the special events into the model would cause overfitting and the model would not give accurate forecasts for the whole data. Therefore, only the Black Friday effect is added to the model as a binary feature. Figure 4.2 shows the total revenue of all products in the dataset. The effect of the Black Friday event can be easily observed in the Figure 4.2 which shows the total revenue ($) of the company during the time horizon.

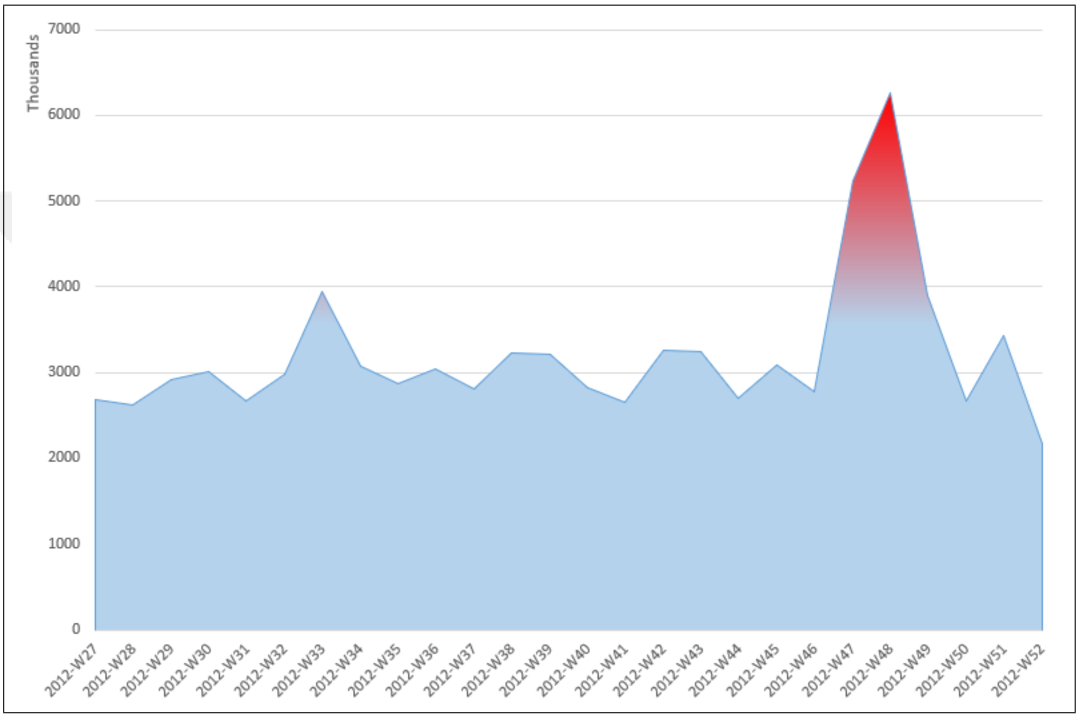

Figure 4.2. Effect of the Black Friday promotion. (Revenue - $)

4.1.3. Method Comparison

For the main forecasting model, different methods are evaluated based on the performances on the data-set. The performance metric is the weighted average of R-Square values with different weights for each product. The weights are calculated using the corresponding total revenues of the products.

The training and test data are sampled randomly with 40% of the data-set being used for training and 60% for testing.

Table 4.5. Model Performances

Method	R-Square Value	R-Square Value (Weighted)
Multi-Variable Linear Regression	0.40	0.47
Multi-Variable (1+Log)-Linear Regression	0.42	0.50
RandomForest Regression	0.58	0.67
Extra-Trees Regression	0.79	0.81

As shown on Table 4.5, Extra-Trees regressor outperformed other methods and therefore it has been chosen as the main forecasting method.

4.1.4. Sampling and Cross Validation

Two different test dataset selection methods are used in this study. The first method divides the data into two groups by considering the time series characteristics of the data. For each product-location pair, the time horizon is split into two parts and the data are categorized using the corresponding part of the whole data. The other method is random selection by using only a split ratio. The data are categorized as test and train set in a perfectly random way. Evaluation tests are performed with the Extra-Tree regressor. The second method provided about 30% higher weighted R-Square value and therefore it is used as main sampling method.

After selecting the forecasting and sampling methods, the model parameters are determined with a cross-validation process. 5 parameters for Extra-Trees Regressor are examined with different values.

Table 4.6. Cross Validation Parameters

Number	Parameter Name	Values
1	# of estimators	50, 100, 250
2	Max # of features	2, 3, 4
3	Max Depth	1, 4, 9, 35, 50
4	Bootstrap	True, False

Same parameter values are not used for each product but a cross validation is performed instead. Therefore, for each product we got different models with different parameter values.

The forecasts generated by the proposed model are used in the fulfillment process which will be covered in the following sections. The special day effects, sales price effect, time series characteristics of the products and location sales performances are used in the proposed forecasting model.

4.2. Fulfillment Model

4.2.1. Overview

Fulfillment problems take customer orders as inputs and allocate available resources to meet customers' expactations. The main problem in this thesis is solving the fulfillment and pricing problems together. The main reason behind that is the pricing process and fulfillment operations are highly correlated and should be assumed as a single process. In this section, optimization model for fulfillment problem is discussed.

In this thesis, different settings for initial inventory levels are examined and based on different settings model outputs analyzed to get an optimal strategy. The company is trying to allocate its inventory resource to the fulfillment centers in a way that total profit is maximized.

4.2.2. Model Assumptions

Some necessary assumptions are made to simplify the general problem and to get more robust results. These assumption are related to both modeling and general fulfillment procedure.

(i) Orders can be fulfilled by any fulfillment center that shipped any product in past.
(ii) All orders have only one distinct product and delivered independently of the other

orders.

(iii) Order quantities are assumed to be continuous, not integer.

(iv) Fulfillment costs are generated by the shipment-costs in the actual data-set.

(v) The only operational cost during the fulfillment process is the shipment cost.

(vi) Inventory of a product can be carried by FCs during the time horizon.

(vii) Inventory distribution cost is charged with a cost multiplied with initial inventory level.

(viii) Inventory distribution cost is assumed to be linear to the weight of a product.

(ix) There is no inventory holding cost in the model. Remaining stocks are assumed to be sold in the following periods.

(x) There is no delay between order time and delivery time.

(xi) The only restriction for an FC is its available on-hand inventory level.

(xii) An unfulfilled order assumed to be lost.

(xiii) Inventories cannot be replenished during the optimization horizon.

(xiv) Transshipment between FCs is not allowed.

4.2.3. The Mathematical Model

Two different modeling approaches will be discussed in this section. Firstly the main model will be described where all Demand-Price pairs have similar likelihood to happen. Then, coefficient of prescriptiveness will be used to add the effect of stochasticity to the model. The first model is a stochastic problem with random forecast values. Then the model is turned into an LP and final model is utilized to optimize total profit of the company.

In the mentioned problem there are K products to be fulfilled and J locations where products can be shipped to. In Section 4.1, generating forecasts for these K products in J locations is described. Based on these forecasts, the company can fulfill the incoming orders via I - 1 fulfillment centers during the optimization horizon, T. One FC is used to handle lost sale procedure by assigning a respectively high fulfillment cost on it.

Demands of Product-FC pairs are defined in the model as deterministic parameters, D, with three dimensions: Period t, product k and FC i. A fulfillment matrix, Z, is used to represent the Order-FC mappings and corresponding order quantities.

Constraint 4.1 exhibits the constraint that represents the relationship between D and Z.

$$\sum_{i=0}^{I} Z_{ijk}^{t} = D_{jk}^{t} \qquad \forall j \in J, \qquad \forall k \in K, \qquad \forall t \in T \tag{4.1}$$

As described, the model can carry inventory throughout the time horizon and replenishment is not permitted. Therefore, sum of initial inventories of the FCs are added to the model as hard constraints. These inventory levels are represented with C.

The dummy FC, which is used for lost orders, has ample capacity to get rid of infeasibility. Constraint 4.2 shows the FC capacity constraint.

$$\sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{j=1}^{J} Z_{ijk}^{t} \leq \sum_{i=1}^{I} C_{ik} \qquad \forall k \in K \tag{4.2}$$

The objective function maximizes total profit which consists of three main parts: Total inventory cost, total revenue of the sales and corresponding fulfillment costs.

Inventory distribution cost (4.3) is not defined as time phased but it is used in a way where the model is forced to minimize initial inventory levels. Last FC is ignored while calculating the total inventory distribution cost.

$$\sum_{i=1}^{I-1} \sum_{k=1}^{K} s_{ik} C_{ik} \tag{4.3}$$

The revenue part of the objective function (4.4) is added to the model as sum of the multiplication of corresponding prices and demands of the products. In the optimization part, there is no pricing effect on the demands. That effect is added to the model in the forecasting phase. Lost sale correction is handled by subtracting the lost sales from original demand forecast D.

$$\textstyle\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J} p_k^t(D_{jk}^t - Z_{Ijk}^t) \tag{4.4}$$

Different FCs can meet customer demands with different costs. These costs are product specific and do not differ in time. Fulfillment cost in the objective function (4.5) handles the trade-off between fulfilling an order with high cost versus keeping high level of inventory at the beginning of the time horizon.

$$\textstyle\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} c_{ijk}Z_{ijk}^t \tag{4.5}$$

The generated optimization model is a 2-stage stochastic program where revenue and distribution costs are optimized to increase total profit with a random demand variable.

$$\min_{C_{ik}} \sum_{i=1}^{I-1}\sum_{k=1}^{K} s_{ik}C_{ik} + \max_{Z_{ijk}^t} \mathbf{E}\left[\sum_{j=1}^{J} p_k^t(D_{jk}^t - Z_{Ijk}^t) - \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} c_{ijk}Z_{ijk}^t\right]$$

st.

$$\textstyle\sum_{i=0}^{I} Z_{ijk}^t = D_{jk}^t$$
$$\forall j \in J, \forall k \in K, \forall t \in T$$

$$\textstyle\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{J} Z_{ijk}^t \leq \sum_{i=1}^{I} C_{ik}$$
$$\forall k \in K$$

$$\tag{4.6}$$

At the first stage, C stock is distributed to FCs with a cost of s. At the second stage, the model tries to maximize total expected profit by using corresponding demand forecasts and prices.

To use an LP rather than two stage SP, scenario based analysis are performed. The LP equivalent of the two stage SP is stated below where r index stands for the corresponding scenario(4.7).

$$\begin{array}{ll} \textbf{min.} & \sum_{i=1}^{I-1}\sum_{k=1}^{K} s_{ik}C_{ik} \\ - & \sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J} p_{rk}^{t}(D_{rjk}^{t} - Z_{rIjk}^{t})\pi_{rjk}^{t} \\ + & \sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} c_{ijk}Z_{rijk}^{t}\pi_{rjk}^{t} \\ & \\ \textbf{st.} & \sum_{i=0}^{I} Z_{rijk}^{t} = D_{rjk}^{t} \\ & \forall r \in R, \forall j \in J, \forall k \in K, \forall t \in T \\ & \\ & \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{J} Z_{rijk}^{t} \leq \sum_{i=1}^{I} C_{ik} \\ & \forall r \in R, \forall k \in K \end{array} \tag{4.7}$$

where π stands for corresponding probability of the specific scenario.

4.2.4. Coefficient of Prescriptiveness

The mathematical model described in the previous section has a stochasticity bacause of the generated demand forecasts. These demands can be described as random variables but in the optimization part, mean demands are used as input variables. The forecasting model generates demand forecasts with different robustness along with their probabilities for each scenarios, because some cases are not seen in the past data so hard to predict. To determine optimal initial inventory levels for FCs, the *robustness* *effect* of the forecasts is added to model with the coefficient of prescriptiveness.

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\{\text{no. of instances in node k, tree t}\}}{\text{Total no. of instances in tree t}} \tag{4.8}$$

Formula 4.8 exhibits the prescriptiveness calculation for the generated forecast data. The tree-based ensemble model sends instances to different nodes and their reliability in a stochastic environment can be measured by its neighbour instances which share same node with the corresponding data point. For each tree, this ratio is calculated and their mean value is used as final ratio for the forecast model.

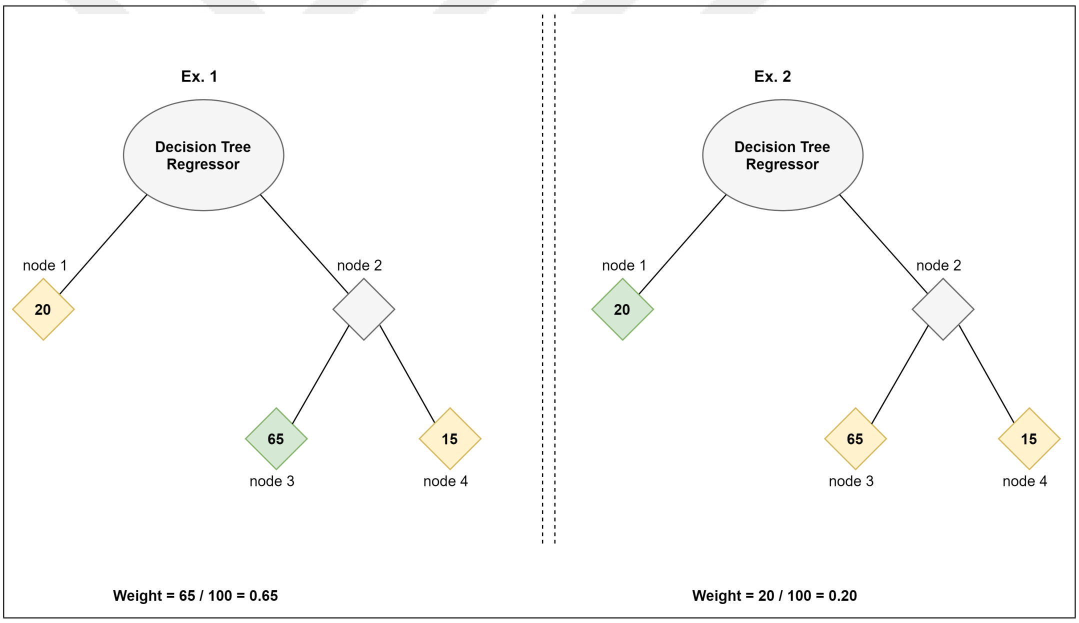

Figure 4.3. An example of weight calculation for 1 Decision Tree

In Figure 4.3, weight calculation using only 1 decision tree for 1 output value is visualized. The tree is trained using 100 data points. The number inside a leaf node shows the number of training data points in corresponding node. In the first example, that 1 output value is calculated using green node which has 65 training points. Simply, the weight can be calculated as $65/100 = 0.65$. Similarly, in the second example this ratio appears as $20/100 = 0.20$. For multiple forecast results, these ratios are summed up. After taking these sums for each decision trees, final ratio for this output set is calculated by simply taking the average of them.

4.2.5. Input Setting Creation

Different input cases are created to analyze the main fulfillment problem. Two main inputs are used to create these cases: price level, capacity level.

In the fulfillment part, price level has no direct effect on the optimization results because prices are added to the model as parameters. However, price level changes the forecasted quantity so it changes the objective values in that way. To analyze these effects, three levels for prices are used: Lower, Mid, Upper.

These price levels are generated for each product using real sales data. Upper, mid and lower limits are determined by using the average sales prices of products in 5%, 50% and 95% quantiles of cumulative sales ordering by price. Table 4.7 shows price levels for each product (Prices are rounded to the nearest integer value.).

Table 4.7. Price Levels of Products

ProductID	High	Mid	Low
1008	120	120	118
1049	125	123	106
219	230	226	215
287	60	53	40
33	220	219	210
2932	12	10	8
322	130	110	100
3785	55	51	50
416	110	107	82
496	30	28	20
528	50	50	50
570	110	104	87
179	340	327	300
187	50	48	40
227	36	33	30
305	70	64	55
600	6	5	5

In this thesis, a tree-based ensemble forecasting model is used, therefore extreme values, which are not available in the training data, are hard to use. As seen in Table 4.7, some products have same value for different price levels. It is because these products have no data for different prices so same values are used in different price levels.

The other scenario input is the capacity level of a product which is initial inventory level in the stated model. Again 3 levels are used to create different settings for capacity values. These levels are determined by using the sales data of last two months for each product.

(i) *Low capacity level:* Daily average sales * 3.5
(ii) *Medium capacity level:* Daily average sales * 7
(iii) *High capacity level:* Daily average sales * 20

Initial purchase quantity is not an operational decision for the company. It needs large amount of budget because of the high values of the products in sale. Therefore, increasing the capacity at its highest level may be very costly.

The medium capacity level for the company is the most common situation for its normal business. Therefore, mid capacity case is assumed to be the baseline for the optimization results. High and low capacities are discussed by considering the optimal results in medium inventory capacity level.

After creating price and capacity levels, 9 different settings are obtained for each product to analyze. Results for these cases are discussed in the Section 5.

4.3. Solution Methodology and Technology

For forecasting part, Scikit-Learn library is used with python interface [22]. Forecast generation process has about 133.4 seconds runtime for 17 products on average. Cross validation, data preparation and forecast generation times are included in this duration.

The optimization problem is solved for 17 products. To create optimal results for the fulfillment problem, multiple optimization runs are performed for differnet input sets. Gurobi optimization library(version 8.00) is used in Python interface to perform these optimization runs [23]. The problem is solved optimally for each input instance. Model creation and solving time differs by the input data. For 17 products, 1 price level, and 1 capacity level has a solution time about 72.3 seconds in average.

5. RESULTS

For different input set combinations, results are evaluated with different metrics: Lost sales, total profit, weighted total profit.

The stated forecasted model creates fitted values with a random effect. Therefore, to make the results converge to a stable result 10 independent runs are used for each combination of price and inventory levels. The results shown in this section shows simply the average of the total costs and profits.

During this optimization process, all sales data are used and following 7 days are optimized with the stated fulfillment model. Therefore, the sales performances of both products and states are highly effected by the sales in the last periods.

5.1. Price Effects on High Inventory Levels

In real data, most of the products are sensitive to the price change. Therefore high price makes the customers buy less but yields a higher margin. Therefore there is a natural equilibrium between sales volume and price levels.

In high capacity case, almost all demands can be fulfilled by the FCs. Therefore, the objective value is expected to be higher respectively.

Table 5.1 shows average total profits of products in high capacity level.

Table 5.1. Total Profit in $, High Capacity - Different Price Levels

ProductID	Low Price	Mid Price	High Price
33	358,351	352,548	319,213
179	1,719,049	855,808	898,217
187	6,729	26,073	22,664
219	973,666	944,575	966,985
227	59,716	50,922	51,590
287	24,140	25,522	37,478

Table 5.1. Total Profit in $, High Capacity - Different Price Levels (cont.)

ProductID	Low Price	Mid Price	High Price
305	123,741	121,401	84,133
322	183,237	147,133	208,866
416	220,407	231,485	197,530
496	24,937	52,800	43,376
528	82,568	82,432	82,565
570	145,747	174,065	167,479
600	6,665	6,688	5,351
1008	282,567	214,890	215,325
1049	184,537	208,444	185,126
2932	4,875	7,262	9,564
3785	39,146	46,716	53,813

In this case, inventory level is not a binding constraint for most cases. Therefore, as seen in Table 5.1, most of the products have higher profits on lower price levels because of respectively high demands on low prices. However some products have shortage for all price levels even in high capacity situation(Table 5.2). At this point, higher prices can be charged or more inventories can be carried at the starting of the planning horizon.

Table 5.2. Lost sales in $, High Capacity - Different Price Levels

Products	Low Price	Mid Price	High Price
33	-	-	-
179	-	-	-
187	-	-	-
219	-	-	-
227	-	-	-
287	-	-	-
305	61	-	-
322	-	-	-
416	32,362	-	-
496	-	-	-
528	-	-	-
570	-	-	-
600	-	-	-
1008	-	-	-

Table 5.2. Lost sales in $, High Capacity - Different Price Levels (cont.)

Products	Low Price	Mid Price	High Price
1049	800	-	-
2932	6,582	4,276	2,688
3785	278	8,725	11,769

On the other hand, 12 out of 17 products are not stock-out even in low price levels. It means, the inventory levels can be reduced to get lower initial investment costs.

5.2. Price Effects on Mid Inventory Levels

Most of the time, companies keep their inventories with a stable cover level. Therefore, medium level capacity is the most common case for the company. The results of optimization model for different price levels on medium capacity case is reported in Table 5.3.

Table 5.3. Total Profit in $, Medium Capacity - Different Price Levels

Products	Low Price	Mid Price	High Price
33	189,833	198,062	198,946
179	638,507	696,008	723,888
187	22,940	32,641	34,683
219	685,739	720,920	733,736
227	37,416	41,419	45,488
287	24,927	42,286	52,679
305	48,022	56,123	61,419
322	88,712	97,598	115,395
416	77,347	101,352	104,228
496	23,300	35,066	37,791
528	48,788	48,788	48,788
570	94,880	113,723	120,357
600	5,009	5,009	5,610
1008	110,431	112,300	112,299
1049	67,363	78,405	79,695
2932	1,865	2,747	3,649
3785	17,678	18,567	20,949

As seen in Table 5.3, optimal results for the fulfillment problem force the company to sell all of its products if there is enough demand. Therefore, low price strategy is not a good choice for the pricing process. Table 5.4 shows how the lost sale increases as the price level decreases.

Table 5.4. Lost sales in $, Medium Capacity - Different Price Levels

Products	**Low Price**	**Mid Price**	**High Price**
33	174,440	159,585	125,083
179	1,090,360	166,353	184,706
187	30,553	41,558	34,235
219	294,768	228,995	239,663
227	29,027	15,625	12,030
287	73,299	49,481	50,087
305	79,568	68,509	25,649
322	95,457	49,270	94,217
416	178,763	134,648	96,818
496	10,989	27,988	15,008
528	33,816	33,815	33,688
570	53,120	62,712	49,595
600	2,036	2,006	3
1008	178,060	108,021	107,978
1049	121,475	133,582	108,994
2932	13,291	12,820	12,865
3785	49,857	63,185	70,676

5.3. Price Effects on Low Inventory Levels

The most restricted settings in the fulfillment model is the low inventory capacity level situations. In that case, almost all of the inventories are over after the fulfillment process. Therefore, increasing the price always affects the total profit in a positive way. This effect can easily seen at Table 5.5.

Table 5.5. Total Profit in $, Low Capacity - Different Price Levels (Cont.)

Products	**Low Price**	**Mid Price**	**High Price**
33	94,995	99,114	99,561
179	319,379	348,233	362,156
187	12,208	17,124	18,251
219	342,989	360,587	366,992

Table 5.5. Total Profit in $, Low Capacity - Different Price Levels (cont.)

Products	Low Price	Mid Price	High Price
227	18,818	20,841	22,892
287	13,348	22,472	27,522
305	24,054	28,108	30,791
322	44,363	48,808	57,706
416	38,703	50,716	52,156
496	11,857	17,714	19,102
528	24,396	24,396	24,396
570	47,494	56,913	60,231
600	2,512	2,512	3,031
1008	55,271	56,214	56,214
1049	33,719	39,240	39,887
2932	966	1,415	1,867
3785	9,196	9,551	10,745

Because of the tight constraint on the inventory levels, lost sales are respectively high in these low capacity cases (Table 5.6). Even highest price levels cannot preserve the company to get shortage in its stocks.

Table 5.6. Lost sales in $, Low Capacity - Different Price Levels

Products	Low Price	Mid Price	High Price
33	269,531	259,708	225,542
179	1,424,306	518,657	548,036
187	55,061	70,136	64,382
219	636,053	593,185	606,943
227	49,709	38,298	36,579
287	103,129	89,514	94,071
305	103,701	98,256	57,338
322	139,117	98,554	152,105
416	216,811	185,792	149,076
496	25,438	48,715	36,699
528	58,180	58,127	58,233
570	101,057	120,078	110,417
600	4,618	4,613	2,727
1008	235,754	164,724	165,667
1049	155,909	173,606	149,749
2932	15,166	15,094	15,590
3785	64,650	78,148	86,502

5.4. Profit Comparison Between Different Input Cases

The price effects show, almost all products are price sensitive and corresponding profits are highly related with the price level. Similarly, fulfillment and inventory distribution costs are also affected by these price changes. Figure 5.1 compares different cases with the corresponding profits and lost sales. A clear relation between capacity, price and lost sales can be easily seen in Figure 5.1.

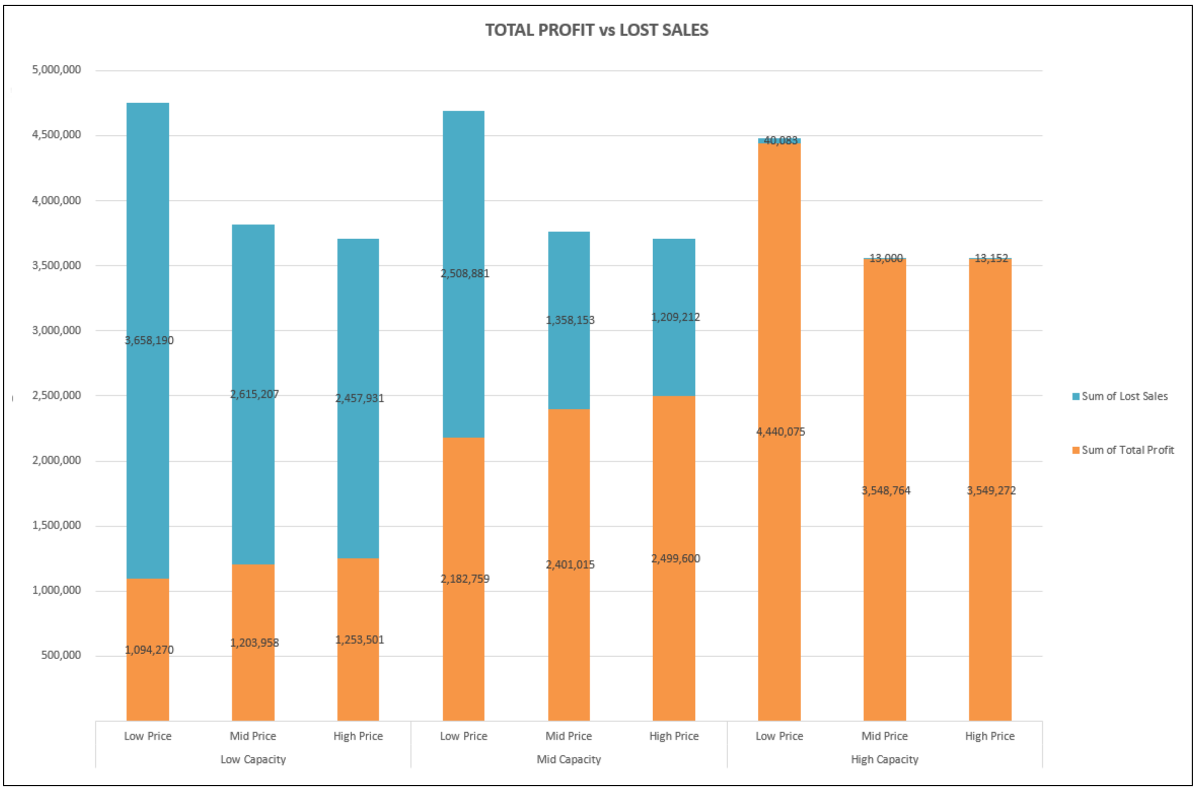

Figure 5.1. Total Profits & Lost Sales for Different Input Sets

In low and mid capacity levels, higher prices create higher profit. However as the capacity level exceeds the maximum demand then, low price level creates more profit for the company.

The differences between low-mid and mid-high price levels on different capacity cases are compared using one-tailed t-tests. Table 5.10, shows the p-values for the null Hypotheses . The critical value for the test is determined as 0.05 so all of the results show that, different price levels create statistically different profits.

Table 5.7. p-values for Null Hypotheses

	Low-Mid Prices	Mid-High Prices
Low Capacity	1.69327E-30	9.15842E-24
Mid Capacity	3.01241E-29	1.31749E-22
High Capacity	4.84541E-14	0.000108587

These results do not include the prescriptiveness level of the forecasting model. To add this effect into the optimization results, weighted results are used. Table 5.8 shows the weighted profits for high capacity case. These weights are calculated as discussed in Section 4.2.4 and reported in Appendix B.

Table 5.8. Weighted Total Profits, High Capacity

Products	Low Price	Mid Price	High Price
33	61,883	108,606	97,798
179	199,620	89,732	94,771
187	887	3,394	4,261
219	87,528	107,016	161,391
227	3,904	5,639	4,803
287	2,575	3,774	6,407
305	22,425	21,379	15,428
322	19,010	27,093	25,635
416	27,358	43,360	45,868
496	3,030	7,397	8,633
528	21,222	20,612	23,641
570	28,400	40,079	46,602
600	1,054	1,022	808
1008	83,525	77,414	82,459
1049	35,245	54,774	52,430
2932	3,112	4,915	5,746
3785	21,633	25,785	26,087

Results in Table 5.8 suggests that, as the price level increases, model creates more accurate forecasts. The main reason behind that is, at high price level, tree-base ensemble model can find more robust results. Non-weighted results(Table 5.1) declares that low price would increase corresponding demand, and also the total profit. However, in a stochastic environment this is not a reliable result without any confidence coefficient.

As mentioned, sampling ratio is highly correlated with the confidence level of the forecasting model. Therefore, as the train set on the forecasting procedure increases, the robustness of the results increases. This fact forces the optimization results to converge a stable value as the train data ratio gets closer to the 100%.

Besides, Figure 5.2 shows that the gap between original and weighted profits gets higher as the test sample ratio gets higher. Therefore, using only weighted profits is not an accurate process.

The weighted results should be utilized for each sampling ratio separately because, for larger test ratios, forecast results tend to get closer to each other and it yields a high weight. It is because of the proposed weight calculation in Equation 4.8. The results are reported for the 20% test and 80% training case. And all weighted profits are compared using the results with these ratios.

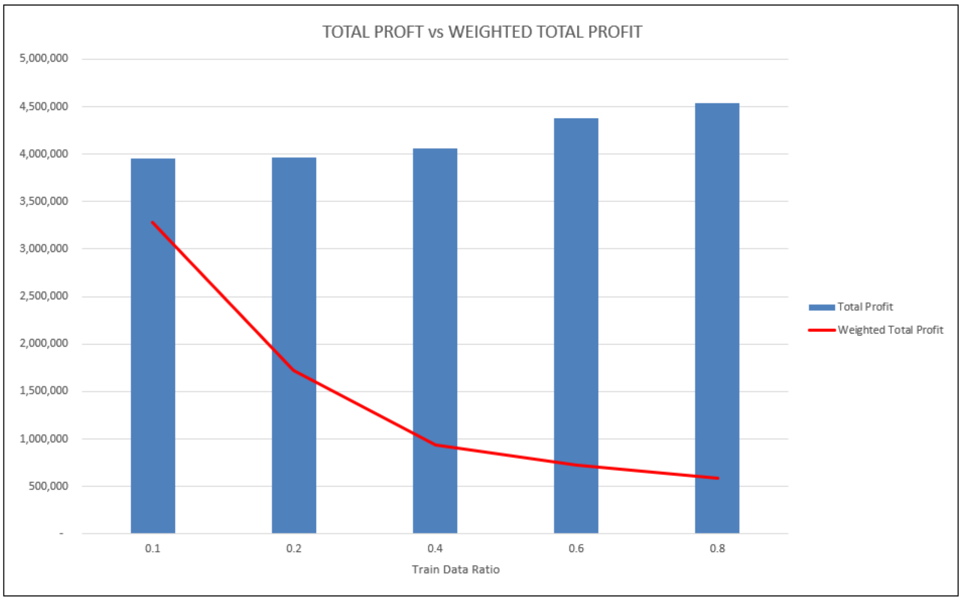

Figure 5.2. Total Profit vs Weighted Total Profit (High Cap., Low Price level) for Different Train Data Ratios

Similarly, medium and low capacity cases are evaluated and optimal pricing strategies for different capacity levels are determined. Table 5.9 clearly shows that as the inventory constraint becomes binding, the prices tend to be higher.

Table 5.9. Optimal Pricing Strategies for Weighted Profit

Products	**Low Capacity**	**Mid Capacity**	**High Capacity**
33	High Price	High Price	Mid Price
179	Mid Price	High Price	Low Price
187	High Price	High Price	High Price
219	High Price	High Price	High Price
227	High Price	High Price	Mid Price
287	High Price	High Price	High Price
305	High Price	High Price	Low Price
322	Mid Price	Mid Price	Mid Price
416	High Price	High Price	High Price
496	High Price	High Price	High Price
528	Low Price	Low Price	High Price
570	High Price	High Price	High Price
600	High Price	Mid Price	Low Price
1008	Mid Price	High Price	Low Price
1049	High Price	High Price	Mid Price
2932	High Price	High Price	High Price
3785	High Price	High Price	High Price

5.5. Sales and Forecast Distribution

The stated forecasting model uses realized sales to generate demand forecasts. The chosen method, ExtraTrees Regressor, tries to imitate the distribution of the real sales. The fitted decision trees can generate different demand forecasts with same input data-set. Therefore, the votes of the trees has a distribution which is expected to be similar to the real sales data.

Figure 5.3 exhibits sales and forecast distributions of a product for different price levels where x axis corresponds to the sales quantity. As seen from the histograms, distributions of the generated forecasts are similar to the distribution of the realized sales.

Outlier points are removed from both sales and forecast data. Histograms for other products are reported in Appendix 6.1. The titles of the figures include product, price level and data type information with a format *'ProductID - PriceLevel - DataType'*.

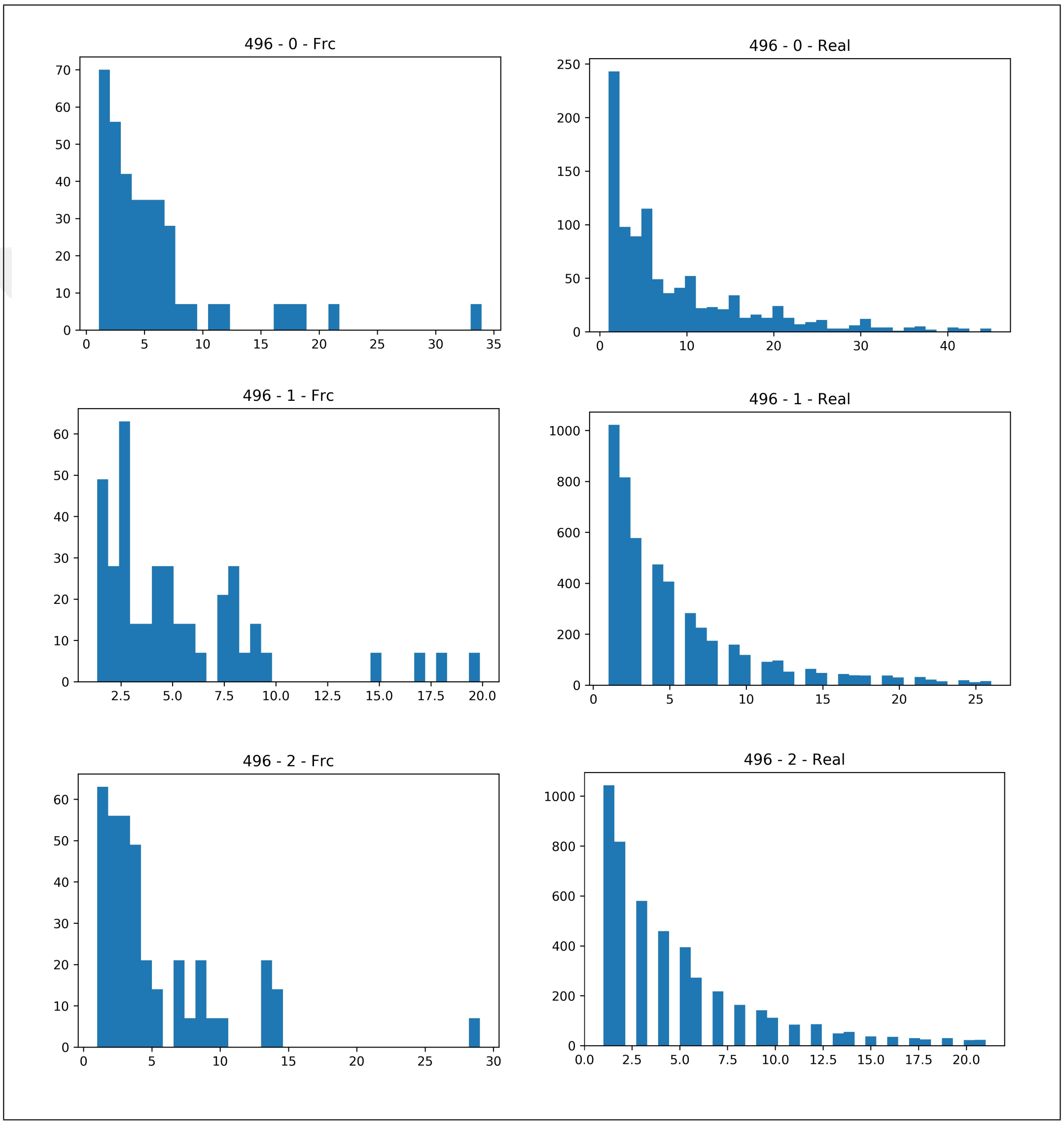

Figure 5.3. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=496

5.6. Inventory Allocation Between Fulfillment Centers, Optimal Results

Stated fulfillment model, determines inventory allocation between fulfillment centers. In the first period of the planning horizon, company should allocate its resources through different locations. The optimization model gives optimal results for that allocation problem. Table 5.10 summarizes total allocated inventory quantities across all products and all runs.

Table 5.10. Total Allocated Inventory Value Across All Runs(in $)

StateID	State	Total Allocated Inventory
4	**AZ**	7,985,894
5	**CA**	2,426,554
10	**FL**	6,846,496
11	**GA**	3,041,688
13	**IA**	116,665,412
15	**IL**	34,263,433
16	**IN**	16,230,450
18	**KY**	22,861,366
25	**MO**	13,772,842
26	**MS**	31,853,276
35	**NY**	77,994,852
36	**OH**	13,850,792
37	**OK**	1,954,485
38	**OR**	4,175,756
39	**PA**	1,580,726
44	**TN**	2,437,372
45	**TX**	7,995,204

The fulfillment model selects the FCs according to their corresponding fulfillment costs. As mentioned these costs are driven from the real shipment data. Figure 5.4 shows the sum of allocated inventory quantities at the beginning of the planning horizon.

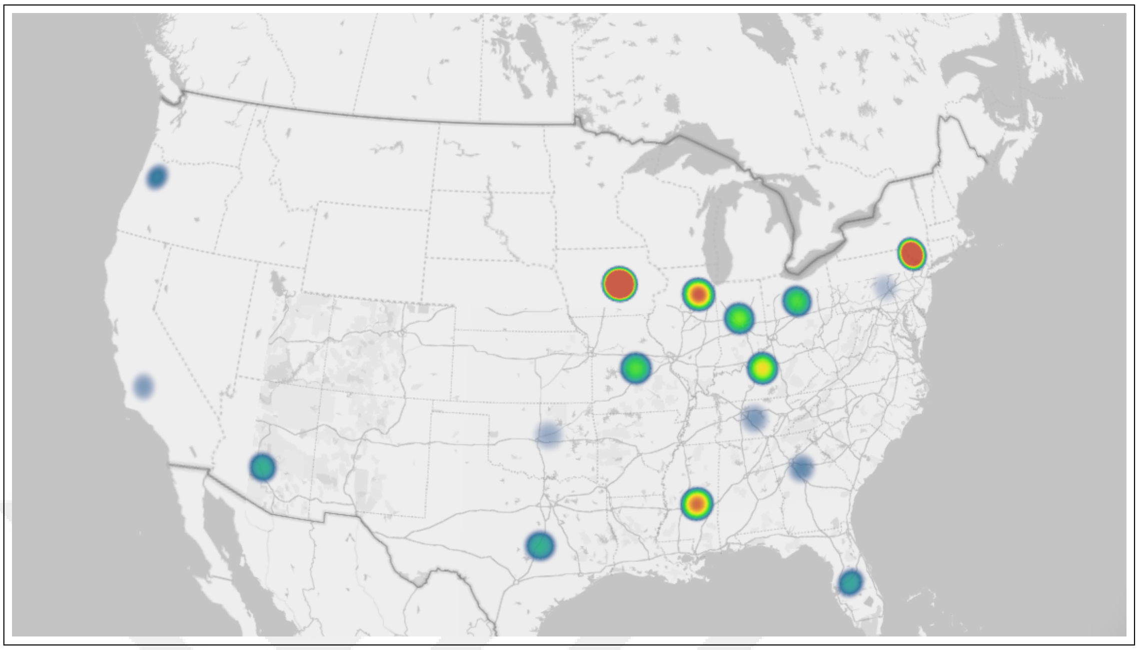

Figure 5.4. Heatmap for Inventory Allocation Between FCs in Different States(in $)

5.7. Summary

In different capacity levels, total profit can be increased by in a range of 4-25%. Table 5.11, compares the mid price level with the optimal price level that maximizes the total profit. The optimization result with mid price level is utilized as a baseline for the comparison process.

Table 5.11. Total Profits for Optimal Results($)

Capacity Level	Mid Price Level	Optimal Price Level	Improvement
0	1,203,958	1,253,501	4.115%
1	2,401,015	2,499,601	4.106%
2	3,548,764	4,608,933	29.874%

As mentioned in Section 5.4, not all the optimization runs have same confidence level bacause of the generated demand forecasts. By comparing the weighted total profits, optimal price levels are determined. Table 5.12 compares that weighted total profit with the mid price level case.

Table 5.12. Total Profits for Optimal Results($)

Capacity Level	Mid Level Profit	Weighted Max Profit	Improvement
0	1,203,958	1,230,680	2.220%
1	2,401,015	2,481,202	3.340%
2	3,548,764	4,472,522	26.030%

As seen from tables above, using only mean demands may provide better optimization results. However, to implement these results in real world, prescriptiveness level of the results should be used. Therefore, results seen in Table 5.12 is more realistic and applicable than the results in Table 5.11.

6. CONCLUSION

In this study, a joint pricing and fulfillment model is established and discussed for an online retailer company.

The forecasting part shows that, machine learning approaches outperform traditional time-series methods in most cases. The offered model handles and normalizes the sales in promotion duration without any manual manipulation. And also, time series effects can be added into such models to use obvious sales patterns and seasonality effects.

The pricing process indicates that, changing the price helps company to reach customers with different willingness to pay levels. However, even a small change on price may change all inventory allocations across different FCs. Therefore, solving these two problems at the same time is very crucial to maximize total profit of the company. The shipment costs, price levels, ship-to locations and other model parameters/constraints are generated using real sales and shipment data. Therefore, after implementing such solution to an online retailer, it would become a sustainable model for future periods.

The optimal results offer about 5% to 25% improvement on total profit(Table 5.11, Table 5.12). Especially on high capacity situations, pricing effect has a huge impact on the total profit. It is because, there is a binding lost sales situation for low and mid capacity levels.

Table 5.9 suggests that, according to the available inventory capacity, correct pricing strategy would increase the total profit of the company. Also the initial inventory levels can be decided by analyzing the lost sales reported in Table 5.2, Table 5.4 and Table 5.6. Some products have shortage in low capacity levels even with high price levels. Capacity levels for these products can be decreased, or higher prices can be charged to get rid of a possible shortage. On the other hand, most of the products have lost sales in mid capacity level. Increasing their initial inventory levels, would increase the expected total profit of the company.

To sum up, pricing and capacity levels can be an instrument for the company to manage its service level and inventory capacities. By using such a model, the company can manage its resources to get an optimal pricing and fulfillment strategy, which maximize its profit.

6.1. Future Work

Replenishment and transsshipment between FCs during the planning horizon is not permitted in this study. Future works may focus on such a solution that handles these two operations. Also some model assumptions are made for the optimization model(continuous demands, no backlogs etc). An advance version of the offered study may be constructed and discussed to relax these assumptions. The final results are compared by using also a model output in this thesis. For future studies, a real-time simulation can be implemented and tested for the forecasting method. This advanced model may also offer more dynamic pricing and fulfillment process for the company.

REFERENCES

1. Richter, F., *The Growing Weight of Amazon's Logistics Costs*, 2019, https://www.statista.com/chart/12893/amazon-fulfillment-and-shipping-costs/, accessed in June 2019.

2. of the Department of Commerce, T. C. B., *Quarterly Retail E-Commerce Sales 1st Quarter 2019*, https://www2.census.gov/retail/releases/historical/ecomm/19q1.pdf, accessed in July 2019.

3. Huang, J., "Demand Functions in Decision Modeling: A Comprehensive Survey and Research Directions", *Decision Sciences*, Vol. 44, No. 3, pp. 557–609, 2013, https://onlinelibrary.wiley.com/doi/abs/10.1111/deci.12021.

4. Kuo, R. and K. Xue, "Fuzzy neural networks with application to sales forecasting", *Fuzzy Sets and Systems*, Vol. 108, No. 2, pp. 123 – 143, 1999, http://www.sciencedirect.com/science/article/pii/S0165011497003266.

5. Chinthalapati, V., N. Yadati and R. Karumanchi, "Learning dynamic prices in MultiSeller electronic retail markets with price sensitive customers, stochastic demands, and inventory replenishments", *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, Vol. 36, pp. 92 – 106, 02 2006.

6. Ghiasnezhad Omran, N. and S. Filizadeh, "Location-Based Forecasting of Vehicular Charging Load on the Distribution System", *IEEE Transactions on Smart Grid*, Vol. 5, No. 2, pp. 632–641, March 2014.

7. Ramanathan, U. and L. Muyldermans, "Identifying demand factors for promotional planning and forecasting: A case of a soft drink company in the UK", *International Journal of Production Economics*, Vol. 128, No. 2, pp. 538 – 545, 2010, http://www.sciencedirect.com/science/article/pii/S092552731000232X, supply Chain Forecasting Systems.

8. Ferreira, K., B. Hong Alex Lee and D. Simchi-levi, "Analytics for an Online Retailer: Demand Forecasting and Price Optimization", *Manufacturing Service Operations Management*, Vol. 18, 11 2015.

9. Xia, C. H. and P. Dube, "Dynamic Pricing in e-Services under Demand Uncertainty", *Production and Operations Management*, Vol. 16, No. 6, pp. 701–712.

10. Fisher, M., S. Gallino and J. li, "Competition-Based Dynamic Pricing in Online Retailing: A Methodology Validated with Field Experiments", *Management Science*, Vol. 64, 06 2017.

11. Lin, F.-r. and M. Shaw, "Reengineering the Order Fulfillment Process in Supply Chain Networks", *International Journal of Flexible Manufacturing Systems*, Vol. 10, pp. 197–229, 07 1998.

12. Acimovic, J. and S. Graves, "Making Better Fulfillment Decisions on the Fly in an Online Retail Environment", *Manufacturing Service Operations Management*, Vol. 17, pp. 34–51, 02 2015.

13. Torabi, S., E. Hassini and M. Jeihoonian, "Fulfillment source allocation, inventory transshipment, and customer order transfer in e-tailing", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 79, pp. 128 – 144, 2015, http://www.sciencedirect.com/science/article/pii/S136655451500085X.

14. Ishfaq, R. and N. Bajwa, "Profitability of online order fulfillment in multichannel retailing", *European Journal of Operational Research*, Vol. 272, No. 3, pp. 1028 – 1040, 2019, http://www.sciencedirect.com/science/article/pii/S0377221718306696.

15. Lei, Y. M., S. Jasin and A. Sinha, "Joint Dynamic Pricing and Order Fulfillment for E-commerce Retailers", *Manufacturing & Service Operations Management*, Vol. 20, No. 2, pp. 269–284, 2018, https://doi.org/10.1287/msom.2017.0641.

16. Otero, D. F. and R. Akhavan-Tabatabaei, “A stochastic dynamic pricing model for the multiclass problems in the airline industry”, *European Journal of Operational Research*, Vol. 242, No. 1, pp. 188 – 200, 2015, http://www.sciencedirect.com/science/article/pii/S0377221714007772.

17. Marand, A. J., H. Li and A. Thorstenson, “Joint inventory control and pricing in a service-inventory system”, *International Journal of Production Economics*, Vol. 209, pp. 78 – 91, 2019, http://www.sciencedirect.com/science/article/pii/S0925527317302256, the Proceedings of the 19th International Symposium on Inventories.

18. Ahn, H.-s., M. Gümüş and P. Kaminsky, “Pricing and Manufacturing Decisions When Demand Is a Function of Prices in Multiple Periods”, *Operations Research*, Vol. 55, No. 6, pp. 1039–1057, 2007, https://doi.org/10.1287/opre.1070.0411.

19. Bertsimas, D. and N. Kallus, 2014, ”From Predictive to Prescriptive Analytics”, 2014.

20. Liaw, A. and M. Wiener, “Classification and Regression by RandomForest”, *Forest*, Vol. 23, 11 2001.

21. Geurts, P., D. Ernst and L. Wehenkel, “Extremely Randomized Trees”, *Machine Learning*, Vol. 63, pp. 3–42, 04 2006.

22. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.

23. Gurobi Optimization, L., *Gurobi Optimizer Reference Manual*, 2018, http://www.gurobi.com, accessed in July 2019.

APPENDIX A: SALES AND FORECAST DISTRIBUTIONS FOR PRODUCTS

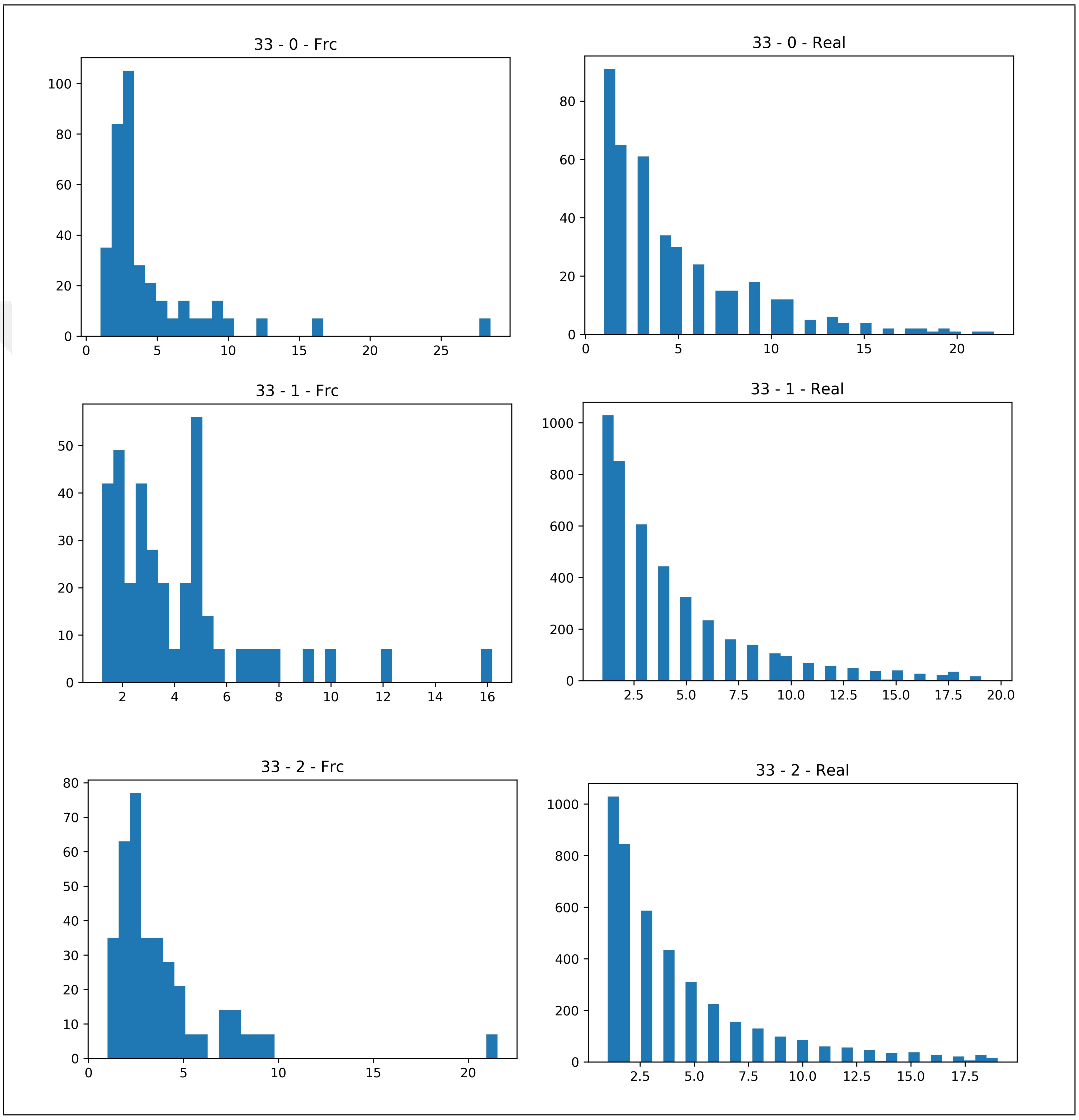

Figure A.1. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=33

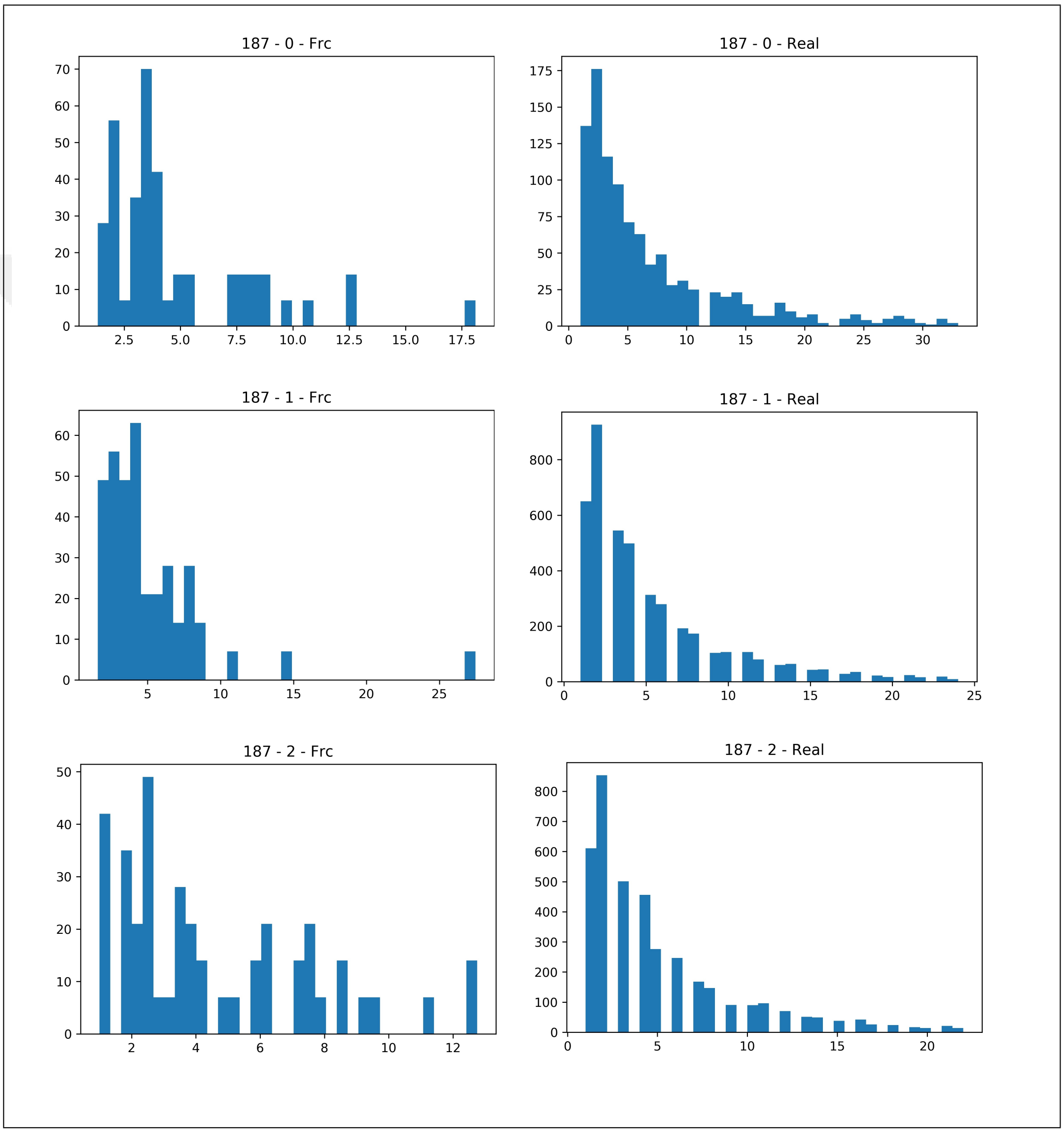

Figure A.2. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=187

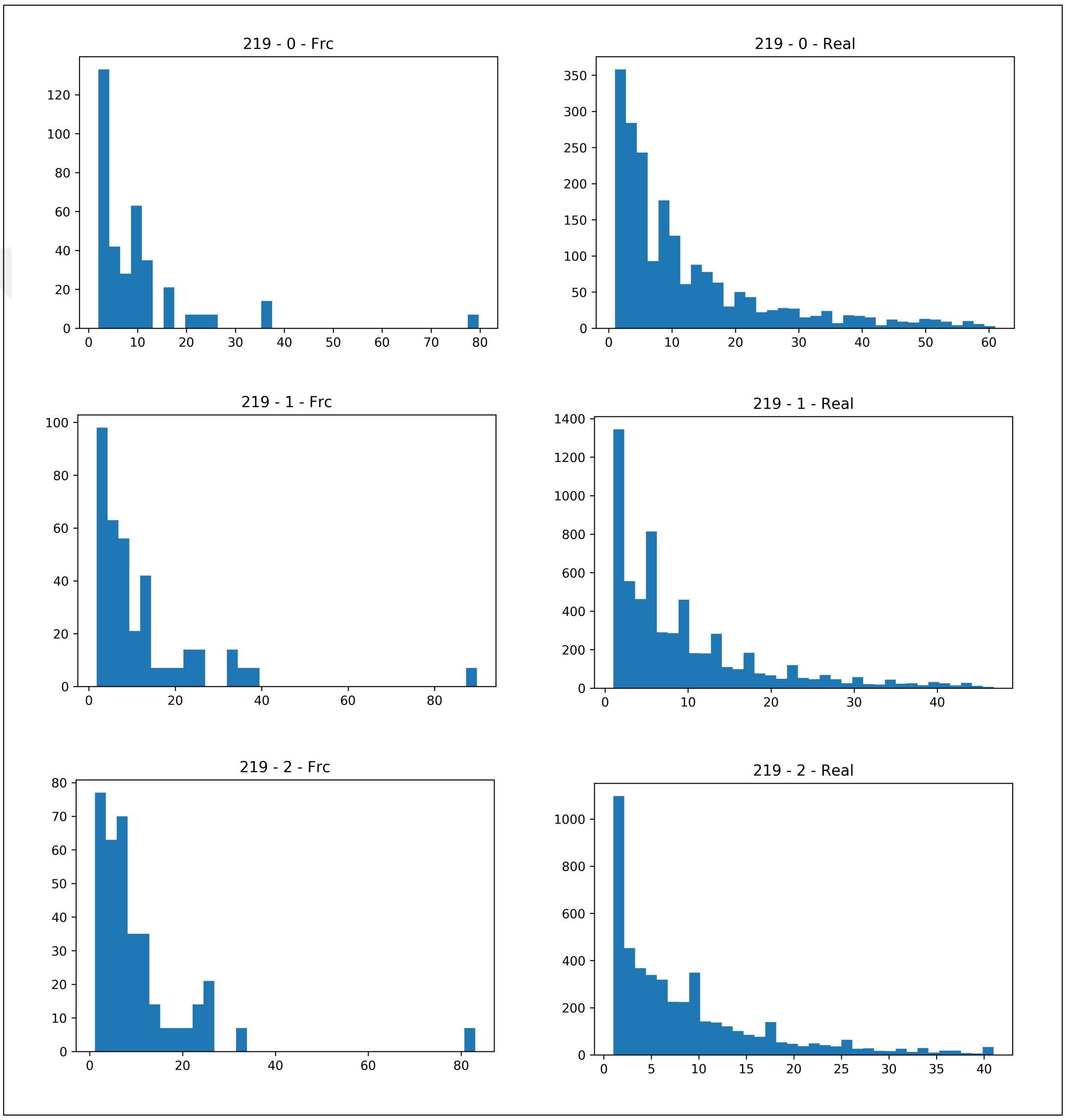

Figure A.3. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=219

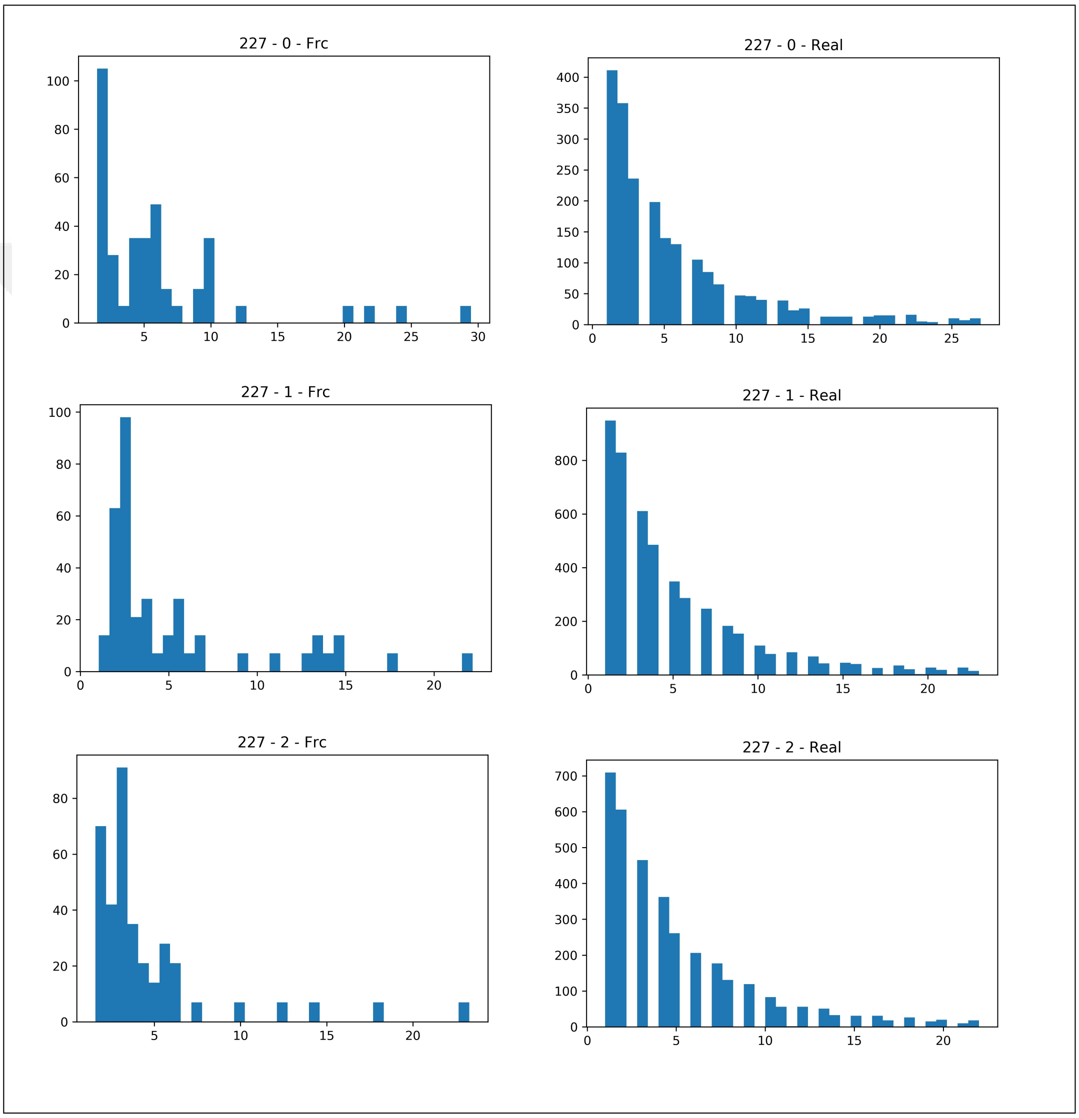

Figure A.4. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=227

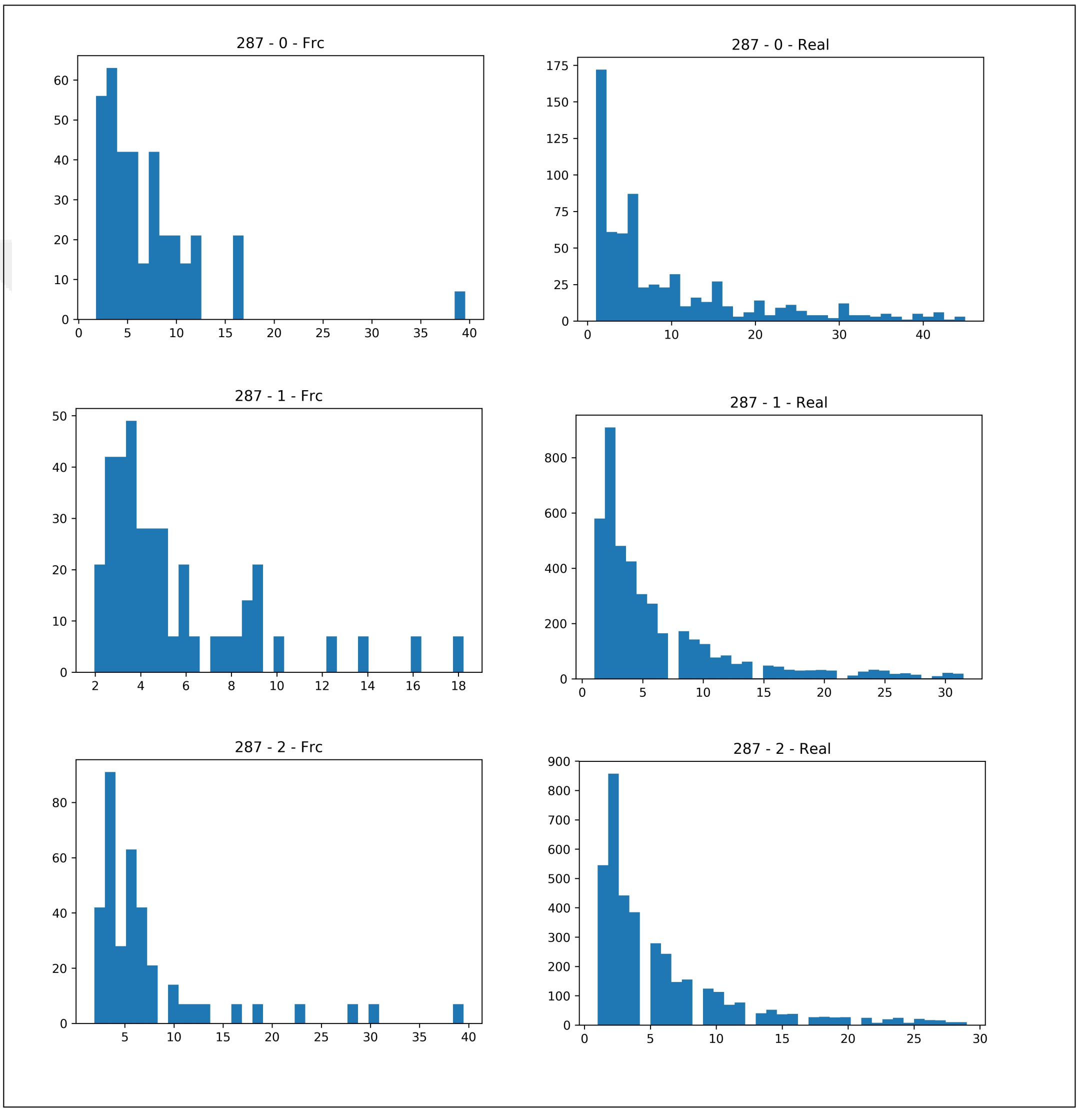

Figure A.5. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=287

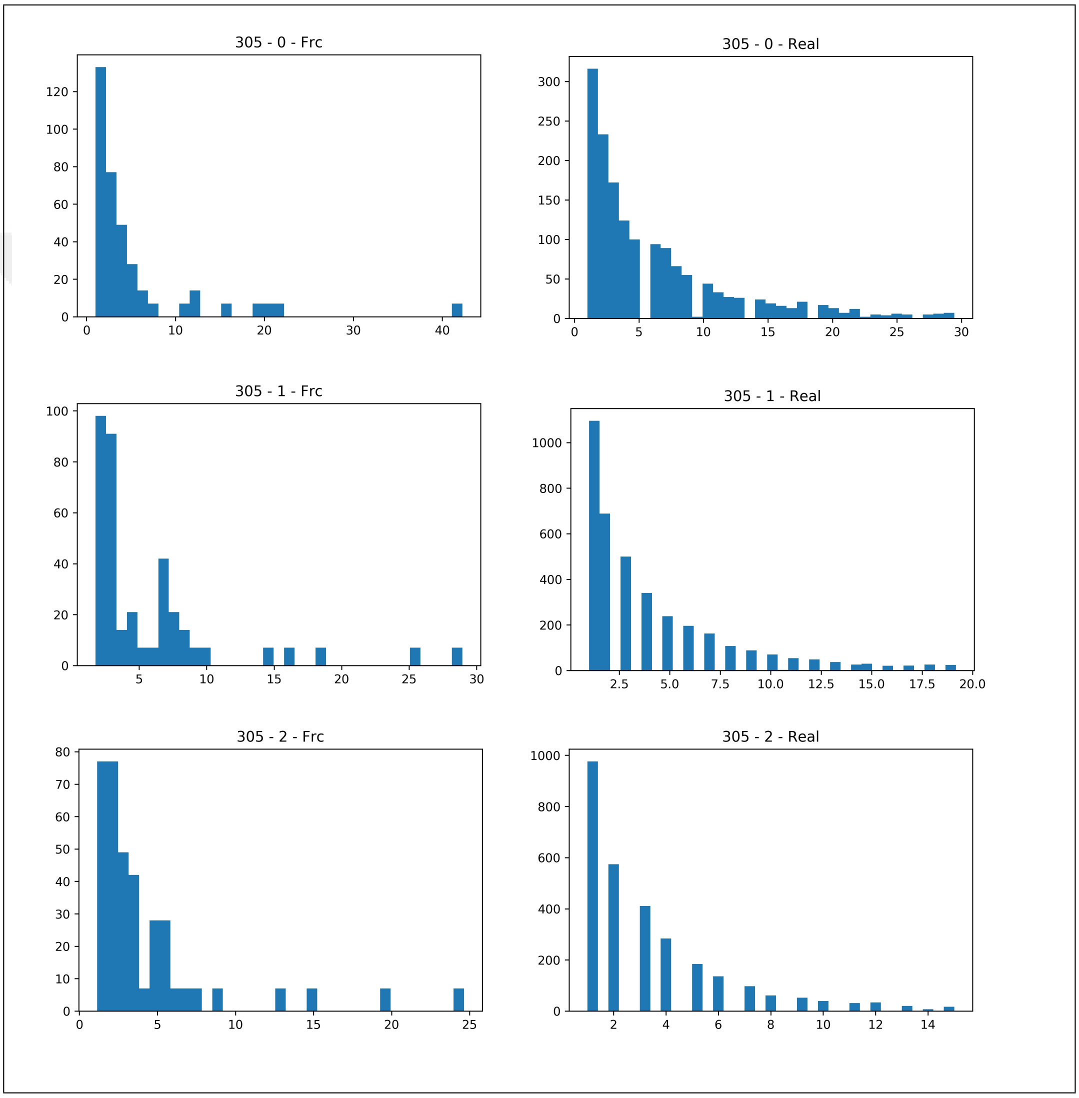

Figure A.6. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=305

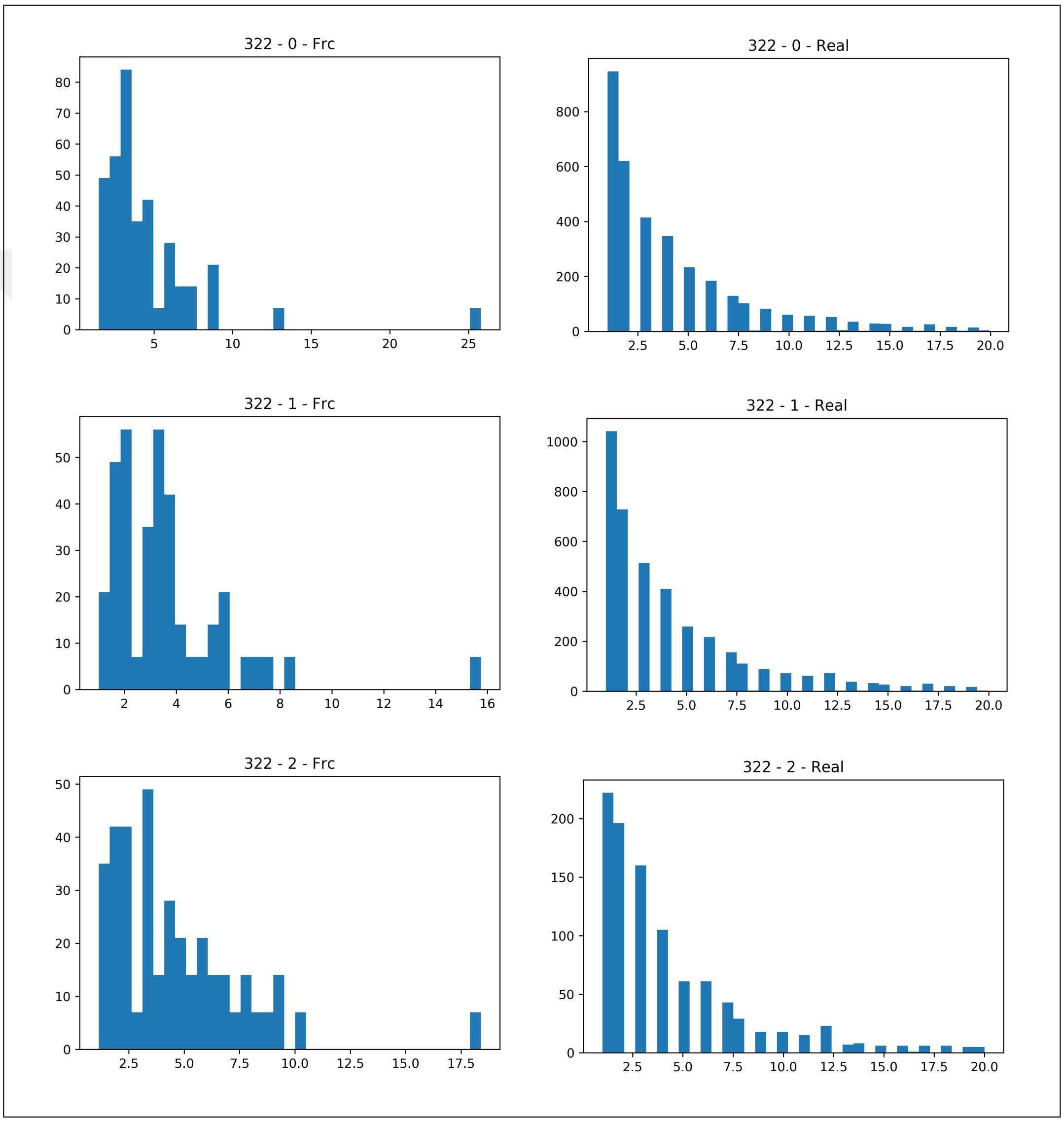

Figure A.7. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=322

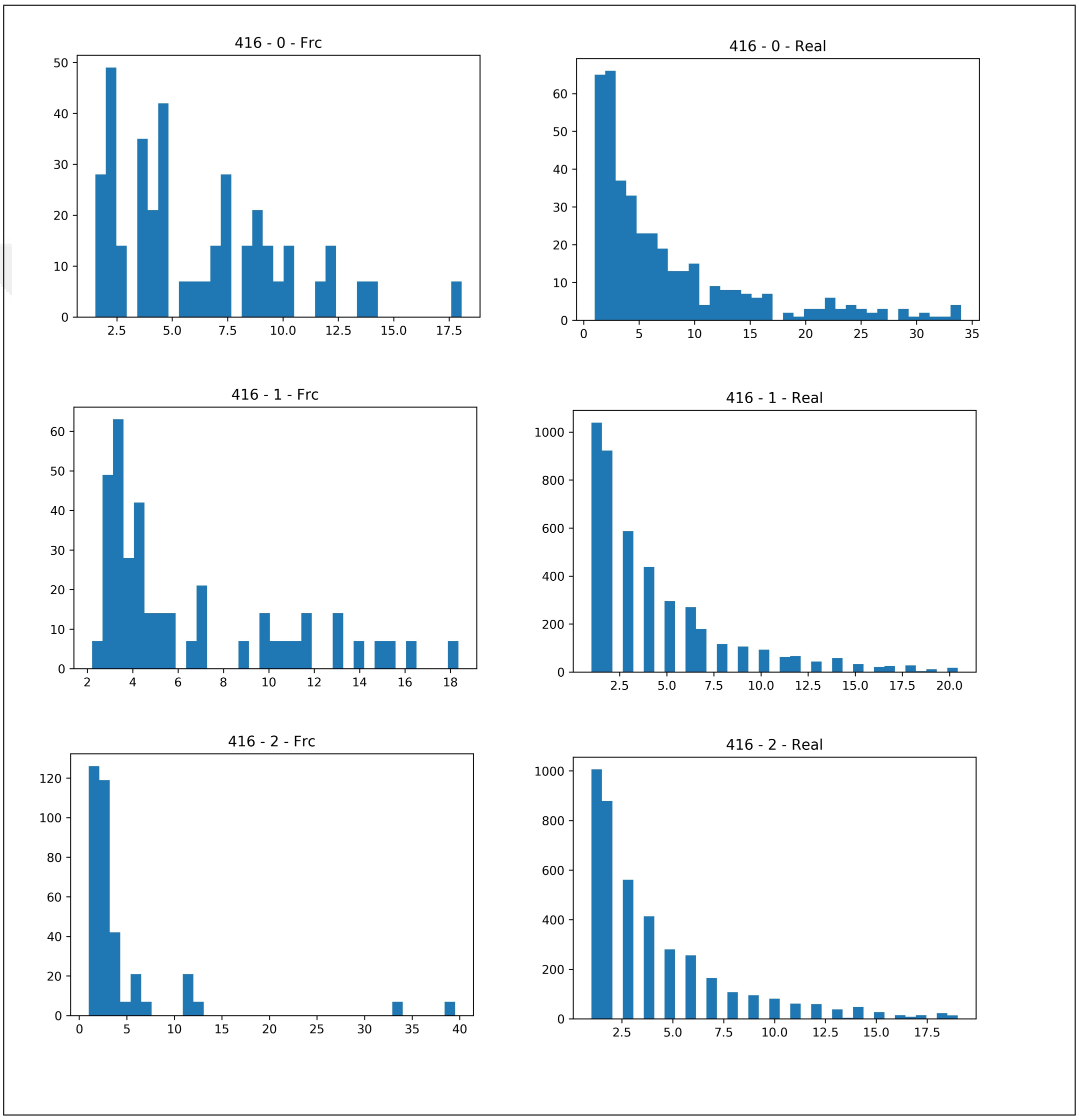

Figure A.8. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=416

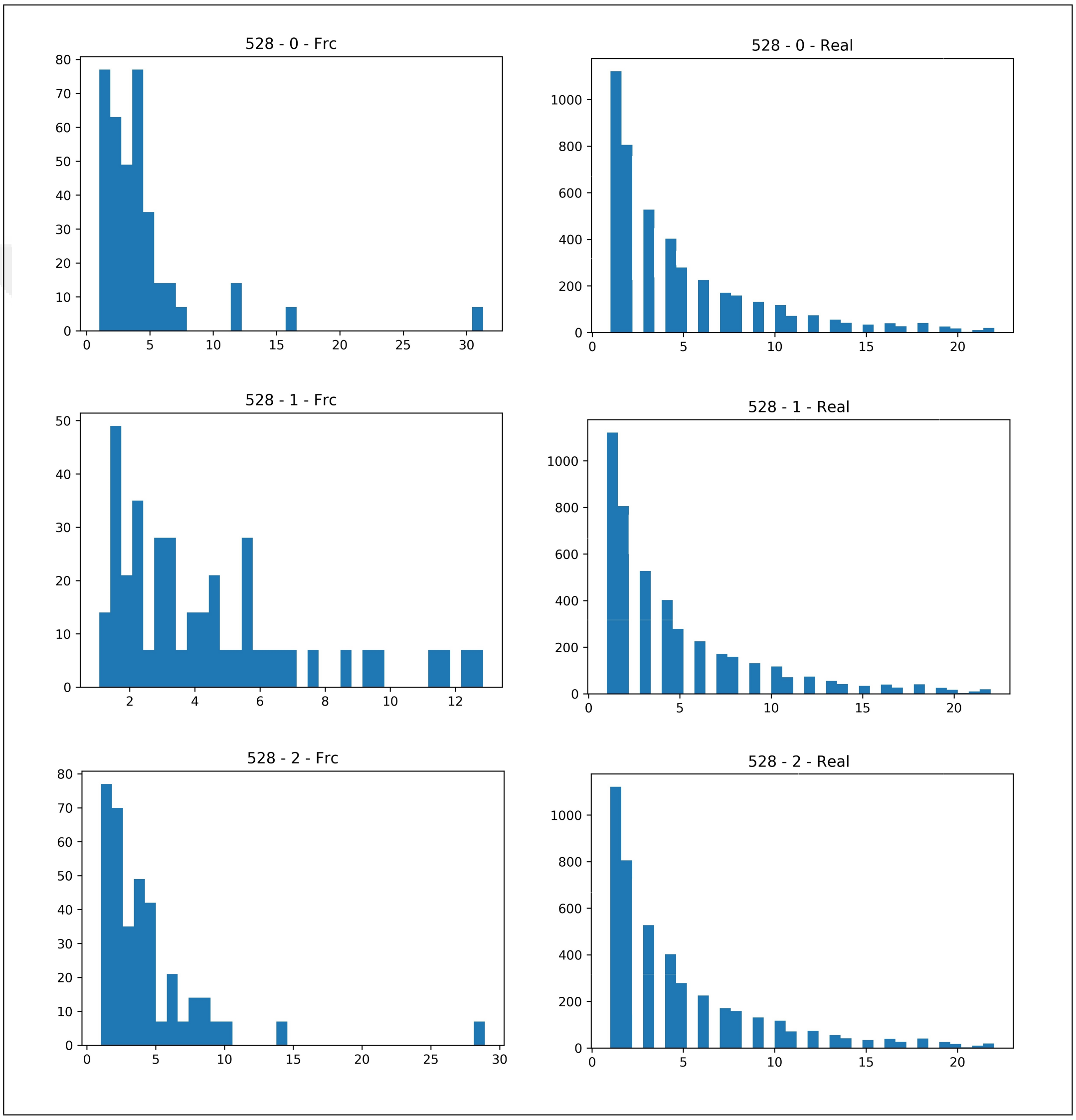

Figure A.9. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=528

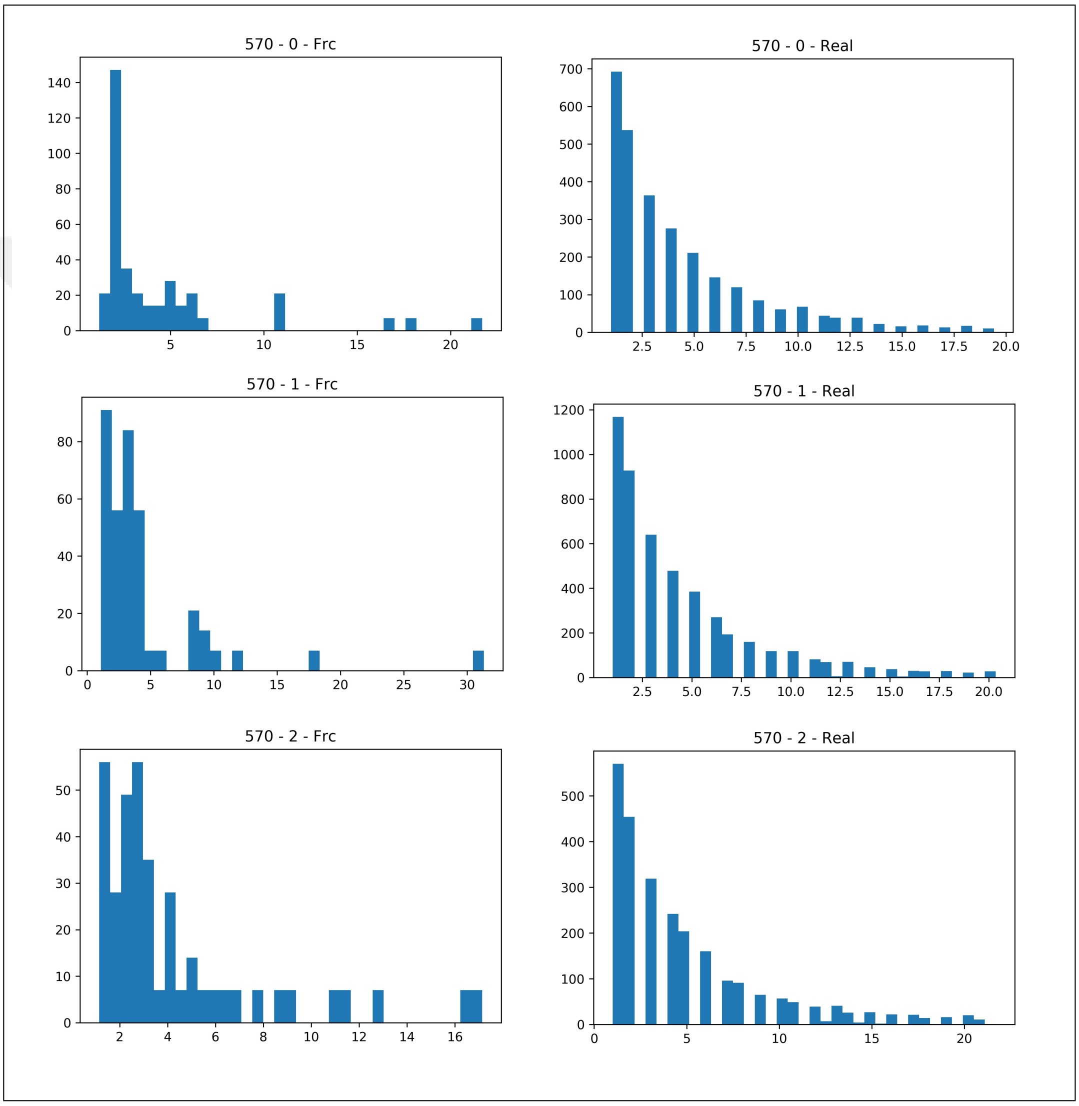

Figure A.10. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=570

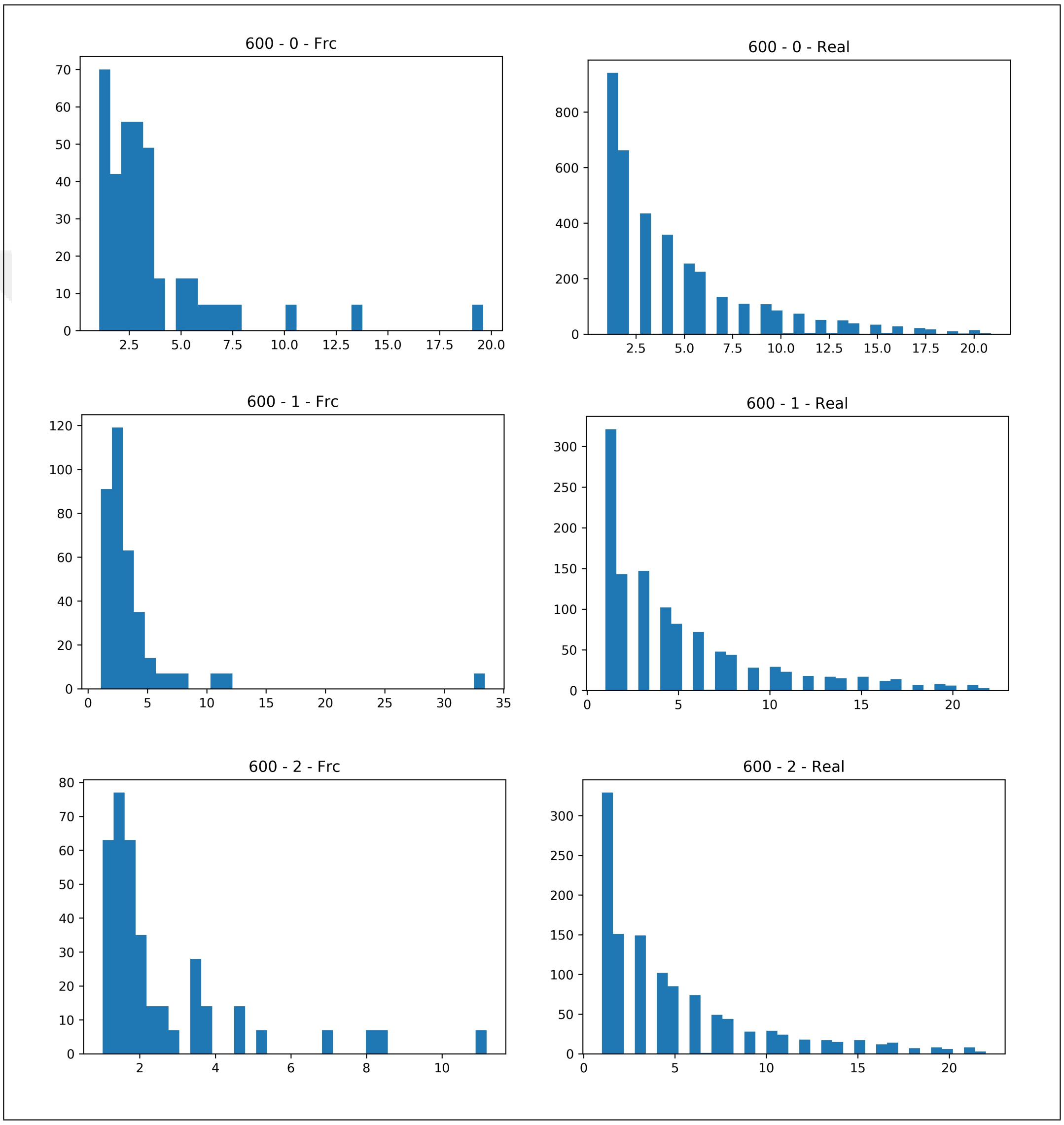

Figure A.11. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=600

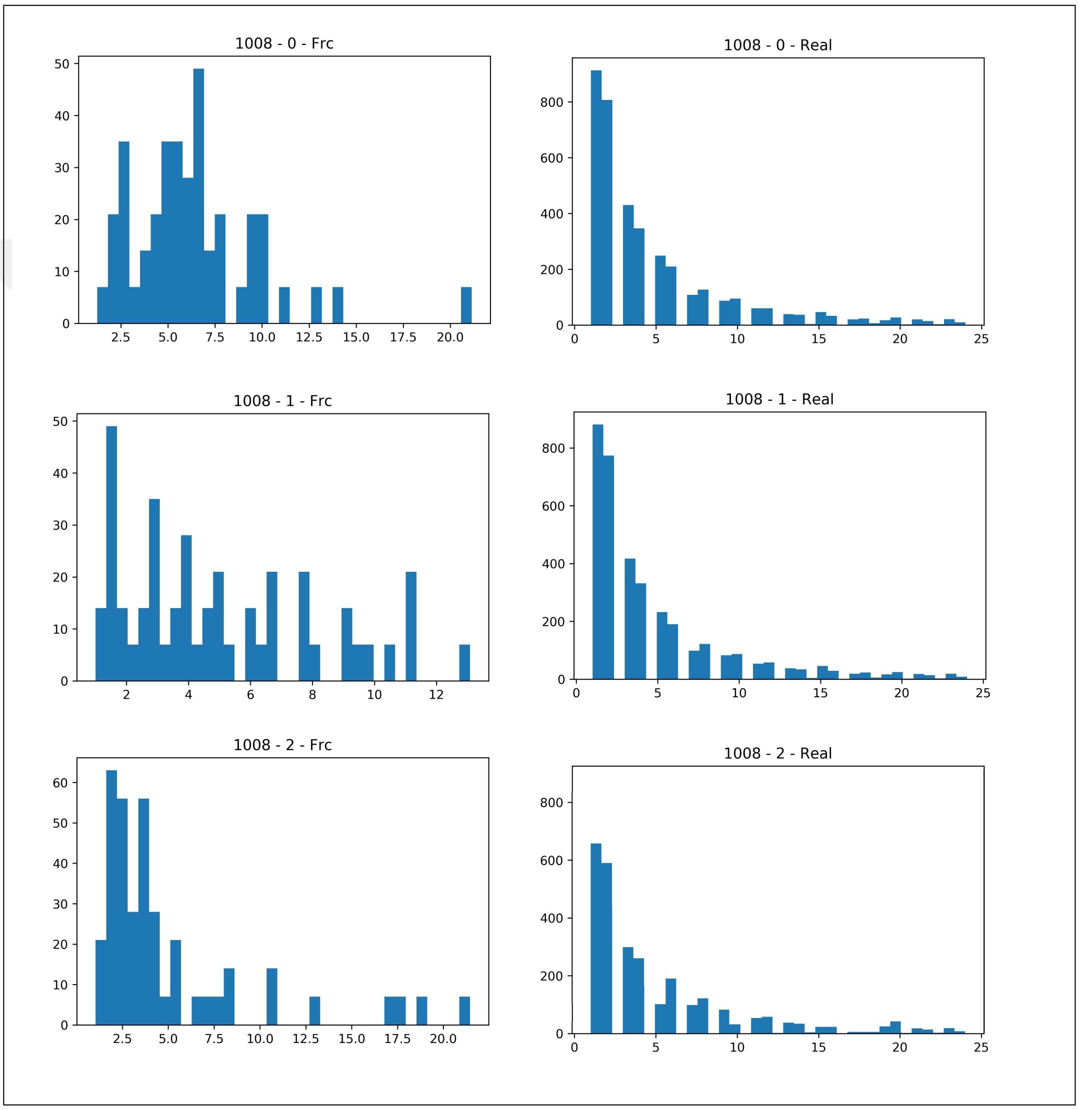

Figure A.12. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=1008

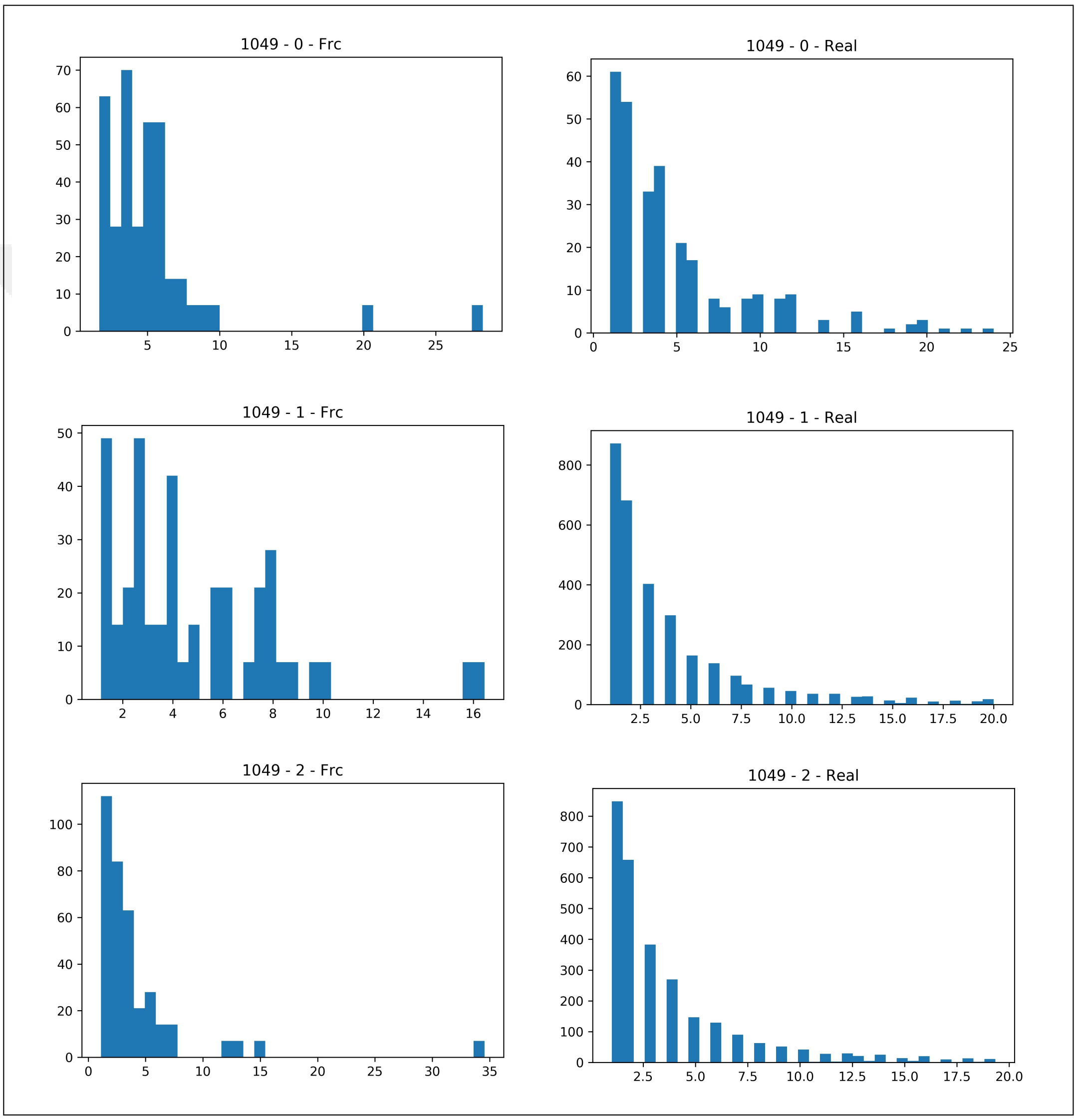

Figure A.13. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=1049

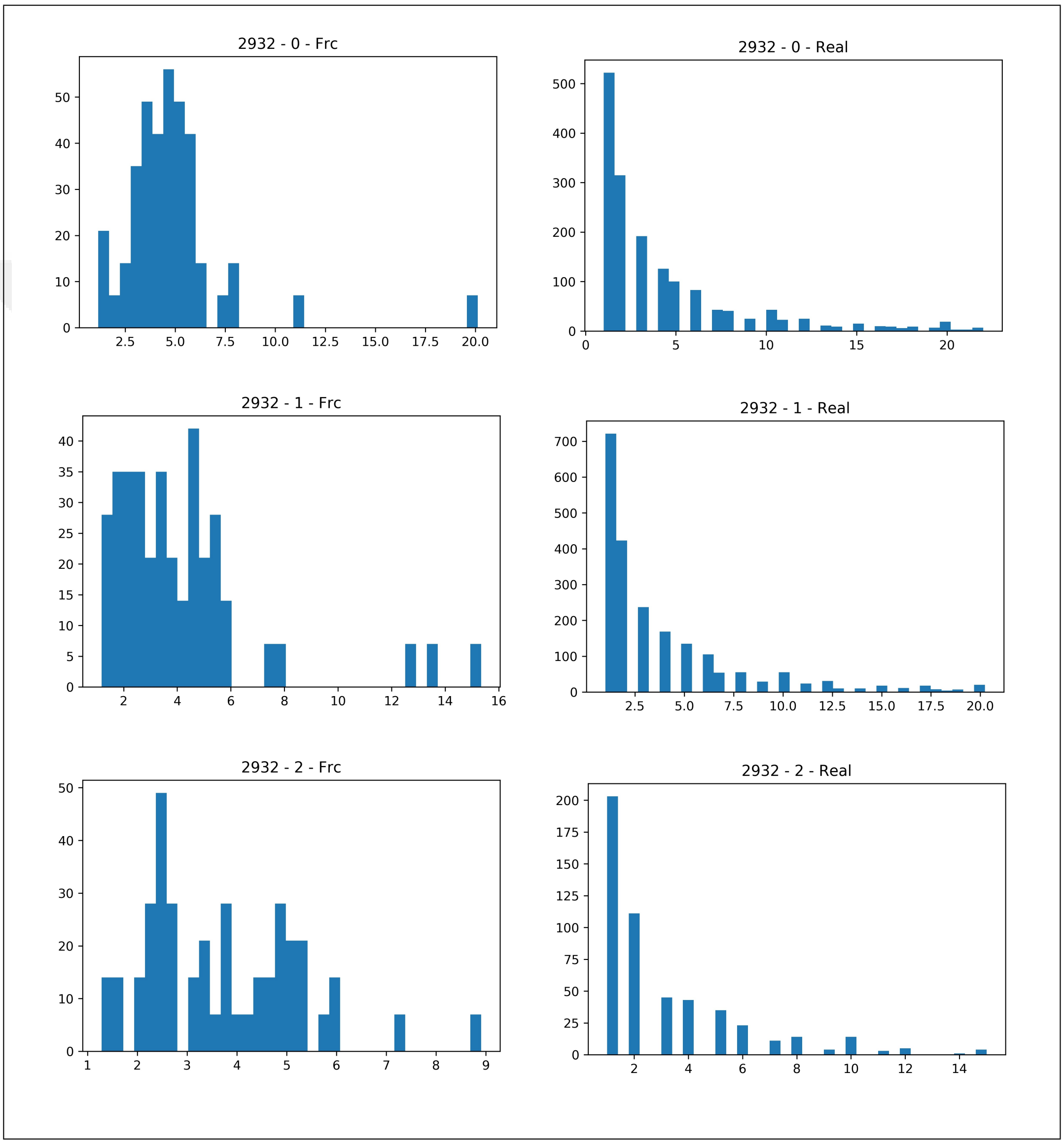

Figure A.14. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=2932

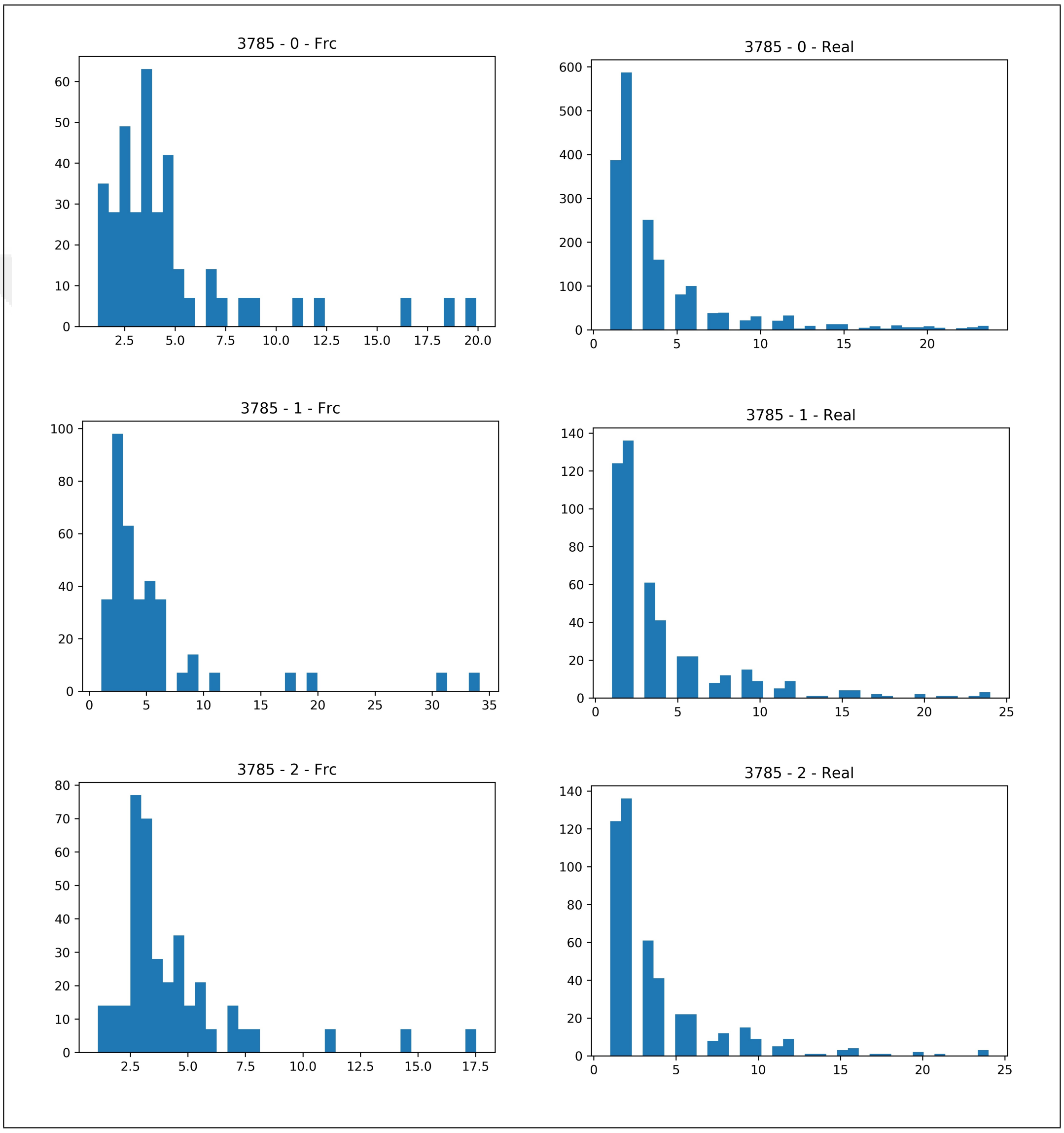

Figure A.15. Sales and Forecast Frequency Histograms for Different Price Levels, ProductID=3785

APPENDIX B: AVERAGE WEIGHTS OF THE FORECASTING RUNS

Weights for the generated forecasts summarized in the following table.

Table B.1. Avg. Weights of the Forecasting Runs

RunID	PriceLevel	Low Capacity	Mid Capacity	High Capacity
0	**Low Price**	0.1777	0.2171	0.2208
1	**Low Price**	0.2108	0.1912	0.1957
2	**Low Price**	0.2007	0.1992	0.2121
3	**Low Price**	0.1749	0.2011	0.1776
4	**Low Price**	0.1731	0.2102	0.2348
5	**Low Price**	0.1878	0.2187	0.1972
6	**Low Price**	0.2170	0.2174	0.1964
7	**Low Price**	0.2046	0.2062	0.2163
8	**Low Price**	0.2108	0.1987	0.1983
9	**Low Price**	0.2190	0.1902	0.2177
10	**Low Price**	0.2070	0.1704	0.1992
11	**Low Price**	0.2049	0.2416	0.2144
12	**Low Price**	0.2208	0.1837	0.1895
13	**Low Price**	0.2164	0.2146	0.1994
14	**Low Price**	0.2094	0.1993	0.2118
15	**Low Price**	0.1954	0.2097	0.2109
16	**Low Price**	0.2240	0.1821	0.2113
17	**Low Price**	0.1962	0.2055	0.2097
18	**Low Price**	0.2300	0.1816	0.2029
19	**Low Price**	0.1948	0.2192	0.2071
0	**Mid Price**	0.2301	0.2249	0.2271
1	**Mid Price**	0.2353	0.2392	0.2440
2	**Mid Price**	0.2365	0.2401	0.2378
3	**Mid Price**	0.2363	0.2330	0.2626
4	**Mid Price**	0.2491	0.2277	0.2351
5	**Mid Price**	0.2443	0.2357	0.2254
6	**Mid Price**	0.2352	0.2400	0.2361
7	**Mid Price**	0.2349	0.2450	0.2521
8	**Mid Price**	0.2566	0.2189	0.2585
9	**Mid Price**	0.2062	0.2247	0.2301
10	**Mid Price**	0.2483	0.2376	0.2303
11	**Mid Price**	0.2217	0.2342	0.2443
12	**Mid Price**	0.2436	0.2285	0.2252

Table B.1. Avg. Weights of the Forecasting Runs (cont.)

RunID	PriceLevel	Low Capacity	Mid Capacity	High Capacity
13	**Mid Price**	0.2073	0.2295	0.2428
14	**Mid Price**	0.2411	0.2405	0.2316
15	**Mid Price**	0.2274	0.2157	0.2240
16	**Mid Price**	0.2445	0.2486	0.2684
17	**Mid Price**	0.2142	0.2204	0.2359
18	**Mid Price**	0.2466	0.2415	0.2486
19	**Mid Price**	0.2489	0.2294	0.2493
0	**High Price**	0.2424	0.2444	0.2665
1	**High Price**	0.2370	0.2332	0.2535
2	**High Price**	0.2632	0.2587	0.2244
3	**High Price**	0.2281	0.2486	0.2354
4	**High Price**	0.2429	0.2354	0.2299
5	**High Price**	0.2603	0.2698	0.2671
6	**High Price**	0.2563	0.2739	0.2523
7	**High Price**	0.2187	0.2608	0.2287
8	**High Price**	0.2557	0.2250	0.2479
9	**High Price**	0.2515	0.2656	0.2610
10	**High Price**	0.2528	0.2446	0.2889
11	**High Price**	0.2569	0.2541	0.2539
12	**High Price**	0.2664	0.2451	0.2348
13	**High Price**	0.2521	0.2456	0.2321
14	**High Price**	0.2554	0.2445	0.2625
15	**High Price**	0.2591	0.2578	0.2609
16	**High Price**	0.2565	0.2523	0.2377
17	**High Price**	0.2420	0.2652	0.2413
18	**High Price**	0.2587	0.2206	0.2493
18	**High Price**	0.2330	0.2529	0.2540

APPENDIX C: OPTIMIZATION RESULTS FOR DIFFERENT INPUT CASES

Table C.1. Results for Low Capacity, Low Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	1,094,233	78,168	48,444
1	1,094,231	173,529	48,445
2	1,094,273	147,061	48,403
3	1,094,394	94,287	48,283
4	1,094,159	91,338	48,518
5	1,094,445	106,918	48,231
6	1,094,169	198,891	48,507
7	1,094,451	140,867	48,225
8	1,094,276	163,925	48,400
9	1,094,306	121,629	48,370
10	1,094,293	137,191	48,384
11	1,094,293	143,306	48,383
12	1,094,273	127,691	48,404
13	1,094,301	135,775	48,375
14	1,094,163	188,204	48,513
15	1,094,337	102,936	48,340
16	1,094,376	137,972	48,300
17	1,094,001	121,935	48,675
18	1,094,151	187,744	48,525
19	1,094,282	106,328	48,395

Table C.2. Results for Low Capacity, Mid Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	1,203,743	239,991	49,179
1	1,203,978	147,815	48,945
2	1,203,775	203,541	49,147
3	1,204,145	202,692	48,778
4	1,203,953	225,594	48,969
5	1,204,130	215,080	48,792
6	1,204,372	214,944	48,550
7	1,203,823	183,786	49,099
8	1,203,984	266,380	48,938
9	1,203,792	135,792	49,130
10	1,203,818	173,537	49,104
11	1,203,906	162,570	49,016
12	1,204,001	174,237	48,921
13	1,204,070	143,884	48,852
14	1,203,906	187,366	49,016
15	1,203,915	235,065	49,008
16	1,204,010	229,831	48,912
17	1,203,966	184,349	48,956
18	1,203,869	170,911	49,053
19	1,204,007	210,804	48,915

Table C.3. Results for Low Capacity, High Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	1,253,406	221,749	49,397
1	1,253,342	176,729	49,461
2	1,253,519	268,842	49,283
3	1,253,582	181,598	49,221
4	1,253,507	188,190	49,295
5	1,253,708	242,312	49,095
6	1,253,804	248,958	48,999
7	1,253,423	163,932	49,379
8	1,253,475	242,050	49,328
9	1,253,231	230,820	49,572
10	1,253,408	201,259	49,395
11	1,253,658	264,856	49,145
12	1,253,415	282,048	49,388
13	1,253,532	179,669	49,271
14	1,253,473	254,088	49,330
15	1,253,591	242,632	49,212
16	1,253,452	211,357	49,350
17	1,253,509	176,719	49,293
18	1,253,492	240,617	49,311
19	1,253,492	210,500	49,310

Table C.4. Results for Mid Capacity, Low Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	2,182,358	341,169	102,995
1	2,182,744	302,500	102,609
2	2,182,518	338,519	102,834
3	2,183,500	180,769	101,853
4	2,182,459	372,312	102,894
5	2,183,559	371,663	101,794
6	2,182,524	417,701	102,829
7	2,183,121	263,228	102,232
8	2,182,683	283,301	102,669
9	2,182,792	211,326	102,561
10	2,183,285	193,416	102,068
11	2,182,738	389,071	102,615
12	2,182,607	251,366	102,746
13	2,183,015	305,541	102,338
14	2,182,127	382,883	103,226
15	2,183,000	287,827	102,352
16	2,183,301	250,406	102,052
17	2,181,403	319,515	103,950
18	2,182,421	233,200	102,932
19	2,183,035	356,006	102,318

Table C.5. Results for Mid Capacity, Mid Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	2,400,710	407,583	105,135
1	2,400,996	451,167	104,848
2	2,400,592	408,277	105,253
3	2,401,544	451,983	104,300
4	2,400,939	337,193	104,905
5	2,401,093	390,756	104,751
6	2,401,906	409,492	103,938
7	2,400,715	431,138	105,129
8	2,401,139	278,146	104,705
9	2,400,846	428,963	104,998
10	2,400,869	384,974	104,975
11	2,400,866	457,616	104,979
12	2,401,250	390,934	104,594
13	2,401,234	334,211	104,610
14	2,400,581	421,767	105,263
15	2,400,876	366,300	104,968
16	2,401,111	408,559	104,733
17	2,401,006	385,741	104,838
18	2,400,651	423,585	105,193
19	2,401,370	389,325	104,474

Table C.6. Results for Mid Capacity, High Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	2,499,420	507,891	105,700
1	2,499,202	402,173	105,852
2	2,499,482	602,591	105,571
3	2,499,315	519,050	105,784
4	2,499,322	390,773	105,928
5	2,499,787	528,581	105,357
6	2,501,040	542,696	104,565
7	2,499,330	377,971	105,837
8	2,499,795	466,465	105,370
9	2,498,931	417,108	106,316
10	2,499,758	475,722	105,514
11	2,500,459	545,529	105,005
12	2,499,350	475,161	105,809
13	2,499,673	480,255	105,446
14	2,499,202	395,476	105,880
15	2,500,095	571,620	105,147
16	2,499,388	486,179	105,933
17	2,499,337	555,365	105,758
18	2,499,308	452,195	105,815
19	2,499,801	428,318	105,259

Table C.7. Results for High Capacity, Low Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	4,546,799	718,573	313,396
1	4,517,300	757,704	317,929
2	4,523,772	706,844	316,076
3	4,425,393	528,998	318,737
4	4,389,712	837,296	313,119
5	4,433,357	511,268	317,379
6	4,365,420	577,761	312,416
7	4,506,189	548,504	314,109
8	4,393,370	542,206	316,766
9	4,373,880	841,721	311,945
10	4,462,363	576,966	317,546
11	4,308,942	691,887	312,574
12	4,489,922	490,757	315,321
13	4,518,652	469,826	316,405
14	4,321,758	636,568	315,412
15	4,336,291	745,090	315,990
16	4,561,823	565,397	317,476
17	4,416,811	700,168	311,710
18	4,407,276	580,909	315,683
19	4,502,475	419,802	314,342

Table C.8. Results for High Capacity, Mid Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	3,521,834	679,698	305,374
1	3,620,073	575,899	305,294
2	3,542,648	660,703	303,652
3	3,556,916	672,713	306,333
4	3,548,812	546,955	304,889
5	3,529,999	666,874	304,102
6	3,544,785	582,593	309,648
7	3,551,423	710,380	304,149
8	3,551,540	781,031	307,115
9	3,601,169	523,450	303,634
10	3,579,514	732,999	306,175
11	3,481,606	628,720	304,134
12	3,573,965	499,790	305,551
13	3,475,903	651,106	304,479
14	3,545,306	475,711	303,352
15	3,480,103	590,016	307,380
16	3,584,371	724,679	303,419
17	3,517,384	641,465	305,341
18	3,563,792	694,117	303,283
19	3,604,131	800,916	307,434

Table C.9. Results for High Capacity, High Price, All Products

RunNo	Profit	Weighted Profit	Total Cost
0	3,461,742	790,648	302,697
1	3,614,617	755,086	298,962
2	3,555,086	642,055	297,702
3	3,413,739	624,578	301,419
4	3,568,808	731,817	298,552
5	3,521,673	719,426	297,975
6	3,539,087	643,563	303,562
7	3,512,873	804,356	299,945
8	3,591,840	731,034	301,246
9	3,553,916	695,361	297,585
10	3,590,317	871,798	300,999
11	3,567,022	650,328	299,798
12	3,606,101	632,127	298,958
13	3,454,706	548,075	300,207
14	3,538,622	723,599	299,057
15	3,453,117	771,391	301,815
16	3,572,283	725,428	294,732
17	3,602,358	568,395	298,713
18	3,640,856	793,744	298,741
19	3,626,680	632,576	302,154

APPENDIX D: WEIGHTED PROFITS BY PRODUCT

Table D.1. Weighted Total Profits, Low Capacity

Products	Low Price	Mid Price	High Price
33	14,963	28,288	30,165
179	25,800	37,484	37,459
187	1,820	2,507	3,319
219	32,174	42,141	54,085
227	2,175	2,186	3,661
287	1,560	3,687	4,021
305	3,356	4,425	5,525
322	4,877	8,741	7,322
416	4,553	7,739	12,415
496	1,311	2,764	3,939
528	6,371	6,340	6,118
570	8,678	12,364	15,708
600	355	374	449
1008	15,763	20,820	18,959
1049	5,552	9,388	11,648
2932	635	915	1,219
3785	5,340	5,245	5,434

Table D.2. Weighted Total Profits, Mid Capacity

Products	Low Price	Mid Price	High Price
33	39,238	60,423	66,720
179	70,710	63,790	89,398
187	2,949	4,695	5,526
219	71,205	106,787	121,540
227	3,436	4,265	5,192
287	2,014	5,688	7,865
305	7,480	7,018	10,751
322	10,837	18,392	12,775
416	10,005	17,313	28,781
496	2,585	4,930	7,256
528	13,547	12,107	12,231
570	13,128	24,259	35,337
600	728	865	799
1008	31,287	37,702	43,393
1049	12,250	17,869	21,143
2932	1,145	1,848	2,296
3785	10,041	9,934	10,054

APPENDIX E: SAMPLING RATIO TEST RUNS, ALL PRODUCTS

Table E.1. Sampling Results for All Product Aggregation

Train Ratio	Capacity Level	Price Level	Profit (Weighted)	Profit	Total Cost
0.1	Low Capacity	Low Price	878,902	1,094,068	48,609
0.1	Low Capacity	Mid Price	1,004,984	1,203,912	49,010
0.1	Low Capacity	High Price	1,054,382	1,253,556	49,246
0.1	Mid Capacity	Low Price	1,750,416	2,182,434	102,919
0.1	Mid Capacity	Mid Price	2,007,233	2,401,330	104,514
0.1	Mid Capacity	High Price	2,099,970	2,499,967	105,561
0.1	High Capacity	Low Price	3,284,463	3,952,285	311,598
0.1	High Capacity	Mid Price	3,027,809	3,434,604	303,413
0.1	High Capacity	High Price	3,162,439	3,602,792	300,650
0.2	Low Capacity	Low Price	456,966	1,094,091	48,586
0.2	Low Capacity	Mid Price	543,631	1,203,878	49,044
0.2	Low Capacity	High Price	579,494	1,253,389	49,414
0.2	Mid Capacity	Low Price	917,157	2,182,294	103,058
0.2	Mid Capacity	Mid Price	1,083,354	2,400,648	105,196
0.2	Mid Capacity	High Price	1,152,338	2,499,012	106,593
0.2	High Capacity	Low Price	1,720,705	3,963,403	309,237
0.2	High Capacity	Mid Price	1,623,852	3,414,210	301,456
0.2	High Capacity	High Price	1,680,316	3,499,508	299,732
0.4	Low Capacity	Low Price	241,272	1,094,131	48,545
0.4	Low Capacity	Mid Price	304,330	1,204,065	48,857
0.4	Low Capacity	High Price	337,034	1,253,720	49,083
0.4	Mid Capacity	Low Price	480,914	2,182,180	103,172
0.4	Mid Capacity	Mid Price	607,975	2,401,170	104,674
0.4	Mid Capacity	High Price	671,477	2,499,896	105,176
0.4	High Capacity	Low Price	932,017	4,060,186	311,662
0.4	High Capacity	Mid Price	965,121	3,526,859	302,062
0.4	High Capacity	High Price	1,037,560	3,665,263	298,872
0.6	Low Capacity	Low Price	177,364	1,093,896	48,780
0.6	Low Capacity	Mid Price	233,203	1,203,991	48,931
0.6	Low Capacity	High Price	260,393	1,253,635	49,168
0.6	Mid Capacity	Low Price	353,021	2,181,903	103,450
0.6	Mid Capacity	Mid Price	462,618	2,401,070	104,774
0.6	Mid Capacity	High Price	518,982	2,500,455	105,151
0.6	High Capacity	Low Price	724,692	4,374,904	313,302
0.6	High Capacity	Mid Price	721,605	3,465,814	303,150
0.6	High Capacity	High Price	778,426	3,550,084	302,531
0.8	Low Capacity	Low Price	139,338	1,094,234	48,443
0.8	Low Capacity	Mid Price	198,504	1,203,911	49,011
0.8	Low Capacity	High Price	230,203	1,253,563	49,240
0.8	Mid Capacity	Low Price	278,778	2,182,527	102,825
0.8	Mid Capacity	Mid Price	396,549	2,401,133	104,711
0.8	Mid Capacity	High Price	458,358	2,499,785	105,458
0.8	High Capacity	Low Price	587,090	4,534,190	312,735
0.8	High Capacity	Mid Price	644,084	3,621,819	305,616
0.8	High Capacity	High Price	700,655	3,557,266	301,520