Essays in Empirical Operations Management: Bayesian Learning of Service Quality and Structural Estimation of Perishable Product Pricing and Inventory Management

by

Yan Shang

Department of Business Administration
Duke University

Date: ______________________

Approved:

__________________________
Jing-Sheng Song, Supervisor

__________________________
Andrés Musalem, Supervisor

__________________________
Bora Keskin

__________________________
Robert Swinney

__________________________
David Dunson

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University 2016
ABSTRACT

Essays in Empirical Operations Management: Bayesian Learning of Service Quality and Structural Estimation of Perishable Product Pricing and Inventory Management

by

Yan Shang

Department of Business Administration
Duke University

Date: __________________________
Approved:

__________________________
Jing-Sheng Song, Supervisor

__________________________
Andrés Musalem, Supervisor

__________________________
Bora Keskin

__________________________
Robert Swinney

__________________________
David Dunson

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2016
Abstract

This dissertation contributes to the rapidly growing empirical research area in the field of operations management. It contains two essays, tackling two different sets of operations management questions which are motivated by and built on field datasets from two very different industries — air cargo logistics and retailing.

The first essay, based on the dataset obtained from a world leading third-party logistics company, develops a novel and general Bayesian hierarchical learning framework for estimating customers’ spillover learning, that is, customers’ learning about the quality of a service (or product) from their previous experiences with similar yet not identical services. We then apply our model to the dataset to study how customers’ experiences from shipping on a particular route affect their future decisions about shipping not only on that route, but also on other routes serviced by the same logistics company. We find that customers indeed borrow experiences from similar but different services to update their quality beliefs that determine future purchase decisions. Also, service quality beliefs have a significant impact on their future purchasing decisions. Moreover, customers are risk averse; they are averse to not only experience variability but also belief uncertainty (i.e., customer’s uncertainty about their beliefs). Finally, belief uncertainty affects customers’ utilities more compared to experience variability.

The second essay is based on a dataset obtained from a large Chinese supermarket chain, which contains sales as well as both wholesale and retail prices of unpackaged
perishable vegetables. Recognizing the special characteristics of this particularly product category, we develop a structural estimation model in a discrete-continuous choice model framework. Building on this framework, we then study an optimization model for joint pricing and inventory management strategies of multiple products, which aims at improving the company's profit from direct sales and at the same time reducing food waste and thus improving social welfare.

Collectively, the studies in this dissertation provide useful modeling ideas, decision tools, insights, and guidance for firms to utilize vast sales and operations data to devise more effective business strategies.
I dedicate my dissertation work to my advisors, Prof. Jeannette Song, who supports my research from the beginning to the end and believes in me no matter how many mistakes I made. I can never finish this Ph.D. study without Prof. Jeannette Song.

I also dedicate my dissertation work Prof. Andrés Musalem, without whom I cannot finish my PhD dissertation. I thank Prof. Yiting Deng for helping me with my dissertation and supporting me in life in the last three years we spent together as doctoral students at Duke University.

I thank my dissertation committee members, Prof. Bora Keskin, Prof. Robert Swinney and Prof. David Dunson for their great support!

I dedicate my dissertation work to my father, my mother and my husband. Thank you for supporting me in all these years.
# Contents

Abstract iv  
List of Tables x  
List of Figures xi  
Acknowledgements xii  

1 Introduction 1  

2 Estimating Customer Spillover Learning of Service Quality: A Bayesian Hierarchical Approach 6  
2.1 Introduction 6  
2.2 Contributions in the Literature 11  
2.2.1 Contributions to the Empirical Service Operations Management Literature 11  
2.2.2 Contributions to the Customers’ Bayesian Learning Literature 13  
2.2.3 Contributions to the Airline Customer Satisfaction Literature 16  
2.3 Data Description 16  
2.3.1 Air Cargo Shipping Process 17  
2.3.2 Shipping Panel Data Descriptives 18  
2.3.3 Price and Other External Data Descriptives 20  
2.3.4 Exploratory Analysis 22  
2.4 Modeling Framework 23  
2.4.1 Demand Arrival 23
3 Pricing and Inventory Management of Perishable Produce: An Empirical Perspective

3.1 Introduction ............................................. 54

3.2 Data Description ........................................ 57

3.2.1 Demand and Supply Data ............................ 57

3.2.2 Wholesale and Retail Price .......................... 59

3.2.3 Instruments of Endogenous Variables ............... 61

3.2.4 Food Waste ........................................... 61

3.3 Modeling Framework ...................................... 62

3.3.1 Consideration Arrival Process ....................... 62

3.3.2 Discrete-Continuous Choice Model ................ 63

3.4 Estimation and Identification ........................... 64

3.4.1 Discrete-Continuous Choice Model ................ 64

3.4.2 Model Identification ................................ 67

3.5 Results and Discussions ................................. 67

3.6 Counterfactual Experiments and Managerial Implications .... 69
List of Tables

2.1 Data Summary Statistics .................................................. 20
2.2 Price Data Descriptives ................................................... 21
2.3 Price Forecast Model Estimates ......................................... 21
2.4 Linear Regression of Purchase Probability on Delays .............. 22
2.5 Model Comparison and Goodness of Fit ................................ 40
2.6 Learning model parameters’ population mean and 90% quantile interval 42
2.7 Rank of Goodness of Fit on Alternative Specifications of the Shipping Quality Function .................................................. 45
2.8 Estimation Results for Shipping Choice Model with Simple Bayesian Hierarchical Learning ................................................. 53
3.1 Data Summary Statistics ................................................... 58
3.2 Estimation Results ............................................................ 68
A.1 Pre-estimation Priors and Model Priors ............................... 73
A.2 Actual vs. Estimated Values of the Discrete Choice Model Parameters 77
A.3 Log-Likelihood of Alternative Model Specifications .................. 78
List of Figures

2.1 Top: updates of perceived shipping quality on each route $\mu_{at}^E$, $\mu_{bt}^E$, $\mu_{ct}^E$ and overall shipping quality represented by the hyper-parameter $\mu_t^E$. One observation on period $t = 62$ is marked; Bottom: updates of experience variability $(\sigma^2)_t^E$ and belief uncertainty $(\xi^2)_t^E$. .......................... 33

2.2 Change of purchase probability on route A and B under different learning rules. Top: spillover learning; Middle: independent learning; Bottom: information pooling learning. ................................. 49

3.1 Price vs. Wholesale Price (lettuce, broccoli, zucchini, sweet potato, ginger, onion) .................................................. 60

3.2 Distribution of the Percentage of Order Quantity Wasted .......... 62

A.1 Actual (Y axis) vs. Estimated (X axis) Values of the Multinomial Distribution Parameters ................................. 77
Acknowledgements

I am gratefully for the generous support of industry sponsors Dr. Prof. Rod Franklin and Michael Webber. I thank Prof. Yongrui Duan and Prof. Jiazhen Huo for helping me obtain the data that supports the research work of my second essay.
Introduction

With increasingly advanced information and communication technologies, we have witnessed unprecedented availability of business data on operations and marketing. As a result, both practitioners and academics are keen to know how to use these data to improve decision making. Consequently, empirical studies in the field of operations management have grown rapidly in recent years. This dissertation contributes to this emerging research area through two essays, tackling two different sets of research questions which are motivated by and built on field dataset obtained from two different industries. Because the data and the research agendas of these two essays are completely independent, below I will summarize each essay separately.


This research was motivated by the field data obtained from AlphaShip (the true name of the company is disguised for confidential reasons). AlphaShip is a world leading freight forwarder company. I obtained the first piece of data from the company in June 2012 and the last piece of data in August 2015, which spans more than 3 years. During this period of time, we had numerous discussions with the
company to identify research topics that not only are of interest to the academics but also benefit AlphaShip.

Indeed, I have explored the earlier data from AlphaShip to complete my master thesis in Statistics from the Department of Statistical Science of Duke University, on my way to the Ph.D. degree. This thesis developed a novel nonparametric Bayesian statistics method to help assess and predict transport risks, which is the deviation of the actual arrival time from the planned arrival time.

While focus of my master’s thesis was on the supply side of air cargo logistic service, Essay 1 in this dissertation focuses on the demand side. Specifically, we investigate the effect past service experiences on customers’ willingness-to-purchase from the same service provider in the future. We develop a novel and general Bayesian hierarchical learning framework for estimating customers’ spillover learning, i.e., customers’ learning about the quality of a service (or product) from their previous experiences with similar yet not identical services. We then apply our model to a one-year shipping/sales historical data from AlphaShip to study how customers’ experiences from shipping on a particular route affect their future decisions about shipping not only on that route, but also on other routes serviced by the same logistics company.

Our new method provides a parsimonious approach to model complex spillover learning processes. Our empirical application reveals several important findings:

1. Customers indeed borrow experiences from similar but different services to update their quality beliefs that determine future purchase decisions. This result is significant when estimating the overall impact of a service improvement or failure, since the corresponding impact may not be restricted to one particular service but may also spread to other related services. In this regard, our model can be used to identify the overall effect of an operational changes in all related products/services, which might be different when ignoring learning spillovers.
2. Service quality beliefs have a significant impact on the decisions about purchasing shipping services from the logistics company. Furthermore, this effect is asymmetric — customers are more sensitive to “poor” than to “good” service experiences — and that these consequences are economically significant.

3. Customers are risk averse; they are not only averse to experience variability but also belief uncertainty (i.e., customer’s uncertainty about their beliefs) and belief uncertainty affects customers’ utilities more compared to experience variability. This is an important contribution, since to the best of our knowledge, we are the first to separately estimate the effect of these two types of service quality uncertainty on customer behavior. We show that these results have important implications for designing service improvement strategies, which should be based on a careful balance between changes or improvements in the mean and variance of service quality while subjecting costs to remain within a certain level.

Essay 2 (Chapter 3): Pricing and Inventory Management of Perishable Produce: An Empirical Perspective

This research was motivated by the data obtained from BetaMart (the true name of the company is disguised for confidential reasons), a large Chinese supermarket chain. The data set contains the point of sales (POS) data of fresh vegetables from 92 stores, and the order and fulfillment information from 4 distribution centers to these stores. The observation period ranges from January 2013 to January 2014.

The data are special because we have both inventory and wholesale information. There are many empirical Operations and Marketing papers on store/product choices, but they only have information on sales. Given that freshness of perishable produce is highly valued by consumers, products in this category exhibits strong substitutability. On the other hand, because of Chinese cooking, which typically
serve several dishes in a meal with different kinds of vegetables, many of the products in the data set are also complementarity. Based on these unique characteristics of this data set, our research focus on developing structural estimation methods to aid pricing and inventory management strategies for complementary products, with a special attention to perishable products.

We develop a discrete-continuous framework that allows us to estimate the effect of produce’s freshness, price and other characteristics on demand. Based on the knowledge of demand, we further develop innovative pricing and inventory strategies to help with increasing profits through direct sales while reducing food wastes. Specifically, instead of the current one-price-all-time practice, we develop pricing strategies that are tailored to different produce. For instance, if the current high price is designed to compromise high waste cost of fast-perishable produce, then lower the price can help sell as much fresh produce at its freshest, which not only increases the overall profit for the company but is also beneficial to customers. Alternatively, we can also design price discount strategies at the certain age of the produce, estimated by the model, so that the price decrease would compromise the quality decrease perceived by customers. As a result, the produce is also sold out without being thrown. Meanwhile, we also assess the changes of current inventory management strategies by increasing or decreasing order quantity or frequencies to affect product sales and wastes. By combining the smarter inventory management strategy and pricing mechanism, we aim to reduce cost and increase revenue.

The contributions of our work to the literature are several folds:

1. Theoretical: First, on the demand side, we model store choice and product choice decisions in a coherent model framework, allowing both substitution and complementarity among perishable goods. Second, on the supply side, we obtain insights on inventory/pricing policy for perishable goods, instead of
consumer packaged goods.

2. Methodological: We develop a Poisson arrival process to capture store visit timing. This technique helps us to overcome a limitation of the data set, from which we only observe customer’s shopping histories in the focal supermarket chains but we do not know when the customers shop at competitor supermarkets.

3. New insights: Using our model and counterfactual experiments, we study the effects of freshness on demand, the degree of substitution/complementarity among perishable goods, the implications on joint inventory/pricing strategies for substitutes/complements, and the impact on waste reduction for perishable goods.

In summary, this dissertation analyzes different operations management problems from an empirical perspective, using a variety analytical tools, such as Bayesian statistics, structural estimation, maximum likelihood estimation, nonlinear optimization and simulation. Together, these works contribute to modeling, decision tools, insights, and guidelines for effective management of logistics planning, quality control, service operations and retailing operations.
2.1 Introduction

The service industry is diverse and complex and it continues to grow in prominence. Data from the U.S. Bureau of Labor Statistics indicates that the number of jobs in the service industry was twice as large as the number of jobs in manufacturing in 1999. By January 2016, this ratio has increased to 6:1, with service jobs accounting for 86% of the total U.S. employment. Furthermore, this industry represents 63% of the global economy and 80% of the U.S. GDP. As a result, the need for effective service operations management strategies has grown rapidly in the past decades.

One of the central goals of service operations management is to offer high levels of customer satisfaction while maintaining a low cost. In this regard, the choice of service levels can be seen as a balancing act: higher service levels improve customer satisfaction and demand, however these are also likely to increase service costs. Therefore, understanding how different service levels affect customer satisfaction and service costs is critical. Most work in the operations management literature has fo-
cused on developing models that are useful to quantify and minimize the operational costs of attaining a given level of service quality. Because these operational costs are more salient, it is frequent in practice to observe service operations rules designed to attain a quantifiable target service level at a minimum cost. For example, in the air cargo shipping industry, which provides the empirical setting for this paper, all international cargo shipments sent to the United States need to be screened before being loaded into the respective aircraft. Because of this requirement, it is common to allocate additional staff for cargo security screening, in case the screening time exceeds a certain target service level, so that shipments can be loaded on time. However, there is not much research focusing on how to choose an appropriate target service level. This requires measuring the value that customers assign to objective service level improvement and how this additional value translates into incremental revenues and profits. The focus of this paper is therefore to measure the effect of service levels — in particular, international air cargo on-time delivery — on customers’ decisions in a context where customers learn about the quality of the service through repeated encounters with the service provider.

An important feature of services is that typically their quality can only be fully observed when the service is finished. In our air cargo shipping context, customers do not know whether their cargo will be delivered to the destination on time since the actual delivery performance can only be assessed after the cargo has been delivered. This uncertainty makes most random utility frameworks unsuitable since they typically assume that customers know the attributes of all alternatives perfectly before making a choice, such as in Aksin et al. (2013). In contrast, a learning model is more suitable to our context (e.g., Ching et al. 2013), since they postulate that customers have incomplete information about product attributes. Specifically, customers form expectations about these attributes from one or several information sources such as past experiences with the product, word of mouth and advertisements, and then use
their beliefs about service performance in their purchase decisions. In this paper, we follow this approach and study customers’ learning process about air cargo shipping service quality (i.e., on-time delivery). We use one year of shipping and sales data from individual customers of a world-leading third-party logistics company. The data set contains information about the quality of the service (e.g., delivery times and lengths of delays) received by customers each time they interacted with this provider. This is an advantage over most quality learning models in the empirical literature, where service quality for each experience is not observed and hence is treated as a latent construct (e.g., Erdem 1998).

Furthermore, most studies using learning models assume that the information about a certain product comes exclusively from experiences with the same product. Such assumption works well when a consumer has had multiple and frequent experiences with each product. When this assumption is not likely to hold, consumers might learn from multiple experiences with similar, but not necessarily identical products. We refer to this belief updating mechanism as “spillover” learning. In our application, assume for example that a customer has ordered cargo shipping services several times from a logistics company on route\(^1\) from airport \(A\) to airport \(B\) and that the flights have been either delayed at the departure or the cargo has failed to be loaded onto the flight. Then, when the customer considers sending a new shipment from airport \(A\) to airport \(C\), she may have an expectation that this shipment might also be delayed since the two paths, \(A \rightarrow B\) and \(A \rightarrow C\), share the same origin airport. If instead her past experiences shipping on the \(A \rightarrow B\) path had been mostly on time, then she might expect the shipment from \(A\) to \(C\) will be more likely to be delivered on time. In the extant literature, a similar phenomenon has been studied (e.g., Erdem (1998); Sridhar et al. (2012)). For example,

\(^1\) In this study, a route is defined as a directed pair from origin airport \(A\) to destination airport \(B\), \(A \rightarrow B\), and it is different from \(B \rightarrow A\).
in Erdem (1998), the author considers a model where the priors are correlated across “umbrella brands” (i.e., brands that operate in multiple categories) through the use of a covariance matrix, and finds evidence that consumers learn from experiences across umbrella brands in the toothpaste and toothbrush categories. In this paper, we explore a new methodology that relies on a Bayesian hierarchical model, which yields a more parsimonious formulation and can be easily extended to accommodate a variety of learning spillover situations. This is also possible since we have access to quality information and hence we don’t need to impute quality information as in the case of a latent variable.

Another feature of our data set is that the service quality measures correspond to objective metrics (e.g., delivery times). This is important since most of the extant research studying customer sensitivity to service quality relies on survey data (e.g., Taylor 1994). Even though surveys are useful since they provide information about how customers perceive the quality of the service, they are less helpful when making operational decisions (e.g., deciding service capacity or service levels) since it is not always evident how to link subjective service quality assessments to an operational decision or service level. Finally, another interesting aspect of our application is that our dataset considers a business to business (B2B) setting, while most of the existing literature has consider business to consumer empirical settings.

As previously mentioned, we apply our methodology to model customers in the international air cargo transport industry. International cargo logistics management plays an increasingly important role in the global economy as global trade has grown considerably in recent decades and many companies now have overseas facilities and supply chain partners. Air transport delivers goods, that are time-sensitive, expensive, perishable or used in just-in-time supply networks, at competitive prices to customers worldwide. Indeed, the air cargo industry transports goods worth in excess of $6.4 trillion annually. This is approximately 35% of world trade by value.
This industry, including express traffic, is forecasted by Boeing to grow on average 4.7% annually in the next two decades, reaching a total of more than twice the number of revenue ton-kilometers (RTK) logged in 2013. However, attention paid to this industry is surprisingly small: air cargo industry ‘.. has remained the poor cousin to the more glamorous passenger side of the business (passenger air transport industry)’ (Morrell, 2011). According to a 2011 survey of the global management consulting firm PRTM, 69% of companies mentioned improving delivery performance as their top supply chain management strategy. In a 2010 report of Infosys, “carrier delays and non-performance on delivery” is ranked as the leading risk in the logistics industry. Furthermore, in a 2014 survey conducted by the International Air Transport Association (IATA) to major freight forwarders and their customers, low reliability is perceived as the second most important factor (next to transportation cost) driving the transport model shift from air to sea.

As we will show in our empirical application, our study suggests that customers indeed borrow experiences from similar but different services to update their quality beliefs that determine future purchase decisions. This result is important when estimating the overall impact of a service improvement or failure, since the corresponding impact may not be restricted to one particular service but may also spread to other related services. In this regard, our model can be used to identify the overall effect of an operational changes in all products/services, which might be different when ignoring learning spillovers. In addition, we also find that service quality beliefs have a significant impact on the decisions about purchasing shipping services from the logistics company. Furthermore, our results imply that this effect is asymmetric — customers are more sensitive to “poor” than to “good” service experiences — and that these consequences are economically significant. Similar to the results in the previous literature (e.g., Erdem 1998), we also find customers to be risk averse; customers are not only averse to experience variability but also belief uncertainty
(i.e., customer’s uncertainty about their beliefs, see §4.3 for the definition and details) and belief uncertainty affects customers’ utilities more compared to experience variability. This is an important contribution of our paper, since to the best of our knowledge, we are the first to separately estimate the effect of these two types of service quality uncertainty on customer behavior. We show that these results have important implications for designing service improvement strategies, which should be based on a careful balance between changes or improvements in the mean and variance of service quality while subjecting costs to remain within a certain level.

The rest of the paper is organized as follows. We discuss the relevant literature in Section 2.2. In Section 2.3 we describe the empirical setting (e.g., data, exploratory analysis). Section 2.4 presents the learning and demand model and discusses the identification of the model parameters. In Section 2.5 we report the results. Section 2.6 illustrates managerial implications of our model and empirical results through counterfactual experiments; and Section 2.7 concludes the paper and discusses future research directions.

2.2 Contributions in the Literature

Our study contributes to three streams of research. In this section we provide a brief review of the literature that is most closely related to our study in each of these streams.

2.2.1 Contributions to the Empirical Service Operations Management Literature

Recent empirical studies have explored the problem of estimating the effect of service quality on demand. Brown et al. (2005) and Aksin et al. (2013) study customers’ abandonment behavior in the context of call centers, whereas Batt and Terwiesch (2015) study patients abandonment behavior in the context of hospital emergency room. In addition, Allon et al. (2011) quantify the effect of waiting time on sales,
using market level data from the fast food industry; and Lu et al. (2013) study the effect of waiting time and queue length on customers’ purchase decisions in a supermarket deli. Other studies have focused on different service quality measures including delivery time in quick service restaurants (Cho et al., 2015), signal quality in video streaming services (Sridhar et al., 2012), on-time performance of airlines (Grewal et al., 2010) and order fill-rate in a supplier-retailer supply chain (Craig et al., 2014).

Our paper differs from the previous studies in that we explicitly model the customers’ learning process about service quality, such that their beliefs about the quality of future experiences may depend on past experiences. The previous empirical work in service operations management typically uses a static approach to consider customers’ beliefs about service quality. For example, Lu et al. (2013) assume that customers form beliefs about service quality when they arrive at the physical queue based on the current observed length of the queue; Aksin et al. (2013) assume the customers (i.e. callers) correctly anticipate service quality (i.e., the probability of receiving service in a certain period) based on their past experiences and that this service quality is common knowledge among all customers; Batt and Terwiesch (2015) considers a hospital emergency department and argues that it is not clear what customers (i.e. patients) can learn from what they observe while waiting, so they investigate how the factors that customers observe and experience in the waiting room impacts their abandonment behavior. Instead of a relying on this static assumption about consumer beliefs, we explicitly allow consumer expectations to evolve over time as each customer gains additional experiences with the service provider. Furthermore, allowing the customer to form beliefs based only on what he observes immediately before choosing a service provider can be a reasonable assumption when the service environment is observable (e.g., when the customer observes the length of the queue as in Lu et al. (2013)). However, this is not always the case as in the case
of call centers, where a caller does not directly observe the state of the queue, and also in our air cargo application, where a shipper chooses a service provider without observing the congestion level that his shipment will face once the cargo is picked up. Hence, our paper contributes to the empirical operations management literature by considering and modeling customer learning, which is an important driver of customer behavior in many service operations scenarios, particularly in cases where the service environment is not observable to the customer.

2.2.2 Contributions to the Customers’ Bayesian Learning Literature

Our paper uses a customer Bayesian learning modeling framework. Ching et al. (2013) provide an extensive literature review and identifies four categories of recent developments. Our work falls into the category: models of correlated learning. Ching et al. (2013) summarize the few existing studies in this stream (e.g., Erdem 1998; Coscelli and Shum 2004; Sridhar et al. 2012; Chan et al. 2013) and indicates that correlated learning means “… learning about a brand in one category by using the same brand in another category and/or learning about one attribute (e.g., drug potency) from another (e.g., side effects). This occurs if priors and/or signals are correlated across products or attributes …”. Previous literature has explored the information spillover across multiple categories under the same brand (e.g., Erdem 1998) or type (e.g., Sridhar et al. 2012), or the information spillover among attributes of a multi-attribute product (e.g., Coscelli and Shum 2004; Chan et al. 2013). Our study explores the information spillover from past shipping service experiences, specifically the shipping experiences on multiple routes, of the same logistics company.

Our paper differs from previous studies of correlated learning in several important aspects. First, we directly observe objective service level metrics (e.g., the actual and planned time when the cargo is delivered at the destination) while none
of the previous studies observe this information. Being able to observe objective service level significantly lowers our computational burden, compared to the previous studies, by separating the estimation of customer learning from the estimation of purchase model and eliminating the need of integrating out unobserved service quality. Therefore we are able to estimate the learning process about the quality of $1000^+$ products, a dramatically larger number compared to $2 \sim 7$ products usually seen in the learning model literature, and obtain results within a reasonable computational time. Moreover, having access to quality data enables us to explore the potential impact of different quality measures (see §4 for details), beyond the classic measures of perceived quality mean and risk, on customers’ purchase decisions.

Second, the Bayesian hierarchical model provides a parsimonious and structural (i.e., theory-based) approach to characterize the information spillover process, and enables us to model the correlation among the true qualities of the products for the first time in the literature.

We note that previous studies of correlated learning rely on estimating the covariance matrix of customers’ quality priors (e.g., Erdem 1998; Coscelli and Shum 2004) and/or a covariance matrix of noise in quality signals (e.g., Sridhar et al. 2012; Chan et al. 2013; Coscelli and Shum 2004). The number of parameters in these approaches rapidly (i.e., quadratically) grows with the number of products for which spillovers are allowed. In contrast, our model provides a more parsimonious approach to allow for information spillovers among large numbers of products. Furthermore, in our approach it is straightforward to allow for information spillovers based on multiple product characteristics (e.g., distance and weight). This can be easily modeled embedding a linear model of quality as a function of product characteristics within the Bayesian learning framework. Finally, and in contrast with previous approaches, because we have access to quality data we can allow the true quality of the same service (or product) to be different for different customers. This is relevant in services, since
a provider may set different service levels for different customers (e.g., giving priority to high volume customers). Note that without access to quality data it would be very challenging to separately estimate heterogeneity in service quality sensitivity and heterogeneity in true service quality. In fact, to our knowledge all previous approaches have assumed that the quality of a product or service is homogeneous across all customers.

Third, we relax the constant experience variability and belief uncertainty assumption that are used in the previous correlated learning literature. Specifically, existing correlated learning models assume that consumers are uncertain about the mean quality of a product or service, but that the precision of quality signals (e.g., usage experience, advertising) is constant and known to consumers. Therefore, consumers update their mean quality beliefs after new information arrives (e.g., Erdem 1998; Coscelli and Shum 2004; Sridhar et al. 2012; Chan et al. 2013). The implication of this assumption is that as a consumer receives more information, his uncertainty about the mean quality of the product is lower (i.e., the consumer’s perceived risk associated with the product). In general, this is true when the new information that a customer receives about a product is relatively congruent with his or her prior knowledge of the product. However, when this new information is markedly inconsistent with the customer’s prior beliefs, it could raise rather than reduce the consumer’s uncertainty associated with the mean quality level. This is relevant for our application since severe transport disruptions (e.g., a delay longer than a couple of days) happen much more often than large quality changes for physical goods.

Accordingly, our model allows consumers to update their beliefs about not only the mean product quality but also about the precision of the quality signals. This feature yields a richer and more realistic description of the customers’ learning process.
2.2.3 Contributions to the Airline Customer Satisfaction Literature

There is a large body of literature aiming to relate airline service quality to microeconomic factors (e.g., price, competition, firm merging) and another important stream that seeks to forecast flight delays using statistical methods (refer to Deshpande and Arikan 2012 for a detailed literature review). Different from our interests in air cargo transport service, almost all of these studies have focused on passenger airlines, with Shang et al. (2016) as an exception, in which the authors also studied air cargo transport delivery performance and develop a Bayesian nonparametric method aiming to help forwarders to forecast transport delays thus selecting better service suppliers.

There are a few studies sharing a similar interest to our paper in terms of studying the customer-side consequences of airline transport service performance. Taylor (1994) relies on a survey of delayed airline passengers and finds that delays decrease service evaluations by invoking uncertainty and anger reactions. Forbes (2008) analyzes the impact of airline delays on customer complaints, showing that customer expectations play an important role mediating this effect. Anderson et al. (2008) uses survey data to investigate customer attributions during service failures (i.e., flight delays) and find customer satisfaction is significantly lower when delays are of internal origin (i.e., airline's fault) and the customer blames the service provider (i.e., the airline). Our paper contributes to this literature by providing evidence for the air cargo industry, which suggests that customers and their beliefs are more sensitive to "poor" rather than "good" service quality. We also provide evidence that customers are averse to service experience variability.

2.3 Data Description

We first describe the air cargo shipping process in §3.1, identifying the different parties that participate in this industry and their roles. In §3.2, we describe the main
features of data set that was made available to us by a world leading freight forward company, hereafter *AlphaShip* (the true company name is disguised for confidential reason). Additional information was collected by us from Internet sources, and described in §3.3. We conclude the data description in §3.4 providing a exploratory analysis that presents model-free evidence of spillovers in customer learning.

2.3.1 *Air Cargo Shipping Process*

The air cargo industry, though a crucial part of global operations, is less known to the public because it operates behind the scenes. Typically, an air cargo transport involves four parties: shippers (e.g., manufacturers), freight forwarders (forwarders in short), carriers (i.e., airlines) and consignees (e.g., downstream manufacturers or distributors). These four parties form a chain structure, commonly called the air transport supply chain. The shipping process starts with a request from the shipper to the forwarder with certain shipping needs, such as origin and destination cities, collection and delivery date, and cargo information (pieces, weight and volume). The forwarder, who typically has reserved spaces from airline partners, is often able to tell the shipper immediately whether it has shipping space available that fits the customer’s requirements; and if so, the forwarder will provide the customer with a route map, i.e. a shipping proposal that includes flight numbers and airline information. The route map may be modified a few times before both the shipper and the forwarder agree on a final version as well as on the shipping fee. Then, the shipping process may start, which consists of three stages: (1) the door to airport stage (D2A): the forwarder picks up cargoes from the shipper at the required time, consolidate cargoes sharing the same route if possible, and then sends cargoes to the selected airline at an origin airport; (2) the airport to airport stage (A2A): the airline is in charge of the cargoes until they arrive at the destination airport; (3) the airport to door stage (A2D): the forwarder accepts cargoes at the destination, and delivers
them to the consignees. As can be seen from this process, the shipper decides which forwarder company to choose and pays the shipping fees (in most of the cases), so we referred to it as the customer in what follows.

As an intermediate controller, the forwarder not only has power but also bears responsibility. In this regard, the forwarder is the service provider and direct contact for its customers. Its responsibilities range from making the first route map to updating the customer with alternative shipping plans if the shipment does not proceed as planned (such as if the flight took off without loading the cargoes as planned). As a result, if cargo is not delivered on-time and hence customer satisfaction drops, the forwarder faces the risk of losing the opportunity of conducting business with this customer in the future. At the same time, the forwarder relies on the services provided by airlines. Accordingly, the forwarder has no control over the A2A shipping stage, however it bears the consequences of poor shipping service quality (e.g., customer complaints). So from the perspective of the forwarder, managing customer satisfaction through the entire shipping process is an important operational goal.

2.3.2 Shipping Panel Data Descriptives

Our data is provided by a world leading freight forwarder company, AlphaShip, and records its air freight shipments in 2013 from January to December. One unique feature of this data set is that it follows the Cargo 2000 standard, an air cargo industry information system standard that was initiated in 1997 and currently widely used across the entire industry (adopted by 80 major airlines, freight forwarders, GHA’s, trucking companies and IT providers). The Cargo 2000 standard was designed to enable industry-wide participants “to implement processes, backed by quality standards, that are measurable and supported by data, thereby improving the efficiency

---

2 Since more than 90% of the air cargo shipments are handled by forwarders, we do not discuss the less frequent situations where the airlines deal with air cargo shipments directly without an intermediate forwarder.
of air cargo, enhancing customer service levels and reducing operational costs ...” (IATA, 2014). A major advantage of this data set is that it records not only the actual delivery time at each milestone, but also the planned latest-by time for each milestone (see Shang et al. 2016 for more details). Thus, by comparing the actual delivery time against the planned delivery time, we can easily obtain the objective on-time delivery performance for each shipment: whether the shipment is delayed and how many hours it is delayed. Borrowing the concept in Shang et al. (2016), we use a key (service) performance measure, transport risk, as the primary measure of service quality of a shipment. Since this term might be misinterpreted as related to uncertainty, we will refer to this measure as transport delay. Accordingly, for each shipment, the transport delay, is defined as the deviation of the actual final delivery time at the destination from the planned delivery time:

\[
\text{transport delay} = \text{actual delivery time} - \text{planned delivery time}
\]

Neither earliness (i.e., a negative transport delay) nor tardiness (i.e., a positive transport delay) is desirable for customers. While tardiness causes delays in production and/or product delivery to all downstream customers, earliness incurs additional storage and handling costs.

The data set contains customer IDs, which allows us to track customer purchases and model customers’ learning about transport delay from their experiences with AlphaShip. After data cleaning and selecting customers with enough observations for model estimation (see Appendix A.1.1 for details), there are 725 customers and 26,045 shipments left in the data. The cargoes are transported from 53 countries to 153 countries on 2,897 routes all over the world. Table 2.1 provides more details on the data, including the choice predictors we use in our models (refer to §4 for details) and presents customer level statistics. As we can see from Table 2.1, even though the average transport delay is negative (i.e. cargoes delivered earlier than
Table 2.1: Data Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (std. dev)</th>
<th>Per customer Statistics</th>
<th>Mean (std. dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport delay (hours)</td>
<td>-0.77 (3.51)</td>
<td>Number of shipments</td>
<td>41.83 (13.78)</td>
</tr>
<tr>
<td>Chargeable weight (kg)</td>
<td>1274.07 (4785.73)</td>
<td>Number of frequent routes</td>
<td>3.25 (1.00)</td>
</tr>
<tr>
<td>Pieces</td>
<td>1.13 (0.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance (km)</td>
<td>7758.25 (3341.17)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

planned time), the variance is significantly large, meaning that the actual shipping service levels vary dramatically from shipment to shipment. Actually, 26.0% of the shipments are delayed and 9.2% of the shipments are delayed more than 2 hours (regarded as transport disruption by industry standard, according to our discussion with the company executives).

2.3.3 Price and Other External Data Descriptives

Due to its sensitive nature, AlphaShip cannot provide us the shipping fee associated with each shipment. In order to reconstruct prices, we train a forecasting model by merging two external datasets. The first is a sample of the official price data provided by IATA. The data contains price information on routes from six airports in the Netherlands to almost all the other airports in the world (846 airports), spanning 12 months in 2013; in addition, the chargeable weight and weight break associated with the prices are also included. However, the number of routes included in the IATA sample is much smaller than that in our data, so we seek additional data from a shipping service quotation website www.worldfreightrates.com. Specifically, we build a Python program to crawl shipping prices from the website using the input (i.e. route, chargeable weight) extracted from our Cargo 2000 data. Thus, we are

---

3 The IATA official price might be slightly different than the actual shipping price however should be very close.

4 Chargeable weight = max{volume weight (kg), actual weight (kg)}; volume weight = volume (cubic meter)/6.
Table 2.2: Price Data Descriptives

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean (std. dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(Price) ($)</td>
<td>1</td>
<td>6.7 (1.4)</td>
</tr>
<tr>
<td>Distance (km)</td>
<td>1</td>
<td>7201.8 (3534.2)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>1</td>
<td>962.4 (5077.1)</td>
</tr>
<tr>
<td>Weight_break</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Month</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>To_country</td>
<td>220</td>
<td></td>
</tr>
<tr>
<td>Pieces</td>
<td>71</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3: Price Forecast Model Estimates

<table>
<thead>
<tr>
<th></th>
<th>Estimate (std. err)</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 ) (Intercept)</td>
<td>4.50 (7.69e-03)***</td>
<td>584.90</td>
</tr>
<tr>
<td>( \gamma_1 ) (Distance)</td>
<td>1.15e-5 (3.26e-07)***</td>
<td>35.32</td>
</tr>
<tr>
<td>( \gamma_2 ) (Weight)</td>
<td>1.65e-6 (1.65e-07)***</td>
<td>10.00</td>
</tr>
</tbody>
</table>

F-statistics (331, 473,520) = 1.62e+04  
\( R^2 = 0.914; \) Adjusted \( R^2 = 0.914 \)

\* \( p < 0.05; \) \* \* \( p < 0.01; \) \* \* \* \( p < 0.001. \)

We are able to obtain price information for the majority of our routes and weight range. We crawled this website twice, in September 2014 and February 2016 respectively, and each process took around 15 days. Then, we merged the data from the two sources to train the price forecasting model. The following equation shows the best fitting forecasting model:

\[
\ln(P) = \gamma_0 + \gamma_1 \cdot Distance + \gamma_2 \cdot Weight + Weight\_break + Month + To\_country + Pieces \quad (2.1)
\]

More details about the data and the estimation results can be found in Table 2.2 and Table 2.3. By using the coefficients estimated from Equation (2.1), we forecast the price for each shipment obtained from AlphaShip.

We note that the Distance variable is not included in either the Cargo 2000 data nor the IATA price sampler. Instead, we obtained latitude and longitude information for all the airports in our data set and calculated the great-circle distance between the
original and destination airport as the approximate distance for that route. Although this great-circle distance is shorter than the actual flight distance (especially if the shipping process includes multiple flights), it proves to be useful not only in predicting shipping price but also in specifying the learning and shipping choice model that we describe next.

2.3.4 Exploratory Analysis

In this subsection, we are interested in investigating how customers use past experiences, especially experiences from similar yet not identical services, to form quality beliefs which further affect their purchase decisions later. Before formulating a structural learning model (explained in the next section), we would like to provide evidence from simple descriptive models. More specifically, for each customer $i$, we pick route $r_i$ that this customer has shipped the most times during the 1-year sample period, then construct the dependent variable as a Bernoulli variable, $y_{it}$, where $y_{it} = 1$ if customer $i$ ships on route $r_i$ during period $t$, and zero otherwise. Then we use the average transport delay customer $i$ has experienced on route $r_i$ until period $t$, $\text{averag}_{TR\_on\_this\_route}$, and the average transport delay customer $i$ has experienced on all other routes, $\text{averag}_{TR\_on\_other\_routes}$ as independent variables. Besides, we control for customer heterogeneity by adding customer fixed effect (724 of them) and control by seasonality by adding month fixed effect. Results are reported in Table 2.4. It is not surprising that high average transport delay (e.g., more delays or less earliness) on route $r_i$ lowers customer $i$’s likelihood of shipping on

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate (std. err)</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{averag}_{TR_on_this_route}$</td>
<td>-9.92e-1 (3.89e-02)**</td>
<td>1.92</td>
</tr>
<tr>
<td>$\text{averag}_{TR_on_other_routes}$</td>
<td>-9.65e-1 (1.72e-02)***</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Pseudo $R^2 = 0.08$

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. 

Table 2.4: Linear Regression of Purchase Probability on Delays
this route in the future. What is interesting is that experiences on other routes also affect the customer’s purchase decisions on the focal route \( r_t \). This provides evidence consistent with customers forming quality beliefs using past experiences, which in turn affect their future decisions. In addition, the relevant experiences are not only those from the same service, since experiences with other services can spillover and affect future purchasing decisions.

2.4 Modeling Framework

We formulate a model comprised of three interrelated components: demand arrival, shipping choice decision and Bayesian learning about service quality. We explain the details of each component in §4.1, §4.2 and §4.3, respectively.

2.4.1 Demand Arrival

One limitation of our data is that it only includes the observed purchases of shipping services performed by AlphaShip. In other words, we do not observe data for cases where the customer contacted AlphaShip to find out fees for shipping services but decided not to rely on AlphaShip (i.e., using a competitor or simply not shipping during that period). Similarly, our data does not separately identify periods where a customer did not have a shipping need. Both scenarios from the perspective of AlphaShip correspond to a no-purchase observation.

In order to overcome this difficulty, we first borrow the modeling approach used in Newman et al. (2014). Specifically, we divide the whole time horizon into a series of small discrete time slices during which a customer arrival may or may not be observed (an “arrival” means contacting AlphaShip for a shipping service quote, which does not necessarily lead to a final purchase). The slices are small enough so that the probability of two or more shipping demands from one customer arriving during the same time period is negligible. Arrivals are exogenous and modeled as a
Bernoulli process, and the arrival probability (denoted by $\lambda_i$, where $i$ is the subscript of customers) is the probability that a customer arrives to AlphaShip in a given time period within the booking horizon. In this study, we choose half a week (i.e., 3.5 days) as a period. This choice leads to a low probability of multiple arrivals\(^5\).

A customer $i$ arriving at AlphaShip may be interested in shipping on any route in his “frequent routes set”, $\mathcal{Y}_i \subset \Delta$, where $\Delta$ is the set of all the routes (2000+ routes) in the data set. As discussed in Section 2.3, we focus on customers with a set between 2 and 10 frequent routes. Furthermore, upon arrival the probability that he is interested in shipping on route $j$ is given by $m_{ij}$, where $\sum_{j \in \mathcal{Y}_i} m_{ij} = 1$ and $m_{ij} = 0$ for $j \in \Delta/\mathcal{Y}_i$ (i.e., the choice of routes follows a multinomial distribution).

Defining $d_{ijt}$ as an indicator which equals to 1 if customer $i$ is interested in shipping on route $j$ in period $t$ using AlphaShip and 0 otherwise, then:

$$\text{Prob} (d_{ijt} = 1) = \lambda_i m_{ij}. \quad (2.2)$$

### 2.4.2 Shipping Choice Model

A customer $i$ with a shipping demand on route $j$ in period $t$ needs to decide whether to use AlphaShip or not (i.e., to use a competitor or not to ship). The utility that customer $i$ derives from using AlphaShip is given by (we use bold symbols for parameter vectors and regular symbols for scalars):

$$U_{ijt} = \beta_i^0 + \beta^p \cdot \text{Price}_{jt} + \beta^x \cdot X_{jt} + f(I_{it}, \beta_i^q) + e_{ijt}, \quad (2.3)$$

while the utility of the outside option (i.e., not using AlphaShip) is normalized to $U_{i0t} = e_{i0t}$. Here $e_{ijt}$ captures idiosyncratic preferences of the customer unobserved to the researcher. In Equation (2.3), we use an individual-level intercept $\beta_i^0$ to control for customer heterogeneity. We use the shipping price, $(\text{Price}_{ijv})$, as a factor influencing

\(^5\) For periods when multiple arrivals from the same customer are observed (only 10.5% among all the customer-period), we choose to keep the route with the highest total arrival times from that customer.
customers’ utility, which allows us to estimate customers’ price sensitivity and to put a dollar tag on the cost of service quality. Since in most cases, customers call the forwarder’s local branch or use the forwarder’s online price quotation tool to get shipping prices before making purchase decisions, we assume that prices are known to customers. However, due to lack of pricing information, we use the price forecasted by model (2.1) in §3 to impute prices. Furthermore, $X_{ijt}$ is a vector of controls which include: (i) customer consumption rate $CR_i$; (ii) cargo chargeable weight ($Weight$); (iii) monthly dummies (e.g., April, May) to control for seasonality. In addition to the random intercepts in the model, we control for each customer’s average purchasing propensity by including a covariate measuring the average consumption rate, $CR_i$, using a pre-estimation sample as done by Lu et al. (2013). Accordingly, the random intercepts $\beta_0^i$, measure unobserved heterogeneity across customers in their propensity to use AlphaShip after controlling for differences in the consumption rate. One important element in Equation (2.3) is the parametric function, $f(\cdot)$, that captures the impact of customer’s beliefs about service quality contained in his information set $I_{it}$. $\beta_q^i$ is the function’s parameter vector to be estimated. We will discuss the functional form for $f(\cdot)$ in §4.3.3 after having described the customers’ service quality learning process in §4.3.1 and §4.3.2.

Assuming a standard extreme value distribution for $e_{ijt}$, the random utility model described by Equation (2.3) becomes a random coefficient logit model. Let $y_{ijt}$ be an indicator that is set to 1 if consumer $i$ purchases shipping services on route $j$ at time $t$, and 0 otherwise. Then the probability of a purchase, $y_{ijt} = 1$, given a demand $d_{ijt}$ is:

$$P_{ijt} = \text{Prob}(y_{ijt} = 1 | d_{ijt} = 1) = e^{U_{ijt}}/(1 + e^{U_{ijt}}), \quad (2.4)$$

where $\bar{U}_{ijt}$ is the deterministic component of $U_{ijt}$: $\bar{U}_{ijt} = U_{ijt} - e_{ijt}$. The model in Equation (2.3) includes customer-specific (random) intercept and coefficients for
the terms associated with the effect of the shipping service quality \((\beta_i^0, \beta_i^q)\). These random coefficients are assumed to follow a multivariate normal distribution with mean \((\beta^0, \beta^q)\) and for simplicity a diagonal covariance matrix \(\Omega\), which we seek to estimate from the data.

2.4.3 Customer Learning Spillover

In this subsection we describe information spillovers under which a customers’ experience with one particular service offered by the firm may offer relevant information about the service quality of other similar (yet not identical) services from the same firm. Throughout, we model this learning process as being \textit{independent across customers}. Therefore, we describe the learning mechanism for customer \(i\), omitting the subscript \(i\) in most of the equations below for ease of exposition.

As discussed in §3.4, we focus on transport delay, denoted by \(Q\), as the primary service quality measure that the customer learns about. We will refer to \(Q\) as tardiness if \(Q > 0\) and earliness if \(Q < 0\) to avoid possible confusions. Since the transport process consists of a series of multiple stochastic events (i.e. multiple connected flights, loading and unloading at the airports) and since \(Q\) can be either positive or negative, we assume that \(Q\) follows a normal distribution:

\[
Q_{jt} \sim N(\mu_{jt}, \sigma_{jt}^2)
\] 

where \(\mu_{jt}\) and \(\sigma_{jt}^2\) are the true mean and variance of service quality on route \(j\) for a customer’s visit during period \(t\); \(Q_{jt}\) is the customer’s experienced service quality, which he observes after the cargo has been delivered at the destination. Note that different customers may have different true quality mean and variance (we have omitted the consumer subscript in \(Q_{jt}\) and its moments only for ease of exposition). We assume that customers do not know the mean quality level \(\mu_{jt}\) due to “inherent product variability” (Ching et al., 2013) \(\sigma_{jt}^2\), which is referred to as “experience
variability" in the learning literature. Customers form beliefs about both the mean quality level $\mu_{jt}$ and experience variability $\sigma^2_{jt}$ from the signals $Q_{jt}$ received from their own usage experiences. Different from previous research data used in learning models, both customers and the researchers (us) observe the objective transport delay, $Q_{jt}$. This data advantage enables us to explore more flexible and general learning models and also allows us to relax strong assumptions used in many previous literature, such as rational prior information (Chan et al., 2013) or full knowledge of experience variability $\sigma^2$ (Sridhar et al., 2012). We will explain these in more details in §4.3.1 and §4.3.2.

**A Simple Bayesian Hierarchical Model for Customers’ Learning**

First, we consider a simple Bayesian Hierarchical learning model (referred as the “simple model” in the following context) to characterize customer learning spillovers. Throughout, we adapt the indexing convention where subscript $t$ denotes the beginning of period $t$. As a simplification of Equation (2.5), customers assume that the true mean is constant across periods, i.e. $\mu_{jt} = \mu_j$ and hence it is only differentiated across routes $j$. Moreover, we assume that the variance $\sigma^2_{jt} = \sigma^2$ is the same for all service types and periods. Accordingly:

$$Q_{jt} \sim N(\mu_j, \sigma^2)$$  \hspace{1cm} (2.6)

Since the service on all routes is arranged by the same logistics firm, one might expect that the reliability of the service for one route might offer information about the reliability of the service on another route. Mathematically, a Bayesian hierarchical model allows for information borrowing among the true qualities for different routes by adding one more layer to the model. Here we assume that the true qualities $\mu_j, \forall j \in \mathcal{Y}_i$, are generated exchangeably from a common population, with a
distribution governed by the hyper-parameters $\mu$ and $\xi$:

$$
\mu_j \sim N\left(\mu, \xi^2\right), \forall j \in \mathcal{Y}_i.
$$

The hyper-parameter $\mu$ provides the grand mean of qualities across all routes, while $\xi^2$ measures the degree of heterogeneity in service quality across routes. This parameter is related to the degree of shrinkage of route qualities towards the grand mean. Smaller values of this parameter imply that route qualities are more similar and hence experiences for one route become more informative about the quality of another route.

Furthermore, we assume these hyper-parameters ($\mu$ and $\xi^2$) are not known by consumers. Hence, their prior beliefs about these parameters are modeled using a Normal-Gamma distribution (this facilitates computations due to its conjugacy properties). The consumer also doesn’t know how variable experiences are ($\sigma^2$), so an inverse gamma prior is used to model consumer beliefs about this variability. Accordingly, at the beginning of period $t = 1$, customer $i$ has the following initial beliefs:

$$
\sigma^2 \sim IG(\alpha_\sigma, \delta_\sigma), \quad \mu \sim N\left(\mu_0, \sigma^2_\mu\right), \quad \xi^2 \sim IG(\alpha_\xi, \delta_\xi)
$$

Here $IG(\alpha, \beta)$ represents the inverse-Gamma distribution such that for $x \sim IG(\alpha, \beta)$ the pdf is \(f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\beta/x)\) and the expectation is $E[x] = \frac{\beta}{\alpha-1}$.

Since our data presents a left truncation problem (i.e., most of the customers have shipped with AlphaShip before our observation period starts), we cannot assume the customers’ priors to be the same during the first period in our data. Borrowing a solution used in previous studies (e.g., Mehta et al. 2004 and Zhao et al. 2011), we use the first 24 periods (around 3 months) of data as a pre-estimation sample and assume that all customers have the same prior beliefs about experience variability ($\sigma^2$), the grand mean of product quality ($\mu$) across routes and the quality heterogeneity
across routes ($\xi^2$) at the beginning of the pre-estimation sample, such that $\sigma^2 \sim IG(1.25, 100)$, $\mu \sim N(0, 30^2)$ and $\xi^2 \sim IG(1.25, 100)$. Here, contrary to the widely used assumption of “rational expectation of priors” (e.g., Chan et al. 2013), we use proper but relatively flat priors. This adds very weak assumptions to the estimation while also avoiding the Bartlett-Lindley paradox caused by improper priors where frequentist and Bayesian estimations can yield qualitatively different conclusions. Based on this formulation, we calibrate individual learning using the pre-estimation sample and calculate the posterior distribution of $\{\sigma^2, \mu, \sigma^2\}$ for each customer at the end of the 24th period. Then we use these posteriors as the corresponding prior for $\{\sigma^2, \mu, \sigma^2\}$ for every customer in the 25th period (see Table A.1 in Appendix A.2.1 for details) and estimate the full model using the remaining 80 periods in our data.

The evolution of a customer’s beliefs over time can then be updated each period given every new piece of information acquired (i.e., shipping experiences in our application) using Bayes’ rule. Different from the previous literature in which the posterior distribution of the learning parameters has a closed-form expression, after adding a hierarchy to our model to allow for learning spillovers (i.e., $\mu_j \sim N(\mu, \xi^2)$), a closed-form expression is not available. Thus we resort to Gibbs sampling for calculating the posterior distribution of all quality moments ($\mu_j, \sigma^2, \mu$ and $\xi^2$). A detailed discussion on the need of using a simulation-based method such as Gibbs sampling and its advantages over other simulation methods such as Metropolis-Hastings can be found in Appendix A.2.2.

Accordingly, let $I_t$ denote the information set of a customer at the beginning of time period $t$ ($25 \leq t \leq T$). Then $I_t = \{Q_{j,p-1}, \mu_j^E, \sigma_p^E, \xi_p^E, \forall j \in T_i$ and $25 \leq p \leq t\}$, which contains both (1) the customer’s shipping experiences $Q_{j,p}$ in the past periods $1 \sim (t - 1)$; and (2) his expectation for shipping service quality at the beginning of each period. Here, we use superscript $E$ to denote a customer’s “estimated” value for a quantity of interest (in our model, this estimated value

29
corresponds to the posterior mean). For example, \( \mu_{jp}^E \) is a customer’s estimate of \( \mu_j \) at the beginning of period \( p \); similar definitions apply to \( \sigma_p^E \) and \( \xi_p^E \). The customer’s shipping service quality estimates (e.g., \( \mu_{jp}^E \)) are updated if he has new shipments delivered by AlphaShip in the previous period; otherwise, the estimates are the same as before. Assuming customer \( i \) has shipments finished in period \( t \), then he updates his quality beliefs at the beginning of period \( t + 1 \) using the updated information set \( I_{t+1} \). Specifically, a Gibbs sampler is implemented using the following full-conditional posterior distributions of the model parameters, and this sampling process is iterated multiple times until convergence (we use “…” to indicate all the other parameters and data):

\[
\mu_j \mid \ldots \sim N \left( \frac{n_j \xi^2 Q_j + \sigma^2 \mu}{n_j \xi^2 + \sigma^2}, \frac{\xi^2 \sigma^2}{n_j \xi^2 + \sigma^2} \right),
\]

\[
\sigma^2 \mid \ldots \sim IG \left( \alpha_\sigma + \frac{1}{2} \sum_{j \in \Upsilon_i} n_j, \delta_\sigma + \frac{1}{2} \sum_{j \in \Upsilon_i} \sum_{p=1}^I y_{jp}^* (Q_{jp} - \mu_j)^2 \right),
\]

\[
\mu \mid \ldots \sim N \left( \frac{J \sigma^2 \bar{\mu} + \xi^2 \mu_0}{J \sigma^2 + \xi^2}, \frac{\sigma^2 \xi^2}{J \sigma^2 + \xi^2} \right),
\]

\[
\xi^2 \mid \ldots \sim IG \left( \alpha_\xi + \frac{J}{2}, \delta_\xi + \frac{1}{2} \sum_{j \in \Upsilon_i} (\mu_j - \mu)^2 \right).
\]

where the number of routes is denotes as \( J = |\Upsilon_i|, \bar{\mu} = \frac{1}{J} \sum_{j \in \Upsilon_i} \mu_j \). \( y_{ij}^* \)\(^6\) is an indicator that is set to 1 if customer \( i \) had a shipment on route \( j \) delivered during period \( t \), and 0 otherwise, \( \bar{Q}_j = \frac{1}{n_j} \sum_{p=1}^t Q_{jp} y_{jp}^* \) and \( n_j = \sum_{p=1}^t y_{jp}^* \).

The Gibbs sampling outcome is a set of values for all the parameters (i.e., \( \mu_j, \sigma^2, \mu \) and \( \xi^2 \)) drawn from their posterior distribution. To construct estimates of mean quality beliefs such as \( \mu_{j,t+1}^E \), we assume the customers take the average of the (posterior)

\(^6\) Note that \( y_{ij}^* \) is a shipment delivery indicator and \( y_{ij} \) (§4.2) is a shipment start indicator. A shipment is does not always finished in the same period when it starts, thus \( y_{ij} \) and \( y_{ij}^* \) are often different.
draws for $\mu_j$, which gives an estimate of the posterior mean of this parameter. The same approach is used for all other parameters. The estimates, $\mu_{t+1}^E$, $\mu_{t+1}^E$, $\sigma_{t+1}^E$ and $\xi_{t+1}^E$, are in turn used as predictors in the quality function $f(I_{it}, \beta_i^q)$ that describes how quality beliefs affect a customer’s decision to rely on AlphaShip for shipping services (see §4.3.3).

It is important to note that besides being an efficient computational tool to estimate the posterior distribution of quality beliefs, the Gibbs sampling steps are also useful to understand how learning spillovers shape these beliefs. Specifically, a new quality (i.e., delay) observation on route $j$ changes the customer’s beliefs about the mean transport delay on route $j$, which in turn moves the customer’s mean quality ($\mu$) about all routes served by AlphaShip. Through this iterative process, the new beliefs about overall service quality shifts the beliefs of all other routes (and more strongly when the degree of route quality heterogeneity $\xi^2$ is small, see equation 2.9). Consequently, information about one product (route) is used not only to update beliefs about that product, but also about other products.

Furthermore, new experiences not only change consumer beliefs about the mean qualities ($\mu_j$ and $\mu$), but also about the variability of the experiences. More specifically, experience variability $(\sigma^2)^E_i$ reflects the variance of service quality $Q_{jt}$ around its mean ($\mu_j$). In this regard, it is useful to rely on the properties of the inverse-gamma distribution, which imply that the consumer’s estimate of experience variability $(\sigma^2)^E_p$ can be expressed as:

$$E[\sigma^2 | \cdots] = \left[ \delta_\sigma + \frac{1}{2} \sum_{j \in \mathcal{Y}_i} \sum_{p=1}^t y_{jp}^t (Q_{jp} - \mu_j)^2 \right] / \left( \alpha_\sigma - 1 + \frac{1}{2} \sum_{j \in \mathcal{Y}_i} n_j \right)$$

(2.13)

Note that the denominator in this equation increases with the arrival of new information, since $\sum_{j \in \mathcal{Y}_i} n_j$ increases with more information. However, the numerator in Equation (2.13) also increases when new information arrives. For example, if the
new delay information is very inconsistent with prior quality perception $\mu_j$ (e.g., after experiencing a unusually long delay), beliefs about experience variability will increase dramatically.

A similar pattern is observed when focusing on the evolution of beliefs about route quality heterogeneity $(\xi^2)_t^E$, which measures how different service quality is across routes. When new information about a route is dramatically different from prior beliefs (e.g., an unusually early delivery), this will shift mean quality beliefs about that route, and if doing so the estimated quality of that route $(\mu^E_{jt})$ shifts away from AlphaShip’s grand mean quality ($\mu$), then the estimated route heterogeneity $(\xi^2)_t^E$ will increase.

It is important to note that the standard learning models in the literature (e.g., Erdem 1998) assume experience variability $\sigma^2$ to be a constant known by the customer. In addition, the variance of prior (i.e., initial) beliefs about $\mu_j$ is also assumed to be constant (e.g., consumers might start with very vague beliefs). This last assumption is reasonable when modeling the learning process about the quality (or other attributes) of a single product in isolation. However, this assumption also makes beliefs about the mean quality of a product to monotonically become more precise as the customer gains a new experience. Zhao et al. (2011) relax this assumption by assuming an inverse-gamma distribution for belief uncertainty. Our model relaxes the assumption about the experience variability $\sigma^2$ being known to the consumer and also allows the uncertainty about mean quality beliefs to potentially increase or decrease as new information is gained by the customer. Accordingly, to the best of our knowledge our paper is the first study in correlated learning to relax both assumptions, thus providing a more flexible quality learning structure.

To help readers better understand the information spillover process, we estimate the learning process for a customer (in our data) who has three frequent routes $\{a,$
Figure 2.1: Top: updates of perceived shipping quality on each route $\mu_{at}^E$, $\mu_{bt}^E$, $\mu_{ct}^E$ and overall shipping quality represented by the hyper-parameter $\mu_t^E$. One observation on period $t = 62$ is marked; Bottom: updates of experience variability $(\sigma^2)_t^E$ and belief uncertainty $(\xi^2)_t^E$.

$b, c}$. We plot the changes to his estimated mean quality beliefs for each route $\mu_{at}^E, \mu_{bt}^E, \mu_{ct}^E$, grand mean quality represented by the hyper-parameter $\mu_t^E$, experience variability $\sigma_t^E$ and quality heterogeneity across routes $\xi_t^E$ throughout the observation period. As we can see from the top panel in Figure 2.1, the estimated mean quality for each route changes over time and so does the grand mean quality $(\mu_t^E)$ which takes values between the highest and lowest of the route quality beliefs. To maintain the readability of the figure, we did not plot the quality of all shipping experiences.
Instead, let us focus on the events corresponding to $t = 59$. We use two vertical lines in the top and bottom panels to highlight the period in which this “experience” was gained ($t = 59$) and the following period ($t = 60$). The top panel shows a large increase of estimated quality (i.e., delay) on route $c$ in period 60. This is caused by a severe delay of almost 9 hours on this route in period 59. This severe delay also causes the estimated grand mean quality $\mu^E_t$ to increase in period 60 (see the top plot). Furthermore, as we explained before, the large inconsistency between the experience in period 59 and the estimated mean quality $\mu^E_{59}$ causes a big surge in experience variability $\sigma^E_t$ following the severe transport disruption as depicted in the bottom plot.

With this large increase in experience variability $\sigma^E_t$ and a relatively small increase in quality heterogeneity across routes $\xi_t^E$, the estimates of the mean quality for route $c$ (i.e., $\mu^E_{ct}$) place more weight on the grand mean quality $\mu^E_t$ compared to the average observed quality on that route ($\bar{Q}_j$). Mathematically, this can be seen using the Gibbs sampling formula $E[\mu | \ldots] = \frac{n_j \xi^2 \bar{Q}_j + \sigma^2 \mu}{n_j \xi^2 + \sigma^2}$ (Equation (2.9)). Note that this expectation is a weighted average between the average observed quality on route $j$ ($\bar{Q}_j$) and the overall mean quality across all routes ($\mu$). As a result, when quality exhibits large variability ($\sigma^E_t$), the estimated mean qualities on routes $a$ and $b$, $\mu^E_{at}$ $\mu^E_{bt}$ respectively, more strongly shift towards the overall quality $\mu^E_t$. If instead, heterogeneity across routes $\xi_t^E$ increased while experience variability $\sigma^E_t$ remains almost unchanged, the relative importance of observed qualities on each route $\bar{Q}_j$ would increase, implying that the estimated mean quality on that route ($\mu^E_{jt}$) would change more strongly towards the customer’s own average experiences on that route ($\bar{Q}_j$).
Bayesian Regression Hierarchical Model for Customers’ Learning

We now consider a setting where customers learn from their shipping experiences at AlphaShip taking into account the characteristics of different routes when using information about one route to learn about the quality on another route. This approach is based on embedding a hierarchical Bayesian regression within the learning model. Accordingly, instead of modeling consumer beliefs in a given period $t$ for route $j$ as $\bar{Q}_{jt} = \mu_j$ in Equation (2.5), we allow $\mu_{jt}$ to not only change with route $j$ in §4.3.1 but also with the specific characteristics of that route. Focusing on distance as one of those characteristics, $\mu_{jt}$ becomes a linear function of route distance:

$$\mu_{jt} = \theta_j + \gamma \cdot Distance_j$$  \hfill (2.14)

where $\theta_j$ is a route-level mean quality intercept for route $j$; $Distance_j$ is the great-circle distance of route $j$; and $\gamma$ measures the effect of distance on the mean transport delay. Due to data constraints, especially the missing no-purchase data as explained in §4.1, we use only distance as a predictor of mean quality. However, our model and estimation method can be easily extended to include more route characteristics if more observations were available. As before, we add a hierarchy to these parameters to allow for information sharing:

$$\theta_j \sim N(\mu, \xi^2) \forall j \in \Upsilon_i$$

Once again, we adopt a Normal-Gamma distribution as the hyper-prior. In particular, we assume that customer $i$ has the following vague priors $\theta \sim N(\mu_0, \sigma_\theta^2)$; $\gamma \sim N(\gamma_0, \sigma_\gamma^2)$; $\xi^2 \sim IG(\alpha_\xi, \delta_\xi)$; $\sigma^2 \sim IG(\alpha_\sigma, \delta_\sigma)$. We solve the left truncation problem in the same way as before by using the first 24 periods data as pre-estimation sample and adopting the same initial beliefs except for the distance coefficient for which we assume a vague prior: $\gamma \sim N(0, 30^2)$. Finally, under this learning model each customer’s information set becomes $I_t = \{Q_{j,p-1}, \mu_{jp}^E, \theta_{jp}^E, \gamma_p^E, \sigma_p^E, \xi_p^E \forall j \in \Upsilon_i$
and $25 \leq p \leq t$, $\forall 25 \leq t \leq T$. For detailed Gibbs sampling formulas, please refer to Appendix A.2.3 and A.2.4.

**Service Quality Related Factors in the Choice Model**

In this section we discuss how shipping service quality beliefs may affect customers’ shipping decisions through $f(I_{jt}, \beta)$ in Equation (2.3) (subscript $i$ is omitted for expositional clarity). In particular we focus on three characteristics of this function.

First, we consider the potentially asymmetric effect of the estimated mean quality $\bar{Q}_{jt}$. In our application positive and negative quality values have a different meaning, they imply tardiness and earliness, respectively. Even though both of these might be undesirable, they may have generate different costs to the customer. Therefore, we will allow positive and negative values of the mean quality to have a different impact on the utility function that determines a customer’s shipping decisions.

Second, we explore the possible nonlinear effect of the estimated mean quality (i.e., transport delay). Customers might respond differently to minor delays (e.g., 2 hours) than to disruptive delays (e.g., 2 days). For example, compared to the smaller inconvenience brought by a 2-hour day, a delay of 2 days may probably disrupt the customer’s production or delivery plan and cause more severe loss of business. Accordingly, we consider a quadratic function of the estimated mean quality $\mu_{jt}$ to account for this possibility.

Finally, similar to several previous learning models in the literature (e.g., Erdem 1998; Sridhar et al. 2012), we consider that customers might be risk averse. The uncertainty about service quality $Q_j$ has two components: $Var[Q_j] = E[(\sigma^2)_{jt}] + Var[\mu_{jt}]$. The first component $E[(\sigma^2)_{jt}]$ measures the estimated degree of experience variability and originates from inherent variability in the delivery of the service and thus cannot be reduced. The second component $Var[\mu_{jt}]$ measures cus-
tomers’ uncertainty about the true value of the mean quality and it originates from a customer’s limited knowledge. The effect of these two uncertainty components on customer utility may differ. Previous studies have explored customers’ risk aversion by either combining the two components (i.e., estimating the effect of the total uncertainty \( \text{Var} [Q_j] \), such as Erdem (1998)) or only focusing on the effect of the uncertainty associated with learning the true value of the mean quality \( \text{Var} [\mu^E_{jt}] \) (e.g., Coscelli and Shum 2004; Sridhar et al. 2012; Zhao et al. 2011). In contrast, in our model we try to separately investigate and compare the effect of these two sources of uncertainty.

2.4.4 Model Identification

Since customers directly observe the service quality \( Q_{ijt} \) for each of their shipments with AlphaShip, the identification of the learning model (see §4.3) is a standard Bayesian statistics problem: the parameters will always have proper posterior distributions given the prior distributions that we use, and the magnitude of the posterior variance depends on the information contained in the Cargo 2000 shipping data. So we focus on the identification of the demand arrival (see §4.1) and shipping choice models (see §4.2).

From Equation (3.1) and (2.4), we obtain the probability of observing a purchase from customer \( i \) on route \( j \) at time \( t \):

\[
P(y_{ijt} = 1) = P(y_{ijt} = 1 \mid d_{ijt} = 1) \cdot P(d_{ijt} = 1) = \lambda_i m_{ij} \cdot e^{\bar{U}_{ijt}}/(1 + e^{\bar{U}_{ijt}})
\]

In the above equation, the parameter \( \lambda_i \) determines the overall demand level of customer \( i \), since \( P(y_{ijt} = 1) \propto \lambda_i \). On the other hand, the customer-level intercept \( \beta_0^i \) also plays the role of adjusting the overall probability of shipping: if \( \beta_0^i \) increases then \( e^{\bar{U}_{ijt}}/(1 + e^{\bar{U}_{ijt}}) \) increases. As a result, separately identifying both \( \lambda_i \) and \( \beta_0^i \) is very difficult, especially when we only have very limited amount of data (around
20 observations for each customer). For example, the data generated from a high $\lambda_i$ and low $\beta^0_i$ customer can be very similar to that from a customer with low $\lambda_i$ and high $\beta^0_i$. As a result, we set $\lambda_i = 1$ for all customers and focus on estimating $\beta^0_i$.

Now consider $m_{ij}$, which is a parameter of a multinomial distribution. These parameters must satisfy the following constraints $0 < m_{ij} \leq 1$ for $j \in \Upsilon_i$ and $\sum_{j \in \Upsilon_i} m_{ij} = 1$ for customer $i$. To enforce these constraints, we estimate $\bar{m}_{ij}$ instead of $m_{ij}$, where $m_{ij} = \exp(\bar{m}_{ij}) / \sum_{j \in \Upsilon_i} \exp(\bar{m}_{ij})$ and $-\infty < \bar{m}_{ij} < \infty$. Without loss of generality, we normalize the value of $\bar{m}_{ij}$ for the first route ever chosen by customer $i$ to 0. The identification of the remaining model parameters (i.e. $\beta^p$) is standard in discrete choice models and will be omitted for brevity.

Finally, we conducted a Monte Carlo simulation study to test our methodology. We simulated the shipping behavior of 80 customers for 100 periods. The arguments, such as shipping service quality, used in the simulation test are set close to those in the real data. Results show that our parameters can be reasonably recovered from the simulated data. These results provide evidence that our estimation approach can recover the true parameters of the data generating process (detailed results are available in the online appendix). For more details about the estimation, please refer to Appendix A.1.2 and A.3.1.

2.5 Results and Discussion

We first discuss different learning models and compare their goodness-of-fit in §5.1. In §5.2, we then provide the results of the demand arrival and shipping choice model, particularly, we provide results for the choice models based on different learning model assumptions.
2.5.1 Quality Learning Model

Different from most models in the empirical quality learning literature, after every experience customers in our study obtain an objective measure of service quality (i.e. transport delay). Since we also observe these service quality metrics, the estimation of each customers’ learning process can be performed independently from the estimation of the remaining model parameters. Before presenting the estimates of the learning model, we first compare the goodness-of-fit of several competing learning models using the shipping service quality data.

Learning Model Comparison

In addition to the Bayesian hierarchical model explained in section 4.1, we further consider several benchmark learning models. The first benchmark is a short-memory learning model — customers only rely on the most recent experience when anticipating the service quality they will receive. Specifically, we let

\[ f(I_t, \beta^q) = Q_{ij,t-1} \]  

(2.15)

where \( Q_{ij,t-1} \) is the actual shipping service quality experienced by customer \( i \) on route \( j \) in period \( t - 1 \) (\( Q_{ij,t-1} \) is the same as \( Q_{ij,t-2} \) if there is no new experience in period \( t - 1 \)), and is set to 0 if the customer has not chosen route \( j \) by period \( t \) to rely on the services of AlphaShip.

The second benchmark model describes an independent learning process — a customer only updates his beliefs about shipping quality on a route using the usage experience on that specific route, without borrowing information from other routes he has experienced before. Recall that we model \( Q \) using a normal distribution, i.e. \( Q_{jt} \sim N(\bar{Q}_{jt}, \sigma_{jt}^2) \) (see Equation (2.5) in §4.3.1). In the independent learning model, customer beliefs about the mean quality \( \mu_{jt} \) can be expressed as follows:

\[ \mu_{jt} \sim N(\mu_{0j}, \xi_j^2) \]  

(2.16)
Due to the lack of data for each customer on each route, we do not consider an independent learning model with predictors here.

The third benchmark model relies on full information pooling — for each customer the information coming from all routes is equally informative to learn about the mean quality of a particular route. Specifically, we consider two sub-models: benchmark 3A\(^7\) (equation (2.17)) and benchmark 3B (equation (2.18)), where model 3B has an additional predictor \(\text{Distance}\) compared to model 3A:

\[
\begin{align*}
\mu_{jt} &= \mu, & \mu &\sim N(\mu_0, \xi^2) \\
\mu_{jt} &= \theta + \gamma \cdot \text{Distance}, & \theta &\sim N(\mu_0, \xi^2), & \gamma &\sim N(\gamma_0, \sigma_\gamma^2)
\end{align*}
\]  

(2.17) (2.18)

Table 2.5 provides fit measures: negative log-likelihood \((-LL)\) and Deviance Information Criterion \((DIC)\) of the benchmark models, the simple hierarchical model (see \(\S 4.3.1\)) and the regression hierarchical model (see \(\S 4.3.2\)).\(^8\) For both \(-LL\) and \(DIC\), the smaller the values, the better the model fits the shipping experience data. Note that the dependent variable in this model is the quality (i.e., delays) of the service provided by AlphaShip. Since benchmark model 1 does not involve learning, it is not included in Table 2.5. Among all the models, the simple hierarchical model provides the best fit in terms of both \(-LL\) and \(DIC\). A closer look into Table 2.5

\(^7\) In both benchmark model 2 and 3, we use the first 24 period pre-estimation sample to obtain the customer-level priors for the rest of the periods.

\(^8\) DIC is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion, and is particularly useful in Bayesian model selection.
reveals that adding the *Distance* predictor actually decreases model fit, and this is shown by the fact that the corresponding model *without* this predictor performs better for both the hierarchical model (simple hierarchical model vs. full model) and the information pooling models (models 3A vs. 3B). Contrary to a linear regression where adding more predictors does not decrease in-sample fit, in a learning model this is not necessarily the case. In a learning model the importance of a predictor ($\gamma$ in our model) changes over time based on the gained experiences. So all previous experiences determine the weight of a predictor, however this weight is then used to predict a new experience.\(^9\)

Furthermore, the hierarchical model systematically fits the data better than the information pooling model. Note that the fit of the simple hierarchal model is better than that of model 3A and that the fit of the regression hierarchical model is also better than that of model 3B. Also note that the fit of independent learning model is similar to that of the information pooling model 3A. The model comparison result justifies our primary interests of using hierarchical learning models in the paper. However, the model that provides the best fit to the shipping quality data may not necessarily be the actual learning model that customers use. We will explore this issue in §5.2 where we will compare how the ability of the different learning models to provide useful predictors of customer choices. Before comparing these models, we discuss in more detail the features of the best-fitting learning model in the next subsection.

\(^9\) To illustrate ideas, suppose that all previous experiences with long distance shipments were satisfactory (no delays), while those with short distances were not. Then this will give a strong weight to distance when the consumer anticipates the quality of future experiences. However, if the next experiences reverse this pattern (e.g., they involve a satisfactory short distance shipment and an unsatisfactory long distance shipment), then using distance will not make quality beliefs more accurate for these new experiences.
Table 2.6: Learning model parameters’ population mean and 90% quantile interval

<table>
<thead>
<tr>
<th>Bayesian Learning Model Parameters</th>
<th>Mean</th>
<th>Percentiles (5%, 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{ijT}$</td>
<td>-0.727</td>
<td>(-2.863, 1.560)</td>
</tr>
<tr>
<td>$\sigma^2_{ijT}$</td>
<td>2.592</td>
<td>(1.150, 5.122)</td>
</tr>
<tr>
<td>$\xi_{ijT}$</td>
<td>1.759</td>
<td>(1.280, 2.572)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand Arrival Model Parameters</th>
<th>Max$<em>{jeT_i}$ $(m</em>{ij})$</th>
<th>Min$<em>{jeT_i}$ $(m</em>{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.540 (0.319, 0.748)</td>
<td>0.178 (0.036, 0.452)</td>
</tr>
</tbody>
</table>

*Estimation of Bayesian Hierarchical Model*

Since the simple hierarchical model (see §4.3.1) provides the best fit of the shipping quality data, in what follows we discuss the results for this model, while the results of other learning models are similar.

Under this model, each customer learns from his own experiences without using information from other customers. Therefore there are multiple parameters $\{\mu_{ij}, \mu_i, \sigma^2_i, \xi_i^2, \forall j \in \Upsilon_i, i = 1, \cdots, N\}$ to estimate for each customer, leading to more than 2,000 parameters in total. Due to the large number of individual-level parameters, in Table 2.6, we only provide the population mean and 90% interval of the grand mean quality $\mu^{E}_{iT}$, experience variability $(\sigma^2_{ijT}^{E})$ and route quality heterogeneity $(\xi^2_{ijT}^{E})$ as estimated by customers during the last period $T$ in our dataset. The summary statistics of the customer-route level quality perception, $\mu^{E}_{ijT}$, are not included in Table 2.6 due to the varying number of routes $|\Upsilon_i|$, across customers.

From the results in 2.6, we observe considerable heterogeneity among customers. In previous learning models, the heterogeneity in customers’ learning processes mainly comes from the limited number of trials (e.g., Coscelli and Shum 2004). Due to experience variability, customers can only observe noisy signals of the true quality. When the number of experiences is small, customers may form beliefs which are far from the true mean quality, thus resulting in different opinions about the service quality.
Some customers thought the service is good if they gain a few good experiences, while others may not agree if they happen to receive several bad outcomes. To the best of our knowledge, our model for the first time in the literature does not restrict the underlying true mean quality $\mu_{ijt}$ to be the same for all customers. Therefore, except for “finite trials”, the varying levels of quality experienced by different customers serve as another source of customer heterogeneity.

In addition, we find that the population mean of $\mu_T$, $E[\mu_{iT}] = -0.727$ is close to the empirical shipping quality average of -0.77 hours (see Table 2.1), and the population mean of quality standard deviation, $E_i[\sigma_{iT}] = 2.56$ is slightly smaller than the empirical variance of shipping quality 3.51. This shows that after one year of learning, the average customers’ quality beliefs is close to true empirical value from the data, showing a considerable learning from the flat priors.

In sum, from the estimates of the customer Bayesian learning models, we conclude that considerable learning can be achieved on a period-by-period basis with the arrival of new information and that the learning processes greatly vary across customers.

2.5.2 Demand Arrival

For the demand arrival, the parameters to be estimated are $\bar{\mathbf{m}}_i = \{\bar{m}_{ij}, \forall j \in Y_i\}$. Similar to the estimation of the learning parameters, we estimate the demand arrival parameters $\bar{\mathbf{m}}_i$ independently for each customer $i$, rendering 3000+ parameters in total. Since customers have different numbers of frequent routes, we only provide the population summary of each customer’s most and least frequent route in Table 2.6. To facilitate the identification and estimation of the multinomial distribution parameters, we restricted our customers to those whose most frequent route accounts for no more than 70% of his total shipments during the year, and we verify that the $\max_{j \in Y_i} (m_{ij})$ is close to that value. On average, the top route for a given customer
is three times more popular than the least demanded route.

2.5.3 Choice Model

We now consider the estimation of the utility function that determines customer choices to use AlphaShip as a service provider. This estimation depends on i) which learning model customers use to predict quality and ii) the shape of the shipping quality function (i.e., asymmetry, non-linearity and risk aversion). Table 2.7 reports a rank of goodness-of-fit for alternative specifications of the shipping quality function (see §4.3.3) and learning models (see Table 2.5). Here we use all the predictors (e.g., intercept, price) explained in §4.2, since we find they are all significant for all the models and that the values of the coefficients mostly remain unchanged across models.

As previously mentioned, the models we compare differ on two major dimensions: (1) the learning model; and (2) the shape of the service quality function \( f(I_t, \beta^q) \). Specifically, we compare different learning assumptions, including no learning (i.e., benchmark model 1), independent learning (i.e., benchmark model 2), information pooling learning (i.e., benchmark model 3A) and spillover learning (i.e., simple hierarchical model). Here we do not compare the regression counterparts of the models since Distance is not a useful predictor based on the results in §5.1. As we have discussed in §4.3.3, the specification of service quality function \( f(I_t, \beta^q) \) may differ in terms of three aspects: (1) whether the effect of expected service quality mean \( \bar{Q}^E \) is symmetric or asymmetric; (2) whether the effect of expected service quality \( \bar{Q}^E \) is linear or nonlinear; (3) whether the customers are risk averse or risk neutral, and if the customers are risk averse, whether the customers are averse to either one of or both of sources of uncertainty (experience variability and uncertainty about mean quality).

We test all the specification combinations and report the likelihood of selective models in Table A.3 in the Appendix and provide a goodness of fit rank in Table 2.7.
Table 2.7: Rank of Goodness of Fit on Alternative Specifications of the Shipping Quality Function

<table>
<thead>
<tr>
<th>LL/AIC</th>
<th>Benchmark Model 1 (short memory)</th>
<th>Benchmark Model 3A (information pooling)</th>
<th>Benchmark Model 2 (independent learning)</th>
<th>Simple Hierarchical Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>S*</td>
<td>16/16**</td>
<td>15/13</td>
<td>14/14</td>
<td>12/12</td>
</tr>
<tr>
<td>A</td>
<td>——</td>
<td>13/15</td>
<td>10/10</td>
<td>9/8</td>
</tr>
<tr>
<td>A + ERA</td>
<td>——</td>
<td>11/11</td>
<td>8/9</td>
<td>7/7</td>
</tr>
<tr>
<td>A + ERA + BUA</td>
<td>——</td>
<td>6/5</td>
<td>4/3</td>
<td>2/1</td>
</tr>
<tr>
<td>A + ERA + BUA + Q</td>
<td>——</td>
<td>5/6</td>
<td>3/4</td>
<td>1/2</td>
</tr>
</tbody>
</table>

*: We use S for symmetric; A for Asymmetric; ERA for Experience risk averse; BUA for belief uncertainty averse; Q for Quadratic.

**: The rank of models’ goodness-of-fit by log-likelihood is before the slash, and that by Akaike information criterion (AIC) is after the slash. 1 means the model fits the best under the criterion, and 16 means the model fits the worst.

The results of these goodness of fit comparisons show how these measures improve with the properties of the learning model and service quality function. Specifically, for each row of Table 2.7, the rank decreases from left to right, meaning that the simple hierarchical learning model (see Table 2.8 for the estimates) fits the data better than the independent learning assumption under the same specification of the quality function $f(I_t, q^t)$ (i.e., within the same row). In addition, the fit of the independent learning model is in turn better than that of the information pooling learning model. Overall, these results provide evidence consistent with information spillovers driving customer choices.

Next, if we look at each column, we find the rank by model log-likelihood improves from top to bottom, showing that as the model becomes more flexible, goodness-of-fit improves under the same learning model (i.e., the same column). This model comparison, however, does not consider the number of parameters of each model. Accordingly, it is useful to note that the ranks in the last two rows reverse under the criterion of AIC, since the improvement by adding the quadratic structure is not enough to compensate for the penalty associated with adding the quadratic parameters. Considering both model fit and model parsimony, we choose the asymmetric
risk averse model (i.e., \( A + ERA + BUA \)) as the best model, and use its coefficients under the simple Bayesian learning model (Benchmark Model 4) for the following analysis and simulation studies in section 6.

**Results for Shipping Choice Model**

The coefficients estimated for alternative models under different learning models are very similar (i.e., the same signs and close magnitudes), so we only provide the results under the simple hierarchical learning specification (see Table 2.8). In this table the variance of the random coefficients for a particular predictor are indicated with the \( \Omega \) symbol (e.g., \( \Omega(\text{Intercept}) \) denotes the variance of the random intercepts across customers).

We can see that as expected the log-likelihood of customer choices increases from the left to right as the utility model becomes more flexible through the service quality function. Specifically, by comparing model 2 and 3 (counting from the left), we find that allowing for an asymmetric effect of the predicted mean quality \( Q \) not only improves model fit but also highlights customers’ higher sensitivity to tardiness than earliness. Note that in the case of the average customer, the coefficients for both tardiness \((|\mu_t|^+)\) and earliness \((|\mu_e|^-)\) are both negative, however the magnitude of the first coefficient is estimated to be greater. This implies that customers exhibit a negative response to tardiness, a positive response to earliness, but that the negative responses to tardiness are stronger than the positive responses to earliness. We also find substantial heterogeneity in terms of customer responses to earliness, implying that even though earliness is desirable for the average customer, there is a substantial fraction of customers that would prefer to avoid earliness, perhaps due to the incremental storage and handling costs associated with shipments arriving too early.

In terms of the impact of uncertainty, a comparison between models 3, 4 and 5 shows that customers are averse to experience variability \((\sigma^2)\), but even more so to
belief uncertainty \((\text{Var}(\mu_j))\). This finding again confirms the importance of taking into account both sources of uncertainty. In contrast with previous approaches, due to the lack of service quality data, to the best of our knowledge none of the previous learning models in the literature separately studies the effects of these two uncertainty sources. Our results show that customers are averse to both risks, while the impact of belief uncertainty is almost twice as big as that of the experience variability. Finally, by comparing the last two models, we find that adding quadratic terms brings almost no change to the likelihood, and this is consistent with the quadratic coefficients being non-significant.

2.5.4 Discussion

In addition to using routes as choice alternatives for each customer, we have also used a more aggregated definition based on destination countries. Since the customers in our data have fixed physical locations and usually rely on only a single origin airport, different routes of a customer are actually differentiated by different destination airports. To test the model robustness to different levels of product aggregation, we aggregated the destination airports into destination countries and consistent with our spillover formulation allowed customers’ experiences shipping to one country to affect their expectations about service quality when shipping to other countries. After re-estimating all models we obtain virtually the same results: the information spillover model not only outperforms other learning models at predicting shipping service qualities but also in terms of predicting customers’ purchase decisions. These results again provide evidence consistent with information spillovers driving customer choices.
2.6 Policy Simulations and Managerial Implications

The results of the previous section suggest that customer choices are consistent with the use of quality information from other products to form beliefs about a particular product. In addition, customers are sensitive to not only the predicted mean service quality but also to quality uncertainty. In this section, we discuss several managerial insights directly related to these findings.

We first study the consequences of unsatisfactory levels of service quality. When service quality decreases on certain routes (e.g., due to higher average transport delay or larger experience variability), our model can be used to predict the corresponding impact on customer behavior.

For simplicity, we consider customers with frequent routes $A$ and $B$ and whose demand on the two routes follows a multinomial $(0.5, 0.5)$ distribution. Transport delays on these two routes, $Q_A$ and $Q_B$ respectively, follow the same normal distribution, $Q_A \sim N(0, 5^2)$ and $Q_B \sim N(0, 5^2)$. Customers decide whether to ship based on the utility of shipping compared to that of the outside option and then update their quality beliefs if new information is acquired (i.e., if they used AlphaShip services).\textsuperscript{10}

For the policy experiment about examining the effect of average transport delay increase, we change the transport delay on route $A$ to $Q_A \sim N(5, 5^2)$ starting from period 20. The top panel of Figure 2.2 presents the average purchase probability on the two routes over 40 periods.\textsuperscript{11,12} As we can see, before the change on route $A$ in period 20 happens (marked by the vertical line in Figure 2.2), the purchase probabilities on routes $A$ and $B$ are almost the same and both increase over periods, where the increase of purchase probability is caused by the lowered learning uncertainties as

\textsuperscript{10} When calculating utilities we set other exogenous variables (e.g., price) to their empirical means.

\textsuperscript{11} The average purchase probability is the mean of purchase probability of 200 simulated customers, and the customers’ quality sensitivities are the corresponding random coefficients’ mean.

\textsuperscript{12} The first 5 periods data are dropped as learning burn-in.
Figure 2.2: Change of purchase probability on route A and B under different learning rules. Top: spillover learning; Middle: independent learning; Bottom: information pooling learning.

the customers collect more information over the periods. However, after the average transport delay on route $A$ increase from 0 to 5 in period 20, the average purchase probability on route $A$ drops rapidly due to the deteriorating transport service quality. Interestingly, the purchase probability on route $B$ also drops at the same time. Since the service quality on route $B$ never changes, the drop in purchase probability is solely caused by the spillover of customers’ experiences from deteriorating service quality on route $A$. If we use the average empirical price of $2,288 as the revenue obtained from a shipment on route $A$ and $B$, then the increase of transport delay on route $A$ causes a direct expected loss of $103.0 on route $A$ and the indirect expected loss of $9.2 on route $B$ in the last period.

When the learning rule changes, the effect of service quality deterioration also changes. Specifically, in the middle and bottom panel of Figure 2.2 are the changes of purchase probability on route $A$ and $B$ under independent learning and information pooling learning respectively. As we can see, under independent learning, the quality deterioration on route $A$ won’t affect the demand on route $B$; on the contrary, as acquiring more information, the purchase probability on route $B$ actually increases as time goes by. While under the information pooling learning, experiences from route $A$ and $B$ are treated with no difference thus the deceases of purchase probability on these two routes are the same. Meanwhile, since the experiences with low service
quality on route A are diluted by the good experiences on route B, the decrease of purchase probability on route A is smaller than that under information spillover learning or independent learning. Like before, we use average empirical price $2,288 as the revenue obtained from a shipment on route A and B. Under the independent learning, the increase of transport delay on route A only causes a direct expected loss of $119.0 on route A, while there is no indirect loss on route B. If the customer adopts information pooling rule of learning, the expected loss of revenue on route A and B are both $53.5.

From the examples above we can see that when making decisions about whether to improve service quality and particularly how much budget to allocate for quality improvements, operations managers should consider not only the direct effect on the service to be improved, but also on other products affected by information spillovers.

2.7 Conclusions, Limitations and Future Direction

In this paper, we used a one-year shipping/sales historical data of individual customers of a world-leading third-party logistics company to study how customers learn about shipping service quality from their past experiences, especially from experiences of similar yet not identical service encounters. This learning behavior is referred to as “correlated learning” or “spillover learning” in the learning models literature.

Most of the previous studies in this literature do not directly observe quality or rely on surveys for gathering subjective measures of product or service quality. Our dataset instead contains objective service quality measures. This advantage allows us to separately estimate customers’ learning and shipping choice. This further enables us to estimate a learning model with more than 2,000 products (i.e., routes), as compare to no more than 10 focal products as it is typically the case in the literature.
An important contribution of this paper is methodological. Contrary to frequentist methods that are widely used for correlated learning, we use a Bayesian hierarchical methodology to model the learning process. The Bayesian hierarchical model provides a natural way to represent the information borrowing process and can be easily applied to problems with more complex correlation structures without adding too many parameters.

In terms of our results, we compare the Bayesian hierarchical model with three benchmark learning models: (1) a short-memory learning model; (2) a model of independent learning; and (3) a model of information pooling. Our results shows that the Bayesian hierarchical model not only outperforms the alternative model in forecasting transport delays, but also provides the best goodness-of-fit when predicting customer purchase decisions. This result confirms that choices are best predicted by relying on a learning model that allows experiences about one product to influence beliefs about other products. In addition, we found asymmetries in terms of how customers react to earliness and tardiness. More specifically, negative responses to delays are greater than positive responses to the early arrival of a shipment. We further obtain evidence of customers being risk averse, as in previous papers in the learning literature. However, we are also able to separate the effect of experience variability from belief uncertainty, with the latter having a greater impact on customer choices.

We further examine the managerial implications of our findings using counterfactual experiments. Specifically we show that quality deterioration on one product would not only affect the revenues from that product, but also the revenues from other products through learning spillovers. Therefore, decisions about operational quality improvements should account for both the direct and indirect effects of these improvements on customer behavior.

Finally, our study could be extended in several ways in future research. One
possible extension is to measure how observable customer characteristics — such as company demographics — are related to the learning process used by customers, e.g. spillover learning versus independent learning. This would be useful, for example, to develop targeted service improvement strategies based on the approach that customers use to learn about the quality of different services. Competition could also be an important aspect to consider; this would probably require data from multiple leading companies in the market to study how the market structure mediates the effect of service quality on customer purchases. On a final note, this study highlights the importance of bringing advanced methodologies in statistics, economics and marketing into the fields of operations management. We hope that this work stimulates further research on the interface between these academic disciplines.
Table 2.8: Estimation Results for Shipping Choice Model with Simple Bayesian Hierarchical Learning

<table>
<thead>
<tr>
<th>Variable</th>
<th>Null</th>
<th>S\textsuperscript{\dagger}</th>
<th>A</th>
<th>A + ERA</th>
<th>A + ERA + BUA</th>
<th>A + ERA + BUA + Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.65e-1***</td>
<td>-1.66e-1***</td>
<td>-1.93e-1***</td>
<td>-1.57e-1***</td>
<td>-1.21e-1***</td>
<td>-8.16e-2***</td>
</tr>
<tr>
<td>(\Omega(\text{Intercept}))</td>
<td>1.38e-1***</td>
<td>1.37e-1***</td>
<td>1.48e-1***</td>
<td>1.36e-1***</td>
<td>1.19e-1***</td>
<td>6.38e-2***</td>
</tr>
<tr>
<td>Price</td>
<td>-1.44e-1***</td>
<td>-1.44e-1***</td>
<td>-1.46e-1***</td>
<td>-1.44e-1***</td>
<td>-1.42e-1***</td>
<td>-1.42e-1***</td>
</tr>
<tr>
<td>Weight</td>
<td>3.64e-1***</td>
<td>3.64e-1***</td>
<td>3.64e-1***</td>
<td>3.63e-1***</td>
<td>3.62e-1***</td>
<td>3.62e-1***</td>
</tr>
<tr>
<td>CR</td>
<td>6.11e-1***</td>
<td>6.10e-1***</td>
<td>6.09e-1***</td>
<td>6.08e-1***</td>
<td>5.95e-1***</td>
<td>5.95e-1***</td>
</tr>
<tr>
<td>(\mu_{jt})</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\Omega(\mu_{jt}))</td>
<td>—</td>
<td>1.01e-1***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>([\mu_{jt}]^+)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\Omega([\mu_{jt}]^+))</td>
<td>—</td>
<td>—</td>
<td>8.20e-2**</td>
<td>1.28e-1***</td>
<td>1.26e-1***</td>
<td>1.21e-1***</td>
</tr>
<tr>
<td>([\mu_{jt}]^-)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\Omega([\mu_{jt}]^-))</td>
<td>—</td>
<td>—</td>
<td>3.97e-1***</td>
<td>4.05e-1***</td>
<td>4.02e-1***</td>
<td>3.97e-1***</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\Omega(\sigma^2))</td>
<td>—</td>
<td>—</td>
<td>1.95e-2***</td>
<td>1.17e-2***</td>
<td>1.15e-2***</td>
<td>1.15e-2***</td>
</tr>
<tr>
<td>(\text{Var}(\mu_{jt}))</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(\Omega(\text{Var}(\mu_{jt})))</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2.37e-2***</td>
<td>2.36e-2***</td>
<td>2.36e-2***</td>
</tr>
<tr>
<td>((\mu_{jt})^+)^2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>((\mu_{jt})^-)^2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

\(\text{LL}\) | -53643.6 | -53639.9 | -53634.2 | -53632.3 | -53605.9 | -53605.7

\(\dagger\): We use S for symmetric; A for Asymmetric; ERA for Experience risk averse; BUA for belief uncertainty averse; Q for Quadratic.

\(*p<0.1, **p<0.05, ***p<0.01\)

Note: All the predictors are normalized to have zero mean and approximate 1 standard deviation. Specifically, \(\sigma^2\) and \(\text{Var}(\mu_{jt})\) are scaled by 10; \(\mu_{jt}\), \(\mu_{jt}^+\) and \(\mu_{jt}^-\) are scaled by 2; Price is scaled by 5000; Weight is scaled by 3000; CR is scaled by 6.
3.1 Introduction

It is challenging to manage perishable products such as produce at the retail level. This is because the quality of perishable products degrades over time, depending on environmental conditions of storage, handling and customer pick. This time-changing product quality further affects consumers’ willingness to buy, thus demand. Traditional inventory management models treat the demand distribution as exogenously given. Few studies, if any, have explored the management problem with age-dependent demand. In this paper, we take the first step of studying the effect of product age on demand, and then develop inventory and pricing strategies based on the knowledge of demand for perishable products.

This work is facilitated by a unique dataset covering inventory and sales of unpackaged fresh vegetables in a supermarket chain. Compared to the packaged produce, unpackaged fresh produce is associated with better consumer shopping experiences, by allowing consumers to touch, smell and pick. However, it poses more
difficulties for operations management because fragile produce can be easily damaged by customers’ picks. As a result, these damaged items can be easily left unsold.

Due to the difficulties in managing fresh produce, huge wastes are being witnessed. Specifically, 52% of fresh fruits and vegetables are lost before reaching consumers, while over half of the food losses occur in distribution (e.g., supermarkets) and consumption (e.g., after customers have purchased the produce), and this ratio is even higher in middle/high-income regions (Gunders, 2012). Among wastes in the whole food supply chain, the USDA estimates that supermarkets lose $15 billion annually in unsold fruits and vegetables alone. Especially as explained in Gunders (2012):

...... Unfortunately, the retail model views waste as a part of doing business. According to a former President of Trader Joe’s, “the reality as a regional grocery manager is, if you see a store that has really low waste in its perishables, you are worried. If a store has low waste numbers it can be a sign that they aren’t fully in stock and that the customer experience is suffering.” Industry executives and managers view appropriate waste as a sign that a store is meeting quality control and full-shelf standards, meaning that blemished items are removed and shelves are fully stocked.

41 In 2005 and 2006, annual supermarket losses averaged 11.4 percent for fresh fruit and 9.7 percent for vegetables, with losses varying from 0.6 percent for sweet corn to as high as 63 percent for mustard greens ......

Meanwhile, the world continues to face a challenge to feed its people sustainably; around one billion people globally are malnourished (Naylor, 2011).

Despite the difficulties of managing perishable produce, fresh produce has been gaining more and more attentions from policy makers and retailers. Statistics from the United States Department of Agriculture (USDS, 2007) suggest that the per-capital consumption of fresh vegetables has been steadily increasing since the early
1980s, while the per capita consumption of traditional crops such as wheat and other grains have increased at a much slower pace and even decreased over the last few years. Such increases in fresh vegetable consumption can be attributed not only to population growth, but also to market changes, such as the increasing public awareness of the benefits of healthier diets (McLaughlin et al., 1999). From the perspective of the supermarkets, the fresh produce section can attract many consumers planning shopping trips that includes fresh produce, thereby improving the supermarket competency. Furthermore, a well managed and decorated fresh produce section can greatly contribute to a pleasant shopping environment, thus improving store images and attracting customers.

In this study, we first develop a discrete-continuous framework that has been previously applied in studies of consumer demand in marketing and other fields (e.g., Kim et al. 2002; Bhat 2008). This allows us to estimate the effect of produce’s freshness, price and other characteristics on demand. Based on demand, we further develop innovative pricing and inventory strategies that lead to profit improvement and waste reduction. Specifically, instead of the current one-price-all-time practice, better pricing strategies tailored to different produce can be developed. For example, if the current high price is designed to offset high waste cost of fast-perishable produce, the supermarket can potentially be more profitable by adopting a lower price to sell as much fresh produce at its freshest. We can also design dynamic price discount strategies based on consumers’ tradeoff on price and freshness. In addition, we also propose optimal inventory management strategies associated with order quantity and frequency. By combining the smarter inventory management strategy and pricing mechanism, we aim to reduce cost and increase revenue.

The contributions of our work to the literature are several folds:

1. Theoretical: first, on the demand side, we model store choice and product
choice decisions in a coherent model framework. Second, on the supply side, we obtain insights on inventory/pricing policy for perishable goods, instead of consumer packaged goods;

2. Methodological: We develop a Poisson arrival process to capture store visit timing. This technique helps us to overcome a limitation of the data set, from which we only observe customer’s shopping histories in the focal supermarket chains but we do not know when the customers shop at competitor supermarkets;

3. New insights: Using our model and counter-factual experiments, we study the effects of freshness on demand, the degree of substitution/complementarity among perishable goods, the implications on joint inventory/pricing strategies for substitutes/complements, and the impact on waste reduction for perishable goods.

3.2 Data Description

This data are provided by BetaMart (the true name of the company is disguised for confidential reasons), a large Chinese supermarket chain. The data set contains the point of sales (POS) data of fresh vegetables from 92 stores, and the order and fulfillment information from 4 distribution centers to these stores. The observation period spanning the whole year of 2013. The data are special because we have both inventory and wholesale information.

3.2.1 Demand and Supply Data

The unique customer ID, provided by the loyalty customer program of the supermarket chain, enables us to track the shopping history of each customer. Specifically, we observe 369,209 trips by 6,832 consumers recorded at the minute-level in a day,
Table 3.1: Data Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><em><em>Demand Data: Trip-level Statistics</em> (322,428 trips)</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total spending</td>
<td>103.63</td>
<td>114.18</td>
<td>0.81</td>
<td>1880.95</td>
</tr>
<tr>
<td>Spending on veggies</td>
<td>9.42</td>
<td>10.63</td>
<td>0.30</td>
<td>223.71</td>
</tr>
<tr>
<td>Percent of spendings on veggies</td>
<td>16.42%</td>
<td>20.03%</td>
<td>0.08%</td>
<td>100%</td>
</tr>
<tr>
<td>Number of categories</td>
<td>1.74</td>
<td>1.06</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td><strong>Demand Data: Customer-level Statistics (6,828 customers)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of total trips</td>
<td>47.19</td>
<td>28.68</td>
<td>10</td>
<td>277</td>
</tr>
<tr>
<td>Number of veggie shopping trips</td>
<td>18.58</td>
<td>16.79</td>
<td>5</td>
<td>193</td>
</tr>
<tr>
<td>Number of stores visited**</td>
<td>1.11</td>
<td>0.38</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td><strong>Demand Data: Veggie Category-level Statistics (220,804 trip-category)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchase quantity (kg)</td>
<td>0.85</td>
<td>0.82</td>
<td>0.03</td>
<td>33.52</td>
</tr>
<tr>
<td>Purchase amount (¥)</td>
<td>5.41</td>
<td>5.04</td>
<td>0.3</td>
<td>136.68</td>
</tr>
<tr>
<td>Unit price</td>
<td>9.38</td>
<td>9.62</td>
<td>0.6</td>
<td>126.04</td>
</tr>
<tr>
<td><strong>Supply Data: Day-level statistics (365 days)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount ordered (¥)</td>
<td>9100.58</td>
<td>3475.62</td>
<td>3872.40</td>
<td>28,554.65</td>
</tr>
<tr>
<td>Quantity ordered (kg)</td>
<td>1938.45</td>
<td>800.75</td>
<td>722.7</td>
<td>5137.60</td>
</tr>
<tr>
<td>Categories ordered</td>
<td>9.90</td>
<td>0.32</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

*: calculated using the shopping trips containing veggie purchases
**: calculated using all the data from 92 stores

and Table 3.1 provides summary statistics of sales at the customer level. As we can see from Table 3.1, the average portion of spending on fresh produce, 16.41%, is relatively large. Furthermore, since the stores of BetaMart are geographically separated, we find customers usually shops at only one BetaMart stores. Specifically, more than 99% customers shop at only one store in the one year observation period. So we don’t consider the competition among BetaMart’s stores.

The fresh produce is send to each store directly and put onto the shelves each morning without being stored in the warehouses, unlike most of the other packaged non-perishable products. The statistics of the vegetable supplies are also included in Table 3.1. The detailed cleaning steps of demand and supply data are provided in Appendix B.1.1.
*Aggregated Vegetable Categories*

We aggregate all the vegetables into 10 categories, based on the following reasons: (1) there are hundreds of different produces sold through the year and it will cause huge computational burdens if we include all of them into the discrete-continuous multiple choice model; (2) a lot of these produces are seasonal and are only provided in a very short period, however after aggregating them into a broader category, the sales of the 10 categories are almost stable. So we choose to use the 11 categories based on the existing category labels contained in the data: (1) leafy greens; (2) cabbage family (e.g., broccoli); (3) pepper; (4) gourd (e.g., cucumber); (5) common beans; (6) potato (solanum genus); (7) onion genus (e.g., onion, garlic); (8) lotus root; (9) mushroom; (10) others (e.g., sliced gourd).

In particular, we can see that the average number of categories ordered every day is 9.9, close to the total number of categories 10, and this confirms that the perishable veggies are refilled everyday and the aggregated categories provide a way to stabilize delay demand and supply. Moreover, we found customers to purchase on-average 1.74 categories of veggies per trip (see Table 3.1), which provide evidences for the basket purchasing behavior of customers.

### 3.2.2 Wholesale and Retail Price

Figure 3.1 shows the wholesale and retail prices of 6 veggies (lettuce, broccoli, zucchini, sweet potato, ginger and onion) in six randomly selected stores. As we can see, the retail price is always above the wholesale price. Moreover, retail prices are at their “regular” level most of the time, and deviations tend to be downward and short-lived. Meanwhile, we find the wholesale prices to be more volatile than the retail prices in most cases. The high volatility in wholesale prices confirms the importance of using directly observed wholesale prices in pricing and inventory optimization.\(^1\)

\(^1\) Most previous empirical studies rely on inferred wholesale prices since they are not available.
Figure 3.1: Price vs. Wholesale Price (lettuce, broccoli, zucchini, sweet potato, ginger, onion)
Weighted Average Price for Aggregated Categories

In the spirit of Shriver and Bollinger (2015), we compute category \( k \) specific price indices for a day \( t \) as weight average of SKU-level prices:

\[
P_{jt} = \frac{\sum_{j \in k} w_{jt} p_{jt}}{\sum_{j \in k} w_{jt}}
\]

where \( w_{jt} \) and \( p_{jt} \) are the daily sales quantity (oz) and price of SKU \( j \) during day \( t \). The applied weight \( w_{jt} \), naturally reflect more popular products in the resulting index.

3.2.3 Instruments of Endogenous Variables

One important goal of this paper is to study the effect of freshness and fullness of shelf on demand. However, the freshness and how full the shelves are all depend on the inventory refill quantity/frequency, which are further made based on demand forecast. As a result, the estimated impact of the observed factors (e.g., freshness, fullness of the shelves) captures not only that factors’s effect but also the effect of the unobserved factors that are correlated with it.

Control Function Approach to Endogeneity

In order to solve this problem, we use the control function method that is introduced in ?. We will postpone the details of control function approach to the model section. Specifically, we use the weather information as the instrument variable, which we purchased online.

3.2.4 Food Waste

A unique feature about produce is their perishability. Because of this feature, some ends up being wasted. To assess the magnitude of waste, we compute the annual
waste of the top 200 vegetables in each store, which end up with 7,911 vegetable-store pairs in 96 stores. Here we use the difference between annual total quantity received from suppliers and annual total quantity sold as a proxy of the annual waste of each vegetable in each store. Figure 3.2 shows the empirical distribution of the percentage of “waste” in total quantity received for each item-store pair. As we can see, the average waste is around 35%, and this high rate of waste confirms the need for economic and effective management strategies.

3.3 Modeling Framework

In this section, we develop the model consisting of two major parts, the first is the consideration arrival process, and the second is the discrete-continuous choice model.

3.3.1 Consideration Arrival Process

One limit of the data is that we only observe a customer’s shopping histories in the focal supermarket chain and hence do not know when the customer shops at competitor supermarkets. To overcome this issue, we assume that for each customer
i, his consideration of shopping trips to a store of the BetaMart at day $t$ follows a Poisson arrival process with rate parameter $\lambda_i$, controlling for unobserved customer demand heterogeneities. As a result, the gap between a customer’s two arrivals (measured by day), $g_{it}$, follows an exponential distribution with mean $\lambda_i$

$$\text{Prob}(g_{it} = x) = \frac{1}{\lambda_i} \exp\left(-\frac{x}{\lambda_i}\right)$$

(3.1)

### 3.3.2 Discrete-Continuous Choice Model

The consumption of goods is central to most utility maximization problems and is described in this section. In the following, we utilize a discrete-continuous model of demand following Wales and Woodland (1983) and Kim et al. (2002). The discrete-continuous model has the advantage of handling corner solutions and continuous quantities across products flexibly while retaining a clear link to a direct utility function (please see Chintagunta and Nair (2011) for a review). Within this set-up, we use Bhat (2005) parameterization of utility, which has been documented to fit scanner panel data well.

On day $t$, customer $i$ decides whether and how much to consume each of $J + 1$ goods by maximizing a direct utility function subject to an expenditure budget and conditional upon trip budget $b_i$:

$$\max_{x_{it}} U(x_{it}) \text{ s.t. } p_t \cdot x_{it} \leq b_i$$

where $x_{it} = (x_{i0t}, \ldots, x_{ijt})'$ is a vector produce quantities purchased, $p_t = (p_{0t}, \ldots, p_{Jt})$ is a vector of produce prices. Some amount of the outside good ($j = 0$) is always consumed, and its price is normalized to one. The total direct utility from consumption is being divided into two sub-utility functions as shown below, where $U_0$ captures the direct utility from consuming the outside good and $U_1$ captures the direct utility
from consumption of the remaining goods:

\[ U(\mathbf{x}_{it}) = U_0(\mathbf{x}_{i0t}) + U_1(\mathbf{x}_{i1t}, \ldots, \mathbf{x}_{ijt}) \]

where

\[ U_0(\mathbf{x}_{i0t}) = e^{\xi_{i0t}} \cdot \mathbf{x}_{i0t} \]

\[ U_1(\mathbf{x}_{i1t}, \ldots, \mathbf{x}_{ijt}) = \sum_{j=1}^{J} \frac{1}{\alpha_{ij}} \cdot \exp(\gamma_j + \beta_j^T \mathbf{Z}_{ij} + \varepsilon_{ijt}) \cdot [(x_{ijt} + 1)^{\alpha_{ij}} - 1] \]

where \( \alpha_{ij} \leq 1 \), is a satiation parameter. Above, \( \varepsilon_{i0t}, \ldots, \varepsilon_{ijt} \) are i.i.d. stochastic shocks that are known to consumer \( i \) but not to the econometrician. \( \varepsilon_{ijt} \) captures idiosyncratic (unobserved) characteristics that impact the baseline utility for good \( j \). \( \mathbf{Z}_{ij} \) is a set of attributes characterizing alternative \( j \) and the decision maker, including (1) weekday dummies; (2) quantities of the customer’s last purchase in this category to control for the stock at home; (3) freshness measured by the order quantity at the category-day level. The utility function is additively separable across products, which are assumed to be substitutes. The implied marginal utility of consumption is:

\[ \frac{\partial U(\mathbf{x}_{it})}{\partial x_{ijt}} = \exp(\gamma_j + \beta_j^T \mathbf{Z}_{ij} + \varepsilon_{ijt}) \cdot (x_{ijt} + 1)^{\alpha_{ij}-1} \]

So the marginal utility of consuming \( x_{ijt} \) depends on the quantity purchased \( x_{ijt} \).

### 3.4 Estimation and Identification

In this section, we discuss about how the model estimates are estimated. Finally, we discuss about the identification of the model parameters.

#### 3.4.1 Discrete-Continuous Choice Model

For notation simplification, we remove the index \( i \) of each customer in the following expression except explicitly specified. The agent maximizes her utility subject to
the budget constraint in the domain of positive quantities. The Lagrangian for the utility maximization problem is,

\[
\max_{x_t \geq 0, \lambda_t \geq 0} L(x_t, \lambda) = U(x_t) + \lambda_t \left( E_t - \sum p_{jt} \cdot x_{jt} \right)
\]

which can be solved by the Karush-Kuhn-Tucker (KKT) optimality conditions. One goal of the demand model is to accommodate the mixed distribution of zeros (corner solution) and position quantities observed in the data. A generic observation comprises no-purchase of some produce, and positive quantities of the others by a household in a given day. Without loss of generality, let the first \( K \) goods be consumed (don’t include the outside good), and the consumption of the rest to be zero. Then the KKT conditions for the observation imply that

\[
\begin{align*}
\frac{\partial L}{\partial x_{jt}} & = 0, \quad \text{for } x_{jt} > 0, j \leq K \\
\frac{\partial L}{\partial x_{jt}} & \leq 0, \quad \text{for } x_{jt} = 0, j > K \\
x_{0t}, x_{jt}, \lambda_t & \geq 0, \quad j \in J
\end{align*}
\]

The problem above can be simplified by solving the optimality conditions with respect to \( \lambda_t \) and taking logarithms. Following the standard procedure of differencing with respect to the outside good (see for e.g., Chintagunta and Nair 2011), it follows that at the optimum

\[
\begin{align*}
V_{0t} - V_{jt} & = (\varepsilon_{jt} - \varepsilon_{0t}) / \phi, \quad x_{jt}^* > 0, j \leq K \quad (3.3) \\
V_{0t} - V_{jt} & \geq (\varepsilon_{jt} - \varepsilon_{0t}) / \phi, \quad x_{jt}^* = 0, j > K \quad (3.4)
\end{align*}
\]

where

\[
\begin{align*}
V_{0t} & = 0 \quad (3.5) \\
V_{jt} & = \gamma_j + \beta^T Z_j + (\alpha_j - 1) \log (x_{jt} + 1) - \log (p_{jt}) \quad (3.6)
\end{align*}
\]
We can now derive the implied joint probability of purchasing and density of quantities by integrating the random utility errors, \( \varepsilon_{0t}, \cdots, \varepsilon_{Jt} \), over the implied KKT conditions in Equation (3.3) and Equation (3.4). Assuming \( \varepsilon_{0t}, \cdots, \varepsilon_{Jt} \) follow a Type-1 extreme-value distribution (i.e., the pdf is \( f(\varepsilon) = e^{-\varepsilon} \phi e^{-\varepsilon} \phi e^{-\varepsilon} \) with a scale parameter of \( \sigma \) (\( \sigma \) can be normalized to one if there is no variation in unit price across goods), the mixed discrete-continuous density of the outcomes for an agent at day \( t \) conditional on a vector of price becomes,

\[
l(x_{0t}^*, \cdots, x_{Kt}^*, 0, \cdots, 0 | p_t) = K! \frac{\prod_{j=0}^{K} \exp(\phi V_{jt})}{\left( \sum_{j=0}^{J} \exp(\phi V_{jt}) \right)^{K+1}} |\mathcal{J}| \tag{3.7}
\]

where \( V_{jt} \) are as defined in Equation (3.5) and Equation (3.6) and \( |\mathcal{J}| \) is the determinant of the Jacobian induced by the nonlinear change in variables transformation from the density of the error terms to the density of the purchased quantity. The elements of \( \mathcal{J} \) are

\[
|\mathcal{J}|_{l,k} = \frac{\partial (\phi V_{0t} - \phi V_{lt} + \varepsilon_{0t})}{\partial x_{kt}^*} = 1 \quad (l = k) \quad \frac{\phi (1 - \alpha_k)}{1 + x_{kt}^*}, \quad l, k = 1 \cdots K
\]

and its determinant is \( \left[ \prod_{k=1}^{K} f_{kt} \right] \), where \( f_{kt} = \frac{\phi(1-\alpha_k)}{1+x_{kt}^*}, \quad k = 1, \cdots, K \). see Appendix A.2 for a full derivation.

We can further allow for a more flexible error structure by using the Mixed MDCEV model, partitioning the error term as \( \varepsilon_{jt} = \varepsilon_{jt}^1 + \varepsilon_{jt}^2 \), where \( \varepsilon_{jt}^1 \) is i.i.d. Type I extreme value distributed with scale parameter \( \sigma \), and \( \varepsilon_{jt}^2 \) is allowed to be correlated across items, which further allows us to estimate the probability of a basket of items. As with the mixed logit model, the integration over the mixing distribution (\( \varepsilon^1 \)) must be performed numerically — we employ simulation methods to compute this integral in our estimation procedure.

The reader will note the absence of the prices from the model in Equation (3.2), which is customary when specifying a direct utility function. Under the direct utility
approach, the consumer is assumed to maximize her utility subject to a budget constraint, which incorporate product prices. Prices affect demand through the (binding) budget constraint and are subsequently reflected in the optimal consumption quantities obtained from solving the constrained optimization problem.

3.4.2 Model Identification

Finally, the consideration arrival rate parameters are identified by the average shopping trip incidence rate conditional the season of the year.

In a non-linear multi-stage model such as the one presented here, functional form and distributional assumptions inevitably plan an important role in parameter identification. Nevertheless, key patterns of variation in the data link unambiguously to certain parameters. Here we discuss those patterns as well as any parameter normalizations required for identification.

The demand model involves three types of parameters: the baseline utility intercepts ($\gamma_j$), the satiation parameters ($\alpha_j$) and the variance of the extreme values shocks common to all categories ($\sigma_\varepsilon$). The $\gamma_j$ values are primarily pinned down by variation in purchase incidence rates of different veggies, while the $\alpha_j$ values relate most directly to the quantity purchased conditional on choosing the veggies. As in the purely discrete choice setting, neither the level nor scale of utility is separately identified. Normalizing the level of utility requires normalizing one of the $\gamma_j$ parameters — for this, we set $\gamma_1 = 0$ in all specifications.

3.5 Results and Discussions

In Table 3.2 are the results of Poisson arrival model and multiple discrete-continuous choice model. From the results of Poisson demand arrival process we can see that the average gap between a customer’s two arrivals is around 9 to 10 days. However, the standard deviations is relatively large ($\Omega(\lambda)$ is around 6 days), which means there
Table 3.2: Estimation Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>$\Omega(\cdot)$</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>9.35***</td>
<td>6.57e-2</td>
<td>5.83***</td>
<td>4.74e-2</td>
</tr>
<tr>
<td>$\beta$</td>
<td>9.39e-1***</td>
<td>2.36e-2</td>
<td>8.82e-2***</td>
<td>6.54e-3</td>
</tr>
<tr>
<td>$\phi$</td>
<td>4.67e-1***</td>
<td>8.60e-3</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-6.18***</td>
<td>1.11e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-11.00***</td>
<td>3.04e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-7.88***</td>
<td>1.43e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>-7.20***</td>
<td>1.41e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_5$</td>
<td>-6.81***</td>
<td>1.19e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_6$</td>
<td>-9.23**</td>
<td>1.92e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_7$</td>
<td>-4.59***</td>
<td>8.51e-2</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_8$</td>
<td>-5.51***</td>
<td>1.14e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_9$</td>
<td>-8.43***</td>
<td>1.64e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>-5.55***</td>
<td>1.01e-1</td>
<td>1.54***</td>
<td>6.54e-2</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>3.05e-1***</td>
<td>1.42e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.73e-1***</td>
<td>8.53e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>3.31e-1***</td>
<td>1.51e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>2.53e-1***</td>
<td>1.68e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>3.70e-1***</td>
<td>1.31e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_6$</td>
<td>2.24e-1***</td>
<td>2.18e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_7$</td>
<td>3.22e-1***</td>
<td>1.32e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_8$</td>
<td>6.28e-2***</td>
<td>1.93e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_9$</td>
<td>2.39e-1***</td>
<td>1.87e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>3.27e-1***</td>
<td>1.35e-2</td>
<td>1.08e-1***</td>
<td>2.43e-2</td>
</tr>
</tbody>
</table>

are large differences between customers’ shopping trip gaps: some customers may come to BetaMart everyday, while others may come average half a month.

The estimation of $\gamma_j$ and $\alpha_j$ ($j=1,2,\cdots,10$) are from the aggregate sales of 10 categories of produces. $\alpha$ is the satiation parameters, which controls for the quantity sold in each category in a shopping trip. $\gamma$ reflects the baseline marginal utility, which controls for the frequency that customers purchase each category. As we can see, customers purchase potatoes with the highest frequency and purchase onions with the largest quantity, while purchase high grade leafy greens with both the lowest frequency and quantity.

We use the quantity of newly arrived produces in each category in every day as
a measure of freshness. As we can see, the coefficient $\beta$ is positive and significant, meaning that freshness can help increase demand. However, we realize there is endogeneity problem: the large quantity of order in a category in a day may be induced by the high anticipated demand. Thus, we need further modification of the variable or model to control for the endogeneity problem.

3.6 Counterfactual Experiments and Managerial Implications

In this section, we discuss how our demand estimation can be used to develop smart operations management strategies.

3.6.1 Retail Price Optimization

We consider the profit optimization problem of supermarket operations managers. The profit is consisted with two parts: (1) the direct profit from selling fresh veggies to consumers; (2) the waste incurred by throwing out veggies at the end of their shelf lives. So we can write the optimization problem as follows

$$
\max_{\{p_{jt}\}_{j,t}} \sum_{i,t} \int_{\{q_{ijt}\}_j} \left\{ \sum_j q_{ijt} (p_{jt} - w_{jt}) \right\} dF \{q_{ijt}\}_j - \sum_{j,t} W_{jt}w_{jt},
$$

where $i$ denotes consumer, $w_{jt}$ is the wholesale price of product $j$, $q_{ijt}$ is the purchase quantity and $W_{jt}$ is the quantity of product $j$ that is discarded at time $t$. Here cumulative function $F \{q_{ijt}\}$ can be empirically computed from $P (q_{1t}, q_{2t}, \ldots, q_{Kt}, 0, \ldots, 0)$ which is obtained from the demand model. The waste $W_{jt}$ is also a function of $F \{q_{ijt}\}$ and inventories decisions made before.

3.7 Next Step

This study is not finished yet. The next step is to find the right instrument variable(s) to help identify the effect of: (1) freshness and (2) shelf stock, on demand. To find
the right instrument variables could be challenging; currently we are considering the following factors as potential instruments (a) weather; (b) order information from other stores; (c) delayed or cancelled orders. After estimating the impact of these two factors, with the help of instrument variables, we would like to study an optimization problem of pricing and inventory management so as to not only optimize profit but also reduce food waste (see §3.6.1 for the optimization equation).
Appendix A

Appendix of Essay 1

A.1 Data Selection and Missing Data Interpolation

This section explains the data manipulation steps we used to generate the data set for model estimation.

A.1.1 Data Selection

We focus on customers with at least 5 shipments during January to March (to construct the pre-estimation sample that is explained in §4.3.1) and at least 15 shipments in the last 9 months in 2013 (to construct the data for model estimation). Among the customers that have shipped no less than 5 times in the year, only 12\% of them have only shipped on one route. Since we are interested in customer’s learning spillover, specifically across multiple routes, and to limit the computational complexity, we limit our focus to customers with shipping experiences on 2 ~ 10 routes. For the same reason, we only consider customers whose most frequent route accounts for no more than 70\% of his total shipments during the year. In addition, we also exclude customers with too many shipments (exceeding 100 shipments in our case), because
these customers may have deep long-term business partnership with AlphaShip and share integrated information system that make them very reluctant to change service suppliers in a short period (such as the one year time in our data) even after experiencing poor services. We further exclude data with obvious errors (e.g., shipments with arrival times earlier than their departure times).

A.1.2 Missing Data Interpolation

Our data, the purchase data, records the shipping information of each shipment, as described in §3. We only observe these information of the purchased shipping services, but in order to estimate the choice model (Equation (2.3)), we need to reconstruct the data series to include the control variables $X_{ijt}$ and $Price_{ijt}$ for the periods when there is no purchase and for all frequent routes of each customer. The method to reconstruct price data is described in §3.3. For the cargo-related information in $X_{ijt}$ — cargo volume and pieces — for simplicity we use the customer-route level average to interpolate the missing data.

A.2 Supplementary Material for Customers’ Bayesian Learning Model

This section includes the supplementary material for the estimation of customers’ Bayesian learning model.

A.2.1 Pre-estimation for Customer-level Priors

At the beginning of period 1, all the customers have the same priors of quality evaluation (see column $t = 1$ in Table A.1). After calibrating the individual learning using the pre-estimation sample (i.e., the first 24 periods data), customers have their individual priors at the beginning of period 25, the column $t = 25$ of Table A.1 provides the summary statistics of customer-level priors.
Table A.1: Pre-estimation Priors and Model Priors

<table>
<thead>
<tr>
<th></th>
<th>$t = 1$</th>
<th></th>
<th>$t = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Quantile (5%, 95%)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_\sigma$</td>
<td>1.05</td>
<td>5.39</td>
<td>(3.55, 8.55)</td>
</tr>
<tr>
<td>$\delta_\sigma$</td>
<td>10</td>
<td>40.95</td>
<td>(14.13, 118.03)</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>0</td>
<td>-0.77</td>
<td>(-3.26, 1.77)</td>
</tr>
<tr>
<td>$\sigma_\mu$</td>
<td>30</td>
<td>1.15</td>
<td>(0.36, 2.17)</td>
</tr>
<tr>
<td>$\alpha_\xi$</td>
<td>1.05</td>
<td>2.37</td>
<td>(2.05, 3.05)</td>
</tr>
<tr>
<td>$\delta_\xi$</td>
<td>3</td>
<td>9.51</td>
<td>(5.65, 16.13)</td>
</tr>
</tbody>
</table>

A.2.2 Advantages of Gibbs Sampling

In Bayesian statistics, the posterior distribution is often not available analytically, typically because the posterior distribution requires the computation of a normalizing constant which is typically not available in closed form. For this reason, posterior calculations usually rely on Markov chain Monte Carlo (MCMC) sampling. The basic idea in MCMC sampling is to construct a Markov chain having a stationary distribution corresponding to the joint posterior distribution of the model parameters, with this done in a manner that avoids ever having to calculate the intractable constant.

In order for the Markov chain to have the appropriate behavior, the Markov transition kernel needs to be carefully chosen, with usual choices corresponding to either Metropolis–Hastings (MH) or Gibbs sampling. MH can involve a lot of tuning in models with many parameters, while Gibbs avoids tuning by sampling sequentially from the conditional posterior distributions of subsets of parameters given current values of the other parameters. Gibbs sampling relies on a property known as conditional conjugacy. Focusing on a subset of the model parameters and conditioning on the other parameters, the prior probability distribution is conditionally conjugate if the conditional posterior distribution takes the same form as the prior. The specific choices of our model form and prior distributions are motivated by retaining conditional conjugacy (see Equation (2.5), (2.6), (2.7) and (2.8)). In addition, the update
mechanism of Gibbs sampling is also more intuitive than that of MH, thus is suitable for our model assumption that customers learn from their experiences approximately following their intuitions and rough calculations.

A.2.3 Gibbs Sampling of the Regression Hierarchical Model

Gibbs sampling of the full proposed model in §4.3.3 can be conducted by using the following full conditional distributions of the model parameters, and the sampling process is iterated multiple times until convergency (we use “⋯” to indicate all the other parameters and data):

\[
\begin{align*}
\theta_j | ⋯ & \sim N \left( \frac{n_j \xi^2 G_j + \sigma^2 \mu}{n_j \xi^2 + \sigma^2}, \frac{\xi^2 \sigma^2}{n_j \xi^2 + \sigma^2} \right), \\
\mu^d | ⋯ & \sim N \left( \frac{\sigma^2 \mu^d G^d + \sigma^2 \mu_0}{n^d \sigma^2 \mu^d + \sigma^2}, \frac{\sigma^2 \mu \sigma^2}{n^d \sigma^2 \mu^d + \sigma^2} \right), \\
\sigma^2 | ⋯ & \sim IG \left( \alpha_\sigma + \frac{1}{2} \sum_{j \in Y_i} n_j, \delta_\sigma + \frac{1}{2} \sum_{j \in Y_i} \sum_{p=1}^t y^*_{jp} \left( Q_{jp} - \theta_j - \mu^d \cdot Distance_j \right)^2 \right), \\
\mu | ⋯ & \sim N \left( \frac{J \sigma^2 \bar{\theta} + \xi^2 \mu_0}{J \sigma^2 \mu + \xi^2}, \frac{\sigma^2 \mu \xi^2}{J \sigma^2 \mu + \xi^2} \right), \\
\xi^2 | ⋯ & \sim IG \left( \alpha_\xi + \frac{J}{2}, \delta_\xi + \frac{1}{2} \sum_{j \in Y_i} (\mu_j - \mu)^2 \right). \quad (A.1)
\end{align*}
\]

where \( y^*_{jt} \) is an indicator that is set to 1 if customer \( i \) has a shipment on route \( j \) delivered during period \( t \), and 0 otherwise. Specifically, \( J = |Y_i|, \bar{\theta} = \frac{1}{J} \sum_{j \in Y_i} \theta_j, G_j = \frac{1}{n_j} \sum_{p=1}^t \left[ y^*_{jp} \cdot (Q_{jp} - \mu^d \cdot Distance_j) \right], n_j = \sum_{p=1}^t y^*_{jp}, n^d = \sum_{j \in Y_i} \sum_{p=1}^t (y^*_{jp} \cdot Distance_j^2) \) and \( G^d = \sum_{j \in Y_i} \sum_{p=1}^t \left[ y^*_{jp} \cdot (Q_{jp} - \mu_j) \cdot Distance_j \right] \).

A.2.4 Computation

The Gibbs sampling estimation of the simple (see §4.3.2) and regression model (see §4.3.3) is carried out in the same way: for each customer in each period, if new
information is available to the customer (i.e., the customer observes the service quality of a shipment with AlphaShip), 1000 samples are iteratively drawn from the full-conditional posterior distributions, where the first 500 samples are discarded as burn-in. To speed up convergence, for each period, we use the last period’s estimate (e.g. \( \mu_{E_{i,t-1}}, \mu_{E_{i-1}}, \sigma_{p_j}^E \)) as a starting point. Finally, since customers learn service quality independently, parallel computing (32 threads) is adopted, which significantly reduces the computational time. Accordingly, each of the 32 threads generates the posterior draws for a different customer. This code was implemented in Matlab, and the longest running time was 50h on a 2.96-GHz Intel Xeon E5-2690 computer with 32 cores.

A.3 Supplementary Material for Choice Model Estimation

This section provides supplementary material for the computation details, simulation experiments to test and validate our methodology and estimation results of the shipping choice model.

A.3.1 Shipping Choice Model Estimation

The estimation of the demand arrival and shipping choice model parameters (Equation (3.1) and (2.3)) is implemented using a simulated maximum likelihood estimation technique (Train, 2009). The goal is to estimate (i) the distribution of the intercept, price and service quality sensitivity parameters, which are governed by \( \beta \) and \( \Omega \); and (ii) the vector of parameter, \( \beta^X \), related to the control variables. The likelihood function is then given by:

\[
L = \int \prod_i \prod_t \left[ \left( \prod_{j \in \Gamma_i} (\lambda_i m_{ij} P_{ijt})^{y_{ijt}} \right) \left( 1 - \sum_{j \in \Gamma_i} \lambda_i m_{ij} P_{ijt} \right)^{1 - \sum_{j \in \Gamma_i} y_{ijt}} \right] dF(\theta)
\]
where $\theta$ is the vector of draws from a multivariate normal distribution with mean $\beta$ and $\Omega$. We resort to Halton draws to simulate the above integral over the parameter spaces in order to keep the simulation error low (Train, 2009). We run the simulated maximum likelihood estimation for 100 draws for each of the individual-level parameters of the model.

A.3.2 Simulation Experiment

In this subsection, we test the proposed methodology using simulated data. We generate purchase and service experience data for 100 customers and 50 periods. Specifically, each customer can ship from 2 to 6 routes in the 50 periods, and we simulate his shipping demand in each period from a multinomial distribution. Then, we simulate the service experience (i.e., delays) data from a $N(p_0, 1)$ distribution. In each period, within the arrival of each new experience, the customer updates his belief about delays following the simple Bayesian hierarchical model and their prior beliefs are flat priors as we use in §A.1. Then, the beliefs about service quality (represented by $x_1$), together with the intercept and another variable (represented by $x_2$) are used in the utility function, where the values of $x_2$ are generated from $N(0, 1)$. We use random coefficients for the intercept and transport delay $x_1$, and assume fixed (i.e., constant across customers) coefficients for $x_2$. Accordingly, The coefficient of the intercept, $x_1$ and $x_2$ are generated from a multivariate normal distribution with mean $\bar{\theta} = [-0.5, 0.3, -0.4]$ and diagonal variance matrix $\Sigma$ in which elements equal to 0.6 and 0.5 for the intercept and $x_1$, respectively, and the element equals to zero for $x_2$.

Since there are too many parameters in the multinomial distribution demand model (see §???), instead of listing all the estimation values in a table, we plot the estimates versus the true values into Figure A.1. As we can see from Figure A.1, the estimates are close to the true value (i.e., the dots spread evenly around the $y = x$
Table A.2: Actual vs. Estimated Values of the Discrete Choice Model Parameters

<table>
<thead>
<tr>
<th></th>
<th>True value</th>
<th>Estimated value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. err.</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>-5e-1</td>
<td>-5.5e-1</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>3e-1</td>
<td>2.8e-1</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>-4e-1</td>
<td>-4.2e-1</td>
</tr>
<tr>
<td>$\Omega (\theta_1)$</td>
<td>6e-1</td>
<td>5.7e-1</td>
</tr>
<tr>
<td>$\Omega (\theta_2)$</td>
<td>5e-1</td>
<td>4.6e-1</td>
</tr>
</tbody>
</table>

Line). Table A.2 provides the estimates of the discrete choice model parameters, where we can see the estimates are close to the true values and the estimation standard errors are very small compared to the true value. The results in Figure A.1 and Table A.2 show that our method recovers well the original parameters.

A.3.3 Supplementary Material for Results

In Table A.3 are the log-likelihoods under different combinations of the learning model and the quality function specification.
Table A.3: Log-Likelihood of Alternative Model Specifications

<table>
<thead>
<tr>
<th></th>
<th>Benchmark Model 1 (short memory)</th>
<th>Benchmark Model 3A (information pooling)</th>
<th>Benchmark Model 2 (independent learning)</th>
<th>Simple Hierarchical Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>S*</td>
<td>-54,498.1</td>
<td>-53,638.4</td>
<td>-53,638.7</td>
<td>-53,636.7</td>
</tr>
<tr>
<td>A</td>
<td>——</td>
<td>-53,638.3</td>
<td>-53,635.2</td>
<td>-53,632.5</td>
</tr>
<tr>
<td>A + ERA</td>
<td>——</td>
<td>-53,635.7</td>
<td>-53,631.7</td>
<td>-53,630.1</td>
</tr>
<tr>
<td>A + ERA + BUA</td>
<td>——</td>
<td>-53,629.0</td>
<td>-53,607.4</td>
<td>-53,603.8</td>
</tr>
<tr>
<td>A + ERA + BUA + Q</td>
<td>——</td>
<td>-53,627.6</td>
<td>-53,607.3</td>
<td>-53,603.6</td>
</tr>
</tbody>
</table>

*We use S for symmetric, A for asymmetric, ERA for experience risk averse, BUA for belief uncertainty averse, Q for quadratic.
Appendix B

Appendix of Essay 2

B.1 Data Manipulation

This section explains the data manipulation steps we used to generate the data set for model estimation.

B.1.1 Data Cleaning and Selection

We first drop sales records with obvious error, including observations with non-positive sales quantity/price, observations with unreliable numbers, such as extremely small purchase amount (i.e., less than 0.2 yuan) or quantity (less than 0.01 kilogram) or extremely high prices (i.e., more than 150 yuan per kilogram) etc. Then, we restrict our study to customers with loyalty cards, which allows us to track the shopping history of each customer. Specifically, since we are interested in estimating the arrival process of each customer, we restrict our study to customers who came to BetaMart during at least 10 months in 2013. In addition, since we focus on estimating the demand of fresh unpackages veggies, we further restrict our sample to customers who have shopped veggies in at least 5 trips in 2013. As a results, we have
6,832 customers for analysis.

For the inventory order and fill data, we drop the observations with obvious error, including non-positive order quantity or order amount.

B.2 Supplementary Material for Model Estimation

In this section, we provide the supplementary materials for understanding the multiple discrete-continuous extreme value (MDCEV) model and its estimation.

B.2.1 Derivation of the MDCEV Model Likelihood

From Equation () of the text:

\[
P(x_{0t}, \cdots, x_{Kt}, 0, \cdots, 0)
\]

\[
= |\mathcal{J}| \frac{1}{\sigma^K} \left[ \prod_{i=1}^{K} e^{\phi(V_1 - V_i)} \right] \left[ \int_{\varepsilon_0 = -\infty}^{\varepsilon_0 = +\infty} 
  e^{-\phi \varepsilon_0} K^j e^{-\sum_{i=1}^{J} \phi(V_1 - V_i) - \frac{1}{\sigma} e^{-\phi \varepsilon_0} d\varepsilon_0} \right]
\]

Now, consider the last term within the integral in the expression above, and let \( t = e^{-\phi \varepsilon_0} \). Then \( dt = -e^{-\phi \varepsilon_0} \cdot \phi d\varepsilon_0 \), and we can write the integral as

\[
- \int_{t=+\infty}^{0} t^M \cdot e^{-t \sum_{k=1}^{K} e^{-\phi(V_1 - V_k)}} dt = \frac{M!}{\left[ \sum_{k=1}^{K} e^{-\phi(V_1 - V_k)} \right]^M}
\]

putting this back to B.1 we get

\[
P(x_{0t}, \cdots, x_{Kt}, 0, \cdots, 0)
\]

\[
= \phi^K |\mathcal{J}| \left[ \prod_{j=1}^{K} e^{-\phi(V_1 - V_j)} \right] \left[ \frac{1}{\left( \sum_{j=1}^{J} e^{\phi(V_1 - V_j)} \right)^K} \right] \left[ \frac{K!}{\left( \sum_{i=1}^{I} e^{\phi V_i} \right)^K} \right]
\]

\[
= \phi^K \left[ \prod_{i=1}^{K} c_i \right] \left[ \frac{1}{c_i} \right] \left[ \frac{\prod_{i=1}^{K} e^{\phi V_i}}{\left( \sum_{j=1}^{J} e^{-\phi V_i} \right)^M} \right] \left[ \frac{K!}{\left( \sum_{j=1}^{J} e^{-\phi V_i} \right)^M} \right]
\]

80
where \( c_i = \left( \frac{1 - \alpha_i}{e_i^{\gamma_i} + \gamma_i p_i} \right) \).

### B.2.2 Simulated Maximum Likelihood Estimation

From Equation (B.1), the log-likelihood function of observing the data set can be written as (we should work with precision here):

\[
\log L = \sum_i \sum_t \left\{ \sum_{k=1}^{K_{it}} \left[ \ln \phi \left( \frac{1 - \alpha_k}{1 + x_{ikt}} \right) + \ln (k) + \phi V_{ikt} \right] - (K + 1) \ln \left( \sum_{j=0}^{J} \exp (\phi V_{ijt}) \right) \right\}
\]

(B.1)

Since \( \alpha = 1 - e^\delta \), \( d\alpha = (\alpha - 1)d\delta \). Using the facts that \( \frac{\partial V}{\partial \beta} = Z \), \( \frac{\partial V}{\partial \gamma} = 1 \), and \( \frac{\partial V}{\partial \delta} = (\alpha - 1) \log (1 + x_{jt}) \), the gradient can be written as

\[
\begin{align*}
\frac{\partial \log L}{\partial \beta} &= \sum_i \sum_t \left\{ \sum_{k=1}^{K_{it}} \phi Z_{ikt} - \eta \sum_{j=1}^{J} \phi Z_{ikt} \exp (\phi V_{ijt}) \right\} \\
\frac{\partial \log L}{\partial \gamma_{ij}} &= \sum_i \sum_t \left\{ \phi 1 \{ j \leq K \} - \eta \phi \exp (\phi V_{ijt}) \right\} \\
\frac{\partial \log L}{\partial \delta_{ij}} &= (\alpha - 1) \sum_i \sum_t \left\{ \phi \ln (1 + x_{ijt}) - \frac{1}{1 - \alpha_{ikt}} \right\} \{0 \leq j \leq K\} \\
&\quad - \eta \phi \ln (1 + x_{ijt}) \exp (\phi V_{ijt}) \\
\frac{\partial \log L}{\partial \sigma} &= \sum_i \sum_t \left\{ \sum_{k=1}^{K_{it}} \left( V_{ikt} + \frac{1}{\phi} \right) - \eta \sum_{j=1}^{J} V_{ijt} \exp (\phi V_{ijt}) \right\}
\end{align*}
\]

where \( \eta = \frac{K+1}{\sum_{j=0}^{J} \exp (\phi V_{ijt})} \).

The estimation of the demand arrival and shipping choice model parameters (Equation (3.1) and (3.2)) is implemented using a simulated maximum likelihood estimation technique (Train, 2009). We resort to Halton draws to simulate the above integral over the parameter spaces in order to keep the simulation error low (Train, 2009). We run the simulated maximum likelihood estimation for 100 draws for each of the individual-level parameters of the model.
Bibliography


Biography

Yan Shang was born in June 18, 1988 in Luoyang, China. She is the only child of Youlin Shang and Min Yan. She is the wife of Yuli Hu.

Yan attended Tsinghua University in Beijing, China, where she earned her Bachelor of Science in Electronic Engineering in 2010. She then attended Duke University, where she earned her Master of Science in Statistics in 2014 and Doctor of Philosophy in Business Administration in 2016.

After graduating from Duke in 2016, she will join Facebook Inc. at Menlo Park, CA as a Research Scientist.