Springer Series in Supply Chain Management

Saibal Ray
Shuya Yin *Editors*

Channel Strategies and Marketing Mix in a Connected World

Springer

Springer Series in Supply Chain Management

Volume 9

Series Editor
Christopher S. Tang
University of California
Los Angeles, CA, USA

More information about this series at http://www.springer.com/series/13081

Saibal Ray • Shuya Yin
Editors

Channel Strategies and Marketing Mix in a Connected World

Springer

Editors
Saibal Ray
Bensadoun School of Retail Management
McGill University
Montreal, QC, Canada

Shuya Yin
Paul Merage School of Business
University of California, Irvine
Irvine, CA, USA

ISSN 2365-6395 ISSN 2365-6409 (electronic)
Springer Series in Supply Chain Management
ISBN 978-3-030-31732-4 ISBN 978-3-030-31733-1 (eBook)
https://doi.org/10.1007/978-3-030-31733-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

In the process of selling a product or service, firms usually consider various components of the marketing mix and value proposition to influence consumers' purchase behaviors, such as product design, advertising, delivery and convenience, pricing, and promotions. This mix varies depending on consumer characteristics in the market that the firm is targeting, the specific distribution channel(s) and related strategy, the level of product information disclosure, or firms' environmental concerns. Keeping in mind the growing digitalization of business processes in the retail world and the move towards omni-channel retailing, this book aims to revisit the "traditional" interactions between channel strategies and the marketing mix in a connected world. By collecting state-of-the-art academic studies along these dimensions, this book would enhance our understanding of the potential impact that the new technologies and strategies can have on practice in the near future. In particular, we divide the chapters in the books along how digitilization of the retail channel affects the following three aspects: consumers, products, and sustainability.

1. ***Consumers*** represent the demand side of the value chain. Consumer characteristics and especially their behaviors in the Internet era shape how products or services should be distributed, how prices should be set, and how market uncertainty is formed, etc. Therefore, when firms consider different channel strategies, it is important to have a proper understanding of a "modern" consumer's decision-making process and his/her utility function. This might involve their impulsiveness or patience when making purchase decisions, their price-sensitivity, how willing they are to collaborate for consumption purposes, how they react to environmental cues, how sensitive they are to health or environmental concerns, etc. Subsequently, firms should take this understanding into account when constructing their value chains and make their decisions accordingly. The first three chapters (Chapters 1–3) of this book focuses on consumers.

   Chapter 1 demonstrates that demand of a particular product may be uncertain and hard to predict, especially when there are competing brands in the market. Failure to correct the errors in demand estimation may create biases which could further lead to miscalculation of operational decisions such as stocking levels. The uncertain demand could be caused by a new innovative product for which consumers are unsure about the utility to be gained from consuming the product. Given such a product, Chapter 2 analyzes firms' optimal promotion and pricing strategies when consumers' anxiety can be mitigated by learning from experiences of early adopters of this product. Chapter 3 also discusses a number of instruments that firms may adopt to disclose product information to consumers so as to reduce uncertainty in consumers' valuation of a certain product and hence to increase the benefit they can obtain from it. These discussions indicate that understanding consumers' psychological dimension when they are considering purchasing a product can help firms improve their operational decisions and distribution channel strategies.

2. ***Products*** are the core to success for any business and represent the supply side of the value chain. Given a product type, its value proposition is affected by various pricing, promotion, and advertising strategies that firms in a channel might use to influence consumers' purchasing decisions. Building on some of the strategies examined in Chapters 2 and 3, the book presents five chapters (Chapters 4–8) that address how these strategies need to be modified when consumers are "connected" via various media.

   Chapter 4 studies promotion planning for supermarket retailers that sell a large number of different items whose sales are interrelated cross-sectionally due to their complementarity or substitutability and also longitudinally across time periods due to consumers' stockpiling behavior caused by promotions. Chapter 5 provides a different perspective on the topic of advertising. It specifically reviews the academic work on how the promotion or advertising strategies can be implemented online, given the rapid explosion of the digital advertising industry. The Internet era also leads to the prevalence of e-commerce, which provides retailers an additional vehicle to reach out to their consumers. Chapter 6 focuses on the omni-channel retailing strategy that retailers adopt to sell their goods to consumers through both online and offline channels, taking into consideration consumer behavior in the digital era. This chapter shows that the omni-channel strategy can help mitigate two key problems in retailing: stockouts and product misfit. Subsequently, Chapter 7 explores the adoption of on-demand customization technology (such as additive manufacturing or 3D printing) in retailing and studies its impact on the online and offline distribution channels in terms of the product variety offered and the pricing strategies. In addition to the impact of new technologies on retail channels, Chapter 8 focuses on price matching strategies offered by competing retailers to price-sensitive consumers, where the online information search has made it much easier for consumers to compare prices charged by competing sellers. This chapter examines how consumer behavior and channel structure influence the effectiveness of such strategies.

3. ***Sustainability*** In addition to pricing, promotion, advertising, and distribution convenience, there are other factors (mostly non-pricing) that also affects consumer purchase behavior. Collaborative strategies among micro-retailers in developing countries and the consideration of retail sustainability are among them. The book uses the last two chapters (Chapters 9 and 10) to cover

these topics. Chapter 9 explores the collaborative strategies that entities in public or private sectors can take in order to help micro-retailers in developing countries coordinate their inventory replenishment strategies and serve their communities better. The study unveils several key trade-offs associated with these collaborative strategies. Finally, starting with an overview of the most common and significant environmental impacts of retail, Chapter 10 examines the origin of sustainable planning and operations in retail, the business case for sustainability programs, and the maturation of retail sustainability programs. This chapter includes business-actionable steps for retail sustainability practitioners and describes the critical programmatic components for a strong retail sustainability program.

We would like to extend our special thanks to Christophe S. Tang for giving us this opportunity to edit this refereed book as part of the Springer Series in Supply Chain Management. We also thank the following reviewers for their helpful and constructive comments in the review process: George Cai, Rachel Chen, Nagesh Gavirneni, Luyi Gui, Ho-Yin Mak, Arcan Nalca, Karthik Ramachandran, Nathan Yang, John Turner, and Wei Qi.

Montreal, QC, Canada — Saibal Ray
Irvine, CA, USA — Shuya Yin

Contents

Chapter 1
Estimating Demand with Constrained Data and Product Substitutions

Mark E. Ferguson

Abstract The inventory and revenue management models most commonly taught in the operations management and industrial engineering disciplines typically assume that the demand for a product is easily estimated and is independent of competing products offered through the same channel. In this chapter, we show why this is rarely a good assumption and provide a review of the statistical techniques that have been developed to correct product demand distribution estimates that are biased due to truncated demand or product substitution effects. Failure to correct product demand estimates for these biases has been shown to result in significantly missed opportunities in meeting customers' true demand due to incorrectly calculated optimal product stocking levels.

Keywords Demand estimation · Product substitutability · Consumer preferences · Choice behavior

1.1 Introduction

Every organization that sells a physical product or service faces out-of-stock events and, potentially, customer demand substitution effects. The impact of these occurrences on the accuracy of a firm's historical demand data is often called the truncated demand problem, a problem that is prevalent across a wide range of industries. Airlines and hotels, for example, deliberately ration the capacity made available to lower-fare customers to protect their remaining capacity for the more profitable, higher-fare customers. Often times, however, the capacity protected for these customers is insufficient to serve all those who would like (and could afford) to buy the ticket at that price. Hence, a portion of the customers that are willing to buy at the previously provided price is frequently denied a reservation due to a

M. E. Ferguson (✉)
Management Science Department, University of South Carolina, Columbia, SC, USA
e-mail: mark.ferguson@moore.sc.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_1

lack of available capacity (at least at the previous price). While most retailers do not intentionally remove their available merchandise from their store shelf or website, they also face stock-out situations whenever demand exceeds their predetermined stocking levels (typically set based on forecast demand plus some safety stock). Consider, for example, what happens when a particular stock keeping unit (SKU) of toothpaste, Tartar Control with Teeth Whitening Colgate toothpaste, at a Kroger grocery store stocks out. While some customers who would otherwise purchase the product and counted on it being available in the store will substitute for another product type or brand, others will leave disappointedly and most likely purchase their preferred product at a competing store.

The airline and toothpaste examples above are different in regard to how out-of-stocks form and manifest but they also share some important similarities. First, out-of-stocks (or sell-outs) sometimes lead to lost sales which erode the already thin profit margins that many organizations face. Second, unless controlled for, the out-of-stock/substitution events provide a distorted view of the demand for a product that will often lead to biased forecasts for future demand and, subsequently, suboptimal future performance of decisions that use historical demand data as an input. Conlon and Mortimer (2013) describe a study involving the tracking of inventory levels of the various products sold through campus vending machines and record when products sold in the machines stock-out before a scheduled replenishment takes place (typically every 4 h in their setting). These stock-out events are not random, however, as they occur when the demand is higher than normal. Conlon and Mortimer (2013) show that the average sales rate during a period that includes stock-outs is nearly three times the sales rates of the periods that did not include any stock-outs. To help demonstrate the intricacies and complexities of the truncated demand problem, we begin with one of the most common and basic inventory management problems—the newsvendor problem.

Practically every undergraduate student pursuing a degree in industrial engineering or business administration learns of the problem facing a newsvendor who must decide how many newspapers to bring to their newsstand each morning, without knowing a priori exactly what the demand for the newspapers will be for that day. We are taught, in this problem, that the newsvendor should first fit a distribution to her historical demand data for newspapers, recorded over many days of operating the newsstand and documenting the cumulative demand at the end of each day. The newsvendor should then choose the amount of newspapers to order for an upcoming day by calculating a critical fractile ((cost of being under demand)/(cost of being under plus the cost of being over)), which is always some number between 0% and 100%, and choosing the order quantity that represents that fractile in the cumulative demand distribution mentioned before.

There is a reason the newsvendor problem is taught in practically every undergraduate operations management class. Since unsold newspapers become worthless at the end of each day, this problem presents a clean and easy-to-visualize example of one of the main trade-offs found in stochastic inventory management theory, i.e., the cost of balancing an unequal marginal cost of ordering too many units versus ordering too few. What is rarely discussed about this problem, however, is the more fundamental issue of how the newsvendor is supposed to estimate her distribution of demand that is required to make this calculation.

Let us give the newsvendor a name, call her Susan. Susan runs a newsstand on a busy corner of town from 7 a.m. to 5 p.m. each day. She must purchase all the newspapers that she plans on selling each day from the local newspaper's printing company by 6:30 a.m., and she does not have the opportunity to replenish her stock during the day. On days where she sells all of her newspapers before 5 p.m., she simply closes the stand and goes home early. Susan buys her newspapers for $0.25 each and sells them for $1 each. Thus, each unit of unmet demand costs Susan $0.75 in missed opportunity.

Suppose Susan purchases 100 newspapers at the beginning of each day. At the end of the first day, she ends up with five unsold newspapers. At the end of the second day, she ends up with no unsold newspapers. Since Susan has taken an operations management class and learned about the newsvendor problem, she dutifully records her sales each day and has the following observations of demand after her first 10 days: [95, 100, 100, 92, 100, 88, 96, 100, 98, 100]. Notice that she never observes sales of more than 100, which makes sense because she only has a maximum of 100 newspapers to sell each day. This common problem of recording sales data, rather than actual demand data, is known as "constrained" demand, since the data is constrained by the maximum amount of inventory that was available to sell.

So why is constrained demand data a problem? First, it distorts the estimated demand distribution, resulting in a lower estimated mean and variance than those of the true demand distribution. Since the "optimal" order quantity is calculated using this estimated demand distribution, this means that the newsvendor equation will result in a lower order quantity than what it should have been if the true demand distribution was used to calculate it. While this is a bad outcome, it gets worse over time because a lower order quantity means that any new estimates of the demand distribution will also be truncated at this lower order quantity. Thus, the newsvendor continues to update their estimate of the demand distribution using increasingly truncated data and, thus, continues to calculate lower and lower values for the optimal order quantity. This results in profits that continue to decrease over time, diverging further from the theoretical maximum profit predicted by the newsvendor model when the true demand distribution is known. If Susan continues along this path, she will soon be out of business. This situation was definitely not discussed in her operations management class.

So what can be done about this unfortunate situation? Theoretically, Susan could stay in her newsstand until the end of each day and record the number of customers that stopped by and asked for a paper, even on days when she had sold out newspapers hours before her normal closing time. This would take a very dedicated effort on Susan's part but, at least theoretically, it is possible. One challenge with directly observing constrained demand is that the customers who intended to buy a newspaper may observe that there are none left and not even bother to stop by the stand to inquire. Another challenge is that not every customer who stops by the stand is willing to pay the price being asked for the newspaper; i.e., they may simply be comparison shopping the price at this newsstand versus another newsstand in a different part of town. While it is relatively straightforward to distinguish the true buyers from the mere shoppers when the stock is on hand, this task becomes considerably more difficult when an item is out of stock. The problem of identifying constrained demand becomes even more challenging in most realistic settings where the retailer carries more than one type of product. In addition, while it is reasonably straightforward to quantify the marginal cost of not having enough newspapers to meet demand on a given day ($0.75), it becomes considerably more difficult to estimate this cost when customers will sometimes substitute another product from the same retailer. In this case, Susan does not lose the $0.75 newspaper profit margin that she did in the single product case because some of those customers that wanted to buy a newspaper (but found it out of stock) will switch to another product instead.

Now suppose that Susan decides to expand her portfolio of products and begins offering a daily news magazine, in addition to the newspaper that she already offers. The daily news magazine has some similar attributes to the newspaper that she already sells. They both cover the local news from the previous day and cost a similar amount. They also differ in some attributes because the newspaper covers a wider variety of stories than the magazine, but the magazine includes more in-depth coverage of the (fewer number of) topics in each issue. The magazine faces the same ordering restriction as the newspaper—Susan must purchase a set number of magazines at the beginning of each day and all unsold copies are disposed of (for zero value) at the end of each day. Susan knows that this new offering should bring in some new customers, who prefer the magazine to the newspaper. Because of the overlap in some of the product attributes, however, she also expects that there will be some existing customers who will switch from the newspaper to the magazine. Thus, Susan changes her daily order quantity of newspapers down to 80 per day, and decides to order 50 magazines each day. Recording her sales of (newspaper, magazine) combinations over the next 5 days, Susan records the following: (74, 41), (80, 48), (79, 50), (68, 38), and (80, 50).

Casually observing these sales numbers indicates that there appears to be a positive correlation between the sales of newspapers and magazines—that is, when the sales for one is high, then the sales for the other also appears to be high. There are two potential explanations for this positive correlation. The first is that the "foot traffic" of potential customers might be higher on certain days, so there are more potential newspaper customers and potential magazine customers on some days than on the other days. This certainly seems to be the case on the fifth day, when both products sold out. We will use the term "arrival rate" to represent this overall foot traffic of customers later in this chapter. A second explanation is that some customers who preferred either a newspaper or a magazine may, upon finding their preferred item out of stock, have switched to the remaining product that was still in stock. Thus, on the second day, the "primary" demand for magazines may not have been 48 because some of this demand may have come from customers who wanted to purchase a newspaper but found that they were already sold out. The same can be said for the demand for newspapers in week three since it appears that the magazines sold out before the end of that day. We will use the term “spillover” demand to represent the observed demand that a product receives from customers who purchased a product only because their first choice was not available at the time of purchase. Compared to the earlier scenario, where Susan only sold newspapers, determining what the primary demand is for newspapers and for magazines just got considerably more complex.

As the example with Susan’s newsstand shows, just adding a second product to a firm’s portfolio further complicates an already challenging problem. In fact, most firms offer hundreds, thousands, or even hundreds of thousands of different products in their portfolio, each of which most likely has correlated demand with a myriad of other products within the same portfolio. At any particular point in time, some of these products will be out of stock (or voluntarily withheld) from the market, so that the primary demand for these unavailable products either goes away or is spilled-over to other products that are still available. While complex, the situation is not hopeless. There has been considerable progress made in how to estimate this primary demand using statistical methods, although this science is still far from being fully developed. In this chapter we review some of these methods, starting with the problem of estimating demand when only a single product is offered and concluding with the case where there is a portfolio of products. In Sect. 1.2, we begin with the most straightforward case of a single product that faces no substitution from other products. In Sect. 1.3, we extend our review to the case with two partially substitutable products where one or both may incur out-of-stocks. In Sect. 1.4, we extend further to the case with more than two partially substitutable products, any of which may incur instances of being out-of-stock.

1.2 Estimating Demand for a Single Product with Out-of-Stocks

We start with Susan’s original problem where she only sold newspapers at her newsstand. The term commonly used for converting observed but “constrained” sales data into an estimate of the true demand distribution for a product is called “unconstraining.” In Susan’s original setting, she ends each day with either a positive or zero inventory of newspapers. In order to accurately apply her newsvendor model to determine how many newspapers she should purchase each morning, we need to estimate a demand distribution using only the historical sales data that she recorded each day. Thus, some of these sales amounts are constrained (typically the ones where sales $= 100$) while others are not (the observations on days when the sales are less than 100). While there is a number of techniques that can be used for the single product problem, we will discuss the Expectation-Maximization method because it is one of the more commonly used ones and because we will build upon this method when we discuss the problem of unconstraining the demand for a portfolio of products. For an alternative approach using Bayesian updating, see Mersereau

(2015). For a more comprehensive review of the unconstraining research in the revenue management field, see Guo et al. (2012).

1.2.1 Estimating Constrained Demand Using the Expectation-Maximization Algorithm

The Expectation-Maximization (EM) algorithm (Dempster et al. 1977) is a statistics-based technique which, for a variety of incomplete-data problems, alternates between an expectation step and a maximum likelihood estimation step until some given convergence criteria are met. The EM algorithm is used in many different statistical applications, where, due to the existence of grouped, censored, or truncated data, the estimation of the maximum likelihood parameters is made difficult by the structure of the corresponding (log-) likelihood function. In such cases, direct optimization over the incomplete-data (log-) likelihood function is problematic for most distributions.

In general, the EM algorithm computes the maximum likelihood estimates for an incomplete-data problem by formulating an associated complete-data problem that is much simpler to solve. By iteratively revising the maximum likelihood estimates for the simpler problem, the EM algorithm ultimately computes the maximum likelihood estimates for the original incomplete-data problem. To get a sense for what the EM algorithm does, we discuss it in a context of Susan's problem of estimating a demand distribution for her newspapers. We assume that the demand for newspapers each day is represented by a series of observations $x = (x_1, \ldots, x_n)$ which are independent and identically distributed and come from a continuous distribution with probability density function $f(x)$. For this set of observations, the complete-data log-likelihood function is given by $\log(L^c(\theta|x)) = \log\left(\prod_{i=1}^{n} f(x_i)\right)$, where $\theta = \left(\theta_1, \ldots, \theta_p\right)$ is a vector of unknown parameters that describe $f\ (x)$ such as the mean and standard deviation. Consider further that a and b are vectors that consist of all uncensored (fully observed values in a) and censored (constrained values in b) observations of x. Recall that Susan's observations of demand after her first 10 days were [95, 100, 100, 92, 100, 88, 96, 100, 98, 100]. For this dataset, vector a is [95, 92, 88, 96, 98] and vector b is [100, 100, 100, 100, 100].

Given this separation of our data, we can write the corresponding incomplete-data log-likelihood function as $\log(L(\theta|a)) = \log\left(\int L^c(\theta|a, b)\, db\right)$. Since the observations in b are censored, the direct maximization of $\log(L(\theta|a))$ to estimate θ is problematic. Instead of this direct approach to computing θ, the EM algorithm solves the incomplete-data problem indirectly by iteratively employing the complete-data log-likelihood function $\log(L^c(\theta|x))$. In particular, during the Expectation step (or, the E-step) of each iteration k of the EM algorithm, the complete-data expected conditional log-likelihood function is computed as

$$Q\left(\theta|\theta_{k-1}\right) = E\left\{\log\left(L^c\left(\theta|x\right)\right)\middle| a, \theta_{k-1}\right\}$$
$$= \int \log\left(L^c\left(\theta|a, b\right)\right) p\left(b|a, \theta_{k-1}\right) db, \tag{1.1}$$

where θ_{k-1} are the estimates for θ revised at iteration $k-1$ and $p\left(b|a, \theta_{k-1}\right)$ is the probability density function of b. In Eq. (1.1), we take the expectation of $\log\left(L^c\left(\theta|x\right)\right)$ given the fully observed data a and the previously estimated parameter estimates θ_{k-1} (vs. $\log\left(L^c\left(\theta|x\right)\right)$) because $\log\left(L^c\left(\theta|x\right)\right)$ is unobservable. With $Q\left(\theta|\theta_{k-1}\right)$ expressed analytically, the Maximization step (or, the M-step) of each iteration k of the EM algorithm estimates θ_k such that $Q\left(\theta|\theta_{k-1}\right)$ is maximized. Upon its completion, the EM algorithm converges to a local maximum, but it has been shown to typically converge (under fairly general conditions) to the global maximum, of the incomplete-data log-likelihood function (Boyles 1983; Dempster et al. 1977; Redner and Walker 1984; Wu 1983).

We next demonstrate the EM method for the specific case of the Normal distribution. For clarity, we slightly alter the notations around x, n, a, b, and θ as follows. We assume that the available $M+N$ observations z_i, $1 \le i \le M+N$, are realizations of a sequence of independent and identically distributed normal random variables. The parameters $\theta = \left(\hat{\mu}, \hat{\sigma}\right)$ of the underlying normal distribution are unknown and must be estimated. Of the $M+N$ observations, M are constrained. For these observations, $z_i = d_i$, where d_i, the available inventory, constrains the true value of z_i. Since the demands for each day are assumed to be independent of each other, we reorder the observations such that the first M observations in z_i are constrained, while the other N are exactly specified. For Susan's data, this results in the vector of observed sales: [100, 100, 100, 100, 100, 95, 92, 88, 96, 98].

If none of these observations were constrained, the complete-data likelihood and log-likelihood functions $L^c\left(\hat{\mu}, \hat{\sigma}|z_i\right)$ and $\log\left(L^c\left(\hat{\mu}, \hat{\sigma}|z_i\right)\right)$ would be

$$L^c\left(\hat{\mu}, \hat{\sigma}\middle| z_i\right) = \prod_{i=1}^{M+N} \frac{1}{\hat{\sigma}\sqrt{2\pi}} \cdot \exp(-(z_i - \hat{\mu})^2/(2\hat{\sigma}^2))$$
$$= \left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\right)^{M+N} \cdot \exp\left(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{M+N}\left(z_i - \hat{\mu}\right)^2\right) \tag{1.2}$$

and

$$\log\left(L^c\left(\hat{\mu}, \hat{\sigma}\middle| z_i\right)\right) = -\frac{M+N}{2} \cdot \ln\left(2\pi\right) - (M+N) \cdot \ln\hat{\sigma} - \frac{\sum_{i=1}^{M+N}\left(z_i - \hat{\mu}\right)^2}{2\hat{\sigma}^2} \tag{1.3}$$

which can be maximized by the closed-form parameter estimates $\hat{\mu}$ and $\hat{\sigma}$:

$$\hat{\mu} = \frac{1}{M+N} \cdot \sum_{i=1}^{M+N} z_i = \frac{1}{M+N} \cdot \left(\sum_{i=1}^{M} z_i + \sum_{i=M+1}^{M+N} z_i \right)$$
$$\hat{\sigma} = \left(\frac{1}{M+N} \cdot \sum_{i=1}^{M+N} \left(z_i - \hat{\mu} \right)^2 \right)^{1/2}. \tag{1.4}$$

If we incorrectly assume that the observed values of newspaper sales from Susan's newsstand were not truncated, we would simply apply the parameter estimators above to the sales data $x = (100, 100, 100, 100, 100, 95, 92, 88, 96, 98)$ to get $\hat{\mu} = 96.9$ and $\hat{\sigma} = 3.96$. Because some of the observations are constrained, however, these estimates for the mean and standard deviation are biased and underestimate the true demand mean and standard deviation. The reason is that in Eq. (1.2), the true z_i values that correspond to the M constrained demand observations are unobserved. Hence, we cannot directly use Eq. (1.4) to compute $\hat{\mu}$ and $\hat{\sigma}$ for the incomplete-data problem. However, by conditioning on a current set of parameter estimates $\hat{\mu}_{k-1}$ and $\hat{\sigma}_{k-1}$, we can replace the values of z_i by their expected values $Z_i^{(k)}$ such that $E\left\{\log\left(L^c\left(\hat{\mu}, \hat{\sigma}|z_i\right)\right) \mid \left(z_i,\ M < i \le M+N\right), \hat{\mu}_{k-1}, \hat{\sigma}_{k-1}\right\}$ can be computed analytically (the E-step). We can then use formulas similar to Eq. (1.4) to suggest revised estimates $\hat{\mu}_k$ and $\hat{\sigma}_k$ for the parameters of the assumed underlying normal demand distribution (the M-step). In formal terms (see, for example, Swan 1969; Wolynetz 1979) the iterative application of the E- and M-steps of the EM algorithm proceeds as follows:

Initialization Initialize $\hat{\mu}$ and $\hat{\sigma}$ to be the sample mean and standard deviation (uncorrected for the degrees of freedom) of all the unconstrained observations. Thus, we express $\hat{\mu}$ and $\hat{\sigma}$ as $\hat{\mu} = \hat{\mu}_0 = (1/N) \cdot \sum_{i=M+1}^{M+N} z_i$ and $\hat{\sigma} = \hat{\sigma}_0 = ((1/N) \cdot \sum_{i=M+1}^{M+N} (z_i - \hat{\mu}_0)^2)^{1/2}$. It should be noted that the EM method does not work when all observations are constrained, so other unconstraining methods are needed to handle these cases.

***E*-Step at Iteration *k*, *k* ≥ 1** Replace z_i, $1 \le i \le M$, with their expected values $Z_i^{(k)}$. The values of $Z_i^{(k)}$ are computed assuming that they are the expected values of the normal distributions $N\left(\hat{\mu}_{k-1}, \hat{\sigma}_{k-1}^2\right)$ left truncated at $d_i = z_i$. Specifically, if X is a normally distributed random variable with mean $\hat{\mu}_{k-1}$ and standard deviation $\hat{\sigma}_{k-1}$, $Z_i^{(k)} = E[X|X > d_i, X \sim N(\hat{\mu}_{k-1}, \hat{\sigma}_{k-1}^2)] = \hat{\mu}_{k-1} + \hat{\sigma}_{k-1} \cdot S(h_i)$ (or, equivalently, $Z_i^{(k)} = E[TN(\hat{\mu}_{k-1}, \hat{\sigma}_{k-1}^2, d_i, +\infty)] = \hat{\mu}_{k-1} + \hat{\sigma}_{k-1} \cdot S(h_i))$, where h_i equals $(z_i - \hat{\mu}_{k-1})/\hat{\sigma}_{k-1}$ and $S\left(h\right)$, the generic hazard function of a normal distribution, can be expressed as $S\left(h\right) = \phi\left(h\right) / \left(1 - \Phi\left(h\right)\right)$. The values $S\left(h_i\right)$ are computed by replacing the generic element h with the corresponding h_i, whereas ϕ and Φ in the expression of the hazard function are the standard normal density function and the standard normal cumulative distribution function, respectively. Alternatively, $Z_i^{(k)}$ can be computed through numerical integration (by using, for example, the function integrated in R) as $\int_{d_i}^{+\infty} x \cdot f\left(x\right)\, dx / \int_{d_i}^{+\infty} f\left(x\right)\, dx$, where $f\left(x\right)$ is the normal probability density function of X.

***M*-Step at Iteration *k*, k ≥ 1** Revise the parameter estimates for $\hat{\mu}$ and $\hat{\sigma}$ by maximizing the expected conditional log-likelihood function

$$E\left\{\log\left(L^c\left(\hat{\mu},\hat{\sigma}|z_i\right)\right)\middle|\left(z_i,\ M<i\leq N\right),\hat{\mu}_{k-1},\hat{\sigma}_{k-1}\right\}. \tag{1.5}$$

The estimates for $\hat{\mu}_k$ and $\hat{\sigma}_k$ are computed using

$$\hat{\mu}_k = \frac{1}{M+N}\cdot\left(\sum_{i=1}^{M} Z_i^{(k)} + \sum_{i=M+1}^{M+N} z_i\right),\quad \text{and,}$$

$$\hat{\sigma}_k = \left(\frac{1}{\sum_{i=1}^{M} T\left(h_i\right)+N}\cdot\left(\sum_{i=1}^{M}(Z_i^{(k)}-\hat{\mu}_k)^2 + \sum_{i=M+1}^{M+N}\left(z_i-\hat{\mu}_k\right)^2\right)\right)^{1/2},$$

respectively, where $h_i = \left(d_i - \hat{\mu}_k\right)/\hat{\sigma}_{k-1}$ and $T\left(h_i\right) = S\left(h_i\right)\cdot\left(S\left(h_i\right)-h_i\right)$.

Convergence Test If $\left|\hat{\mu}_k - \hat{\mu}_{k-1}\right| < \varepsilon$ and $\left|\hat{\sigma}_k - \hat{\sigma}_{k-1}\right| < \varepsilon$, stop; otherwise, proceed with iteration $k+1$ of the EM algorithm. The tolerance ε is typically a small number such as 0.000001. If convergence is reached, $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the parameter estimates of $\hat{\mu}$ and $\hat{\sigma}$. Similarly, $Z_i^{(k)}$ (or, $Z_i^{(k+1)}$) are then used as the unconstrained values of z_i, $i \leq M$.

After running the EM algorithm on Susan's truncated data for the sales of newspapers, we obtain (after 15 iterations and a stopping criteria of 0.000001) new estimates for the demand distribution of to get $\hat{\mu} = 99.59$ and $\hat{\sigma} = 6.93$. Note that while the estimated mean did not change much from the original estimate ($\hat{\mu} = 96.9$) that was based on the truncated data, the estimated standard deviation is approximately 75% larger than the original estimate ($\hat{\sigma} = 3.96$). Since safety stock decisions are often set as a direct multiple of the product demand distribution's standard deviation, the new estimate (based on the unconstrained data) will result in a significantly larger suggested stocking level.

1.2.2 A Different Way of Estimating Demand

A different way of thinking about any demand estimation problem is to consider it as the combination of two different random variables. The first random variable is the probability that a customer "arrives" and is available to make a purchase. The second random variable is, for each customer that arrives, the probability that a particular customer will decide to purchase given the set of products that are presented to him or her. While it is typical to estimate the cumulative demand during some aggregate period of time, such as a day in Susan's case, one can also break a time window into sufficiently small slivers of time such that it is very unlikely that more than one customer will arrive during that sliver of time. For Susan's newsstand business, that could be something like 15 s intervals. Thus, for each 15 s interval, the probability of a single unit of demand (assuming each customer only considers purchasing one unit) can be represented by the probability of a customer arriving, times the probability that a customer, after arriving, makes a purchase. If we estimate this combined probability and then sum it up over all the 15 s intervals included during the working hours that the newsstand is open for a given day, then we should end up with the same daily demand distribution that we typically use for our newsvendor calculations.

So what value do we gain from this new way of estimating demand? In Susan's case when she was only selling one product, the short answer is not much. Susan only faces one real demand-related decision each day, how many newspapers to purchase each morning. Thus, there is little value in breaking down the demand for her newspapers between customer arrivals and a customer's decision to purchase or not. What she needs in order to apply her newsvendor calculation is an estimate of the demand distribution for daily newspaper demand, which she can estimate by applying the EM method on the data she has collected of the daily sales of her newspapers. The real value of breaking demand down into these two separate components comes when we need to estimate the individual product demand for a single product when it is sold alongside similar products that are offered to the customer at the time of purchase. In such a case, customers will often substitute another product from the portfolio when their first choice is unavailable. How to estimate the demand for each product in the portfolio is the topic of our next two sections.

1.3 Estimating Demand for a Portfolio of Two Products with Out-of-Stocks and Substitution Effects

Let us return to the scenario where Susan has added a second product—magazines—to her newsstand. Susan now must estimate the demand for both newspapers and magazines, given that one or both products may have been out of stock at some point during the time when she collected sales data. If Susan employs one of the single product unconstraining techniques discussed in the previous chapter, such as the EM method, she will most likely overestimate the demand for each product. The reason for it is that some customers are likely to choose a substitute product when their first choice is not available. Thus, the observed demand for newspapers will contain some substitute demand during the periods when magazines were out of stock. A similar problem arises when she tries to estimate the demand for her magazines. In Conlon and Mortimer (2013)'s study involving the tracking of inventory levels of the various products sold through campus vending machines, they find that the products that stock-out between replenishments are often the most popular products offered in the machines. Failure to account for substitute demand during these periods of stock-outs results in an underestimation of the demand for the products that stock-out and an overestimation of the "true" demand for the products that do not stock-out.

When attempting to estimate the demand for a product that is part of a larger portfolio of products offered to each potential customer, one must be careful to distinguish between "primary" demand for the product versus "spillover" demand, which occurs when the customer's first choice is not available during the time of purchase. In Susan's case, spillover demand may occur for magazines after her newspapers have sold out, or vice versa. To help distinguish between primary and spillover demand, it is helpful to think of demand in terms of the separate arrival and purchase-choice components described at the end of the last section. That is, demand for a product occurs only when a customer first arrives, evaluates the portfolio of products available, and then chooses to purchase that product.

Anupindi et al. (1998) were among the first to tackle the problem of estimating primary demand for a product when a portfolio of products are offered and the demand data for at least one of the products is constrained. Their setting is a vending machine that sells two products, A and B. For Susan's newsstand setting with the two products, two things are required to make the primary demand estimation for each product straightforward. The first is that the overall customer arrival rate is constant over time and can be represented by a Poisson distribution. That is, the number of potential customers walking by the newsstand is the same at 8 a.m. as it is at 4 p.m. The second required thing is that Susan is meticulous in her record keeping such that she records the exact time during each day whenever a product stock-out occurs. Anupindi et al. (1998) also provide a methodology for estimating primary and spillover demand when the exact timing of the stock-outs is not recorded. This methodology is useful when the inventory levels of the products are only observed periodically, such as those managed by a periodic review system. For brevity, however, we will only describe the first case, where the exact times of the stock-outs are known.

Given that the two requirements above are met, a brief explanation of Anupindi et al. (1998)'s method is as follows. In any given period of time between replenishments (1 day in Susan's case), there can be one of five possible outcomes: Neither product A or B stocks out: (A, B), only Product A stocks out $(\overline{A}, B)$, only Product B stocks out $(A, \overline{B})$, first Product A stocks out and then Product B stocks out $(\overline{A}, \overline{B})$, and first Product B stocks out and then Product A stocks out $(\overline{B}, \overline{A})$. This results in five different intervals of time that can be observed, as each period starts with both product A and B in stock and then either reaches the end of the period in the same state or transitions to one of the other states. If the total amount of time in the period of interest is T, then we can break this into five smaller intervals of time described above as follows. Let $T_{A,B}$ represent the total time that both products were available, $T_{A,\overline{B}}$ represent the total time that only Product A was available, $T_{\overline{A},B}$ represent the total time that only Product B was available, and $T_{\overline{A},\overline{B}}$ represent the total time that neither product was available. Since Susan has recorded the exact times of the stock-outs, she can take several weeks' worth of sales data for her newsstand and break the total time period (for example, 5 weeks $\times$ 5 days/week $\times$ 10 h/day $= 250$ h) down into the exact number of hours that fell into each of the five intervals described above.

In addition to the time interval data, Susan needs to record how much of each product was sold during that time interval. For example, if 10 h of the total time block of 250 h corresponds to time interval $T_{A,\overline{B}}$, then Susan needs to record how many units of Product A were sold during this time interval (there should be no recorded sales of Product B in this interval since it is assumed that Product B was out of stock). Let D_A and D_B represent the total demand observed for Products A and B, respectively, during the time interval $T_{A,B}$. Let $D_{A,\overline{B}}$ and $D_{\overline{A},B}$ represent the total demand observed for Products A and B during the time intervals $T_{A,\overline{B}}$ and $T_{\overline{A},B}$, respectively.

In order to estimate the primary demand for newspapers, for example, Susan needs to estimate the customer arrival rate of the customers whose first choice is for newspapers. The data she has available is the number of items sold for each product during each time block. This is where the requirement of a time-invariant Poisson process comes in. Let the overall customer arrival rate be Λ, which includes all customers that arrive and purchase and those that arrive but do not purchase. Note that the arrival and demand rates are the same in this setting because we assume that each customer only chooses one unit of one option (they choose either Product A, B or the no-purchase option). This overall arrival rate can be broken down into subcomponents as follows. When both products are in stock, the primary arrival rate (demand rate) for Product A can be represented by λ_A, for Product B as λ_B, and for the no-purchase option as $\Lambda - \lambda_A - \lambda_B$. Now assume that Product A is out of stock. During this period of time, the arrival rate for Product B increases to $\lambda_{\overline{A},B}$ and the arrival rate for the no-purchase option increases to $\Lambda - \lambda_{\overline{A},B}$. A similar observation holds for the time periods where only Product B is out of stock.

We now have all the notation we need for estimating the primary demand rates. The primary demand rate for Product A is estimated from $\lambda_A = D_A/T_{A,B}$. The primary demand rate for Product B is estimated from $\lambda_B = D_B/T_{A,B}$. Finally, the primary plus spillover demand rates for Products A and B are estimated from $\lambda_{A,\overline{B}} = D_{A,\overline{B}}/T_{A,\overline{B}}$ and $\lambda_{\overline{A},B} = D_{\overline{A},B}/T_{\overline{A},B}$.

Anupindi et al. (1998) show that the overall customer arrival rate can also be estimated in this setting, as long as the arrivals follow a Poisson distribution. To see how, we first need to calculate the probabilities that a given customer will choose a given option. Let the probabilities that an arriving customer will choose Product A when both products are in stock be p_A, that they choose Product B be p_B, that they choose Product A when Product B is not in stock be $p_{A,\overline{B}}$, and that they choose Product B when Product A is not in stock be $p_{\overline{A},B}$. These probabilities can be estimated as follows: $p_A = \lambda_A/\Lambda$, $p_B = \lambda_B/\Lambda$, $p_{A,\overline{B}} = \lambda_{A,\overline{B}}/\Lambda$, and $p_{\overline{A},B} = \lambda_{\overline{A},B}/\Lambda$. The overall arrival rate can now be estimated from

$$\Lambda = \frac{D_A + D_B + D_{A,\overline{B}} + D_{\overline{A},B}}{T_{AB}(p_A + p_B) + T_{A,\overline{B}}p_{A,\overline{B}} + T_{\overline{A},B}p_{\overline{A},B}}. \tag{1.6}$$

Anupindi et al. (1998)'s estimation of the overall customer arrival rate was quite groundbreaking, as it was one of the first attempts to estimate both the sales that did occur as well as the sales that *could have* occurred, but were not observed due to a lack of product available to meet the demand. This is important because it allows us to *quantify how much demand we actually lose* when an item is out of stock. To see how, consider the following example for Susan's newsstand. Let us assume that the overall arrival rate, Λ, was estimated to be 30 customers per hour. When both the newspapers and the magazines are in stock, the arrival rate of customers whose first choice is a newspaper is estimated to be $\lambda_A = 10$ customers per hour. Now suppose that the newsstand stocks out of the newspaper 1 h before the stand closes for the day but the magazine does not stock-out. What did this stock-out cost Susan in terms of lost profits?

To answer this question, we need to estimate how many of those 10 newspaper customers that arrived during the last hour of the newsstand's opening hours and found that the newspapers were out of stock decided to purchase a magazine instead. Let us assume that the profit margin for the magazine is also \$0.75 per unit and that the primary demand rate for magazine customers is estimated to be $\lambda_B = 5$ customers per hour. The arrival rate of customers that are willing to purchase a magazine when the newspaper is out of stock is estimated to be $\lambda_{\overline{A},B} = 8$ customers per hour. With these estimates, Susan can estimate the number of lost sales during that hour that the newspaper was out of stock to be $\lambda_A + \lambda_B - \lambda_{\overline{A},B} = 10+5-8 = 7$ newspapers. Multiplying the unit number of lost sales times the profit margin of the newspaper, Susan can estimate that she lost $7 \times \$0.75 = \$5.25$ from the stock-out of newspapers on that day. A similar analysis can be done for cases where only the magazine is out of stock, or even when both the newspaper and the magazine are out of stock.

Of course, Anupindi et al. (1998)'s methodology only works under a restrictive set of assumptions, the combination of which are unlikely to hold in most situations. The most important of these assumptions is the requirement for a homogeneous Poisson arrival rate and a portfolio of only two products. As we will see, estimating an arrival rate of no-purchase customers remains one of the primary challenges of more recent and sophisticated estimation methodologies.

1.4 Estimating Demand for a Portfolio of *K* Products with Out-of-Stocks and Substitution Effects

In their appendix, Anupindi et al. (1998) extend their model to the case of more than two products, with the restriction that each product has only one potential substitute product, out of the portfolio of products that are offered. This restriction is fine for the case where Susan only sells two products, a newspaper and a magazine, at her newsstand. Suppose, however, that Susan expanded her portfolio to add a third product, say a different daily magazine that is focused on only covering business news. Without a substantial analysis, it is unclear that this new magazine will serve as a substitute for only the newspaper or only the original magazine. Most likely, it will serve as a partial substitute for both. The methodology provided in Anupindi et al. (1998) no longer suffices for this expanded product portfolio. The inclusion of multiple substitution products requires a different estimation methodology, which we describe next.

1.4.1 Using Discrete Choice Models for Estimating the Market Shares for More than Two Products (Most General Model)

Discrete choice models were developed to model how customers choose among a set of multiple product options. As opposed to the more aggregate measures of demand, such as 60 newspapers and 40 magazines on some given day, discrete choice models capture the market share (percentage) of demand that each product in a portfolio captures during some time period. Thus, the same demand example represented by a discrete choice model with only newspapers and magazines as options would be that any given arriving customer would have a 60% probability of purchasing a newspaper and a 40% probability of purchasing a magazine. To equate these probabilities into actual quantities, we need to estimate a customer arrival rate over that same period of time, similar to the Anupindi et al. (1998) methodology. If the number of customer arrivals during that same time period was 100 and every customer who arrived had made a purchase, then the resulting estimated demand quantities for each product would match the aggregate demand of 60 newspapers and 40 magazines.

What if the number of customer arrivals was estimated to be 200, only 100 of which ended up making a purchase of either product. That is, not all arriving customers end up making a purchase. To capture this scenario, it helps to add a third option to our customer choice set—the "no-purchase" option. Thus, the customer's choice set now includes all the products in the portfolio of products for sale, along with an option of not purchasing anything. With an arrival rate of 200 customers but only 100 total purchases, our new market share estimates are 30% market share for newspapers, 20% for magazines, and 50% for the no-purchase option. With our new, more complete, choice set, we next provide a more formal explanation of how a discrete choice model attempts to capture the customer buying decision.

Discrete choice models are based on the concept of utility that represents the "value" an individual places on different products and different product attributes, thus capturing how individuals make trade-offs among a portfolio of products. Individual customers are assumed to select the alternative that provides them with the maximum utility. Alternative j is chosen by customer i if the utility that customer obtains from alternative j at time t, $v_{i,j,t}$, is greater than the utility for all other alternatives available at time t in the choice set S_t. The utility $U_{i,j,t}$ has an observed component, $v_{i,j,t}$, and an unobserved component, $\epsilon_{i,j,t}$, commonly referred to as an "error term." The observed component, often called the systematic or representative component of utility, is typically assumed to be a linear-in-parameters function of attributes that vary across individuals and alternatives. Formally, the observed component of utility for customer i of product j at time t is

$$v_{i,j,t} = \alpha_{i,j,t} + \boldsymbol{\beta}\mathbf{x}_{i,j,t}, \tag{1.7}$$

where $\mathbf{x}_{i,j,t}$ represents the vector of k attributes of product j exposed to customer i at time t and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ is an unknown vector of coefficients corresponding to the set of k product attributes. The first term on the right hand side, $\alpha_{i,j,t}$, is an alternative-specific constant (ASC). From a modeling perspective, including ASCs in a discrete choice model is similar to including an intercept term in a regression model. From an interpretation perspective, ASCs capture the average effect of all unobserved factors left out of the model. Due to underlying identification rules, given a total of k alternatives in the universal choice set, at most $k - 1$ parameters can be estimated.

Different discrete choice models are derived via assumptions on the error terms. One of the most common discrete choice models is the multinomial logit (MNL) model. The MNL model is derived by assuming $\epsilon_{i,j,t}$'s are i.i.d. Gumbel random variables. The attributes of the no-purchase option is assumed to not vary across customers or over time. Thus, $v_{i,j,t} = v_0$ for this choice. Under these assumptions, the choice probability that customer i chooses alternative j at time t and the no-purchase probability are given, respectively, as

$$P_{i,j,t} = \frac{e^{v_{i,j,t}}}{\sum_{j \in S_t} e^{v_{i,j,t}} + e^{v_0}} \quad \text{and} \quad P_0 = \frac{e^{v_0}}{\sum_{j \in S_t} e^{v_{i,j,t}} + e^{v_0}}. \tag{1.8}$$

To estimate (1.8) based on observed sales and customer transaction data, we can use maximum likelihood estimation. We shall assume we have data describing $i = N$ customer choice decisions. Define Customer i's purchase decision by the values

$$\delta_{i,j,t} = \begin{cases} 1, & \text{if customer } i \text{ purchases product } j \text{ at time } t, \\ 0, & \text{otherwise} \end{cases} \tag{1.9}$$

where

$$\sum_{j=1}^{J} \delta_{i,j,t} \leq 1.$$

Then, the customer utility model can be expressed as

$$v_{i,j,t} = \alpha_j + \boldsymbol{\beta}\mathbf{x}_{i,j,t}. \tag{1.10}$$

The model (1.10) indicates that each customer may have a different utility for a given product and that this utility may vary across time periods as the product attributes vary over time. This most general model is appropriate for settings such as with online retailers, where observed attributes such as the product price may vary across different customers and over time. Since the retailer may change the price of a product over time, the relative utility that this product provides to a customer will depend on how much higher or lower the price is compared to the prices of the other products in the available assortment as well as any outside option.

For utility model (1.10), the data consists of the values of $\delta_{i,j,t}$: $i = 1, \ldots, N$, $t = 1, \ldots, T$, $j = 1, \ldots, K$ together with the choice sets $S_{i,t}$: $i = 1, \ldots, N$, $t = 1, \ldots, T$. The choice sets can be determined, for example, by tracking the on-hand inventory status prior to each purchase.

1.4.2 Special Case Where We Know the Customer Arrival Rate

We start with a special case where we know the true customer arrival rate, Λ. Assume a discrete choice model for which an observation represents a customer, a vector of attributes associated with the customer and the products, and the chosen product. The problem of interest is to solve for the parameters $\boldsymbol{\beta}^*$ given a random sample of observations. Maximum likelihood estimation solves for the values of $\boldsymbol{\beta}$ that maximize the likelihood function

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{j \in S_{i,t}} \left(P_{i,j,t}\right)^{\delta_{i,j,t}}. \tag{1.11}$$

Since we know the true arrivals of each customer, we can track the buy or no buy decision of each customer $i = 1, .., N$. Upon arrival, each customer observes a choice set $S_{i,t}$, where $j \in S_{i,t}$ are alternative products in the choice set for customer i in time period t, $P_{i,j,t}$ is the probability that individual i selects alternative j in time period t given a sample of attributes $\mathbf{x}_{i,j,t}$ and parameter estimates $\boldsymbol{\beta}$, and $\mathbf{x}_{i,j,t}$ is the vector of attributes associated with alternative j and individual i in time period t.

Computationally, it is easier to maximize the logarithm of the likelihood function, i.e., the log-likelihood (LL) function

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j \in S_{i,t}} \delta_{i,j,t} \ln P_{i,j,t}. \tag{1.12}$$

The $\boldsymbol{\beta}$ parameter estimates are obtained by using optimization algorithms that maximize the log-likelihood function. In the case of the multinomial logit models, the log-likelihood function is globally concave. This can be verified by examining its first and second derivatives with respect to $\boldsymbol{\beta}$. The log-likelihood can be further expanded as follows:

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j \in S_i,t} \delta_{i,j,t} \ln\left(\frac{e^{v_{i,j,t}}}{\sum_{j \in S_{i,t}} e^{v_{i,j,t}}}\right) \tag{1.13}$$

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j \in S_{i,t}} \delta_{i,j,t}\left(\boldsymbol{\beta}^T \mathbf{x}_{i,j,t} - \ln \sum_{j \in S_{i,t}} e^{\alpha_j + \beta x_{i,j,t}}\right). \tag{1.14}$$

Let d_j represent the expected demand for product j, $j = 1, \ldots, K$. The expected demand for product j can be estimated by simply multiplying the aggregate customer arrival rate times the estimated probability that a customer will choose product j should product j be available in the choice set at the time of their arrival

$$d_j = \sum_{i=1}^{N}\sum_{t=1}^{T} \Lambda P_{i,j,t}. \tag{1.15}$$

1.4.3 Estimating the Choice Probabilities and the Customer Arrival Rate

For the remainder of this section, we do not assume that we know the overall customer arrival rate, Λ. Thus, the estimation problem becomes much more complex, as we now have to estimate the coefficients of the choice model ($\boldsymbol{\beta}$) and the customer arrival rate (Λ) simultaneously. This is the equivalent constrained demand problem for multiple substitute products.

Despite the prevalence of the multi-product unconstraining problem in retail settings, there has been comparatively little work that specifically focuses on a retail setting. Instead, much of the work on this problem has been performed on the single-leg airline revenue management problem, where an airline will offer multiple variations of its core product (an airline seat from a particular origin to a particular destination at a particular time and date). While the physical product is the same (a coach-class seat on an airplane), airlines typically create different versions of this product by imposing restrictions (i.e., refundable tickets vs. non-refundable tickets). An airline will offer these different product variations up until the time of the flight's departure, with sales occurring at various time points and the airline making some variations of the product available or not available at different time points in the booking curve.

Talluri and van Ryzin (2004), hereafter referred to as TvR, were the first to propose modeling a single-leg choice-based revenue management problem by directly integrating customer choice behavior into the objective function. Specifically, the objective is to find subsets of the K total products to offer, S_t, at each point t in time, for each remaining capacity to maximize revenue. The booking curve (timeline of purchases up to the day of departure) is broken into T discrete time periods, starting at period 0 and ending at period T. The arrival rate is assumed to be constant over the booking curve and is represented as $\lambda^\star(t) = \Lambda/T$. Dynamic programming (using backward induction) is used to solve the objective function, given as

$$V_t(x) = \max_{S_t \in K} \Bigg\{ \sum_{j \in S_t} \{\lambda^\star(t) P_j(S_t)[p_j + V_{t-1}(x-1)] \\ + [\lambda^\star(t) P_0(S_t) + (1 - \lambda^\star(t))] V_{t-1}(x) \Bigg\}. \tag{1.16}$$

In the TvR formulation, the utility associated with the no-purchase alternative is assumed to be $\nu_0 = 0$, and the arrival rate is interpreted as the probability that a customer arrives to the system in a given time window. Importantly, the population of arriving customers includes those who arrive and purchase a product from the portfolio offered to them, as well as those who arrive and decide not to purchase any product from the available portfolio. As in our previous examples, the sales data used to estimate their model typically only includes the observations of customers who arrive and purchase a product. Thus, it is impossible to distinguish between time periods where a customer arrived and did not purchase a product (i.e., purchased from a competitor *or* did not purchase any product) and periods in which no customer arrived. To estimate this potential market size, or the total number of "potential" customers, TvR use a variation of the Expectation-Maximization (EM) method described in the previous section.

Note that TvR estimate a special case of (1.10) where the product attributes (such as price) change over time, but all customers arriving in the same time period observe the same choice set and the same product attributes. This setting is appropriate for most revenue management applications because contractual obligations between the airline or hotel with their third-party distribution networks (such as Expedia or Orbitz) require that product pricing be consistent across platforms. It may be less appropriate in online retailing applications, where certain customer segments receive personalized pricing offers at the same time that other customer segments observe different prices. We conclude with the most restrictive case, where the product attributes do not vary across customers or over time. While being the most restrictive, this model captures a wide range of practical scenarios, one example of which is the newsstand example from the beginning of this chapter.

Table 1.1 summarizes the three model versions we discuss in this section. While describing a method for estimating the most general model (1.10) is beyond the scope of this chapter, the reader is referred to Im et al. (2018) for a discussion of how to estimate this model. Instead, we will focus on the two special cases where either the product attributes only vary over time or they do not vary at all.

Table 1.1 A summary of tested parameter values

Product attributes	Probability function
Vary by time and customer	$P_{ij}(S_{it}) = \frac{\exp(\nu_{ijt})}{\sum_{l \in S_{it}} \exp(\nu_{ilt}) + \exp(\nu_{i0})}$, $j \in S_{it}$, where $\nu_{ijt} = \alpha_j + \boldsymbol{\beta}\mathbf{x}_{ijt}$
Vary only by time	$P_j(S_t) = \frac{\exp(\nu_{jt})}{\sum_{l \in S_t} \exp(\nu_{lt}) + \exp(\nu_0)}$, $j \in S_t$, where $\nu_{jt} = \alpha_j + \boldsymbol{\beta}\mathbf{x}_{jt}$
Are constant	$P_j(S) = \frac{\exp(\nu_j)}{\sum_{l \in S_t} \exp(\nu_l) + \exp(\nu_0)}$, $j \in S$, where $\nu_j = \alpha_j + \boldsymbol{\beta}\mathbf{x}_j$

All combinations of these values constitute 13,689 numerical experiments

1.4.3.1 Special Case Where Product Attributes Change Over Time but not Across Customers in the Same Time Period

In the case where the attributes of a given product are the same for all customers within each time period. Let S_t denote the set of alternatives available to all arriving customers in the time window t. Define a given customer's purchase decision by the values:

$$\delta_{j,t} = \begin{cases} 1, & \text{if an arriving customer purchases product } j \text{ at time } t, \\ 0, & \text{otherwise.} \end{cases} \tag{1.17}$$

Then, the customer utility model simplifies to

$$\nu_{j,t} = \alpha_j + \boldsymbol{\beta}\mathbf{x}_{j,t} \tag{1.18}$$

The reduced model (1.18) indicates that each customer has the same utility for a given product but the utility can still vary across time periods as the product attributes vary over time. This model is appropriate for settings such as a traditional brick-and-mortar retailer, where observed attributes such as the product price are the same for all the customers that are in the store during the same time window. The retailer may change the price of a product over time, however, so the relative utility that this product provides to a customer will depend on how much higher or lower the price is compared to the prices of the other products in the available assortment as well as any outside option.

The choice probability for alternative j, and the probability for the no-purchase alternative, for a customer arriving at time t are given, respectively, as

$$P_{j,t}(\boldsymbol{\beta}, S_t, \mathbf{X}_t) = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{j,t}\right)}{\sum_{j \in S_t} \exp\left(\boldsymbol{\beta}'\mathbf{x}_{j,t}\right) + \exp\left(V_{0,t}\right)}, \quad j \in S_t, \tag{1.19}$$

and

$$P_{j=0,t}(\boldsymbol{\beta}, S_t, \mathbf{X}_t) = \frac{\exp\left(V_{0,t}\right)}{\sum_{j \in S_t} \exp\left(\boldsymbol{\beta}'\mathbf{x}_{j,t}\right) + \exp\left(V_{0,t}\right)}. \tag{1.20}$$

These probabilities are a function of the available choice set S_t available at time t, the matrix of attributes $\mathbf{X}_t$ of the available alternatives in that choice set, attributes of the choice itself (e.g., the day of the week when the purchase occurs), and a vector of parameters $\boldsymbol{\beta}$ that needs to be estimated. For convenience, we denote the complement of (1.20) as P_t (i.e., $P_t(\boldsymbol{\beta}, S_t, \mathbf{X}_t) = 1 - P_{0,t}(\boldsymbol{\beta}, S_t, \mathbf{X}_t)$), which represents the probability of purchasing some product from the firm. We also define a set of choice probabilities for the alternatives conditioned on the fact that some purchase is made,

$$P_{j,t|\delta=1}\left(\boldsymbol{\beta} \mid t, S_t, \mathbf{X}_t\right) = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{j,t}\right)}{\sum_{j \in S_t} \exp\left(\boldsymbol{\beta}'\mathbf{x}_{j,t}\right)}. \tag{1.21}$$

Note that (1.21) specifically excludes the no-purchase alternative.

It is well known that utility maximization models are only unique up to the scale of utility. Adding a constant to every utility value will result in an identical set of probabilities. Thus it is not possible to generate unique parameter estimates for ASCs for each alternative, and at least one alternative in the model needs to be set as the reference and given a fixed constant utility value to ensure these parameter estimates are unique. TvR choose the reference alternative to be the no-purchase alternative by setting $V_{j=0} = 0$, but, as it will become evident in the next section, that choice will become restrictive. Instead, we parameterize the utility of the no-purchase alternative, denoting it as γ, and assume that the modeler imposes other suitable restrictions (see Ferguson et al. 2012) within the $\boldsymbol{\beta}$ vector to ensure the parameter estimates are unique.

For notational clarity, the probabilities indicated by (1.19) are denoted as $P_{j,t}(\boldsymbol{\beta}, \gamma)$, the choice probability of the no-purchase alternative and its complement are denoted as $P_{j,t}(\boldsymbol{\beta}, \gamma)$ for $j \in \{1, \ldots, K\}$ and $P_{j=0,t}(\boldsymbol{\beta}, \gamma)$, respectively, and the choice probabilities conditional on some purchase given in (1.21) are denoted as $P_{j,t|\delta=1}(\boldsymbol{\beta})$. Importantly, $P_{j,t|\delta=1}$ is a function of $\boldsymbol{\beta}$ but not of γ. In each case, the conditionality upon the observed data and available choice set is implied.

Direct Estimation Approach

A direct estimation of the likelihood-maximizing parameters of the arrival and discrete choice models is complicated by missing data. Formally, when there are K possible products for sale, the customer arrival and purchase processes can result in one of $K+2$ mutually exclusive and collectively exhaustive outcomes: K discrete outcomes associated with a customer arriving and choosing to purchase one of the K products offered by the firm, one discrete outcome associated with a customer arriving who chooses not to purchase any product from the firm, and one discrete outcome associated with no customer arriving.

An uncensored data observation would uniquely observe which of these $K+2$ outcomes occurred in each discrete time slice. A censored data observation differs in that the no-purchase and no-arrival outcomes are indistinguishable, jointly representing the same observable outcome. The complexity of the log-likelihood function for censored data arises because the censoring requires the summation of the no-purchase and no-arrival probabilities in calculating the likelihood, before taking the logarithm.

At an abstract level, the log-likelihood function is given as

$$LL(\boldsymbol{\beta}, \gamma, \lambda) = \sum_{t \in T} \log\left(\Pr_t\left(\delta_{j,t} \mid \boldsymbol{\beta}, \gamma, \lambda\right)\right), \tag{1.22}$$

with T as the set of all discrete time slices in the estimation data, $\delta_{j,t}$ as the observed outcome in time slice t (for example, the purchase of a Y-class ticket in time slice t), and $\Pr_t(j \mid \boldsymbol{\beta}, \gamma, \lambda)$ as the probability of outcome j occurring in time slice t, given the model and parameters $\boldsymbol{\beta}$, γ, and λ. For the time slices in which a sale is observed (denote this set of time slices as $\mathscr{P}$), the value $\Pr_t(\delta_t = 1 \mid \boldsymbol{\beta}, \gamma, \lambda)$ is given by the joint probability that a customer arrives and that the arriving customer chooses to purchase at least one of the available product options $j \in S_t$ from the firm, giving

$$\Pr_t(\delta_t = 1 \mid \boldsymbol{\beta}, \gamma, \lambda) = \lambda P_{t,\delta_t=1}(\boldsymbol{\beta}, \gamma), \quad \forall t \in \mathscr{P}.$$

For all other time slices (denote this set of time slices as $\bar{\mathscr{P}}$), the outcome is that a sale is not observed. The probability of that outcome is the sum of the probabilities of the two possible reasons: no purchase and no arrival, giving

$$\Pr_t(\delta_t = 0 \mid \boldsymbol{\beta}, \gamma, \lambda) = \underbrace{\lambda P_{t,\delta_t=0}(\boldsymbol{\beta}, \gamma)}_{\text{no purchase}} + \underbrace{(1-\lambda)}_{\text{no arrival}}, \quad \forall t \in \bar{\mathscr{P}}.$$

Thus, the resulting log-likelihood function for the model with censored data can be written as

$$LL(\boldsymbol{\beta}, \gamma, \lambda) = \sum_{t \in \mathscr{P}} \left[\log\left(\lambda P_{t,\delta_t=1}(\boldsymbol{\beta}, \gamma)\right)\right] + \sum_{t \in \bar{\mathscr{P}}} \left[\log\left(\lambda P_{t,\delta_t=0}(\boldsymbol{\beta}, \gamma) + (1-\lambda)\right)\right]. \tag{1.23}$$

Maximum likelihood estimators for $\boldsymbol{\beta}$, γ, and λ can be found by maximizing this log-likelihood function. However, this function is not generally concave and TvR suggest it may be hard to maximize. Nevertheless, the size and scope of some RM problems may allow for the direct maximization of this log-likelihood function. If a particular problem is found to be computationally tractable using regular maximum likelihood estimators, then that approach should be preferentially employed. If, however, directly maximizing (1.23) is found to be difficult, slow, or if there is concern about converging to local optima, then other approaches can be considered.

Application of the Expectation-Maximization Method for Estimating the Arrival Rate

Instead of directly maximizing (1.23), TvR use the Expectation-Maximization (EM) method. In this application, the EM method is used to estimate the market size (i.e., the total number of “potential” customers) without having actual observations of no-purchase customers. The EM method avoids calculating the actual censored data log-likelihood given in (1.23), and instead focuses on calculating the expected value of that log-likelihood, which incorporates observed data as well as the expected value of unobserved data. In their Eq. (1.15), TvR write an expected log-likelihood function for the sales model (assuming $\gamma = 0$) as

$$\begin{aligned} E\left[LL\left(\boldsymbol{\beta}, \lambda\right)\right] = &\sum_{t \in \mathscr{P}} \left[\log(\lambda) + \log\left(P_{t,\delta_t=1}(\boldsymbol{\beta}, \gamma = 0)\right)\right] \\ &+ \sum_{t \in \bar{\mathscr{P}}} \Big[\hat{a}(t)\left(\log(\lambda) + \log\left(P_{t,\delta_t=0}(\boldsymbol{\beta}, \gamma = 0)\right)\right) \\ &+ \left(1 - \hat{a}(t)\right)\log\left(1 - \lambda\right)\Big], \end{aligned} \tag{1.24}$$

with $\hat{a}(t)$ as the expected value of $a(t)$, an indicator variable that takes on a value of 1 if an arrival occurred at time t and 0 otherwise. Where $t \in \bar{\mathscr{P}}$, the value of $a(t)$ cannot be observed, as either a customer arrives but chooses to purchase nothing or no customer arrives. These two outcomes are indistinguishable in the data. The EM algorithm iterates between updating the expected log-likelihood function given a distribution for the missing data (essentially, replacing $a(t)$ in (1.24) with its expected value) and maximizing that function, a process which is widely accepted as effective but slow. As previously discussed, when the EM algorithm converges, it converges to a local stable point of the likelihood function. Often, this represents a local maximum for the underlying censored data log-likelihood, but it could also be a saddle point or a local minimum of that function. Maximizing the function given in (1.24) is appealing because $\boldsymbol{\beta}$ and λ are fully separable, and both components are easy to maximize globally concave sub-problems. However, this approach requires multiple iterations between the expectation and maximization steps.

The application of the EM method to a choice model that includes product attributes has some significant limitations. First, an unlucky choice for the time slices can lead to inconsistent parameter estimates (Talluri 2009). TvR recommend that the time slices be chosen such that the probability of two arrivals occurring in the same period is small, although they do not provide any guidance on what exactly constitutes small. Newman et al. (2014) show that time periods as small as 15 s may be required for the EM method to achieve reasonable accurate estimates. Of course, smaller time windows result in longer computation times needed to estimate a model, since the EM iterates between the expectation and maximizing steps. If the initial values for the parameter estimate are far from the point of convergence, the estimation times take even longer Newman et al. (2014), resulting in unacceptably long times for a firm that must estimate thousands of different models. Third, the

EM method is not guaranteed to converge to a global maximum, leading to an identification issue (Dempster et al. 1977). Fourth, the exact product portfolio needs to be recorded during the time period for which no purchase occurred. Thus, the firm must make a conscious effort to collect this information, since it is typically not included in the regular sales transaction data.

To rule out the first two limitations described above, Newman et al. (2014) propose an alternative estimation approach, which they term a two-step (TS) approach. Compared to TvR, the time slices for the arrivals are converted from discrete time to continuous time, by modeling the arrival process as a Poisson distribution. In the first step of their methodology, the choice model parameters (excluding the no-purchase option) are estimated by maximizing a partial likelihood function. The no-purchase utility parameter and the arrival rate are estimated in the second step, dramatically speeding up the computation times. One drawback of their method, however, is that the second step sometimes becomes too unstable to provide consistent estimates. A second limitation, which is shared by the TvR formulation, is that customers are assumed to arrive at a constant rate throughout the booking curve.

Subramanian and Harsha (2017) present a loss minimization (LM) method which minimizes an objective function of prediction errors based on the assumed demand and market size models. To obtain the parameter estimates, the authors use an optimization algorithm instead of a maximum likelihood estimation. Although their approach allows for non-constant arrival rates, the known covariates for the market size model are necessary conditions for estimating the arrival rate. The LM method may not work well when the number of arrivals is relatively small within each time window.

The joint estimation of a choice model and an arrival rate with constrained demand where the product attributes change over time remains a challenging problem. While the papers described above have made progress in solving this problem, there does not appear to be an established methodology that has been universally adopted. Instead, some researchers have shifted their focus to a further relaxation of this problem. In the next subsection, we discuss some work on the variation of this problem when the product attributes do not change over time.

1.4.3.2 Special Case Where the Product Attributes Do Not Change Across Customers or Over Time

While the more general approach described above can be used to estimate the true demand for newspapers and magazines at Susan's newsstand, we can take advantage of the fact that Susan never changes the price on either of these two products. The fact that the product attributes of the newspaper and the magazine do not change over time allows for a much simpler formulation, that we call a *ranked-preference approach*. In a ranked-preference approach, each customer is assumed to have a ranked ordering of products, starting with their most desirable (highest utility) product down to their least desirable product. In the simple case of Susan's newsstand, there are three product options: a newspaper (N), magazine (M), and no purchase (NP). A customer with a ranked-preference list of (N, M, NP) means that that customer will purchase a newspaper if one is available, a magazine if one is available but a newspaper is not available, and will only choose not to purchase anything if both newspapers and magazines are not available. Thus, there are 3! possible ranked-preference combinations: (N, M, NP), (N, NP, M), (M, N, NP), (M, NP, N), (NP, N, M), and (NP, M, N). If a customer with a ranked-preference list of (N, M, NP) purchases a newspaper, that demand is considered to be "primary" demand for newspapers, since it was the customer's first choice. If the same customer arrived and found that all the newspapers were sold out, the customer would purchase a magazine, which would be considered "spillover" demand for the magazine and unfulfilled "primary" demand for the newspaper.

In this scenario, both the customer and the product attributes are constant across each time period so the customer utility model can be simplified to $v_j = \alpha_j + \beta x_j$. When the product attributes are also assumed constant, the utility model further simplifies to a constants only model

$$v_j = \alpha_j. \tag{1.25}$$

Vulcano et al. (2012) suggest for the constants only model an unconstraining method that employs the EM method and an outside estimate of the market share of the no-purchase option. The market share information is used for model parameter identification, and a non-homogeneous Poisson arrival process is assumed for the arrival rate. An outline of their approach is as follows. Let $P_j(S, \mathbf{v})$ denote the probability that a customer chooses product $j \in S$ when S is offered and the preference weights are given by vector $\mathbf{v}$. Then,

$$P_j(S, \mathbf{v}) = \frac{v_j}{\sum_{j \in S} v_j + 1}. \tag{1.26}$$

The no-purchase option is assumed to have a preference weight equal to 1 (similar to the assumption of zero utility in the TvR formation) such that the no-purchase probability is

$$P_0(S, \mathbf{v}) = \frac{1}{\sum_{j \in S} v_j + 1}. \tag{1.27}$$

One major difference between the method proposed in Vulcano et al. (2012) and the methods discussed in the previous subsection is that the time periods in Vulcano et al. (2012) can be much larger and contain aggregated demand, whereas the time periods in TvR (and subsequent other methods where the product attributes can change over time) are required to be small enough, so that the likelihood of two arrivals in the same period is very small. Indeed, the time periods in Vulcano et al. (2012) need only be small enough that the same subset of products be made available and the arrival rate remains constant. Thus, a time subscript is added to both the arrival rate and the choice set notation: λ_t and S_t.

The statistical challenge in Vulcano et al. (2012) is to estimate the parameters of their model. Both the vector of preference weights, $\mathbf{v}$, and the vector of arrival rates $\boldsymbol{\lambda}$ must be estimated using only sales transaction data. Let $z_{j,t}$ denote the primary demand for product j in time period t. Note that we do not actually observe the primary demand for each product (it may not have been available in that period, or the sales for product j may have included some spillover demand from other products that were not available). Next, define z_j as the total primary demand for product j, where $z_j = \sum_{t=1}^{T} z_{j,t}$ and z_0 is the primary demand for the no-purchase option. If the primary demands were known, the log-likelihood function would simply be

$$LL(\mathbf{v}) = \sum_{j=1}^{K} z_j \log \left(\frac{v_j}{\sum_{l=1}^{K} v_l + 1} \right) + z_0 \log \left(\frac{1}{\sum_{l=1}^{K} v_l + 1} \right). \tag{1.28}$$

Of course, the primary demands are not known and must be estimated. Vulcano et al. (2012) do so by utilizing an estimate of the total market share for the firm (the market share the firm would have if all its products were available in every time period). Combining this total market share with the observed number of total purchases provides an estimate for the total customer arrival rate in each time period. They then employ a variation of the EM method, starting with initial estimates of the preference weights, $\mathbf{v}$, to compute estimates for the primary demand for each product (the E-step). They then use these estimates to maximize the log-likelihood function (1.28) (the M-step) until convergence is achieved. For details, see Sect. 3.5.2 in Vulcano et al. (2012).

1.5 Conclusions

As teachers, researchers, and practitioners of operations management and industrial engineering, our community has historically concentrated on the more tractable aspects of problems involving the setting of capacity, staffing, or inventory levels to efficiently meet an uncertain future demand. In this chapter, I have presented an argument that more focus should be placed in both our teaching and research on the less-tractable, but perhaps even more important, problem of estimating this future demand. While estimation methods have primarily focused on time-series forecasting and other predictive analytics techniques, there is a common underlying assumption in these methods that our historical data represents the true demand for the product or service that we are trying to estimate. In most applications, however, this is simply not the case. The recording of sales data (rather than demand data) and the reality of product substitution behavior results in truncated data available for estimation purposes. Ignoring this fact when estimating demand distributions for our optimization models can result in significant cost or profit implications—the best optimization models cannot compensate for poor quality inputs.

All is not lost, however, as there has been some work devoted to improving the estimates of demand distributions by first "unconstraining" the underlying sales data. The earliest work in this area assumed that the demand for a product was independent of all other products sold through the same channel or store. Some of these methodologies employed statistical concepts such as the Expectation-Maximization algorithm. Subsequent work extended these concepts to simple variations of product substitution scenarios, where two products may partially substitute for each other when the other is out of stock. In such cases, it is assumed that the attributes of the two products (such as product prices) do not change over time. Thus, some early applications of these methodologies were for sales channels where these assumptions held, such as estimating the demand for products sold through vending machines.

As our community became more aware of the reality of product substitution effects, we started incorporating this behavior into our optimization problems. One early summary of this work by Mahajan and van Ryzin (1999) provides a review of how the newsvendor model can be adjusted to allow for product substitution effects using these new demand estimates. As one may expect, including substitution effects results in optimal inventory stocking solutions where the firm should stock relatively more of popular products and less of unpopular products than a traditional newsvendor analysis indicates, due to excess substitution demand combined with reduced underage costs from having substitute variants as backups. As with the earlier work on inventory management, however, the development of more sophisticated optimization models incorporating various product substitution effects outpaced the development of new estimation techniques required to provide the inputs for these optimization models.

The extension of inventory models with product substitution effects to scenarios with more than two substitute products in the portfolio often required the use of consumer-choice models, such as the MNL model. Choice models do not estimate the total demand distribution directly, but rather estimate the probability that an arriving customer will choose any particular product out of the assortment of products made available to him or her. Thus, choice models must be combined with estimates of an arrival rate in order to provide an estimate of total demand for a product. This creates a challenge as both the choice probabilities and the customer arrival rates must be estimated from the same truncated sales transaction data, resulting in a rather complex multi-parameter estimation problem. While some progress has been made in this area, these methodologies are not yet equipped to solve the increasing common problem of estimating product demand for online retailers who regularly change product assortments, product prices, and other product features based on such triggers as timing, available stocks, and individual customer characteristics. While this chapter presents a summary of the work to date, much more research is needed before anyone can claim to have provided an efficient solution to these new retail settings.

References

Anupindi R, Dada M, Sachin G (1998) Estimation of consumer demand with stock-out based substitution: an application to vending machine products. Market Sci 17(4):406–423

Boyles RA (1983) On the convergence of the EM algorithm. J R Stat Soc Ser B Methodol 45(1):47–50

Conlon C, Mortimer J (2013) Demand estimation under incomplete product availability. Am Econ J Macroecon 5(4):1–30

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Methodol 39(1):1–38

Ferguson ME, Garrow LA, Newman JP (2012) Application of discrete choice models to choice-based revenue management problems: a cautionary note. J Revenue Pricing Manag 11(5):536–547

Guo P, Xiao B, Li J (2012) Unconstraining methods in revenue management systems: research overview and prospects. Adv Oper Res 2012:1–23. https://doi.org/10.1155/2012/270910

Im J, Cho S, Ferguson ME, Pekgun P (2018) Robust demand estimation with customer choice-based models for sales transaction data. Working paper, University of South Carolina

Mahajan S, van Ryzin GJ (1999) Retail inventories and consumer choice. In: Tayur S, Ganeshan R, Magazine M (eds) Quantitative models for supply chain management, chap 17. Springer, Boston, pp 491–551. https://doi.org/10.1007/978-1-4615-4949-9_17

Mersereau A (2015) Demand estimation from censored observations with inventory record inaccuracy. Manuf Serv Oper Manag 17(3):335–349

Newman JP, Ferguson ME, Garrow LA, Jacobs TL (2014) Estimation of choice-based models using sales data from a single firm. Manuf Serv Oper Manag 16(2):184–197

Redner RA, Walker HF (1984) Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev 26(2):195–239

Subramanian S, Harsha P (2017) Demand modeling in the presence of unobserved lost sales. Working paper, IBM TJ Watson Research Center, Yorktown Heights, NY

Swan AV (1969) Algorithm AS 16: maximum likelihood estimation from grouped and censored normal data. J R Stat Soc Ser C Appl Stat 18(1):110–114

Talluri K (2009) A finite-population revenue management model and a risk-ratio procedure for the joint estimation of population size and parameters. Technical report, Universitat Pompeu Fabra, Barcelona

Talluri K, van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. Manag Sci 50(1):15–33

Vulcano G, van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. Oper Res 60(2):313–334

Wolynetz MS (1979) Algorithm AS 138: maximum likelihood estimation from confined and censored normal data. J R Stat Soc Ser C Appl Stat 28(2):185–195

Wu CFJ (1983) On the convergence properties of the EM algorithm. Ann Stat 11(1):95–103

Chapter 2
Selling Innovative Products to Anxious Consumers

Yufei Huang, Bilal Gokpinar, Christopher S. Tang, and Onesun Steve Yoo

Abstract When deciding whether to adopt an innovative product, consumers often experience different levels of anxiety that prompt them to resist purchase. In some cases, consumers' anxiety is mitigated by "validation" through *externality* (e.g., the number of early adopters). To reduce consumers' anxiety, firms can also invest in "familiarization" through promotion (e.g., free trials). In this chapter, we conceptualize an innovative product as a product that engenders anxiety, and present a model that employs a consumer utility model focusing on the psychological dimension. We examine the firm's profit-maximizing promotion and pricing decisions when selling to forward-looking consumers in the presence of externality. Our equilibrium analysis reveals that, unlike the conventional wisdom for promoting new version of an existing product, for anxiety-inducing innovations with externality, accelerating the speed of adoption through promotion can actually be detrimental to the firm.

Keywords Innovative product introduction · Adoption anxiety · Promotion · Externality

2.1 Introduction

When selling an innovative product, a firm needs to examine the underlying factors that drive consumers' purchasing decisions. Hellofs and Jacobson (1999) find that consumers often evaluate both *objective functional* and the *subjective psychological*

Y. Huang
Trinity Business School, Trinity College, Dublin, Ireland
e-mail: y.huang@bath.ac.uk

B. Gokpinar · O. S. Yoo
UCL School of Management, University College London, London, UK
e-mail: b.gokpinar@ucl.ac.uk; o.yoo@ucl.ac.uk

C. S. Tang (✉)
UCLA Anderson School of Management, University of California, Los Angeles, CA, USA
e-mail: chris.tang@anderson.ucla.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_2

benefits and costs before they make purchasing decisions.[1] Unlike a new version of an existing product, we define an innovative product as a product that engenders anxiety—subjective psychological factors that play a prominent role in deciding whether to purchase an innovative product. (Castano et al. 2008) argue that an innovative product brings about different anxieties: (1) the disruption of established habits (Ram 1989; Ram and Sheth 1989); (2) the need to learn about the new technology; (3) the anxiety of letting go of the old technology (Fournier 1998; Hoeffler 2003; Castano et al. 2008); and (4) the anxiety over privacy and cyber security of internet-based products/services. While consumers with low anxiety level will likely become early adopters, those with higher anxiety level may delay their purchase until their anxiety level decreases (Ram and Sheth 1989). Consider the following examples:

1. *Satellite TV or TV-set top boxes such as Apple TV or ROKU.* These innovative products allow consumers to watch TV and stream various programs using a single source at a lower cost than other alternatives including cable TV.
2. *Contactless payment technology that was recently introduced by Transport for London.* This innovative service allows customers to tap their debit/credit cards at the turnstile gates and therefore eliminates the need to wait in line to purchase traditional tickets or travel cards (e.g., an Oyster card).
3. *Uber or Lyft.* These innovative mobile app services allow consumers to access Taxi service within minutes via a smart phone from arbitrary locations at a reduced cost.

Despite the seemingly superior objective functional benefits[2] (simpler user interfaces: simple Apple TV control versus complex TV remote control, convenience: no need to top up the value of Oyster card, and ease of use: users can hail a ride without the need to explain the pick-up and drop-off locations) that these new-generation products or services offer over existing alternatives, many consumers are still reluctant to adopt them due to various anxieties including: "Will it work with my other products?", "Is it secure enough?", and "Will I be able to use it easily?" Therefore, when launching an innovative product, a firm must consider ways to manage consumer anxiety in order to improve adoption speed and maximize profit.

[1]The subjective component of utility is different from models of bounded rationality, which considers consumers with cognitive limitations or psychological biases. We do not assume that consumers use simple heuristics to make complex decisions, or display certain intrinsic psychological tendencies. See Ren and Huang (2018) for a recent review of modeling bounded rationality in operations management.

[2]These objective functional benefits are essentially the additional willingness to pay over comparable traditional products. This additional valuation can be measured via lab experiments and it is well examined by consumer behavior researchers (Sheth et al. 1991; Dahl and Moreau 2002; Mukherjee and Hoyer 2001).

The diffusion literature suggests that new product adoption is driven by: (1) an internal anxiety factor that can be reduced through externality, and (2) an external anxiety factor that can be mitigated through familiarization and pricing decisions (Bass 1969; Gatignon and Robertson 1985; Chatterjee and Eliashberg 1990; Peres et al. 2010). First, through externality, consumers can reduce their anxiety levels by observing other adopters. For example, an anxious consumer who is initially reluctant to use Uber (or contactless payment or TV top boxes) may become more comfortable adopting it after observing many other users. Second, through familiarization, a firm can help consumers reduce their anxiety levels by offering free trials or training videos/seminars.

These observations motivated us to develop a model that captures both the *internal* factor (i.e., validation through externality) and the *external* factor (i.e., familiarization through promotion) so that we examine whether a firm should promote its innovative product and how to set its prices over two time periods (the launch period and the post-launch period), when consumers experience the above anxiety factors. Specifically, we consider a situation in which the new product's objective quality is known, but its adoption engenders anxiety. We employ a consumer utility model that focuses on subjective anxiety and present a stylized model that captures three salient features: (1) consumers are heterogeneous in their anxiety levels, (2) psychological anxiety can be reduced by firm's promotion strategy (due to familiarization effect) as well as the number of early adopters (due to externality effect), and (3) consumers are forward-looking (i.e., they may delay their purchase decisions to lower their anxiety or in anticipation of price drops).

The analysis of our two-period model enables us to obtain two main results. First, we find that decreasing the selling price over time (price markdown) is always optimal for the firm when externality is absent. However, in the presence of externality, increasing the selling price over time can be optimal for the firm. This is because, lowering the price in the launch period increases the number of early adopters, which reduces the remaining consumers' anxiety and raises their valuation of the product. Thus, firms can benefit from the markup pricing by capturing the late adopters at a premium price. This is consistent with penetration pricing strategy of networked goods (Dhebar and Oren 1985). Second, we find that, in the presence of externality, promoting the new product too hard during the launch period can lead to *lower* profit for the firm. This is because externality enables the firm to reap reward from charging late adopters a higher price during the post-launch period. Therefore, employing familiarization strategy through promotion may offset this benefit because it increases the speed of adoption and decreases the number of late adopters.

The rest of the chapter is organized as follows. Section 2.2 discusses the externality mechanism by reviewing the related literature. In Sect. 2.3, we present a stylized model of consumer adoption. In Sect. 2.4, we analyze the impact of the firm's familiarization strategy and externality on consumer adoption and the firm's optimal pricing and profit. We conclude in Sect. 2.5. All proofs are provided in Huang et al. (2018).

2.2 Positive Externality

Consumer anxiety can be classified into four groups (2×2) based on positive vs. negative externality, and whether it involves functional vs. psychological component of utility (Huang et al. 2018). Specifically, we address the case when consumers have subjective psychological anxiety about the new product, which can be reduced due to the presence of certain "positive externality." Positive externality exists not only for networked goods, e.g., telephone networks (Katz and Shapiro 1985), but it is also present for many non-networked products or services, when an increase in the number of adopters increases the consumer's subjective psychological utility (due to reduced anxiety level) by appealing to consumers' penchant for *conformity*. Becker (1991) suggests that the pleasure from a good (e.g., food from a restaurant) may be inherently greater when more people want to consume it. Greater number of early adopters can influence the confidence about a product (Hellofs and Jacobson 1999), creating a perception as a standard (Brynjolfsson and Kemerer 1996), indicating fashion trends which is important for customer subgroups who seek conformity from peers (Abrahamson and Rosenkopf 1993; Moretti 2011), or increasing reputation (Gabszewicz and Garcia 2008).

We focus on positive externality that enables consumers to reduce their "subjective psychological" anxiety as the number of early adopters increases. However, we do not consider the issue of social learning under which consumers can learn about the "objective value" of the product through customer ratings or word-of-mouth (Banerjee 1992; Bikhchandani et al. 1998; Yu et al. 2016; Papanastasiou and Savva 2016). This allows us to examine the interaction of promotion and externality, when both factors affect the dynamics between early adopters and late adopters. Specifically, our equilibrium analysis reveals that, unlike the conventional wisdom for promoting new products, for anxiety-inducing innovations with positive externality, accelerating the speed of adoption through familiarization can be detrimental to the firm.

2.3 The Model

A firm introduces a new innovative product with an "objective valuation" v that is known and common to all consumers. We consider a two-period model in which the firm seeks to maximize its profit over two periods by making the following decisions. In period 1 (the launch period), the firm determines whether to launch a promotion campaign to familiarize the consumers with the product, then sets the selling price p_1. In period 2 (post-launch period), after observing the realized demand in period 1, D_1, the firm sets the price p_2. For ease of exposition, we scale the firm's discount factor to 1 and production cost to 0.

All consumers are rational and determine the timing of their purchases to maximize their expected utilities. In our model, a consumer's expected utility is governed by both "objective" and "subjective" components. The objective component consists of the value of the product v (constant over two periods) and prices p_1 and p_2. The subjective component pertains to consumers' psychological anxiety about adopting the innovation. The psychological discomfort is represented as a *subjective disutility*. Consumers are heterogeneous in their subjective anxiety levels (Hoeffler 2003; Wang 1997), and consumer i's initial anxiety level x_i is uniformly distributed over $[0, v]$. For simplicity, we normalize the market size to 1 (i.e., demand is equivalent to proportion of consumers).[3]

Consumer i's anxiety level in periods 1 and 2 depends on his or her initial anxiety level x_i and the two effects of familiarization (F) and externality (E). First, the familiarization (F) effect occurs when the firm invests in a promotion campaign to alleviate consumers' anxiety level. In our model, the firm can invest K at the beginning of period 1 so that the anxiety levels are reduced by a factor $\alpha < 1$. As a result, consumer i's initial anxiety level reduces from x_i to $\alpha \cdot x_i$. No investment for familiarization corresponds to the case when $K = 0$ and $\alpha = 1$.[4] We will refer to the parameter α as the *effectiveness* of the firm's familiarization effort.

Second, the externality (E) effect occurs because consumers' period 2 anxiety levels would decrease in D_1, the number of early adopters in period 1. Specifically, the anxiety level in period 1 x_i will be revised to $(\beta/D_1) \cdot x_i$ in period 2 for any consumer i. This simple adjustment rule captures the notion of positive externality because the anxiety level $(\beta/D_1) \cdot x_i$ is adjusted downwards if the number of adopters in period 1 satisfies $D_1 > \beta$. (We use this simple adjustment rule to capture the notion of positive externality so that the anxiety level is decreasing in D_1.[5] Clearly, one can easily model the revised anxiety level in period 2 as $(\beta/D_1 - \beta) \cdot x_i$ so that it equals ∞ when $D_1 = 0$ and equals 0 when $D_1 = 1$.) Here, the parameter $\beta \in (0, 1)$ is an exogenously given parameter and it can be interpreted as the consumers' *reference point* regarding D_1 and is associated with the characteristics of the product market environment (Oliver 1980; Thaler 1985; Hoch and Loewenstein 1991), and the value of β can be estimated through lab experiments.

Third, the joint familiarization and externality (F+E) effect occurs when the firm promotes in period 1 and the positive externality takes effect in period 2. In this case,

---

[3]Our model is easily generalizable to the case where the x_i is uniformly distributed over $[0, R]$ for any R.

[4]We treat K and α as parameters for ease of analysis, but α, in principle, can be a function of K. Specifically, because our focus is on the pricing decisions, we shall consider the case when K and α are exogenously given to simplify our analysis. However, if investment K is the focused decision, then one needs to model K as a function of α, and the functional form and the associated value can be estimated through lab experiments. We shall relegate this issue to a future research topic.

[5]It can also be adjusted upward if $D_1 < \beta$, i.e., if there are few early adopters, then it can actually raise the anxiety levels for the remaining consumers. This characterization is consistent with the notion that consumers are more (less) willing to enter a restaurant when it is full (empty).

Table 2.1 Consumer i's utilities based on initial anxiety level x_i, presence of F or E, and prices p_1, p_2

	Period 1	Period 2
F only	$u_i = v - \alpha x_i - p_1$	$u_i = \delta(v - \alpha \cdot x_i - p_2)$
E only	$u_i = v - x_i - p_1$	$u_i = \delta(v - (\beta/D_1) \cdot x_i - p_2)$
F+E	$u_i = v - \alpha x_i - p_1$	$u_i = \delta(v - (\alpha\beta/D_1) \cdot x_i - p_2)$

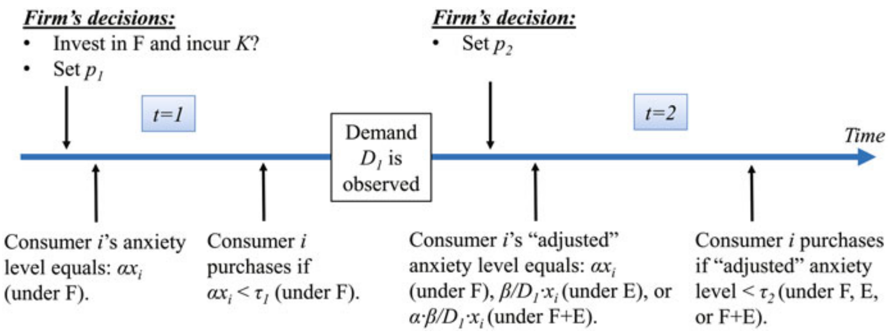

Fig. 2.1 Sequence of events

consumer i's anxiety level is reduced from x_i to αx_i in period 1, and in period 2 it is further revised to $\alpha(\beta/D_1)x_i$. By assuming that each consumer has a unit demand and consumes the product at the time of purchase, and that all consumers have a common discount factor $\delta < 1$ due to delayed consumption, we can summarize consumer i's utility for the new product in period 1 and period 2 without and with the presence of externality in Table 2.1.

Figure 2.1 summarizes the sequence of our model. In the beginning of period 1, the firm first determines whether to launch a promotion campaign by investing K (or not) to familiarize consumers and lower their anxiety by a factor of $\alpha \leq 1$. It then sets price p_1. The consumers decide whether to purchase in period 1 or not by comparing their corresponding expected utilities between purchasing now, later, or leaving the market, i.e., $v - \alpha x_i - p_1 \geq \max\{\delta(v - \alpha(\beta/D_1)x_i - p_2), 0\}$. Because both future price p_2 and anxiety level $\alpha(\beta/D_1)x_i$ are not known at the time of decision in period 1, consumers form rational expectations about those values to evaluate their utilities. At the end of period 1, demand D_1 is realized, which is observed by the firm and the remaining consumers. Notice that the delayed purchasing decision in period 2 is driven not only from the anticipation of a lower price, but also from the anticipation of a lower anxiety level, which depends on both firm's pricing decision and consumers' rational expectation. In the rational expectation equilibrium, consumer i's period 1 expectations about p_2 and D_1 coincide with the realized p_2 and D_1. In period 2, the firm sets price p_2 and the remaining consumers use their corresponding utilities in period 2 as stated in Table 2.1 to decide whether to purchase the product.

2.4 Analysis

To examine the implications of familiarization, externality, and their interaction on the equilibrium outcomes, we first examine independently the setting with familiarization (F) only (Sect. 2.4.1) and externality (E) only (Sect. 2.4.2). Then, we examine the equilibrium outcomes when both mechanisms (F+E) are present (Sect. 2.4.3), and examine the managerial implication of their interaction (Sect. 2.4.4).

2.4.1 Firm's Familiarization Effort (F Only)

We begin by examining the case where the firm invests K to promote familiarization of the new product to consumers at the time of launch in period 1. Due to the familiarization effect induced by the promotion, consumer i's anxiety level is reduced from x_i to αx_i with $\alpha < 1$ throughout periods 1 and 2 as captured by those utilities in Table 2.1.

To build intuition of our equilibrium analysis based on backward induction, consider the case when $\alpha = 1$ (i.e., no promotion). We begin by analyzing consumers' purchasing decision in period 2. We shall show that all consumers will adopt the following purchasing strategy in equilibrium. Consumers with lower anxiety levels (below a threshold τ_1) will adopt the new product early (in period 1), while consumers with higher anxiety levels will delay their decision (until period 2) and purchase in period 2 if their anxiety levels are below a threshold, τ_2.[6] Because x_i is uniformly distributed over $[0, v]$, the demand in period 1 is $D_1 = \tau_1/v$. Therefore, all remaining consumers in period 2 have an anxiety level $x_i \in (\tau_1, v]$. For any given selling price p_2, a consumer remaining in period 2 will purchase the new product if and only if the consumer's utility $u_i = v - x_i - p_2 > 0$, i.e., when the anxiety level $x_i \leq \tau_2 \equiv v - p_2$. In this case, the demand in period 2 is $D_2 = (\tau_2 - \tau_1)/v = (v - p_2 - \tau_1)/v$.

Anticipating the demand D_2 in period 2 as given above, the firm sets price p_2 to maximize its period 2 profit (π_2),

$$\pi_2^* = \max_{p_2}\{p_2 \cdot D_2(p_2)\} = \max_{p_2}\left\{p_2 \cdot \frac{v - p_2 - \tau_1}{v}\right\}.$$

Taking the first-order condition with respect to p_2, we obtain the following optimal price, demand, and profit in period 2 as a function of τ_1:

[6]These thresholds will be determined endogenously as we determine the equilibrium purchasing strategy.

$$p_2^*(\tau_1) = \frac{v - \tau_1}{2}; \quad D_2^*(\tau_1) = \frac{v - \tau_1}{2}\frac{1}{v}; \quad \pi_2^*(\tau_1) = \left(\frac{v - \tau_1}{2}\right)^2 \frac{1}{v}.$$

We next examine the period 1 problem, where the firm selects p_1 to maximize its total profit in both periods. First, observe that in a rational expectation equilibrium, a consumer will purchase the new product in period 1 if and only if her surplus from purchasing in period 1 is non-negative and is higher than the surplus from purchasing in period 2, i.e., $v - x_i - p_1 \geq 0$ and $v - x_i - p_1 \geq \delta(v - x_i - p_2)$. Hence, in equilibrium, a consumer with anxiety level $x_i = \tau_1$ is indifferent between purchasing in period 1 or period 2, so that $v - \tau_1 - p_1 = \delta(v - \tau_1 - p_2)$. By using the fact that $p_2^*(\tau_1) = (v - \tau_1)/2$, the equilibrium price p_1 can be expressed as a function of τ_1,

$$p_1(\tau_1) = \left(1 - \frac{\delta}{2}\right)(v - \tau_1). \tag{2.1}$$

Second, recall that the demand in period 1 is $D_1(p_1) = \tau_1/v$. Because the firm's profit in period 2 is given as $\pi_2^*(\tau_1) = [(v - \tau_1)/2]^2 \cdot (1/v)$, the firm's problem for period 1 can be formulated as:

$$\pi^* = \max_{p_1} \left\{p_1 \cdot \frac{\tau_1}{v} + \pi_2^*(\tau_1)\right\},$$
$$\text{s.t.} \quad (2.1) \quad \text{holds.}$$

Using (2.1) to transform the decision variable from p_1 to τ_1, the firm's problem is reformulated to

$$\pi^* = \max_{\tau_1 \leq v}\left\{\left(1 - \frac{\delta}{2}\right)\frac{(v - \tau_1)\tau_1}{v} + \left(\frac{v - \tau_1}{2}\right)^2 \frac{1}{v}\right\}. \tag{2.2}$$

Because the objective function is concave in τ_1, one can use the first-order condition to determine the optimal value of τ_1, which can in turn be used to retrieve the equilibrium outcomes for both periods—p_1^*, p_2^*, D_1^*, and D_2^*—in closed form.

This argument can be extended to the case where $\alpha < 1$ by replacing x_i with αx_i. When α is sufficiently small, the analysis must take into account the boundary condition that demand across both periods is limited to 1 ($D_1^* + D_2^* = 1$), i.e., market saturation occurs (see Huang et al. 2018 for details). We obtain the following results.

Proposition 2.1 (Familiarization Only) *When the firm invests K in the familiarization effort to alleviate consumer anxiety by a factor of α, the equilibrium prices (p_1^*, p_2^*), demands (D_1^*, D_2^*), and total profit (π^*) can be expressed as follows: Let* $\hat{\alpha}(\delta) \equiv (3 + \delta)^{-1}(2 + \sqrt{4 - (3 + \delta)(2 - \delta)^2/(3 - 2\delta)})$,

1. *When* $\alpha \geq \hat{\alpha}(\delta)$*,*

$$p_1^* = \frac{v}{2}\frac{(2-\delta)^2}{3-2\delta}, \qquad p_2^* = \frac{v}{2}\frac{2-\delta}{3-2\delta},$$
$$D_1^* = \frac{1}{\alpha}\frac{1-\delta}{3-2\delta}, \qquad D_2^* = \frac{1}{2\alpha}\frac{2-\delta}{3-2\delta},$$
$$\pi^* = \frac{v}{4\alpha}\cdot\frac{(2-\delta)^2}{3-2\delta} - K;$$

2. *When* $\alpha < \hat{\alpha}(\delta)$*,*

$$p_1^* = v\left(1-\alpha\frac{1+\delta}{2}\right), \qquad p_2^* = (1-\alpha)v,$$
$$D_1^* = \frac{1}{2}, \qquad D_2^* = \frac{1}{2},$$
$$\pi^* = \frac{v}{4}(4-\alpha(3+\delta)) - K.$$

Proposition 2.1 suggests that, when the familiarization effect is moderate (i.e., $\alpha \geq \hat{\alpha}(\delta)$), the firm does not capture the whole market (i.e., $D_1^* + D_2^* < 1$) and the optimal profit π^* is convex and decreasing in α. However, when the familiarization effect is strong (i.e., $\alpha < \hat{\alpha}(\delta)$), the firm saturates the market (i.e., $D_1^* + D_2^* = 1$) and the optimal profit π^* is linearly decreasing in α. Corollary 2.1 examines the structural properties of the equilibrium outcomes under familiarization.

Corollary 2.1 (Structural Properties of Equilibrium Outcomes (F Only)) *The equilibrium outcomes exhibit the following characteristics. For any effectiveness level* α*,*

(i) *the firm will reduce its selling price in period 2, i.e.,* $p_1^* \geq p_2^*$*;*
(ii) *the majority of consumers will purchase in period 2, i.e.,* $D_1^* \leq D_2^*$*;*
(iii) *the firm's profit* π^* *is decreasing in the discount factor* δ*,* $\forall \alpha$*.*

Corollary 2.1 reveals that, in the absence of externality, the price markdown strategy is always optimal. This is because, due to the familiarization effect, the level of anxiety is reduced by the same factor α in both periods, thus all remaining consumers in period 2 will still have relatively higher anxiety level than the early adopters who have purchased in period 1. To entice them to purchase and generate demand in period 2, the firm has to reduce its price in period 2. Also, due to the markdown strategy, it is intuitive that more consumers will delay their purchase until period 2. Finally, as the discount factor δ increases, more consumers are willing to postpone their purchasing decision until period 2 to purchase at a lower price, resulting in lower profit for the firm.

Figure 2.2 illustrates the optimal profit π^* with respect to $\alpha \in [0, 1]$ when the promotion requires investment of $K = 2$. The firm's optimal profit is continuous

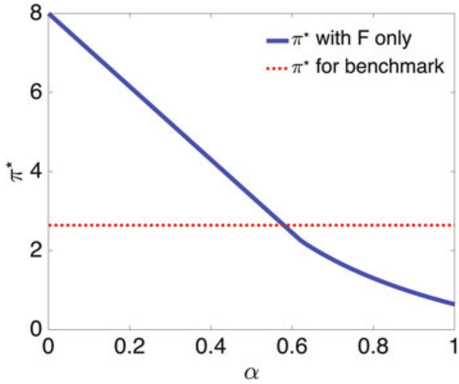

Fig. 2.2 The optimal firm profit with familiarization with respect to its effectiveness (α) (in this plot, $\delta = 0.7$, $\hat{\alpha}(\delta = 0.7) = 0.6$, and $K = 2$)

and convex decreasing in α. The profit without promotion ($K = 0$ and $\alpha = 1$) is represented by the horizontal dotted line. Figure 2.2 confirms the intuition that investing in familiarization is beneficial only when, relative to cost (K), the familiarization effect is sufficiently strong (α sufficiently low). This intuitive result is also consistent with broad marketing literature (Lilien et al. 1992), which highlights the beneficial effects of consumers' familiarization with new products (e.g., offering free samples/trials, advertising) on firms' profits.

2.4.2 Externality Among Consumers (E Only)

We now examine the case when only externality (E) is present. In this case, Table 2.1 shows that the initial anxiety level x_i of each remaining consumer i in period 2 will be adjusted to $(\beta/D_1) \cdot x_i$. Thus, in the presence of externality, consumers may delay their purchase decisions also in anticipation of observing a large number of early adopters D_1 and lowering their anxiety levels. When taking this positive externality into consideration, the firm may want to lower its price in period 1 to generate higher early demand D_1, which will reduce the anxiety for those remaining consumers in period 2. However, selling at reduced price to too many consumers in period 1 is also not desirable because it leaves too few remaining consumers for the firm to sell to in period 2. Therefore, the optimal pricing strategy in equilibrium is far from obvious. By using the same approach as in the previous section, we get:

Proposition 2.2 (Externality Only) *When consumers adjust their anxiety level from x_i to $(\beta/D_1) \cdot x_i$ according to externality at the beginning of period 2, the equilibrium prices (p_1^*, p_2^*), demands (D_1^*, D_2^*), and the total profit for both periods (π^*) can be expressed as follows: $\exists\hat{\beta}(\delta)$ available in implicit form such that,*

1. *If* $\beta < \hat{\beta}(\delta)$*, then*

$$p_1^* = v(1+\delta\beta) - \frac{\delta\beta v^2}{\tau_1^*} - \tau_1^*, \qquad p_2^* = v - \frac{\beta v^2}{\tau_1^*},$$
$$D_1^* = \frac{\tau_1^*}{v}, \qquad D_2^* = \frac{v - \tau_1^*}{v},$$
$$\pi^* = p_1^* D_1^* + p_2^* D_2^*,$$

where

$$\tau_1^* = \max\left\{\frac{v\delta(1+\beta)}{2}, v\left(\frac{\delta\beta}{6} + (M_1 + \sqrt{M_2})^{1/3} + (M_1 - \sqrt{M_2})^{1/3}\right)\right\}$$
$$M_1 := \left(\frac{\delta\beta}{6}\right)^3 + \frac{\beta}{4}, \quad M_2 := \frac{\beta}{2}\left(\frac{\delta\beta}{6}\right)^3 + \left(\frac{\beta}{4}\right)^2;$$

2. *If* $\beta \geq \hat{\beta}(\delta)$*, then*

$$p_1^* = \frac{v}{2}\left(1 - \frac{\delta(1-\beta)}{2} - \frac{(1-\beta)^2}{4\beta}\right), \qquad p_2^* = \frac{v}{2}(1-\beta),$$
$$D_1^* = \frac{1}{2}\left(1 - \frac{\delta(1-\beta)}{2} + \frac{(1-\beta)^2}{4\beta}\right),$$
$$D_2^* = \frac{1-\beta}{4\beta}\left(1 - \frac{\delta(1-\beta)}{2} + \frac{(1-\beta)^2}{4\beta}\right),$$
$$\pi^* = \frac{v}{4}\left(\frac{1-\delta}{2} + \frac{2\delta+1}{4}\beta + \frac{1}{4\beta}\right)^2.$$

When the consumer reference point β is not large ($\beta < \hat{\beta}(\delta)$), the externality effect is favorable to the firm. This is because the anxiety level would be more easily adjusted downwards by the remaining consumers in period 2, leading to higher demand in period 2. In this case, the firm is able to capture the entire market, i.e., $D_1^* + D_2^* = 1$. In the second case when β is large ($\beta \geq \hat{\beta}(\delta)$), however, the externality effect is unfavorable to the firm and the firm does not capture the whole market, $D_1^* + D_2^* < 1$. The following corollary examines the structural properties of the equilibrium outcomes and the impact of externality.

Corollary 2.2 (Structural Properties of Equilibrium Outcomes (E Only)) *In the presence of externality where consumers adjust their anxiety level from* x_i *to* $(\beta/D_1) \cdot x_i$ *at the beginning of period 2,*

(i) *the firm employs a markup pricing* ($p_1^* < p_2^*$) *when* β *is lower and a markdown pricing* ($p_1^* \geq p_2^*$) *when* β *is higher;*

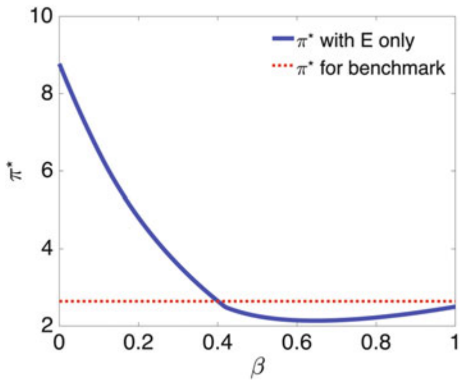

Fig. 2.3 The optimal firm profit in the presence of externality with respect to consumer reference point β (in this plot, $\delta = 0.7$ and $\hat{\beta}(\delta = 0.7) = 0.42$)

(ii) *the equilibrium demands are so that* $D_1^* < D_2^*$ *when* β *is lower and* $D_1^* \geq D_2^*$ *when* β *is higher;*

(iii) *the firm's profit* π^* *is decreasing in the discount factor* δ, $\forall \beta$.

Corollary 2.2 reveals that the optimal strategy hinges on the consumer reference point β. Specifically, when the consumer reference point β is lower (i.e., when the externality effect is favorable), the first statement reveals that markup pricing is optimal (i.e., $p_1^* < p_2^*$). This is because a low initial price p_1 will generate sufficient demand in period 1 (i.e., D_1) that would reduce the anxiety of the remaining consumers in period 2. As a result, consumers in period 2 would have higher utilities for the product, and the firm can afford to charge a higher premium price in period 2. We obtain the opposite result when the consumer reference point β is higher. These results are consistent with previous studies which investigate dynamic pricing with externality (Dhebar and Oren 1985; Bensaid and Lesne 1996). Namely, if the externality effect is strong enough, markup pricing is optimal; otherwise, markdown pricing is optimal (Bensaid and Lesne 1996). We can interpret all other statements in a similar manner.

Figure 2.3 plots the optimal profit in the presence of externality with respect to the consumer reference level $\beta \in (0, 1)$. The optimal firm profit π^* is continuous and convex in β, but no longer monotonic. The profit with no externality (and no familiarization effort by the firm, $K = 0$ and $\alpha = 1$) is illustrated again by the horizontal dotted line. As discussed, we observe that externality is beneficial when the consumer reference level β is lower (favorable), and detrimental when it is higher (unfavorable).

So far, we have learned that a firm can benefit from either an effective familiarization effort (lower α) as shown in Fig. 2.2 or the presence of favorable externality (lower β) as shown in Fig. 2.3. These two observations raise the following question: Will the firm benefit from both the familiarization and externality effects? We shall examine this question next.

2.4.3 Combined Effect of Familiarization and Externality (F+E)

We now examine the case when both anxiety-mitigating mechanisms of familiarization (F) and externality (E) are present. By considering the utilities as given in Table 2.1 and by using the same approach as before, we get:

Proposition 2.3 (Familiarization and Externality) *When the firm invests K in familiarization effort and externality is present among consumers, the equilibrium prices (p_1^*, p_2^*), demands (D_1^*, D_2^*), and total profit (π^*) can be expressed as follows: $\exists\hat{\beta}(\alpha, \delta)$ available in an implicit form such that,*

1. *If $\beta \geq \hat{\beta}(\alpha, \delta)$, then*

$$p_1^* = \frac{v}{2}\left(1 - \frac{\delta(1-\alpha\beta)}{2} - \frac{(1-\alpha\beta)^2}{4\alpha\beta}\right), \qquad p_2^* = \frac{v}{2}(1-\alpha\beta),$$

$$D_1^* = \frac{1}{2\alpha}\left(1 - \frac{\delta(1-\alpha\beta)}{2} + \frac{(1-\alpha\beta)^2}{4\alpha\beta}\right),$$

$$D_2^* = \frac{1}{4\alpha}\frac{(1-\alpha\beta)}{\alpha\beta}\left(1 - \frac{\delta(1-\alpha\beta)}{2} + \frac{(1-\alpha\beta)^2}{4\alpha\beta}\right),$$

$$\pi^* = \frac{v}{4\alpha}\left(\frac{1-\delta}{2} + \frac{2\delta+1}{4}\alpha\beta + \frac{1}{4\alpha\beta}\right)^2 - K;$$

2. *If $\beta < \hat{\beta}(\alpha, \delta)$, then*
   (a) *If $\alpha \geq \delta/(2-\delta\beta)$,*

$$p_1^* = v(1+\delta\alpha\beta) - \frac{\delta\alpha\beta v^2}{\tau_1^*} - \alpha\tau_1^*, \qquad p_2^* = v - \frac{\alpha\beta v^2}{\tau_1^*},$$

$$D_1^* = \frac{\tau_1^*}{v}, \qquad D_2^* = \frac{v-\tau_1^*}{v},$$

$$\pi^* = p_1^* D_1^* + p_2^* D_2^* - K,$$

where

$$\tau_1^* = \max\left\{\frac{v\delta(1+\alpha\beta)}{2\alpha}, v\left(\frac{\delta\beta}{6} + (M_1 + \sqrt{M_2})^{1/3} + (M_1 - \sqrt{M_2})^{1/3}\right)\right\}$$

$$M_1 := \left(\frac{\delta\beta}{6}\right)^3 + \frac{\beta}{4}, \quad M_2 := \frac{\beta}{2}\left(\frac{\delta\beta}{6}\right)^3 + \left(\frac{\beta}{4}\right)^2$$

   (b) *If $\alpha < \delta/(2-\delta\beta)$, then*

$$p_1^* = v\left(1 - \frac{\delta(1-\alpha\beta)}{2} - \frac{2\alpha^2\beta}{1+\alpha\beta}\right), \qquad p_2^* = v\left(1 - \frac{2\alpha^2\beta}{\delta(1+\alpha\beta)}\right)$$

$$D_1^* = 1, \qquad D_2^* = 0,$$

$$\pi^* = v\left(1 - \frac{\delta(1-\alpha\beta)}{2} - \frac{2\alpha^2\beta}{1+\alpha\beta}\right) - K.$$

Proposition 2.3 is consistent with the results as stated in both Propositions 2.1 and 2.2. However, statement 2(b) of Proposition 2.3 additionally suggests that it is optimal for the firm to accelerate adoption to capture the whole market in period 1 (i.e., $D_1^* = 1$) when familiarization is sufficiently effective (i.e., when $\alpha < \delta/(2 - \delta\beta)$). The following results reveal the structural properties of equilibrium outcomes and how familiarization and externality interact to influence them.

Corollary 2.3 (Structural Properties of Equilibrium Outcomes (F+E)) *Consider the case when the firm invests K to induce the familiarization effect in the presence of externality.*

(i) *For any β, as α decreases, p_1^* decreases and p_2^* increases, shifting the optimal pricing from markdown towards markup pricing;*
(ii) *For any β, as α decreases, demands in both periods D_1^* and D_2^* increase in case 1 of Proposition 2.3; D_1^* increases to 1 and D_2^* decreases to 0 in case 2(a); and remain unchanged at $D_1^* = 1$ and $D_2^* = 0$ in case 2(b);*
(iii) *The firm's profit π^* is decreasing in the discount factor δ, $\forall \alpha, \beta$.*

Unlike Corollary 2.1 stating that markdown pricing is always optimal when externality is not present, the first part of Corollary 2.3 reveals that externality does affect the optimal pricing strategy. Specifically, in the presence of externality, more effective familiarization campaign decreases the optimal period 1 price (p_1^*) and increases the optimal period 2 price (p_2^*), thereby reducing the extent of the markdown. Furthermore, depending on the externality effect due to the consumer reference point β, the firm's optimal pricing strategy can switch from markdown pricing to markup pricing. We can interpret other statements in a similar fashion.

2.4.4 Managerial Implications of Interaction Between Familiarization and Externality

We now examine the implication of the interaction between familiarization (F) and externality (E) on the firm's decision to invest in a familiarization campaign, by comparing the optimal firm profit with both familiarization and externality (π^*(F+E)) to that with externality only (π^*(E)). By using the closed form expressions of equilibrium profits as stated in Propositions 2.2 and 2.3, we are able

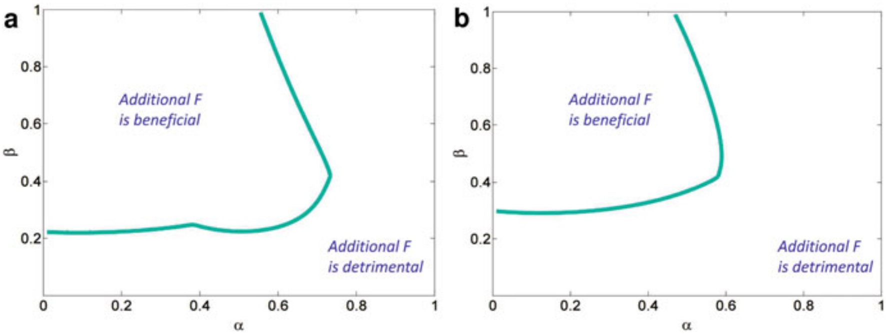

Fig. 2.4 Plots of regions where additional familiarization in the presence of externality helps or hurts firm profit for $K = 2$ ((**a**) π^* for $\delta = 0.7$ and (**b**) $\delta = 0.95$)

to examine the conditions (in terms of α, β, and δ) under which $\pi^*(\text{F+E}) > \pi^*(\text{E})$. Figure 2.4 characterizes the region in the (α, β) space where additional familiarization in the presence of externality benefits or hurts the firm profit, for $\delta = 0.7$ (left panel) and $\delta = 0.95$ (right panel).

First, observe from the upper left region in both panels that running a familiarization campaign is beneficial when externality is unfavorable (i.e., when the consumer reference point β is high) and the familiarization campaign is effective (i.e., when α is low). This is because, when β is high, running a familiarization campaign can reduce the initial anxiety level drastically (due to low α), which will make familiarization campaign beneficial. Second, when this condition does not hold, it can be seen from Fig. 2.4 that running a familiarization campaign is detrimental to the firm in all other regions. Specifically, consider the case when externality is favorable (i.e., when β is low), both panels reveal that, regardless of the effectiveness of the familiarization campaign, running such a campaign is detrimental to the firm even when the campaign is costless, i.e., when $K = 0$.[7] Why would additional familiarization effort exerted by the firm ever backfire when externality effect is favorable? To understand this, observe that when externality is favorable (β is lower), the anxiety level reduces and consumer utility increases significantly in period 2. As such, the firm can charge those late adopters a premium price in period 2. However, if the firm runs a familiarization campaign in period 1, it will accelerate the adoption rate (shifts demand from period 2 to period 1), leaving fewer late adopters for the firm to charge at a premium price. This implies that, when externality is favorable (i.e., when the consumer reference point β is low), running a familiarization campaign in period 1 is detrimental to the firm.

[7]When $K = 0$, the two regions will be delineated by a curve resembling a horizontal line around $\beta = 0.2$. The detrimental effect of familiarization when β is low holds even when α is a decreasing function of K.

2.5 Summary and Discussion

While existing literature focuses on the issue of social learning by examining the case when consumers are unsure about the value of a new innovative product, we consider the case where consumers are reluctant to make a purchase because they experience psychological anxiety of adoption associated with disrupting established routine or habit, learning the necessary steps to use the innovation effectively, or parting from an old technology they are emotionally attached to. Such anxiety can be mitigated by the firm's effort to help consumers to become familiarized with the product and by the number of early adopters, i.e., externality. We investigated how these two effects interact and examined how they affect the firm's optimal promotion and pricing strategies. Unlike the conventional wisdom for promoting new products, we find that for anxiety-inducing innovations with externality, accelerating the speed of adoption through promotion can be detrimental to the firm. Specifically, our equilibrium results reveal that, (1) when externality is absent, price markdown is always optimal for the firm; however, when externality is present and becomes more favorable, price markup is optimal; (2) the firm should carefully devise its promotion campaign because it can be detrimental when externality is present and favorable. Our model is based on a theoretical construct and our equilibrium results can only serve as hypotheses for further examination. Therefore, it is of interest to test these hypotheses via empirical studies of firms who launch innovative products or via lab experiments to deepen our understanding about the issues of consumer anxiety and (positive or negative) externalities. We relegate these studies to future research studies.

Various extensions to our model are possible. Our model currently assumes a discount factor for consumers but not for the firm. One can incorporate the firm's discount factor in our model to examine its impact on firm's strategy, which may be relevant in an entrepreneurial setting. Externalities can also be negative, for example when consumers seek exclusivity or status-seeking behavior (Gao et al. 2016). As such, an interesting future work would be to examine the effect of negative externality. Moreover, how should firms make decisions when the anxiety-alleviating effects α and β are uncertain, or how the results might change when the distribution of the anxiety level x_i is altered (Johnson and Myatt 2006) can also be interesting future enquiries. It is our hope that this chapter will stimulate researchers to further explore consumer psychological externality and its impact on firms' operational decisions and strategies. We also believe that there are numerous opportunities to examine the issue of consumer psychological externalities via behavioral experiments.

References

Abrahamson E, Rosenkopf L (1993) Institutional and competitive bandwagons: using mathematical modeling as a tool to explore innovation diffusion. Acad Manag Rev 18(3):487–517
Bandejee AV (1992) A simple model of herd behavior. Q J Econ 107(3):797–817
Bass FM (1969) A new product growth for model of consumer durables. Manag Sci 15(5):215–227
Becker GS (1991) A note on restaurant pricing and other examples of social influences on price. Eur J Polit Econ 99(5):1109–1116
Bensaid B, Lesne JP (1996) Dynamic monopoly pricing with network externalities. Int J Ind Organ 14(6):837–855
Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the behavior of others: conformity, fads, and informational cascades. J Econ Perspect 12(3):151–170
Brynjolfsson E, Kemerer CF (1996) Network externalities in microcomputer software: an econometric analysis of the spreadsheet market. Manag Sci 42(12):1627–1647
Castano R, Kacker M, Sujan H (2008) Managing consumer uncertainty in the adoption of new products: temporal distance and mental simulation. J Market Res 45(3):320–336
Chattergee R, Eliashberg J (1990) The innovation diffusion process in a heterogeneous population: a micromodeling approach. Manag Sci 36(9):1057–1079
Dahl DW, Moreau P (2002) The influence and value of analogical thinking during new product ideation. J Market Res 39(1):47–60
Dhebar A, Oren SS (1985) Optimal dynamic pricing for expanding networks. Market Sci 4(4):336–351
Fournier S (1998) Consumers and their brands: developing relationship theory in consumer research. J Consum Res 24(4):343–353
Gabszewicz JJ, Garcia F (2008) A note on expanding networks and monopoly pricing. Econ Lett 98(1):9–15
Gao SY, Lim WS, Tang CS (2016) Entry of copycats of luxury brands. Market Sci 36(2):272–289
Gatignon H, Robertson TS (1985) A propositional inventory for new diffusion research. J Consum Res 11(4):849–867
Hellofs LL, Jacobson R (1999) Market share and customers' perceptions of quality: when can firms grow their way to higher versus lower quality? J Market 63(1):16–25
Hoch SJ, Loewenstein GF (1991) Time-inconsistent preferences and consumer self-control. J Consum Res 17(4):492–507
Hoeffler S (2003) Measuring preferences for really new products. J Market Res 40(4):406–420
Huang Y, Gokpinar B, Tang CS, Yoo OS (2018) Selling innovative products in the presence of externalities. Prod Oper Manag 27(7):1236–1250
Johnson JP, Myatt DP (2006) On the simple economics of advertising, marketing, and product design. Am Econ Rev 96(3):756–784
Katz ML, Shapiro C (1985) Network externalities, competition, and compatibility. Am Econ Rev 75(3):424–440
Lilien GL, Kotler P, Moorthy KS (1992). Marketing models. Prentice Hall, Englewood Cliffs, NJ
Moretti E (2011) Social learning and peer effects in consumption: evidence from movie sales. Rev Econ Stud 78(1):356–393
Mukherjee A, Hoyer W (2001) The effect of novel attributes on product evaluation. J Consum Res 28(3):462–472
Oliver RL (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. J Market Res 17(4):460–469
Papanastasiou Y, Savva N (2016). Dynamic pricing in the presence of social learning and strategic consumers. Manag Sci 63(4):919–939
Peres R, Muller E, Mahajan V (2010) Innovation diffusion and new product growth models: a critical review and research directions. Int J Res Market 27(22):91–106
Ram S (1989) Successful innovation using strategies to reduce consumer resistance: an empirical test. J Prod Innov Manag 6(1):20–34

Ram S, Sheth JN (1989) Consumer resistance to innovation: the marketing problem and its solutions. J Consum Market 6(2):5–14
Ren H, Huang T (2018) Modeling customer bounded rationality in operations management: a review and research opportunities. Comput Oper Res 91:48–58
Sheth JN, Newman BI, Gross BL (1991) Why we buy what we buy: a theory of consumption values. J Bus Res 22(2):159–170
Thaler R (1985) Mental accounting and consumer choice. Market Sci 4(3):199–214
Wang H (1997) Treatment of "don't-know responses" in contingent valuation surveys: a random valuation model. J Environ Econ Manag 32(2):219–232
Yu M, Debo L, Kapuscinski R (2016) Strategic waiting for consumer-generated quality information: dynamic pricing of new experience goods. Manag Sci 62(2):410–435

Chapter 3
Buyer Valuation Uncertainty and Firm Information Provision Strategies

Jane Z. Gu and Rachel R. Chen

Abstract This chapter reviews research on buyer valuation uncertainty originated from information asymmetry between the firm and consumers, and the firm's information provision strategy. Before purchase, consumers could be uncertain about the product's vertical attributes, i.e., quality uncertainty, and/or the product's horizontal attributes, i.e., fit uncertainty. For each type of uncertainty, we discuss the firm's inventive and instruments to disclose information, as well as other mechanisms to reveal information that help consumers resolve such valuation uncertainty. We then review recent literature on advance selling and opaque selling strategies, where the firm benefits from creating consumer valuation uncertainty. We conclude the chapter with discussions on future research directions.

Keywords Buyer valuation uncertainty · Information provision · Quality disclosure · Fit revelation · Information asymmetry

3.1 Overview

Consumers are commonly uncertain about the value of a product prior to purchase and such valuation uncertainty impedes their purchase intentions. Endowed with more product information, firms may have incentive to disclose such information to help resolve consumer valuation uncertainty originated from information asymmetry. In this chapter, we review research works that investigate motivations and consequences of firms' information provision activities.

J. Z. Gu
School of Business, University of Connecticut, Storrs, CT, USA
e-mail: jane.gu@uconn.edu

R. R. Chen (✉)
Graduate School of Management, University of California at Davis, Davis, CA, USA
e-mail: rachen@ucdavis.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_3

A product can be viewed as a collection of vertical and horizontal attributes. Vertical attributes, such as material and craftsmanship, constitute the product's quality, which can be measured and compared on a one-dimensional scale. Consumers have homogeneous preferences for product vertical quality in the sense that they all derive a greater consumption value from a higher-quality product, despite their different willing to pay for the quality premium. For example, consumers are likely to agree that a flute made with superior material and craftsmanship has a high value, but not everyone accepts its hefty price tag. The homogeneous nature of consumer quality preferences leads consumer valuations of a product to *converge* upon firm disclosure of the product's vertical attributes. In particular, disclosure of a product quality higher (lower) than expected shifts all consumers' product valuations upward (downward). As such, a firm that offers a higher product quality has stronger incentive to disclose vertical attributes of its product. Moreover, since quality can be measured and compared on a scale universally agreed upon, an individual consumer can learn about a product's vertical quality from other consumers' product experiences. In recent years, the growing prevalence of online review platforms has greatly alleviated consumer quality uncertainty.

While early research has mainly focused on issues related to disclosure of product vertical attributes, recent research has focused on issues related to disclosure of product horizontal attributes. Horizontal attributes, such as color and flavor, differentiate products even when they have the same quality. Consumers have heterogeneous preferences for product horizontal attributes in the sense that they are endowed with heterogeneous tastes, which lead to their different "fit" with a product. For example, some consumers like red color, whereas some others like green; similarly, some consumers like sweet flavor, whereas some others like spicy flavor. A product that provides a good fit or a high value for some consumers may be perceived as offering a "bad fit" or a low value for other consumers. This heterogeneous nature of product fit preferences leads consumer valuations of a product *diverge* upon firm disclosure of the product's horizontal attributes, posing a sharp contrast to the consequence of firm disclosure of a product's vertical attributes.

Moreover, note that a consumer's perceived product fit is specific to the individual consumer as well as the particular product. While the consumer perceives fit uncertainty because of her lack of information on product horizontal attributes, the firm also perceives uncertainty about how its product fits the consumer, owing to its lack of information on the consumer's taste. When the firm discloses horizontal attributes of its product, the former type of fit uncertainty perceived by the consumer is resolved, but the latter type of fit uncertainty perceived by the firm remains. That is, disclosing product horizontal attributes actually puts the firm at an information disadvantage, which suppresses its disclosure incentive. Another implication of the idiosyncratic nature of product fit is that knowing other people's fit with a product does not necessarily help a consumer evaluate her own product fit. The helpfulness of third-party reviews in resolving consumer fit uncertainty is thus limited, providing firms' opportunities to manipulate consumers' fit search via marketing tools. These complexities about consumer fit uncertainty add to the richness of this research area, which has drawn attention of scholars in the field of economics, marketing, operations management, and information systems as well.

Our review focuses on buyer valuation uncertainty that originated from information asymmetry between consumers and profit-maximizing firms. This can be mitigated through various information provision activities of the firm. A product's value may be subject to factors beyond the firm's knowledge or control. For example, a highly rated refrigerator may arrive with a defect after a rough delivery; a well-planned vacation may be ruined by unexpected weather conditions. In these cases, consumer valuation uncertainty remains despite the firm's full information disclosure, and can be mitigated though warranty, insurance, or compensation policies (e.g., Chen et al. 2009; Png and Wang 2010). In some institutional purchase contexts where the firm and the buyer co-create customized products such as production equipment, architecture design, or software systems, valuation uncertainty arises when the buyer is unable to articulate their needs ex ante and can be mitigated via the adoption of interactive communication tools (e.g., Terwiesch and Loch 2004). Our discussion does not cover this type of consumer uncertainties.

In the following, we first review earlier research related to firm disclosure of product's vertical attributes and then move on to review more recent research related to firm disclosure of product horizontal attributes. We then discuss the literature on firm's advance selling and opaque selling strategies, which create consumer valuation uncertainty by withholding product information. We conclude this chapter with discussions on future research directions.

3.2 Firm Disclosure of Product Vertical Attributes

The issue of consumer quality uncertainty has caught research attention since the 1970s. Akerlof (1970) considers less developed markets where truthful, credible disclosure is prohibitively expensive and concludes that all sellers would misrepresent quality. Milgrom (2008) and Dranove and Jin (2010) provide excellent reviews of early economic literature on quality disclosure and certification. Our discussion focuses on business contexts where information on product quality can be truthfully and credibly communicated to end customers. We review three major streams in this type of literature. The first stream investigates a firm's incentive to disclose product quality in various market structures. The second stream of literature examines direct and indirect instruments for quality disclosure. And the third stream of literature empirically investigates issues related to vertical quality disclosure.

3.2.1 Firm Incentive to Disclose Product Quality

(a) Monopoly Market Grossman (1981) considers a model where a monopolistic seller knows the true quality of its product and can claim either the exact quality level (full disclosure), or a range of its product quality that can be verified ex post at negligible cost. Examples of quality statements verifiable ex post at negligible cost include "*the seller is selling boxes of oranges... states that there are ten oranges in a box*" and "*the seller states that the diamond weighs one ounce*." In this market, if the seller claims a quality range, consumers with rational expectations will assume that the product's true quality is the lowest of the given range. The monopolist, anticipating this, makes a full disclosure in equilibrium, that is, to disclose the exact quality level of its product. Grossman and Hart (1980) and Milgrom (1981) obtain similar results in the context of takeover bids.

Jovanovic (1982) considers a different setup where a seller does not know its product's true quality, but observes a private signal drawn from a distribution with its mean being the true quality. For example, the true quality of a used car can be the average quality of all of the car's components, and the private signal the seller observes is how these components function on his particular driving habit. The seller can withhold the private signal, or disclose the signal truthfully and credibly to the buyer after incurring a cost. In this market, disclosure can happen only if the cost of doing so is not too high, and the seller that observes a higher quality signal has a stronger incentive to disclose.

(b) Competitive Market Guo and Zhao (2009) considers a duopoly market where the seller knows the true quality of its product, but not the true quality of its rival's product. A higher disclosure cost shifts the threshold for quality disclosure toward the high end, and consequently elates consumers' expectation on a product's quality when no vertical attribute information of the product are disclosed. This effect suppresses the incentive of a high-quality seller to disclose quality. Moreover, competition reduces sellers' expected benefits from quality disclosure, further inhibiting their disclosing incentive. While Guo and Zhao (2009) assumes that each firm does not know its rival's product quality, Board (2009) obtains similar results that competition inhibits firms' incentive to disclose quality information under the assumption that firms know each other's product quality.

Kuksov and Lin (2010) consider a duopoly market where consumers ex ante are not only uncertain about the quality of competitive products, but uncertain about their quality preferences, or how much they are willing to pay for a quality premium. While the former type of uncertainty is specific to the product, the latter type is specific to the consumer. Each firm endogenously chooses the quality level of its product. A firm then decides whether to disclose the quality of its own product, and whether to provide information that helps consumers find their quality preferences that apply to both products. The study shows that in equilibrium, the two firms differentiate in their product quality levels as well as the type of information they provide. In particular, the high-quality firm is more likely to disclose its product's quality and the low-quality firm is more likely to provide information that helps consumers to find their quality preferences. Extending Kuksov and Lin (2010), Lin and Pazgal (2016) considers the case when exogenously determined product quality enter the consumer market sequentially. The study shows that the first entrant always discloses its product quality. A late entrant with a superior product may choose not to inform consumers of its better quality, but instead provide information to help consumers to find their quality preferences. On the other hand, a late entrant offering an inferior product may wish to admit so.

(c) Distribution Channel Guo (2009) considers a distribution channel where a manufacturer sells its product through a retailer under the wholesale price contract. Both channel members know the true quality of the product. The manufacturer can disclose the quality directly to end customers (e.g., through national advertising), or leave to the retailer to decide whether to disclose (e.g., through free samples and returns, sales assistance, in-store media). The study shows that more information is revealed under retailer disclosure than under manufacturer disclosure. This is because the manufacturer can, through wholesale price cuts, partially absorb the retailer's effective disclosure cost, which elevates the retailer's disclosing incentive.

Guan and Chen (2017) considers a similar channel structure and examines the case when the monopolistic manufacturer has private information about its product quality, but has less information about the consumer's quality preferences than the retailer has. The study shows that the manufacturer's decisions to disclose information on its product quality and to acquire information on consumers' quality preferences interact, and together influence the retailer's rational inference about the product quality level and channel relationship.

3.2.2 Firm Instruments to Disclose Product Quality

Studies that investigate firm incentive to disclose quality commonly assume that such information can be fully and truthfully communicated to end consumers. In contrast, studies on firm's instruments to disclose quality typically consider a more realistic setting where quality cannot be fully disclosed through direct communication. In this case, various signaling mechanisms such as pricing, uninformative advertising, warranty, and money-back guarantee can be leveraged to help consumers differentiate a high-quality product from the low-quality one.

Starting with the pioneering work of Nelson (1974), a large body of literature has examined how price and conspicuous advertising can help firms signal product quality to imperfectly informed consumers (e.g., Kihlstrom and Riordan 1984; Milgrom and Roberts 1986; Bagwell and Riordan 1991; Linnemer 2002). Warranty has also been recognized as an effective quality signal since it is very costly for low-quality firms to mimic the terms offered by high-quality firms (e.g., Spence 1977; Cooper and Ross 1985; Gal-Or 1989). Besides serving as a quality signal, warranty can also be used to sort consumers based on their heterogeneous risk preferences (Kubo 1986), provide protection against product failures (Heal 1977; Courville and Hausman 1979; Menezes and Currim 1992; Padmanabhan and Rao 1993), or incentivize the seller to improve product quality (Prosser 1943). These four functions of warranty are summarized in Emons (1989) and later empirically tested in Chu and Chintagunta (2011). A related literature stream examines design and profitability of extended service contracts offered by retailers and/or manufacturers beyond the basic warranty (e.g., Lutz and Padmanabhan 1995; Padmanabhan 1995; Chen et al. 2009; Jiang and Zhang 2011).

Moorthy and Srinivasan (1995) considers a market where consumers are uncertain about whether the seller is of high or low quality, and demonstrates that a high-quality seller can effectively use money-back guarantee as a quality signal supplemental to other quality signals such as price or uninformative advertising. The superiority of money-back guarantee over other quality signals resides in its high cost to the low-quality seller, which inhibits the low-quality seller from mimicking the high-quality seller's strategy.

Bhardwaj et al. (2008) consider a context where a firm discloses its product's vertical attributes to end consumers through a salesperson, but the limited bandwidth in sales communication only allows the salesperson to transmit a subset of all vertical attributes. The focal research question concerns the format of sales communication: should the firm choose the attributes to show to consumers (i.e., seller-initiated communication) or should it let consumers choose which attributes they want to see (i.e., buyer-initiated communication)? While seller-initiated communication grants the firm more control over quality disclosure, buyer-initiated communication credibly signals that the firm has nothing to hide, or that the product has a high quality.

Mayzlin and Shin (2011) examines a context where a firm discloses its product's vertical attributes to end consumers through advertising, which is nonetheless ineffective at disclosing all vertical attributes. Consumers may conduct a costly search for the true product quality, and the extent of search is endogenously determined by the content of advertising. The focal research question concerns the format of advertising: Should the firm use informative advertising that emphasizes product vertical attributes or uninformative advertising that makes vague claims (or no claims) about product vertical attributes? Compared to informative advertising, uninformative advertising motivates consumers to search for the true quality themselves, thus leading to a more accurate quality valuation ex ante. A high-quality firm, thus, has a stronger incentive than a low-quality firm to invite consumer search through uninformative advertising. As such, uninformative advertising, when coupled with consumer search, can be used by a monopolistic firm to signal its high quality.

3.2.3 Empirical Research on Firm Quality Disclosing Strategies

Early empirical research on firm quality disclosing strategies focuses on testing the signaling effect of uninformative advertising. Tellis and Fornell (1988) uses PIMS (Profit Impact of Market Strategies) dataset to examine how advertising spending affects product quality measured with the difference between the sales percentage of products superior to those of the rivals and products inferior to. Caves and Greene (1996) and Moorthy and Zhao (2000) construct brand quality measurement using Consumer Reports survey and examine how brands' advertising spending affects their quality scores. Similar investigations were conducted by Thomas et al. (1998) by using data from the US automobile industry and Horstmann and MacDonald (2003) by using data from compact disc players industry. Some researchers use experiments to investigate how manipulated conditions of advertising spending affects participants' perceived product quality (e.g., Kirmani and Wright 1989; Kirmani 1990; Moorthy and Hawkins 2005). Animesh et al. (2010) test the advertising-quality relationship by using the online paid search advertising data (e.g., Animesh et al. 2010).

Recent empirical research has focused on examining quality uncertainty in online markets. A product sold online can be viewed as a bundle of the core product (i.e., the product's physical attributes) and the extended product (i.e., service provided by the online seller), and quality uncertainty can arise from either. While theoretical studies typically treat the bundled product as a whole, empirical research has tried to distinguish the two sources of quality uncertainty. In online markets, consumer uncertainties about the vertical attributes of the core product often come from the difficulty for the seller to describe the product's physical attributes. This uncertainty is more severe for used goods, whose wearing conditions can vary significantly. Using data from the motor vehicle industry and the computer industry, Heiman and Muller (1996) shows that the number and the length of product demonstration affect product acceptance by mitigating consumers' perceived product quality uncertainty.

Consumer quality uncertainties about the extended product concern two roles the seller fulfills: providing product information and delivering products. Prior to purchase, consumers are likely to be uncertain about the vertical quality of the seller, particularly for unfamiliar ones, such as whether the seller would intentionally misrepresent the product to increase sales or whether the seller will deliver the wrong product due to negligence or incompetence. In online markets, such uncertainty is exacerbated, because buyers are unable to infer seller characteristics by observing social cues from personal interactions or body language (Gefen et al. 2003). On the other hand, the online market provides opportunities to mitigate seller quality uncertainty through information systems such as online feedback ratings (e.g., Ba and Pavlou 2002; Dellarocas 2003), user-generated textual comments (e.g., Ghose and Ipeirotis 2011; Pavlou and Dimoka 2006), third-party escrows (e.g., Pavlou and Gefen 2004), and product diagnostic tools (Jiang 2007).

Some researchers compare the effect of quality uncertainty in the core product and the extended product. Ghose (2009) examines used goods trading data in multiple product categories and shows that both seller-related (e.g., seller characteristics) and product-related (e.g., condition of used cars) quality uncertainty lead to adverse selection, which does not completely disappear even with mechanisms such as seller reputation feedback and product quality disclosure. Dimoka et al. (2012) examine auction data on used cars and shows that, compared to seller-related quality uncertainty, product-related quality uncertainty has more adverse effect on price premium. The study also shows that both types of uncertainties can be reduced by

IT-enabled solutions such as diagnostic product descriptions and third-party product assurances.

3.3 Firm Strategy to Disclose Product Horizontal Attributes

Lewis and Sappington (1994) is one of the earliest works that recognize the heterogeneous consumer valuations for product horizontal attributes and investigate firm's incentive to help consumers to learn their idiosyncratic fit with a product. Research on firm strategies to disclose product's horizontal attributes has developed rapidly in recent years accompanying the growing popularity of online third-party reviews. It is generally believed that it has effectively alleviated consumers' uncertainty about quality. The efficacy of online reviews in resolving consumers' quality uncertainty resides in consumers' homogeneous preferences for product quality, which allows different consumers to measure and compare quality on a scale universally agreed upon. A consumer can conveniently infer a product's quality level from its "valence score" or average review rating. Moreover, the "valence score," as a numerical quality indicator, is easy to understand and process, which encourages consumer usage of the score. In contrast, consumers' heterogeneous preferences for a product's horizontal attributes make it difficult for an individual to infer her own fit from others' perceived fit. As such, consumer fit uncertainty persists despite the presence of third-party reviews. Moreover, compared to product vertical attributes such as material and craftsmanship, product horizontal attributes such as color and style are often hard to describe or quantify, making such information hard to communicate. Commonly, a personal inspection is necessary to find out the consumer's true fit with a product.

Below we review three main streams of this literature. The first stream examines a firm's incentive to disclose fit-revealing information in various market structures. The second stream of literature investigates direct and indirect instruments that help disclose product horizontal attributes. Finally, the third stream of literature empirically explores issues related to firms' fit-revealing strategies.

3.3.1 Firm Incentive to Disclose Product Horizontal Attributes

(a) Monopolistic Market Lewis and Sappington (1994) considers a model where consumers are ex ante identical in their expected valuations on a monopolistic firm's product, and knowledge of product horizontal attributes leads to differentiation in consumer product evaluations. Disclosing product horizontal attributes creates "targeting" opportunities for the firm, allowing it to sell to a segment of the market at a price higher than the average valuation. Nonetheless, the firm has to abandon the segment of market with below-than-average valuations. This tension between pursuing a higher margin or a larger demand is at the core of the firm's incentive to disclose product horizontal attributes. The firm's optimal strategy is either not to provide any information, which ensures the full advantage of the demand-oriented strategy, or to provide full information, which ensures the full advantage of the margin-oriented strategy. Johnson and Myatt (2006) models differentiation in consumers' product valuations as the result of a rotation of the demand curve, which can be induced by marketing mix variables such as advertising and product design. Echoing Lewis and Sappington (1994), the study shows that a monopolistic firm obtains maximized profit when differentiation in consumers' product valuations is either very high to facilitate an effective margin-oriented niche strategy, or very low to facilitate an effective demand-oriented mass-market strategy.

Bar-Issac et al. (2010) deviate from Lewis and Sappington (1994) and Johnson and Myatt (2006) by considering a market where consumers are ex ante heterogeneous in their expected valuations of a monopolistic firm's product, with one consumer segment exhibiting consistently higher willingness to pay for the same fit level, than the other consumer segment. Prior to purchase, consumers find their true fit with the product through a costly inspection, and the firm can manipulate the inspection cost to induce inspection by none, some, or all consumers. The study shows that an intermediate information disclosure strategy can be optimal. It would induce only consumers with low willingness to pay to inspect, but not those with high willingness to pay. That is, the intermediate information disclosure strategy is used as a non-price means to discriminate between different consumer types. Bhargava and Chen (2012) considers a similar setup where ex ante the smaller consumer segment has consistently higher willingness to pay for the same product fit level than the larger segment. The firm can disclose information on its product's horizontal attributes, which allow all consumers to find their fit with the product, or withhold such information, which will leave all consumers' fit uncertain. The study shows that full disclosure is profitable when consumer heterogeneity in willingness to pay ex ante is moderate, but non-disclosure is profitable when such heterogeneity is very low or very high. Lahiri and Dey (2018) considers versioning as a way to disclose product fit information in the context of information goods. The study shows that if a fraction of consumers is fully aware of their true valuations ex ante, information provision through versioning can be more profitable than keeping consumers in the dark.

Chen and Xie (2008) considers a monopolistic firm's strategy to offer either partial or full information on product horizontal attributes when consumers never buy with null fit information and can acquire additional attribute information from third-party reviews. The study assumes the existence of a segment of novice consumers who can only process information provided by third-party reviews but not information provided by the seller, and shows that the availability of third-party reviews may reverse the firm's optimal disclosure strategy. In particular, without third-party reviews, a firm with a low production cost enjoys a high margin and has incentive to pursue a demand-oriented strategy by disclosing only partial information. Third-party reviews, however, forces the firm to switch to a margin-oriented strategy and disclose full information. On the other hand, without third-party reviews, a firm with a high product cost pursues a margin-oriented strategy by disclosing full information. Third-party reviews, however, help inform novice customers, which allows the firm to switch to a demand-oriented strategy and disclose only partial information.

Sun (2011) models a product as containing both vertical and horizontal attributes and examines how a monopolistic firm's fit disclosure strategy is moderated by its product quality level. The study shows that when the product quality level is high, the firm enjoys a high margin and optimally pursues a demand-oriented strategy through non-disclosure. On the other hand, when the product quality level is low, the firm collects a low margin and optimally pursues a margin-oriented strategy through full disclosure. Anderson and Renault (2013) obtains similar results in a market where consumers are uncertain about a product's quality and price, in addition to being uncertain about the product's horizontal attributes.

A set of research studies examines the sustainability of fit disclosure as a perfect Bayesian equilibrium outcome in an incomplete information game. Anderson and Renault (2006) considers a model where consumers are uncertain about not only the horizontal attributes, but the price of a monopolistic seller's product, and shows the firm should never fully disclose the product's horizontal attributes without disclosing its price. Koessler and Renault (2012) considers a general modeling framework where the monopolistic firm has perfect and private information about the product's attributes, which can be vertical and horizontal, and the single buyer has perfect and private information about her own taste. Their study shows that full disclosure is always an equilibrium when product and consumer types are independently distributed. Çelik (2014) characterizes conditions under which a monopoly seller fully reveals the location of its product on the consumer preference spectrum when an individual consumer's preference is privately known only to herself.

(b) Competitive Market In a competitive market, disclosing product horizontal attributes allows a firm to create product differentiation from its competitor, which alleviates price competition. Firms that occupy different competitive status demonstrate different incentives to disclose product horizontal information. Anderson and Renault (2009) considers two competing firms that offer two products differentiated in both vertical and horizontal attributes. Consumers have full knowledge about the two products' quality and price, but are uncertain about their fit with either product. A firm can advertise its own products' horizontal attributes to end customers, and can also disclose its rival's horizontal attributes through comparative advertising. A key finding is that when the quality difference between the two products is large, the high-quality firm never discloses any information; the low-quality firm does not disclose either, if comparative advertising is banned, but otherwise will disclose the fit information about its own product as well as that of the rival product.

Gu and Xie (2013) also considers a setting where consumers face the choice between two competing products with differentiated vertical quality, as well as differentiated horizontal attributes. A consumer's perceived fits with the two products are independent. A firm can help resolve consumers' fit uncertainty regarding its own product through costly marketing activities (e.g., offering free samples or proving free trials), but such activities do not help resolve consumer fit uncertainty regarding the other product. A key finding is that the firm offering the high-quality product implements fit-revealing activities with greater intensity than its low-quality rival, if both products' qualities are sufficiently high and their quality difference is small. This result poses an interesting contrast to the finding in the monopolistic market where the low-quality firm has a stronger incentive to disclose fit. Jing (2016) considers a model where consumers have knowledge about the quality as well as price of two competing products, but can only find their fit with the product through a costly inspection. Each firm determines the level of customer learning investments (CLI), and a higher investment induces more consumers to inspect its product prior to purchase. In a fully covered market, the firm that makes a larger CLI enjoys a higher demand as well as a higher price. Echoing Gu and Xie (2013), the study shows in equilibrium the firm with a greater relative production efficiency invests more in CLI to facilitate customer fit search.

Boleslavsky et al. (2017) model competition between an innovative firm that offers a new product with unknown horizontal attributes and an established firm that offers a product for which the consumers have full information. The result shows that the innovative firm benefits from fully disclosing horizontal attributes of its product through demonstration to resolve consumer fit uncertainty if pricing policy is flexible, but partial disclosure is optimal if the price decision has to be made prior to the demonstration decision.

(c) Distribution Channel Hao and Tan (2017) considers a vertical channel composed of a supplier and a retailer and demonstrates that the format of channel contract affects channel members' fit-disclosing incentive. Under the agency pricing contract, the revenue sharing mechanism leads the supplier to benefit from more fit disclosure but the retailer to suffer from it. On the other hand, under the wholesale pricing contract, potential misalignment of channel members' interests regarding fit disclosure disappears, if the demand is linear. If the demand is log-concave and derived from common valuation distributions like normal or logistic distributions, however, misalignment reappears, with the retailer benefiting and the supplier suffering from more fit disclosure.

3.3.2 Firm Instruments to Disclose Product Horizontal Attributes

Studies that investigate firm incentive to disclose product horizontal attributes typically focus on disclosure instruments implemented by manufactures, such as advertising and sampling. In business reality, retailers that carry an array of horizontally differentiated products can leverage various instruments to manipulate consumers' fit knowledge ex ante. As such, the literature on fit disclosure instruments has merged into the retailing literature.

(a) Disclosing Product Horizontal Attributes Through In-Store Sales Communication Wernerfelt (1994) considers a model where a seller offers two products with differentiated horizontal attributes to a fit-uncertain buyer and shows that a knowledgeable salesperson can effectively and truthfully match the customer with the product that suits her needs through a "dialogue" or an interactive communication with the buyer. Ofek et al. (2011) find that a monopolistic retailer that sells through both online and offline channels offers less sales assistance in its offline store to match consumers with suitable products, than in the case it operates only an offline channel, because the existence of the online store reduces consumer traffic to the physical store. In a competitive market, however, a dual channel retailer may offer more in-store sales assistance than a pure offline retailer to combat competition. Gu and Liu (2018) extends Wernerfelt (1994) and models a retailer that sells through a salesperson to end consumers two horizontally differentiated products offered by two competing manufacturers. The study shows that the retailer has incentive to demotivate its salesperson from advising consumers, if the effectiveness of such sales advising is too high or too low in helping consumers learn their true fit with products.

(b) Disclosing Product Horizontal Attributes Through Manipulating Consumer In-Store Fit Search Gu and Liu (2013) considers a retailer that sells horizontally differentiated products offered by competing upstream manufacturers and examines the retailer's optimal in-store display decision: whether to display competing products in the same location so that consumers can inspect multiple choice alternatives all at once, or display them in distant locations so that consumers have to inspect one product first and then decide whether to incur a travel cost to inspect another product. The study finds that the former display format is more profitable for product categories with overall high fit probability (e.g., home appliances), whereas the latter display format is more profitable for product categories with overall low fit probability (e.g., apparel).

Branco et al. (2016) model how consumers evaluate their fit with a product offered by a monopolistic seller through a sequential search on the products' multiple attributes. Consumers check one attribute at a time after incurring a search cost, learn about the attribute on which the seller provides information, and then decides whether to check more attributes or make a choice decision without further search. The seller decides on which attributes to provide information, but does not know the order of consumers' attribute searches. The study shows that the seller's optimal strategy is to provide information on an intermediate number of attributes. Providing too much information makes the search less informative, and providing too little information makes consumers believe there is less positive information about the product; both strategies will deter consumer search and lower seller profit.

Gu and Tayi (2017) considers a retailer that operates both an online and an offline store and aims at maximizing the omni-channel profit. The retailer carries horizontally differentiated products, and decides whether to sell the products through both channels, or through the online channel only. Through an in-store inspection, consumers learn about their fit with the product offered at the store offline, and make inferences about their fit with the product offered through the online store only. The study shows it can be profitable for the retailer to offer the full assortment through the online store, but only partial assortment through the offline store, and that the retailer benefits more from selling higher-quality or higher-demand products through the online store only.

(c) Mitigating Fit Uncertainty Through Product Return Policy When consumers have difficulty evaluating their fit with a product prior to purchase, retailers often use product return policies to mitigate fit uncertainty. Davis et al. (1995) demonstrate that money-back guarantee can be used to reduce the perceived risk of fit-uncertain consumers and enhance their willingness to pay. As opposed to the warranty, money-back guarantee allows consumers to return a product for a full refund, even if the product has no quality defect. Shulman et al. (2009) indicate that a higher restocking fee will reduce fit-uncertain consumers' purchase intention and a firm in devising the optimal level of restocking fee should consider this impact in addition to the recouping cost associated with product returns. Heiman et al. (2001) show that money-back guarantee and demonstrations can be complements or substitutes in revealing product horizontal attributes.

Gu and Tayi (2015) examines a pure online seller's optimal return policy in a context where consumers, uncertain about their fit with a product ex ante if finding a misfit after purchase, can choose between making a costly return or self-mending to assure a proper fit. The study shows that an online retailer can benefit from tightening the return policy and maintaining a reasonable return cost for consumers, because such a policy motivates consumers to self-mend a misfit product and consequently eases the firm from the burden of handling returns. Moreover, accompanying the tighter return policy, the firm charges a lower price, which can enhance consumer surplus.

Shulman et al. (2015) consider a model where a consumer's perceived ex post utility of a product is reference-dependent on a consumer's ex ante expectation and makes product return decisions based on the perceived ex post utility, rather than the true product value. In this case, ex ante fit uncertainty leads to consumers with true high product valuations to form low ex ante expectations, which elevates their perceived ex post utilities and reduces their return tendency. On the other hand, pre-purchase fit disclosure increases these consumers' ex ante expectation and may increase product returns. These theoretical insights are further supported by controlled behavioral experiments as well as econometric analysis of archival data.

(d) Revelation of Product Horizontal Attributes Through Third-Party Reviews While the valence score of third-party reviews reveals product vertical quality, the text content of reviews often provides useful information about a product's horizontal attributes. As such, research that investigates the impact of third-party reviews in revealing product horizontal attributes typically also considers the impact of such reviews in revealing the product quality. Kwark et al. (2014) consider a distribution channel where two competing manufacturers sell through a common retailer, and shows that reviews that disclose products' horizontal attributes by enhancing product differentiation soften manufacturer competition and hurt the retailer. On the other hand, reviews that disclose products' qualities by reducing product differentiation, intensify manufacturer competition and benefit the retailer. Extending this work, Kwark et al. (2017) further show that a retailer can benefit from third-party reviews on product horizontal attributes by adopting the commission scheme, and can benefit from third-party reviews on product vertical attributes by adopting the wholesale pricing scheme, rather than the commission scheme.

Jiang and Guo (2015) examines a monopolistic firm's optimal design of a review system. Hosting a review system facilitates disclosure of product vertical attributes, which is optimal when the product quality is sufficiently high. Moreover, offering "granular reports" that review specific product attributes facilitates disclosure of product horizontal attributes, which is profitable when misfit significantly reduces in a consumer's willingness to pay. Loginova and Mantovani (2015) examines competing firms' incentive to join an online review aggregator's website (e.g., tripadvisor.com), which expands consumer demand by reducing fit uncertainty but intensifies price competition.

Li (2017) considers a monopolistic online seller that offers two products differentiated in both vertical and horizontal attributes. The menu page shows the list of products, and a consumer has to click a link to go to an individual product's page, where detailed third-party reviews are displayed and consumers can learn both vertical and horizontal attributes of the product. The focal question is whether the online seller should show the products' aggregated valence score in the menu page to disclose product quality before consumers check individual products. The result shows that not showing the quality score can be optimal when consumers have highly heterogeneous preferences for the low-quality product's horizontal attributes.

3.3.3 Empirical Research on Firm Strategy to Mitigate Consumer Fit Uncertainty

Empirical research has generally supported the theoretical predictions that consumer fit uncertainty adversely affects firm profit and that firms' fit disclosing strategies help alleviate such a problem. A stream of research examines how consumers' fit uncertainty affects their purchase intentions. In an empirical study of hundreds of product categories, Kim and Krishnan (2015) show that consumer fit uncertainty inhibits online purchase of higher-priced products, and that accumulated online shopping experience will encourage consumers to purchase more of the cheaper products. The study also shows that online sellers can mitigate consumer fit uncertainty using technology such as digitized video commercials. Using a series of randomized field experiments, Gallino and Moreno (2018) shows that offering virtual fit information in online apparel retail increases conversion, basket sizes, average price of purchased products, and revisits to the site, and also reduces fulfillment costs related to returns and home try-ons. These benefits are more pronounced for products that are more expensive or available in more sizes. Moreover, virtual fitting technology increases customer engagement and loyalty, resulting in a spillover effect for products beyond those available for virtual fitting. Ball et al. (2018) use quasi-experimental methods to assess the effect of opening showrooms for a business that operates as an online-only business. The result shows that opening showrooms has a positive impact on the overall demand, and results in spillovers between the online and the offline channels. This finding is consistent with the notion that offline channels are more effective at providing product fit information than online channels (e.g., Lal and Sarvary 1999).

Another stream of empirical research examines the impact of consumer fit uncertainty on product returns. Consumers who buy a product under ex ante fit uncertainty are likely to make a return if finding a misfit ex post, which causes a return handling cost for the firm. Using consumer survey data from eBay, Hong and Pavlou (2014) shows that product fit uncertainty has more adverse effect on product returns than product quality uncertainty. The study also shows online product forums are more helpful in alleviating product fit uncertainty, whereas website media on product pages are more effective in mitigating product quality uncertainty, and both activities reduce product returns. Using a transaction level dataset, Sahoo et al. (2017) demonstrate that the availability of product reviews online leads to higher sales and fewer product returns.

Some empirical studies find that a firm's effort to provide information on product horizontal attributes can have a negative impact on that firm's profit. Jain et al. (1995) show that sampling might have a negative impact on the firm's profit, if the firm cannot control the type and number of consumers who receive samples. Bawa and Shoemaker (2004) points out that offering samples may cannibalize the sales of products with low repurchase rate. Shulman et al. (2015) demonstrate, with both theory and field experiments, that fit-revealing information provided before the purchase can actually increase decision reversals. Arora et al. (2017) show that the practice of offering free versions of paid apps is negatively associated with adoption speed of apps, and the association is stronger for hedonic apps and in the later life stages of paid apps. This result is consistent with the literature that offering free versions of information goods is suboptimal (Bhargava and Choudhary 2001, 2008; Jones and Mendelson 2011), especially when consumer uncertainty is high and price is low (Lahiri and Dey 2013).

3.4 Firm Strategy to Create Consumer Valuation Uncertainty

While the information disclosure literature focuses on incentive and consequences of firm activities that help resolve consumer valuation uncertainty, related literature on advance/opaque/probabilistic selling examines incentives and consequences of firm's activities that create consumer valuation uncertainty in product transactions. These two streams of literature can be viewed as two sides of the same coin that address the same question: How much product information the firm would like its consumers to have prior to purchase?

Shugan and Xie (2000) and Xie and Shugan (2001) study advance selling strategies, under which a firm sells its product in advance of actual time of consumption. When consumers' utilities from a product or service (such as air travel, hotels, and cruises) depend on their idiosyncratic consumption state such as mood, health, and personal work schedule, the separation between the (advance) time of purchase and the time of consumption results in consumer valuation uncertainty at the purchase time. Because of such uncertainty, consumers make their purchase decisions in the advance period based on their expected valuations, which tend to be more homogeneous. In contrast, their valuations become more heterogeneous at a later time period after their idiosyncratic consumption states are realized, which makes it harder for the firm that lacks capability of conducting direct price discrimination to extract consumer surplus. The key insight that more information leads to heterogeneous consumer preferences is consistent with Lewis and Sappington (1994) and the literature on firm's strategy to disclose product horizontal attributes ever since. In a more recent study, Yu et al. (2015) examine the effects of interdependent consumer valuations and seller's capacity on the firm's advance selling decisions.

While advance selling is essentially a pricing strategy that takes advantage of consumers' valuation uncertainty in the advance time, opaque selling and probabilistic selling add uncertainty to consumer valuations by withholding product information. Under opaque selling, one or more product attributes are deliberately hidden from the buyer until payment has been made. Under probabilistic selling, a firm creates a "virtual" product or service, i.e., a probabilistic good that offers consumers a probability of getting any one of a set of multiple distinct items (Fay and Xie 2008, 2010, 2015; Fay et al. 2015). While the potential choice set is well defined under probabilistic selling (five identical shirts that only differ in color), it may not be as clear under opaque selling (a 3-star hotel in San Francisco downtown district). Theoretically, these two concepts are similar and sometimes indistinguishable in the literature (Huang and Yu 2014); we will treat these two terms interchangeable in our discussion.

Under opaque selling, the undisclosed attributes introduce an element of "damaged goods" to the opaque product, which allows the firm to segment the market based on consumers' tolerance level for uncertainty. In many markets, consumers differ in their tolerance level towards product uncertainty. For example, when shopping for hotel rooms, business travelers are likely to have specific location requirements and thus favor hotels with known addresses (i.e., the transparent product) despite its high price, whereas leisure travelers are likely to be more flexible in location choices and thus favor hotels with addresses undisclosed (i.e., the opaque product) as long as the price is sufficiently low. Offering a hotel room with hidden location thus allows the intermediary to price discriminate between buyers with little tolerance for uncertainty and hence low price sensitivity (e.g., business travelers) and those with high tolerance of uncertainty and hence high price sensitivity (e.g., leisure travelers). This strategy was first introduced by online travel agencies (e.g., Priceline and Hotwire) for selling leftover capacity for airlines and hotels and has gained popularity among consumers and service providers (Post and Spann 2012).

Jiang (2007) and Fay (2008) examine the incentive for a monopoly firm to sell opaque or probabilistic goods. Shapiro and Shi (2008) shows that the ability to price discriminate allows a firm to profit from offering an opaque product, even in a competitive market where the opaque feature virtually erases product differentiation and thus intensifies competition. Fay (2008) considers a setting where two firms use a common intermediary to sell the opaque product and find that an opaque product magnifies price competition, if there is little brand-loyalty in an industry, and curtails price competition, if there is significant brand-loyalty in the industry. Another motivation for firms to adopt opaque selling is that it helps reduce supply–demand mismatches, especially in industries with little flexibility in supply, e.g., airline, hotel, and car rental industries. Gallego and Phillips (2004) examines the optimal design of a probabilistic product to balance the benefit of increasing overall demand and enabling better capacity utilization at the cost of potentially cannibalizing high-fare demand for specific products. Jerath et al. (2010) consider an opaque intermediary who sells last-minute capacity for competing providers facing stochastic demand for the aggregate market. Chen et al. (2014) compare posted-price and Name-Your-Own-Price (NYOP) as two pricing mechanisms to dispose of excess inventory in an opaque distribution channel. In this stream of research, the leftover inventory in the opaque channel is subject to the demand shock in the direct channel, which adds another layer of availability uncertainty to the opaque product. Huang et al. (2017) highlight the impact of inventory and time on equilibrium prices, expected profit, and channel strategy in the presence of an opaque channel. Cai et al. (2013) consider a retailer's strategy of mixing products from competing suppliers to generate a probabilistic good and shows that introducing the probabilistic good is beneficial for the channel members. Additional benefits of opaque selling include softening price competition (Shapiro and Shi 2008). Recently, Huang and Yu (2014) shows that opaque selling may soften price competition and increase the industry profits as a result of consumer bounded rationality, providing a behavioral rational for opaque selling.

Whereas most of the opaque selling literature focuses on withholding horizontal attributes, a growing body of literature examines opaque selling with vertically differentiated products. Biyalogorsky et al. (2005) consider the airline industry where some firms offer tickets that can be upgraded to higher-class, depending on the availability of higher-class products at the service delivery time, and characterize conditions under which such strategy is profitable. Zhang et al. (2015) show that probabilistic selling in quality-differentiated markets can be profitable by disposing excess capacity, even when quality levels are endogenously determined. Halbheer et al. (2018) show that deliberately randomizing service quality can benefit the provider and society because heterogeneity in customer damages from service failures allows the provider to profit from selling damage prevention services or offering compensation to high-damage customers.

3.5 Future Research Directions

In recent years, researchers' focal interest has shifted from issues related to quality disclosure to issues related to fit revelation. This shift is partially driven by the prevalence of third-party reviews online. Nevertheless, quality uncertainty for services remains a challenge. Moreover, compared to quality uncertainty, consumer fit uncertainty appears to be a much richer construct, determined not only by a product's horizontal attributes known to the firm, but also consumers' idiosyncratic tastes privately known to themselves. Such complexities of consumer fit uncertainty provides abundant research opportunities.

3.5.1 Service Quality Uncertainty

Consumers' quality uncertainty about a standardized product or service can be effectively mitigated when they learn about the product/service's vertical attributes from the seller or other consumers. Interestingly, quality uncertainty remains a concern for professional services such as car repair, real estate sales, and health care. Consumers' needs for professional services are often highly idiosyncratic, causing high variation in both service procedure and outcome. For example, among patients with the same disease, some may have pre-existing conditions that interfere with the treatment. In addition, the completion of a professional service typically involves not only the service provider, but the customer and sometimes third-party players who could potentially bring more shocks to the service quality. For example, not every patient follows the doctor's advice closely, and a dentist's service quality may depend on which nurse is assisting. Moreover, the complexity of professional services often makes it difficult for consumers to evaluate a service's true quality and consumers may form false quality perception based on cosmetic aspects of the service process. For example, a consumer may believe a car repair service is of high quality because of a free car wash, or believe a dentist provides high-quality service because of her pleasant attitude. In a recent work, Liu et al. (2017) show that it is inappropriate to interpret reviews for professional services the same way as reviews for commodity. Moreover, deadweight loss in social welfare can occur if the service provider is allowed to select customers to ensure favorable reviews. Future research can investigate mechanisms that facilitate disclosure of product quality in the professional service context.

3.5.2 Uncertainty in Consumer Fit Preferences

While consumers always prefer a higher product vertical quality, their preferences for a product's horizontal attributes may be context-dependent and time-variant.

For example, a consumer's preference for fashion products may well depend on her mood. Even if the firm discloses its product's horizontal attributes prior to purchase, the consumer may still find a misfit after purchase due to changes in her mood. This instability of consumer fit preference is more pronounced for hedonic products than for utilitarian products. Moreover, consumers may be unaware of their true preferences for product horizontal attributes at the point of purchase. For instance, when booking a vacation package to Amazon rainforest, a consumer knows that a package that offers more events is of better value. Having never visited a rainforest, however, the consumer may have trouble evaluating which events she is likely to enjoy. Reading descriptions and reviews beforehand cannot fully eliminate pleasant surprises or disappointments. This uncertainty in consumers' preferences for product horizontal attributes is more severe for experience products than for search goods. In these cases, disclosing product horizontal attributes cannot fully resolve consumer fit uncertainty. Firms may use interactive communication tools to learn about consumers and predict their future preferences. Firms may also adopt new technologies such as virtual reality to help consumers learn about their own tastes. Gu and Tayi (2015) considers a firm's strategy to offer a product that can be customized after purchase to ensure a proper fit. Alptekinoğlu and Ramachandran (2018) examines a dynamic model where consumer preferences may change across time periods and identifies conditions under which offering a consumer-customizable product is more valuable than offering a portfolio of standard products. We encourage future research to explore innovative marketing tools that can be used to alleviate consumer uncertainty on fit preferences.

3.5.3 Interactions Between a Firm's Fit Disclosure Strategy and Other Strategic Decisions

Consumers' knowledge about product attributes changes their information set and may thus affect how they respond to firm strategies on distribution channel, advertising, product line design, etc. Deng et al. (2017) show that a firm's strategy to disclose product horizontal attributes can be complementary or substitutable with advertising and can also be influenced by the firm's quality provision decision. We encourage future research that explores interactions between information revelation and other marketing mix strategies.

The wide adoption of online social media platforms encourage consumers to seek information from their social connections in evaluating a product's fit. For instance, when shopping for vacation packages, consumers are likely to pay more attention to recommendations from their friends than anonymous reviews on Tripadvisor.com. As such, the structural properties of consumer social network can impact consumers' ex ante product valuations. Fainmesser et al. (2018) consider a two-period model where first-period consumers learn about product fit through advertising and second-period consumers learn about product fit through first-period buyers' product reviews. The study shows that the firm has a greater incentive to advertise in the first period when the social connections between the first and the second-period consumers exhibit greater homophily. We encourage future research on how a firm can leverage the power of social media to assist consumers' fit search.

References

Akerlof G (1970) The market for lemons: quality uncertainty and the market mechanism. Q J Econ 84(3):488–500
Alptekinoğlu A, Ramachandran K (2018) Flexible products for dynamic preferences. Prod Oper Manage 28(6):1558–1576. https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.12990
Anderson SP, Renault R (2006) Advertising content. Am Econ Rev 96(1):93–113
Anderson SP, Renault R (2009) Comparative advertising: disclosing horizontal match information. RAND J Econ 40:558–581
Anderson SP, Renault R (2013) The advertising mix for a search good. Manage Sci 59(1):69–83
Animesh A, Ramachandran V, Viswanathan S (2010) Quality uncertainty and the performance of online sponsored search markets: an empirical investigation. Inf Syst Res 21(1):190–201
Arora S, Hofstede F, Mahajan F (2017) The implications of offering free versions for the performance of paid mobile apps. J Market 81(6):62–78
Ba S, Pavlou PA (2002) Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. MIS Q 26(3):243–268
Bagwell K, Riordan MH (1991) High and declining prices signal product quality. Am Econ Rev 81(1):224–239
Bar-Issac H, Caruana G, Cuñat V (2010) Information gathering and marketing. J Econ Manag Strat 19(2):375–401
Bawa K, Shoemaker R (2004) The effects of free sample promotions on incremental brand sales. Market Sci 23(3):345–363
Bell DR, Gallino S, Moreno A (2018) Offline showrooms in omni-channel retail: demand and operational benefits. Manag Sci 64(4):1629–1651
Bhardwaj P, Chen Y, Godes D (2008) Buyer-initiated vs. seller-initiated information revelation. Manag Sci 54(6):1104–1114
Bhargava HK, Chen RR (2012) The benefit of information asymmetry: when to sell to informed customers? Decis Support Syst 53(2):345–356
Bhargava HK, Choudhary V (2001) Information goods and vertical differentiation. J Manag Inf Syst 18(2):89–106
Bhargava HK, Choudhary V (2008) When is versioning optimal for information goods? Manag Sci 54(5):1029–1035
Biyalogorsky E, Gerstner E, Weiss D, Xie J (2005) The economics of service upgrades. J Serv Res 7(3):234–244
Board O (2009) Competition and disclosure. J Ind Econ 57(1):197–213
Boleslavsky R, Cotton CS, Gurnani H (2017) Demonstrations and price competition in new product release. Manag Sci 63(6):2016–2026
Branco F, Sun M, Villas-Boas JM (2016) Too much information? Information provision and search costs. Market Sci 35(4):605–618
Cai G, Chen Y, Wu C, Hsiao L (2013) Probabilistic selling, channel structure, and supplier competition. Decis Sci 44(2):267–296
Caves RE, Greene DP (1996) Brands' quality levels, prices, and advertising outlays: empirical evidence on signals and information costs. Int J Ind Organ 14(1):29–52
Çelik L (2014) Information unraveling revisited: disclosure of horizontal attributes. J Ind Econ 62(1):113–136

Chen Y, Xie J (2008) Online consumer review: word-of-mouth as a new element of marketing communication mix. Manag Sci 54(3):477–491
Chen RR, Gerstner E, Yang Y (2009) Should captive sardines be compensated? Serving customers in a confined zone. Market Sci 28(3):599–608
Chen T, Kalra A, Sun B (2009) Why do consumers buy extended service contracts? J Consum Res 36(4):611–623
Chen RR, Gal-Or E, Roma P (2014) Opaque distribution channels for competing service providers: posted price vs. name-your-own-price mechanisms. Oper Res 62(4):733–750
Chu J, Chintagunta OK (2011) An empirical test of warranty theories in the U.S. computer server and automobile markets. J Market 75(2):75–92
Cooper R, Ross T (1985) Warranties and double moral hazard. RAND J Econ 16(1):103–113
Courville L, Hausman WH (1979) Warranty scope and reliability under imperfect information and alternative market structures. J Bus 52:361–370
Davis S, Gerstner E, Hagerty M (1995) Money back guarantees in retailing: matching products to consumer tastes. J Retail 71(1):7–22
Dellarocas C (2003) The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag Sci 49(10):1407–1424
Deng S, Wu L, Chen RR (2017) Fit-revelation sampling and advertising: complementary or substitutable? Working paper, University of California, Davis
Dimoka A, Hong Y, Pavlou PA (2012) On product uncertainty in online markets: theory and evidence. MIS Q 36(2):395–426
Dranove D, Jin GZ (2010) Quality disclosure and certification: theory and practice. J Econ Lit 48(4):935–963
Emons W (1989) The theory of warranty contracts. J Econ Surv 3(1):43–57
Fainmesser IP, Lauga D, Ofek E (2018) Ratings, reviews, and the marketing of new products. Working paper, Johns Hopkins University, Baltimore
Fay S (2008) Selling an opaque product through an intermediary: the case of disguising one's product. J Retail 84(1):59–75
Fay S, Xie J (2008) Probabilistic goods: a creative way of selling products and services. Market Sci 27(4):674–690
Fay S, Xie J (2010) The economics of buyer uncertainty: advance selling vs. probabilistic selling. Market Sci 29(6):1040–1057
Fay S, Xie J (2015) Timing of product allocation: using probabilistic selling to enhance inventory management. Manag Sci 61(2):474–484
Fay S, Xie J, Feng C (2015) The effect of probabilistic selling on the optimal product mix. J Retail 91(3):451–467
Gal-Or E (1989) Warranties as a signal of quality. Can J Econ 22(1):50–61
Gallego G, Phillips R (2004) Revenue management of flexible products. Manuf Serv Oper Manag 6(4):321–337
Gallino S, Moreno A (2018) The value of fit Information in online retail: evidence from a randomized field experiment. Manuf Serv Oper Manag 20(4):767–787
Gefen D, Karahanna E, Straub D (2003) Trust and TAM in online shopping: an integrated model. MIS Q 27(1):51–90
Ghose A (2009) Internet exchanges for used goods: an empirical analysis of trade patterns and adverse selection. MIS Q 33(2):263–291
Ghose A, Ipeirotis P (2011) Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. IEEE Trans Knowl Data Eng 23(10):1498–1512
Grossman S (1981) The informational role of warranties and private disclosure about product quality. J Law Econ 24(3):461–483
Grossman S, Hart O (1980) Disclosure laws and takeover bids. J Financ 35(2):323–334
Gu Z, Liu Y (2013) Consumer fit search, retailer shelf layout, and channel interaction. Market Sci 32(4):652–668
Gu Z, Liu Y (2018) Why would a big retailer refuse to collaborate on manufacturer SPIFF? Quant Market Econ 16(4):441–472

Gu Z, Tayi GK (2015) Investigating firm strategies on offering consumer-customizable products. Inf Syst Res 26(2):456–468
Gu Z, Tayi GK (2017) Consumer pseudo-showrooming and omni-channel placement strategies. MIS Q 41(2):583–606
Gu Z, Xie Y (2013) Facilitating fit-revelation in the competitive market. Manag Sci 59(5):1196–1212
Guan X, Chen Y-J (2017) The interplay between information acquisition and quality disclosure. Prod Oper Manag 26(3):389–408
Guo L (2009) Quality disclosure formats in a distribution channel. Manag Sci 55(9):1513–1526
Guo L, Zhao Y (2009) Voluntary quality provision and market interaction. Market Sci 28(3):488–501
Halbheer D, Gärtner D, Gerstner E, Koenigsberg O (2018) Optimizing service failure and damage control. Int J Res Market 35(1):100–115
Hao L, Tan Y (2017) Who wants consumers to be informed? Facilitating information disclosure in a distribution channel. Inf Syst Res 30(1):1–349. https://doi.org/10.1287/isre.2017.0770
Heal G (1977) Guarantees and risk-sharing. Rev Econ Stud 44(3):549–560
Heiman A, Muller E (1996) Using demonstration to increase new product acceptance: controlling demonstration time. J Market Res 33(4):422–430
Heiman A, McWilliams B, Zilberman D (2001) Demonstrations and money-back guarantees: market mechanisms to reduce uncertainty. J Bus Res 54(1):71–84
Hong Y, Pavlou PA (2014) Product fit uncertainty in online markets: nature, effects, and antecedents. Inf Syst Res 25(2):328–344
Horstmann I, MacDonald G (2003) Is advertising a signal of product quality? Evidence from the compact disc player market, 1983–1992. Int J Ind Organ 21(3):317–345
Huang T, Yu Y (2014) Sell probabilistic goods? A behavioral explanation for opaque selling. Market Sci 33(5):743–759
Huang X, Sošić G, Kersten G (2017) Selling through Priceline? On the impact of name-your-own-price in competitive market. IISE Trans 49(3):304–319
Jain D, Mahajan V, Muller E (1995) An approach for determining optimal product sampling for the diffusion of a new product. J Prod Innov Manag 12(2):124–135
Jerath K, Netessine S, Veeraraghavan SK (2010) Revenue management with strategic customers: last-minute selling and opaque selling. Manag Sci 56(3):430–448
Jiang Y (2007) Price discrimination with opaque products. J Revenue Pricing Manag 6(2):118–134
Jiang Z, Benbasat I (2007) The effects of presentation formats and task complexity on online consumers' product understanding. MIS Q 31(3):475–500
Jiang Y, Guo H (2015) Design of consumer review systems and product pricing. Inf Syst Res 26(4):714–730
Jiang B, Zhang X (2011) How does a retailer's service plan affect a manufacturer's warranty? Manag Sci 57(4):727–740
Jing B (2016) Lowering customer evaluation costs, product differentiation, and price competition. Market Sci 35(1):113–127
Johnson JP, Myatt DP (2006) On the simple economics of advertising, marketing, and product design. Am Econ Rev 96(3):756–784
Jones R, Mendelson H (2011) Information goods vs. industrial goods: cost structure and competition. Manag Sci 57(1):164–176
Jovanovic B (1982) Truthful disclosure of information. Bell J Econ 13(1):36–44
Kihlstrom RE Riordan M (1984) Advertising as a signal. J Polit Econ 92(3): 427–450
Kim Y, Krishnan R (2015) On product-level uncertainty and online purchase behavior: an empirical analysis. Manag Sci 61(10):2449–2467
Kirmani A (1990) The effect of perceived advertising costs on brand perceptions. J Consum Res 17(2):160–171
Kirmani A, Wright P (1989) Money talks: perceived advertising expense and expected product quality. J Consum Res 16(3):344–353

Koessler F, Renault R (2012) When does a firm disclose product information? RAND J Econom 43(4):630–649
Kubo Y (1986) Quality uncertainty and guarantee—a case of strategic market segmentation by a monopolist. Eur Econ Rev 30:1063–1079
Kuksov D, Lin Y (2010) Information provision in a vertically differentiated competitive marketplace. Market Sci 29(1):122–138
Kwark Y, Chen J, Raghunathan S (2014) Online product reviews: implications for retailers and competing manufacturers. Inf Syst Res 25(1):93–110
Kwark Y, Chen J, Raghunathan S (2017) Platform or wholesale? A strategic tool for online retailers to benefit from third-party information. MIS Q 41(3):763–785
Lahiri A, Dey D (2013) Effects of piracy on quality of information goods. Manag Sci 59(1):245–264
Lahiri A, Dey D (2018) Versioning and information dissemination: a new perspective. Inf Syst Res 29(4):965–983
Lal R, Sarvary M (1999) When and how is the Internet likely to decrease price competition? Market Sci 18(4):485–503
Lewis TR, Sappington DEM (1994) Supplying information to facilitate price discrimination. Int Econ Rev 35(2):309–327
Li X (2017) Revealing or non-revealing: the impact of review disclosure policy on firm profitability. MIS Q 41(4):1335–1345
Lin Y, Pazgal A (2016) Hide supremacy or admit inferiority—market entry strategies in response to consumer informational needs. Consum Needs Solut 3(2):94–103
Linnemer L (2002) Price and advertising as signals of quality when some consumers are informed. Int J Ind Organ 20(7):931–947
Liu Y, Feng J, Xie J (2017) Review for service: a blessing or a curse? Working paper, City University of Hong Kong, Hong Kong
Loginova O, Mantovani A (2015) Information and online reviews. Working paper, University of Bologna, Bologna
Lutz NA, Padmanabhan V (1995) Why do we observe minimal warranties? Market Sci 14(4):417–441
Mayzlin D, Shin J (2011) Uninformative advertising as an invitation to search. Market Sci 30(4):666–685
Menezes MAJ, Currim IS (1992) An approach for determination of warranty length. Int J Res Market 9(2):177–195
Milgrom P (1981) Good news and bad news: representation theorems and applications. Bell J Econ 12(2):380–391
Milgrom P (2008) What the seller won't tell you: persuasion and disclosure in markets. J Econ Perspect 22(2):115–131
Milgrom P, Roberts J (1986) Price and advertising signals of product quality. J Polit Econ 94(4):796–821
Moorthy S, Hawkins SA (2005) Advertising repetition and quality perception. J Bus Res 58(3):354–360
Moorthy S, Srinivasan K (1995) Signaling quality with a money-back guarantee: the role of transaction costs. Market Sci 14(4):442–466
Moorthy S, Zhao H (2000) Advertising spending and perceived quality. Market Lett 11(3):221–233
Nelson P (1974) Advertising as information. J Polit Econ 82(4):729–754
Ofek E, Katona Z, Sarvary M (2011) "Bricks and clicks": the impact of product returns on the strategies of multichannel retailers. Market Sci 30(1):42–60
Padmanabhan V (1995) Usage heterogeneity and extended warranties. J Econ Manag Strateg 4(1):33–53
Padmanabhan V, Rao RR (1993) Warranty policy and extended service contracts: theory and an application to automobiles. Market Sci 12:230–247

Pavlou PA, Dimoka A (2006) The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. Inf Syst Res 17(4):391–412
Pavlou PA, Gefen D (2004) Building effective online marketplaces with institution-based trust. Inf Syst Res 15(1):37–59
Pavlou PA, Liang H, Xue Y (2007) Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective. MIS Q 31(1):105–136
Png PL, Wang H (2010) Buyer uncertainty and two-part pricing: theory and applications. Manag Sci 56(2):334–342
Post D, Spann M (2012) Improving airline revenues with variable opaque products: "Blind booking" at Germanwings. Interfaces 42(4):329–338
Prosser WL (1943) The implied warranty of merchantable quality. Minnesota Law Rev 27:117–168
Sahoo N, Dellarocas C, Srinivasan S (2017) The impact of online product reviews on product returns. Working paper, Boston University, Boston
Shapiro D, Shi X (2008) Market segmentation: the role of opaque travel agencies. J Econ Manag Strateg 17(4):803–837
Shugan SM, Xie J (2000) Advance pricing of services and other implications of separating purchase and consumption. J Serv Res 2(3):227–239
Shulman JD, Coughlan AT, Savaskan RC (2009) Optimal restocking fees and information provision in an integrated demand supply model of product returns. Manuf Serv Oper Manag 11(4):577–594
Shulman JD, Cunha M, Saint Clair JK (2015) Consumer uncertainty and purchase decision reversals: theory and evidence. Market Sci 34(4):590–605
Spence M (1977) Consumer misperceptions, product failure and producer liability. Rev Econ Stud 44(3):561–572
Sun M (2011) Disclosing multiple product attributes. J Econ Manag Strateg 20(1):195–224
Tellis GJ, Fornell C (1988) Advertising and quality over the product life cycle: a contingency theory. J Market Res 15(1):64–71
Terwiesch C, Loch C (2004) Collaborative prototyping and the pricing of custom-designed products. Manag Sci 50(2):145–158
Thomas L, Shane S, Weigelta K (1998) An empirical examination of advertising as a signal of product quality. J Econ Behav Organ 37(4):415–430
Wernerfelt B (1994) On the function of sales assistance. Market Sci 13(1):68–82
Xie J, Shugan SM (2001) Electronic tickets, smart cards, and online prepayments: when and how to advance sell. Market Sci 20(3):219–243
Yu M, Kapuscinski R, Ahn H-S (2015) Advance selling: effects of interdependent consumer valuations and sellers capacity. Manag Sci 61(9):2100–2117
Zhang Z, Joseph K, Subramaniam R (2015) Probabilistic selling in quality-differentiated markets. Manag Sci 61(8):1959–1977

Chapter 4
Optimizing Promotions for Multiple Items in Supermarkets

Maxime C. Cohen and Georgia Perakis

Abstract Promotion planning is an important problem for supermarket retailers who need to decide the price promotions for thousands of items. One of the key reasons retailers use promotions is to increase sales and profits by exploiting relations among the different items. We formulate the promotion optimization problem for multiple items as a nonlinear Integer Program (IP). Our formulation captures several business requirements, as well as important economic factors such as the post-promotion dip effect (due to the stockpiling behavior of consumers) and cross-item effects (substitution and complementarity). Our demand models are estimated from data and are typically nonlinear, hence rendering the exact formulation intractable. In this chapter, we discuss a class of IP approximations that can be applied to any demand function. We then show that for demand models with additive cross-item effects, it is enough to account for unilateral and pairwise deviations, leading to an efficient method. In addition, when the products are substitutable and the price ladder is of size two, we show that the unconstrained problem can be solved efficiently by a linear program. This result is unexpected as the feasible region of the formulation is not totally unimodular. Next, we derive a parametric worst-case guarantee on the accuracy of the approximation relative to the optimal solution. Finally, we test our model on realistic real-world instances and show its performance and practicality. The model and tool presented in this chapter allow retailers to solve large realistic instances and to improve their promotion decisions.

Keywords Promotion optimization · Dynamic pricing · Integer programming · Retail operations

---

M. C. Cohen (✉)
Desautels Faculty of Management, McGill University, Montreal, QC, Canada
e-mail: maxime.cohen@mcgill.ca

G. Perakis
MIT Sloan School of Management, Cambridge, MA, USA
e-mail: georgiap@mit.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_4

4.1 Introduction

This chapter presents some recent developments in retail promotions. In many retail settings such as supermarkets, promotions are a key driver to boost profits. Promotions are used on a daily basis in most retail environments including supermarkets, drugstores, fashion retailers, electronics stores, online retailers, convenience stores, etc. For example, a typical supermarket sells several thousand products and needs to decide the price promotions for all products at each time period. These decisions are of primary importance, as using the right promotions can significantly enhance the business' bottom line. In today's economy, retailers offer hundreds or even thousands promotions simultaneously. Promotions aim to increase sales and traffic, enhance awareness when introducing new items, clear leftover inventory, bolster customer loyalty, and improve competitiveness. In addition, promotions are often used as a tool for price discrimination among customers.

Surprisingly, many retailers still employ a manual process based on intuition and past experience to decide the depth and timing of promotions. The unprecedented volume of data that is now available to retailers presents an opportunity to develop decision support tools that can help retailers improve promotion decisions. The promotion planning process typically involves a large number of decision variables, and needs to ensure that the relevant business constraints (called promotion *business rules*) are satisfied (more details can be found in Sect. 4.3.2). In this chapter, we discuss how analytics can help retailers decide the promotions for multiple items while accounting for many important modeling aspects observed in retail data. In particular, we consider practical models that are motivated by a collaboration between academia and industry. Most of the material discussed in this chapter is inspired by the results in Cohen et al. (2017) and in Cohen et al. (2018). For more details on the specifics of the algorithms, the proofs of the analytical results, and the managerial insights, we refer the reader to those papers.

Several recent advances in Operations Management and Marketing have focused on developing new methods to improve the process of deciding retail promotions. The ultimate goal is to increase the total profit by promoting the right items at the right time using the right price points. At a high level, retail promotions can be categorized as follows: (1) manufacturer versus retailer promotions, (2) markdowns versus temporary price discounts, (3) targeted versus mass campaigns, and (4) price reductions versus alternative promotion vehicles. We next discuss these four categorizations.

Manufacturer Versus Retailer Promotions In retail settings, the brand manufacturer (e.g., Coca-Cola and Kellogg's) can directly offer a price discount either to the retailer or to the end consumer. These incentives are often called trade funds, vendor funds, or manufacturer coupons. These types of promotion are usually driven by long-term negotiations between the manufacturer and the retailer and involve several contractual terms. For example, a manufacturer can offer a rebate to the retailer if the cumulative sales during the quarter exceed a certain target level. In exchange, the retailer will place the manufacturer's products in preferred locations (e.g., end-cap displays). A second example is a shared promotion contract in which the manufacturer subsidizes some portion of the price discount offered to consumers. A third example occurs when a manufacturer offers a coupon to end-consumers who then need to claim the discount (at the store or online). Typically, retailers need to decide when to accept such vendor funds and under what conditions. In many situations, manufacturers tend to be aggressive on these contractual terms by imposing long-term commitments, high volumes, and sometimes exclusivity restrictions (e.g., not allowing promoting competing brands).

Markdowns versus Temporary Price Discounts Markdowns typically refer to the practice of decreasing the price of an item at the end of the selling season. The regular price is decreased in order to clear leftover inventory. Note that in such a case, the price may be reduced several times but cannot be increased back to the regular price. This is common practice in the fashion and tourism industries as well as in the business of selling tickets for media events (e.g., concerts). For example, summer apparel may be discounted toward the end of the season if remaining inventory is higher than anticipated. On the other hand, temporary price discounts are used in different contexts. A well-known such context is Fast-Moving Consumer Goods (FMCG) such as processed foods, soft drinks, and frequently purchased household products (e.g., laundry detergent and toothpaste). These products are usually non-perishable and have a long shelf life. Such purchases are recurring and retailers do not need to clear remaining inventory. To increase profit, it is common for retailers to use temporary price reductions (e.g., 20% off the regular price during 1 week).

Targeted versus Mass Campaigns Retailers can decide either to send promotions to a few targeted customers or to simply decrease the price of a particular product for all potential buyers. Targeted marketing campaigns can be implemented via email-redeemable coupons or by using advanced geo-localization techniques. Online retailers often use targeted promotions by tracking potential customers using cookies and by sending promotional offers to selected sets of customers (e.g., active members that made a recent purchase). On the other hand, mass promotions are price discounts that apply to all customers. Brick-and-mortar retailers such as supermarkets mainly employ mass-promotion campaigns.

Price Reductions versus Alternative Promotion Vehicles Retailers promote products in many ways. The most straightforward method is to use a price discount in which the item is temporarily priced below its regular price. Other options include "buy one get one free" offers, in-store flyers, coupons, tasting stands, placing products at the end of an aisle (end-cap display), sending out flyers, broadcasting TV commercials, and radio advertisements (these are often called promotion vehicles). Typically, a retailer can choose among 5–40 different promotion vehicles at each point in time.

In this chapter, we focus on the mass pricing promotion optimization problem faced by a retailer who sells FMCG products. Namely, we consider a retailer (e.g., a supermarket) who needs to decide which items to promote, at which price points, and when to schedule the promotions of the different items. The problems of setting the right manufacturer incentives, optimizing markdowns, designing targeted promotions, and optimizing promotion vehicles are also important retail questions, but are beyond the scope of this chapter. We will briefly refer to some of the relevant literature on these problems in Sect. 4.2.

The amount of money spent on promotions for FMCG products can be significant—it is estimated that FMCG manufacturers spend about $1 trillion annually on promotions (Nielsen 2015). In addition, promotions play an important role in the FMCG industry as a large proportion of sales is made during promotions. For example, retail data indicates that 12–25% of supermarket sales in five European countries were made during promotions (Gedenk et al. 2006). The market research group IRI found that more than half of all goods (54.6%) sold to UK shoppers in supermarkets and major retailers were on promotion.[1]

The promotion planning process faced by a medium to large size retailer is challenging for several reasons. First, one needs to carefully account for cross-item effects in demand (cannibalization and complementarity). When promoting a particular item, the demand for some other products may also be affected by the promotion. Consequently, one needs to decide the promotions of all items in the category while accounting for those effects that can be directly estimated from data. Second, retail promotions are often constrained by a set of business rules specified by the retailer and/or the product's manufacturer. Examples of business rules include prices chosen from a set of discrete values, limiting the number of promotions (both per time period and for each item), and cross-item business rules that restrict the relationship between prices of different items (more details are provided in Sect. 4.3.2). Third, demand usually exhibits a post-promotion-dip effect. This effect is induced by *promotion fatigue* (i.e., repeating the same promotion may have a low marginal impact) and by the stockpiling behavior of consumers. More precisely, for certain categories of (non-perishable) products, customers tend to stockpile during promotions by purchasing larger quantities for future consumption. This ultimately leads to a reduced demand following the promotion period. Fourth, the problem is difficult due to its large scale. As we mentioned, an average supermarket offers several thousand SKUs (Stock Keeping Units), and the number of items on promotion at any time can be very large. Consequently, this leads to a large number of decisions that need to be made by the retailer.

Retail promotions can have a significant impact on boosting sales and on influencing customers. For example, a study from the International Council of Shopping Centers shows that 90% of adult consumers claim to be influenced by promotions in terms of the amount they spend and the items they purchase.[2] Despite the complexity of the promotion planning process, it is still to this day performed manually in many supermarket chains. This motivates us to design and study

[1] https://www.theguardian.com/business/2015/nov/02/majority-of-goods-sold-in-uk-stores-on-promotion-finds-study-multi-buys.

[2] https://retailleader.com/brick-and-mortar-makes-grade-back-school-shopping.

promotion optimization models that can make promotion planning more efficient and automated. The goals of this line of research include the following:

- *Formulate the Promotion Optimization Problem for Multiple Items (Labeled as Multi-POP).* This formulation is directly motivated from practice, holds for general demand models (estimated from data), and can incorporate the relevant business rules.
- *Discuss how the Formulation Captures Several Important Economic Factors Present in Retail Environments.* These factors include the post-promotion dip effect (due to the stockpiling behavior of consumers), cross-item effects, and demand seasonality.
- *Develop an Efficient Approximate Solution Approach to Solve the Problem.* We propose a discrete linearization method that allows the retailer to solve a large-scale instance of the problem within seconds. We also convey that our solution approach provides a parametric worst-case bound on the quality of our approximation relative to the optimal solution (which cannot be found due to computational limitations).
- *Present a Beginning-to-End Application of the Entire Process of Optimizing Retail Promotions.* We divide the process into five steps that the retailer needs to follow-from collecting and aggregating the data to computing the future promotion decisions.
- *Discuss the Potential Impact of Using Data Analytics and Optimization for Retail Promotions.* We convey that in our tested examples (calibrated with real data), using the promotions suggested by our model can yield a 2–9% profit improvement. Such an increase is significant given that retail businesses typically operate under small margins.

This chapter is organized as follows: In Sect. 4.2, we review some of the related literature. In Sect. 4.3, we report the notation, assumptions, and problem formulation. In Sect. 4.4, we present a class of approximation methods to efficiently solve the promotion optimization problem. In Sect. 4.5, we use our model and solution approach to draw practical insights on promotion planning, and present a summary of how to apply our model to real-world retail environments. Finally, we report our conclusions in Sect. 4.6. As mentioned, additional details on the technical results and on the insights can be found in Cohen et al. (2017, 2018).

4.2 Literature Review

The topic of retail promotions has been an active research area both in academia and industry. In particular, our problem is related to several streams of literature, including dynamic pricing, promotions in Marketing, and retail operations.

Dynamic Pricing Dynamic pricing has been an extensive topic of research in the Operations Management community. Comprehensive reviews can be found in the books and review papers by Bitran and Caldentey (2003), Elmaghraby and Keskinocak (2003), Talluri and Van Ryzin (2006), Özer and Phillips (2012), as well as the references therein. A large number of recent papers study the problem of dynamic pricing under various contexts and modeling assumptions. Examples include Ahn et al. (2007), Su (2010), and Levin et al. (2010), just to name a few. In Ahn et al. (2007), the authors propose a demand model in which a proportion of customers strategically wait k periods and purchase the product once the price falls below their willingness to pay. They then formulate a mathematical programming model and develop solution techniques. In Su (2010), the author studies a model with multiple consumer types who may differ in their holding costs, consumption rates, and fixed shopping costs. The author solves the dynamic pricing model by computing the rational expectation equilibrium, and draws several managerial insights. In Levin et al. (2010), the authors consider a dynamic pricing model for a monopolist who sells a perishable product to strategic consumers. They model the problem as a stochastic dynamic game and prove the existence of a unique subgame-perfect equilibrium pricing strategy. A very prominent topic in the dynamic pricing literature is the setting in which consumers are strategic (or forward-looking) (see, e.g., Aviv and Pazgal 2008; Cachon and Swinney 2009; Levina et al. 2009; Besbes and Lobel 2015; Liu and Cooper 2015; Chen and Farias 2015). The problem considered in this chapter is in the same spirit as the dynamic pricing problem. Nevertheless, we focus on a setting where the demand model is estimated from historical data, and the optimization formulation includes the simultaneous promotion decisions of several interconnected items. In addition, we require the dynamic pricing decisions to satisfy several business rules.

Promotions in Marketing Sales promotions are an important area of research in Marketing (see Blattberg and Neslin (1990) and the references therein). However, the focus in the Marketing community is typically on modeling and estimating dynamic sales models (econometric or choice models) that can be used to draw managerial insights (Cooper et al. 1999; Foekens et al. 1998). For example, Foekens et al. (1998) study econometric models based on scanner data to examine the dynamic effects of sales promotions. It is widely recognized that for certain products, promotions may have a pantry-loading or post-promotion dip effect, i.e., consumers tend to purchase larger quantities during promotions for future consumption (stockpiling behavior). This effect leads to a decrease in sales in the short term. To capture the post-promotion dip effect, many of the dynamic sales models in the Marketing literature posit that the demand is not only a function of the current price, but also of the past prices (see, e.g., Ailawadi et al. 2007; Macé and Neslin 2004). Finally, note that several prescriptive studies examine the impact of retail coupons in the context of sales promotions (see, e.g., Heilman et al. 2002). The demand models used in this chapter also consider that demand depends explicitly on current and past prices as well as on prices of other items.

Retail Operations Several academic papers study the topic of retail promotions from an empirical descriptive perspective. Van Heerde et al. (2003) and Martínez-

Ruiz et al. (2006) use panel data to study how retail promotions induce consumers to switch brands. The recent work by Felgate and Fearne (2015) uses supermarket loyalty card data from a sample of over 1.4 million UK households to analyze the effect of promotions on the sales of specific products across different shopper segments. Another line of research discusses field experiments on pricing decisions implemented at retailers. A classical successful example is the implementation at the fashion retail chain Zara (see Caro and Gallien 2012). In their work, the authors report the results of a controlled field experiment conducted in all Belgian and Irish stores during the 2008 fall-winter season. They assess that the new process has increased clearance revenues by approximately 6%. An additional recent work can be found in Ferreira et al. (2015) in which the authors collaborated with Rue La La, a flash sales fashion online retailer. They propose a non-parametric prediction model to predict future demand of new products and develop an efficient solution for the price optimization problem. They estimate a revenue increase for the test group of approximately 9.7%. Pekgün et al. (2013) describe a collaboration with the Carlson Rezidor Hotel Group. In this study, the authors show that demand forecasting and dynamic revenue optimization consistently increased revenue by 2–4% in participating hotels relative to non-participating ones.

Other Types of Promotions As mentioned before, retail promotions can be divided into several categories. While the models presented in this chapter focus on the mass pricing promotion optimization problem faced by a retailer who sells FMCG products, other studies have considered alternative promotion types. Several papers consider the problem of vendor funds in the context of promotion planning (see, e.g., Silva-Risso et al. 1999; Nijs et al. 2010; Yuan et al. 2013; Baardman et al. 2017). As mentioned before, an additional related topic is the one of markdown pricing or markdown optimization. In this problem, the seller needs to decide when to decrease the price of the item(s) so as to clear remaining inventory by the end of the season. There are a large number of academic papers that propose different models and methods to solve the markdown pricing problem. Examples include Yin et al. (2009), Mersereau and Zhang (2012), Zhang and Cooper (2008), Vakhutinsky et al. (2012), and Caro and Gallien (2012). As we explained before, the promotion optimization problem for FMCG products differs from the markdown optimization problem by the structure of the pricing policy and by the lack of inventory expiration. The topic of designing targeted promotions has recently attracted a lot of attention from both academics and practitioners. Given that sending promotions to existing or new customers can be expensive and often results in low conversion rates, several firms aim to develop quantitative methods that exploit large historical data to design targeted promotion campaigns. For example, retailers often need to decide which types of customer to target, and what are the most important features (e.g., geo-localization, demographics, and past behavior). Targeted marketing campaigns (email and mobile offers) have been extensively studied in the academic literature (see, e.g., Arora et al. 2008; Fong et al. 2015; Andrews et al. 2015; Jagabathula et al. 2018). Finally, in addition to price promotions, retailers typically need to decide how to assign the different vehicles (e.g., flyers and TV commercials). The recent work in Baardman et al. (2018) addresses the problem of optimally scheduling promotion vehicles for a retailer.

Methodology From a methodological standpoint, the tools used in this chapter are related to the literature on nonlinear and integer optimization. We formulate the promotion optimization problem as a nonlinear mixed integer program (NMIP). Due to the general classes of demand functions we consider, the objective function is typically non-concave, and such NMIPs are generally difficult to solve from a computational standpoint. Under certain special structural conditions (see, e.g., Hemmecke et al. (2010) and the references therein), there exist polynomial time algorithms for solving NMIPs. However, many NMIPs do not satisfy these conditions and are solved using techniques such as Branch and Bound, Outer-Approximation, Generalized Benders, and Extended Cutting Plane methods (Grossmann 2002).

In the special instance of the promotion optimization problem with linear demand and continuous prices, one can formulate our problem as a Cardinality-Constrained Quadratic Optimization (CCQO) problem. It has been shown in Bienstock (1996) that this problem is NP-hard. Thus, tailored heuristics have been developed to solve this type of problem (see, for example, Bienstock 1996; Bertsimas and Shioda 2009). The general instance of our problem has discrete variables and considers a general demand function. Note that our problem was also shown to be NP-hard (Cohen et al. 2016). Our solution approach is based on approximating the objective function by exploiting the discrete nature of the problem. Given that we consider general demand functions, it is not possible to use linearization approaches such as in Sherali and Adams (1998). Our main approximation method results in a formulation which is related to the field of Quadratic Programming. Such problems were extensively studied in the literature (see, e.g., Frank and Wolfe 1956; Balinski 1970; Rhys 1970; Padberg 1989; Nocedal and Wright 2006).

4.3 Problem Formulation

In this section, we formulate the promotion optimization problem (labeled as Multi-POP). We first introduce our notation and assumptions. We then discuss the various business rules that the retailer needs to satisfy when deciding price promotions. Finally, we present the resulting optimization formulation.

Consider a retailer who sells several FMCG products. Very often, retailers decide the price promotions of their products for each category separately. Consequently, we focus our presentation on a single category (e.g., ground coffee and soft drinks) composed of N items (or SKUs). The goal of the category manager is to maximize the total profit over a selling horizon composed of T periods (for example, one quarter of 13 weeks). We denote by p_t^i the price of item i at time t.[3] We also denote by c_t^i the (exogenous) cost of a single unit of item i at time t. In other words, we assume that the cost of each item at each time is known, and that the retailer needs to decide the prices of all N items during all T time periods. A summary of our notation can be found at the end of this section.

4.3.1 Assumptions

To gain tractability, we impose the following assumptions.

Assumption 4.1

1. *The retailer decides all price promotions at the beginning of the season.*
2. *The retailer carries enough inventory to meet demand for each item in each time period.*[4]
3. *Demand is expressed as a deterministic time-dependent nonlinear function of prices.*
4. *The demand function of item i depends explicitly on self-current and past prices (i.e., p_t^i and $p_{t-\ell}^i$ for $\ell = 1, \ldots, M^i$) and on cross-current prices (i.e., the vector of prices of all items but i at time t denoted $\mathbf{p}_t^{-i}$).*

Here, M^i represents the memory parameter of item i, that is, the number of past prices that affect current demand. We next briefly discuss the validity of the above assumptions. Assumption 4.1.1 applies to a setting where the retailer needs to commit upfront for the entire selling season. For example, such restrictions can emerge from vendor funds or can be imposed by sending out flyers through different advertising channels.

Note that Assumption 4.1.2 does not apply to all products and retail settings (e.g., very often in the fashion industry, limited inventory is produced to induce scarcity). Unlike fashion items which may be seasonal, FMCG products are typically available all year round. These products have a long shelf life and customers have been conditioned to always find them in stock at retail stores. Since FMCG products are usually easy to store and have a high degree of availability, FMCG retailers typically do not stock out. In Cohen et al. (2017), the authors analyze 2 years of supermarket data for FMCG products and convey that: (1) demand forecast accuracy for this type of product is often high (good out-of-sample R^2 and MAPE) and (2) inventory is not an issue as very few stock-outs occurred over a 2-year period. This can be justified

[3]Throughout this chapter, the subscript (resp. superscript) index corresponds to the time (resp. item).

[4]We therefore use the words *demand* and *sales* interchangeably.

by the fact that supermarkets have a long experience with inventory decisions, and collect large data sets allowing them to deploy sophisticated demand forecasting tools to support ordering decisions (see, e.g., Cooper et al. 1999; Van Donselaar et al. 2006). Finally, grocery retailers are aware of the negative effects of being out of stock for promoted products (see, e.g., Corsten and Gruen 2004; Campo et al. 2000). However, for settings where inventory is limited, one needs to consider a different formulation than the one presented in this chapter.

Assumption 4.1.3 translates to denoting demand of item i at time t by $d_t^i(\mathbf{p})$, where $\mathbf{p}$ is a vector of current and past prices (see more details below). We assume that demand is a deterministic function as we observed a high out-of-sample prediction accuracy using our data. Extending our model when demand is a stochastic function is an interesting direction for future research (e.g., by using learning algorithms).

Assumption 4.1.4 implies that the demand function does not explicitly depend on cross-past prices. In other words, the demand of item i does not depend on the past prices of other items in the category. This assumption was validated by running demand prediction models using retail data (more details can be found in Cohen et al. 2018). Consequently, demand of item i at time t can be any nonlinear and time-dependent function of the form: $d_t^i(p_t^i, p_{t-1}^i, \ldots, p_{t-M^i}^i, \mathbf{p}_t^{-i})$. Note that in practice M^i is estimated from historical data and can be item-dependent.

As discussed, demand of item i at time t depends on several factors:

- The self-current price p_t^i—This captures the price sensitivity of consumers toward the item.
- The self-past prices $(p_{t-1}^i, \ldots, p_{t-M^i}^i)$—This captures the post-promotion dip effect (induced by the stockpiling behavior of consumers).
- The cross-current prices $\mathbf{p}_t^{-i}$—This captures the cross-item effects on demand (substitution and complementarity).
- Other potential features such as demand seasonality (weekly, monthly, or quarterly), trend factor, store effect, holiday boosts, etc.

Concrete demand models such as the log-log demand function can be found in Cohen et al. (2017).

In most product categories, a promotion for a particular item affects its own sales but also the sales of other items in the category. As mentioned, we capture these cross-item effects by assuming that demand of item i depends on the prices of other items (at the same time period). The standard examples of substitutable items are competing brands such as Coke and Pepsi. In this case, it is clear that promoting a Coke product potentially increases Coke's sales but it may also decrease Pepsi's sales. Mathematically, one can assume that if items i and $j \neq i$ are substitutes, then $\partial d_t^i / \partial p_t^j \geq 0$ and $\partial d_t^j / \partial p_t^i \geq 0$ for some t. Two products i and j are complements if the consumption of i induces customers to purchase item j (and vice versa), e.g., shampoo and conditioner. Mathematically, one can assume that if items i and $j \neq i$ are complements, then $\partial d_t^i / \partial p_t^j \leq 0$ and $\partial d_t^j / \partial p_t^i \leq 0$ for some t.

4.3.2 Business Rules

In the retail setting that we consider, there are typically two classes of business rules: (1) business rules on each item separately (called *self business rules*) and (2) business rules that impose joint pricing constraints on several items (called *cross-item business rules*). The self business rules are identical to the ones presented in Cohen et al. (2017), whereas the cross-item business rules are similar to Cohen et al. (2018).

Self Business Rules

1. *Prices are Chosen from a Discrete Price Ladder.* For each product, there is a finite set of permissible prices. In particular, we consider that each item $i = 1, \ldots, N$ can take several prices: the regular price denoted by q^{i0}, and $K^i = |Q_i| - 1$ promotion prices denoted by q^{ik}, $k = 1, \ldots K^i$. The total number of price points for item i is called the size of the price ladder (denoted by $|Q_i|$).[5] Consequently, the price of item i at time t can be written as $p_t^i = \sum_{k=0}^{K^i} q^{ik} \gamma_t^{ik}$, where the binary decision variable γ_t^{ik} is equal to 1 if the price of item i at time t is q^{ik} and 0 otherwise.
2. *Limited Number of Promotions.* The retailer may want to limit the promotion frequency for a product to preserve the image of their store, and not train customers to be deal-seekers. For example, the retailer may wish to promote item i at most $L^i = 3$ times during the quarter. This requirement for item i is captured by the following constraint: $\sum_{t=1}^{T} \sum_{k=1}^{K^i} \gamma_t^{ik} \leq L^i$.
3. *Separating Periods Between Successive Promotions (No-Touch Constraint).* A common additional requirement is to space out two successive promotions by a minimal number of separating periods, denoted by S^i. This constraint also helps retailers preserve their store image and discourage consumers to be deal-seekers. In addition, this type of requirement may be dictated by the manufacturer that sometimes restricts the frequency of promotions to preserve the brand image. Such a requirement for item i translates to adding the following constraint: $\sum_{\tau=t}^{t+S^i} \sum_{k=1}^{K^i} \gamma_\tau^{ik} \leq 1 \; \forall t$.

[5]For simplicity, we assume that the elements of the price ladder are time-independent but our results still hold when this assumption is relaxed. In addition, we assume without loss of generality that the regular non-promotion price $q^{i0} = q^0$ is the same across all items $i = 1, \ldots, n$ and all time periods (this assumption can be relaxed at the expense of a more cumbersome notation).

Cross-Item Business Rules

1. *Total Limited Number of Promotions.* The retailer may want to limit the total number of promotions throughout the selling season. For example, at most $L_T = 20$ promotions may be allowed during the season. Mathematically, one can impose the following constraint:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K^i} \gamma_t^{ki} \leq L_T. \tag{4.1}$$

Note that L_T should satisfy $\sum_{i=1}^{N} L^i > L_T$ for this constraint to be relevant.
2. *Inter-Item Ordinal Constraints.* Several price relations can be dictated by business rules. For example, smaller size items should have a lower price relative to similar larger-sized products; and national brands must be more expensive when compared to private labels. These constraints can be captured by linear inequalities in prices (e.g., if item i should be priced no higher than item j, one can add the constraint: $p_t^i \leq p_t^j\ \forall t$).
3. *Simultaneous Promotions.* Sometimes, retailers require particular items to be promoted simultaneously as part of a manufacturer incentive or a special promotional event. If items i and j should be promoted simultaneously, one can impose: $\gamma_t^{0i} = \gamma_t^{0j}\ \forall t$, where γ_t^{0i} (resp. γ_t^{0j}) is a binary variable that is equal to 1 if item i (resp. item j) is not promoted at time t.
4. *Limited Number of Promotions in Each Period.* One can impose a limitation on the number of promotions in each time period. For example, at most $C^t = N/10$ promotions may be allowed, i.e., at most 10% of the items. Mathematically, we have:

$$\sum_{i=1}^{N}\sum_{k=1}^{K^i} \gamma_t^{ki} \leq C^t\ \forall t. \tag{4.2}$$

5. *Cross No-Touch Constraints.* An additional requirement can be to space out promotions of a set of similar items by a minimal number of separating periods, denoted by S^c. As before, this is motivated by the wish of preserving the store image and to mitigate the incentives for consumers to be deal-seekers. In this case, we need to separate successive promotions for two (or more) products. Mathematically, one can impose:

$$\sum_{i}\sum_{\tau=t}^{t+S^c}\sum_{k=1}^{K^i} \gamma_\tau^{ki} \leq 1\ \forall t,$$

where the sum on i can be over any given subset of items in the category. Note that when $S^c = 0$, this corresponds to never promoting the items simultaneously to impose an exclusive offer (very common in practice).

4.3.3 Problem Formulation

In what follows, we present the promotion optimization problem for multiple items (Multi-POP):

$$\max_{\gamma_t^{ik}} \sum_{i=1}^{N}\sum_{t=1}^{T}(p_t^i - c_t^i)d_t^i(p_t^i, p_{t-1}^i, \ldots, p_{t-M^i}^i, \mathbf{p}_t^{-i})$$

$$\begin{aligned}
\text{s.t. } & p_t^i = \sum_{k=0}^{K^i} q^{ik}\gamma_t^{ik} \quad \forall i && \text{(Prices are chosen from a discrete price ladder)}\\
& \sum_{t=1}^{T}\sum_{k=1}^{K^i}\gamma_t^{ik} \le L^i \quad \forall i && \text{(Limited number of promotions)}\\
& \sum_{\tau=t}^{t+S^i}\sum_{k=1}^{K^i}\gamma_\tau^{ik} \le 1 \quad \forall i, t && \text{(No-touch constraint)}\\
& \sum_{k=0}^{K^i}\gamma_t^{ik} = 1 \quad \forall i, t && \text{(Only a single price is selected)}\\
& \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K^i}\gamma_t^{ki} \le L_T && \text{(Total limited number of promotions)}\\
& \sum_{i=1}^{N}\sum_{k=1}^{K^i}\gamma_t^{ki} \le C^t \quad \forall t && \text{(Limited number of promotions in each period)}\\
& \gamma_t^{ik} \in \{0, 1\} \quad \forall i, t, k && \text{(Binary decision variables).}
\end{aligned}$$

In this problem, the objective is to maximize the total profit from all N items during the selling season. The objective function of the (Multi-POP) problem is denoted $MPOP$ whereas the objective function of the single-item problem is denoted $SPOP$. Note that in the above formulation, we have included all self business rules, as well as the constraints on the total limited number of promotions from (4.1), and on the limited number of promotions in each period from (4.2). One can naturally include additional cross-item business rules into the formulation, depending on the requirements. It is worth mentioning that even in the absence of cross-item business rules, the N items are linked through cross-item effects present in the demand functions.

Summary of Notation

T	Length of the selling season
N	Number of different items in the category
c_t^i	Cost of item i at time t (assumed to be known)
p_t^i	Price of item i at time t (decision variable)
$\mathbf{p}_t^{-i}$	Vector of prices of all items but i at time t
$d_t^i(p_t^i, p_{t-1}^i, \ldots, p_{t-M^i}^i, \mathbf{p}_t^{-i})$	Demand of item i at time t assumed to be a function of the self-current and past prices as well as cross-current prices (estimated from data)
M^i	Memory parameter of item i, i.e., the number of past prices that affect current demand (estimated from data)
L^i	Limitation of the number of promotions for item i
S^i	No-touch period for item i, i.e., the minimal number of periods between two successive promotions
K^i	Number of promotion prices in the price ladder of item i
q^0	Regular price (assumed to be the same across all items)
$\|Q_i\| = K^i + 1$	Total number of possible prices for item i
q^{ik}	Price point k for item i ($k = 1, \ldots, K^i$)
γ_t^{ik}	Binary decision variable to indicate if the price of item i at time t is equal to q^{ik}
$MPOP$	Objective function of the (Multi-POP) problem, i.e., the total profit generated by all items at all times
$SPOP$	Objective function of the single-item problem

4.4 Solution Approach

Our goal is to solve the (Multi-POP) optimization problem. Since the problem is a nonlinear Integer Program, solving the formulation efficiently is not straightforward. Consequently, we develop an approximate solution approach. The requirements are twofold: (1) the solution method needs to be efficient and run fast, and (2) the approximate solution needs to be near-optimal. In retail settings, retailers typically solve the (Multi-POP) problem for a large number of items. In addition, retailers often solve several instances of the problem to test the robustness of the solution before implementing it. More precisely, these routine tests are called *what-if scenarios*. They consist of solving perturbed versions of the nominal optimization problem, where some of the demand parameters and some of the business rules are slightly modified (more details are discussed in Sect. 4.5.2). In what follows, we describe the solution approaches developed in Cohen et al. (2017, 2018).

4.4.1 Single-Item Setting

We first present an efficient solution approach to solve the single-item problem. While the most interesting and relevant case is the problem with multiple items, the single-item setting is used as a starting point for the presentation, and is interesting in its own right. In certain retail categories, the different items can be independent, i.e., the demand function of each item depends solely on its prices and not on the prices of other items. In this case, the (Multi-POP) problem decomposes into N independent single-item problems (assuming that there are no cross-item business rules), and one can solve each problem separately.

Even in the case of a single item, the problem is hard to solve (the problem is shown to be NP-hard in Cohen et al. 2016). We observe that the constraints in the (Multi-POP) formulation are linear. However, the objective function is nonlinear, and usually neither concave nor convex, as we do not want to impose restrictions on the structure of the demand functions. This motivates us to propose a way to approximate the objective function by using a linear approximation and by exploiting the discrete nature of the problem. In particular, we approximate the objective function by the sum of the marginal contributions of having a single promotion at a time. For example, if the item is on promotion at times 2, 3, and 7, we approximate the objective by the sum of the marginal deviations of having a single promotion at time 2, a single promotion at time 3, and a single promotion at time 7. We next present this approach, called $App(1)$, in more detail.

The $App(1)$ approximation method ignores the second-order interactions between promotions and captures only the direct effect of each promotion. Since we consider the same set of constraints as in the original problem, the solution remains feasible. We next introduce some additional notation. We consider a particular item, and hence we drop the superscript i in the remainder of this subsection. For a given price vector $\mathbf{p} = (p_1, \ldots, p_T)$, we define the corresponding total profit (of the item under consideration) throughout the season:

$$SPOP(\mathbf{p}) = \sum_{t=1}^{T} (p_t - c_t) d_t(\mathbf{p}_t).$$

Next, we define the price vector (of dimension T) $\mathbf{p}_t^k$ such that the promotion price q^k is used at time t, and the regular price q^0 (no promotion) is used for all remaining periods. We denote the regular price vector (of dimension T) by $\mathbf{p}^0 = (q^0, \ldots, q^0)$, for which the regular price is set at all time periods. We define the coefficients b_t^k as follows:

$$b_t^k = SPOP(\mathbf{p}_t^k) - SPOP(\mathbf{p}^0). \tag{4.3}$$

The coefficients in Eq. (4.3) represent the unilateral deviations in total profit obtained by using a single promotion at a single point in time. One can compute these TK coefficients before starting the optimization procedure so that it does not affect the complexity of the method. The approximated objective function is then given by:

$$SPOP(\mathbf{p}^0) + \max_{\gamma_t^k} \sum_{t=1}^{T} \sum_{k=1}^{K} b_t^k \gamma_t^k, \tag{4.4}$$

while the set of constraints is the same as in the original problem. Consequently, the approximation optimization problem is linear and can be solved using a solver. As mentioned before, two important requirements for our solution approach are (1) a low running time and (2) a near-optimal solution. We next summarize the properties (both theoretical and practical) for the single-item setting.

Summary for the Single-Item Setting We solve the promotion optimization problem for a single item using the $App(1)$ approximation. This approximation linearizes the objective solution by computing the sum of the marginal contributions of each promotion separately. The following properties hold:

- The formulation is integral, i.e., one can solve the problem efficiently by considering the Linear Programming (LP) relaxation.
- Under two general demand models which are discussed below (multiplicative and additive), we derive a parametric worst-case bound on the quality of the approximation relative to the optimal profit (the expressions can be found in Theorems 1 and EC.1 in Cohen et al. 2017).
- In many tested instances (calibrated with real data), the approximation yields a solution which is optimal or very close to optimal.

We next discuss the implications of the above summary. Since one can get a solution by solving an LP, the approach is efficient (we can solve large instances in milliseconds). Consequently, the retailer can use this approach in practical settings. The approach works for a general demand function and for any objective function. If we further impose some structure on the demand function, we can derive a parametric bound on the quality of the approximation. We do so by considering two general classes of demand functions:

1. *Multiplicative Demand:*

$$d_t(p_t, p_{t-1}, \ldots, p_{t-M}) = f_t(p_t) \cdot g_1(p_{t-1}) \cdot g_2(p_{t-2}) \cdot \ldots \cdot g_M(p_{t-M}), \tag{4.5}$$

that is, the demand function (of the item under consideration) can be written as the product of M functions that each depends on a single price. Note that since we consider a single-item setting, demand does not depend on the prices of other items. The class of demand functions in (4.5) includes the log-log and log-linear functions, which are commonly-used in retail.

2. *Additive Demand:*

$$d_t\big(p_t, p_{t-1}, \ldots, p_{t-M}\big) = f_t(p_t) + g_1(p_{t-1}) + g_2(p_{t-2}) + \ldots + g_M(p_{t-M}), \tag{4.6}$$

that is, the demand function (of the item under consideration) can be written as the sum of M functions that each depends on a single price. The class of demand functions in (4.6) includes the linear function as a special case.

For these two classes of demand functions, one can derive bounds on the quality of the *App*(1) approximation. These bounds explicitly depend on the problem parameters and depict a very high performance on all the instances we tested (based on retail data). More details can be found in Cohen et al. (2017).

4.4.2 Multiple-Item Setting

In this section, we consider the more general setting where the retailer needs to decide the prices of N interconnected items by solving the (Multi-POP) problem. Recall that in this case, a promotion in item i may have an effect on demand of item $j \neq i$. The cross-item effects on demand can be directly estimated from data. A potential simple approach can be the following: Solve the (Multi-POP) problem by applying the *App*(1) solution approach, i.e., approximate the objective by the sum of the marginal contributions of each item at each period ignoring cross-item effects (as discussed in Sect. 4.4.1). We tested this approach and observed a poor performance (especially in cases where cross-item effects are significant). In particular, it fails to accurately capture cross-item effects and may find a promotion strategy far from optimal. For example, it may suggest to promote two items simultaneously, whereas this pair of items highly cannibalize each other. As a result, one needs to develop an alternative solution approach that can capture cross-item effects, and at the same time, remains efficient. We introduce the following sequence of methods, *App*(κ), for any given $\kappa = 1, 2, \ldots, N$.

- *App*(1) is the approximation applied to (Multi-POP) in a similar fashion as in the single-item setting discussed in Sect. 4.4.1. Specifically, it approximates the objective function by the sum of the marginal contributions of a single promotion for each item and period separately. As previously discussed, in the case of multiple items, it will generally provide a poor performance guarantee relative to the optimal solution.
- *App*(2) is an alternative approximation applied to (Multi-POP) that includes the marginal contributions (same as *App*(1)), as well as the pairwise contributions (i.e., having two items promoted at the same time). *App*(2) is described in full details below.

- More generally, $App(\kappa)$ for any $\kappa = 3, \ldots, N$ is an alternative approximation that includes the marginal contributions, the pairwise contributions, the three-way, four-way, up to k-way contributions.

There is a clear trade-off between simplicity (as well as speed) and performance (in terms of accuracy of the approximation relative to the optimal solution). On one extreme, $App(1)$ is a simple approach that only requires computing the marginal contributions of having a single promotion at a time, but can perform poorly as it does not capture cross-item effects at all. On the other extreme, $App(N)$ is clearly more accurate, as it successfully captures all cross-item effects. However, this benefit comes at the expense of being more complex, as one needs to compute the marginal contribution of each possible combination of items that could be promoted simultaneously. More precisely, it requires us to compute an exponential number of coefficients and to solve an Integer Program (IP) that grows exponentially with the number of items. Note that when $T = 1$ or $M^i = 0\ \forall i$, $App(N)$ is exact as it captures accurately all cross-item effects. Nevertheless, for a general dynamic problem with $T > 1$ periods and non-zero memory parameters, $App(N)$ is still not an exact algorithm, as it approximates the time effects induced by past prices. We next describe $App(2)$ in more details as we will use it subsequently.

As we previously mentioned, $App(2)$ approximates the objective of (Multi-POP) by the sum of unilateral deviations (i.e., having a single promotion at a time) and the pairwise contributions (i.e., having two items promoted simultaneously). More precisely, the approximated objective is:

$$MPOP(\mathbf{p}^0) + \max_{\boldsymbol{\gamma}} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K^i} b_t^{ki} \gamma_t^{ki} + \sum_{i,j:i>j}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K^i} \sum_{\ell=1}^{K^j} b_t^{k\ell ij} \gamma_t^{k\ell ij} \right\}, \tag{4.7}$$

where the coefficients b_t^{ki} and $b_t^{k\ell ij}$ are formally defined in Eqs. (4.8) and (4.9) respectively. We denote the regular price vector (of dimension NT) by $\mathbf{p}^0 = (q^0, \ldots, q^0)$, which means that the regular price is set for all items at all times. The first term, denoted $MPOP(\mathbf{p}^0)$, represents the total profit generated by all items throughout the selling season, without any promotion. The second term captures all the marginal contributions of having a single promotion, i.e., for one item in one time period. More precisely, we define the price vector $\mathbf{p}_t^{kj}$ (of dimension NT) as follows:

$$(\mathbf{p}_t^{kj})_{\tau,i} = \begin{cases} q^{kj}, & \text{if } \tau = t \text{ and } i = j, \\ q^0, & \text{otherwise.} \end{cases}$$

In other words, the vector $\mathbf{p}_t^{kj}$ has the promotion price q^{kj} for item j at time t, and the regular price q^0 (no promotion) is used for all remaining periods for item j, and for all other items at all times. The coefficient b_t^{kj} is then given by:

$$b_t^{kj} = MPOP(\mathbf{p}_t^{kj}) - MPOP(\mathbf{p}^0), \tag{4.8}$$

and represents the marginal contribution in total profit of having a single promotion for item j at time t, using price q^{kj}.

The third term in Eq. (4.7) represents all the pairwise contributions of having two items on promotion at the same time. More precisely, we define the price vector (of dimension NT) $\mathbf{p}_t^{k\ell ju}$ for all pairs of items $j > u$ as follows:

$$(\mathbf{p}_t^{k\ell ju})_{\tau,i} = \begin{cases} q^{kj}, & \text{if } \tau = t \text{ and } i = j, \\ q^{\ell u}, & \text{if } \tau = t \text{ and } i = u, \\ q^0, & \text{otherwise.} \end{cases}$$

In other words, the vector $\mathbf{p}_t^{k\ell ju}$ uses the promotion price q^{kj} for item j at time t, the promotion price $q^{\ell u}$ for item u at time t, and the regular price q^0 for items j and u in all remaining periods, and for all other items at all times. The coefficient $b_t^{k\ell ju}$ is given by:

$$b_t^{k\ell ju} = MPOP(\mathbf{p}_t^{k\ell ju}) - MPOP(\mathbf{p}_t^{kj}) - MPOP(\mathbf{p}_t^{\ell u}) + MPOP(\mathbf{p}^0), \tag{4.9}$$

and represents the marginal pairwise contribution in total profit of having two simultaneous promotions. Finally, to make the formulation consistent, we should ensure that when both items i and j are on promotion, we count the pairwise contribution but also both unilateral deviations, i.e., for each pair of items i and $j < i$, $\gamma_t^{ki} = \gamma_t^{\ell j} = 1$ if and only if $\gamma_t^{k\ell ij} = 1$ for each t and k, ℓ. One can encode this set of conditions by incorporating the following constraints into the formulation for each pair of items i, $j < i$, each t, and each promotion prices q^{ki} and $q^{\ell j}$:

$$\begin{aligned} \gamma_t^{k\ell ij} &\le \gamma_t^{ki}, \\ \gamma_t^{k\ell ij} &\le \gamma_t^{\ell j}, \\ \gamma_t^{k\ell ij} &\ge 0, \\ \gamma_t^{k\ell ij} &\ge \gamma_t^{ki} + \gamma_t^{\ell j} - 1. \end{aligned}$$

When maximizing the objective of the approximated problem in Eq. (4.7), the decisions are the binary variables γ. Specifically, there is one such variable for each item/time/price (i.e., there are $NT(K+1)$ variables, assuming for simplicity that $K^i = K \; \forall i$), and one such variable for any pair of items $i > j$ at each time/price (i.e., $(N(N-1)/2)TK^2$ variables). As mentioned, for $App(N)$, this number grows exponentially with N and K and hence, it may not be practical to go beyond $App(3)$ or $App(4)$. We next summarize the main results for the multiple-item setting.

Summary for the Multiple-Item Setting We solve the promotion optimization problem for multiple items by using the $App(2)$ approximation. The following properties hold[6]:

- Assuming that the cross-item effects for each item are additively separable, i.e.,

$$d_t^i(p_t^i, p_{t-1}^i, \ldots, p_{t-M^i}^i, \mathbf{p}_t^{-i}) = h_t^i(p_t^i, p_{t-1}^i, \ldots, p_{t-M^i}^i) + \sum_{j \neq i} H_t^{ji}(p_t^j), \tag{4.10}$$

  then $App(2) = App(3) = \ldots = App(N)$.
- If we further assume that the function $h_t^i(p_t^i, p_{t-1}^i, \ldots, p_{t-M^i}^i)$ is additively separable for each item, i.e.,

$$h_t^i(p_t^i, p_{t-1}^i, \ldots, p_{t-M^i}^i) = f_t^i(p_t^i) + g_1^i(p_{t-1}^i) + \ldots + g_{M_i}^i(p_{t-M^i}^i), \tag{4.11}$$

  then the $App(2)$ solution is optimal.
- Consider the class of demand models in (4.10) and $K^i = 1$ (i.e., a single promotion price). For substitutable items, the $App(2)$ formulation can be solved efficiently in the absence of business rules (i.e., the formulation is always integral and can be solved as an LP).
- Under two general demand models (multiplicative and additive price dependence), we derive a parametric bound on the quality of the approximation relative to the optimal profit.
- In many tested instances, the approximation yields a solution which is optimal or very close to optimal.

We next discuss the implications of the above summary. Interestingly, for demand functions with additively separable cross-item effects (several demand models satisfy this property), it is sufficient to consider $App(2)$ as opposed to including higher-order terms. In the special case where each item can take two prices, the $App(2)$ approximation can be solved efficiently when all items are substitutable. Having two prices is common in practice as the promotion price is often negotiated upfront with the manufacturer. In the more general case, where the retailer can choose among several promotion prices, we observed computationally that one can still solve the IP in low runtimes for realistic-size instances. It is worth mentioning that for most categories of supermarket items, the products within a category are either independent (i.e., no cross-item effects) or substitutable. In particular, for categories such as coffee, tea, and chocolate, we could not find any complementarity effects in the data we analyzed. Note also that even if some of the products are complements, we observed by extensive testing that solving the relaxation of the $App(2)$ formulation yields an optimal integer solution very often. More details on such computational tests are presented in Cohen et al. (2018).

[6]More details can be found in Cohen et al. (2017, 2018).

4.5 Insights and Practical Impact

In this section, we summarize the insights we could draw by solving the (Multi-POP) problem using our solution approach. We then describe how to concretely apply our model to a real-world retail setting.

4.5.1 Insights

We briefly discuss several insights that were drawn by using our promotion optimization model. Very often, retailers want to infer the impact of promoting different items at different time periods. Our solution approach can easily be used to test various promotion strategies, so as to reach a better understanding of the impact of retail promotions. As mentioned, several economic factors are present in the context of our problem: cross-item effects on demand, the post-promotion dip effect, seasonality, and the presence of business rules. It is clearly valuable for retailers to learn the tradeoffs between these different effects and to understand how they impact promotion decisions. Our model can help retailers to deepen their knowledge of the following topics:

- **Understanding the Structure of Cross-Item Effects.** In a given category of items, retailers need to decide the price promotions by accounting for cross-item effects on demand. Using our model, retailers can infer the impact of promoting a specific item on the demand level of each item in the category. This can ultimately allow retailers to carefully decide which set of items should be promoted simultaneously, and which should not. For example, when two (or more) items have strong substitution effects (i.e., promoting an item increases its own sales but also decreases the sales of other items), retailers should not promote those items simultaneously. More details on such insights can be found in Cohen et al. (2018).
- **Inferring the Strength of the Post-Promotion Dip Effect.** It is well-known that promoting a FMCG product induces a boost in its current demand, as well as a potential decrease in its future demand, due to the stockpiling behavior of consumers and the promotion fatigue effect. The strength of the post-promotion dip effect can vary significantly depending on the category under consideration. For example, in Cohen et al. (2017), the authors found that the number of past prices that affect current demand (which is one possible way to measure the post-promotion dip effect) highly depends on the item and on the category. For example, the post-promotion dip effect tends to be weaker for perishable products and for luxury/expensive brands, as expected.
- **Identifying the Presence of a Loss-Leader Effect.** The loss leader is a common phenomenon in which one item is priced below its cost to extract significant profits on complementary items (see, e.g., Hess and Gerstner 1987). It is reported in Cohen et al. (2018) that the model considered in this chapter can identify

the presence of a loss-leader effect. This can be a very important information for retailers that can use one (or more) items to profitably deploy a loss-leader strategy.

- **Learning the Impact of the Business Rules.** As discussed in Sect. 4.4, retailers can easily solve several instances of the problem, with and without the presence of some of the business rules. Consequently, this allows retailers to quantify the profit impact of relaxing some of the business requirements. This can ultimately help them decide which vendor funds to accept and under what terms.

In practice, retailers often solve the (Multi-POP) problem for large-scale instances that involve a number of different factors. It does not seem possible for managers, as experienced as they are, to understand and anticipate the impact of all the conflicting tradeoffs. Using an optimization tool calibrated with data can take into account all the different tradeoffs and compute a close-to-optimal solution for the promotion planning problem.

4.5.2 Practical Impact

We next consider a concrete application of the (Multi-POP) optimization problem. We propose a generic process that can be used by retailers who seek to improve promotion planning decisions. This process consists of the following five steps:

1. Data collection, cleaning, and aggregation
2. Store and product clustering
3. Demand estimation
4. Optimization and sensitivity analysis
5. Quantifying the impact.

We next describe each step in more detail.

Data Collection, Cleaning, and Aggregation The first step is to collect and store the relevant data. In our context, retailers need to simply collect the data from past transactions. Each observation typically includes the store, the date/time, the items purchased, the prices, the promotion vehicles that were used, as well as various features of the item (brand, size, flavor, etc.). After gathering a large enough dataset, one needs to carefully clean the data and perform the appropriate aggregations. Various techniques exist for cleaning and aggregating data but this is beyond the scope of this chapter (see, e.g., the book by Han et al. 2011). At a high level, one wants to deal with the missing data, remove outliers, and perform basic statistical tests. Once the data is cleaned, one needs to decide the level of aggregation. Depending on the context, one can perform the analysis at the brand, item, or category level. Similarly, one can aggregate the data at the week, day, or hour level. Once the data is cleaned and aggregated at the right level, one can start using it for estimation and prediction purposes. For example, in Cohen et al. (2017), the authors decided to aggregate the data at the brand-week level.

Store and Product Clustering In many retail settings, available historical data can be sparse. As a result, one needs to combine the data from multiple sources in order to obtain more reliable forecasts. Two common techniques widely used in retail consist of merging several stores together or clustering similar products. The idea is to use the data from several stores that share similar features (e.g., geographical location, size, management team). Similarly, items from the same brand (e.g., different sizes or flavors) can often be clustered together to improve the prediction accuracy of the models.

Demand Estimation This step is the actual first stage of using our promotion optimization model. As an input to the optimization, one first needs to estimate the demand models. The modeler has several degrees of freedom: choice of the demand function (e.g., log-log, log-linear), selection of the dependent variables, choice of the instrumental variables (if any), and choice of the estimation procedure. In many applications, one can simply run a linear regression (e.g., ordinary or weighted least squares, ridge regression, lasso). The typical process also includes splitting the data for out-of-sample testing. The demand estimation step is completed once the prediction model yields a high out-of-sample prediction accuracy. In practice, one needs to test different models and assumptions in order to reach a good and robust prediction model. In Cohen et al. (2017), the authors present a prediction model for two coffee brands based on using ordinary least squares to predict a log-log model that includes past prices, weekly seasonality, and trend effect. The resulting out-of-sample R^2 (resp. MAPE) was around 0.90 (resp. 0.11).

Optimization and Sensitivity Analysis Once the demand models are accurately estimated from data, one can use them as an input to the (Multi-POP) problem. The retailer also needs to specify the business rules that need to be satisfied, the number of time periods in the selling season, and the cost of each item at each period. At this point, one can use the $App(2)$ approximation method presented in Sect. 4.4.2 to solve the problem. As discussed before, this yields a near-optimal solution by computing the price promotions of all items during each period of the selling season. Usually, retailers want to check the robustness of the solution prior to a potential implementation. To this end, one can re-solve the (Multi-POP) problem by perturbing several input parameters (e.g., estimated demand coefficients, business rules' parameters). If the suggested solution appears to be robust with respect to variations in the problem input, this provides a higher confidence on the validity of the solution.

Quantifying the Impact The last step is to assess the potential impact of the entire process. For example, one can compare the simulation results obtained by using the optimized promotion prices relative to the profit generated using the original promotion prices set by the retailer. In our experience, by applying our model to several retailers, we observed a profit improvement of 2–9%, depending on the product category and the store under consideration.

4.6 Conclusions

Retail promotions are important decisions faced by most retailers. Promoting the right set of items at the right time using the right price points can have a significant impact on retailers' bottom line. In settings such as supermarkets, retailers need to simultaneously decide the price promotions for multiple items throughout the selling season. Historically, many retailers were designing their promotion strategies based on past experience and on trial-and-error processes. The unprecedented volume of available data has now changed the picture. Using past data, retailers can improve demand forecasting accuracy. They can also exploit the data to develop quantitative tools for promotion planning. In particular, the combination of data analytics and optimization allows retailers to decide promotions in a more systematic and profitable fashion. In this chapter, we considered a retailer selling FMCG products who needs to decide the (mass) price promotions for all items in a category. We first formulated the problem as a nonlinear integer program. This formulation holds under general demand functions estimated from data and includes several practical business rules which typically apply to price promotions. Given that the resulting formulation is hard to solve, we presented an approximate solution approach. This approach can solve the problem in short timeframes, and admits a parametric worst-case bound on the quality of the approximation. We first considered the single-item setting, and then extended the presentation to the more general instance with multiple items. In each case, we presented the model, the approximate solution approach, and its analytical and practical properties. Overall, the methods presented in this chapter run fast and provide a near-optimal solution for many tested instances (calibrated with real data).

We then summarized an application of this model to a real-world setting. In particular, we proposed a beginning-to-end process for retailers that consists of five steps: (1) data collection, cleaning, and aggregation, (2) store and product clustering, (3) demand estimation, (4) optimization and sensitivity analysis, and (5) quantifying the impact. By following these steps, retailers can potentially improve their promotion planning process. In our own experience, we observed a profit improvement of 2–9%, which is a significant impact in the retail industry.

While most of the results presented in this chapter are borrowed from previous publications (mainly from Cohen et al. 2017, 2018), it provides a summary of this line of research by presenting the two complementary studies in a single report. This chapter has focused on the mass pricing promotion optimization problem faced by a retailer who sells FMCG products. As mentioned in Sect. 4.1, several alternative promotion problems are also important in the retail industry. Interesting research directions can be the development of new data-driven decision tools for those practical retail problems.

References

Ahn H, Gümüş M, Kaminsky P (2007) Pricing and manufacturing decisions when demand is a function of prices in multiple periods. Oper Res 55(6):1039–1057
Ailawadi KL, Gedenk K, Lutzky C, Neslin SA (2007) Decomposition of the sales impact of promotion-induced stockpiling. J Mark Res 44(3):450–467
Andrews M, Luo X, Fang Z, Ghose A (2015) Mobile ad effectiveness: hyper-contextual targeting with crowdedness. Market Sci 35(2):218–233
Arora N, Dreze X, Ghose A, Hess JD, Iyengar R, Jing B, Joshi Y, Kumar V, Lurie N, Neslin S, et al (2008) Putting one-to-one marketing to work: personalization, customization, and choice. Mark Lett 19(3–4):305
Aviv Y, Pazgal A (2008) Optimal pricing of seasonal products in the presence of forward-looking consumers. Manuf Serv Oper Manag 10(3):339–359
Baardman L, Panchamgam K, Perakis G (2017) Pass-through constrained vendor funds for promotion planning. Working paper, MIT, Chennai
Baardman L, Cohen MC, Panchamgam K, Perakis G, Segev D (2018) Scheduling promotion vehicles to boost profits, Manag Sci 65(1), published online April 2018
Balinski ML (1970) On a selection problem. Manag Sci 17(3):230–231
Bertsimas D, Shioda R (2009) Algorithm for cardinality-constrained quadratic optimization. Comput Optim Appl 43(1):1–22
Besbes O, Lobel I (2015) Intertemporal price discrimination: structure and computation of optimal policies. Manag Sci 61(1):92–110
Bienstock D (1996) Computational study of a family of mixed-integer quadratic programming problems. Math Program 74(2):121–140
Bitran G, Caldentey R (2003) An overview of pricing models for revenue management. Manuf Serv Oper Manag 5(3):203–229
Blattberg RC, Neslin SA (1990) Sales promotion: Concepts, methods, and strategies. Prentice Hall, Upper Saddle River
Cachon GP, Swinney R (2009) Purchasing, pricing, and quick response in the presence of strategic consumers. Manag Sci 55(3):497–511
Campo K, Gijsbrechts E, Nisol P (2000) Towards understanding consumer response to stock-outs. J Retail 76(2):219–242
Caro F, Gallien J (2012) Clearance pricing optimization for a fast-fashion retailer. Oper Res 60(6):1404–1422
Chen Y, Farias VF (2015) Robust dynamic pricing with strategic customers. In: Proceedings of the sixteenth ACM conference on economics and computation, pp 777–777
Cohen MC, Gupta S, Jeremy, Perakis G (2016) An efficient algorithm for dynamic pricing using a graphical representation. Working paper
Cohen MC, Leung NHZ, Panchamgam K, Perakis G, Smith A (2017) The impact of linear optimization on promotion planning. Oper Res 65(2):446–468
Cohen MC, Kalas J, Perakis G (2018) Optimizing promotions for multiple items in supermarkets. Working paper
Cooper LG, Baron P, Levy W, Swisher M, Gogos P (1999) Promocast: a new forecasting method for promotion planning. Market Sci 18(3):301–316
Corsten D, Gruen T (2004) Stock-outs cause walkouts. Harvard Bus Rev 82(5):26–28
Elmaghraby W, Keskinocak P (2003) Dynamic pricing in the presence of inventory considerations: research overview, current practices, and future directions. Manag Sci 49(10):1287–1309
Felgate M, Fearne A (2015) Analyzing the impact of supermarket promotions: a case study using tesco clubcard data in the UK. In: The Sustainable Global Marketplace, Springer, Berlin, pp 471–475
Ferreira KJ, Lee BHA, Simchi-Levi D (2015) Analytics for an online retailer: demand forecasting and price optimization. Manuf Serv Oper Manag 18(1):69–88

Foekens EW, SH Leeflang P, Wittink DR (1998) Varying parameter models to accommodate dynamic promotion effects. J Econ 89(1):249–268
Fong NM, Fang Z, Luo X (2015) Geo-conquesting: Competitive locational targeting of mobile promotions. J Mark Res 52(5):726–735
Frank M, Wolfe P (1956) An algorithm for quadratic programming. Nav Res Logist (NRL) 3(1–2):95–110
Gedenk K, Neslin SA, Ailawadi KL (2006) Sales promotion. In: Retailing in the 21st century, Springer, Berlin, pp 345–359
Grossmann IE (2002) Review of nonlinear mixed-integer and disjunctive programming techniques. Optim Eng 3(3):227–252
Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier, Waltham
Heilman CM, Nakamoto K, Rao AG (2002) Pleasant surprises: consumer response to unexpected in-store coupons. J Mark Res 39(2):242–252
Hemmecke R, Köppe M, Lee J, Weismantel R (2010) Nonlinear integer programming. In: Jünger M, Liebling TM, Naddef D, Nemhauser GL, Pulleyblank WR, Reinelt G, Rinaldi G, Wolsey LA (eds) 50 Years of integer programming 1958–2008, Springer, Berlin, pp 561–618
Hess JD, Gerstner E (1987) Loss leader pricing and rain check policy. Mark Sci 6(4):358–374
Jagabathula S, Mitrofanov D, Vulcano G (2018) Customized individual promotions: model, optimization, and prediction. Working paper, New York University, Stern School of Business, New York
Levin Y, McGill J, Nediak M (2010) Optimal dynamic pricing of perishable items by a monopolist facing strategic consumers. Prod Oper Manag 19(1):40–60
Levina T, Levin Y, McGill J, Nediak M (2009) Dynamic pricing with online learning and strategic consumers: an application of the aggregating algorithm. Oper Res 57(2):327–341
Liu Y, Cooper WL (2015) Optimal dynamic pricing with patient customers. Oper Res 63(6):1307–1319
Macé S, Neslin SA (2004) The determinants of pre- and post-promotion dips in sales of frequently purchased goods. J Mark Res 41:339–350
Martínez-Ruiz MP, Mollá-Descals A, Gómez-Borja MA, Rojo-Álvarez JL (2006) Assessing the impact of temporary retail price discounts intervals using SVM semiparametric regression. Int Rev Retail Distrib Consum Res 16(02):181–197
Mersereau AJ, Zhang D (2012) Markdown pricing with unknown fraction of strategic customers. Manuf Serv Oper Manag 14(3):355–370
Nielsen (2015) The path to efficient trade promotions. http://www.nielsen.com/us/en/insights/reports/2015/the-path-to-efficient-trade-promotions.html
Nijs V, Misra K, Anderson ET, Hansen K, Krishnamurthi L (2010) Channel pass-through of trade promotions. Market Sci 29(2):250–267
Nocedal J, Wright SJ (2006) Sequential quadratic programming. Springer, Cambridge
Özer Ö, Phillips R (2012) The Oxford handbook of pricing management. Oxford University Press, Oxford
Padberg M (1989) The Boolean quadric polytope: some characteristics, facets and relatives. Math Prog 45(1):139–172
Pekgün P, Menich RP, Acharya S, Finch PG, Deschamps F, Mallery K, Van Sistine J, Christianson K, Fuller J (2013) Carlson Rezidor hotel group maximizes revenue through improved demand management and price optimization. Interfaces 43(1):21–36
Rhys J (1970) A selection problem of shared fixed costs and network flows. Manag Sci 17(3):200–207
Sherali HD, Adams WP (1998) A reformulation-linearization technique for solving discrete and continuous nonconvex problems, vol 31. Springer, Berlin
Silva-Risso JM, Bucklin RE, Morrison DG (1999) A decision support system for planning manufacturers' sales promotion calendars. Market Sci 18(3):274–300
Su X (2010) Intertemporal pricing and consumer stockpiling. Oper Res 58(4-part-2):1133–1147
Talluri KT, Van Ryzin GJ (2006) The theory and practice of revenue management, vol 68. Springer Science and Business Media, Boston

Vakhutinsky A, Kushkuley A, Gupte M (2012) Markdown optimization with an inventory-depletion effect. J Revenue Pricing Manag 11(6):632–644

Van Donselaar K, Van Woensel T, Broekmeulen R, Fransoo J (2006) Inventory control of perishables in supermarkets. Int J Prod Econ 104(2):462–472

Van Heerde HJ, Gupta S, Wittink DR (2003) Is 75% of the sales promotion bump due to brand switching? No, only 33% is. J Mark Res 40(4):481–491

Yin R, Aviv Y, Pazgal A, Tang CS (2009) Optimal markdown pricing: implications of inventory display formats in the presence of strategic customers. Manag Sci 55(8):1391–1408

Yuan H, Gómez MI, Rao VR (2013) Trade promotion decisions under demand uncertainty: a market experiment approach. Manag Sci 59(7):1709–1724

Zhang D, Cooper WL (2008) Managing clearance sales in the presence of strategic customers. Prod Oper Manag 17(4):416–431

Chapter 5
Optimization of Operational Decisions in Digital Advertising: A Literature Review

Narendra Agrawal, Sami Najafi-Asadolahi, and Stephen A. Smith

Abstract The digital advertising industry has witnessed an impressive explosion since its inception. With Internet advertising revenues already at over $200 billion, the projections are for continued growth. The format and technological underpinnings of digital advertising make it a fascinating subject of study for practitioners and academics alike, and distinguish it from traditional advertising in many ways. In contrast to traditional advertising, online advertising offers significantly more channels, a lower cost alternative, greater targeting and personalization capabilities, and dynamic pricing capabilities. While these digital technologies offer unprecedented opportunities to marketers to maximize the ROI on their marketing budgets, the data-rich environment also presents a unique set of managerial, operational, and intellectual challenges. In this chapter, we introduce the reader to some of these challenges. Our goal is to identify some of the salient operational challenges facing advertisers and publishers in this digital advertising environment, and summarize the state of the art of the published research that attempts to address these challenges. Our hope is that practitioners, academics, and graduate students will find this to be a valuable resource in their various endeavors.

Keywords Digital supply chain · Advertisers · Publishers · Ad exchanges · Display advertising · Search advertising · Inventory allocation · Ad scheduling · Ad pacing · Ad pricing

5.1 Introduction

"In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients.

N. Agrawal · S. Najafi-Asadolahi (✉) · S. A. Smith
ISA Department, Leavey School of Business, Santa Clara University, Santa Clara, CA, USA
e-mail: nagrawal@scu.edu; snajafi@scu.edu; ssmith@scu.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_5

Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it." This famous quote by Herbert Simon in 1971 very aptly captures a key challenge faced by advertisers—grabbing attention of the target consumers in an effective manner. This is particularly true in today's era of sensory and information overload that characterizes all markets and consumers today. Digital advertising has created unprecedented opportunities and challenges for companies and consumers alike in this rapidly evolving arena.

Legend has it that the first digital advertisement appeared in 1994 as a HotWired banner ad from AT&T, which asked: "Have you ever clicked your mouse right here? You will." While conversion of the traditional ads to the electronic media format as faxes and emails had already begun by then, this ad was a sign of things to come. The digital advertising industry has witnessed an impressive explosion since then. According to the PwC IAB Report of 2017, out of a total advertising spend of nearly \$200 billion, Internet advertising revenues in the USA totaled \$88 billion for the year 2017, which was a 21.4% increase over 2016. The figure was about \$49.5 billion during the first 6 months of 2018, which represented a 23.1% increase over the same period in 2017. Industry experts expect 2018 revenues to exceed 2017 revenues by a significant margin. About 46% of these revenues derive from search-related ads, 46% from display (32% banner ads and 14% video ads), and the rest from other categories such as classified ads, lead generation, and audio ads.

The number of channels available for traditional advertising, even aggregated across print media and other platforms, is small. In comparison, with over a billion websites and nearly five billion mobile device users, the number of channels for digital advertising is potentially very large. Advertising in traditional formats can be very expensive. For example, broadcasting a 30 s commercial ad on CNN typically costs between \$6000 and \$7000 during the day and could go up to \$30,000 in prime time. Moreover, the effectiveness of reaching the intended audience can be low. In comparison, posting an ad on CNN's website can cost as little as 2.5 cents each time the ad is displayed to a viewer. As compared to the traditional print and media advertising, the Internet enables a considerably more granular targeting of the consumer that advertisers intend to reach. This allows ads to be personalized and targeted to specific individuals, platforms, and contexts, delivered at any time and at any location, updated frequently in a cost-effective manner, and priced almost in real-time. For example, over 100 billion impressions are now served daily, with hundreds of variables being used to qualify the viewer of each impression (publisher site, URL, ad characteristics, demographic characteristics, etc.). The total number of resulting combinations of values that characterize impression opportunities is in the millions of billions, a level of segmentation unimaginable in the traditional advertising environment. Digital advertising also offers the opportunity to update the ads served based on the users' interactions with what has been previously displayed.

Although technology offers unprecedented opportunities to marketers to maximize the ROI on their marketing budgets, the data-rich environment presents a unique set of managerial, operational, and intellectual challenges. It is some of these challenges that we intend to explore throughout this chapter. Our goal is to identify some of the salient operational challenges facing advertisers and publishers in this digital advertising environment and summarize the state of the art of the published research that attempts to address these challenges. Our hope is that practitioners, academics and graduate students will find this to be a valuable resource in their various endeavors.

Our focus is on operational problems in the digital advertising industry which lend themselves to optimization techniques. We do not, for example, focus on the vast literature (much of it in the domain of marketing journals) that aims to characterize user behavior in online settings, or estimate various parameters that would be relevant to subsequent decision making. For example, papers that attempt to estimate click-through rates and display ad effectiveness including Chatterjee et al. (2003), Rutz and Bucklin (2012), Johnson et al. (2017). Ghose and Yang (2009), Agarwal et al. (2011), and Narayanan and Kalyanam (2015) examine the role of an ad's location on a page on viewer actions in the context of sponsored search. These and other related articles are discussed in Bucklin and Hoban (2017). Even within our scope, ours is not the only survey paper on this vast topic. We want to draw the reader's attention to excellent reviews by Korula et al. (2016), Choi et al. (2017), and Wang et al. (2017). In Korula et al. (2016), the focus of the review is on the mechanisms for selling display advertising, which include reservation contracts and real-time bidding on exchanges. They discuss the key terms included typically in reservation contracts, and theoretical as well as practical challenges in specifying these terms. They also discuss some of the issues in exchange markets—design of the auction, the fee structure for the exchange, and the role of intermediaries. Finally, they also point to challenges faced when contracts and bidding happen simultaneously,[1] including adverse selection and fair allocation, and discuss approaches to address such issues. In contrast, Choi et al. (2017) focus on issues related to display advertising only, but cover a broader range of topics, and do so from the perspective of advertisers, publishers, and intermediaries (DSPs, SSPs, ad exchanges, ad networks, and data aggregators). In particular, their discussion on intermediaries is very interesting. Intermediaries possess information spanning advertisers and publishers and can, therefore, assess the competitive landscape more completely than a single advertiser or publisher. However, their incentives for sharing data may differ from those of the advertisers or publishers. Therefore, while intermediaries can lead to improved market efficiency, publishers and advertisers must design appropriate contracts to address the incentive incompatibility issues. Wang et al. (2017) focus primarily on real-time bidding and provide an overview of the infrastructure and solution algorithms related to topics such as user response prediction, bid landscape forecasting, bidding algorithms, revenue optimization, statistical arbitrage, dynamic pricing, and ad fraud detection.

[1]For example, if the value of impressions to advertisers who bid on the exchange and who buy contracts is correlated, the "valuable" impressions will be cherry-picked by advertisers on the exchange.

While these survey papers cover a broad range of topics, in contrast, our focus is on a deeper description of the published literature on optimization of some of the most important operational decisions involved in digital advertising. In other words, our focus is primarily on the methodological and algorithmic aspect of decision making in this environment. In this sense, our survey provides a nice complement to those by Korula et al. (2016), Choi et al. (2017), and Wang et al. (2017).

We begin with a brief discussion on the literature on traditional advertising in Sect. 5.2. We then describe the online advertising ecosystem as a digital supply chain and overview the industry structure in Sect. 5.3. A detailed description of the key operational problems facing each major entity in this supply chain and a review of optimization models for these problems are presented in Sect. 5.4. We conclude with some observations about open research opportunities in Sect. 5.5.

5.2 Traditional Advertising

We begin with a short discussion of traditional (i.e., nondigital) promotions. The published literature related to this topic is vast, and thus it does not seem practical or desirable to attempt to survey it in this chapter. Instead, a few papers that address certain specific issues will be discussed to provide insights that may apply to digital promotions as well. Traditionally, promotions have targeted retail sales and services across a wide range of merchandise including grocery, package goods, housewares, apparel, auto parts and home improvement, as well as many types of specialty items. Traditional promotions can be planned and funded by the retailer, by the brand, or by the manufacturer, as well as through cooperative arrangements among these parties. With the growth of digital advertising, traditional and digital promotions may overlap, either compete with or complement each other, when promoting the same or similar merchandise. One of the key characteristics that distinguishes the various types of promotions is their ability to target certain customer groups or types, or in some cases to target specific customers.

Nearly $200 billion was spent on advertising in the USA in 2016, and that number is projected to increase substantially in the coming years. Broadcast television advertising has traditionally been the largest category at $73 billion in 2016, but in 2017 digital advertising is projected to surpass TV advertising in expenditure.[2] The relative effectiveness of TV advertising has long been the subject of some debate. Rubinson (2009) applied various hypothesis tests to empirical data sets to conclude that TV advertising is superior to other advertising media for generating brand awareness, and also that TV advertising effectiveness has not declined over time, despite consumers' increased ability to skip through commercials in recorded programs, or otherwise multi-task. This study did not include digital advertising specifically, and found that targeted print promotions were the best medium for generating customer intent to purchase.

[2]Source: Statistica, the Statistics Portal.

While TV advertising is the largest component of traditional advertising, broadcast television has only limited ability to target specific consumers. For example, targeting has traditionally been based on the projected viewer type for certain program content, e.g., advertising beer and pickup trucks during football games. However, with the explosion in the number of cable channels with their known subscriber databases, TV advertising can now be targeted more specifically to groups of viewers. The effectiveness, of course, still depends on whether viewers actually watch the ads.

Traditional retail promotions that are associated with specific merchandise items are often referred to as "deals," and have been the subject of many marketing studies. An excellent survey of the state of the art at the time is in Blattberg et al. (1995). This article found empirical studies that supported a number of hypotheses, while others remained open questions. The ones clearly supported by empirical evidence, according to this article, include: (1) product specific promotions (price reductions as well as advertising) lead to a short-term sales spike, (2) higher market share brands experience less response to deals, due to less switching behavior by consumers, (3) increasing the frequency of deals lowers the deal spike. Other questions where the empirical evidence was less conclusive include: (1) What is the interaction between feature advertising and in store displays? (2) How does advertising affect store traffic? (3) What are the cross-product effects of promotions (these seem to vary by product type)? and (4) What are the possible negative long-term effects of promotions?

A subsequent study of the long-term effects of periodic promotions using a database for consumer package goods is in Jedidi et al. (1999). They concluded that a pattern of regular temporary price cuts produced short-term spikes, but reduced total profits over the longer term, and increasing frequency of promotions was uniformly negative. Advertising, on the other hand, was found to have positive long-term effects on profits, but only for some brands. Using different data, Taylor and Neslin (2005) found that a "buy one get one free" type of reward program for frequent shoppers had positive long-term effects.

In an editorial, Levy et al. (2004) argued that American retailers were losing more than $200 billion per year due to temporary markdowns, with as much 78% of all apparel sold by major department store chains on markdown. The article argues that this is partly because the short- and long-term interactions between temporary price cuts, advertising, competition, and other factors are not well understood by retailers. Carpenter and Moore (2008) studied nonprice promotions using panel data to estimate how their effectiveness depends on demographic characteristics.

Many of the questions surrounding the effectiveness of traditional retail promotions and their corresponding interaction effects were revisited in Ailawadi et al. (2009). They point to evidence that in the majority of cases, frequent price promotions have a long-term negative effect. They also raise the issue of interactions between traditional and digital promotions, which appear to have largely unknown effects. The interactions between digital and traditional promotions are also discussed in Grewal et al. (2011) and Lewis and Reiley (2014).

For traditional promotions, major advances in customer targeting have been achieved as a result of the data obtained through retailer loyalty programs. The data collected on individual customers' purchases can be used to target print advertising, catalogs, coupons, and other incentives that are mailed directly to the customer. See Grewal et al. (2011) for a survey of various promotion planning methods associated with loyalty programs. Targeted retailer promotions can also involve the use of digital media in some cases. For example, in a novel combination of in-store sales and digital advertising, retailers can track shoppers' smartphones while they are in the store and then send specific ads to their phones as they pass by the targeted merchandise.

Retailers use several promotional tools inside the store as well. A recent study by Leischnig et al. (2011) considers the effectiveness of store events in promoting brands through creating a positive customer experience. End of aisle displays can increase sales for the items placed there, as well as placing items in more prominent shelf locations, or with larger shelf facings. Of course, these strategies may merely result in the substitution of the prominently displayed items for other similar items in the store. This is one of the points raised in Levy et al. (2004).

A key question is to what extent do the insights obtained for traditional promotions carry over to digital promotions. It seems almost certain that frequent price deals will have negative long-term impacts on profits with digital advertising as well. Similarly, substitution of promoted items for purchases of other nonpromoted items clearly can dilute the effects of digital promotions. There is also good reason to believe that digital media will exacerbate the problem of deal competition, because consumers now have more ways to compare prices and deals. For example, shoppers can use web crawlers to search for the best deals before shopping, compare store prices to the posted online prices, or use their smartphones while shopping to compare prices at competing retailers. Many retailers have promised to match competitors' prices in these situations. There are some indications that a high frequency of nonprice promotions can have negative effects as well, but this needs further study. It seems natural to combine the customer information obtained from traditional loyalty programs with digital advertising, but the effectiveness of this strategy does not appear to have been studied. Thus, the understanding of the interaction effects between traditional and digital promotions is still evolving because the various digital media are evolving so rapidly. The best allocation of advertising budgets across the various digital and traditional media will remain an open question for some time, and provides many opportunities for future research.

5.3 Digital Advertising

The key goal in digital advertising is to find the best *match* between viewers and the advertisers who wish to target them so that the advertisers' objectives are realized effectively. The architecture that helps achieve this supply-demand match, which we call as the *digital advertising supply chain*, is described below, followed by a discussion on the key operational challenges facing each participant in this supply chain.

5.3.1 Digital Advertising Supply Chain

A typical digital advertising supply chain, shown in Fig. 5.1, consists of the *buy side* (*demand side*) and the *sell side* (*supply side*). In this environment, demand consists of a request by an advertiser (e.g., INFORMS) to show its ads to viewers with specific attributes (*targeting*). Depending upon the advertiser's objective, this request may be merely to serve an impression to build awareness (*branding advertiser*) or to seek a particular measurable action by the viewer such as a click or an actual purchase (*performance advertiser*). In addition, the advertiser may seek any viewer or only those who have already visited its website previously (*retargeting*). More sophisticated advertisers attempt to develop strategies for long-term data ownership, integration, analysis and enablement using their own properties, partners or advertising campaigns (Kosorin 2016). Metrics used by advertisers include click-through rate (CTR), conversion rate, ROI, reach,[3] and frequency.

Digital media is merely a fraction of advertisers' overall advertising strategy, along with print media, television, radio, billboards, etc. Larger advertisers might have in-house teams to create and execute their advertisements. However, more often, they work with intermediaries called ad agencies (e.g., AKQA) that perform the advertising processing functions on behalf of the advertisers.

Advertisers or ad agencies typically access ad inventory through a Demand Side Platform (DSP), and place ads using its agency trading desk. A DSP is a system (i.e., software) that allows buyers of digital advertising inventory to manage multiple ad exchange and data exchange accounts through one interface.[4] DSPs allow advertisers to manage their bids for the ads and the pricing for the data that they are layering on to target their audiences while optimizing key performance indicators such as effective cost per click (eCPC) and effective cost per action (eCPA). Examples of DSPs include MediaMath, AOL, Doubleclick, and Rocket Fuel.

The sell-side consists of the publishers who control the digital content and the ad inventory. Prominent media brands (e.g., CNN.com) all the way down to individual bloggers are examples of publishers. These publishers sell their inventory either directly to advertisers (or their agents) using their in-house sales teams, or through ad networks and exchanges. Initially, the premium ad inventory was sold directly and the rest indirectly, but this allocation of inventory has increasingly become dynamic. Similar to a DSP, a supply-side (or sell-side) platform (SSP) is a technology platform that enables publishers to manage their ad spaces and receive revenue.[5] This system allows advertisers to show ads to a targeted audience.

---

[3]Reach is the total number of people that the ad is exposed at least once during a given period.

[4]Source: Wikipedia, 10/6/2017, https://en.wikipedia.org/wiki/Demand-side_platform.

[5]Source: Wikipedia, 10/6/2017, https://en.wikipedia.org/wiki/Supply-side_platform.

SSPs send potential impressions into ad exchanges, where DSPs purchase them on marketers' behalf, depending on specific targeting attributes. Examples of SSPs include PubMatic, AppNexus, OpenX, AOL, and Google's AdX.

Ad agencies can place their ads either through an ad exchange, or through an ad network. Ad networks typically pre-purchase and aggregate inventory from multiple publishers and sell it at a marked up price (Jerath and Savary 2017). They often have a fixed cost per mille (cost per 1000 impressions, or CPM) compared with the changing prices on ad exchanges. Vertical ad networks are transparent regarding the publisher sites on which the ads are posted. Horizontal ad networks do not offer such a transparency (*blind networks*) but offer lower prices.

Ad exchanges, such as Google's Double Click ad exchange and AppNexus, facilitate real-time auctions for ads to be displayed on publisher websites (this process is described later in this chapter). In addition to running the auctions, they also facilitate payments to the publishers, and get paid a fee for providing these services. They add value to the digital supply chain by serving as a clearing house for supply and demand, thereby reducing the need for inefficient negotiations between multiple pairs of publishers and advertisers (Korula et al. 2016). The inherent variability in user traffic can lead to publishers incurring contractually specified penalties for unmet demand, or opportunity cost due to unsold inventory. Consequently, following the principles of revenue management, publishers contract to sell less than the traffic that they forecast, and use ad exchanges to obtain additional revenue from traffic that exceeds what they have contracted for. Finally, exchanges allow for unbundling of impressions into individual ones, which facilitates highly granular targeting, individually customized ads for users, and more efficient allocation of ads.

We note that the digital advertising industry is more complicated than indicated by Fig. 5.1,[6] with many additional participants offering specialized services, and with some players performing overlapping functions. Figure 5.2 describes a snap-

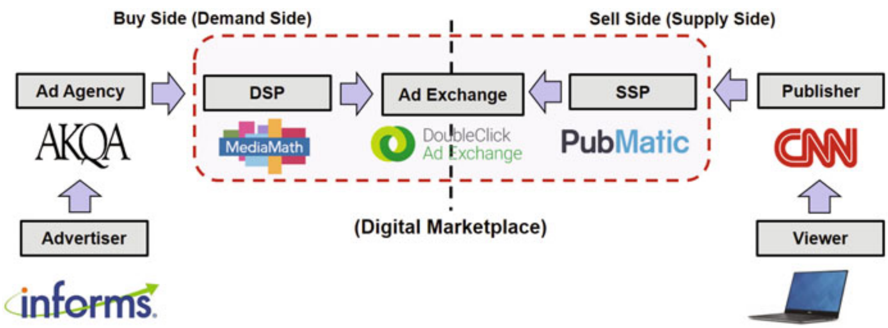

Fig. 5.1 The digital advertising supply chain

[6]Source: http://digitaladblog.com/2015/02/19/online-advertising-ecosystem-explained/.

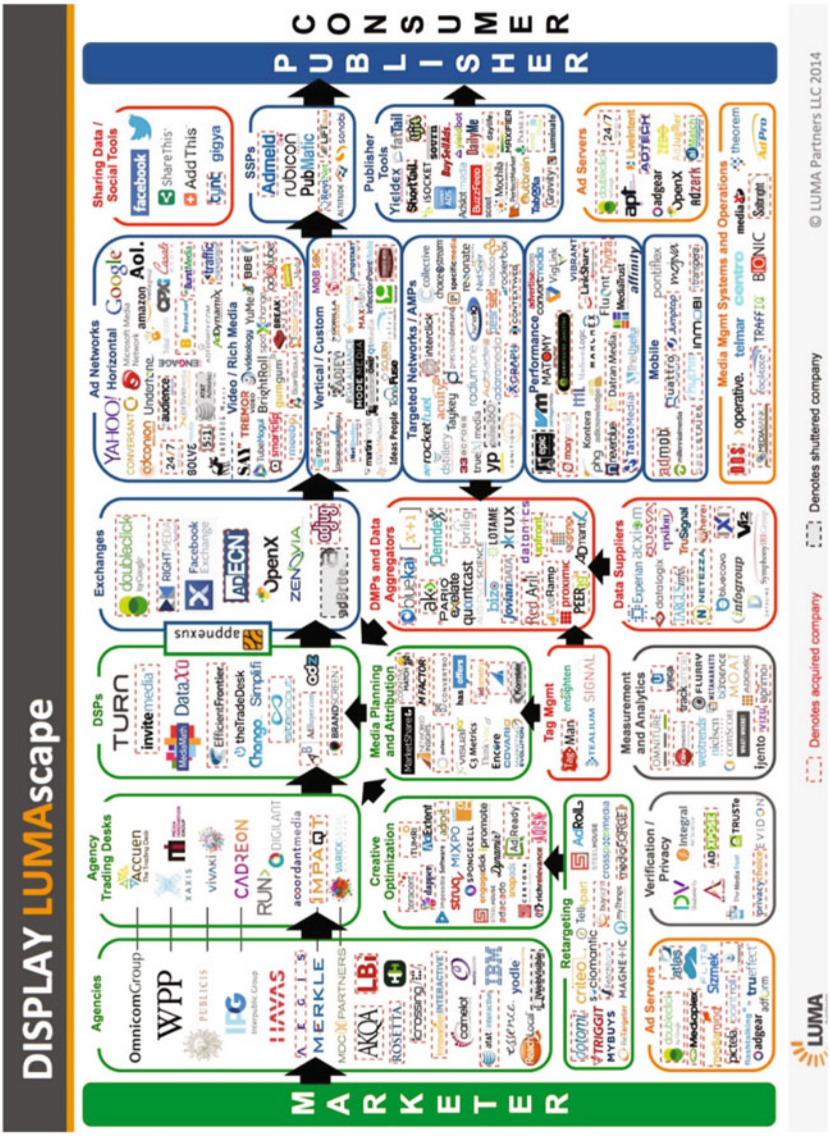

Fig. 5.2 The snapshot of companies in the digital advertising supply chain

shot of some of the main companies in this industry. In particular, there are a variety of companies that specialize in providing data that advertisers and publishers use to augment data that they already possess. For example, advertisers have data about the purchase patterns and behavior of viewers who have visited their websites. However, they can augment their knowledge by combining this information with data about viewers' income, job history, home ownership, mortgage payments, etc. (Choi et al. 2017) to develop better target profiles. These additional data on viewers must be purchased from third parties. Similarly, publishers can use such data to offer better targeting attributes to advertisers. *Data Suppliers*, therefore, perform a very useful function in digital supply chains. *Data Management Platforms* and *Data Aggregators* provide the necessary knowhow and technological platforms to access such data from multiple sources.

5.3.2 Types of Digital Ads

In the digital advertising parlance, two major ad contexts exist, namely, *sponsored search* and *display*. Sponsored search ads, also called as *search ads*, are displayed in response to *queries* from viewers. When we search using a search engine like Google for the keyword *laptops*, the search engine returns the results of the query but also displays ads from vendors interested in selling laptops. If the viewer clicks on such an ad, it may ultimately translate into a sale. In contrast, display ads such as banner ads on a website are served based not on specific queries by a viewer, but on the known or presumed characteristics of the viewer. In fact, in case of display ads, the viewer may be browsing for a completely unrelated activity. In search ads, the goal is to guide the viewer toward a purchase, while display ads are intended to attract more users. Display ads, therefore, can use a variety of alternatives to target viewers—contextual targeting (viewers browsing a contextually relevant site), behavioral targeting (based on the viewers browsing history), or demographic targeting. Given where a viewer is in terms of his or her purchase decision, search is much more competitive and the keywords tend to be very valuable. In contrast, since an impression requires little action from the viewer, display ads tend to be less valuable. For the same reasons, the content of the ad tends to be very different for the two types of ads.

Sponsored search and display ads also use different pricing mechanisms. Search auctions are often cached or hosted auctions. Advertisers pre-specify what they are willing to pay and this information resides on the search engine's server. Thus, the optimization is controlled by the search engine (e.g., Google). In display and in other real-time bidding auctions, the optimization is maintained by the buying technology/platform, and the buyers have more influence on the auction.

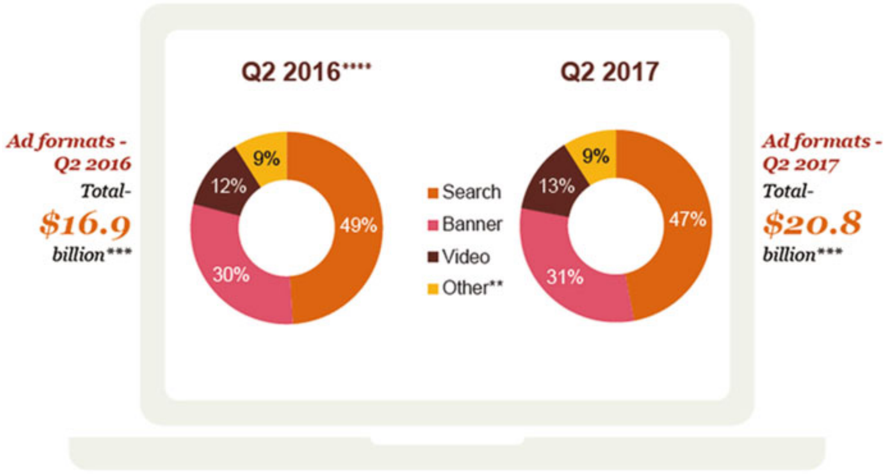

Fig. 5.3 Digital advertising revenue for Q2, 2016 and Q2, 2017

Revenues from search ads dwarf those from other forms of digital advertising (Fig. 5.3),[7] totaling about $19.1 billion during the first 6 months of 2017, while display ads—including banner ads and video ads together—accounted for about $17.6 billion in revenues.

Classified, lead generations and audio ads accounted for about $3.4 billion in revenues. Another way to dissect the market for online advertising is by the device to which the ads are served, e.g., desktop vs. mobile devices. As expected, the volume of ads delivered to mobile devices has increased steadily, and made up about 54% of total online ad revenues in the second quarter of 2017 (about $21.7 billion in the first 6 months of 2017), with a 10-year CAGR of about 15.4%. The split between mobile and desktop ad revenues (for half year) since 2012 is shown in Fig. 5.4.[8]

5.3.3 Types of Contracts in the Digital Supply Chain

Broadly speaking, there are two types of contracts for display ads: *Guaranteed delivery (GD)* contracts and *Non-guaranteed delivery (NGD)* contracts.

In GD contracts, publishers agree to show the ads to a particular number of targeted viewers, and charge advertisers a certain negotiated price based on the cost-per-impression (CPM) or cost per click (CPC) pricing scheme. Publishers are responsible for any shortfall in the promised impressions or clicks. GD contracts are the standard form for selling display ads at premium prices. As a result, the quality

[7] Source: IAB Internet Advertising Revenue Report, December 2017.

[8] See footnote 7.

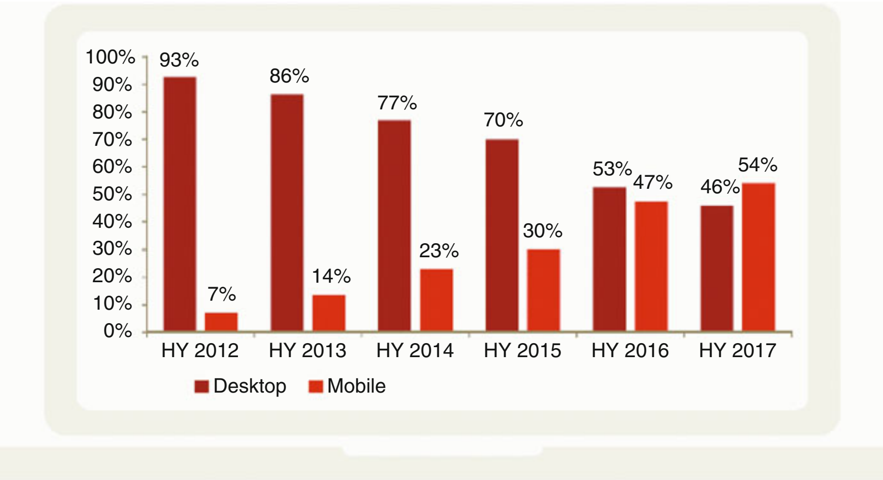

Fig. 5.4 Desktop vs mobile ad revenues

and relevance of viewers in GD contracts are usually higher than NGD contracts. A typical GD contract contains the following elements (Bharadwaj et al. 2010): (1) *Viewer targeting:* The advertiser can target viewers based on their attributes, including (a) the slot size and format with the page content that viewers visit, (b) the keywords searched by the viewers, (c) the demographic information of the targeted viewers (e.g., their gender, age, ethnicity, education, income, geographical location), (d) the device and operating systems that viewers use (e.g., PCs, laptops, and smartphones), and (e) other (behavioral) attributes that the ad network may learn by tracking viewers' activities through the *cookie files* posted in their devices (e.g., viewers with potential interest in photography). (2) *Campaign Duration:* The start and end times of the contract. (3) *Impression Goal:* The number of viewers who are shown the ad. (4) *Contract and Penalty Costs:* The CPM or CPC price and any shortfall penalty.

In NGD contracts, ads are displayed to viewers who are won by advertisers through auctions conducted on ad exchanges. Since the outcome of such auctions cannot be predicted, it is not possible to guarantee to deliver a specific number of impressions. However, as these auctions are conducted for one viewer at a time, they can facilitate a high degree of targeting by advertisers. At the same time, they also serve as a mechanism for publishers to dispose of their excess inventory (which cannot be sold through GD contracts) in an effective manner. Auctions are used both in NGD contracts as well as in sponsored search contracts. The most common form of the auction used in these contracts is the *generalized second price* (GSP) auction, as will be discussed in Sect. 5.4.3. We also refer the reader to Wang et al. (2017) for a recent discussion of theoretical underpinnings for this auction, and for further references on auctions.

The Sequence of Events for an Ad Auction

Auction-based transactions result from structured communications between DSP and SSP. The information exchanged typically include bid requests, bid responses, win notices, and ad markup. The sequence of events in a typical transaction made between an ad agency and a publisher, shown in Fig. 5.5 is as follows: (1) A viewer visits a website (or a mobile app) (i.e., CNN.com). The page starts to load on the viewer's browser. (2) The publisher's SSP (e.g., PubMatic) shares the viewer profile (e.g., male, CEO, 46, from Palo Alto, etc.) as well as the available ad slot information (e.g., a leaderboard ad slot on top of the viewer's page) with the ad exchange and makes a bid request. (3) The ad exchange shares this information with the ad agency's DSPs and asks them to place their bids. (4) Each agency's DSP evaluates the bid request and sends in a bid response. (5) The ad exchange runs a second-price auction and selects the winning DSP. It sends the winning ad's markup and the price to the publisher's ad server. (6) The publisher's ad server sends the ad to the viewer's browser and posts it on the designated ad slot. All these steps within the real-time bidding (RTB) auction occur within around 360 ms for every impression!

Although GD and NGD contracts describe the broadest category of digital ad contracts, multiple variations of these contracts are observed in practice. The nature of transactions between advertisers and publishers depends upon how the price is set (fixed or auction-based) and the type of inventory of ads (reserved or unreserved). When ad inventory is reserved and the price is fixed, the situation resembles the traditional model of direct sales, where the terms are negotiated between the advertiser and the publisher's sales team. The only difference is that the transactions are conducted programmatically. In this method, also called *automated guaranteed*, advertisers generally do not have the flexibility of picking individual impressions. The main advantage of this model for advertisers is that they have an assurance that their campaign will achieve the target number of impressions. Reserved inventory is typically not sold through auctions. When inventory is unreserved, prices may be fixed, or determined through an auction. Fixed price deals with unreserved fixed inventory, may occasionally receive priority over auction-based deals in this instance. When prices are determined through auctions, the auctions may be open or invitation-only. Invitation-only auctions (or a private marketplace), available only to a restricted set of advertisers, offer greater information transparency, more premium inventory, and priority in auctions. Open auctions are the most common form of auctions and accessible to all buyers.

5.4 Key Operational Problems in the Digital Supply Chain

In this section, we describe the typical challenges and decisions faced by the key participants in any digital supply chain—the publishers, advertisers, and the ad exchange. Figure 5.6 summarizes some of the main decisions made by these participants.

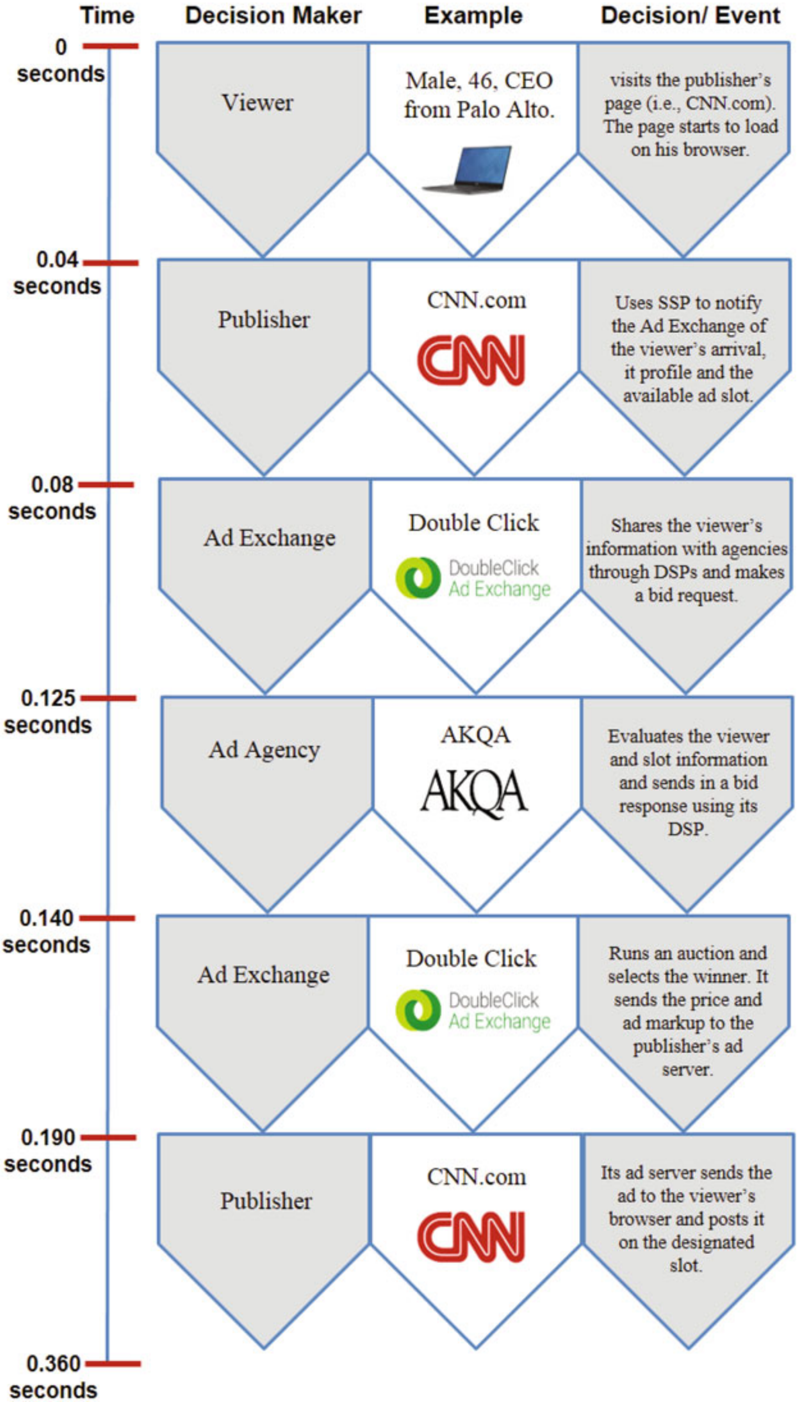

Fig. 5.5 The sequence of events in the digital supply chain

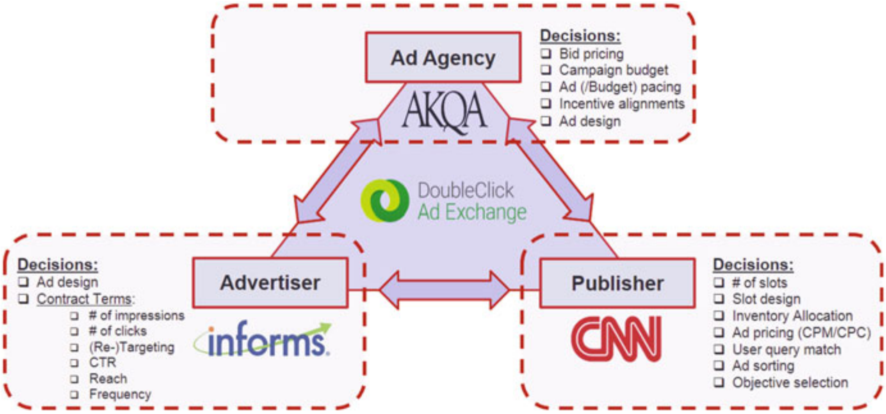

Fig. 5.6 Some of the main decisions made in the digital supply chain

5.4.1 Inventory Allocation (Ad Scheduling)

One of the key operational problems to be solved by publishers (or ad networks) is *inventory allocation*, which is the decision about how to optimally allocate the supply (or inventory) of user visits to the demand, represented by advertising campaigns.

There are several aspects of the problem that make it complex. Any inventory allocation must conform to requirements specified by each campaign about user characteristics. So, for example, when multiple campaigns target the same viewer attribute(s), the inventory allocation policy must assign users to campaigns in such a way that each campaign's requirements about users of specific types are met. The problem becomes particularly difficult because arrivals and viewer responses are uncertain. For this reason, we find that a vast majority of papers solve the deterministic version of the problem. Second, unlike in search advertising, the goals of display advertising can range from brand and awareness building to achieving performance, such as a specific number of clicks, impressions, or action. Thus, a second question relevant to this decision is how does a publisher allocate inventory across diverse advertiser goals and payment types (CPM, CPC, CPA or a combination thereof) so that advertiser and publisher objectives are met. Another source of complexity in answering this question arises because there are different types of contracts that can be sold for ad inventory—guaranteed delivery and non-guaranteed delivery. As noted earlier, GD contracts are sold to advertisers far in advance of actual impressions, while under NGD contracts, the advertisers bid for impressions in ad exchanges in real-time. Clearly, the same impression qualifies for both types of contracts. Therefore, the publisher must decide how to allocate inventory to both GD and NGD advertising campaigns (on spot markets), while still ensuring that the guaranteed advertiser objectives are met, and publisher revenue is maximized.

The environment within which these decisions are made also presents several technical challenges. Clearly, the size of the problems in terms of the number of decision variables and constraints is large. Therefore, in order for the problem to be solvable in real-time, solution methods must leverage any special properties of the mathematical structure, and any demands on the computing environment must also be addressed (for example, memory considerations). The large number of viewer segments makes parameter estimation a particularly challenging problem (our review does not dive into this aspect, since it is likely to be a substantial review in itself). Obviously, any dynamic environment like digital advertising must rely on effective and timely feedback about the outcomes of decisions made. For example, how has each campaign performed with respect to its goals? How effective have the campaign bidding strategies been? Did the impressions served actually conform to the campaign objectives? Unfortunately, our understanding is that technological constraints preclude access to such information in a timely manner and at the desired level of granularity.

We will review the literature on inventory allocation issues separately for the cases of display and search ads. Both types of ads have been analyzed by deterministic and stochastic model formulations. Since some of these models involve bidding, some of these papers are emphasized again in Sect. 5.4.3, where we discuss bidding strategies.

5.4.1.1 Display Ads Inventory Allocation: Deterministic Models

Turner (2012) focuses on the problem faced by an ad network that manages guaranteed targeted display advertising. The model in this paper is representative of the approaches adopted in many other papers that consider deterministic display ad inventory allocation. The author develops a single period planning model to allocate impressions to multiple audience segments by formulating a transportation problem with a quadratic objective, which is used to deal with audience uncertainty. The goal is to develop a plan that has a high reach (i.e., a large number of unique viewers) and low variance (i.e., since the actual number of impressions is a random variable, its variance should be low). The idea is that advertisers want impressions from all audience segments that match the targeting requirements specified, not just the particular subset of the audience that is easiest (or cheapest) for the ad network to deliver. In this context, minimizing a quadratic objective spreads impressions across audience segments by minimizing the L^2 distance to the "most representative" allocation; i.e., one that gives each campaign an equal proportion of each audience segment.

There are multiple viewer types whose supply is random. There is uncertainty in the arrival of viewers and the number of ads on a page. Turner (2012) replaces the supply for each viewer type by a deterministic value equal to its mean, which allows him to develop a deterministic formulation. Let K be the set of ad campaigns. Each campaign $k \in K$ requests its ad to be displayed to g_k unique viewers (impressions goal). Let V be the set of viewer types. Each viewer type $v \in V$ refers to a distinct *partition* (segment) in the viewers' population and has a deterministic supply s_v. Let V_k be the set of viewer types targeted by campaign k and K_v be the set of campaigns that target viewer type v. When a viewer of type v visits, only ads from the set K_v are displayed. The set K_v might contain several ad campaigns that target viewer type v. For example, if v is a male living in California interested in photography, then he could be shown an ad targeting anyone living in California, anyone interested in photography, or a male interested in photography. Let x_{vk} be the number of impressions of viewer type v assigned to campaign k and p_{vk} be the proportion of the expected number of impressions of viewer type v assigned to campaign k. Then, the deterministic problem can be formulated in two different ways shown below. In both formulations, the constraints are that impression goals must be met and the number of impressions delivered must be no more than the viewer supply. The objective functions are minimizing the weighted squares of allocations to ensure uniformity. The resulting equivalent formulations are

$$\text{(Impression Formulation)} \quad \min_{x_{vk} \geq 0} \sum_{k \in K, v \in V_k} \frac{x_{vk}^2}{s_v}, \tag{5.1}$$

s.t.

$$\sum_{v \in V_k} x_{vk} = g_k, \quad k \in K, \text{ (impression goals)}, \tag{5.2}$$

$$\sum_{k \in K_v} x_{vk} \leq s_v, \quad v \in V, \text{ (supply constraints)}. \tag{5.3}$$

$$\text{(Proportion Formulation)} \quad \min_{p_{vk} \geq 0} \sum_{k \in K, v \in V_k} s_v p_{vk}^2, \tag{5.4}$$

s.t.

$$\sum_{v \in V_k} s_v p_{vk} = g_k, \quad k \in K, \text{ (impression goals)}, \tag{5.5}$$

$$\sum_{k \in K_v} p_{vk} \leq 1, \quad v \in V, \text{ (supply constraints)}. \tag{5.6}$$

Once the solution of the transportation problem is determined, it is used to compute the mean and variance of metrics such as the actual number of impressions served to a campaign. He also develops methodologies to aggregate the audience segments into a smaller number of clusters to render the problem size manageable (due to the large number of audience segments in practice).

Bharadwaj et al. (2012) develop a similar formulation, with an objective function that is the sum of penalties associated with under-delivery of impressions to contracts and a cost associated with deviation from the target representativeness, i.e., the total L^2 distance between the actual and target allocation to impressions. Their main goal, however, is to develop a solution methodology that is efficient, scalable, and robust.

Deza et al. (2015) also formulate a problem with a quadratic objective that determines optimal proportions of viewers allocated to campaigns, but model a chance constraint, a constraint that specifies that the probability that a campaign's target is unfilled is smaller than a specified amount. They develop a sample approximation program with a branching heuristic, and convex approximations under Normal and distribution-free viewer supply assumptions, with an iterative method for improving feasible solutions.

In general, advertising contracts do not differentiate between multiple impressions of an ad made by the same viewer, and distinct impressions made by different viewers. However, increasingly, advertisers are getting sensitized to not only the number of unique users (reach) who view their ads, but also the frequency with which individuals are served the same ad. Hojjat et al. (2017) consider a new form of contract that allows advertisers to specify reach, as well as frequency. They develop an optimizing methodology that also incorporates constraints to ensure minimal under-delivery and appropriate spread of each campaign across its targeted viewer types. Their method employs pre-generated patterns to schedule the exact sequence of ads for each viewer, and can be implemented efficiently using a two-phase algorithm that employs column generation in a hierarchical scheme with three parallelizable components.

Mookerjee et al. (2017) take a different approach and develop a threshold based policy to determine when an ad should be served to a viewer to maximize revenue earned through clicks. They develop a predictive model of a visitor clicking on a given ad, which is used to determine a threshold to decide whether or not to show an ad to the visitor. Their decision model's objective is to maximize the advertising firm's revenue subject to a click-through-rate constraint. They study and contrast two competing solutions: (1) a static solution and (2) a rolling-horizon solution that resolves the problem at certain points in the planning horizon. Their solution was also implemented at an Internet advertising firm.

5.4.1.2 Search Ads Inventory Allocation: Deterministic Models

Langheinrich et al. (1999) is one of the early papers on the inventory allocation problem in search advertising. Their model is representative of the approaches used by many others for this problem. In contrast to personalized ad service strategies, which rely on a significant amount of information about the viewer or user, the ad allocation model presented in Langheinrich et al. (1999) (called ADWIZ) relies primarily on the keywords used by the user, or the URL of the page where the user has arrived.

Suppose that m ads $A_1, \dots, A_m$ must be displayed depending on which of the search keywords $W_1, \dots, W_n$ is input. Let h_j be the desired display rate for A_j in the next period defined as the number of remaining impressions for A_j, divided by the number of remaining periods, normalized such that $\sum_m h_j = 1$. Also, let k_i be the probability that W_i is searched and c_{ij} be the click-through rate (CTR) of A_j when W_j is searched. The decision variable, d_{ij}, is the probability of showing A_i when W_j is searched. The optimization problem is

$$\min_{d_{ij} \geq 0} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} k_i d_{ij}, \tag{5.7}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} k_i d_{ij} = h_j, \ \ \forall j, \qquad \sum_{j=1}^{n} d_{ij} = 1, \ \ \forall i. \tag{5.8}$$

Depending on special restrictions on ads or keywords, additional constraints can be added. If some d_{ij} are set to zero, then this may make an estimation of the ad-keyword combination's CTR value become difficult. Therefore, the authors set a minimum display rate for each ad-keyword combination, which is gradually lowered as a function of the sample size, by adding the constraint $d_{ij} \geq 1/(2m\sqrt{D_{ij}+1})$, where D_{ij} is the number of times that A_j has been displayed for W_i so far.

In their evaluation, they compared their learning method with two other simple methods: the method of selecting an ad randomly (the random selection method) and the method of selecting an ad with the maximum click-through rate for the given keyword in the current data (the max click rate method). While it is true that the max click-through rate method achieves the highest total click-through rate, with this method, half the ads did not get displayed enough. Thus, this method is most likely not usable in practice. The number of displays for the ads is balanced for both the random selection method and our method, but the total number of clicks yielded by the LP-based method is significantly higher.

Nakamura and Abe (2005) extend the formulation in Langheinrich et al. (1999) to address a number of technical challenges associated with the estimation of click-through rates and optimization of display probabilities. In order to address the exploration-exploitation trade-off inherent in the LP approach used in Langheinrich et al. (1999), instead of using click-through rate as the objective function, they use the Gittins Index. To address estimation challenges due to the sparseness of data, they devise a clustering methodology that groups the attributes that have similar click-through statistics. Tomlin (2000) points out the problems associated with LP optimal solution. Since such solutions are corner-point solutions, some words may be assigned to a majority of ads and others to none at all (called over-targeting). He proposes an entropy-based modeling approach to smoothen the objective function that prevents any all-or-nothing solutions.

Chickering and Heckerman (2003) extend the Langheinrich et al. (1999) model by incorporating the impression context to gain additional insights about the viewer. A publisher faces a supply of viewers consisting of I types where each viewer type is a disjoint audience segment or partition in the viewers' population. The publisher sells the impressions to J advertising campaigns (also known as *contracts*). Each campaign j ($j = 1, \ldots, J$) requests that its ad is shown to q_j viewers from the viewer types targeted by the campaign. Let x_i be the overall supply of viewers of type i ($i = 1, \ldots, I$) and y_{ij} be the number of viewers of type i allocated to campaign j. All campaigns start simultaneously at time 0 and end at time T. The authors determine the optimal y_{ij} to maximize the overall expected click-through probability of the website. The objective problem is

$$\max_{y_{ij} \geq 0} \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \frac{y_{ij}}{N}, \tag{5.9}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} y_{ij} \leq x_i, \quad \sum_{i=1}^{I} y_{ij} \geq q_j. \tag{5.10}$$

In this formula, p_{ij} is the probability that a viewer of type i clicks an ad from campaign j and N is the total number of viewers arriving during T periods.

Rusmevichientong and Williamson (2006) sort keywords based on their profit-to-cost ratio while selecting keywords, and show that their algorithm performs better than typical multi-armed bandit approaches. Özlük and Cholette (2007) also consider a deterministic formulation to allocate an advertising budget among multiple keywords for an advertiser. They show that the ratio of the CTR values of the keywords and the price elasticities of the response functions determines the bid for any keyword. They also examine when an advertiser should increase the number of keywords used, and compute the impact of an additional keyword under the assumption of constant elasticity.

Zhao and Nagurney (2005) determine the optimal marketing strategies when an advertiser is displaying its ad in n websites. Let e_i denote the number of impressions allocated to the ad on website i with the CPM price. Let $r_i(e_i)$ be the CTR on website i and C be the advertiser's budget. Then, the advertiser seeks to maximize the CTR, subject to the budget constraint as

$$\max_{e_i \geq 0} \sum_{i=1}^{n} r_i(e_i), \quad \text{s.t.} \sum_{i=1}^{n} c_i e_i \leq C. \tag{5.11}$$

McAfee et al. (2013) study the optimal allocation of viewers to different campaigns such that each ad is displayed to sufficiently large (i.e., representative) numbers of viewers from each type it has targeted. Let V and K be the set of viewer types and ad campaigns, respectively. Let s_v be the supply of viewers of type $v \in V$ and y_{vk} be the number of viewers of type v allocated to campaign $k \in K$. Let $\tau_{vk} = 1$ if viewers v are targeted by campaign k and $\tau_{vk} = 0$, otherwise. Then, the total number of viewers serving campaign k is $Y_k = \sum_{v \in V} \tau_{vk} y_{vk}$. Campaign k specifies Y_k up-front (impressions goal). Let $\tau_{vk} s_v / \sum_{v' \in V} \tau_{v'k} s_{v'}$ be the *representative* (fair) market share of viewers of type v, directed to campaign k, and $\tau_{vk} s_v / \sum_{v' \in V} \tau_{v'k} y_{v'k}$ be the proportion of viewers of type v allocated to campaign k. In addition, let campaign k have priority W_k.The optimization problem is minimizing the weighted squared difference of the allocated and representative shares for each campaign as

$$\min_{y_{vk}\geq 0} \frac{1}{2}\sum_{k\in K} W_k Y_k \sum_{v\in V} \frac{\sum_{v'\in V} \tau_{v'k}s_{v'}}{\tau_{vk}s_v}\left(\frac{\tau_{vk}s_v}{\sum_{v'\in V}\tau_{v'k}s_{v'}} - \frac{\tau_{vk}s_v}{Y_k}\right)^2 \tag{5.12}$$

s.t.

$$\sum_{k\in K} y_{vk} \leq s_v, \text{ for all } v \in V \text{ (supply constraints)}, \tag{5.13}$$

$$Y_k = \sum_{v\in V} \tau_{vk} y_{vk}, \text{ for all } k \in K \text{ (demand constraints)}, \tag{5.14}$$

in which the term $(\sum_{v'\in V} \tau_{v'k}s_{v'})/\tau_{vk}s_v$ is the given weight to each campaign's representative allocation.

5.4.1.3 Stochastic Models

The papers discussed above treat the deterministic inventory-allocation problem. The underlying problem is a dynamic resource allocation problem. In the computer science literature, this is also referred to as the online stochastic optimization problem. Feldman et al. (2010) consider a general version of this problem, of which the display ads problem is a special case. In the display ads allocation (DA) problem, there is a set J of m advertisers who have paid a web publisher for their ads to be shown to visitors of the website. The contract bought by advertiser j specifies an integer upper bound on the number $n(j)$ of impressions that j is willing to pay for. A set I of impressions arrives online, each impression i with a value $w_{ij} \geq 0$ for advertiser j. Each impression can be assigned to at most one advertiser, i.e., there are m options for each impression, and if the impression i is allocated advertiser j, then it is denoted with $a_{ij} = 1$. The goal is to maximize the value of all the assigned impressions. The authors consider a random-order stochastic model, where the order in which impressions arrive is random, but no other prior information is known. Their training-based primal-dual algorithm for the stochastic packing LP problem observes the first ε fraction of the input and then solves an LP on this instance. For each advertiser, the corresponding dual variable β_j extracted from this LP serves as a (posted) price for the remaining impressions, and the algorithm assigns an impression to advertiser j that maximizes $w_{ij} - \beta_j$. They prove that this algorithm provides a $(1 - \varepsilon)$ percent of the optimal solution.

Online optimization methods have also been considered by Feldman et al. (2009a) and Feldman et al. (2009b) for display advertising, and by Vazirani et al. (2005), Buchbinder et al. (2007), Goel and Mehta (2008), and Devanur and Hayes (2009) for the ad words problem. Other papers that address ad inventory scheduling include Roels and Fridgeirsdottir (2009) who formulate the display ad scheduling problem as a stochastic dynamic program, but solve its deterministic equivalent. Kumar et al. (2006) develop a heuristic and a genetic algorithm approach for the search ads case.

5.4.1.4 Guaranteed and Non-guaranteed Delivery Contracts

Because of the increasing role of ad exchanges in real-time transactions of impressions, publishers must decide whether to assign an arriving impression to one of their GD campaigns or sell it in the spot market through the ad exchange. Therefore, the publisher must evaluate, in real-time, the trade-off between the short-term and potentially higher revenue from the ad exchange (NGD contracts), and the long-term benefit of making good on the promise of delivery to GD contracts.

Yang et al. (2010) develop a deterministic formulation of the problem as a bipartite graph. Their way of modeling is typical among the papers that study deterministic contracts. Demand nodes are ad campaigns and supply nodes are viewers targeted by campaigns. Let V and K be the set of viewer types and ad campaigns, respectively. Let s_v be the supply of viewers of type $v \in V$ and r_v be the payout for $v \in V$ by NGD campaigns. Let g_k be the impressions goal of campaign k and V_k be the set of viewer types that it targets. Let y_{vk} be the amount of viewers type v allocated to the GD campaign k. Then, $z_v = s_v - \sum_k y_{vk}$ will be the number of viewers of type v allocated to the NGD campaigns. The objective function for the NGD campaigns is to maximize the publisher's revenue by allocating the viewers with the highest values to the NGD contracts—that is, $\max_{y_{vk}} \sum_{v \in V} r_v z_v$. The objective of GD campaigns is to maximize the brand awareness/reach or performance (clicks, conversions) by ensuring each campaign k is shown to a representative number of viewers from each type $v \in K_v$ (i.e., $\theta_{vk} = (s_v / \sum_{i \in K_v} s_i) g_k$). Define W_k as the importance of the representative allocation to campaign k. Then, the authors define the objective function for the GD campaigns as

$$\min_{y_{vk}} \sum_{k \in K} \sum_{v \in K_v} \frac{W_k}{2\theta_{vk}} \left(y_{vk} - \theta_{vk}\right)^2 . \tag{5.15}$$

Constraints are specified for demand (for GD contracts), supply (each viewer can be allocated to only one campaign), and non-negativity similar to the deterministic problem formulation in (5.1).

In contrast to Yang et al. (2010), Balseiro et al. (2014) include an ad exchange besides the publisher's own pages. Advertisers can buy display ad placements by negotiating guaranteed contracts directly with publishers. Since the publisher signs many such contracts, it must then decide how to allocate arriving impressions to contracts that correspond to the attributes of the viewer. The assignment is done so as to maximize placement quality: an example of a metric for this is click-through rate. Since guaranteed contracts are booked in advance, they suffer from an inability to respond to instantaneous changes in market conditions. As mentioned earlier, ad exchanges (like DoubleClick, OpenX, and AdECN) deal with real-time market conditions by offering a spot market for ads. If the option to use ad exchanges is present, the publisher's problem becomes complicated. For each viewer, it must decide whether to assign the viewer to an advertiser on the exchange, and at what price, or to an advertiser with whom a guaranteed contract has been booked. The revenue may be higher on the exchange but the quality of the advertiser may be lower.

They formulate the problem as a combination of a capacity allocation problem and a dynamic pricing problem. They assume a single ad slot, and a known number of N viewers who arrive one at a time. Each arriving viewer is first offered to the ad exchange at a reservation price p, which is determined by optimizing Yield $=$ revenue(AdX) $+ \gamma \times$ quality(advertisers), where γ is a parameter that can assist in balancing the trade-off between revenue from the ad exchange and the quality of advertisers. Bids are accepted with cumulative probability $F(p, u)$, where u is the user's information. The optimization problem becomes $\max_{p \geq 0}\{(1 - F(p, u)) + F(p; u)c\}$, where c is the publisher's opportunity cost for selling its ad inventory in the exchange. If none of the prices bid on the exchange exceed this reservation price, the viewer is assigned to the guaranteed contract whose placement quality exceeds its bid price by the largest amount. Ties among contracts are resolved by randomizing according to a probability distribution that can be predetermined by solving an assignment problem.

The overall problem can be formulated as a dynamic program indexed by the number of remaining viewers. Because of its large state space, they approximate the problem with its deterministic version in which (1) the policy is independent of the history but dependent on the realization of the vector of attributes (recall that placement qualities are deterministically determined based on the attributes), (2) capacity constraints are met in expectation, and (3) controls are allowed to randomize. They derive an efficient solution policy and prove the asymptotic optimality of this policy in terms of any arbitrary trade-off between the quality of delivered reservation ads and revenue from the exchange, and show that their policy approximates any Pareto-optimal point on the quality-versus-revenue curve. Experimental results on data derived from real publisher inventory confirm that there are significant benefits for publishers if they jointly optimize over both channels.

Celis et al. (2014) propose an auction sales mechanism in online display advertising in which bidders can "buy-it-now" at a posted price, or "take-a-chance" in an auction where the top $d > 1$ bidders are equally likely to win (random allocation). Mookerjee et al. (2016) provide an approach to manage an on-going Internet ad campaign that substantially improves the number of clicks and the revenue earned from clicks. They describe the problem faced by Chitika, which contracts with publishers to place relevant ads over a specified period on publisher websites. Ad revenue accrues to the firm and the publisher only if a visitor clicks on an ad. This might imply that all visitors to the publisher's website be shown ads. However, this is not the case if the publisher imposes a CTR constraint on the advertising firm. This performance constraint captures the publisher's desire to limit ad clutter on the website and hold the advertising firm responsible for the publisher's opportunity cost of showing an ad that did not result in a click. The paper develops a predictive model of a visitor clicking on a given ad. Using this prediction of the probability of a click, the authors develop a decision model that uses a threshold to decide whether to show an ad to the visitor. The decision model's objective is to maximize the advertising firm's revenue subject to a click-through-rate constraint.

Chen (2017) proposes a stylized model in which an online publisher sells its display ad space sequentially in two markets. In the first market, a set of long-term guaranteed advertisers negotiate with the publisher; once they enter an agreement, the publisher commits to deliver a pre-specified number of impressions within a fixed time frame through a guaranteed contract. In the second market, new advertisers arrive and express their interest in the display ads. This "spot" market runs an auction to allocate the display ad spaces every period among the short-lived advertisers. Given the dynamic nature and the unpredictable supply in the display ads industry, even in the GD contract, the publisher cannot promise that a certain number of impressions will be delivered within a time frame for sure. Thus, the GD contract specifies a per unit penalty if the promised number of impressions are not delivered by the end date specified in the contract. In addition, the publisher partially shares the instantaneous benefit from the guaranteed advertiser upon successfully delivering the promised impressions over time. This benevolence may be rationalized by the fact that the publisher is engaged in the long-term contractual relationship with the guaranteed advertiser. The paper uses the mechanism design approach to characterize the optimal dynamic selling scheme.

5.4.1.5 Number of Ad Slots and Ad Placement on Webpages

Another question related to how inventory allocation decisions are made is how many ad slots are displayed on any page. This question is the focus of Kim et al. (2012), where the authors study the challenge that search engines face in determining the number of ad slots to display on their page. Obviously, this decision is related to those of advertiser slot assignments, and the payments per click based on the advertiser's bids and the quality of their ads. More ad slots potentially generate incremental revenue gains because of the additional ads displayed. However, it may increase the clutter, leading to a reduction in the number of clicks to any individual ad, and therefore, its value. Thus, the number of ad slots has a significant impact on the search engine's profit. They further explore whether the number of ad slots should be keyword specific.

The problem is modeled as a two-stage game between the search engine and advertisers. In the first stage, the search engine sets up a GSP auction to sell ads and announce the number of ad slots for a keyword to N potential advertisers. The search engine knows the distribution of the advertisers' valuations and click-through rates, but not their values. The position specific click-through rate (CTR) is not advertiser specific, and is known. In the second stage, the values of the valuation and CTR are revealed and the advertisers participate in the GSP by bidding per click for their ads.

Based on Nash equilibrium results, they characterize the optimal number of slots. If additional slots do not affect clicks from existing slots, a global optimum number of slots may exist if the valuation distribution has an increasing hazard rate or is log-concave. In comparison, if additional slots cannibalize clicks, the optimal number of slots may be greater or smaller.

This paper extends the earlier work in Feng et al. (2007), Balachander et al. (2009), and Katona and Sarvary (2010). Katona and Sarvary (2010) consider the interaction between the list of organic search results and the list of advertisers featured on the results page of the search engine and show that highly relevant advertisers who tend to be featured high on the organic list may not bid high enough to be featured highly on the sponsored list. They generate normative guidelines for both advertisers and the search engine on how to buy and sell sponsored links. For instance, they find that in some cases, a search engine can attain higher revenues by displaying fewer sponsored links. In contrast, Feng et al. (2007) and Balachander et al. (2009) assume that the number of ad slots are pre-specified, and compare alternative GSP auction policies for determining bidders' ranks and payments per click. These and related papers build on early work done on GSP auctions by Edelman et al. (2007) and Varian (2007).

Since advertisements on the web are specified by geometry and display frequency, Kumar et al. (2006) consider both of these factors in developing a solution to the advertisement scheduling problem. Given a set of ads, a schedule of the ads specifies which ads are to be displayed at the same time. Each ad must be displayed with the correct frequency, allocated enough space for the specified geometry, and it must be possible for all the ads to be displayed simultaneously to be arranged in the space available for advertising. They show the problem to be NP-hard, develop a heuristic called LSMF to solve the problem, and then combine it with a genetic algorithm (GA) to develop a hybrid GA. Adler et al. (2002) determine the optimal schedule by finding a solution to a new variant of the bin packing problem, where there is a number of copies of each item to be placed into the bins, and they provide an efficient algorithm for the new bin packing problem. Amiri and Menon (2006) extend it to a more realistic setting, where the customer is allowed to specify a set of acceptable display frequencies. The Lagrangian decomposition-based solution approaches presented in this chapter are observed to provide good schedules in a reasonable period of time. Deane and Agarwal (2012) allow for variable display frequencies in the context of banner ads.

5.4.2 Ad Pacing/Budget Pacing

A very important consideration for advertisers in managing digital advertising is the rate with which ads are delivered to viewers over time. Obviously, this directly affects the rate with which the advertising budget is consumed over the same period. Therefore, ad pacing and budget pacing are related problems. For NGD contracts, the bidder (the advertiser or its agent) must address this challenge. In

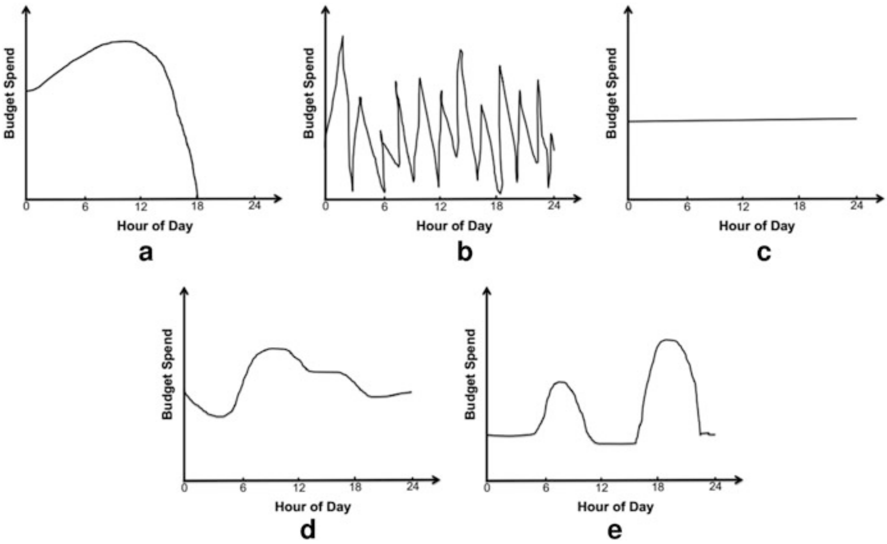

Fig. 5.7 Budget pacing strategies. (**a**) Premature stop. (**b**) Fluctuating budget. (**c**) Uniform pacing. (**d**) Traffic-based pacing. (**e**) Performance-based pacing

case of GD contracts, recall that the publisher must deliver a certain number of impressions (or actions) for each campaign over a given length of time, within a fixed budget. Advertisers expect that these campaigns are managed so the budget is not consumed before the end of the campaign, which is referred to as Premature Campaign Stop (Fig. 5.7a, Lee et al. 2013). This is a challenge because the advertisers may miss out on valuable opportunities in the remaining portion of the campaign duration. Advertisers also attempt to analyze the outcomes from their campaigns. This becomes difficult if the delivery of impressions (hence budget spend) fluctuates widely over any period of time (called as Fluctuation in Spend, Fig. 5.7b). Premature campaign stops also lead to a skewed representation of target audience segments. Moreover, as certain campaigns end prematurely, market competition reduces, which may result in reduced revenue. Finally, because of technical delays in reporting on the performance of campaigns, some campaigns may over-spend and over-deliver, which may be suboptimal overall (Agrawal et al. 2014). Thus, depending on the type of contract, pacing is a problem for advertisers as well as publishers.

A commonly sought-after strategy is the Uniform Pacing policy (Fig. 5.7c), which attempts to spend the available ad budget uniformly. Unfortunately, there are two main challenges with this implementing strategy. The viewer traffic may not be uniform during a campaign period, which means that the budget pacing might have to mirror traffic patterns, as seen in Fig. 5.7d. Alternatively, the quality of the traffic, as measured by metrics such as CPC, CTR or others, may not be uniform. This implies that the budget spend might have to mirror the trajectory of these performance metrics, as seen in Fig. 5.7e.

In general, there are two approaches to control the rate at which ad budget is spent, which is referred to as *throttling* in the literature. One approach is bid price throttling, i.e., modifying bid prices. Bidding a smaller amount reduces the amount spent per bid and the likelihood that the bid wins, which reduces the expected spending rate. An alternative approach is probabilistic throttling, where the rate at which bids are submitted is randomly influenced. Both approaches are described in the research reviewed below.

In both cases, a key responsibility of campaign managers is to see if campaigns are tracking to plan, and take actions to accelerate campaigns that are falling behind, and decelerate those that may be ahead. This also means that the bidding strategy and the budget pacing strategy are tightly connected. Since we will devote Sect. 5.4.3 to reviewing the literature on bidding, here, we will only provide a brief overview of approaches used to influence the budget pacing. Details about some of these papers will be shown in Sect. 5.4.3.

Finally, we note that one could also view the pacing problem as a variant of the inventory allocation problem because pacing essentially implies allocation of ad inventory to campaigns over time. Therefore, the reader will note some similarities between the approaches used in papers described earlier and those described here.

Zhang et al. (2012) consider the problem of jointly optimizing bid prices and allocation of budget to specific campaigns in the context of sponsored search. Suppose that an advertiser has m campaigns $C_1, C_2, \ldots, C_m$. Campaign C_i is denoted with $C_i = \{g_i^{(0)}, D_i, K_i\}$ in which $g_i^{(0)}$ is the original periodical (e.g., monthly) budget set by the advertiser, D_i denotes the set of ads included in campaign C_i, and K_i denotes the set of keywords in campaign C_i. The ad set D_i can be written as $D_i = \{d_{i,1}, \ldots, d_{i,l_i}\}$ where l_i is the number of ads in C_i and $d_{i,s}$ $(s = 1, 2, \ldots, l_i)$ denotes the sth ad in C_i. The bid keyword set K_i can be written as $K_i = \{K_{i,1}, \ldots, K_{i,n_i}\}$, with $K_{i,t} = (k_{i,t}, b_{i,t}^{(0)}, v_{i,t})$ $(t = 1, 2, \ldots, n_i)$, where $k_{i,t}$ is a keyword chosen by C_i, $b_{i,t}^{(0)}$ is the original bid submitted for it, and $v_{i,t}$ is its value per click. When a viewer does a search, the search engine might match $K_{i,t}$ from campaign C_i with it and display a related ad $d_{i,s}$ in C_i. The information of this candidate item for the auction can be summarized by $\omega_{i,s,t} = (d_{i,s}, K_{i,t})$ $(t = 1, \ldots, n_i,\ s = 1, 2, \ldots, l_i)$, which the authors call the *order item* $\omega_{i,s,t}$, or more generally the item $\omega \in C_i$. Let Φ be the maximum number of slots on each search result page. In addition, let ρ_ϕ $(\phi = 1, \ldots, \Phi)$ be the position of the ad on the page. Let θ be the number of auctions $(\theta = 1, \ldots, \Phi)$. Furthermore, let $\tau_\phi r_{\omega,\theta}$ be CTR in which $r_{\omega,\theta}$ denotes the quality score of the item ω in auction θ, and τ_ϕ denotes the position bias at slot ρ_ϕ. Let $p_\omega\left(\rho_\phi | b_\omega\right)$ be the probability for item ω to be ranked in slot ρ_ϕ when its bid is b_ω. In addition, let $c_{\omega,\phi,\theta}$ be the cost for a click on ω in auction θ when it is ranked on position ρ_ϕ. Let g_i $(i = 1, \ldots, m)$ and b_ω denote the variables of campaign budgets and keyword bid prices. Furthermore, let ϵ_g and ϵ_b be the minimum campaign budget and bid price. The authors determine the optimal prices and budgets allocated to campaigns for keywords using sequential quadratic programming. An important input in this analysis is the probability of winning for a given bid price, $p_\omega\left(\rho_\phi|b_\omega\right)$, which is determined by fitting a normal distribution to the history of winning prices. The formulation in any iteration of the sequential method is

$$\max_{\substack{g_i\in[\epsilon_g,+\infty),\\ b_\omega\in[\epsilon_b,v_\omega]}} \left\{\sum_{i=1}^{m}\sum_{\omega\in C_i}\sum_{\theta=1}^{\Theta_\omega}\sum_{\phi=1}^{\Phi} p_\omega(\rho_\phi|b_\omega)(\tau_\phi r_{\omega,\theta})(v_\omega - c_{\omega,\phi,\theta})\right\} \tag{5.16}$$

$$\text{s.t.}\ \sum_{i=1}^{m} g_i = \sum_{i=1}^{m} g_i^{(0)}, \tag{5.17}$$

$$0 \le \sum_{\omega\in C_i}\sum_{\theta=1}^{\Theta_\omega}\sum_{\phi=1}^{\Phi} p_\omega\left(\rho_\phi|b_\omega\right)\tau_\phi r_{\omega,\theta} c_{\omega,\phi,\theta} \le g_i, \quad i=1,\ldots,m. \tag{5.18}$$

The objective function is the product of three terms: the probability for an ad to be ranked in position ρ_ϕ when its bid price is b_ω, the actual probability of an ad being clicked, and the expected revenue from a clicked ad. The first constraint indicates that the total campaign budget should be the same between consecutive iterations of the optimization steps. The second constraint specifies that the actual amount spent on a campaign should be less than its budget. The third constraint specifies a threshold limit on a campaign's budget, and the last constraint specifies bounds on the bid. The problem is formulated as a single period problem. Therefore, the implied assumption is that the solution is applied in the steady-state, or that it will be determined at the beginning of each period within the campaigns' planning horizon. Using simulation on the sponsored search log from a commercial search engine, the authors show that their proposed methodology can effectively help advertisers improve their campaign performance with respect to metrics, such as click number, cost per click (CPC), and advertiser revenue while also helping the search engine to increase its revenue.

Agarwal et al. (2014) have developed an algorithmic approach to implementing budget pacing for campaigns during any day, where the day is assumed to consist of discrete time periods. The algorithm begins with a forecast of the cumulative number of eligible impressions for each campaign by the start of each time period within the day. The goal is to ensure that the proportion of the total daily budget for a campaign that is spent by a time period is the same as the ratio of the cumulative forecast of eligible impressions by that time period and the total forecast for the campaign. This goal is implemented by probabilistically throttling the bidding rate. This probability is adjusted in every period (although this paper does not report on an analytical or optimization-based approach for determining this). The authors assess the impact of their methodology on advertiser-centric measures such as campaign lifetime and the number of campaigns served, publisher-centric measures such as the cost-per-request and overdelivery, and member-centric measures such as the number of unique campaigns served. Note that the approach in this chapter contrasts with those in Abrams et al. (2007), Borgs et al. (2007), and Vazirani et al. (2005), which propose throttling the bid. With bid price throttling, it may require bid prices lower than reserve prices set for bidding. This restriction may prevent extension of the life of campaigns. Also, probabilistic throttling does not require a recalculation of bids, which must be done under the bid throttling scheme.

Xu et al. (2015) focus on the problem faced by DSPs (governed by ad agencies) that are involved in managing multiple campaigns on behalf of advertisers by participating in RTB in ad exchanges. Different advertisers have different goals. For branding campaigns, the goal is to reach a broad audience by spending out the budget, while ensuring campaign performance metrics are as good as possible. Performance campaigns focus primarily on performance goals (e.g., effective cost per click $=$ total cost/number of clicks), and spend as much of the budget as possible. Like Agarwal et al. (2014), they use probabilistic throttling as well. The key decision variable is r_i, called the point pacing rate, for the ith viewer, which determines $s_i = Bernoulli(r_i)$, the probability with which the ad is allowed to participate in the bidding for that viewer. The spending plan for the planning period $1, \ldots, K$, is described by the vector, $B = (B^{(1)}, \ldots, B^{(K)})$, and the corresponding actual spending pattern by $C = (C^{(1)}, \ldots, C^{(K)})$. The deviation between the plan and actual spending can be quantified by a variance-type penalty function such as $\Omega(C, B) = \sqrt{(1/K)\sum_{t=1}^{K}(C^{(t)} - B^{(t)})^2}$. The authors develop separate formulations for each case, where the goal is to optimize performance goals with respect to r_i subject to budget spend $\sum_{t=1}^{K} C^{(t)} = \sum_{t=1}^{K} B^{(t)}$ and $\Omega(C, B) \leq \epsilon$ (for a given tolerance ϵ), or minimize $\Omega(C, B)$ subject to ensuring certain performance goals and the constraint on the budget as $\sum_{t=1}^{K} B^{(t)} - \sum_{t=1}^{K} C^{(t)} \leq \epsilon$. Since the resulting problem formulations are difficult to optimize, they develop heuristic solution algorithms and test them on real data.

Balseiro et al. (2017) compare the system equilibria of different budget management mechanisms provided by advertising platforms to control the expenditures of advertisers. Throttling controls ad expenditures by precluding a buyer from bidding. Thresholding allows a buyer to participate if its bid is above a fixed threshold. Reserve pricing is similar to thresholding, but the winner is charged the maximum of the second-highest bid and its reserve price, which leads to higher payments. Bid shading allows the buyers to participate in all auctions, but each bid is reduced, i.e., shaded, by a factor. The multiplicative boosting mechanism modifies the allocation rule which leads to higher payments. Thus, these mechanisms control expenditures by reducing bids (bid shading), modifying the allocation (multiplicative boosting), excluding buyers (throttling and thresholding), or imposing reserve prices (reserve pricing). The authors show that from the seller's perspective, imposing reserve prices and excluding buyers are more effective in controlling expenditures, and budgets can be depleted with fewer items sold, leading to higher seller profits. From the buyers' perspective, lower bids are more beneficial than fewer competitors, and bid shading leads to the highest buyer utility.

The mentioned papers above build upon earlier work done by several authors. For example, Lee et al. (2013) present an online approach to the smooth budget delivery problem while optimizing the conversion performance by selecting high-quality impressions and adjusting bid prices based on prior performance distribution by distributing the budget optimally across time. Zhou et al. (2008) also consider the budget constraint, albeit in the context of sponsored search and model it as an online knapsack problem. Charles et al. (2013) present a game-theoretic model for the outcome of an ideal budget smoothing algorithm. The authors propose the notion of regret-free budget smoothing policies whose outcomes throttle each advertiser optimally, given the participation of the other advertisers. They show that regret-free budget smoothing policies always exist, and with single slot auctions they can give a polynomial-time smoothing algorithm. Yang et al. (2014) extend the basic problem to include multiple campaigns that may be “coupled” due to overlaps in the campaign content, duration of the promotional period or target regions. They develop an optimal control to maximize the total payoff from advertising activities subject to budget constraints.

5.4.3 Ad Pricing

In both sponsored search and display advertising contexts, publishers often face uncertain demand from advertisers to post their ads and uncertain traffic of viewers whose click behavior is also uncertain. In the attempt to monetize this traffic in such an inherently uncertain environment, pricing of ads is one of the most challenging operational decisions that web publishers face. In GD contracts, the publishers (or ad networks) make the ad price decisions (with or without negotiating with advertisers). However, in NGD and sponsored search contracts, the price that each advertiser pays for posting its ad is the result of an auction.

Sponsored search contracts are typically based on CPC, while in display contracts CPM is the more common pricing scheme. The decision regarding whether the pricing scheme must be CPC or CPM depends on the campaign’s goal. CPM is standard when the campaign’s goal is to strengthen the advertiser’s brand and not necessarily to drive clicks. For instance, when CNN.com posts its own ads (house ads) on its website, its goal is not that viewers click on them; rather it is to strengthen its brand. CPC is used when an advertiser is focused on driving clicks. This suggests that, an advertiser’s ad might get many free impressions between every two clicks, but in reality, these “free impression” prices are indirectly accounted for in the CPC price. As shown in some of the papers described in this section, in general, if an advertiser’s ad has a low CTR, then CPC might be the preferred pricing option to avoid unnecessary payments for impressions, but if its ad performs very well (i.e., its CTR is higher than some threshold), paying based on CPC can become very expensive and the advertiser might be better off switching to CPM. A common approach in practice to switch between the two pricing schemes is by assuming $CTR \approx CPI/CPC$, where CPI is the price of an impression (i.e.,

$CPI = CPM/1000$) and CTR is the average long-run number of clicks made on an ad divided by the number of times it is seen by viewers. This conversion approach works well when the market's demand and supply conditions are very stable, so the average CTR does not change much, but as shown by some authors, e.g., Najafi-Asadolahi and Fridgeirsdottir (2014), in general, this conversion rule could be suboptimal and leads to a substantial revenue loss for the publishers because the CTR value can change significantly and instantaneously over time depending on the supply and demand rates and the number of ads served in each point in time.

5.4.3.1 Ad Pricing Models in Sponsored Search

As mentioned earlier, pricing of sponsored search and NGD display ads involves using GSP auctions. A GSP auction is an alternative approach for dynamically changing the ad prices in online advertising to match the viewers' supply and advertisers' demand. The closest classic auction to GSP is the second-price (also known as Vickrey) auction. Pricing of ads through GSP auctions differs from GD's ad pricing in that advertisers submit the maximum prices they are willing to pay (bids) and then the publisher charges each advertiser based on the submitted bids according to the auction's mechanism. As we will explain, the prices of ads depend on both the number of bids and each advertiser's valuation for the keyword. Hence, the more advertisers there are, or the higher their valuation for particular keywords are, the higher is the generated CPC or CPM price by the auction when a viewer queries those keywords. In this way, the CPC or CPM prices of ads effectively adapt to the market condition.

In the simplest GSP auction, for a specific keyword, advertisers submit bids stating their maximum willingness to pay for a click made by a viewer (i.e., user). Recall that multiple ads may be displayed on a page. With multiple ads displayed on a page, the ad appearing at the topmost positions is most valuable. When a viewer searches for a keyword, results are displayed along with sponsored links in descending order of bids. That is, the ad with the ith highest bid is displayed at the ith top position on the page. If the viewer then clicks on an ad in position i, that advertiser pays the $(i + 1)$st highest bid. One distinguishing feature of GSP auctions is that, advertisers can affect the rate of the clicks they receive by changing their bids at any time. Hence, due to this continuous nature of the bidding process, the advertiser can predict the number of clicks it can generate during a particular time by developing an appropriate price function (e.g., like the one used by Zhang et al. 2012). GSP requires that advertisers submit only one bid for each keyword although they might have different ads for different products. The one-bid requirement makes sense if we assume that the value of each position on the page is proportional to its CTR, while viewers who click on ads on different positions must have similar purchase probability. In other words, even though in the GSP environment different products could be shown, viewers are assumed to have the same purchase probability for all the ads. Thus, the most important factor in the GSP auction is a slot's CTR. Zhang et al. (2012) study the CTR of the slots in a search engine and verify that the CTR is decreasing from top to the bottom of the page. Furthermore, Brooks (2004) empirically shows that the purchase probability of viewers is not affected much by the ad's position on the page.

GSP assumes that different ads in the same position have the same CTR while in reality, this may not be the case. Search engines treat this possibility differently. For example, Yahoo! does not consider this possibility and ranks the ads solely in descending order of bids (*bid ranking*). However, Google multiplies an advertiser's bid value by its "quality score" to determine an advertiser's expected revenue. Then, it ranks the ads based on their expected revenues (*revenue ranking*). In an easy way, the expected revenue of the ad in position i is $bid(i) \times CTR(i)$ while it pays $bid(i+1) \times [CTR(i+1)/CTR(i)]$. Edelman et al. (2007) provide a more detailed model for GSP as described below, and show that a GSP auction tends to be more profitable for the publisher compared to a classic second-price auction such as Vickrey–Clarke–Groves.

The GSP Model

Assume that for a given keyword, there are N ad positions on the search page and K bidders (advertisers). The expected number of clicks per unit period for an ad posted in position i is α_i. Advertiser k gains value s_k for each click made on its ad, so advertiser k's payoff from being in position i is $\alpha_i s_k$, minus the payment it makes. Assume that the number of times that position i is clicked does not depend on the ads in this and other position partitions, and that an advertiser's value per click does not depend on its ad's position. We label the positions in descending order so for any i and j with $i < j$, $\alpha_i \geq \alpha_j$. When a viewer enters a given keyword, suppose that advertiser k's last bid before the search was b_k. Had advertiser k placed no bid for the keyword, then $b_k = 0$. Let $b^{(j)}$ be the jth highest bid and $g(j)$ be the advertiser who submitted $b^{(j)}$. $g(j)$'s ad is posted on position $j \in \{1, \ldots, \min\{N, K\}\}$. Each ad is posted at most in one position. If the viewer clicks on ad j, $g(j)$ gets a payoff $s_{g(j)} - b^{(j+1)}$. Hence, $g(j)$'s expected payoff is $\alpha_j(s_{g(j)} - b^{(j+1)})$. The GSP setting is similar to Vickrey–Clarke–Groves (VCG) auction. In a VCG auction $g(i)$'s ad is posted in position i. However, its payment is the negative of the externality it imposes on others. Thus, the payment of the last advertiser allocated a slot is zero if $N > K$, and $\alpha_N b^{(N+1)}$, otherwise. For other positions $1 \leq i < \min\{N, K\}$, $g(i)$ pays $p(i) = p(i+1) + (\alpha_i - \alpha_{i+1})b^{(i+1)}$. It is easy to see that, unlike VCG, truth-telling is not a dominant strategy under GSP. For example, consider three advertisers, with per-click values of \$10, \$4, and \$2, and two positions. Let the click rates of these positions be nearly the same, namely, let position 1 receive 200 clicks per hour, and position 2 gets 199. If all advertisers bid truthfully, then advertiser 1's payoff is equal to 200(\$10 − \$4) = \$1,200. If instead, it bids \$3 per click, it will get position 2, and its payoff will be 199(\$10 − \$3) = \$1592.

Edelman et al. (2007) examine if the bids in the GSP auction converge to stable equilibrium values when all values are common knowledge. A possible simple strategy for an advertiser to increase its payoff is to try to force out the advertiser who occupies the position immediately above. Suppose advertiser k bids b_k and is assigned to position i, and advertiser k' bids $b_{k'} > b_k$ and is assigned to position $(i-1)$. If k raises its bid slightly, its payoff does not change, but the payoff of the player above it will decrease. Player k' can retaliate, and the most it can do is to slightly underbid advertiser k, effectively exchanging positions with k. If advertiser k is better off after such retaliation, it will indeed want to force player k' out, and the vector of bids will change. If the vector converges to a fixed point, an advertiser in position i should not want to exchange positions with the advertiser in position $(i-1)$. Edelman et al. (2007) call such vectors of bids *locally envy-free*. Edelman et al. (2007) show that the game can have multiple locally envy-free Nash equilibria that all provide the publisher with payoffs no less than the one in VCG auction.

Sponsored Search with Budget Constraint

Kitts and Leblanc (2004) analyze decisions for a trading agent for CPC auctions. The agent creates a look-ahead plan of its desired bids, which allows it to exhibit decisions, including the ability to hold back money during expensive periods. In reality, search advertisers can bid on several keywords. Zhou et al. (2008) model the budget-constrained bidding optimization for sponsored search auctions when advertisers are *omniscient* (i.e., they know the bids of all advertisers at each time period) and bid only for a single keyword. If a bidder is *omniscient*, then the best bidding strategy corresponds to solving a knapsack problem. To better illustrate this idea, they consider a simple case where the publisher has only one slot. At each time t, let b_t be the maximum bid on the keyword among all the bidders. The omniscient bidder knows all the bids $\{b_t\}_{t=1}^{T}$. To maximize its profit, the omniscient bidder should bid higher than b_t. Winning at time t costs the bidder $w_t = b_t s_t \alpha$ and earns a profit $v_t = (V - b_t)s_t\alpha$, where V is the ad's per-click value and s_t is the expected number of times the keyword is searched during time t, and α is the CTR (the chance the advertiser's ad is clicked). Thus, the omniscient bidder with the maximum budget B would solve

$$\max_{x_t \in \{0,1\}} \sum_{t=1}^{T} v_t x_t, \quad \text{s.t.} \sum_{t=1}^{T} w_t x_t \leq B. \tag{5.19}$$

Thus keyword bidding corresponds to designing an algorithm for the online knapsack problem.

As explained previously, search engines often post an advertiser's ad only in one position. Thus, if an advertiser has submitted bids for two different keywords that both match the same search query, this conflict must be resolved somehow. For example, if an advertiser has placed bids on the keywords "shoes" and "high-heel," then if a viewer makes the query "high-heel shoes," it will match on two different keywords. The search engine specifies, in advance, a rule for the resolution based on the query, the keyword, and the bid. The most natural rule is to select the keyword with the highest bid. Feldman et al. (2007) model this problem using an undirected bipartite graph $G = (K \cup Q, E)$ where K is a set of keywords and Q is a set of queries. Each $q \in Q$ has an associated landscape, as defined by the cost $c_q(b_q(\mathbf{b}))$ and the clicks $\alpha_q(b_q(\mathbf{b}))$. An edge $(k, q) \in E$ means that keyword

k matches query q. The advertiser can control their individual keyword bid vector $\mathbf{b} \in \mathbb{R}_+^{|K|}$ specifying a bid b_k for keyword $k \in K$. Then, given the vector $\mathbf{b}$, the search engine chooses the advertiser's maximum bid, which is connected to query q as $b_q^*(\mathbf{b}) = \max_{k:(k,q)\in E} b_k$. By submitting the bid vector $\mathbf{b}$, the total amount of payment made by the advertiser and the expected number of clicks it receives are $C(\mathbf{b}) = \sum_{q\in Q} c_q(b_q^*(\mathbf{b}))$ and $\alpha(\mathbf{b}) = \sum_{q\in Q} \alpha_q(b_q^*(\mathbf{b}))$, where $\alpha_q(b_q^*(\mathbf{b}))$ is the number of clicks the advertiser is expected to receive if it bids b_q^* for q and $c_q(b_q^*(\mathbf{b}))$ is the associated cost paid. If the advertiser uses a randomized strategy, then it provides the search engine with a distribution $\mathscr{B}$ over bid vectors $\mathbf{b} \in \mathbb{R}_+^{|K|}$. In this case, the total cost and the clicks are $C(\mathscr{B}) = \mathbb{E}_{\mathbf{b}}[C(\mathbf{b})]$ and $\alpha(\mathscr{B}) = \mathbb{E}_{\mathbf{b}}[\alpha(\mathbf{b})]$. The advertiser's budget optimization problem is defined as

$$\max_{\mathbf{b}} \ \alpha(\mathbf{b}), \quad \text{s.t. } C(\mathbf{b}) \le B. \tag{5.20}$$

The solution to this problem provides the search engine with a distribution $\mathscr{B}$ over bid vectors $\mathbf{b} \in \mathbb{R}_+^{|K|}$. In this case, the total cost and the clicks are $C(\mathscr{B}) = \mathbb{E}_{\mathbf{b}}[C(\mathbf{b})]$ and $\alpha(\mathscr{B}) = \mathbb{E}_{\mathbf{b}}[\alpha(\mathbf{b})]$. The budget optimization problem in its general form is very hard to solve. In addition, it is not easy to justify strategies involving arbitrary distributions over arbitrary bid vectors to advertisers. The reason is that advertisers often like strategies that are easy to understand and evaluate. For this reason, Feldman et al. (2007) propose a "uniform bidding strategy" in which $\mathbf{b} = (b, b, \ldots, b)$ for some $b \ge 0$. In this case, the advertiser would merely bid b until its budget runs out, and the ad serving system would remove it from all subsequent auctions until the end of the day. Feldman et al. (2007) show that always there exists a uniform single-bid strategy that is $\frac{1}{2}$-optimal.[9]

The other work that considers advertisers bidding on several keywords is Borgs et al. (2007) that examines a setting similar to Feldman et al. (2007) in which m advertisers with limited budgets bid for n keywords. Borgs et al. (2007) assume that if advertiser i bids b_{ij} on keyword j then its day-long payment and net utility (i.e., total value minus total payment) on that keyword are $P_j(b_{ij})$ and $U_j(b_{ij})$, respectively. If advertiser i has a budget B_i, the optimization problem becomes

$$\max_{b_{ij}} \sum_j U_j(b_{ij}), \quad \text{s.t. } \sum_j P_j(b_{ij}) \le B_i, \tag{5.21}$$

which is similar to (5.20). At the optimal solution b_{ij}^*, we have $dU_j/dP_j = \lambda_i|_{b_{ij}=b_{ij}^*}$, where λ_i is the Lagrange multiplier (or the corresponding dual variable). This derivative is referred to as the "marginal ROI" and is the change in the advertiser's net utility for one unit of change in its payments. The derivative condition implies that, at the optimal level, advertiser i has the same marginal ROI across all keywords. Since obtaining the marginal ROI is often a difficult

[9] $\frac{1}{2}$-Optimal means that the achieved revenue by the uniform single-bid strategy is half of the revenue that the bidder could have gained by using the optimal bidding policy.

task, in order to solve (5.21), Borgs et al. (2007) propose and study a heuristic approach by determining b_{ij} values such that $ROI_j = U_j(b_{ij})/P_i(b_{ij})$ become approximately equal across all the keywords. As explained previously, Zhang et al. (2012) extend the models presented by Borgs et al. (2007) and Feldman et al. (2007), by considering the model given by (5.16) and optimize it jointly over the bid prices and campaign budgets. Likewise, Abhishek and Hosanagar (2013) develop a different extension of Borgs et al. (2007) in which the authors determine the optimal bids for multiple keywords by an advertiser when each keyword can be searched for a random number of times.

Ye et al. (2014) consider a retailer using sponsored search marketing together with dynamic pricing. The retailer's bid on a search keyword affects the retailer's rank among the search results. The higher the rank is, the higher will be the customers' traffic and their willingness to pay. They study whether a retailer should reduce prices when it bids more to attract customers (to strengthen the bid's effect on demand) or increase prices (to take advantage of higher willingness to pay). They find that the answer depends on the "pace" at which the retailer increases its bid. In particular, as the end of the season approaches, the optimal bid exhibits smooth increases followed by big jumps. The optimal price increases only when the optimal bid increases sharply, including the instances when the bid jumps up. Such big jumps arise, for example, when the customer traffic is an S-shaped function of the retailer's bid.

5.4.3.2 Ad Pricing Models in Display Advertising

Ad Pricing in GD Contracts

We begin by looking at the pricing of impressions or clicks in GD display contracts. These contracts are commonly used between advertisers and online publishers such as Yahoo! for display ads. Such online publishers face uncertain demand from advertisers to display an agreed-upon number of impressions or clicks within a certain period, and uncertain supply of viewers with uncertain click behavior. The main challenge is to determine the price to charge a particular advertiser when its ad is shown to or clicked by a particular viewer, given the supply and demand uncertainties, and the penalty if the promised number of impressions are not delivered by the end date specified in the contract.

The existing models in the published domain tend to be simplistic in many ways. For example, they mostly consider only a single web publisher in isolation. Hence, the demand is a function solely of the publisher's price for showing the ad to a particular viewer. These models also ignore the impact of competition, the possibility of substitution (if advertisers find a particular viewer segment to be overly expensive), and possible strategic behavior of advertisers and viewers over time. Despite these simplifications, these models provide good approximations which can be useful in practice. Even with such simplifying assumptions, the analysis can still be complex due to uncertainties in demand and supply, and constraints on advertisers' budgets.

Deterministic GD Models

Mangani (2003) and Fjell (2009) examine the choice of CPM and CPC pricing schemes when the price is set by the market (the publisher is only a price taker).[10] Mangani (2003) considers a web publisher with a single page who is a price taker. Let c_1 and c_2 be the market's CPM and CPC prices, respectively. Let A be the number of ads on the page. In addition, let $N(A)$ be the number of viewers visiting the page with $\partial N/\partial A < 0$. This assumption was later adopted by many other papers in this stream of literature. Let $c(A)$ be the number of clicks made on one of the ads in a unit time with $\partial C/\partial A < 0$. Let α be the proportion of ads with the CPM pricing scheme and $(1-\alpha)$ be the proportion of ads with the CPC. The publisher is seeking to find the optimal shares of the two contracts by solving the following optimization problem:

$$\max_{\alpha, A} R(\alpha, A) = [\alpha N(A)c_1 + (1-\alpha)c(A)c_2]A. \tag{5.22}$$

Fjell (2009) extends the model in Mangani (2003) by letting the CPM and CPC ad shares depend on c_1 and c_2 explicitly. Specifically, Fjell (2009)'s model assumes that $A = A_1(c_1) + A_2(c_2)$, while (unlike Mangani (2003)) c_1 and c_2 are decision variables. So, this optimization problem becomes $\max_{c_1,c_2} R(\alpha, A)$ with $R(\alpha, A) = c_1A_1(c_1) + A_2(c_2)wc_2$, in which, $w = c(A)/N(A)$ is the CTR.

Asdemir et al. (2012) develop a principal–agent model to capture the decision regarding the choice of CPM or CPC, but from an advertiser's perspective instead of the publisher. The principal (advertiser) hires an agent (publisher) to deliver its ad on its website. The principal is uncertain about which of the publisher's pages are visited by the targeted viewer. The advertiser can induce the publisher to obtain information about pages visited by the targeted viewers by choosing a CPC contract. For the advertiser, a CPC contract requires assigning campaign decisions to the publisher. Therefore, the advertiser has two options to choose from: it can select the CPM model and control the campaign decisions by choosing which pages and how much (i.e., how many impressions, or how long) to advertise, or select the CPC model in which the publisher makes the decisions but also has better information about viewers.

Among the pricing models that consider GD contracts is the work by Bharadwaj et al. (2010) who propose a special pricing model, in which the price of a guaranteed contract is obtained based on the prices of the different viewer types that the campaign targets. The price of each viewer type is, in turn, based on the historical sales prices that were negotiated between salespeople and advertisers. An interesting approach that the authors use to find the GD's contract price is the Weighted Average Pricing (WAP) approach in which the price of each impression must be as close

[10]These works could have been classified as a separate subsection focused on the publisher's "objective selection" decision (see Fig. 5.6). Nevertheless, for brevity, we merely consider them as parts of the publisher's pricing decisions.

as possible to the negotiated price of the past eligible historical contracts made for different ad campaigns (normalized per impression). Specifically, the price of viewer type i, p_i, is obtained as

$$p_i \in \arg\min_{p \geq 0} \left\{ \sum_{j \in \mathscr{H}_i} x_{ij} (q_j - p)^2 \right\}, \tag{5.23}$$

where $\mathscr{H}_i$ is a set of eligible historical contracts that contain viewer type i, and q_j is the negotiated per-impression price of contract j. $x_{ij} \geq 0$ is the weight that captures the importance of contract j in determining the price i. By solving the above problem for p_i, one can find that $p_i = \sum_{j \in \mathscr{H}_i} x_{ij} q_j / \sum_{j \in \mathscr{H}_i} x_{ij}$. For example, suppose there is a viewer's arrival corresponding to a Computer Scientist living in Palo Alto viewing a Sports page. If there are two historical contracts interested in the viewer's arrival at the page, but 60% of such viewer arrivals were given to the first contract (priced at \$1 CPM) while 40% of such viewer arrivals were given to the second contract (priced at \$4 CPM), then WAP would price the viewer's arrival as the weighted average price of the contracts. So, the price would be $0.60(1) + 0.40(4) = \$2.2$ CPM.

Some authors have studied the problem of a web publisher who generates revenues not only from advertising, but also from subscriptions. Baye and Morgan (2000) present a simple model that explains why the mass media (e.g., TV, newspapers, magazines, and online publishers) gain most of their revenues from advertisements rather than subscriptions. Prasad et al. (2003) examine the subscription price and the amount of advertising that must be offered to consumers with several options. They show that pure revenue models such as ad-based or subscription-based only may not be optimal, so web publishers should consider hybrid models. In response to increasing interest in hybrid models, Kumar and Sethi (2009) study the dynamic pricing of web content on a site when revenue is generated from both a subscription fee and display ads. They use optimal control theory to solve the problem and obtain the optimal subscription fee and ad level.

Stochastic GD Models

Najafi-Asadolahi and Fridgeirsdottir (2014) consider a GD model, based on CPC pricing scheme, in which a publisher is faced with uncertain demand for advertising slots (sent through an ad network) and uncertain arrival of viewers to its website with uncertain click behavior. For simplicity, let the web publisher's website have a single page with n similar slots for ads. The publisher deals with one ad campaign and one viewer type. Advertisers arrive (sent by an ad network) according to a Poisson process with rate λ. Each advertiser requests its ad to be posted on the page until clicked by x viewers. Likewise, viewers arrive at the publisher's page according to a Poisson process with rate μ. Each viewer clicks on an ad on the page with probability β or leaves without clicking on any ad with probability $1 - \beta$. Let $\hat{\mu} := \beta\mu$ be the effective rate with which viewers click on one of the ads on the publisher's page. The publisher's goal is to maximize its total revenue rate by determining the appropriate prices to charge per click. The revenue rate consists of the payments made by advertisers multiplied by the "actual" demand rate. Each payment consists of the price per click, p, multiplied by the number of clicks requested, x. The authors capture the price-sensitivity of the advertisers with the inverse-demand (price) function, $p(\lambda, x, n)$, which is assumed to be continuous and (weakly) decreasing in the advertisers' arrival rate, the number of clicks, and the number of slots. As there is a one-to-one relationship between $p(\lambda, x, n)$ and λ, one can optimize the revenue rate with respect to λ and then determine $p(\lambda, x, n)$. The optimization problem of the publisher is

$$\max_{\lambda\in[0,+\infty)} R(\lambda), \quad \text{with } R(\lambda) = \lambda(1 - \mathbb{P}_n(\lambda; x, n, \beta, \hat{\mu}))p(\lambda, x, n)x, \tag{5.24}$$

where $\hat{\mu} = \beta\mu$ is the viewers' effective arrival rate and $\mathbb{P}_n$ is the steady-state probability that all the slots are occupied. Hence, $\lambda(1 - \mathbb{P}_n(\lambda; x, n, \beta, \hat{\mu}))$ is the advertisers' "actual" arrival rate into the system. The authors show that the steady-state probability that i slots are occupied is $\mathbb{P}_i = (rx)^i / \sum_{j=0}^{n}(rx)^j$, $(i = 0, 1, 2, \ldots, n)$. Araman and Fridgeirsdottir (2011) consider a similar setting but focus on direct GD contracts in which advertisers can wait for a particular publisher's page to become available. They study the optimal CPM pricing when the publisher allocates the arriving impressions uniformly (evenly) among the ads. Although ads are displayed uniformly, they can suffer delays due to demand and supply uncertainties. In addition to the optimal CPM price, they determine the optimal display frequency for this uniform allocation policy by using a large-capacity system analysis. Fridgeirsdottir and Najafi-Asadolahi (2017) consider a similar display advertising setting in which a web publisher posts display ads on its website and charges based on the CPM pricing scheme while promising to deliver a certain number of impressions on the ads posted. The publisher is faced with uncertain demand for advertising slots and uncertain supply of visits from viewers. Advertisers specify various attributes of viewers, and request their ads to be displayed only to those viewer types (targeting). They formulate the problem as a queuing system and show that the optimal CPM price can increase in the number of impressions made of each ad, which is in contrast to the quantity-discount commonly offered in practice.

Ad Pricing in NGD Contracts NGD contracts have become popular, largely due to the emergence and success of advertising exchanges. As explained previously, an ad exchange provides a platform for publishers (or ad networks) to conduct auctions to sell a wide range of impressions one at a time to advertisers or their ad agencies (Muthukrishnan 2009). As the delivery of each impression to advertisers is immediate, these markets are also referred to as "spot markets." An immense variety of impressions (related to different viewer types) are sold in an ad exchange. Most significantly, an ad exchange facilitates the trades of impressions among geographically dispersed advertisers and publishers and allows for real-time bidding (RTB) by advertisers or their ad agencies. NGD display contracts have two major differences with sponsored search. First, unlike sponsored search in which advertisers place their bids for predefined keywords beforehand, in RTB advertisers are allowed to use automated algorithms (programmatic buying) to bid for an impression quite shortly after its arrival (e.g., in about 140 ms as shown in Fig. 5.5) (Yuan et al. 2013; Zhang et al. 2014). The bid can differ from one impression to the next based on how much the impressions' attributes are targeted by the advertisers.

To determine the optimal bidding strategy, an advertiser determines the expected utility from displaying the ad and the payment made for it. The expected utility from displaying the ad consists of the winning probability multiplied by the CTR (probability that the ad is clicked) times the expected value of the click (i.e., the purchase probability (conversion rate) times the price of the product or service). Second, unlike the sponsored search in which CPC is generally used, in RTB, the typical pricing scheme is CPM.

Several works have considered RTB for advertisers and the majority consider the budget or spending constraint. Ghosh et al. (2009b) consider a bidding agent that must win d impressions and has a total budget T. Suppose that the bidding agent knows the total supply n of impressions satisfying the targeting requirements. Let $f = d/n$ be the fraction of the supply that the agent must win, and $t = T/d$ be target spend-per-impression won. Suppose that the *highest* bids from other bidders are drawn i.i.d. from a distribution with CDF $\mathscr{P}$ for each impression sold through the auction. If the distribution $\mathscr{P}$ is *known* to the advertiser, then the bid that wins the fraction f of the supply is $z^* = \mathscr{P}^{-1}(f)$. Define p^* such that $\mathbb{E}_{\mathscr{P}}[X|X \leq p^*] = t$. That is, p^* is the bid price that reaches the target spending in expectation. If $z^* \leq p^*$, then bidding p^* with probability $A = f/\mathscr{P}(p^*)$ independently on each impression achieves both the targets on buying of f and the spending of t in expectation (because $\mathscr{P}(p^*)n \cdot f/\mathscr{P}(p^*) = d$, and $\mathbb{E}_{\mathscr{P}}[X|X \leq p^*] = t$). In practice, an advertiser does not know $\mathscr{P}$. So, it needs to learn it to meet the target quantity and spend constraints. However, learning incurs a penalty, leading to an explore-exploit trade-off. The nature of the penalty depends on the assumptions made about the extent of information available to the bidding agent. If the advertiser can fully observe the winning bid, then it can explore the first m opportunities by not bidding and learning the empirical distribution of $\mathscr{P}$ and then use it as an approximation. If the information is not fully observable, the authors assume that the advertiser yet can observe the maximum bid by paying a cost. In that case, the authors suggest a learning algorithm called *Guess-Double-Panic* in which the advertiser should start with a safe bid and increase it exponentially exploring enough with each new bid to learn the distribution below this bid.

Chen et al. (2011) propose an RTB algorithm that enables targeting viewers with real-time conversion data, and adjusts bids according to budget consumption levels. They show that under a linear programming primal-dual formulation, the simple real-time bidding algorithm is an online solver to the original primal problem by taking the optimal solution of the dual problem as input.

Balseiro et al. (2015) consider advertisers with budget constraints who participate in repeated auctions and need bidding strategies to optimize the allocation of their budgets to incoming impressions. Specifically, they consider a continuous-time infinite horizon setting in which viewers arrive at the publisher's page according to a Poisson process with rate η. Upon the arrival of a viewer, the publisher may send the viewer to an ad exchange, where an auction is run to choose the winning advertiser. Advertisers arrive at the exchange according to a Poisson process with rate λ. Advertiser k is characterized by a type $\theta_k = (b_k, s_k, \alpha_k, \gamma_k)$ in which b_k is advertiser k's budget; s_k is its campaign length; α_k is the probability that advertiser k matches a viewer (independently and at random). Conditional on a match, advertiser k has a private (and independent) value for a viewer, characterized by $F_v(\cdot; \gamma_k)$, parameterized by γ_k over the support $[\underline{V}, \overline{V}]$. The advertisers' utility is the sum of the values from the impressions won minus the payments made during the campaign length. The publisher obtains the revenue $c > 0$ for each impression not won by any advertiser in the exchange (opportunity cost). The publisher's payoff is the long-run average profit rate generated by the auctions, where the profit is the difference between the payment from the auction and c when the impression is won by an advertiser. The publisher's objective is to maximize its payoff by adjusting the reserve price r to set for the auctions.

Similar to Ghosh et al. (2009a) and Iyer et al. (2014), the authors assume that—because many advertisers are in the ad exchange market—the distribution of competitors' bids is fully *stationary*. In addition, the authors assume that the budget constraint must be satisfied in *expectation*. The authors call these conditions as *fluid mean-field* (FMF). To win an auction, an advertiser competes against other bidders as well as the reserve price r. Let D be the steady-state *competitors' highest bid*, which is *independent* and *identically distributed* (i.i.d.) across auctions. Then, the problem, for a bidder of type $\theta = (b, s, \alpha, \gamma)$ optimization problem is

$$\max_{w(\cdot)} \ \alpha \eta s \mathbb{E}[\mathbf{1}_{\{D \le w(V)\}}(V - D)], \tag{5.25}$$

$$\text{s.t.} \quad \eta s \mathbb{E}[\mathbf{1}_{\{D \le w(V)\}}(V - D)] \le b. \tag{5.26}$$

5.4.4 Other Operational Challenges for Advertisers and Publishers

In addition to the challenges described thus far, the publisher faces several others, which we have chosen to not review in detail given our scope outlined at the beginning of this paper. However, we briefly outline some of these challenges here.

Ad Design An essential element in any ad is the effectiveness of its content. Thus, the creative element of ad design, including ad formats (e.g., banners, floating ads, rectangles, skyscrapers, and pop-ups), remains an important design decision for advertisers and ad agencies. Studies such as Burns and Lutz (2006) have examined consumer attitudes towards different online ad formats.

Objective Selection We have noted before that, advertisers have a variety of objectives that drive their advertising decisions, some related to long-term and others related to short-term goals. Examples of metrics used typically (many of which were deployed in the papers described earlier) include the number of impressions (Danaher et al. 2010), clicks (Chatterjee et al. 2003), viewer visits (Dalessandro et al. 2015), and purchases (Manchanda et al. 2006). While the chosen objective will drive outcomes and decisions, it is not always clear which is the best objective to use.

User Query Match The publisher must match relevant ads to the query specified by viewers. A viewer's query could be an *exact match* with the keyword that an advertiser pays for. However, more often, the query language and the keywords do not match exactly. The query language might have a partial *match* with the keywords that advertisers pay for. For example, for a search query "How much is the Omega watch?" the relevant ads might be those who have placed bids for the terms "Omega," "watch," "how much," "local watch seller," etc. In such a case, choosing the most relevant ads is quite challenging. Also, the query might not have any direct overlap, but the content might be relevant to the ad campaign. For example, if a viewer misspells "flowers" as "flwers," the search engine might still match ads that have placed bids for "flowers." As another example, for the query "how much tesla," relevant ads may have placed bids for the keyword "price" rather than "how much" (Yin et al. 2016). Search engines conduct an *advanced match* using *query fragment rewriting*. Query fragment rewriting generates a list of paraphrases for a search query (for example, if a viewer types "ping pong rackets," the search engine also considers searching for the alternate wordings generated by its query rewriting algorithm, such as "ping pong paddles," "table tennis rackets," "table tennis racquets," "table tennis paddles," etc.). Search engines have developed scoring functions to determine relevant ads based on *similarity search algorithms* that are extensively studied as parts of the *information retrieval theory*.

Ad Sorting Publishers and search engines must determine which ad to display in which location on a page. This is important as there are multiple locations on a page to display ads, and different locations have different attractiveness for viewers. Higher ranked ads are placed in more attractive locations, and may generate more revenues for the publisher. When a search engine determines the ads with similar relevance levels to the search query, it orders them based on some ordering rule. As noted earlier, the ranking of ads is based either on their bid prices (bid ranking) or on their expected revenues (revenue ranking) (Edelman et al. 2007, Karande et al. 2013).

Information Asymmetry There is a fundamental incentive incompatibility between the publishers and the advertisers because of their differences in availability of information. Publishers have a greater level of detailed information about their viewers. They can share this information with advertisers allowing them to make better decisions about targeting decisions or about bidding decisions in RTB contracts. However, withholding this information can allow publishers to price-discriminate and deal with the uncertainty of viewer arrivals by bundling impressions. Hummel and McAfee (2016) analyze when improved targeting increases revenues, and show that improved targeting increases revenues when there is a sufficiently large number of bidders, but targeting can hurt revenues when there are just a few bidders. Fu et al. (2012) show that for second price auctions with reserve prices, increased information can decrease expected revenue. Choi et al. (2017) provide other references related to this question.

Attribution The goal of any advertising, display or search based, is to ensure that the viewer undertakes the desired action (view, click, purchase, etc.). However, in general, the final action, whenever it happens, may be the result of one of the several interactions that the viewer may have had with the advertiser. It is important to the advertiser as well as the publisher to understand which ad event contributed to this final action and in which way. This challenge of assigning the appropriate level of credit to the appropriate ad event is referred to as "attribution." This is important because it can influence decisions and resources allocated to various ad events (Geyik et al. 2014). Various heuristic (e.g., Kee 2012) as well as analytical model-based (e.g., Anderl et al. 2014) attribution approaches have been used in the literature.

Invalid Traffic and Fraud Issues Correctly identifying users for targeting and retargeting purposes, identifying them across various Internet domains and across platforms continues to be a significant technical challenge for all participants in the digital supply chain. Another important technical challenge that is a serious concern to industry as well as academics is ad fraud, defined by Google as "invalid traffic including both impressions, clicks, and conversions which are not to be the result of the genuine user interests." According to published reports by Interactive Advertising Bureau, ad fraud costs the media industry billions of dollars in lost revenues. Such fraud can be in the form of fake bids generated that advertisers bid on, fake clicks on ads, and fake actions on ads. Such fraudulent activities arise from pay-per-view networks, botnets, competitors, hired spammers, as well as keyword stuffing, impression stuffing and coercive actions by unethical publishers. Therefore, fraud detection and prevention are significant challenges for all participants in the digital advertising industry.

Incentive Alignment A recently emerging topic in digital advertising is "incentive alignment" between the advertisers and the ad agencies that represent them. When bidding on behalf of their clients (i.e., advertisers), the ad agencies aim to maximize their internal revenues, in addition to meeting each campaign's promised goal. Allouah and Besbes (2017) propose a framework to analyze the implications of such a campaign coordination role by ad agencies, taking as a benchmark the case in which each ad agency would manage the bidding process of each advertiser that it represents independently of other clients, a case that the authors refer to as multi-bidding. The authors show that the adoption of multi-bidding by all ad agencies would lead to an increase in both the social welfare and the publisher's revenues.

Incorporating Learning Another topic considered recently is the learning of the distribution of the value of clicks or impressions, or the bidding distribution during the bidding process. Ghosh et al. (2009b) study learning about the distribution of the winning bid. Amin et al. (2013) consider the problem of learning an advertiser's per-click value distribution when it is repeatedly interacting with a publisher through a posted-price mechanism. Jiang et al. (2014), Iyer et al. (2014), and Cai et al. (2017) are other examples that incorporate learning in different ways.

5.5 Conclusion

The digital advertising industry has witnessed exponential growth since its inception. Currently, a $200 billion industry, it serves a critical role in enabling the matching processes between viewers and advertisers to meet the advertisers' objectives effectively. The digital advertising supply chain allows publishers to match the supply of viewers with the demand they receive from advertisers. In addition, its Internet-based infrastructure provides a basis for designing optimization tools for solving the unique operational challenges faced by publishers and advertisers.

In this chapter, we summarized some of the more important challenges faced by advertisers and publishers, and surveyed a broad range of representative papers on digital advertising to highlight the major contributions to the state-of-art research in this field. Given the trends in the digital advertising industry, we believe that there is a growing need for further research in this area, to help to develop new operational decision-making models.

In particular, we believe that the current research on digital advertising can be extended in at least three directions. First, given that multiple entities are involved in the digital advertising supply chain, each can have different and often conflicting objectives. Modeling how the incentive misalignment between the different parties can affect the operational decisions has received minimal attention in the literature, and is a very interesting direction. Second, most of the works we presented in this chapter consider only a monopolistic web publisher. Exploring competition between two or more publishers offering the same or different pricing schemes and having symmetric or asymmetric information structures is another interesting research direction. Finally, almost all papers studying inventory allocation have analyzed the publisher's problem by assuming the uniform delivery of impressions to advertisers across time. Although this uniform allocation policy is prevalent in practice due to its simplicity and ease of use, it might not be optimal. For example, the uniform allocation policy could be directly tied with the advertisers' smooth budget pacing policy discussed earlier. When viewers arrive at the publisher's page at time varying rates, it makes sense for publishers to change the pace of the delivery of the impressions to advertisers dynamically. Optimizing the rate at which the impressions are delivered to advertisers at each point in time is another interesting direction.

References

Abhishek V, Hosanagar K (2013) Optimal bidding in multi-item multislot sponsored search auctions. Oper Res 61(4):855–873

Abrams Z, Mendelevitch O, Tomlin J (2007) Optimal delivery of sponsored search advertisements subject to budget constraints. In: Proceedings of the 8th ACM conference on electronic commerce. ACM, New York, pp 272–278

Adler M, Gibbons PB, Matias Y (2002) Scheduling space-sharing for Internet advertising. J Sched 5(2):103–119

Agarwal A, Hosanagar K, Smith M (2011) Location, location, location: an analysis of profitability of position in online advertising markets. J Mark Res 48(6):1057–1073

Agarwal D, Ghosh S, Wei K, You S (2014) Budget pacing for targeted online advertisements at LinkedIn. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1613–1619

Agrawal S, Wang Z, Ye Y (2014) A dynamic near-optimal algorithm for online linear programming. Oper Res 62(4):876–890

Ailawadi K, Beauchamp J, Donthu N, Gauri D, Shankar V (2009) Communication and promotion decisions in retailing: a review and directions for future research. J Retail 85(1):42–55

Allouah A, Besbes O (2017) Auctions in the online display advertising chain: a case for independent campaign management. Columbia Business School Research Paper No. 17-60. Available at SSRN: https://ssrn.com/abstract=2919665

Amin K, Rostamizadeh A, Syed U (2013) Learning prices for repeated auctions with strategic buyers. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ (eds) Advances in neural information processing systems, vol 26. Curran Associates, Inc., Red Hook, pp 1169–1177

Amiri A, Menon S (2006) Scheduling web banner advertisements with multiple display frequencies. IEEE Trans Syst Man Cybern Part A Syst Hum 36(2):245–251

Anderl E, Becker I, Wangenheim F, Schumann J (2014) Mapping the customer journey: a graph-based framework for online attribution modeling. SSRN Working paper

Araman V, Fridgeirsdottir K (2011) A uniform allocation mechanism and cost-per-impression pricing for online advertising. SSRN Working paper

Asdemir K, Nanda K, Varghese J (2012) Pricing models for online advertising: CPM vs. CPC. Inf Syst Res 23(3-Part-1):804–822

Balachander S, Kannan K, Schwartz D (2009) A theoretical and empirical analysis of alternate auction policies for search advertisements. Rev Mark Sci 7(1):1–49

Balseiro S, Feldman J, Mirrokni V, Muthukrishnan S (2014) Yield optimization of display advertising with ad exchange. Manag Sci 60(12):2886–2907

Balseiro S, Besbes O, Weintraub G (2015) Repeated auctions with budgets in ad exchanges: approximations and design. Manag Sci 61(4):864–884

Balseiro S, Kim A, Mahdian M, Mirrokni V (2017) Budget management strategies in repeated auctions. In: Proceedings of the 26th international conference on World Wide Web, International World Wide Web Conferences Steering Committee, Perth, pp 15–23

Baye MR, Morgan J (2000) A simple model of advertising and subscription fees. Econ Lett 69(3):345–351

Bharadwaj V, Ma W, Schwarz M, Shanmugasundaram J, Vee E, Xie J, Yang J (2010) Pricing guaranteed contracts in online display advertising. In: Proceedings of the 19th ACM international conference on information and knowledge management. ACM, Toronto, pp 399–408

Bharadwaj V, Chen P, Ma W, Nagarajan C, Tomlin J, Vassilvitskii S, Vee E, Yang J (2012) Shale: an efficient algorithm for allocation of guaranteed display advertising. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, Beijing, pp 1195–1203

Blattberg RC, Briesch R, Fox E (1995) How promotions work. Mark Sci 14(3_supplement):G122–G132

Borgs C, Chayes J, Immorlica N, Jain K, Etesami O, Mahdian M (2007) Dynamics of bid optimization in online advertisement auctions. In: Proceedings of the 16th international conference on World Wide Web. ACM, Banff, pp 531–540

Brooks N (2004) The Atlas rank report: how search engine rank impacts traffic. Insights, Atlas Institute Digital Marketing. Technical report

Buchbinder N, Jain K, Naor J (2007) Online primal-dual algorithms for maximizing ad-auctions revenue. In: European symposium on algorithms. Springer, Berlin, pp 253–264

Bucklin RE, Hoban PR (2017) Marketing models for Internet advertising. In: Wierenga B, van der Lans R (eds) Handbook of marketing decision models. Springer, Berlin, pp 431–462

Burns KS, Lutz R (2006) The function of format: consumer responses to six on-line advertising formats. J Advert 35(1):53–63

Cai H, Ren K, Zhang W, Malialis K, Wang J, Yu Y, Guo D (2017) Real-time bidding by reinforcement learning in display advertising. In: Proceedings of the tenth ACM international conference on Web search and data mining. ACM, Cambridge, pp 661–670

Carpenter J, Moore M (2008) Us consumers' perceptions of non-price retail promotions. Int J Retail Distrib Manag 36(2):111–123

Celis E, Lewis G, Mobius M, Nazerzadeh H (2014) Buy-it-now or take-a-chance: price discrimination through randomized auctions. Manag Sci 60(12):2927–2948

Charles D, Chakrabarty D, Chickering M, Devanur NR, Wang L (2013) Budget smoothing for Internet ad auctions: a game theoretic approach. In: Proceedings of the fourteenth ACM conference on electronic commerce. ACM, Philadelphia, pp 163–180

Chatterjee P, Hoffman D, Novak T (2003) Modeling the clickstream: implications for Web-based advertising efforts. Mark Sci 22(4):520–541

Chen Y (2017) Optimal dynamic auctions for display advertising. Oper Res 65(4):897–913

Chen Y, Berkhin P, Anderson B, Devanur N (2011) Real-time bidding algorithms for performance-based display ad allocation. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, pp 1307–1315

Chickering DM, Heckerman D (2003) Targeted advertising on the web with inventory management interfaces. Interfaces 33:71–77

Choi H, Mela C, Balseiro S, Leary A (2017) Online display advertising markets: a literature review and future directions. SSRN Working paper

Dalessandro B, Hook R, Perlich C, Provost F (2015) Evaluating and optimizing online advertising: forget the click, but there are good proxies. Big Data 3(2):90–102

Danaher PJ, Lee J, Kerbache L (2010) Optimal Internet media selection. Mark Sci 29(2):336–347

Deane J, Agarwal A (2012) Scheduling online advertisements to maximize revenue under variable display frequency. Omega 40(5):562–570

Devanur NR, Hayes T (2009) The adwords problem: online keyword matching with budgeted bidders under random permutations. In: Proceedings of the 10th ACM conference on electronic commerce. ACM, Stanford, pp 71–78

Deza A, Huang K, Metel MR (2015) Chance constrained optimization for targeted internet advertising. Omega 53:90–96

Edelman B, Ostrovsky M, Schwarz M (2007) Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords. Am Econ Rev 97(1):242–259

Feldman J, Muthukrishnan S, Pal M, Stein C (2007) Budget optimization in search-based advertising auctions. In: Proceedings of the 8th ACM conference on electronic commerce. ACM, San Diego, pp 40–49

Feldman J, Korula N, Mirrokni V, Muthukrishnan S, Pál M (2009a) Online ad assignment with free disposal. In: Leonardi S (ed) International workshop on internet and network economics. Springer, Berlin, pp 374–385

Feldman J, Mehta A, Mirrokni V, Muthukrishnan S (2009b) Online stochastic matching: Beating $1 - 1/e$. In: 50th Annual IEEE symposium on foundations of computer science. IEEE, Atlanta, pp 117–126

Feldman J, Henzinger M, Korula N, Mirrokni VS, Stein C (2010) Online stochastic packing applied to display ad allocation. In: de Berg M, Meyer U (eds) European symposium on algorithms. Springer, Berlin, pp 182–194

Feng J, Bhargava H, Pennock D (2007) Implementing sponsored search in Web search engines: computational evaluation of alternative mechanisms. INFORMS J Comput 19(1):137–148

Fjell K (2009) Online advertising: pay-per-view versus pay-per-click – a comment. J Revenue Pricing Manag 8(2/3):200–206

Fridgeirsdottir K, Najafi-Asadolahi S (2017) Cost-per-impression pricing for display advertising. Oper Res 66(3):653–672

Fu H, Jordan P, Mahdian M, Nadav U, Talgam-Cohen I, Vassilvitskii S (2012) Ad auctions with data. In: Serna M (ed) Algorithmic game theory. Springer, Berlin, pp 168–179

Geyik S, Saxena A, Dasdan A (2014) Multi-touch attribution based budget allocation in online advertising. In: Proceedings of the eighth international workshop on data mining for online advertising. ACM, New York, pp 1–9

Ghose A, Yang S (2009) An empirical analysis of search engine advertising: sponsored search in electronic markets. Manag Sci 55:1605–1622

Ghosh A, McAfee P, Papineni K, Vassilvitskii S (2009a) Bidding for representative allocations for display advertising. In: Leonardi S (ed) Internet and network economics. Springer, Berlin, pp 208–219

Ghosh A, Rubinstein B, Vassilvitskii S, Zinkevich M (2009b) Adaptive bidding for display advertising. In: Proceedings of the 18th international conference on World Wide Web. ACM, Madrid, pp 251–260

Goel G, Mehta A (2008) Online budgeted matching in random input models with applications to adwords. In: Proceedings of the nineteenth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics, San Francisco, pp 982–991

Grewal D, Ailawadi K, Gauri D, Hall K, Kopalle P, Robertson J (2011) Innovations in retail pricing and promotions. J Retail 87:S43–S52

Hojjat A, Turner J, Cetintas S, Yang J (2017) A unified framework for the scheduling of guaranteed targeted display advertising under reach and frequency requirements. Oper Res 65(2):289–313

Hummel P, McAfee R (2016) When does improved targeting increase revenue? ACM Trans Econ Comput 5(1):4

Iyer K, Johari R, Sundararajan M (2014) Mean field equilibria of dynamic auctions with learning. Manag Sci 60(12):2949–2970

Jedidi K, Mela C, Gupta S (1999) Managing advertising and promotion for long-run profitability. Mark Sci 18(1):1–22

Jerath K, Sarvary M (2017) A primer on programmatic advertising. Columbia CaseWorks, ID number CU180, Columbia CaseWorks, New York

Jiang C, Beck C, Srikant R (2014) Bidding with limited statistical knowledge in online auctions. ACM SIGMETRICS Perform Eval Rev 41(4):38–41

Johnson GA, Lewis RA, Nubbemeyer EI (2017) Ghost ADS: improving the economics of measuring online AD effectiveness. J Mark Res 54(6):867–884

Karande C, Mehta A, Srikant R (2013) Optimizing budget constrained spend in search advertising. In: Proceedings of the sixth ACM international conference on Web search and data mining. ACM, Rome, pp 697–706

Katona Z, Sarvary M (2010) The race for sponsored links: bidding patterns for search advertising. Mark Sci 29(2):199–215

Kee B (2012) Attribution playbook – Google analytics. http://services.google.com/

Kim A, Balachander S, Kannan K (2012) On the optimal number of advertising slots in a generalized second-price auction. Mark Lett 23(3):851–868

Kitts B, Leblanc B (2004) Optimal bidding on keyword auctions. Electron Mark 14(3):186–201

Korula N, Mirrokni V, Nazerzadeh H (2016) Optimizing display advertising markets: challenges and directions. IEEE Internet Comput 20(1):28–35

Kosorin D (2016) Introduction to programmatic advertising. Self-published. Wrocław, Poland

Kumar S, Sethi SP (2009) Dynamic pricing and advertising for web content providers. Eur J Oper Res 197(3):924–944

Kumar S, Jacob VS, Sriskandarajah C (2006) Scheduling advertisements on a web page to maximize revenue. Eur J Oper Res 173(3):1067–1089

Langheinrich M, Nakamura A, Abe N, Kamba T, Koseki Y (1999) Unintrusive customization techniques for web advertising. Comput Netw 31(11–16):1259–1272

Lee K, Jalali A, Dasdan A (2013) Real time bid optimization with smooth budget delivery in online advertising. In: Proceedings of the seventh international workshop on data mining for online advertising. ACM, Chicago, pp 1–13

Leischnig A, Schwertfeger M, Geigenmueller A (2011) Do shopping events promote retail brands? Int J Retail Distrib Manag 39(8):619–634

Levy M, Grewal D, Kopalle PK, Hess J (2004) Emerging trends in retail pricing practice: implications for research. J Retail 80(3):xiii–xxi

Lewis RA, Reiley D (2014) Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo! Quant Mark Econ 12(3):235–266

Manchanda P, Dubé JP, Goh KY, Chintagunta PK (2006) The effect of banner advertising on Internet purchasing. J Mark Res 43(1):98–108

Mangani A (2003) Online advertising: Pay-per-view versus pay-per-click. J Revenue Pricing Manag 2:295–302

McAfee RP, Papineni K, Vassilvitskii S (2013) Maximally representative allocations for guaranteed delivery advertising campaigns. Rev Econ Des 17:83–94

Mookerjee R, Kumar S, Mookerjee V (2016) Optimizing performance-based internet advertisement campaigns. Oper Res 65(1):38–54

Mookerjee R, Kumar S, Mookerjee VS (2017) Optimizing performance-based internet advertisement campaigns. Oper Res 65(1):38–54

Muthukrishnan S (2009) Ad exchanges: Research issues. In: Leonardi S (ed) Internet and network economics. Lecture notes in computer science, vol 5929. Springer, Berlin, pp 1–12

Najafi-Asadolahi S, Fridgeirsdottir K (2014) Cost-per-click pricing for display advertising. Manuf Serv Oper Manag 16(4):482–497

Nakamura A, Abe N (2005) Improvements to the linear programming based scheduling of Web advertisements. Electron Commer Res 5(1):75–98

Narayanan S, Kalyanam K (2015) Position effects in search advertising and their moderators: a regression discontinuity approach. Mark Sci 34(3):388–407

Özlük Ö, Cholette S (2007) Allocating expenditures across keywords in search advertising. J Revenue Pricing Manag 6(4):347–356

Prasad A, Mahajan V, Bronnenbert B (2003) Advertising versus pay-per-view in electronic media. Int J Res Mark 20:13–30

Roels G, Fridgeirsdottir K (2009) Dynamic revenue management for online display advertising. J Revenue Pricing Manag 8(5):452–466

Rubinson J (2009) Empirical evidence of TV advertising effectiveness. J Advert Res 49(2):220–226

Rusmevichientong P, Williamson DP (2006) An adaptive algorithm for selecting profitable keywords for search-based advertising services. In: Proceedings of the 7th ACM conference on electronic commerce. ACM, Ann Arbor, pp 260–269

Rutz O, Bucklin R (2012) Does banner advertising affect browsing for brands? Clickstream choice model says yes, for some. Quant Mark Econ 10(2):231–257

Taylor G, Neslin S (2005) The current and future sales impact of a retail frequency reward program. J Retail 81(4):293–305

Tomlin JA (2000) An entropy approach to unintrusive targeted advertising on the Web. Comput Netw 33(1–6):767–774

Turner J (2012) The planning of guaranteed targeted display advertising. Oper Res 60(1):18–33

Varian H (2007) Position auctions. Int J Ind Organ 25(6):1163–1178

Vazirani U, Vazirani V, Mehta A, Saberi A (2005) Adwords and generalized on-line matching. In: Proceedings of FOCS. IEEE, Pittsburgh

Wang J, Zhang W, Yuan S (2017) Display advertising with real-time bidding (RTB) and behavioural targeting. Found Trends Inf Retr 11(4–5):297–435

Xu J, Lee K, Li W, Qi H, Lu Q (2015) Smart pacing for effective online ad campaign optimization. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, Sydney, pp 2217–2226

Yang J, Vee E, Vassilvitskii S, Tomlin J, Shanmugasundaraml J, Anastasakos T, Kennedy O (2010) Inventory allocation for online graphical display advertising. Yahoo technical report

Yang Y, Qin R, Jansen B, Zhang J, Zeng D (2014) Budget planning for coupled campaigns in sponsored search auctions. Int J Electron Commer 18(3):39–66

Ye S, Aydin G, Hu S (2014) Sponsored search marketing: dynamic pricing and advertising for an online retailer. Manag Sci 61(6):1255–1274

Yin D, Hu Y, Tang J, Daly T, Zhou M, Ouyang H, Chen J, Kang C, Deng H, Nobata C et al (2016) Ranking relevance in yahoo search. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, San Francisco, pp 323–332

Yuan S, Wang J, Zhao X (2013) Real-time bidding for online advertising: measurement and analysis. In: Proceedings of the seventh international workshop on data mining for online advertising. ACM, Chicago, p 3

Zhang W, Zhang Y, Gao B, Yu Y, Yuan X, Liu T (2012) Joint optimization of bid and budget allocation in sponsored search. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining, August 12–16, 2012. ACM, Beijing, pp 1177–1185

Zhang W, Yuan S, Wang J (2014) Optimal real-time bidding for display advertising. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1077–1086

Zhao L, Nagurney A (2005) A network modeling approach for the optimization of internet-based advertising strategies and pricing with a quantitative explanation of two paradoxes. NETNOMICS: Econ Res Electron Netw7(2):97–114

Zhou Y, Chakrabarty D, Lukose R (2008) Budget constrained bidding in keyword auctions and online knapsack problems. In: Papadimitriou C, Zhang S (eds) International workshop on internet and network economics. Springer, Berlin, pp 566–576

Chapter 6
New Models of Strategic Customers in the Age of Omnichannel Retailing

Fei Gao and Xuanming Su

Abstract In the omnichannel era, consumers optimize their shopping experience by exhaustively considering all possible alternatives across both online and offline channels. In this chapter, we present new approaches to model consumer behavior in the omnichannel environment. We start by reviewing traditional models of strategic consumer behavior and then apply them to omnichannel initiatives in the retail industry. These omnichannel strategies help mitigate two key problems in retail: stockouts and product misfit. We hope that our models can inspire future research in the emerging area of omnichannel retailing.

6.1 Introduction

With access to a range of new technologies and a wide variety of online resources, connected consumers are becoming increasingly sophisticated. These consumers optimize their shopping experience by exhaustively considering all possible alternatives across both online and offline channels; we refer to such shoppers as strategic customers. Omnichannel retailing acknowledges this type of consumer behavior by integrating the best of both digital and physical worlds at each step of the customer experience (Rigby 2011).

Different channels provide customers with different shopping experiences. Offline channel allows customers to physically try on the product and therefore reduces the customer's uncertainty over the value of the product. However, shopping in a physical store may involve the risk of encountering stockouts after incurring the hassle of traveling there. In contrast, online channel provides customers with

F. Gao
Kelley School of Business, Indiana University, Bloomington, IN, USA
e-mail: fg1@iu.edu

X. Su (✉)
The Wharton School, University of Pennsylvania, Philadelphia, PA, USA
e-mail: xuanming@wharton.upenn.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_6

an environment to easily find product in-stock status on a website. The downside of shopping online, however, is that customers may not be able to fully evaluate the product in the digital world without the physical touch-and-feel experience, and therefore may end up buying and returning an unwanted product.

Two of the key considerations for retailers are stockouts and product misfit. In this chapter, we present models to study these two issues in retail operations. We first review traditional models of strategic consumer behavior in the presence of stockouts or product misfit; these models mainly focus on a single-channel environment. We then extend them to an omnichannel setting where the retailer operates both online and offline channels in an integrated way while strategic consumers face uncertainty over both inventory availability and product fit.

6.2 Consumer Response to Stockouts

Su and Zhang (2009) builds a newsvendor model where the store makes pricing and inventory decisions and customers strategically decide whether or not to incur the hassle to visit the store given the stockout risk. We present their basic model in this section.

The retailer makes decision about price p and stock level q at unit cost c. Market demand D is random with distribution F. Each individual consumer has valuation v for the product and faces a search cost of h (e.g., transportation cost, shopping time, etc.) in order to visit the retailer. If the product is out of stock, this sunk cost is not recoverable. Price p is observable to customers; but stock level q is not. Before visiting the retailer, consumers have belief about the in-stock probability $\hat{\xi}$. Throughout this chapter, we use the hat notation $\hat{\cdot}$ to denote beliefs. Under this belief $\hat{\xi}$, the consumer expects to earn surplus $(v - p)\hat{\xi} - h$ from visiting the retailer (and zero otherwise). Therefore, given belief $\hat{\xi}$, consumer's behavior follows a threshold rule: They visit the retailer if and only if $p \leq r$, where the reservation price r is given by $(v - r)\hat{\xi} - h = 0$.

To maximize expected profit, the retailer should set the price equal to the consumer's reservation price, because this is the highest price at which consumers are willing to visit the retailer. However, the retailer does not know consumer's reservation prices; instead, he only has a belief $\hat{r}$ about it. Given this belief, the retailer chooses price $p = \hat{r}$ and stocks the critical fractile quantity $q = \bar{F}^{-1}(c/p)$ to maximize the newsvendor profit function $\pi(q) = pE \min(D, q) - cq$, where the operator E means expectation.

To study the strategic interaction between the retailer and the consumers, the rational expectations (RE) equilibrium concept is commonly used in the literature on retail operations. One important feature of a RE equilibrium is that beliefs must be consistent with actual outcomes. In other words, the retailer's belief $\hat{r}$, must coincide with consumers' reservation price r, and consumers' belief over availability probability $\hat{\xi}$ must agree with the actual in-stock probability corresponding to the retailer's chosen quantity q. In fact, this probability is given by $A(q) = E\min(D, q)/ED$ (Deneckere and Peck 1995; Dana 2001).

Definition 6.1 An RE equilibrium consists of the quantities p, q, r, $\hat{\xi}$, $\hat{r}$ satisfying

(i) $r = v - h/\hat{\xi}$;
(ii) $\bar{F}(q) = c/p$, $p = \hat{r}$;
(iii) $\hat{\xi} = A(q)$, $\hat{r} = r$.

The first two conditions (i) and (ii) follow, respectively, from the consumer's optimal decision r given belief $\hat{\xi}$, and the retailer's optimal decisions p, q given belief $\hat{r}$. Condition (iii) requires that both the retailer and consumers' beliefs are correct in equilibrium.

Proposition 6.1 *There exists some $\bar{h} < v - c$ such that an RE equilibrium exists if and only if $h < \bar{h}$. In any RE equilibrium, the retailer's price p^* and quantity q^* satisfy* (i) $\bar{F}(q^*) = c/p^*$ *and* (ii) $(v - p^*)A(q^*) = h$.

This proposition is shown by solving the equilibrium conditions of Definition 6.1. To avoid trivial outcomes and guarantees the existence of an RE equilibrium, the following assumption is imposed.

Assumption 6.1 *Consumer search cost h satisfies $h < \bar{h}$.*

Su and Zhang (2009) further studies the case where the retailer could credibly announce the stock level q in the beginning. With this credible quantity commitment, in any equilibrium, consumer's belief $\hat{\xi}$ must be consistent with the actual availability probability $A(q)$, which is determined by the seller's committed quantity q. Thus, they are willing to visit the retailer as long as price does not exceed the reservation price $r = v - h/A(q)$. Corresponding to quantity q, it is optimal for the retailer to charge price $p(q) = v - h/A(q)$. Then, the optimal order quantity q can be derived by maximizing the following profit function $\pi = p(q)E\min(D, q) - cq = (v - h/A(q))E\min(D, q) - cq = vE\min(D, q) - hED - cq$, and thus by setting q satisfying $\bar{F}(q) = c/v$.

Let us use superscript $\cdot^c$ to denote the case with quantity commitment. The following proposition characterizes market outcomes in this case and is proven similarly as Proposition 6.1.

Proposition 6.2 *Suppose that the retailer can commit to quantity q. In equilibrium, the seller's price p^c and quantity q^c satisfy* (i) $\bar{F}(q^c) = c/v$ *and* (ii) $(v - p^c)A(q^c) = h$.

Besides committing to the initial quantity level q, the retailer may also share real-time inventory information with customers. There is a key difference between quantity commitment and the real-time inventory information. Quantity commitment merely informs consumers about the retailer's *initial* inventory level, and thus consumers may still encounter stockouts when they visit the store. Consider the following numerical example. Suppose the retailer stocks an inventory of 50 units but the random demand realization is 80 customers (each wanting one unit). When these customers show up at the store, 30 of them will face stockouts. In contrast, with *real-time* inventory information, consumers no longer face any availability uncertainty. The first 50 customers will learn that the product is in stock (and thus successfully purchase it) while the remaining 30 will learn that the product is out of stock and thus not go to the store at all. A basic premise behind real-time inventory information is that there is a sequential nature to customer arrivals, so different customers might receive different information as the inventory status evolves in real-time.

Following Gao and Su (2017b), we extend the model in Su and Zhang (2009) to study the case where real-time information is available to customers. Since customers are able to know the inventory availability status in store beforehand, there is no in-store stockout risk for them. As a result, consumer are willing to visit the retailer as long as there is still inventory available and the price does not exceed the reservation price $r = v - h$. Therefore, it is optimal for the firm to charge $p = v - h$. Then, the optimal order quantity q can be derived by maximizing the following newsvendor profit function $\pi = (v - h)E \min(D, q) - cq$.

Let us use superscript $\cdot^i$ to denote the case with real-time inventory information. The following proposition characterizes market outcomes in this case:

Proposition 6.3 *Suppose that the retailer shares real-time inventory information with customers. The seller's optimal price p^i and quantity q^i satisfy* (i) $\bar{F}(q^i) = c/p^i$ *and* (ii) $p^i = v - h$.

So far, we have explored three different models. In the first model, consumers anticipate and respond to potential stockouts; we call this the model with strategic customer behavior (SCB). In the second, the retailer additionally commits to an initial stocking quantity; we call this the quantity commitment (QC) model. In the third model, the retailer provides real-time information (RTI) about inventory status. The equilibrium results of these models are summarized in Table 6.1.

Consumer anticipation of and response to potential stockouts makes it hard for the retailer to attract demand and thus hurts operational performance. Comparing to SCB, Table 6.1 suggests that QC and RTI alleviate the stockouts problem through different mechanisms: QC helps support a high retail service level (as the stockout probability is increased from c/v to c/p), while RTI increases customer willingness-to-pay (from $p = v - h/A(q)$ to $p = v - h$) by eliminating the risk of stockouts.

The following proposition compares the retailer optimal profit in the three scenarios above: base case (π^*), with quantity commitment (π^c), and with real-time inventory information (π^i).

Table 6.1 Comparison of equilibrium results

SCB	$\bar{F}(q) = c/p,\ \ p = v - h/A(q)$
QC	$\bar{F}(q) = c/v,\ \ p = v - h/A(q)$
RTI	$\bar{F}(q) = c/p,\ \ p = v - h$

Proposition 6.4 $\pi^* < \pi^c < \pi^i$.

Proposition 6.4 shows that sharing inventory information (either through QC or through RTI) with customers is valuable to the retailer. Each approach allows the retailer to recover some of the profit loss due to consumer anticipation of potential stockouts. However, they operate in different ways. Under quantity commitment, consumers are better able to assess their chances of securing the product if they do visit the retailer (at a cost of h). All else equal, this encourages consumer visits and increases their willingness-to-pay. This effect in turn increases the retailer's profit. With real-time inventory information, consumer stockout risk is fully eliminated, and thus the retailer is able to capture the entire consumer surplus $v - h$, which further improves the profit. Comparing RTI and QC, the above proposition shows that the former performs better. This is a new result that has not been found in Su and Zhang (2009) and Gao and Su (2017b), and it demonstrates the importance of leveraging current technologies to provide real-time inventory information to consumers.

6.3 Consumer Response to Product Misfit

When making a purchase decision, consumers may also face uncertainty over the value of the product, especially in the online channel due to the lack of physical touch-and-feel experience. Davis et al. (1995) build a model that incorporates consumer valuation uncertainty risk and studies the impact of a specific return policy, i.e., money back guarantee. Below we present a simplified version of the model in Davis et al. (1995).

Consider a retailer who sells a product that can be fully evaluated only after purchase. Consumers are homogeneous ex ante. After the purchase, a fraction θ of customers like the product and obtain a high value v (we refer to them as the "high-type" customers); and the rest $1 - \theta$ do not like the product and obtain 0 value (we refer to them as the "low-type" customers). Thus, consumers realize their types only after purchase. The total market size D is deterministic and normalized to 1. Denote the unit production cost as $c < v$. The firm chooses price p to maximize total profit.

If customers cannot return the product after purchase, then their expected utility from buying the product is $u = \theta v + (1 - \theta)0 - p$. The retailer should set the price $p^* = \theta v$ since this is the highest price at which consumers are willing to make a purchase. Then, the optimal profit is simply given by $\pi^* = (p^* - c)D = \theta v - c$.

If the retailer offers money back guarantee, then consumers can return the product with full refund. Let h_r denote the consumer's hassle cost of returning a product, and let s denote the retailer's salvage value of a returned product. The consumer's decision tree involves two sequential decisions: (i) buy or not, and then, if the consumer buys, (ii) keep or return (after observing his valuation). Clearly, at stage (ii), a low-type consumer (with valuation 0) will return the product if and only if $h_r \leq p$. Thus, each consumer's ex ante expected utility from purchasing the product is $u = \theta(v-p)+(1-\theta)\max(0-p, -h_r)$. Suppose that low-type customers will return the product instead of keeping it, in accordance to the intent of the money back guarantee. Then, each consumer's expected utility is $\theta(v-p) - (1-\theta)h_r$. The retailer should set the price $p^m = v - ((1-\theta)/\theta)h_r$ since this is the highest price at which consumers are willing to make a purchase. Then, the optimal profit is simply given by $\pi^m = p^m\theta D + s(1-\theta)D - cD = \theta v + (1-\theta)(s-h_r) - c$. Comparing π^m and π^*, we find that money back guarantee is more profitable (i.e., $\pi^m > \pi^*$) if and only if the consumer's hassle cost from returning merchandise is smaller than the retailer's salvage advantage (i.e., $h_r < s$). Further, for this to be an equilibrium outcome, low-type consumers must be willing to return the product (as assumed above), i.e., $h_r < p^m = v - ((1-\theta)/\theta)h_r$, which yields $h_r < \theta v$. In summary, money back guarantees are effective if the return hassle cost h_r is smaller than both the consumer's expected valuation θv and the retailer's salvage value s.

Su (2009) further extends the model above by incorporating aggregate demand uncertainty and the possibility of partial refund. Specifically, the market demand D is random with distribution F. Besides the pricing decision p, the retailer also needs to decide the stock quantity q and the refund r to be paid, if consumers choose to return the product (where $r \leq p$). The money back guarantee considered in Davis et al. (1995) is a special case with $r = p$. For simplicity, suppose $h_r = 0$, and thus low-type customers will always return the product and obtain refund r. Therefore, consumer's ex ante expected utility from purchasing the product is $u = \theta v + (1-\theta)r - p$. Note only high-type customers will keep the product. Therefore, the retailer's profit function can be expressed as follows:

$$\begin{aligned}
\pi\,(p,q,r) &= \underbrace{p\theta E\min(D,q)}_{\text{sold}} + \underbrace{(p-r+s)\,(1-\theta)\,E\min(D,q)}_{\text{returned}} \\
&\quad + \underbrace{s\,(q - E\min(D,q))}_{\text{not sold}} - cq \\
&= [(p-s)\,\theta + (p-r)\,(1-\theta)]\,E\min(D,q) - (c-s)\,q,
\end{aligned}$$

where $s < c$ is the salvage value of the returned and unsold products.

Proposition 6.5 *The retailer's optimal price p^*, quantity q^*, and refund r^* are given by $p^* = \theta v + (1-\theta)s$, $\bar{F}(q^*) = (c-s)/[\theta(v-s)]$ and $r^* = s$.*[1]

Proposition 6.5 implies that the optimal policy involves partial refunds (i.e., $r^* < p^*$). Specifically, the optimal partial refund equals the salvage value (i.e., $r^* = s$) as this leads to allocative efficiency: all consumers who value the product at less than the salvage value will return the product. In other words, money back

[1] If consumer valuation takes only two possible values as we assume in this section, then there are actually multiple equilibria (p^*, q^*, r^*) such that $p^* = \theta v + (1-\theta)r^*$, $\bar{F}(q^*) = (c-s)/[\theta(v-s)]$ and $r^* \geq 0$. However, as shown in Su (2009), the equilibrium specified in Proposition 6.5 is the only robust one for a general consumer valuation distribution.

guarantees (or full refunds) offer "too much protection" to consumers, transferring the potential downside of excess inventory and product misfit entirely to the retailer.

Both Davis et al. (1995) and Su (2009) focus on refund policies to mitigate consumer valuation uncertainty risk by reducing their mismatch cost *after* purchase. Another approach that online retailers can take is to provide customers with more product information and therefore reduce the incidence of mismatch *before* purchase. With the development of virtual reality technology, some retailers are able to implement virtual showrooms on their website, so that online shoppers can now try on different products as if they were in the store (Financial Times 2011). For example, on the website of BonLook, an eyewear retailer, consumers can upload their own photos to see how different frames will look on their digital faces. Many advanced technologies are now available from an increasing variety of providers. As another example, Metail provides visualization technology that creates 3D models of shoppers based on a few customized measurements, while Shoefitr uses 3D scanning technology to measure the insides of shoes with accuracy up to a quarter of a millimeter (CNET 2010). Here, we provide a simple way to model the impact of this innovation.

To model virtual showrooms, we follow the approach in Gao and Su (2017b), where consumers receive an imperfect signal of their valuations.[2] For modeling convenience, we assume that the signal is binary, i.e., a group of consumers remain interested in the product while the remaining discover that the product is not for them and leave. Further, we assume that the latter group consists of a fraction $\alpha \in (0, 1]$ of the low-type customers. In other words, if $\alpha = 1$, the virtual showroom offers a perfect signal and all consumers learn their types, but if $\alpha < 1$, the virtual showroom screens out a fraction α of the low types. For those consumers who remain interested in the product, including all potential high-type customers and a fraction $1 - \alpha$ of the potential low-type customers, they update their posterior belief about the probability of being high-type to $\theta' = \theta/[1 - \alpha(1 - \theta)] > \theta$ by Bayes' rule. Then, the two models, with and without virtual showrooms, are very similar. The only difference is that virtual showrooms generate a new consumer pool by filtering away some potential low-type consumers. As a result, the total demand size is $D' = [1 - \alpha(1 - \theta)]D$ and a fraction θ' of them is of high type.

Table 6.2 summarizes the equilibrium outcome under different scenarios. The first row shows the benchmark in which customers are "naive" and simply assume that they are high types. In the second row, customers anticipate potential product misfit (i.e., low types have zero valuation) and thus are willing to pay up to their expected valuation θv. The next two rows summarize the full returns model from Davis et al. (1995) and the partial returns model from Su (2009). The last row presents the virtual showroom model from Gao and Su (2017b).

The possibility of product misfit suppresses customer willingness-to-pay from v to θv, and we have seen two broad ways to address this issue: returns policies and virtual showrooms. Both can raise customer willingness-to-pay above θv.

[2]The model can be extended to other ways of reducing consumer valuation uncertainty; for example, online reviews and sampling.

Table 6.2 Comparison of results

	Prices	Demand
Naive customers	$p = v$	D
No return policy	$p = \theta v$	D
Full returns	$p = \theta v + (1 - \theta)(r - h_r), \quad r = p$	D
Partial returns (if $h_r = 0$)	$p = \theta v + (1 - \theta)r, \quad r = s$	D
Virtual showrooms	$p = \theta' v$	D'

Note: $\theta' = \theta/[1 - \alpha(1 - \theta)] > \theta$, $D' = [1 - \alpha(1 - \theta)]D < D$

However, both may also have some negative impact: generous product returns (e.g., full returns) may lead to a loss in margin, while virtual showrooms are accompanied by a loss in demand, which occurs when customers discover that they are low types.[3] In practice, retailers need to consider the cost of handling returns and the informativeness of the virtual showroom when choosing between the two approaches.

Finally, we compare the retailer's profits under the two strategies discussed in this section, i.e., consumer returns and virtual showrooms. Suppose aggregate demand D is uncertain and the unsold inventory can be salvaged at price s as in Su (2009). We have the following result.

Proposition 6.6 *Compared to the case of partial returns, the retailer obtains higher profit with virtual showrooms if and only if* $\alpha > \bar{\alpha}$ *for some* $\bar{\alpha} \in (0, 1)$.

Therefore, when the retailer implements virtual showrooms (without allowing for product returns), it obtains a higher level of expected profit than that with the return policy specified in Proposition 6.5, as long as the virtual showroom is informative enough (i.e., α is large).

6.4 Omnichannel Model

In Sects. 6.2 and 6.3, we reviewed traditional models of customer strategic shopping behavior in the face of stockout or valuation risks. These models mainly focus on the operations of one single channel (either store or online). Today, however, many retailers have realized the need to integrate their existing channels to enrich customer value proposition and improve operational efficiency. As a result, there is an emerging focus on "omnichannel retailing" with the goal of providing consumers with a seamless shopping experience through all available shopping channels (Rigby 2011). In this section, we present a way to extend the traditional model to the omnichannel environment. The following analysis is based on Gao and Su (2017b).

[3] The latter is similar to how spot selling differs from advance selling, i.e., spot selling could also result in a loss in demand by selling to consumers after they realize their true valuations. Readers can refer to Xie and Shugan (2001) for a detailed discussion.

There is a retailer who sells a product through two channels, store and online, at price p. For simplicity, we assume price is exogenously given. We focus on the retailer's inventory decision in this section.[4] In the store channel, the retailer faces a newsvendor problem: there is a single inventory decision q to be made before random demand is realized. The unit cost of inventory is c, and the salvage value of leftover is zero. The online channel is modeled exogenously: for each unit sold online, the retailer obtains a net profit margin w if it is not returned, and incurs a net loss k otherwise. As shown in Gao and Su (2017b), the model can be extended to the case where online is also a newsvendor problem.

The setup of customer demand is similar as before. The market demand is D with distribution F. There are two types of customers: A fraction of $\theta \in (0, 1)$ are high types, who have positive value v for the product, and the rest are low types with zero valuation for the product. Customers are homogeneous ex ante: they do not know their valuation (or type) beforehand, but θ and v are common knowledge. Customers may learn their valuations before purchase only if they examine the product in store (due to the physical touch-and-feel experience); otherwise, customers learn valuations after purchase which may lead to online returns.

Each consumer makes a choice between shopping online directly or going to the store. If she chooses to buy online directly, she incurs hassle cost h_o (e.g., paying shipping fees or waiting for the product to arrive), and realizes her valuation only after receiving the product. If she likes the product (i.e., she is high type), then she keeps it and receives payoff $v - p - h_o$; if she dislikes the product (i.e., she is low type), then she returns it. Returns are costly to both the retailer and the consumers: each returned unit generates net loss $k > 0$ to the retailer and an additional hassle cost $h_r > 0$ to the consumer. We assume that low-type consumers prefer returning the product to keeping it, i.e., $h_r < p$. Therefore, the consumer's expected payoff from buying online directly is given by

$$u_o = -h_o + \theta(v - p) - (1 - \theta)h_r.$$

On the other hand, if the consumer chooses to go to the store, she has to first incur hassle cost h_s (e.g., traveling to the store or searching for the product in aisles). Once she is in the store, the customer may encounter two possible outcomes: (1) If the store is in stock, then she can evaluate the product on the spot: a high type makes a purchase and receives payoff $v - p$, while a low type leaves without any purchase and receives payoff 0. (2) If the store is out of stock, she cannot resolve her value uncertainty in store, but she can buy the product online and receive an expected payoff u_o instead. Let ξ denote the probability that the store is in stock, and let $\hat{\xi}$ denote consumers' beliefs about this probability. We assume that customers arrive sequentially to the market, but they do not know their own order of arrival. As a result, everyone has the same belief $\hat{\xi}$. Then, given belief $\hat{\xi}$, each consumer's payoff from visiting the store can be expressed as follows:

[4]For the discussions of other firm's decisions in the context of omnichannel retail, readers can refer to other chapters in this book.

$$u_s(\hat{\xi}) = -h_s + \hat{\xi}\theta\,(v - p) + (1 - \hat{\xi})u_o.$$

Consumers compare the expected utility from each channel and chooses accordingly. In the spirit of omnichannel choice, i.e., consumers are willing to consider both channels, we assume v is large enough so that $u_s(1) \geq 0$ and $u_o \geq 0$.

The retailer anticipates that a fraction $\hat{\phi} \in [0, 1]$ of customers will visit the store. Then, if total demand is D, the retailer expects that the number of customers coming to the store will be $\hat{\phi}D$; also, since only high-type customers will eventually make a purchase in the store, given the store inventory level q, the retailer expects that the number of store customers who find that the store is in stock is $D_{\text{in}}\,(q) = \min\left(\hat{\phi}D, q/\theta\right)$, and the remaining $D_{\text{out}}\,(q) = \left(\hat{\phi}D - q/\theta\right)^{+}$ will encounter stockouts when they come to the store. Note that even though the inventory is q, up to q/θ customers may examine the product in the store because only a fraction θ of those customers will buy. Then, the retailer's profit is as follows:

$$\pi\,(q) = p\theta E D_{\text{in}}\,(q) - cq \tag{6.1}$$

$$+ [w\theta - k(1 - \theta)]E D_{\text{out}}\,(q) \tag{6.2}$$

$$+ [w\theta - k(1 - \theta)](1 - \hat{\phi})E D. \tag{6.3}$$

Given the store inventory level q, the newsvendor expected profit from selling the product in the store channel is shown in the first term (6.1) above. The next two terms, respectively, represent profit from customers who switch online after encountering stockouts in the store, and customers who buy online directly. For each unit of online demand, the expected profit is $w\theta - k(1-\theta)$, because a fraction $(1 - \theta)$ of online sales is returned. Again, in the spirit of omnichannel retailing, we assume that $w\theta - k(1 - \theta) > 0$, so that the retailer is willing to operate both channels. Finally, the retailer chooses inventory level q to maximize expected profit.

Definition 6.2 A RE equilibrium $(\phi, q, \hat{\xi}, \hat{\phi})$ satisfies the following:

(i) Given $\hat{\xi}$, if $u_s(\hat{\xi}) > u_o$, then $\phi = 1$; otherwise $\phi = 0$;
(ii) Given $\hat{\phi}$, $q = \arg\max_q \pi(q)$;
(iii) $\hat{\xi} = A(q)$;
(iv) $\hat{\phi} = \phi$.

Conditions (i) and (ii) state that under beliefs $\hat{\xi}$ and $\hat{\phi}$, the consumers and the retailer are choosing the optimal decisions. Conditions (iii) and (iv) are the consistency conditions.

The following proposition gives the (Pareto optimal) equilibrium result; we use the superscript $(\cdot^{\circ})$ to denote the equilibrium outcome for this base scenario.

Proposition 6.7 *There exists a threshold* $\psi^{\circ} \in [0, 1]$ *such that*

- *if* $\theta < \psi^\circ$*, then consumers visit the store and*

$$q^\circ = \theta \bar{F}^{-1}\left(\frac{c}{p - w + k(1-\theta)/\theta}\right) > 0;$$

- *if* $\theta \geq \psi^\circ$*, then consumers buy online directly and* $q^\circ = 0$.

6.4.1 Mitigating Stockouts In-Store: Real-time Inventory Information

Suppose the retailer provides real-time store inventory information on its website. That has been increasingly common in industry. Online shoppers can simply enter a zip code to check the current status of nearby stores, although the information may be presented in different ways: Some retailers, e.g., IKEA, show the exact store inventory level online, while others, e.g., CVS and Walgreens, simply tell their online customers whether or not the product is currently available in store. Recently, some retailers, e.g., Target and Macy's, allow their customers to buy online and pick up their order in the store shortly after. Through the availability of this in-store pickup option, customers can also infer the real-time inventory status in store (Gallino and Moreno 2014). Readers can refer to Gao and Su (2017a) for an analytical model of this new fulfillment method.

With real-time inventory information, the sequence of consumer arrivals matter: consumers who arrive early will see that the store is in stock, whereas consumers who arrive late will encounter stockouts. In the former case, the consumer can go to the store and receive an expected payoff of $u_{s,\,\mathrm{in}} = -h_s + \theta(v - p)$ because she will certainly obtain the product if she realizes high valuation. In the latter case, the consumer will receive nothing from visiting the store and thus will choose to buy online instead. In this model, the format of the shared real-time information does not matter; it could be either a binary indication of inventory availability (as in CVS and Walgreens's case) or the exact number of units in stock (as in IKEA's case). This does not change consumers' shopping behavior in our model, since consumers demand one unit of product and thus care only about whether the store is in stock or not when they arrive in the market.

Each consumer chooses between shopping online directly or going to the store, given the current store inventory availability status. Let ϕ_{in} denote the fraction of customers visiting the store when it is in stock. As before, only high-type consumers will absorb store inventory when they come to the store. Then, the expected number of customers who see that the store is in stock is $D_{\mathrm{in}}(q) = \min(D, q/(\theta\phi_{\mathrm{in}}))$, and the remaining $D_{\mathrm{out}}(q) = (D - q/(\theta\phi_{\mathrm{in}}))^+$ will find that the store is already out of stock when they check availability online. Note that the expressions for $D_{\mathrm{in}}(q)$ and $D_{\mathrm{out}}(q)$ are slightly different from before because all consumers, whether they choose to come to the store or not, will receive the real-time inventory information. Thus, the retailer's profit function is as follows:

$$\pi(q) = p\theta\phi_{\text{in}} ED_{\text{in}}(q) - cq \tag{6.4}$$
$$+ [w\theta - k(1-\theta)](1-\phi_{\text{in}}) ED_{\text{in}}(q) \tag{6.5}$$
$$+ [w\theta - k(1-\theta)] ED_{\text{out}}(q). \tag{6.6}$$

The first two parts of the profit function correspond to the case when the store is in stock: (6.4) is the newsvendor profit from the store, and (6.5) is the profit from those who buy online directly. The last part (6.6) corresponds to the case when the store is out of stock and all consumers buy online.

For the case with real-time inventory information, we use superscript $(\cdot^i)$ to denote the market outcome, which is given in the following proposition.

Proposition 6.8 *With real-time inventory information, there exists a threshold* $\psi^i \in [\psi^\circ, 1]$ *such that the market outcome is given as follows:*

- *If* $\theta < \psi^i$, *then consumers visit store if store is in stock, and buy online directly if store is out of stock:*
$$q^i = \theta\bar{F}^{-1}\left(\frac{c}{p - w + k(1-\theta)/\theta}\right) > 0;$$
- *If* $\theta \geq \psi^i$, *then consumers always buy online directly;* $q^i = 0$.

Moreover, $\pi^i \geq \pi^\circ$.

Comparing Proposition 6.8 (after the provision of real-time inventory information) and Proposition 6.7 (before the provision of real-time inventory information), since $\psi^i \geq \psi^\circ$, we see that providing real-time inventory information may attract consumers to the store (see Fig. 6.1). In the base scenario, consumers bear in-store stockout risk because they incur the hassle of going to the store before finding out whether the store has the product in stock. Now, providing real-time inventory information eliminates such risk: if the store is in stock, consumers are guaranteed availability before incurring any sunk cost.

Similar as in the store-only scenario (see Proposition 6.4), Proposition 6.8 also shows that offering real-time inventory information has no negative effect on an

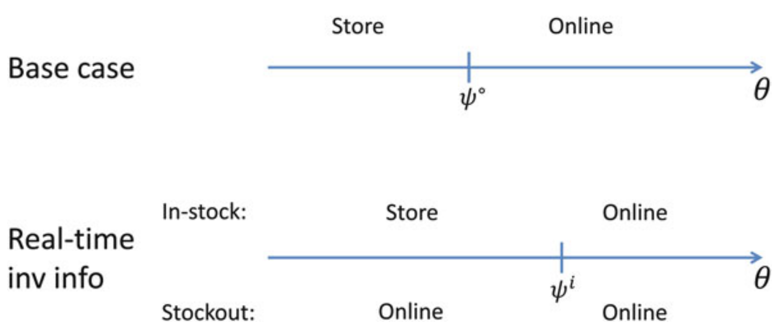

Fig. 6.1 Comparison of consumer behavior with and without real-time inventory information

omnichannel retailer either (i.e., $\pi^i \geq \pi^\circ$). By assuring consumers about inventory availability in store, the retailer can attract more people to the store; consumers can then physically inspect the product and realize their valuation before making any purchase. This helps to reduce potential product returns and increase total profit.

6.4.2 Mitigating Product Misfit Online: Virtual Showrooms

Suppose the retailer implements virtual showrooms in the online channel. To model the impact of virtual showroom on consumer decision making, we adopt the approach described at the end of Sect. 6.3. Specifically, after checking with the virtual showroom, α of the low-type customers discover that the product is not for them and leave, while the rest (including all the high-types and $1-\alpha$ of the low-types) remain interested in the product and update their belief about the probability of being high type to $\theta' = \theta/[1-\alpha(1-\theta)] > \theta$ based on the Bayes' Rule. As a result, the total demand size is $D' = [1-\alpha(1-\theta)]D$ (with cdf $F'(x) = F(x/[1-\alpha(1-\theta)])$ for any x) and a fraction θ' of them is of high type. Thus, similar to Proposition 6.7, the equilibrium outcome is given in the following proposition; we use the superscript $(\cdot^v)$ to denote the equilibrium outcome for the scenario with virtual showrooms. Henceforth, without further specification, we simply refer to this new pool of consumers as the retailer's consumers.

Proposition 6.9 *With virtual showrooms, there exists a threshold $\psi^v \in [0, \psi^\circ]$ such that*

- *if $\theta < \psi^v$, then consumers visit the store and*

$$q^v = \theta' \bar{F'}^{-1}\left(\frac{c}{p-w+k(1-\theta')/\theta'}\right) > 0;$$

- *if $\theta \geq \psi^v$, then consumers buy online directly and $q^v = 0$.*

Comparing Proposition 6.9 (after the provision of virtual showrooms) and Proposition 6.7 (before the provision of virtual showrooms), since $\psi^v \leq \psi^\circ$, virtual showrooms may attract customers to buy online instead of in the store (see Fig. 6.2). This is not surprising. Virtual showrooms give consumers a similar hands-on experience as in the store. With decreased product value uncertainty, shopping online becomes more productive and appealing to consumers.

Proposition 6.10 *Compared to the base model,*

- *if $\theta < \psi^v$ or $\theta \geq \psi^\circ$, then providing virtual showrooms increases total profit, i.e., $\pi^v > \pi^\circ$;*
- *if $\theta \in [\psi^v, \psi^\circ)$, there exists $\bar{w}$ such that providing virtual showrooms increases total profit (i.e., $\pi^v > \pi^\circ$) if $w > \bar{w}$; but reduces total profit (i.e., $\pi^v < \pi^\circ$) if $w < \bar{w}$.*

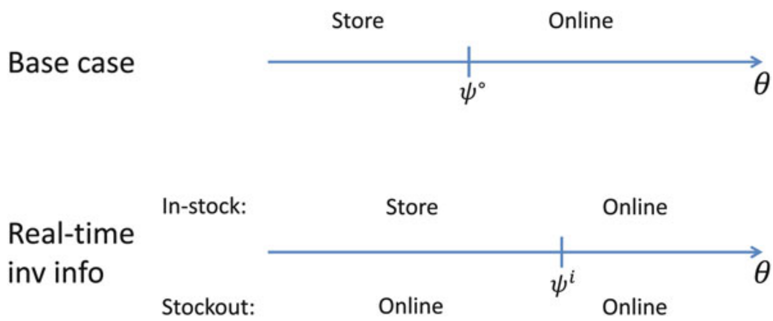

Fig. 6.2 Comparison of consumer behavior with and without virtual showroom

There are three cases discussed in Proposition 6.10. First, if consumers have a small high-type probability (i.e., $\theta < \psi^v$), then they always turn to the physical store for validation before making any purchasing decision. When the store is out of stock, consumers may choose to buy online directly before resolving product value uncertainty. In this case, virtual showrooms help by screening out some low-type consumers beforehand so that the potential number of returns will be smaller.

Second, if consumers have a large high-type probability (i.e., $\theta \geq \psi^o$), then they are comfortable buying online. In this case, virtual showrooms serve as the main source of product information for consumers. By screening out potential low-type customers before they make any purchase on the website, virtual showrooms help to avoid returns and increase profits.

However, when $\theta \in [\psi^v, \psi^o)$, implementing virtual showrooms may backfire. In this case, consumers originally visit the store in the base model, but virtual showrooms attract them to buy online instead. Although the online return rate decreases from $1-\theta$ (without virtual showroom) to $1-\theta'$ (with virtual showroom), total returns may increase. This is because more people now choose to buy online, including low types who are destined to return their purchases. The resulting increase in returns will drive down total profit. Unless the online profit margin is high enough, customer migration from store to online will be unprofitable for the retailer. This result offers a possible explanation for fashion retailer H&M's decision to remove their virtual showroom (called the Dressing Room) from their website even though it has been popular among consumers (H&M 2010). This result also suggests that retailers should be cautious when looking at online return rates and should not neglect the total number of returns as an informative companion metric.

Proposition 6.11 *There exists threshold* $\alpha_1, \alpha_2 \in [0, 1]$ *such that* $\alpha_1 \leq \alpha_2$ *and* $\pi^v < \pi^o$ *if and only if* $\alpha \in (\alpha_1, \alpha_2)$.

Proposition 6.11 demonstrates the importance of the α parameter, i.e., the informativeness of the virtual showroom. Some implementations of virtual showrooms involve relatively rudimentary functions (e.g., picture upload), while others use more advanced virtual reality technologies (e.g., 3D technology) which offer online shoppers a more vivid try-on experience. As virtual showrooms become more informative (i.e., as α increases), Proposition 6.11 shows that it is possible for them to become less attractive (i.e., $\pi^v < \pi^\circ$) if $\alpha \in (\alpha_1, \alpha_2)$. Specifically, virtual showrooms help improve retailer's profit only when α is either sufficiently large ($\alpha \geq \alpha_2$) or sufficiently small ($\alpha \leq \alpha_1$). In other words, attempts to enhance the informativeness of virtual showrooms, if not significant, may reduce profits. The reason is as follows: If α is small, the virtual showroom rarely changes customer's channel choice decisions; specifically, customers may continue to visit the store in which case the virtual showroom acts as a backup source of product information to customers once store is out of stock. As α increases, virtual showrooms tend to attract customers to purchase online directly due to the lower valuation uncertainty risk. However, unless the virtual showroom becomes very effective in screening out low-type customers, it may simply end up attracting online transactions and thus increasing online returns.

6.5 Conclusion

Omnichannel retailing is rapidly becoming the norm in industry. To study the impacts of different types of omnichannel business strategies, from the modeling perspective, it is important to understand how consumers make purchasing decisions among different channel environments which may involve different types of shopping risks. For instance, it may be hard for customers to verify product availability before incurring the hassle of visiting the physical location, while purchases in the digital world may be a bad fit, as customers are unable to fully evaluate the product due to the lack of touch-and-feel experiences. The risks of stockouts and product misfit are major concerns for shoppers and retailers. In this chapter, we present different approaches to model strategic consumer shopping behavior in the presence of both stockouts and product misfit problems. We first reviewed traditional models which focus either on stockouts or product misfit. Then, building upon these models, we introduce a modeling framework to study an omnichannel environment, where customers and the retailer interact through both online and offline channels. Here, both stockouts (in-store) and product misfit (online) are particularly relevant. Omnichannel retail management has received a lot of attention in both industry and academia, we hope the models reviewed in this chapter could contribute to this exciting line of research.

Appendix: Proofs

Proof (Proposition 6.1) This is Proposition 1 in Su and Zhang (2009).

Proof (Proposition 6.2) This is Proposition 2 in Su and Zhang (2009).

Proof (Proposition 6.3) As discussed in the chapter, the optimal price must be $p^i = v - h$. Then, the optimal quantity that maximizes the newsvendor profit $\pi = (v-h)E\min(D,q) - cq$ is simply given by $\bar{F}(q^i) = c/(v-h)$.

Proof (Proposition 6.4) According to Propositions 1–3 in Su and Zhang (2009), we have $\pi^* = vE\min(D,q^*) - cq^* - hED$, $\pi^c = vE\min(D,q^c) - cq^c - hED$ and $\pi^c > \pi^*$. Thus, to prove Proposition 6.4, we simply need to show $\pi^i > \pi^c$. Note $\pi^i = (v-h)\,E\min\left(D,q^i\right) - cq^i > (v-h)\,E\min\left(D,q^c\right) - cq^c > p^c E\min\left(D,q^c\right) - cq^c = \pi^c$, where the first inequality is because q^i is the unique maximizer of $\pi(q) = (v-h)\,E\min\left(D,q\right) - cq$, and the second inequality is due to the fact that $p^c = v - h/A(q^c) < v - h$. This completes the proof.

Proof (Proposition 6.5) This is a special case of Proposition 2 in Su (2009).

Proof (Proposition 6.6) With virtual showrooms, the retailer's profit function is

$$\begin{aligned}\pi(q) &= \theta' vE\min(D',q) + s(q - E\min(D',q)) - cq \\ &= (\theta v - s(1-\alpha(1-\theta)))E\min(D,\tilde{q}) - (c-s)(1-\alpha(1-\theta))\tilde{q} \triangleq \tilde{\pi}(\tilde{q}),\end{aligned}$$

where $\tilde{q} = q/(1-\alpha(1-\theta))$. Denote $\tilde{q}^v = \arg\max_{\tilde{q}}\tilde{\pi}(\tilde{q})$, and thus the optimal profit is $\pi^v = \tilde{\pi}(\tilde{q}^v)$. By Envelope Theorem,

$$\frac{\partial \pi^v}{\partial \alpha} = s(1-\theta)E\min(D,\tilde{q}^v) + (c-s)(1-\theta)\tilde{q}^v > 0.$$

Denote q^* and π^* as the optimal inventory level and profit with partial returns.

If $\alpha = 0$, then

$$\pi^v = (\theta v - s)E\min(D,\tilde{q}^v) - (c-s)\tilde{q}^v < (\theta v - \theta s)E\min(D,\tilde{q}^v) - (c-s)\tilde{q}^v < \pi^*;$$

if $\alpha = 1$, then

$$\begin{aligned}\pi^v &= (\theta v - \theta s)E\min(D,\tilde{q}^v) - \theta(c-s)\tilde{q}^v > (\theta v - \theta s)E\min(D,q^*) - \theta(c-s)q^* \\ &> (\theta v - \theta s)E\min(D,q^*) - (c-s)q^* = \pi^*.\end{aligned}$$

Thus, we can conclude the result.

Proof (Proposition 6.7) This is Proposition 1 in Gao and Su (2017b).

Proof (Proposition 6.8) The equilibrium outcome is Proposition 6 in Gao and Su (2017b). The result that $\pi^i \geq \pi^\circ$ can be derived from Proposition 7 in Gao and Su (2017b).

Proof (Proposition 6.9) This is Proposition 4 in Gao and Su (2017b).

Proof (Proposition 6.10) This is Proposition 5 in Gao and Su (2017b).

Proof (Proposition 6.11) By Proposition 6.9, $\psi^v \leq \psi^o$ for any $\alpha \in [0, 1]$. There are two cases:

- If $\theta \geq \psi^o$: Then by Proposition 6.10, $\pi^v > \pi^o$ for any α. Thus, $\alpha_1 = \alpha_2 = 0$.
- If $\theta < \psi^o$: Note it is easy to check that ψ^v is decreasing in α. Define $\alpha_1 = \arg\min_{\alpha \in [0,1]} |\theta - \psi^v(\alpha)|$. Then, if $\alpha \in [0, \alpha_1)$, then $\psi^v(\alpha) > \theta$ and thus $\pi^v > \pi^o$ by Proposition 6.10. If $\alpha \geq \alpha_1$, then $\psi^v \leq \theta$. Then, we have

$$\pi^\circ = (p - w + k(1-\theta)/\theta)E\min(\theta D, q^\circ) - cq^\circ + (w\theta - k(1-\theta))ED,$$

and

$$\pi^v = wE\theta' D' - kE(1-\theta')D' = wE\theta D - kE(1-\alpha)(1-\theta)D.$$

Define function

$$g(\alpha) = \pi^\circ - \pi^v = (p - w + k(1-\theta)/\theta)E\min(\theta D, q^\circ) - cq^\circ - k\alpha(1-\theta)ED.$$

Note that $\partial g/\partial\alpha = -k(1-\theta)ED < 0$, which implies that there exists $\alpha_2 \geq \alpha_1$ such that $\pi^\circ > \pi^v \Leftrightarrow \alpha < \alpha_2$. Finally, let us show that it is possible to have $\alpha_2 > \alpha_1$. Suppose $\alpha_1 \in (0, 1)$. Thus, $\psi^v(\alpha_1) = \theta$. Since $g(0) > 0$ and $\partial g/\partial\alpha < 0$, there exists $\bar{\alpha} > 0$ such that $g(\bar{\alpha}) = 0$. Thus, if $\theta \in (\psi^v(\bar{\alpha}), \psi^o)$, then we must have $\bar{\alpha} > \alpha_1$, because $\psi^v(\alpha_1) = \theta$ and $\psi^v(\alpha)$ is decreasing in α. Then, this $\bar{\alpha}$ is the α_2 we are looking for.

References

CNET (2010) Shoefitr uses 3D to help buy the right kicks. https://www.cnet.com/news/shoefitr-uses-3d-to-help-buy-the-right-kicks/. Accessed 7 Jan 2019

Dana JD (2001) Competition in price and availability when availability is unobservable. Rand J Econ 32(3):497–513

Davis S, Gerstner E, Hagerty M (1995) Money back guarantees in retailing: matching products to consumer tastes. J Retail 71(1):7–22

Deneckere R, Peck J (1995) Competition over price and service rate when demand is stochastic: a strategic analysis. RAND J Econ 26(1):148–162

Financial Times (2011) Next big trend: virtual fitting rooms. https://www.ft.com/content/57b1fea6-1f55-11e0-8c1c-00144feab49a. Accessed 7 Jan 2019

Gallino S, Moreno A (2014) Integration of online and offline channels in retail: the impact of sharing reliable inventory availability information. Manag Sci 60(6):1434–1451

Gao F, Su X (2017a) Omnichannel retail operations with buy-online-and-pick-up-in-store. Manag Sci 63(8):2478–2492

Gao F, Su X (2017b) Online and offline information for omnichannel retailing. Manuf Serv Oper Manag 19(1):84–98

H&M (2010) H&M annual report 2010. https://about.hm.com/content/dam/hmgroup/groupsite/documents/en/Annual%20Report/Annual_Report_2010_p1_en.pdf

Rigby D (2011) The future of shopping. Harv Bus Rev 89(12):65–76

Su X (2009) Consumer returns policies and supply chain performance. Manuf Serv Oper Manag 11(4):595–612
Su X, Zhang F (2009) On the value of commitment and availability guarantees when selling to strategic consumers. Manag Sci 55(5):713–726
Xie J, Shugan SM (2001) Electronic tickets, smart cards, and online prepayments: when and how to advance sell. Mark Sci 20(3):219–243

Chapter 7
On-Demand Customization and Channel Strategies

Li Chen, Yao Cui, and Hau L. Lee

Abstract In this chapter, we study the impact of two major technological advances in demand fulfillment—the emergence of dual channels in retail (i.e., online and in-store) and the adoption of on-demand customization technology (such as additive manufacturing or 3D printing). Our analysis shows that such technology adoption can have differential impacts to the two channels. The technology leads to increased product variety offered online, as well as allows the firm to charge a price premium for online customers. Yet it also induces the firm to offer a smaller product variety and a reduced price in-store. Moreover, the online demand increases (decreases) if the customer online wait cost is low (high). The firm's profitability with the new technology is driven by the production setup cost of the traditional production technology it replaces and also by how much customers care about the product's custom fit.

Keywords On-demand customization · Dual channels · Product variety · Pricing · Supply chain management

7.1 Introduction

With the advancement of technology, some innovative consumer goods companies such as Adidas and Nike have started to experiment with 3D printing, a form of on-demand customization technology, for real production (Economist 2017). Walmart has been reported to test the 3D printing technology to provide more options for consumers: purchase online for manufacture in-store, or manufacture online and

L. Chen (✉) · Y. Cui
Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY, USA
e-mail: li.chen@cornell.edu; yao.cui@cornell.edu

H. L. Lee
Graduate School of Business, Stanford University, Stanford, CA, USA
e-mail: haulee@stanford.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_7

ship to consumers (Massie 2013). Besides the 3D printing technology, Amazon has recently patented an on-demand apparel manufacturing system, which uses proprietary algorithms to print custom patterns on fabrics, and then cuts and sews them to fill online orders for suits, dresses, and other garments (Wingfield and Couturier 2017). On-demand customization has also been implemented by on-demand book printing platforms (e.g., Espresso Book Machine, Amazon's CreateSpace, BookLocker). The benefits of such on-demand customization are many—companies can offer much more product variety than before, improve customer satisfaction further with perfect fit products, and shorten response times significantly for tailor-made products. This chapter aims at analytically understanding some of these benefits when companies adopt the on-demand customization technology to meet online demand for consumer goods.

Online sales have been increasing worldwide at unprecedented rates (e.g., it was 19.4% at Amazon, 40% at Apple, 9% at Walmart, and 12.5% at Macy's; see Zaczkiewicz 2017). Among the top 10 US retailers with the highest e-commerce sales in 2017, eight of them also sell through their own stores. Dual channels (with sales via both brick-and-mortar stores and online) have become a mainstream means for companies to interact with consumers, which form the basis of our analysis of the impact of the on-demand customization technology. The emergence of such technology will likely be first introduced to factories (such as in the Adidas example cited by Economist 2017 and outlined in the Amazon's patent). When this happens, the factory would serve two kinds of demands: online consumer demand and store replenishment. In the latter, since store replenishment is still in bulk, the factory can continue to use mass production systems for cost efficiency.

The impact of on-demand customization in the presence of dual channels can be analyzed in multiple dimensions. First, from the consumer's point of view, the consumer's choice preferences could be affected by the number of product options offered, ranging from a finite set (under traditional technology) to possibly infinite (under the on-demand customization technology). The consumer's utility, measured by how close the chosen option fits the exact needs of the consumer, the waiting time for the product, and the prices that the consumer has to pay for the product, can be affected as a result of on-demand customization. Second, the merchant's profits could be affected, through its channel strategies of product offerings and the associated prices charged. Third, how would on-demand customization affect the split of sales between the online and in-store channels?

We develop a stylized model to address the above questions. In our model, consumers as customers are heterogeneous in two dimensions: (1) the waiting cost they incur from purchasing online and (2) the fit cost they incur because the product types offered by the merchant may not meet their needs exactly. To obtain some baseline insights for the problem, we assume the total customer demand is deterministic and focus on comparing the firm's optimal channel strategies for product offering and pricing with and without adopting the on-demand customization technology online.

In our model, we assume that the customization technology is deployed at the factory level, so that it is used for online orders only. Currently, this technology is still very expensive, and substantial training investments have to be made, so that it is not yet feasible for it to be implemented at all stores, except for very simple products.

A few key insights from our study are highlighted as follows. First, adopting the on-demand customization technology for online customers gives rise to a *substitution effect* of technological innovation, i.e., the traditional technology is replaced by a better on-demand customization technology (Lee 2007). Such technology substitution leads to the *variety effect*, enabled by the elimination of the production setup cost, and allows the firm to offer perfect customization and charge a price premium for online customers. Second, compared to the base case with traditional technology, adopting on-demand customization to meet online demand will reduce the variety of in-store product offering as well as the associated price. Moreover, the demand segmentation between the online and in-store channels depends on the customers' online purchase waiting cost. If the waiting cost is small, more demand will be steered to the online channel; and if the waiting cost is large, the opposite occurs. Finally, we show that the firm's profitability with this new technology is driven by the production setup cost of the traditional technology it replaces and also by how much customers care about the product custom fit.

The remainder of this chapter is organized as follows. After a review of relevant literature in Sect. 7.2, we describe the model setup and present the main results in Sect. 7.3. We conclude the chapter in Sect. 7.4 with a discussion of future research directions.

7.2 Literature Review

Our study is related to and contributes to the mass customization literature. See Lancaster (1990), Ho and Tang (1998), and Ramdas (2003) for reviews of product variety literature. Lee (1996) studies the product/process postponement design in both built-to-order (BTO) and built-to-stock (BTS) production modes. Lee and Tang (1997) study the optimal point of product differentiation (i.e., the stage after which the products assume their unique identities). Dobson and Yano (2002) uses integer programming to analyze the firm's optimal decisions of which products to offer, how to price them, and whether each should be make-to-stock or make-to-order. Jiang et al. (2006) analyze a mass customization system consisting of an initial BTS stage and a final BTO stage. Alptekinoğlu and Corbett (2010) study the trade-off between the increased ability to precisely meet customer preferences and the increased lead time from order placement to delivery associated with customized products. Desai et al. (2001) and Kim et al. (2013) study the marketing-manufacturing trace-off under component commonality strategies, i.e., while commonality can lower manufacturing cost, such a design may hinder product differentiation. Liu and Cui (2010) show that a manufacturer may choose to extend a horizontally differentiated product line in a decentralized channel but not in a centralized channel. Çil and Pangburn (2017) studies the impact of the market repercussions resulting from increased brand dilution on the firm's mass customization strategy. There are also a series of papers (e.g., Dewan et al. 2003, Syam et al. 2005, Syam and Kumar 2006, Alptekinoğlu and Corbett (2008), Mendelson and Parlaktürk 2008a,b, Xia and Rajagopalan 2009) that study the competition between a firm that uses mass customization and a firm that uses traditional production, or between firms who compete on adoption of mass customization. Different from the existing literature on mass customization, our study considers a dual-channel setting and studies the impact of adopting mass customization in the online channel on the firm's product offering and pricing strategies in both channels.

As an important technology to implement mass customization, 3D printing has started to attract the attention from operations management researchers. There is an emerging stream of research that studies the impact of 3D printing technology on manufacturing and supply chain management. For example, Song and Zhang (2016) develop a queueing model to analyze and quantify the impact of 3D printing on spare parts logistics. Dong et al. (2017) study the impact of 3D printing on a firm's manufacturing strategy and product assortment decision, and show that the implications from adopting the 3D printing technology is significantly different from adopting the traditional flexible technology. Chen et al. (2018) consider two adoption cases of 3D printing in a dual-channel retail setting (adopting in the online channel only, and adopting in both online and in-store channels), and study the firm's integrated decision making regarding product offering, prices for the two channels, as well as inventory decisions.

7.3 Model and Results

We consider a firm that produces and sells products through two channels, a brick-and-mortar store channel and an online channel. Hereafter, without specific mention, "in the store" or "in-store" means in the brick-and-mortar store. We assume that in-store demand and online demand are endogenously determined by customers' channel preferences and product preferences, which are specified as follows. Customers are heterogeneous in two dimensions: (1) the waiting cost they incur from purchasing online, and (2) the fit cost they incur because the product types offered by the firm do not meet their needs exactly.

Specifically, to model customers' heterogeneous channel preferences, we assume that there are three types of customers described as follows.

Type I customers (online-acceptable without waiting cost) For these customers, waiting cost is negligible when purchasing online. For example, some people are prone to shopping online because they want to avoid traveling to the brick-and-mortar store or fear stockout at the store (Gao and Su 2017a,b). Moreover, some people can avoid the shipping fee because they are loyal customers to the firm (e.g., Amazon Prime). From a survey study, Konuş et al. (2008) find that 37% of respondents tend to use the Internet and catalogs for both information search and purchase. In our model, Type I customers correspond to α proportion of the population ($0 < \alpha < 1$).

Type II customers (online-acceptable with waiting cost) These customers incur a finite waiting cost $e > 0$ when purchasing online. For example, by purchasing online, customers forgo the joy of receiving the product immediately, which might create a disutility for some people. The existence of customer disutility from purchasing online has been empirically established by studies such as Bart et al. (2005) and Forman et al. (2009). Type II customers correspond to β proportion of the population ($0 < \beta < 1$ and $\alpha + \beta < 1$).

Type III customers (in-store only) These customers do not consider the online option (i.e., their waiting cost is infinity), and may only purchase from the brick-and-mortar store. For example, there are people who have not accepted the concept of online shopping. In the study of Konuş et al. (2008), 23% of respondents are "store-focused" who reveal favorable attitudes toward brick-and-mortar stores and relatively unfavorable attitudes toward online channels. In our model, Type III customers correspond to the remaining $1 - \alpha - \beta$ proportion of the population.

Notice that for analytical simplicity, we have assumed that the customer waiting costs follow discrete types. The model can be generalized to allow for continuous customer waiting costs without changing the main insights.

Besides customers' heterogeneous channel preferences, we model customers' heterogeneous product preferences using the circular city framework (Salop 1979), which is a variant of the classic Hotelling (1929) model. We assume that customers are located on a circle of unit circumference. Customers are uniformly distributed on the circle. Each customer's location represents her ideal product type (e.g., her size or favorite color of a product), and the arc distance between a product location and the customer location measures the customer's misfit from this product. Each customer may only purchase the product type that is closest to her location on the circle. Given an in-store price p, the customer's utility from purchasing in-store a product that is x arc distance away is $v - p - tx$, where v is the valuation of customers for the ideal product type, t is the fit cost parameter (which corresponds to the transportation cost parameter in the classic Hotelling model) and measures customers' sensitivity to product differences, and tx is the fit cost of customer x, with $t > 0$. Similarly, given an online price p, the customer's utility is $v - p - tx - e$ from purchasing online (note that $e = 0$ for Type I customers and $e = \infty$ for Type III customers). See Fig. 7.1 for an illustration.

To keep the model tractable, we assume that the total customer demand μ from the unit circle is deterministic. This is a fairly common assumption used in the product line design and product differentiation literature. To meet the demand from the two channels, the firm may use the traditional technology or adopt the on-demand customization technology in the online channel. In the former case, the firm uses the traditional technology to produce products sold in both channels. In the latter case, the firm uses on-demand customization to produce products sold online and uses the traditional technology to produce products sold in-store. This would involve adopting the new technology (e.g., installing 3D printers) in the factory. The firm uses both types of technology in its production in the factory, and can

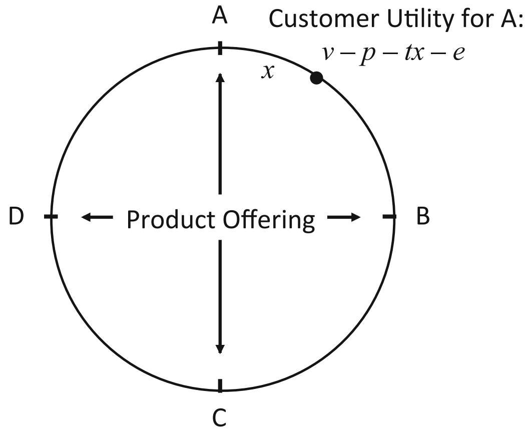

Fig. 7.1 Illustration of the circular city customer utility model

effectively offer "infinite" types of product online to cover the entire customer circle (where customers do not incur any fit costs). We assume that producing one unit of product requires one unit of common raw material under both traditional and on-demand customization technologies, and that the product quality is the same under both technologies.

The firm incurs marginal cost c for each unit of product, regardless of the product type. We further assume that the marginal cost remains the same when the on-demand customization technology is adopted online. When using the traditional technology, the firm incurs a production setup cost s for each product type it offers, due to factors such as switchover and retooling. Thus, given n product types, the total setup cost is sn. Additionally, in order to adopt the on-demand customization technology, the firm incurs a fixed cost k from purchasing new equipment and training employees.

The firm chooses the number of horizontally differentiated products to offer, as well as the prices for products sold in-store and online, to maximize its total profit. In what follows, we analyze the firm's optimal product offering and pricing strategies in each case. Then, by comparing the optimal strategies across the two cases, we obtain how the adoption of the on-demand customization technology in the online channel affects the firm's product types and prices in each channel and develop insights regarding how online on-demand customization creates value to the firm. To avoid trivial scenarios, we make the assumption of $(v - c)\mu \geq \sqrt{2st\mu}$, which indicates that the firm earns a non-negative profit under the traditional technology.

7.3.1 Base Case with Traditional Technology

In this section, we analyze the firm's optimal strategy under the traditional technology. In this case, the firm produces n types of horizontally differentiated products, and chooses price p_o for all products sold online and p_i for all products sold in-store (subscript "o" represents online and subscript "i" represents in-store). Because product types are horizontally differentiated (i.e., differentiated in a dimension other than quality), the firm charges the same price for all product types within each channel. For example, apparel producers usually sell the same style in different sizes and colors and charge the same price for all sizes and colors.

To derive the firm's optimal strategy in this case, we need to first characterize the customer choices. Consider the arc on the customer circle that is centered at the location of any product type and has arc length $\frac{1}{n}$. This arc corresponds to the demand base for this product type. Moreover, the customers' utilities are symmetric on two sides of the product location. Thus, to analyze the customer choices, we focus on the arc on one side of the product location, where the customer's distance from her ideal product type, x, ranges in $0 \le x \le \frac{1}{2n}$. We derive the purchasing decisions of each type of customers as follows.

- Type I customers: Their utility from purchasing online is $v - p_o - tx$, and their utility from purchasing in-store is $v - p_i - tx$. Then, Type I customers purchase online if $v - p_o - tx \ge v - p_i - tx$ and $v - p_o - tx \ge 0$, purchase in-store if $v - p_i - tx > v - p_o - tx$ and $v - p_i - tx \ge 0$, and do not purchase otherwise. Thus, Type I customers purchase online if $p_i - p_o \ge 0$ and $0 \le x \le \min(\frac{v-p_o}{t}, \frac{1}{2n})$, and purchase in-store if $p_i - p_o < 0$ and $0 \le x \le \min(\frac{v-p_i}{t}, \frac{1}{2n})$.
- Type II customers: Their utility from purchasing online is $v - p_o - tx - e$, and their utility from purchasing in-store is $v - p_i - tx$. Then, Type II customers purchase online if $v - p_o - tx - e \ge v - p_i - tx$ and $v - p_o - tx - e \ge 0$, purchase in-store if $v - p_i - tx > v - p_o - tx - e$ and $v - p_i - tx \ge 0$, and do not purchase otherwise. Thus, Type II customers purchase online if $p_i - p_o \ge e$ and $0 \le x \le \min(\frac{v-p_o-e}{t}, \frac{1}{2n})$, and purchase in-store if $p_i - p_o < e$ and $0 \le x \le \min(\frac{v-p_i}{t}, \frac{1}{2n})$.
- Type III customers: Their utility from purchasing in-store is $v - p_i - tx$, so Type III customers purchase in-store if $v - p_i - tx \ge 0$, and do not purchase otherwise. Thus, Type III customers purchase in-store if $0 \le x \le \min(\frac{v-p_i}{t}, \frac{1}{2n})$.

Based on the customer choices, we derive the firm's profit function as the following:

$$
\begin{aligned}
\Pi(n, p_o, p_i) &= (p_o - c)\left[\alpha\mu \min\left(\tfrac{v-p_o}{t}, \tfrac{1}{2n}\right) + \beta\mu \min\left(\tfrac{v-p_o-e}{t}, \tfrac{1}{2n}\right)\right] 2n \\
&\quad + (p_i - c)(1 - \alpha - \beta)\mu \min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right) 2n - sn, \qquad \text{if } p_i - p_o \ge e; \\
&= (p_o - c)\alpha\mu \min\left(\tfrac{v-p_o}{t}, \tfrac{1}{2n}\right) 2n \\
&\quad + (p_i - c)(1 - \alpha)\mu \min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right) 2n - sn, \qquad \text{if } 0 \le p_i - p_o < e; \\
&= (p_i - c)\mu \min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right) 2n - sn, \qquad \text{if } p_i - p_o < 0.
\end{aligned}
$$

Depending on the relationship between p_o and p_i, $\Pi(n, p_o, p_i)$ takes different forms and as is shown above, there are three possible scenarios. Without loss of generality, we restrict the prices to $0 \leq p_o, p_i \leq v$ in all our analyses, because a price higher than v does not yield any sales. Let the triple (n^*, p_o^*, p_i^*) be the optimal solution that maximizes the profit function $\Pi(n, p_o, p_i)$, and Π^* be the optimal profit. The following proposition summarizes the firm's optimal strategy in this case:

Proposition 7.1 *Under the traditional technology, the firm offers* $n^* = \sqrt{\mu t/(2s)}$ *types of product at optimal prices* $p_o^* = p_i^* = v - \sqrt{st/(2\mu)}$. *Under the optimal strategy, all Type I customers purchase online, all Type II and Type III customers purchase in-store. Thus, the firm's online demand is* $\alpha\mu$ *and the in-store demand is* $(1-\alpha)\mu$. *The firm's optimal profit is* $\Pi^* = (v-c)\mu - \sqrt{2st\mu}$.

Proposition 7.1 characterizes the firm's optimal strategy under traditional technology. Consistent with previous literature (e.g., Salop 1979, Riordan 1986, De Groote 1994), we ignore the integer constraint for n. The optimal number of product types is $n^* = \sqrt{\mu t/(2s)}$, and the optimal price is $v - \sqrt{st/(2\mu)}$ for both channels. Under the traditional technology, given the symmetric level of customization in both channels, the firm charges the same price in both channels. The optimal price is $\sqrt{st/(2\mu)}$ below the customer valuation v, which is due to the misfit of product types. Note that $n^* = \frac{t}{2(v-p^*)}$, which indicates that a higher price should be associated with a broader product offering so that the firm does not screen too many customers with the higher price. Moreover, the firm's optimal product offering and pricing strategy induces Type I customers to purchase online, and Type II and Type III customers to purchase in-store.

7.3.2 *Adopting On-Demand Customization Technology Online*

In this section, we analyze the firm's optimal strategy when adopting the on-demand customization technology online. In this case, the firm uses the traditional technology to produce n types of products sold in-store. For online demand, because of on-demand customization, the firm can customize products according to each customer's need and customers do not incur any misfit. Thus, the firm effectively offers infinite product types online.

Same as in the case with traditional technology, in order to characterize the customer choices, we consider the arc on one side of any product location with customer distance ranging in $0 \leq x \leq \frac{1}{2n}$. The analysis is as follows.

- Type I customers: Their utility from purchasing online is $v - p_o$, and their utility from purchasing in-store is $v - p_i - tx$. Then, Type I customers purchase online if $v - p_o \geq v - p_i - tx$ and $v - p_o \geq 0$, purchase in-store if $v - p_i - tx > v - p_o$ and $v - p_i - tx \geq 0$, and do not purchase otherwise. Thus, Type I customers purchase online if $\min(\frac{p_o-p_i}{t}, \frac{1}{2n}) \le x \le \frac{1}{2n}$, and purchase in-store if $0 \le x < \min(\frac{p_o-p_i}{t}, \frac{1}{2n})$. Note that if $p_i - p_o > 0$, all Type I customers purchase online.

- Type II customers: Their utility from purchasing online is $v - p_o - e$, and their utility from purchasing in-store is $v - p_i - tx$. Then, Type II customers purchase online if $v - p_o - e \ge v - p_i - tx$ and $v - p_o - e \ge 0$, purchase in-store if $v - p_i - tx > v - p_o - e$ and $v - p_i - tx \ge 0$, and do not purchase otherwise. Thus, Type II customers purchase online if $p_o \le v - e$ and $\min(\frac{p_o-p_i+e}{t}, \frac{1}{2n}) \le x \le \frac{1}{2n}$, and purchase in-store if $0 \le x < \min(\frac{p_o-p_i+e}{t}, \frac{v-p_i}{t}, \frac{1}{2n})$.
- Type III customers: Same as in Case 1, Type III customers purchase in-store if $0 \le x \le \min(\frac{v-p_i}{t}, \frac{1}{2n})$.

Based on the customer choices, we derive the firm's profit function as the following:

$$
\begin{aligned}
&\Pi(n, p_o, p_i)\\
&= (p_o - c)\alpha\mu\left(\tfrac{1}{2n} - \tfrac{p_o-p_i}{t}\right)^+ 2n\\
&\quad + (p_i - c)\left[\alpha\mu\min\left(\tfrac{p_o-p_i}{t}, \tfrac{1}{2n}\right) + (1-\alpha)\mu\min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right)\right]2n - sn - k,\\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{if } p_i - p_o \le 0 \text{ and } p_o > v - e;\\
&= (p_o - c)\left[\alpha\mu\left(\tfrac{1}{2n} - \tfrac{p_o-p_i}{t}\right)^+ + \beta\mu\left(\tfrac{1}{2n} - \tfrac{p_o-p_i+e}{t}\right)^+\right]2n\\
&\quad + (p_i - c)\left[\alpha\mu\min\left(\tfrac{p_o-p_i}{t}, \tfrac{1}{2n}\right) + \beta\mu\min\left(\tfrac{p_o-p_i+e}{t}, \tfrac{1}{2n}\right)\right.\\
&\quad \left. + (1-\alpha-\beta)\mu\min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right)\right]2n - sn - k, \quad \text{if } p_i - p_o \le 0 \text{ and } p_o \le v - e;\\
&= (p_o - c)\alpha\mu + (p_i - c)(1-\alpha)\mu\min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right)2n - sn - k,\\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{if } p_i - p_o > 0 \text{ and } p_o > v - e;\\
&= (p_o - c)\left[\alpha\mu + \beta\mu\left(\tfrac{1}{2n} - \tfrac{p_o-p_i+e}{t}\right)^+ 2n\right]\\
&\quad + (p_i - c)\left[\beta\mu\min\left(\tfrac{p_o-p_i+e}{t}, \tfrac{1}{2n}\right)\right.\\
&\quad \left. + (1-\alpha-\beta)\mu\min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right)\right]2n - sn - k, \quad \text{if } 0 < p_i - p_o < e \text{ and } p_o \le v - e;\\
&= (p_o - c)(\alpha+\beta)\mu + (p_i - c)(1-\alpha-\beta)\mu\min\left(\tfrac{v-p_i}{t}, \tfrac{1}{2n}\right)2n - sn - k,\\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{if } p_i - p_o \ge e \text{ and } p_o \le v - e.
\end{aligned}
$$

Depending on the relationship between p_o and p_i, $\Pi(n, p_o, p_i)$ takes different forms and as is shown above, there are five possible scenarios. Let $(n^\dagger, p_o^\dagger, p_i^\dagger)$ be the optimal solution that maximizes the profit function $\Pi(n, p_o, p_i)$, and $\Pi^\dagger$ be the optimal profit. The following proposition summarizes the firm's optimal strategy in this case:

Proposition 7.2 *When the firm adopts the on-demand customization technology to serve online demand, there exists a threshold*

$$\bar{e} = \frac{\sqrt{st}}{\sqrt{2\left(1-\frac{\alpha}{4}\right)\mu} + \sqrt{2\left(1-\frac{\alpha}{4}\right)\mu - \frac{2\beta\mu}{\alpha+\beta}}}$$

such that:

(i) *If $e \geq \bar{e}$, the firm offers $n^\dagger = \sqrt{\left(1-\frac{\alpha}{4}\right)\frac{\mu t}{2s}}$ types of product at optimal prices $p_o^\dagger = v - \frac{1}{2}\sqrt{\frac{st}{2(1-\alpha/4)\mu}}$ and $p_i^\dagger = v - \sqrt{\frac{st}{2(1-\alpha/4)\mu}}$. Moreover, $p_o^\dagger > p_i^\dagger$. Under the optimal strategy, Type I customers purchase online if $(p_o^\dagger - p_i^\dagger)/t \leq x \leq 1/(2n^\dagger)$ and in-store if $0 \leq x < (p_o^\dagger - p_i^\dagger)/t$, all Type II and Type III customers purchase in-store. Thus, the firm's online demand is $\alpha\mu/2$ and the in-store demand is $(1-\alpha/2)\mu$. The firm's optimal profit is $\Pi^\dagger = (v-c)\mu - \sqrt{2(1-\alpha/4)st\mu} - k$.*

(ii) *If $e < \bar{e}$, the firm offers $n^\dagger = \sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\mu t/\left[2s - \frac{\beta^2\mu e^2}{(\alpha+\beta)t}\right]}$ types of product at optimal prices*

$$p_o^\dagger = v - \frac{1}{2}\sqrt{\frac{\frac{st}{2} - \frac{\beta^2\mu e^2}{4(\alpha+\beta)}}{\left(1-\frac{\alpha+\beta}{4}\right)\mu}} - \frac{\beta e}{2(\alpha+\beta)} \quad \textit{and} \quad p_i^\dagger = v - \sqrt{\frac{\frac{st}{2} - \frac{\beta^2\mu e^2}{4(\alpha+\beta)}}{\left(1-\frac{\alpha+\beta}{4}\right)\mu}}.$$

Moreover, $p_o^\dagger > p_i^\dagger$. Under the optimal strategy, Type I customers purchase online if $(p_o^\dagger - p_i^\dagger)/t \leq x \leq 1/(2n^\dagger)$ and in-store if $0 \leq x < (p_o^\dagger - p_i^\dagger)/t$, Type II customers purchase online if $(p_o^\dagger - p_i^\dagger + e)/t \leq x \leq 1/(2n^\dagger)$ and in-store if $0 \leq x < (p_o^\dagger - p_i^\dagger + e)/t$, all Type III customers purchase in-store. Thus, the firm's online demand is $\left\{\frac{\alpha+\beta}{2} - \beta e\sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\mu/\left[2st - \frac{\beta^2\mu e^2}{\alpha+\beta}\right]}\right\}\mu$ and the in-store demand is $\left\{1-\frac{\alpha+\beta}{2}+\beta e\sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\mu/\left[2st - \frac{\beta^2\mu e^2}{\alpha+\beta}\right]}\right\}\mu$. The firm's optimal profit is $\Pi^\dagger = (v-c)\mu - \sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\left(2st - \frac{\beta^2\mu e^2}{\alpha+\beta}\right)\mu} - \frac{\beta\mu e}{2} - k$.

Proposition 7.2 characterizes the firm's optimal strategy when adopting on-demand customization online. Depending on whether Type II customers' waiting cost if high (i.e., $e \geq \bar{e}$) or low (i.e., $e < \bar{e}$), the firm's optimal strategy takes different forms and the resulting customer choice is different. Different from the case with traditional technology, when the firm adopts on-demand customization online, the firm should charge a higher price online than in the store. Thus, Type I customers need to trade off the benefit of a better fit from purchasing online and the benefit of a lower price from purchasing in-store. If a Type I customer cannot find a product that is close enough to her ideal type from the firm's in-store offerings (i.e., if $(p_o^\dagger - p_i^\dagger)/t \leq x \leq 1/(2n^\dagger)$), then she purchases online to pursue the improved fit; otherwise she purchases in-store to take advantage of the lower price. Type II customers face the same trade-off as Type I customers, but they also need to factor in their waiting cost from purchasing online. Thus, compared to Type I customers, a smaller proportion of Type II customers choose to purchase online. As stated in Proposition 7.2, if Type II customers' waiting cost is low (i.e., $e < \bar{e}$), then the condition for them to purchase online is $(p_o^\dagger - p_i^\dagger + e)/t \leq x \leq 1/(2n^\dagger)$, which is more stringent than the condition for Type I customers to purchase online. If Type II customers' waiting cost is high (i.e., $e \geq \bar{e}$), then none of them purchases online. As we can see, when the firm adopts on-demand customization online, because the firm offers different levels of customization in different channels, customers' channel choices become more complicated. Although the firm offers infinite types of customized products in the online store, the optimal online price is affected by the misfit from the store and is hence lower than v.

7.3.3 Comparison of Channel Strategies

In this section, we compare the optimal strategies and profits we have obtained in the previous sections. Based on the comparison, we discuss the effect of adopting on-demand customization technology in the online channel on the firm's channel strategies, as well as the value of on-demand customization.

Proposition 7.3

(i) $n^\dagger < n^*$.

(ii) $p_o^\dagger > p_o^*$.

(iii) $p_i^\dagger < p_i^*$.

(iv) *Compared to the case with traditional technology, if* $e > \underline{e}$ *where* $\underline{e} = \frac{(\beta-\alpha)^+}{\beta} \cdot \sqrt{\frac{(\alpha+\beta)st}{2(\alpha+\beta-\alpha\beta)\mu}}$, *the firm's online demand is lower and the in-store demand is higher when it adopts on-demand customization online. If* $e \leq \underline{e}$, *the opposite occurs. Moreover,* $\underline{e} < \bar{e}$.

(v) $\Pi^\dagger \geq \Pi^*$ *if and only if* $k \leq \bar{k}$ *where*

$$\bar{k} = \begin{cases} \sqrt{2st\mu} - \sqrt{2\left(1-\frac{\alpha}{4}\right)st\mu}, & \text{if } e \geq \bar{e}, \\ \sqrt{2st\mu} - \sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\left(2st - \frac{\beta^2\mu e^2}{\alpha+\beta}\right)\mu} - \frac{\beta\mu e}{2} & \text{if } e < \bar{e}. \end{cases}$$

Moreover, $\bar{k} > 0$, *so* $\Pi^\dagger > \Pi^*$ *when* $k = 0$.

(vi) $\Pi^\dagger - \Pi^*$ *is increasing in both t and s.*

In Proposition 7.3, we compare the firm's optimal strategies when it adopts on-demand customization online, and when it uses the traditional technology. When the firm starts to sell customized products online, its online price increases because it can charge a price premium due to perfect customization. At the same time, the firm offers a smaller product selection in-store and correspondingly decreases its price in the store.

Part (iv) of Proposition 7.3 characterizes how the firm's demand in each channel changes when it adopts on-demand customization online. Recall that when the firm uses the traditional technology, Type I customers purchase online and Type II customers purchase in-store. When the firm uses on-demand customization online, it segments customers in a different way. As we have discussed above, both Type I and Type II customers need to trade off the better fit from purchasing online and the lower price from purchasing in-store. Thus, those Type I customers who have low fit costs will switch to purchase in-store and those Type II customers who have high fit costs will switch to purchase online. Moreover, the switching of Type II customers is also determined by their waiting cost e. If their waiting cost is high (i.e., $e > \underline{e}$), not many Type II customers would want to switch to purchase online, so the firm's total online demand becomes lower, whereas the total in-store demand becomes higher when adopting on-demand customization online. On the other hand, if the waiting cost is low (i.e., $e \leq \underline{e}$), then Type II customers who switch to purchase online outnumber the Type I customers who switch to purchase in-store, so the firm's total online demand becomes higher, whereas the total in-store demand becomes lower when the firm adopts on-demand customization online. In addition, notice that when $\alpha \geq \beta$, $\underline{e} = 0$. In this case, the firm's online demand always becomes lower while the in-store demand always becomes higher when adopting on-demand customization online. This is because Type I customers outnumber Type II customers, and the switching of Type I customers to the store dominates the switching of Type II customers to online. Therefore, when adopting on-demand customization online, the firm should have a good understanding of the customers' waiting cost from purchasing online, and then determine the product offering and pricing strategies accordingly.

Part (v) of Proposition 7.3 states that the firm achieves a higher profit by adopting on-demand customization online as long as the fixed cost of technology adoption is not too high (i.e., $k \leq \bar{k}$). Additionally, the threshold $\bar{k}$ is strictly positive, indicating that if we ignore the fixed cost as a sunk cost, the operating profit is always improved when the firm adopts on-demand customization online. Adopting on-demand customization online creates two benefits for the firm. First, the new technology allows the firm to achieve *perfect customization* and eliminate the fit cost for the customers, and hence enables the firm to charge a price premium for products sold online. Second, adopting on-demand customization online reduces the number of product types that the firm offers in the store, and hence reduces the firm's production setup cost. Therefore, the firm achieves another benefit of *setup cost reduction*. Part (vi) of Proposition 7.3 indicates that the benefit of perfect customization becomes stronger when customers' fit cost is higher, and the benefit of setup cost reduction becomes stronger when it is more costly to switch over between product types under the traditional production technology.

7.4 Concluding Remarks

In this chapter, we have presented a baseline model to analyze the impact of on-demand customization on retail product offering and pricing in a dual-channel setting. Overall, we have shown that adopting the on-demand customization technology for the online channel gives rise to a *substitution effect* of technological innovation. Such technology substitution has differential impacts to the online versus offline channels. It allows the firm to offer perfect customization and charge a price premium for online customers. At the same time, the firm offers a smaller product variety in the store at a reduced price. The magnitude of online demands will increase as a result of this adoption, when the cost of waiting for online customers is low. With the trend of online merchants offering faster and faster deliveries to orders, our result shows that customization technology will also help to accelerate the growth of the online markets.

Our work is a first step toward understanding the impact of the on-demand customization technology on retail supply chains. We are currently studying several important extensions of the current baseline model in the context of 3D printing (Chen et al. 2018). First, for analytical tractability, we have assumed the total customer demand as deterministic. Relaxing this assumption to allow for random demand would reveal further insights on the benefits of the build-to-order production enabled by the new technology, over the build-to-stock mode of traditional production technology. Second, with random demand, there will be inventory mismatch at the retail stores. When there is a stockout at a retail store, customers may switch to buy online. It is important to understand how such stockout-based substitution would affect the firm's channel strategies. Third, the on-demand customization technology such as 3D printing may involve a higher marginal production cost than the traditional technology, which might be another factor that affects the firm's channel strategies. Finally, as technological advancement continues or when there is more widespread adoption of the customization technology that drives down the cost of implementing such a technology, it may be possible for on-demand customization technology to be used to meet the in-store demand. When that happens, there will be additional structural changes in the supply chain. It is important to understand how such structural changes would affect the firm's channel strategies.

Appendix: Proofs of Propositions

Proof of Proposition 7.1

To analyze the profit function and derive the optimal strategy, we define three subcases in Case 1: Case 1.1 ($p_i - p_o \geq e$), Case 1.2 ($0 \leq p_i - p_o < e$), Case 1.3 ($p_i - p_o < 0$).

In Case 1.1, depending on n, the profit function becomes

$$
\begin{aligned}
&\Pi_{1.1}(n, p_o, p_i)\\
&\quad = (p_o - c)(\alpha + \beta)\mu + (p_i - c)(1 - \alpha - \beta)\mu - sn, \quad \text{if } \tfrac{1}{2n} < \tfrac{v-p_i}{t};\\
&\quad = (p_o - c)(\alpha + \beta)\mu + (p_i - c)(1 - \alpha - \beta)\mu \left(\tfrac{v-p_i}{t}\right) 2n - sn,\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \tfrac{v-p_i}{t} \le \tfrac{1}{2n} < \tfrac{v-p_o-e}{t};\\
&\quad = (p_o - c)\left[\alpha\mu + \beta\mu\left(\tfrac{v-p_o-e}{t}\right)2n\right]\\
&\qquad + (p_i - c)(1 - \alpha - \beta)\mu\left(\tfrac{v-p_i}{t}\right)2n - sn, \qquad \text{if } \tfrac{v-p_o-e}{t} \le \tfrac{1}{2n} < \tfrac{v-p_o}{t};\\
&\quad = (p_o - c)\left[\alpha\mu\left(\tfrac{v-p_o}{t}\right) + \beta\mu\left(\tfrac{v-p_o-e}{t}\right)\right]2n\\
&\qquad + (p_i - c)(1 - \alpha - \beta)\mu\left(\tfrac{v-p_i}{t}\right)2n - sn, \qquad \text{if } \tfrac{1}{2n} \ge \tfrac{v-p_o}{t}.
\end{aligned}
$$

If $\frac{1}{2n} < \frac{v-p_i}{t}$ which requires n is large enough, $\Pi_{1.1}(n, p_o, p_i)$ is decreasing in n, so $\frac{1}{2n} < \frac{v-p_i}{t}$ is dominated by $\frac{v-p_i}{t} \le \frac{1}{2n} < \frac{v-p_o-e}{t}$. If $\frac{v-p_i}{t} \le \frac{1}{2n} < \frac{v-p_o-e}{t}$ which requires p_o is small enough, $\Pi_{1.1}(n, p_o, p_i)$ is increasing in p_o, so $\frac{v-p_i}{t} \le \frac{1}{2n} < \frac{v-p_o-e}{t}$ is dominated by $\frac{1}{2n} \ge \frac{v-p_o-e}{t}$. Thus, the optimal strategy can only be supported by $\frac{1}{2n} \ge \frac{v-p_o-e}{t}$ in Case 1.1.

In Case 1.2, depending on n, the profit function becomes

$$
\begin{aligned}
&\Pi_{1.2}(n, p_o, p_i)\\
&\quad = (p_o - c)\alpha\mu + (p_i - c)(1 - \alpha)\mu - sn, \qquad \text{if } \tfrac{1}{2n} < \tfrac{v-p_i}{t};\\
&\quad = (p_o - c)\alpha\mu + (p_i - c)(1 - \alpha)\mu\left(\tfrac{v-p_i}{t}\right)2n - sn, \quad \text{if } \tfrac{v-p_i}{t} \le \tfrac{1}{2n} < \tfrac{v-p_o}{t};\\
&\quad = (p_o - c)\alpha\mu\left(\tfrac{v-p_o}{t}\right)2n + (p_i - c)(1 - \alpha)\mu\left(\tfrac{v-p_i}{t}\right)2n - sn,\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \tfrac{1}{2n} \ge \tfrac{v-p_o}{t}.
\end{aligned}
$$

If $\frac{1}{2n} < \frac{v-p_i}{t}$ which requires n is large enough, $\Pi_{1.2}(n, p_o, p_i)$ is decreasing in n, so $\frac{1}{2n} < \frac{v-p_i}{t}$ is dominated by $\frac{v-p_i}{t} \le \frac{1}{2n} < \frac{v-p_o}{t}$. If $\frac{v-p_i}{t} \le \frac{1}{2n} < \frac{v-p_o}{t}$ which requires p_o is small enough, $\Pi_{1.2}(n, p_o, p_i)$ is increasing in p_o, so $\frac{v-p_i}{t} \le \frac{1}{2n} < \frac{v-p_o}{t}$ is dominated by $\frac{1}{2n} \ge \frac{v-p_o}{t}$. Thus, the optimal strategy can only be supported by $\frac{1}{2n} \ge \frac{v-p_o}{t}$ in Case 1.2.

In Case 1.3, depending on n, the profit function becomes

$$
\Pi_{1.3}(n, p_o, p_i) = \begin{cases} (p_i - c)\mu - sn & \text{if } \frac{1}{2n} < \frac{v-p_i}{t},\\ (p_i - c)\mu\left(\frac{v-p_i}{t}\right)2n - sn & \text{if } \frac{1}{2n} \ge \frac{v-p_i}{t}. \end{cases}
$$

If $\frac{1}{2n} < \frac{v-p_i}{t}$ which requires n is large enough, $\Pi_{1.3}(n, p_o, p_i)$ is decreasing in n, so the optimal strategy can only be supported by $\frac{1}{2n} \geq \frac{v-p_i}{t}$ in Case 1.3. Moreover, for any (n, p_o, p_i) in Case 1.3 such that $\frac{1}{2n} \geq \frac{v-p_i}{t}$, we can pick $(\hat{n}, \hat{p}_o, \hat{p}_i) = (n, p_i, p_i)$ which is in Case 1.2 and yields $\Pi_{1.2}(\hat{n}, \hat{p}_o, \hat{p}_i) = \Pi_{1.3}(n, p_o, p_i)$. Thus, Case 1.3 is dominated by Case 1.2, and hence the optimal strategy can only be supported by Case 1.1 with $\frac{1}{2n} \geq \frac{v-p_o-e}{t}$ or Case 1.2 with $\frac{1}{2n} \geq \frac{v-p_o}{t}$.

We now optimize n. In Case 1.1, $\Pi_{1.1}(n, p_o, p_i)$ is linear in n for $\frac{1}{2n} \geq \frac{v-p_o}{t}$ and for $\frac{v-p_o-e}{t} \leq \frac{1}{2n} < \frac{v-p_o}{t}$, so the optimal n is either $n^*(p_o, p_i) = \frac{t}{2(v-p_o)}$ or $n^*(p_o, p_i) = \frac{t}{2(v-p_o-e)}$. With $n^*(p_o, p_i) = \frac{t}{2(v-p_o)}$, the profit function reduces to

$$\Pi_{1.1.1}(p_o, p_i) = (p_o - c)\left[\alpha\mu + \beta\mu\left(\frac{v-p_o-e}{v-p_o}\right)\right] + (p_i - c)(1-\alpha-\beta)\mu\left(\frac{v-p_i}{v-p_o}\right) - \frac{st}{2(v-p_o)}. \tag{7.1}$$

With $n^*(p_o, p_i) = \frac{t}{2(v-p_o-e)}$, the profit function reduces to

$$\Pi_{1.1.2}(p_o, p_i) = (p_o - c)(\alpha+\beta)\mu + (p_i - c)(1-\alpha-\beta)\mu\left(\frac{v-p_i}{v-p_o-e}\right) - \frac{st}{2(v-p_o-e)}. \tag{7.2}$$

Note that without loss of generality, we ignore the possibility of $n^*(p_o, p_i) = 0$, in which case the profit is zero, because the optimal profit is guaranteed to be non-negative. In Case 1.2, $\Pi_{1.2}(n, p_o, p_i)$ is linear in n for $\frac{1}{2n} \geq \frac{v-p_o}{t}$, so the optimal n is $n^*(p_o, p_i) = \frac{t}{2(v-p_o)}$ which reduces the profit function to

$$\Pi_{1.2}(p_o, p_i) = (p_o - c)\alpha\mu + (p_i - c)(1-\alpha)\mu\left(\frac{v-p_i}{v-p_o}\right) - \frac{st}{2(v-p_o)}. \tag{7.3}$$

Next, we optimize p_i. We only need to consider $\Pi_{1.1.1}(p_o, p_i)$, $\Pi_{1.1.2}(p_o, p_i)$ and $\Pi_{1.2}(p_o, p_i)$. First, consider $\Pi_{1.1.1}(p_o, p_i)$. Taking derivative of (7.1) with respect to p_i yields $\frac{\partial \Pi_{1.1.1}}{\partial p_i} = (1-\alpha-\beta)\mu\left(\frac{v+c-2p_i}{v-p_o}\right)$. Thus, $\Pi_{1.1.1}(p_o, p_i)$ is concave in p_i and solving the first-order condition yields $p_i = \frac{v+c}{2}$. Since Case 1.1 requires $p_i - p_o \geq e$, the optimal p_i is

$$p_i^*(p_o) = \begin{cases} \frac{v+c}{2} & \text{for } p_o < \frac{v+c}{2} - e, \\ p_o + e & \text{for } p_o \geq \frac{v+c}{2} - e. \end{cases}$$

With $p_i = p_i^*(p_o)$, $\Pi_{1.1.1}(p_o, p_i)$ reduces to

$$
\begin{aligned}
&\Pi_{1.1.1}(p_o)\\
&= (p_o - c)\left[\alpha\mu + \beta\mu\left(\tfrac{v-p_o-e}{v-p_o}\right)\right] + (1-\alpha-\beta)\mu\left(\tfrac{v-c}{2}\right)^2 \tfrac{1}{v-p_o} - \tfrac{st}{2(v-p_o)},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{for } p_o < \tfrac{v+c}{2} - e; \qquad (7.4)\\
&= (p_o - c)\left[\alpha\mu + \beta\mu\left(\tfrac{v-p_o-e}{v-p_o}\right)\right] + (p_o + e - c)(1-\alpha-\beta)\mu\left(\tfrac{v-p_o-e}{v-p_o}\right) - \tfrac{st}{2(v-p_o)},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{for } p_o \geq \tfrac{v+c}{2} - e. \qquad (7.5)
\end{aligned}
$$

Second, consider $\Pi_{1.1.2}(p_o, p_i)$. Following similar analysis for $\Pi_{1.1.1}(p_o, p_i)$, we can obtain from (7.2) that for $\Pi_{1.1.2}(p_o, p_i)$, the optimal p_i is same as $p_i^*(p_o)$ for $\Pi_{1.1.1}(p_o, p_i)$, and $\Pi_{1.1.2}(p_o, p_i)$ is reduced to

$$
\begin{aligned}
&\Pi_{1.1.2}(p_o)\\
&= (p_o - c)(\alpha+\beta)\mu + (1-\alpha-\beta)\mu\left(\tfrac{v-c}{2}\right)^2 \tfrac{1}{v-p_o-e} - \tfrac{st}{2(v-p_o-e)},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{for } p_o < \tfrac{v+c}{2} - e; \qquad (7.6)\\
&= (p_o - c)(\alpha+\beta)\mu + (p_o + e - c)(1-\alpha-\beta)\mu - \tfrac{st}{2(v-p_o)},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{for } p_o \geq \tfrac{v+c}{2} - e. \qquad (7.7)
\end{aligned}
$$

Third, consider $\Pi_{1.2}(p_o, p_i)$. From (7.3) it is easy to see that $\Pi_{1.2}(p_o, p_i)$ is concave in p_i and the first-order condition yields $p_i = \frac{v+c}{2}$. Since Case 1.2 requires $0 \leq p_i - p_o < e$, the optimal p_i is

$$
p_i^*(p_o) = \begin{cases} p_o + e & \text{for } p_o < \frac{v+c}{2} - e,\\ \frac{v+c}{2} & \text{for } \frac{v+c}{2} - e \leq p_o < \frac{v+c}{2},\\ p_o & \text{for } p_o \geq \frac{v+c}{2}. \end{cases}
$$

With $p_i = p_i^*(p_o)$, $\Pi_{1.2}(p_o, p_i)$ reduces to

$$
\begin{aligned}
&\Pi_{1.2}(p_o)\\
&= (p_o - c)\alpha\mu + (p_o + e - c)(1-\alpha)\mu\left(\tfrac{v-p_o-e}{v-p_o}\right) - \tfrac{st}{2(v-p_o)},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{for } p_o < \tfrac{v+c}{2} - e; \qquad (7.8)\\
&= (p_o - c)\alpha\mu + (1-\alpha)\mu\left(\tfrac{v-c}{2}\right)^2 \tfrac{1}{v-p_o} - \tfrac{st}{2(v-p_o)},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{for } \tfrac{v+c}{2} - e \leq p_o < \tfrac{v+c}{2}; \qquad (7.9)\\
&= (p_o - c)\mu - \tfrac{st}{2(v-p_o)}, \qquad\qquad \text{for } p_o \geq \tfrac{v+c}{2}. \qquad (7.10)
\end{aligned}
$$

Next, we optimize p_o and obtain $\Pi^*_{1.1.1}$, $\Pi^*_{1.1.2}$ and $\Pi^*_{1.2}$. By comparing $\Pi^*_{1.1.1}$, $\Pi^*_{1.1.2}$ and $\Pi^*_{1.2}$, we will obtain which subcase supports the optimal strategy.

First, consider $\Pi_{1.1.1}(p_o)$. For $\frac{v+c}{2} - e \le p_o < \frac{v+c}{2}$, from (7.5) we have

$$\Pi_{1.1.1}(p_o) < (p_o - c)\alpha\mu + (p_o + e - c)(1-\alpha)\mu\left(\frac{v - p_o - e}{v - p_o}\right) - \frac{st}{2(v-p_o)}$$
$$\le (p_o - c)\alpha\mu + (1-\alpha)\mu\left(\frac{v-c}{2}\right)^2 \frac{1}{v-p_o} - \frac{st}{2(v-p_o)} = \Pi_{1.2}(p_o),$$

where the first inequality is straightforward and the second inequality follows from the fact that $(p_o+e-c)(v-p_o-e)$ is maximized at $p_o = \frac{v+c}{2} - e$ and its maximum value is $\left(\frac{v-c}{2}\right)^2$. Moreover, for $p_o \ge \frac{v+c}{2}$, from (7.5) we have

$$\Pi_{1.1.1}(p_o) < (p_o - c)(\alpha+\beta)\mu + (p_o + e - c)(1-\alpha-\beta)\mu\left(\frac{v - p_o - e}{v - p_o}\right)$$
$$- \frac{st}{2(v-p_o)}$$
$$< (p_o - c)(\alpha+\beta)\mu + (p_o - c)(1-\alpha-\beta)\mu - \frac{st}{2(v-p_o)}$$
$$= (p_o - c)\mu - \frac{st}{2(v-p_o)} = \Pi_{1.2}(p_o),$$

where the first inequality is straightforward and the second inequality follows from the fact that $(p_o+e-c)(v-p_o-e)$ is decreasing in w for $p_o \ge \frac{v+c}{2}$ so $(p_o+e-c)\cdot(v-p_o-e) < (p_o-c)(v-p_o)$. Thus, for $p_o \ge \frac{v+c}{2} - e$, $\Pi_{1.1.1}(p_o) < \Pi_{1.2}(p_o)$. Now, consider $p_o < \frac{v+c}{2} - e$. Taking derivative of (7.4) with respect to p_o yields

$$\frac{\partial \Pi_{1.1.1}}{\partial p_o} = (\alpha+\beta)\mu + \frac{1}{(v-p_o)^2}\left[-\beta\mu(v-c)e + (1-\alpha-\beta)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2}\right].$$

If $-\beta\mu(v-c)e + (1-\alpha-\beta)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2} \ge 0$, then $\frac{\partial \Pi_{1.1.1}}{\partial p_o} > 0$ for $p_o < \frac{v+c}{2} - e$, so for $\Pi_{1.1.1}(p_o)$, the optimal p_o is achieved in $p_o \ge \frac{v+c}{2} - e$, and hence we must have $\Pi^*_{1.1.1} < \Pi^*_{1.2}$. If $-\beta\mu(v-c)e + (1-\alpha-\beta)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2} < 0$, then $\frac{\partial \Pi_{1.1.1}}{\partial p_o}$ is decreasing in p_o, so $\Pi_{1.1.1}(p_o)$ is concave in p_o for $p_o < \frac{v+c}{2} - e$. The derivative at $p_o = \frac{v+c}{2} - e$ is

$$\left.\frac{\partial_- \Pi_{1.1.1}}{\partial p_o}\right|_{p_o=(v+c)/2-e} = \frac{1}{((v-c)/2+e)^2}\left[(\alpha+\beta)\mu\left(\frac{v-c}{2}+e\right)^2 - \beta\mu(v-c)e\right.$$
$$\left. + (1-\alpha-\beta)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2}\right]$$

$$= \frac{1}{((v-c)/2+e)^2}\left[\mu\left(\frac{v-c}{2}\right)^2 + \alpha\mu(v-c)e + (\alpha+\beta)\mu e^2 - \frac{st}{2}\right]$$

$$\geq \frac{1}{((v-c)/2+e)^2}\left[\alpha\mu(v-c)e + (\alpha+\beta)\mu e^2\right] > 0,$$

where the inequality follows from the profitability assumption that $(v-c)\mu \geq \sqrt{2st\mu}$. Since $\Pi_{1.1.1}(p_o)$ is concave in p_o for $p_o < \frac{v-c}{2} - e$, we then have $\frac{\partial \Pi_{1.1.1}}{\partial p_o} > 0$ for $p_o < \frac{v-c}{2} - e$, which again indicates $\Pi^*_{1.1.1} < \Pi^*_{1.2}$.

Second, consider $\Pi_{1.1.2}(p_o)$. For $p_o \geq \frac{v+c}{2} - e$, the derivative of (7.7) with respect to p_o is $\frac{\partial \Pi_{1.1.2}}{\partial p_o} = \mu - \frac{st}{2(v-p_o-e)^2}$ which is decreasing in p_o, so $\Pi_{1.1.2}(p_o)$ is concave in p_o for $p_o \geq \frac{v+c}{2} - e$. At $p_o = \frac{v+c}{2} - e$, $\frac{\partial_+ \Pi_{1.1.2}}{\partial p_o}\big|_{p_o=(v+c)/2-e} = \mu - \frac{2st}{(v-c)^2} \geq 0$. Moreover, for $p_o < \frac{v+c}{2} - e$, the derivative of (7.6) with respect to p_o is $\frac{\partial \Pi_{1.1.2}}{\partial p_o} = (\alpha+\beta)\mu + \frac{1}{(v-p_o-e)^2}\left[(1-\alpha-\beta)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2}\right]$. If $(1-\alpha-\beta)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2} \geq 0$, then $\frac{\partial \Pi_{1.1.2}}{\partial p_o} > 0$ for $p_o < \frac{v+c}{2} - e$. If $(1-\alpha-\beta)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2} < 0$, then $\frac{\partial \Pi_{1.1.2}}{\partial p_o}$ is decreasing in p_o. Then, since $\frac{\partial_- \Pi_{1.1.2}}{\partial p_o}\big|_{p_o=(v+c)/2-e} = \mu - \frac{2st}{(v-c)^2} \geq 0$, we have $\frac{\partial \Pi_{1.1.2}}{\partial p_o} > 0$ for $p_o < \frac{v+c}{2} - e$. Thus, we conclude that the optimal p_o is achieved in $p_o \geq \frac{v+c}{2} - e$ and the first-order condition yields $p_o^* = v - e - \sqrt{st/(2\mu)}$. Correspondingly, $\Pi^*_{1.1.2} = (v-c)\mu - (\alpha+\beta)\mu e - \sqrt{2st\mu}$.

Third, consider $\Pi_{1.2}(p_o)$. For $p_o \geq \frac{v+c}{2}$, the derivative of (7.10) with respect to p_o is $\frac{\partial \Pi_{1.2}}{\partial p_o} = \mu - \frac{st}{2(v-p_o)^2}$ which is decreasing in p_o, so $\Pi_{1.2}(p_o)$ is concave in p_o for $p_o \geq \frac{v+c}{2}$. Moreover, $\frac{\partial_+ \Pi_{1.2}}{\partial p_o}\big|_{p_o=(v+c)/2} = \mu - \frac{2st}{(v-c)^2} \geq 0$. For $\frac{v+c}{2} - e \leq p_o < \frac{v+c}{2}$, the derivative of (7.9) with respect to p_o is $\frac{\partial \Pi_{1.2}}{\partial p_o} = \alpha\mu + \frac{1}{(v-p_o)^2} \cdot \left[(1-\alpha)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2}\right]$. If $(1-\alpha)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2} \geq 0$, then $\frac{\partial \Pi_{1.2}}{\partial p_o} > 0$ for $\frac{v+c}{2} - e \leq p_o < \frac{v+c}{2}$. If $(1-\alpha)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2} < 0$, then since $\frac{\partial_- \Pi_{1.2}}{\partial p_o}\big|_{p_o=(v+c)/2} = \mu - \frac{2st}{(v-c)^2} \geq 0$, we have $\frac{\partial \Pi_{1.2}}{\partial p_o} > 0$ for $\frac{v+c}{2} - e \leq p_o < \frac{v+c}{2}$. Thus, $\frac{v+c}{2} - e \leq p_o < \frac{v+c}{2}$ is dominated by $p_o \geq \frac{v+c}{2}$. Finally, for $p_o < \frac{v+c}{2} - e$, taking derivatives of (7.8) yields $\frac{\partial \Pi_{1.2}}{\partial p_o} = \alpha\mu + \frac{(1-\alpha)\mu}{(v-p_o)^2}\left[(v-p_o-e)(v-p_o) - (p_o+e-c)e - \frac{st}{2(v-p_o)^2}\right]$, and $\frac{\partial^2 \Pi_{1.2}}{\partial p_o^2} = -\frac{2(1-\alpha)\mu e(v-c+e)}{(v-p_o)^3} - \frac{st}{(v-p_o)^3} < 0$. So, $\Pi_{1.2}(p_o)$ is concave in p_o for $p_o < \frac{v+c}{2} - e$. Moreover, at $p_o = \frac{v+c}{2} - e$, $\frac{\partial_- \Pi_{1.2}}{\partial p_o}\big|_{p_o=(v+c)/2-e} = \alpha\mu + \frac{1}{((v-c)/2+e)^2} \cdot \left[(1-\alpha)\mu\left(\frac{v-c}{2}\right)^2 - \frac{st}{2}\right] = .\frac{\partial_+ \Pi_{1.2}}{\partial p_o}\big|_{p_o=(v+c)/2-e} > 0$. Note that we have shown that $\frac{\partial \Pi_{1.2}}{\partial p_o} > 0$ for $\frac{v+c}{2} - e \leq p_o < \frac{v+c}{2}$. Therefore, we conclude that the optimal p_o is achieved in $p_o \geq \frac{v+c}{2}$ and the first-order condition yields $p_o^* = v - \sqrt{st/(2\mu)}$. Correspondingly, $\Pi^*_{1.2} = (v-c)\mu - \sqrt{2st\mu} > \Pi^*_{1.1.2}$. We have also shown that $\Pi^*_{1.1.1} < \Pi^*_{1.2}$. Therefore, the optimal strategy is achieved in Case 1.2 and the optimal profit is $\Pi^* = (v-c)\mu - \sqrt{2st\mu}$. Tracing back our analysis for Case 1.2, we obtain that $p_o^* = v - \sqrt{st/(2\mu)}$, $p_i^* = p_o^*$, and $n^* = \frac{t}{2(v-p_o^*)} = \sqrt{\mu t/(2s)}$.

Moreover, under the optimal strategy, all Type I customers purchase online, all Type II and Type III customers purchase in-store. The proof is complete. □

Proof of Proposition 7.2

To analyze the profit function and derive the optimal strategy, we define five subcases in Case 2: Case 2.1 ($p_i - p_o \le 0$ and $p_o > v - e$), Case 2.2 ($p_i - p_o \le 0$ and $p_o \le v - e$), Case 2.3 ($p_i - p_o > 0$ and $p_o > v - e$), Case 2.4 ($0 < p_i - p_o < e$ and $p_o \le v - e$), Case 2.5 ($p_i - p_o \ge e$ and $p_o \le v - e$).

In Case 2.1, depending on n, the profit function becomes

$$
\begin{aligned}
&\Pi_{2.1}(n, p_o, p_i)\\
&\quad = (p_i - c)\mu - sn - k, \qquad \text{if } \tfrac{1}{2n} < \tfrac{p_o - p_i}{t};\\
&\quad = (p_o - c)\alpha\mu\left(\tfrac{1}{2n} - \tfrac{p_o - p_i}{t}\right)2n + (p_i - c)\left[\alpha\mu\left(\tfrac{p_o - p_i}{t}\right)2n + (1-\alpha)\mu\right] - sn - k,\\
&\qquad\qquad \text{if } \tfrac{p_o - p_i}{t} \le \tfrac{1}{2n} < \tfrac{v - p_i}{t};\\
&\quad = (p_o - c)\alpha\mu\left(\tfrac{1}{2n} - \tfrac{p_o - p_i}{t}\right)2n + (p_i - c)\left[\alpha\mu\left(\tfrac{p_o - p_i}{t}\right) + (1-\alpha)\mu\left(\tfrac{v - p_i}{t}\right)\right]2n\\
&\qquad - sn - k, \qquad \text{if } \tfrac{1}{2n} \ge \tfrac{v - p_i}{t}.
\end{aligned}
$$

If $\frac{1}{2n} < \frac{p_o - p_i}{t}$ which requires n is large enough, $\Pi(n, p_o, p_i)$ is decreasing in n, so $\frac{1}{2n} < \frac{p_o - p_i}{t}$ is dominated by $\frac{p_o - p_i}{t} \le \frac{1}{2n} < \frac{v - p_i}{t}$. If $\frac{p_o - p_i}{t} \le \frac{1}{2n} < \frac{v - p_i}{t}$, since $\frac{\partial \Pi}{\partial n} = -\frac{2\alpha\mu(p_o - p_i)^2}{t} - s < 0$, $\frac{p_o - p_i}{t} \le \frac{1}{2n} < \frac{v - p_i}{t}$ is dominated by $\frac{1}{2n} \ge \frac{v - p_i}{t}$. Thus, the optimal strategy can only be supported by $\frac{1}{2n} \ge \frac{v - p_i}{t}$ in Case 2.1.

In Case 2.2, depending on n, the profit function becomes

$$
\begin{aligned}
&\Pi_{2.2}(n, p_o, p_i)\\
&= (p_i - c)\mu - sn - k, \qquad \text{if } \tfrac{1}{2n} < \tfrac{p_o - p_i}{t};\\
&= (p_o - c)\alpha\mu\left(\tfrac{1}{2n} - \tfrac{p_o - p_i}{t}\right)2n + (p_i - c)\left[\alpha\mu\left(\tfrac{p_o - p_i}{t}\right)2n + (1-\alpha)\mu\right] - sn - k,\\
&\qquad\qquad \text{if } \tfrac{p_o - p_i}{t} \le \tfrac{1}{2n} < \tfrac{p_o - p_i + e}{t};\\
&= (p_o - c)\left[\alpha\mu\left(\tfrac{1}{2n} - \tfrac{p_o - p_i}{t}\right) + \beta\mu\left(\tfrac{1}{2n} - \tfrac{p_o - p_i + e}{t}\right)\right]2n\\
&\quad + (p_i - c)\left[\alpha\mu\left(\tfrac{p_o - p_i}{t}\right)2n + \beta\mu\left(\tfrac{p_o - p_i + e}{t}\right)2n + (1-\alpha-\beta)\mu\right] - sn - k,\\
&\qquad\qquad \text{if } \tfrac{p_o - p_i + e}{t} \le \tfrac{1}{2n} < \tfrac{v - p_i}{t};\\
&= (p_o - c)\left[\alpha\mu\left(\tfrac{1}{2n} - \tfrac{p_o - p_i}{t}\right) + \beta\mu\left(\tfrac{1}{2n} - \tfrac{p_o - p_i + e}{t}\right)\right]2n,
\end{aligned}
$$

$$+ (p_i - c)\left[\alpha\mu\left(\tfrac{p_o-p_i}{t}\right) + \beta\mu\left(\tfrac{p_o-p_i+e}{t}\right) + (1-\alpha-\beta)\mu\left(\tfrac{v-p_i}{t}\right)\right]2n - sn - k,$$
$$\text{if } \tfrac{1}{2n} \geq \tfrac{v-p_i}{t}.$$

It can be easily shown that $\Pi(n, p_o, p_i)$ is decreasing in n if $\frac{1}{2n} < \frac{v-p_i}{t}$. Thus, the optimal strategy can only be supported by $\frac{1}{2n} \geq \frac{v-p_i}{t}$ in Case 2.2.

In Case 2.3, depending on n, the profit function becomes

$$\Pi_{2.3}(n, p_o, p_i) = \begin{cases} (p_o - c)\alpha\mu + (p_i - c)(1-\alpha)\mu - sn - k, & \text{if } \frac{1}{2n} < \frac{v-p_i}{t}; \\ (p_o - c)\alpha\mu + (p_i - c)(1-\alpha)\mu\left(\frac{v-p_i}{t}\right)2n - sn - k, & \text{if } \frac{1}{2n} \geq \frac{v-p_i}{t}. \end{cases}$$

Since $\Pi(n, p_o, p_i)$ is decreasing in n if $\frac{1}{2n} < \frac{v-p_i}{t}$, the optimal strategy can only be supported by $\frac{1}{2n} \geq \frac{v-p_i}{t}$ in Case 2.3.

In Case 2.4, depending on n, the profit function becomes

$$\begin{aligned} &\Pi_{2.4}(n, p_o, p_i) \\ &= (p_o - c)\alpha\mu + (p_i - c)(1-\alpha)\mu - sn - k, \qquad \text{if } \tfrac{1}{2n} < \tfrac{p_o-p_i+e}{t}; \\ &= (p_o - c)\left[\alpha\mu + \beta\mu\left(\tfrac{1}{2n} - \tfrac{p_o-p_i+e}{t}\right)2n\right] \\ &\quad + (p_i - c)\left[\beta\mu\left(\tfrac{p_o-p_i+e}{t}\right)2n + (1-\alpha-\beta)\mu\right] - sn - k, \\ &\qquad\qquad \text{if } \tfrac{p_o-p_i+e}{t} \leq \tfrac{1}{2n} < \tfrac{v-p_i}{t}; \\ &= (p_o - c)\left[\alpha\mu + \beta\mu\left(\tfrac{1}{2n} - \tfrac{p_o-p_i+e}{t}\right)2n\right] \\ &\quad + (p_i - c)\left[\beta\mu\left(\tfrac{p_o-p_i+e}{t}\right) + (1-\alpha-\beta)\mu\left(\tfrac{v-p_i}{t}\right)\right]2n - sn - k, \\ &\qquad\qquad \text{if } \tfrac{1}{2n} \geq \tfrac{v-p_i}{t}. \end{aligned}$$

Since $\Pi(n, p_o, p_i)$ is decreasing in n if $\frac{1}{2n} < \frac{p_o-p_i+e}{t}$, the optimal strategy can only be supported by $\frac{1}{2n} \geq \frac{p_o-p_i+e}{t}$ in Case 2.4.

In Case 2.5, depending on n, the profit function becomes

$$\Pi_{2.5}(n, p_o, p_i) = \begin{cases} (p_o - c)(\alpha+\beta)\mu + (p_i - c)(1-\alpha-\beta)\mu - sn - k, & \text{if } \frac{1}{2n} < \frac{v-p_i}{t}; \\ (p_o - c)(\alpha+\beta)\mu + (p_i - c)(1-\alpha-\beta)\mu\left(\frac{v-p_i}{t}\right)2n - sn - k, & \text{if } \frac{1}{2n} \geq \frac{v-p_i}{t}. \end{cases}$$

Since $\Pi(n, p_o, p_i)$ is decreasing in n if $\frac{1}{2n} < \frac{v-p_i}{t}$, the optimal strategy can only be supported by $\frac{1}{2n} \geq \frac{v-p_i}{t}$ in Case 2.5.

Now we optimize n. Note that $\Pi(n, p_o, p_i)$ is piecewise linear in n in all subcases. In Case 2.1, $n^\dagger(p_o, p_i) = \frac{t}{2(v-p_i)}$, and the profit function reduces to

$$\begin{aligned}\Pi_{2.1}(p_o, p_i) = (p_o - c)\alpha\mu\left(\frac{v-p_o}{v-p_i}\right) + (p_i - c)\left[\alpha\mu\left(\frac{p_o-p_i}{v-p_i}\right) + (1-\alpha)\mu\right]\\ - \frac{st}{2(v-p_i)} - k. \qquad (7.11)\end{aligned}$$

In Case 2.2, $n^\dagger(p_o, p_i) = \frac{t}{2(v-p_i)}$, and the profit function reduces to

$$\begin{aligned}\Pi_{2.2}(p_o, p_i) = (p_o - c)\left[\alpha\mu\left(\frac{v-p_o}{v-p_i}\right) + \beta\mu\left(\frac{v-p_o-e}{v-p_i}\right)\right]\\ + (p_i - c)\left[\alpha\mu\left(\frac{p_o-p_i}{v-p_i}\right) + \beta\mu\left(\frac{p_o-p_i+e}{v-p_i}\right) + (1-\alpha-\beta)\mu\right]\\ - \frac{st}{2(v-p_i)} - k. \qquad (7.12)\end{aligned}$$

In Case 2.3, $n^\dagger(p_o, p_i) = \frac{t}{2(v-p_i)}$, and the profit function reduces to $\Pi_{2.3}(p_o, p_i) = (p_o - c)\alpha\mu + (p_i - c)(1-\alpha)\mu - \frac{st}{2(v-p_i)} - k$. Since $\Pi_{2.3}(p_o, p_i)$ is increasing in p_o, Case 2.3 is dominated by Case 2.1. In Case 2.4, the optimal n is either $\frac{t}{2(v-p_i)}$ or $\frac{t}{2(p_o-p_i+e)}$. With $n = \frac{t}{2(v-p_i)}$, the profit function reduces to

$$\begin{aligned}\Pi_{2.4}(p_o, p_i) = (p_o - c)\left[\alpha\mu + \beta\mu\left(\frac{v-p_o-e}{v-p_i}\right)\right]\\ + (p_i - c)\left[\beta\mu\left(\frac{p_o-p_i+e}{v-p_i}\right) + (1-\alpha-\beta)\mu\right]\\ - \frac{st}{2(v-p_i)} - k. \qquad (7.13)\end{aligned}$$

With $n = \frac{t}{2(p_o-p_i+e)}$, the profit function reduces to $\Pi_{2.4}(p_o, p_i) = (p_o - c)\alpha\mu + (p_i - c)(1-\alpha)\mu - \frac{st}{2(p_o-p_i+e)} - k$ which is increasing in p_o, so with $n = \frac{t}{2(p_o-p_i+e)}$, Case 2.4 is dominated by either Case 2.2 or Case 2.3. Thus, the optimal strategy can only be supported by $n^\dagger(p_o, p_i) = \frac{t}{2(v-p_i)}$ in Case 2.4. Finally, in Case 2.5, $n^\dagger(p_o, p_i) = \frac{t}{2(v-p_i)}$, and the profit function reduces to $\Pi_{2.5}(p_o, p_i) = (p_o - c) \cdot (\alpha+\beta)\mu + (p_i - c)(1-\alpha-\beta)\mu - \frac{st}{2(v-p_i)} - k$. Since $\Pi_{2.5}(p_o, p_i)$ is increasing in p_o, Case 2.5 is dominated by either Case 2.3 or Case 2.4.

So far, we have seen that Cases 2.3 and 2.5 are not optimal. Next, we optimize p_o by considering $\Pi_{2.1}(p_o, p_i)$, $\Pi_{2.2}(p_o, p_i)$, and $\Pi_{2.4}(p_o, p_i)$. First, consider Case 2.4. Taking derivative of (7.13) with respect to p_o yields $\frac{\partial \Pi_{2.4}}{\partial p_o} = \alpha\mu + \beta\mu\left(\frac{v-2p_o+p_i-e}{v-p_i}\right)$ which is decreasing in p_o, so $\Pi_{2.4}(p_o, p_i)$ is concave in p_o.

For $p_i \le v - e$, Case 2.4 intersects with Case 2.2 at $p_o = p_i$. Then, since $\frac{\partial_- \Pi_{2.4}}{\partial p_o}\big|_{p_o=p_i} = \alpha\mu + \beta\mu\big(\frac{v-p_i-e}{v-p_i}\big) > 0$, Case 2.4 is dominated by Case 2.2. For $p_i > v - e$, Case 2.4 intersects with Case 2.3 at $p_o = v - e$. Then, since $\frac{\partial_- \Pi_{2.4}}{\partial p_o}\big|_{p_o=v-e} = \alpha\mu + \beta\mu\big(\frac{-v+e+p_i}{v-p_i}\big) > 0$, Case 2.4 is dominated by Case 2.3. Thus, Case 2.4 is not optimal.

Second, consider Case 2.2. Taking derivative of (7.12) with respect to p_o yields $\frac{\partial \Pi_{2.2}}{\partial p_o} = \alpha\mu\big(\frac{v-2p_o+p_i}{v-p_i}\big) + \beta\mu\big(\frac{v-2p_o+p_i-e}{v-p_i}\big)$ which is decreasing in p_o, so $\Pi_{2.2}(p_o, p_i)$ is concave in p_o. Solving the first-order condition yields $p_o = \frac{v+p_i}{2} - \frac{\beta e}{2(\alpha+\beta)}$. Recall that Case 2.2 requires $p_i \le p_o \le v - e$. Also, note that Case 2.2 is valid only for $p_i \le v - e$. Then, we have $\frac{v+p_i}{2} - \frac{\beta e}{2(\alpha+\beta)} > \frac{v+p_i}{2} - \frac{e}{2} = \frac{v-e-p_i}{2} + p_i \ge p_i$. Moreover, $\frac{v+p_i}{2} - \frac{\beta e}{2(\alpha+\beta)} < v - e$ is equivalent to $p_i < v-e-\frac{\alpha e}{\alpha+\beta}$. Thus, for $p_i \ge v-e-\frac{\alpha e}{\alpha+\beta}$, $p_o^\dagger(p_i) = v-e$. For $p_i < v-e-\frac{\alpha e}{\alpha+\beta}$, $p_o^\dagger(p_i) = \frac{v+p_i}{2} - \frac{\beta e}{2(\alpha+\beta)}$, and the profit function reduces to

$$\Pi_{2.2}(p_i) = (p_i - c)\mu + \frac{(\alpha+\beta)\mu(v-p_i)}{4} - \frac{\beta\mu e}{2} + \left[\frac{\beta^2\mu e^2}{4(\alpha+\beta)} - \frac{st}{2}\right]\frac{1}{v-p_i} - k. \tag{7.14}$$

Third, consider Case 2.1. Taking derivative of (7.11) with respect to p_o yields $\frac{\partial \Pi_{2.1}}{\partial p_o} = \alpha\mu\big(\frac{v-2p_o+p_i}{v-p_i}\big)$ which is decreasing in p_o, so $\Pi_{2.1}(p_o, p_i)$ is concave in p_o. Solving the first-order condition yields $p_o = \frac{v+p_i}{2}$. Recall that Case 2.1 requires $p_o > v - e$. $\frac{v+p_i}{2} > v - e$ is equivalent to $p_i > v - 2e$. Thus, for $p_i \le v - 2e$, $p_o^\dagger(p_i) = v - e$. For $p_i > v - 2e$, $p_o^\dagger(p_i) = \frac{v+p_i}{2}$, and the profit function reduces to

$$\Pi_{2.1}(p_i) = (p_i - c)\mu + \frac{\alpha\mu(v-p_i)}{4} - \frac{st}{2(v-p_i)} - k. \tag{7.15}$$

In optimizing p_o, we know that the optimal strategy can only be achieved in Case 2.1 or Case 2.2. Moreover, for $p_i \le v - 2e$, Case 2.1 is dominated by Case 2.2. For $p_i \ge v - e - \frac{\alpha e}{\alpha+\beta}$, Case 2.2 is dominated by Case 2.1. Note that $v - 2e < v - e - \frac{\alpha e}{\alpha+\beta}$.

Next, we optimize p_i and characterize the optimal strategy. First, consider Case 2.1. Taking derivative of (7.15) with respect to p_i yields $\frac{\mathrm{d}\Pi_{2.1}}{\mathrm{d}p_i} = \left(1 - \frac{\alpha}{4}\right)\mu - \frac{st}{2(v-p_i)^2}$. It is decreasing in p_i, so $\Pi_{2.1}(p_i)$ is concave in p_i. If $\frac{\mathrm{d}\Pi_{2.1}}{\mathrm{d}p_i}\big|_{p_i=v-2e} = \left(1 - \frac{\alpha}{4}\right)\mu - \frac{st}{8e^2} > 0$, or equivalently, $e > \sqrt{\frac{st}{8(1-\alpha/4)\mu}} =: e_1$, the optimal p_i in Case 2.1 is given by the first-order condition. Solving the first-order condition yields $p_i^\dagger = v - \sqrt{\frac{st}{2(1-\alpha/4)\mu}}$. Correspondingly,

$$\Pi_{2.1}^\dagger = (v-c)\mu - \sqrt{2\left(1 - \frac{\alpha}{4}\right)st\mu} - k. \tag{7.16}$$

On the other hand, if $e \le e_1$, Case 2.1 is dominated by Case 2.2.

Second, consider Case 2.2. Taking derivative of (7.14) with respect to p_i yields

$$\frac{\mathrm{d}\Pi_{2.2}}{\mathrm{d}p_i} = \left(1 - \frac{\alpha+\beta}{4}\right)\mu + \left[\frac{\beta^2\mu e^2}{4(\alpha+\beta)} - \frac{st}{2}\right]\frac{1}{(v-p_i)^2}.$$

If $\frac{\beta^2\mu e^2}{4(\alpha+\beta)} - \frac{st}{2} > 0$, $\frac{\mathrm{d}\Pi_{2.2}}{\mathrm{d}p_i} > 0$, so Case 2.2 is dominated by Case 2.1. If $\frac{\beta^2\mu e^2}{4(\alpha+\beta)} - \frac{st}{2} \leq 0$, $\Pi_{2.2}(p_i)$ is concave in p_i. Then, if

$$\left.\frac{\mathrm{d}\Pi_{2.2}}{\mathrm{d}p_i}\right|_{p_i=v-e-\alpha e/(\alpha+\beta)} = \left(1 - \frac{\alpha+\beta}{4}\right)\mu + \left[\frac{\beta^2\mu e^2}{4(\alpha+\beta)} - \frac{st}{2}\right]\left(\frac{\alpha+\beta}{2\alpha+\beta}\right)^2\frac{1}{e^2} < 0, \tag{7.17}$$

the optimal p_i in Case 2.2 is given by the first-order condition. Solving the first-order condition yields $p_i^\dagger = v - \sqrt{\left[\frac{st}{2} - \frac{\beta^2\mu e^2}{4(\alpha+\beta)}\right]/\left[\left(1-\frac{\alpha+\beta}{4}\right)\mu\right]}$. Correspondingly,

$$\Pi_{2.2}^\dagger = (v-c)\mu - \sqrt{\left(1 - \frac{\alpha+\beta}{4}\right)\left(2st - \frac{\beta^2\mu e^2}{\alpha+\beta}\right)\mu} - \frac{\beta\mu e}{2} - k. \tag{7.18}$$

Note that (7.17) is equivalent to $e < (\alpha+\beta)\sqrt{\frac{2st}{[(4-\alpha-\beta)(2\alpha+\beta)^2+\beta^2(\alpha+\beta)]\mu}} =: e_2$. Thus, if $e < e_2$, $p_i^\dagger$ and $\Pi_{2.2}^\dagger$ are given above. On the other hand, if $v \geq e_2$, Case 2.2 is dominated by Case 2.1.

We have shown that if $e_1 < e < e_2$, the optimal p_i is given by the first-order condition in both Cases 2.1 and 2.2, and the profits are given by (7.16) and (7.18), respectively. From (7.16) and (7.18), we know $\Pi_{2.1}^\dagger \geq \Pi_{2.2}^\dagger$ is equivalent to

$$\sqrt{2\left(1 - \frac{\alpha}{4}\right)st\mu} - \frac{\beta\mu e}{2} \leq \sqrt{\left(1 - \frac{\alpha+\beta}{4}\right)\left(2st - \frac{\beta^2\mu e^2}{\alpha+\beta}\right)\mu}. \tag{7.19}$$

Note that $e < e_2$ implies $\frac{\beta^2\mu e^2}{4(\alpha+\beta)} - \frac{st}{2} \leq 0$ which in turn implies $\sqrt{2(1-\alpha/4)st\mu} - \frac{\beta\mu e}{2} > 0$. Taking square on both sides of (7.19) and simplifying the resulting inequality yields

$$\frac{\beta\mu}{\alpha+\beta}\cdot e^2 - \sqrt{2\left(1-\frac{\alpha}{4}\right)st\mu}\cdot e + \frac{st}{2} \leq 0. \tag{7.20}$$

Solving (7.20) yields $e_3 \leq e \leq e_4$, where

$$e_3 = \frac{\sqrt{st}}{\sqrt{2\left(1-\frac{\alpha}{4}\right)\mu} + \sqrt{2\left(1-\frac{\alpha}{4}\right)\mu - \frac{2\beta\mu}{\alpha+\beta}}}, \quad \text{and}$$

$$e_4 = \frac{\sqrt{st}}{\sqrt{2\left(1-\frac{\alpha}{4}\right)\mu} - \sqrt{2\left(1-\frac{\alpha}{4}\right)\mu - \frac{2\beta\mu}{\alpha+\beta}}}.$$

To determine when Case 2.1 or Case 2.2 is optimal, we show that the following three conditions hold: (1) $e_3 < e_2$, (2) $e_3 > e_1$, (3) $e_4 > e_2$. Given these conditions, Case 2.1 is optimal if and only if $e \geq e_3 = \bar{e}$, and Case 2.2 is optimal if and only if $e < e_3 = \bar{e}$. First, $e_3 < e_2$ is equivalent to

$$\sqrt{(4-\alpha-\beta)(2\alpha+\beta)^2+\beta^2(\alpha+\beta)}$$
$$< \sqrt{\alpha+\beta}\left[\sqrt{(\alpha+\beta)(4-\alpha)} + \sqrt{\alpha(4-\alpha-\beta)}\right]. \tag{7.21}$$

Taking square on both sides of (7.21) and rearranging terms yields

$$\alpha\left[1+2\alpha-(1-\alpha-\beta)^2\right] < (\alpha+\beta)\sqrt{(\alpha+\beta)(4-\alpha)\alpha(4-\alpha-\beta)}. \tag{7.22}$$

Then, taking square on both sides of (7.22) and rearranging terms yields $4\alpha\beta\cdot[-2\alpha^2\cdot(1-\alpha)-(5\alpha^2+4\alpha\beta+\beta^2)(1-\beta)-\alpha^2-7\alpha\beta-3\beta^2] < 0$ which is true. Second, $e_3 > e_1$ holds because $e_3 > \frac{st}{2\sqrt{2(1-\alpha/4)st\mu}} = e_1$. Third, to show $e_4 > e_2$, it suffices to show $\frac{\alpha+\beta}{2\beta\mu}\sqrt{2\left(1-\frac{\alpha}{4}\right)st\mu} > (\alpha+\beta)\sqrt{\frac{2st}{[(4-\alpha-\beta)(2\alpha+\beta)^2+\beta^2(\alpha+\beta)]\mu}}$, which can be simplified to

$$\sqrt{4-\alpha}\cdot\sqrt{(4-\alpha-\beta)(2\alpha+\beta)^2+\beta^2(\alpha+\beta)} > 4\beta. \tag{7.23}$$

Taking square on both sides of (7.23) and rearranging terms yields $4\alpha[\alpha(\alpha+\beta)^2+7\alpha^2+(8\alpha+11\beta)(1-\alpha)+(\alpha+5\beta)(1-\beta)] > 0$ which is true.

Therefore, we have shown that Case 2.1 is optimal if $e \geq \bar{e}$ and Case 2.2 is optimal if $e < \bar{e}$. When Case 2.1 is optimal, tracing back our analysis for Case 2.1, we obtain that $p_o^\dagger = \frac{v+p_i^\dagger}{2}$ and $n^\dagger = \frac{t}{2(v-p_i^\dagger)}$ which are the ones shown in Proposition 7.2(i). It is easy to see that $p_o^\dagger > p_i^\dagger$. Under the optimal strategy, Type I customers purchase online if $\frac{p_o^\dagger - p_i^\dagger}{t} \leq x \leq \frac{1}{2n^\dagger}$ and purchase in-store if $0 \leq x < \frac{p_o^\dagger - p_i^\dagger}{t}$, all Types II and III customers purchase in-store.

Moreover, when Case 2.2 is optimal, tracing back our analysis for Case 2.2, we obtain that $p_o^\dagger = \frac{v+p_i^\dagger}{2} - \frac{\beta e}{2(\alpha+\beta)}$ and $n^\dagger = \frac{t}{2(v-p_i^\dagger)}$ which are the ones shown in Proposition 7.2(ii). $p_o^\dagger > p_i^\dagger$ can be simplified to

$$e < \frac{\alpha+\beta}{\beta}\sqrt{\frac{st}{2\mu}}. \tag{7.24}$$

To show (7.24) is true if $e < \bar{e}$, it suffices to show $\bar{e} < \frac{\alpha+\beta}{\beta}\sqrt{\frac{st}{2\mu}}$. Then, since $\bar{e} < \sqrt{\frac{st}{2(1-\alpha/4)\mu}}$, it suffices to show

$$\sqrt{\frac{st}{2\left(1-\frac{\alpha}{4}\right)\mu}} < \frac{\alpha+\beta}{\beta}\sqrt{\frac{st}{2\mu}}. \tag{7.25}$$

Taking square on both sides of (7.25) and simplifying the resulting inequality yields $\alpha[-(\alpha+\beta)^2+4\alpha+8\beta] > 0$ which is true. Thus, $p_o^\dagger > p_i^\dagger$. Finally, under the optimal strategy, Type I customers purchase online if $\frac{p_o^\dagger - p_i^\dagger}{t} \le x \le \frac{1}{2n^\dagger}$ and purchase in-store if $0 \le x < \frac{p_o^\dagger - p_i^\dagger}{t}$, Type II customers purchase online if $\frac{p_o^\dagger - p_i^\dagger + e}{t} \le x \le \frac{1}{2n^\dagger}$ and purchase in-store if $0 \le x < \frac{p_o^\dagger - p_i^\dagger + e}{t}$, and all Type III customers purchase in-store. The proof is complete. □

Proof of Proposition 7.3

(i) It is easy to see $n^\dagger < n^*$ if $e \ge \bar{e}$. If $e < \bar{e}$, $n^\dagger < n^*$ can be simplified to

$$e < \frac{\alpha+\beta}{\beta}\sqrt{\frac{st}{2\mu}} \tag{7.26}$$

which is same as (7.24). Thus, $n^\dagger < n^*$.

(ii) It is easy to see $p_o^\dagger > p_o^*$ if $e \ge \bar{e}$. If $e < \bar{e}$, $p_o^\dagger > p_o^*$ is equivalent to

$$(\alpha+\beta)\sqrt{\frac{st\mu}{2}} - \frac{(\alpha+\beta)\mu}{2}\sqrt{\frac{\frac{st}{2} - \frac{\beta^2\mu e^2}{4(\alpha+\beta)}}{\left(1-\frac{\alpha+\beta}{4}\right)\mu}} > \frac{\beta\mu e}{2}. \tag{7.27}$$

From (7.19), we know that $e < \bar{e}(= e_3)$ implies

$$\sqrt{2\left(1-\frac{\alpha}{4}\right)st\mu} - \sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\left(2st - \frac{\beta^2\mu e^2}{\alpha+\beta}\right)\mu} > \frac{\beta\mu e}{2}. \tag{7.28}$$

For (7.28) to sufficiently imply (7.27), we need the left-hand side of (7.27) to be larger than the left-hand side of (7.28), which can be simplified to the following:

$$\left(\sqrt{4-\alpha}-\alpha-\beta\right)\sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\frac{st}{2}} < (2-\alpha-\beta)\sqrt{\frac{st}{2}-\frac{\beta^2\mu e^2}{4(\alpha+\beta)}}. \tag{7.29}$$

First, it is easy to see $\sqrt{4-\alpha}-\alpha-\beta < 2-\alpha-\beta$. Second, (7.26) implies $\sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\frac{st}{2}} < \sqrt{\frac{st}{2}-\frac{\beta^2\mu e^2}{4(\alpha+\beta)}}$. Thus, $p_o^\dagger > p_o^*$.

(iii) It is easy to see $p_i^\dagger < p_i^*$ if $e \geq \bar{e}$. If $e < \bar{e}$, $p_i^\dagger < p_i^*$ can be simplified to (7.26) which we have shown is true. Thus, $p_i^\dagger < p_i^*$.

(iv) Since the firm sells to all customers under the optimal strategy in both cases, it suffices to consider the online demand. It is easy to see that if $e \geq \bar{e}$, the firm's online demand is lower in Case 2 than in Case 1. If $e < \bar{e}$, the condition for the online demand to be lower in Case 2 than in Case 1 is $\alpha > \frac{\alpha+\beta}{2} - \beta e\sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\mu/\left[2st-\frac{\beta^2\mu e^2}{\alpha+\beta}\right]}$ which is equivalent to

$$(\beta-\alpha)\sqrt{\frac{st}{2}-\frac{\beta^2\mu e^2}{4(\alpha+\beta)}} < \beta e\sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\mu}. \tag{7.30}$$

If $\beta < \alpha$, (7.30) holds trivially. If $\beta \geq \alpha$, taking square on both sides of (7.30) and simplifying the resulting inequality yields

$$e > \frac{\beta-\alpha}{\beta}\sqrt{\frac{(\alpha+\beta)st}{2(\alpha+\beta-\alpha\beta)\mu}}. \tag{7.31}$$

Note that if $\beta < \alpha$, the right-hand side of (7.31) is negative and hence (7.31) holds. Thus, (7.30) holds if and only if $e > \underline{e}$. Finally, since (7.30) holds when $e = \bar{e}$, we know $\underline{e} < \bar{e}$.

(v) By comparing Π^* and $\Pi^\dagger$, it is easy to see that $\Pi^\dagger \geq \Pi^*$ if and only if $k \leq \bar{k}$. If $e \geq \bar{e}$, it is easy to see that $\bar{k} > 0$. If $e < \bar{e}$, from (7.19) we know that $\sqrt{2\left(1-\frac{\alpha}{4}\right)st\mu} - \sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\left(2st-\frac{\beta^2\mu e^2}{\alpha+\beta}\right)\mu} - \frac{\beta\mu e}{2} > 0$, which implies $\bar{k} > 0$.

(vi) If $e \geq \bar{e}$, $\Pi^\dagger - \Pi^* = \sqrt{2st\mu}\left(1-\sqrt{1-\frac{\alpha}{4}}\right) - k$ which is increasing in t and s. If $e < \bar{e}$, $\Pi^\dagger - \Pi^* = \sqrt{2st\mu} - \sqrt{\left(1-\frac{\alpha+\beta}{4}\right)\left(2st-\frac{\beta^2\mu e^2}{\alpha+\beta}\right)\mu} - \frac{\beta\mu e}{2} - e$. Then,

$$\frac{\partial(\Pi^\dagger - \Pi^*)}{\partial t} = \frac{s}{t}\left[\sqrt{\frac{\mu t}{2s}} - \sqrt{\frac{(1-(\alpha+\beta)/4)\mu t}{2s-\beta^2\mu e^2/[(\alpha+\beta)t]}}\right] = \frac{s}{t}(n^* - n^\dagger) > 0.$$

Moreover, since $\Pi^\dagger - \Pi^*$ is symmetric in t and s, $\frac{\partial(\Pi^\dagger - \Pi^*)}{\partial s} > 0$. □

References

Alptekinoğlu A, Corbett CJ (2008) Mass customization vs. mass production: variety and price competition. Manuf Serv Oper Manag 10(2):204–217

Alptekinoğlu A, Corbett CJ (2010) Leadtime-variety tradeoff in product differentiation. Manuf Serv Oper Manag 12(4):569–582

Bart Y, Shankar V, Sultan F, Urban GL (2005) Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study. J Mark 69(4):133–152

Chen L, Cui Y, Lee HL (2018) Retailing with 3D printing. Working paper. Cornell University, Ithaca

Çil EB, Pangburn M (2017) Mass customization and guardrails: you can't be all things to all people. Prod Oper Manag 26(9):1728–1745

De Groote X (1994) Flexibility and marketing/manufacturing coordination. Int J Prod Econ 36(2):153–167

Desai P, Kekre S, Radhakrishnan S, Srinivasan K (2001) Product differentiation and commonality in design: balancing revenue and cost drivers. Manag Sci 47(1):37–51

Dewan R, Jing B, Seidmann A (2003) Product customization and price competition on the internet. Manag Sci 49(8):1055–1070

Dobson G, Yano CA (2002) Product offering, pricing, and make-to-stock/make-to-order decisions with shared capacity. Prod Oper Manag 11(3):293–312

Dong L, Shi D, Zhang F (2017) 3D printing vs. traditional flexible technology: implications for operations strategy. Working paper. Washington University, St. Louis

Economist (2017) 3D printers start to build factories of the future. The Economist (June 29). https://www.economist.com/news/briefing/21724368-recent-advances-make-3d-printing-powerful-competitor-conventional-mass-production-3d. Accessed 17 Jul 2017

Forman C, Ghose A, Goldfarb A (2009) Competition between local and electronic markets: how the benefit of buying online depends on where you live. Manag Sci 55(1):47–57

Gao F, Su X (2017a) Omnichannel retail operations with buy-online-and-pick-up-in-store. Manag Sci 63(8):2478–2492

Gao F, Su X (2017b) Online and offline information for omnichannel retailing. Manuf Serv Oper Manag 19(1):84–98

Ho TH, Tang CS (1998) Product variety management: research advances. Kluwer, Springer, New York

Hotelling H (1929) Stability in competition. Econom J 39:41–57

Jiang K, Lee HL, Seifert RW (2006) Satisfying customer preferences via mass customization and mass production. IIE Trans 38(1):25–38

Kim K, Chhajed D, Liu Y (2013) Can commonality relieve cannibalization in product line design? Mark Sci 32(3):510–521

Konuş U, Verhoef PC, Neslin SA (2008) Multichannel shopper segments and their covariates. J Retail 84(4):398–413

Lancaster K (1990) The economics of product variety: a survey. Mark Sci 9(3):189–206

Lee HL (1996) Effective inventory and service management through product and process redesign. Oper Res 44(1):151–159

Lee HL (2007) Peering through a glass darkly. Int Commer Rev 7(1):60–68

Lee HL, Tang CS (1997) Modelling the costs and benefits of delayed product differentiation. Manag Sci 43(1):40–53

Liu Y, Cui TH (2010) The length of product line in distribution channels. Mark Sci 29(3):474–482

Massie C (2013) Walmart wants to add 3D printing to its retail empire. Architect (November 13). https://www.architectmagazine.com/technology/walmart-wants-to-add-3d-printing-to-its-retail-empire. Accessed 28 Sep 2018

Mendelson H, Parlaktürk AK (2008a) Competitive customization. Manuf Serv Oper Manag 10(3):377–390

Mendelson H, Parlaktürk AK (2008b) Product-line competition: customization vs. proliferation. Manag Sci 54(12):2039–2053

Ramdas K (2003) Managing product variety: an integrative review and research directions. Prod Oper Manag 12(1):79–101

Riordan MH (1986) Monopolistic competition with experience goods. Q J Econom 101(2):265–279

Salop SC (1979) Monopolistic competition with outside goods. Bell J Econom 10(1):141–156

Song JS, Zhang Y (2016) Stock or print? Impact of 3D printing on spare parts logistics. Working paper. Duke University, Durham

Syam NB, Kumar N (2006) On customized goods, standard goods, and competition. Mark Sci 25(5):525–537

Syam NB, Ruan R, Hess JD (2005) Customized products: a competitive analysis. Mark Sci 24(4):569–584

Wingfield N, Couturier K (2017) Detailing Amazon's custom-clothing patent. New York Times (April 30). https://www.nytimes.com/2017/04/30/technology/detailing-amazons-custom-clothing-patent.html. Accessed 6 Feb 2018

Xia N, Rajagopalan S (2009) Standard vs. custom products: variety, lead time, and price competition. Mark Sci 28(5):887–900

Zaczkiewicz A (2017) Amazon, Wal-Mart and Apple top list of biggest e-commerce retailers. WWD (April 7). http://wwd.com/business-news/business-features/amazon-wal-mart-apple-biggest-e-commerce-retailers-10862796. Accessed 17 Jul 2017

Chapter 8
Price-Matching Strategy: Implications of Consumer Behavior and Channel Structure

Arcan Nalca, Saibal Ray, and Tamer Boyaci

Abstract Price-matching has become a ubiquitous strategy for retailers both in product and service industries, especially with the growing ease of checking prices online. With this strategy, retailers promise not to be undersold and match competitor's lower price (if any). Price-sensitive consumers tend to be happy with this since they potentially can get the lowest price at their "favourite" retailer. A relatively under-researched topic in this context is the fact that this price convenience normally comes with a number of conditions. We analyze two of the most common ones—the product must be available at the lower priced retailer (availability condition) and the price-match extends only to a competing retailer and not to a direct-to-consumer manufacturer (channel condition)—and investigate their implications for the channel and consumers.

We show that if consumers consider the fact that they might be denied the price-matching benefit based on verification of availability at the competitor's location while making a purchase decision, then they will benefit from a lower price by increasing the competition between the retailers; on the other hand, not considering this fact would allow the retailers to price discriminate and might harm consumers.

As regards channel structure, we prove that if the upstream partner has the power to set the wholesale price, then it does not make sense for a retail channel to price match with a manufacturer selling directly to consumers. But, in sectors where there are dominant retailers who have a say in determination of the wholesale price, price-matching can be an equilibrium strategy for the retail channel, even when

A. Nalca
Smith School of Business, Queen's University, Kingston, ON, Canada
e-mail: arcan.nalca@queensu.ca

S. Ray (✉)
Bensadoun School of Retail Management, McGill University, Montreal, QC, Canada
e-mail: saibal.ray@mcgill.ca

T. Boyaci
European School of Management and Technology, Berlin, Germany
e-mail: tamer.boyaci@esmt.org

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_8

manufacturers are directly selling to consumers. Furthermore, price-matching is also the equilibrium strategy for upstream manufacturers as it redirects demand from the retail to their direct channel.

Keywords Price-matching guarantees · Product availability · Inventory · Demand uncertainty · dual channel

8.1 Introduction

Price-matching guarantees (PMGs) are offers by which retailers promise to match any lower price offered by the competition for the same merchandise. PMGs are quite popular in many sectors, such as consumer electronics (Best Buy, Circuit City, Gateway, Dell), home and office appliances (Home Depot, Staples), leisure and travel (Orbitz), and even in insurance (Bank of Ireland). Some retailers will match the price even if they do not advertise having such a policy. In fact, local stores will often match the price of direct competition upon request as a matter of standard procedure (pcguide.com). Arbatskaya et al. (2004) report on the variety of PMGs for more than 50 categories of products and services.[1,2] Many online discussion groups, forums, and blogs encourage consumers to utilize PMGs and provide guidelines to ensure that their price-match requests are fruitful (e.g., redflagdeals.com).

Offering a PMG shows the persistence of a retailer in being competitive, yet does not necessarily guarantee the "best deal" for the customers. Typically, customers are asked to present proof of a lower price (if any), and the difference is refunded. This proof can be in the form of a weekly flyer, a newspaper advertisement, or website information for an identical brand and model. As such, PMGs only provide an opportunity to receive the lowest price if customers are willing and are able to claim it.

To be able to request a price-match, customers must be aware that the retailer is offering a PMG and know the competitors' prices. These are not challenging tasks, however, given the volume of information available in the online world and social media. PMGs are storewide policies; they are not offered for particular items but for the whole assortment of products at the store. Therefore, from a customer's perspective, it is relatively easy to remember if a certain store offers a PMG (Moorthy and Zhang 2006). Moreover, considering the growing digitalization

[1] Firms sometimes refund more than the difference, so that they "beat" the lower price of the competitor. For ease of exposition, however, we only consider the case where firms match the lower price.

[2] In addition to competitive price-matching guarantees, there is also an alternative price guarantee, less related to our work, that is referred to as an internal price-matching guarantee (IPM). Under IPMs, retailers promise to refund the customer who has already purchased an item if a lower price is offered during some pre-specified time-period. A stream of research identifies that IPMs can facilitate collusion (Cooper 1986; Butz 1990) and mitigate the strategic purchasing behavior of customers (Aviv et al. 2009; Lai et al. 2007; Levin et al. 2007).

of business processes in the retail world, obtaining and comparing prices has never been easier for consumers. For example, a new price app called *Pricerazzi*, launched in October 2017, searches more than 1000 stores over the eligibility term of a given price-match policy. When a price drop is found, the app automatically notifies the customer, so they can request a price-match.

Given the very low cost of price searching, a majority of customers will potentially be informed about the prices and request a price-match when relevant. The economics and marketing literature has examined in detail the implications of PMGs between retail firms. One widely accepted outcome of PMGs between retailers, when it is easy for customers to request a price-match, is tacit collusion, i.e., the ability to maintain high prices without any formal agreement (Arbatskaya 2001; Chen 1995; Edlin 1997; Hay 1981; Logan and Lutter 1989; Salop 1986; Zhang 1995). With a PMG in place, a firm can automatically respond to any price reduction by rivals because its customers request a price-match instead of switching to the lower priced firms. As such, there is no incentive to cut prices; a firm that is matching a competitor's price cannot be undersold. The result is that firms can collectively increase their prices, making PMGs a seemingly enticing strategy.

However, a closer look reveals that many other factors beyond price information play a role in the success rate of a customer's PMG request. For instance, the following excerpt from Best Buy Canada's website lists the necessary conditions that must be met for customers to be granted a price-match. A screenshot of the website is provided in Fig. 8.1. PMGs by other retailers also come with similar clauses.

The product you're comparing must meet each of the following criteria to qualify for a price beat.

1. *Must have the exact same brand name and model number.*
2. *Must be priced in Canadian Dollars and include all environmental fees, shipping costs, and other charges.*
3. *Must be sold and shipped by a retailer AND authorized dealer located within Canada.*
4. *The product must be in stock, available for sale, and cannot be a limited time offer or available only in limited quantity.*
5. *The product's price must not be lower due to an advertising error, misprint, or special sale price.*

In this chapter, our goal is to shed light on how the two conditions listed in these criteria—which are operational and channel related—affect the value proposition of the PMGs offered by retailers. Specifically, we would like to identify how the channel coverage (third item) and the product availability clause (fourth item) in Best Buy's list affect consumers and change decisions in the supply chain.

Note that the fourth item in the list requires that the product under consideration be available for sale at the competing retailer at the time of the price-matching

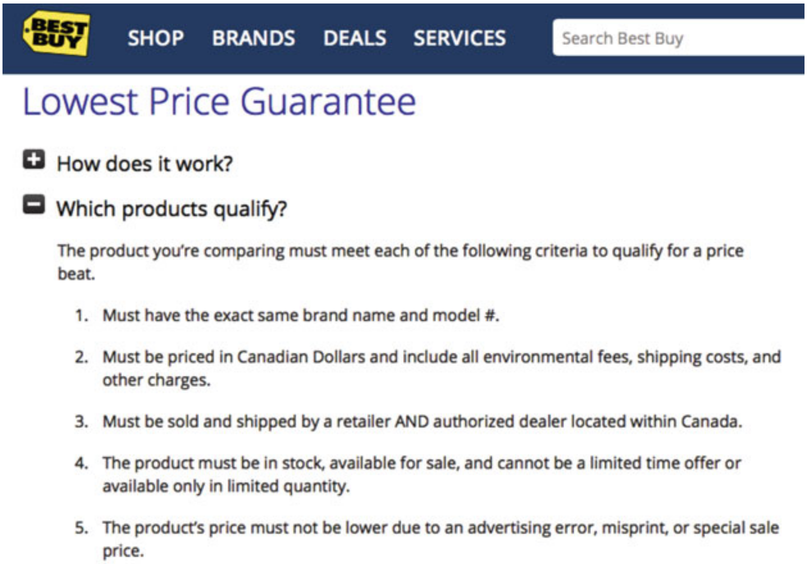

Fig. 8.1 Lowest price guarantee details from Best Buy Canada (retrieved February 20, 2018, from https://www.bestbuy.ca/en-ca/help/lowest-price-guarantee/hc1001.aspx. Screenshot by author)

request, prior to matching the price. Obviously, Best Buy, or any other retailer verifying availability before matching the price, is willing to sacrifice its profit margin only if the customer has a credible alternative to purchase at a lower price. This approach protects Best Buy against a competitor announcing a lower price for a product to attract customers, but deliberately stocking low for that particular product (such a strategy reduces the penalty of low profit margins, but the increased in-store traffic may lead to higher sales of substitute products).

Verification of availability at the competing store cuts down on the number of incidents of sales with low profit margins. At the same time, high stock-out rates may create a scarcity effect in the market, especially in the price-matching context. Consider a customer who asks for a price-match at her preferred store but is declined because the product is sold out at the competing store. She has already traveled to her preferred store and learned there that the product is sold out at the competing store. If she is eager to obtain the product, then she may agree to purchase it at the (high) list price even though she had not been willing to pay that price before learning about the stock-out at the competing store.

The existence of an availability condition reveals the important role of product availability in PMGs as the process of verifying allows retailers to take advantage of the scarcity effect. Given that 100% of the top 20 consumer electronics retailers offering PMGs consider the verification of availability at the competing retailer to be a prerequisite for matching prices, the value that retailers attach to availability is obvious.

Another important element in the list of Best Buy's price-matching conditions is that the product under consideration must be sold by a *retailer*. Dual-channel systems, i.e., those in which manufacturers or suppliers sell a product directly through an integrated channel in addition to traditional retail channels are ubiquitous in today's economy.[3] An interesting property of dual channels is that the supplier becomes the competitor of the retailer, in addition to being a vendor. Moreover, the two channels compete primarily on price to attract customers, in addition to service and quality (Chen et al. 2008).

Best Buy's price-matching policy includes other *resellers* of the product, but excludes manufacturers and suppliers, meaning that their price-matching offer does not extend to the dual channel. Nevertheless, PMGs are sometimes observed in dual channels too. For instance, since February 2013, Air Canada has been promising customers who book tickets on its website that they will be getting the lowest possible fare.[4] A screenshot of the website is provided in Fig. 8.2. If customers find a lower price for the exact same Air Canada itinerary on a travel provider's website (e.g., expedia.ca) or through a travel agency, then the difference will be refunded. Expedia Inc. also promises to match any lower price as a part of their "stress-free travel" program.[5]

In fact, price-matching between travel agencies and suppliers such as hotels and airlines is widely observed. Airlines such as Air Canada, Air France, American Airlines, British Airways, Continental, Delta, United, and US Airways, and hotels such as Choice Hotels, Hilton, Hyatt, Intercontinental Hotels Group, and Marriott offer guarantees to match the price of online-travel-agencies (OTA), such as Expedia, Travelocity, Priceline, Orbitz, and Hotels.com. These OTAs also promise to match the price of airlines, hotels, and competing OTAs.

Despite their popularity in some industries, PMGs are not prevalent in all dual-channel settings. For instance, to the best of our knowledge, in the consumer packaged goods industry there are no incidents of price-matching between a retailer and a supplier; suppliers do not offer a guarantee while retailers do, but they exclude the suppliers from their coverage. We have checked the price-matching policies of big retailers such as WalMart, Target, Staples, Loblaws, Toys R Us, Home-Depot, Home Hardware, Rona, The Brick, Lowe's, and Canadian Tire in the consumer packaged goods, and Best Buy, Radio Shack, The Source, and Fry's in consumer electronics, among many others, and have observed that the guarantee is applicable among retailers only.

Availability verification has both direct and indirect effects on consumers. On the one hand, it can be seen as an extra hassle that consumers need to go through in

[3] In order to keep the model applicable to service industries, we henceforth use the term "supplier" when referring to the vendor even though the term usually refers to a *manufacturing* firm in consumer packaged goods and consumer electronics.

[4] See, for example, "Air Canada guarantees its website has lowest airfares" by Vanessa Lu, *Toronto Star*, February 12, 2013.

[5] See the details of "What is the Best Price Guarantee?" at www.expedia.ca.

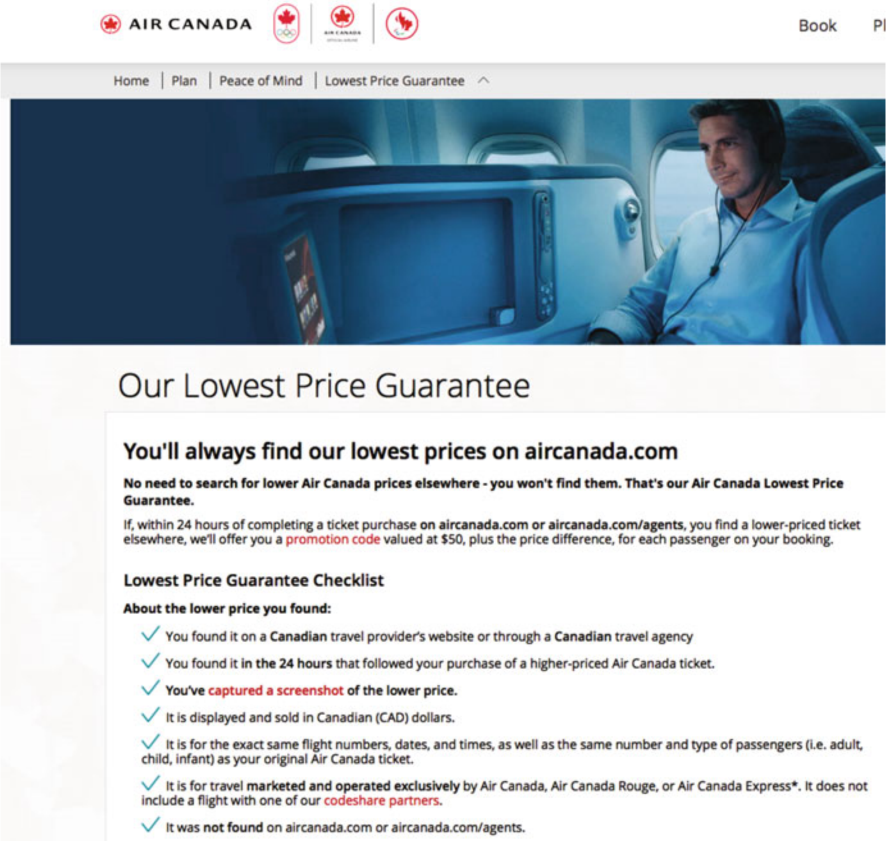

Fig. 8.2 Lowest price guarantee details from Air Canada (retrieved February 20, 2018, from https://www.aircanada.com/ca/en/aco/home/plan/peace-of-mind/lowest-price-guarantee.html. Screenshot by author)

order to request a price-match. Similarly, limiting the coverage of the PMG to other retailers, and excluding the direct channels, also limits the use of PMGs from the consumer's perspective.

On the other hand, availability verification as a price-matching condition will play a role in the profitability of retailers, and therefore will have an impact on their pricing and inventory decisions. Similarly, the presence of a PMG between a retailer and a manufacturer will have an impact on their channel interactions. As such, these conditions (or limitations) for price-matching will also have indirect repercussions on consumers by changing the nature of the interaction between supply chain members.

Our goal in this chapter is to identify how price-matching conditions—such as verification of availability and the extent of channel coverage (traditional versus direct channels)—change the value proposition of price-matching guarantees.

Specifically, we shed light on the following two specific questions: (1) How does availability verification as a condition of price-match affect the value of the guarantee for consumers and retailers? (2) What market conditions produce a wider price-matching coverage in dual channels, and does wider coverage necessarily benefit consumers?

To address these issues, we analyze a duopoly game in a single-period setting. Our modeling framework incorporates the horizontal and vertical differentiation between firms, as well as the product availability at each firm.

Our analysis reveals that, despite being perceived as a hassle by consumers, availability verification as a condition for price-matching benefits consumers as long as they take the availability verification into consideration while making a purchase. Specifically, if consumers take availability verification into account while deciding which retailer to purchase from, then the availability clause increases the degree of price competition between the retailers and, thereby, leads to lower market prices. In contrast, if the consumers neglect the availability verification clause, then the clause allows retailers to price discriminate against consumers.

Our analysis also reveals that the use of PMGs in a dual channel may lead to an increase in market prices. However, the manufacturer's PMG may benefit both consumers and the manufacturer in the presence of a dominant retailer.

In terms of channel coverage, we show that offering a PMG is not beneficial and it is not the equilibrium strategy in a *traditional chain* in which the supplier sets the wholesale price. This is true even under conditions at which competing retail channels (non-dual channel) would benefit from price-matching. The main intuition is that promising to match the price of the retail channel limits the supplier's ability to optimally allocate the demand between the two channels. However, a shift in pricing power toward a dominant retailer—that is, the retailer that dictates the wholesale price—provides an explanation as to why PMGs are observed in dual channels.

Specifically, in a retail dominant chain, the supplier has limited influence on both the pricing of the retail channel and the demand distribution across the two channels. Consequently, the supplier benefits from attracting customers to the direct channel via a PMG. We show that, in a retail dominant chain, all possible outcomes (both parties offer price-matching guarantee, neither offers, only the retailer offers, and only the supplier offers) are possible at equilibrium, depending on the degree of vertical and horizontal differentiation between the retail and direct channels. We also show that, under the condition of both parties offering price-matching, the consumer surplus is reduced compared to the no-PMG scenario. In that sense, it is possible that retail dominance is the underlying reason that retailers and suppliers/manufacturers include each other in their price-matching coverage and this is detrimental to consumer welfare. In contrast, retailers' exclusion of suppliers or manufacturers from their price-matching guarantee coverage—as bothersome as they may be for consumer experience—maintains a higher surplus through the elimination of tacit collusion.

8.2 Review of the Related Literature

There are three streams of research that relate to our work: price-matching guarantees, dual channels, and retail dominance.

8.2.1 Price-Matching Guarantees

A substantial body of economics and marketing research analyzes how PMGs affect retail price competition. The extant literature presents four perspectives on the economical outcome that PMGs postulate: (1) tacit collusion, (2) competition enhancing, (3) price discrimination, and (4) signal theory.

Tacit Collusion Following the seminal works of Hay (1981) and Salop (1986), the first stream illustrates that, despite seemingly acting as a pro-competitive device, price-matching policies actually lead to tacit collusion. Early works in this field deal with identical retailers and assume that price search activity is costless (Edlin 1997; Sargent 1992). Consequently, all customers are informed about the prices and purchase from the lowest priced retailer in the market (Corts 1995; Doyle 1988). This behavior motivates retailers to engage in price competition coupled with a PMG offer so as not to lose customers to competition. Any incentive to undercut the price is eliminated since a retailer that is matching the price of the competitor cannot be undersold, provided that customers are informed about prices. When all retailers offer PMGs, there is no incentive to undercut the prices of rivals; reduced competition then results in high prices, indeed monopoly ones.

The robustness of the tacit collusion argument has been tested under various settings in the following works.

Belton (1987) shows that retailers continue to offer PMGs and set monopoly prices even when they make their decisions sequentially rather than simultaneously.

Zhang (1995) shows that if retailers are allowed to choose their product location (in a Hotelling setting), price-matching guarantee, and prices sequentially and independently, then collusive outcomes are achieved. Interestingly, even if the price competition is deteriorated by the PMGs, the competition on product differentiation is still intense and there is minimal differentiation among the products in equilibrium.

Lin (1988) suppresses the perfect information assumption allowing customers to engage in a search with an increasing marginal cost. In other words, the information level of customers depends on their search level. Under this scenario, PMGs not only allow retailers to act like a monopoly, but to also increase the amount of search by customers.

Logan and Lutter (1989) demonstrate that in case of large cost differences between retailers, the high-cost retailer loses interest in PMGs, and competitive equilibrium prices prevail. Mago and Pate (2009) empirically validate this result.

Arbatskaya et al. (2004) provide empirical evidence of high hassle costs in price-matching policies that come from the restrictions presented in the guarantee itself. The addition of hassle costs can have significant effects on the equilibrium pricing strategy. Hviid and Shaffer (1999) show that hassle costs are sufficient to make any price above marginal cost unsupportable if retailers sell identical products. The reason is that, for any positive hassle cost, customers would strictly prefer the retailer with the lowest price, independent of the PMG. They conclude that if hassle costs of customers are included, any increase in equilibrium prices due to price-matching will be small.

Although experimental studies provide support for collusive behavior in the presence of PMGs (Dugar 2005; Fatas and Mañez 2007), empirical studies are inconclusive. The challenge is the difficulty in estimating what each retailer would be doing if PMGs were not allowed (Arbatskaya et al. 2004).

Hess and Gerstner (1991) are the first to empirically study the effects of PMGs. They study the price decisions of a supermarket before and after adopting a price-matching policy, and find that the average product price increases (slightly) after the integration of the PMG. The follow-up empirical studies conclude that there is no significant difference between the prices of the retailers offering PMGs as compared to prices of the retailers not offering PMGs (Arbatskaya et al. 1999, 2006; Manez 2006).

Recent works, such as Lu and Wright (2010) and Liu (2012) extend the analysis to consider the dynamic implications of PMGs and show the robustness of the collusion argument.

Facilitating Competition By industry observers, PMGs are often interpreted as the initiation of a price war. Despite the general anti-competitive view established by the tacit collusion stream, prior research has shown that price-matching guarantees can indeed facilitate competition; the expected prices and profits of competing retailers can be strictly lower when all stores adopt PMGs, than when they are not allowed to. The basic factors shown to result in non-monopoly prices are customer heterogeneity, retail asymmetry, and the hassle cost of requesting a price-match.

Coughlan and Shaffer (2003) consider two retailers selling multiple products and re-examine the validity of tacit collusion outcome with respect to shelf space constraints. They show that when retailers have unlimited shelf space, they carry all of the products in the market, offer price-matching, and push prices up to monopoly. In the presence of limited shelf space, retailers carry non-overlapping product lines and the equilibrium prices may be lower than the competitive levels, depending on the substitutability of products. If the products are strong substitutes for each other, then retailers prefer overlapping product lines, offer PMGs, and set monopoly prices.

Chen et al. (2001) show that the adoption of PMGs can generate a competition-enhancing effect through customers who prefer to shop at a particular store but are mindful of price-saving opportunities. They characterize the customers by two attributes: store loyalty and price search cost. Store loyalty measures their preferences for shopping at a particular store. The analysis considers four customer segments: switchers, bargain shoppers, opportunistic loyals and loyals. Switchers have a low store loyalty and search by cost, and therefore, they only shop at the lowest priced store. Bargain shoppers, on the other hand, have higher search cost than switchers and therefore they gather limited price information. Loyal customers have high loyalty and high search costs and therefore they always purchase from the same store. Opportunistic loyal customers have high loyalty but a medium search cost which allows them to get information about market prices and request a price-match at their favorite store. Note that bargain shoppers obtain information on competing stores and ask for price-matching at their preferred store. These customers would have paid the full price if the retailer had no PMGs. In other words, PMGs reduce the number of customers paying the list price of the retailer. This competition-enhancing effect can overcome the competition dampening effects of PMGs and thereby lead to competitive price decisions. Ho et al. (2004) consider a model where customers engage in search for low prices after the purchase and show that PMGs in this case actually result in competitive prices.

Price Discrimination The first two research streams suggest that PMGs facilitate collusion or competition. These two streams either consider asymmetric retailers or assume that the willingness-to-pay of customers is identical even if they have different information regarding retail prices. However, the third stream integrates customer heterogeneity both in terms of information level and willingness-to-pay (Varian 1980). This stream claims that retailers utilize PMGs to price-discriminate customers based on their knowledge about market prices (Corts 1997; Png and Hirshleifer 1987).

Png and Hirshleifer (1987) model two customer segments: tourists and locals. Tourists have an infinite search cost and therefore an inelastic demand, whereas locals can collect information about the retailer's pricing strategies at zero cost and have an elastic demand. They show that at equilibrium, both retailers will offer PMG since a retailer loses nothing by offering the guarantee. There exist multiple equilibria to the game. Specifically, in a duopoly, one retailer offers a high (monopoly) price to maximize profits from tourists, while the competitor retailer sets the lower price in the market (effective price) to attract all customers, especially locals, who are informed about the prices.

Levy and Gerlowski (1991) consider a model where the retailer can affect the information level of customers through their advertising decisions. Customers select from the retailers whose advertisements they have received and purchase from the lowest priced retailer. Authors show that PMGs allow price discrimination at the expense of lower advertisement levels and lower profits.

Corts (1997) provides a more general model and shows that PMGs can be used as price discriminating devices which can have both anti-competitive and pro-competitive effects. He builds a competition model with heterogeneous retailers and two customer segments: sophisticated and unsophisticated. Sophisticated customers consider only the effective price (the minimum price in the market), whereas unsophisticated customers consider the list prices. Unsophisticated customers are shared among the retailers according to an unspecified process, which can be considered a function of their market share. Sophisticated customers, on the other hand, buy from the lowest priced retailer. If a number of retailers offer the PMGs, then sophisticated customers will be equally shared among them. Note that in the absence of PMGs, the low-priced retailer captures the sophisticated customers. However, in the existence of PMGs, the high-priced retailers can compete with the low-priced retailer for the sophisticated customer segment. Therefore, the sophisticated customers become (relatively) less important for the low-priced retailer. As a result, the price decision of the low-priced retailer gets closer to its unsophisticated demand optimal price. This may lead to different results. If the sophisticated customer's demand is relatively elastic, the lowest priced retailer increases its price, which in turn leads to a price increase with the high-priced retailer as well. On the other hand, if sophisticated customers have a relatively inelastic demand, then the same reasoning leads to a decrease in prices.

Signalling Evidently, the first three streams of research indicate that customer type and behavior have a direct consequence on the competitive effects of PMGs. However, these streams assume that PMGs do not directly influence the purchase pattern of customers. The fourth stream of research addresses this gap by investigating customer responses to PMGs. It deals with issues such as how search and purchase decisions of customers are affected by PMGs, and whether retailers actually act according to customer expectations when deciding to offer PMGs (Jain and Srivastava 2000; Kukar-Kinney and Grewal 2007; Moorthy and Winter 2006; Moorthy and Zhang 2006).

Jain and Srivastava (2000) provide empirical evidence that customers who are not knowledgeable about the store prices view PMGs as signals of low prices; therefore they only shop at stores that offer PMGs. In contrast, the choice of stores by customers who are knowledgeable about the store prices is dependent on a number of factors. Namely, these are list price, price-matching policy, and store characteristics such as service quality (so, unlike the previous literature mentioned, they do not necessarily purchase from the lowest priced retailer). The theoretical analysis of the authors proves that only retailers with low prices will offer PMGs which is consistent with customer expectations, but contrary to price discrimination arguments.

Moorthy and Winter (2006) support the interpretation of PMGs as signals; they show that low-priced retailers benefit from offering price-matching by indicating their price positions. Moorthy and Zhang (2006) question the incentives for (all) differentiated retailers to offer PMGs. They demonstrate that offering the guarantee signals low service and a low price, while not offering the guarantee is a signal for high-level service and high price. Mamadehussene (2019) explains how homogeneous firms may signal their low prices through PMGs: consumers perceive PMG stores to have lower prices, not because they expect them to have low marginal costs or service quality, but simply because they offer a PMG.

8.2.2 Dual Channels

The second stream of literature related to our work examines the management and coordination of direct channels. Despite many of its advantages, constructing a direct channel places the supplier in face-to-face competition with the retailer (Cattani et al. 2004). As such, many papers in the dual-channel literature determine the conditions under which the supplier and the retailer benefit from the integration of a direct channel while considering prices (Arya et al. 2007; Cai 2010; Chiang et al. 2003) or price and service efforts (Bell et al. 2006; Tsay and Agrawal 2004).

There is also a broad operations management literature focusing on inventory considerations and investigating: (1) the channel inefficiencies arising from demand uncertainty (Chiang 2010; Mahar et al. 2009), (2) the evaluation of alternative distribution strategies (Chiang and Monahan 2005), (3) the value of efficient replenishment (Dong et al. 2007), and (4) the design of coordinating contracts (Boyaci 2005). More recent works identify the optimal dual-channel strategies when customers' purchase channel choice depends on sophisticated metrics, such as inventory availability and delivery time, in addition to price (Chen et al. 2008). Although many of these papers consider horizontal competition as well as vertical, none address the role of price-matching guarantees.

Cattani et al. (2006) is the closest to our work and analyzes a pricing scheme where the supplier commits to fixing the price of the direct channel equal to the price of the retail channel to mitigate the channel conflict arising from the integration of the direct channel. As such, the supplier cannot set a price lower than the retailer, which artificially annihilates price competition. In contrast, the supplier in our setting can set a direct channel price lower than the retail channel—even when offering a price-match guarantee.

8.2.3 Dominant Retailer

The third stream of literature relevant to this paper deals with the supplier–retailer interactions and implications for channel power (Choi 1991, 1996; Lee and Staelin 1997; Moorthy and Fader 2012). While empirical studies aim to measure the power of channel members (Kadiyali et al. 2000; Sudhir 2001), early theoretical works focus on explaining the distribution of profits based on the supply chain structure and the timing of pricing decisions (Choi 1991, 1996; Lee and Staelin 1997; Moorthy and Fader 2012). Over the last decade, the literature has recognized the emergence of dominant retailers due to innovative and efficient operations (Useem et al. 2003), increased use of information technology, and/or intense manufacturer competition (Kadiyali et al. 2000). Research has focused on the impact of dominant retailer on the distribution of channel profits (Chen 2003; Dukes et al. 2006) and product assortment decisions (Dukes et al. 2009), as well as the design of coordinating contracts in the presence of a dominant retailer (Raju and Zhang 2005;

Kolay and Shaffer 2013). Of particular interest to us is Geylani et al. (2007). In that model, a single supplier sells to a duopoly of retailers. One of the retailers is dominant and dictates the wholesale price to the supplier, whereas the other retailer is weak and accepts what the supplier offers. It is shown that the supplier may respond to the dominant retailer by engaging in marketing activities and shift the demand from the dominant retailer channel to the weak retailer channel and at the same time by increasing the wholesale price for the weak retailer. In our model, the supplier cannot engage in marketing activities, yet we find that price-matching guarantee offer by the supplier shifts the demand from the dominant retail channel to the direct channel.

8.3 Demand Model

For analytical tractability, we analyze availability verification and channel coverage separately. In this section, we build a general demand model that can be easily customized to investigate these two cases. We categorize the real-life implementation of PMGs into two policies. Under the first, the retailer guarantees it will match the lower price offered by its competition to any customer demanding a price-match; we call this *simple price-matching policy*, or PM for short. Under the second policy, the retailer verifies the availability of the item at the competitor store, and matches the lower price *only if* the item is available at that store; we call this *price-matching based on availability*, or PMA for short.

Consider two retailers, R_1 and R_2, both selling the same product. Table 8.1 provides the glossary of notations used. The retailers simultaneously announce the

Table 8.1 Notation

i	Index for each channel. In the main model, we have $i \in \{1, 2\}$ and in the dual channel model we have $i \in \{R, S\}$ representing the retailer-owned and supplier-owned channels
R_i	Retailer i
$\mathscr{P}_i$	The PMG policy of channel i where $\mathscr{P}_i \in \{C, PM, PMA\}$
δ_i	Product availability at channel i
q_i	Perceived quality of each channel i
p_i	Announced list price at each channel i
p_e	The minimum of the two list prices, i.e., $\min\{p_1, p_2\}$
t	Per distance travel cost of a customer
$t_i(l)$	Cost of travel for the customer in location l to retailer i
V	Customers valuation of the ideal product
$u_i^{\mathscr{P}_i}(l)$	The net utility gained by a customer located at point l by visiting channel i
$\mathscr{S}_i$	Strategy of channel i consisting of the PMG policy and price
$d_i(\mathscr{S}_i, \mathscr{S}_j)$	Demand for channel i
$\pi_i(\cdot)$	Profit for channel i

type of price-matching guarantee that they are offering before the pricing decisions. They have three options available: no price-matching guarantee (C), price-matching (PM), and price-matching based on availability (PMA). The retailers simultaneously set prices p_1 and p_2, to maximize profits, given their availability levels and the PMG policy (C, PM, or PMA). We define the availability of a retailer as the probability that a customer will find the item on the shelf any time she or he visits the store. Availability levels for retailers are exogenous parameters, denoted by $\delta_i \in (0, 1]$, $i = \{1, 2\}$. Essentially, this means that retailers have already specified and committed to a desired fill-rate[6] (or in-stock percentage) for the product. Since fill-rate decisions typically are made in advance compared to pricing policies, it is reasonable to assume availability levels as given parameters. In Sect. 8.4.4, we relax this assumption and endogenize the fill-rate decisions of the firms. We assume that the unit cost of the product is identical for the two retailers and we normalize it to zero.

On the demand side, customers make a store choice based on the utility provided by the offerings of each retailer. Each retailer's demand is assumed to be equal to the aggregate demand multiplied by its market share, which depends on the PMG policy, price, and fill-rate. The total number of customers who want the product (i.e., the aggregate demand) is normalized to 1.

The retailers are located at the end points of a unit line and potential customers are uniformly distributed between the retailers. Each customer has a constant valuation V for the ideal product. We allow for vertical differentiation between the retailers through their *perceived quality*. In our model, we denote the quality of the retailers as q_1 and q_2. Each customer purchases zero units or one unit of the product, and incurs a travel cost of t per unit distance traveled. Note that the travel cost represents the degree of horizontal differentiation between the retailers. That is, a high travel cost represents highly differentiated retailers—which could be due to distinct store experiences such as sales features and store design—or due to large geographical distance between them, and vice versa (Tirole 1988, p. 279).

All the retail characteristics (i.e., price, quality, fill-rate, and PMG policy) are observable by individual customers. Customers cannot verify product availability before visiting a retailer, but they know the fill-rate, i.e., the anticipated probability of finding the product in stock at each retailer. Each customer makes a store choice in order to maximize her/his expected net utility. We assume that the cost of visiting a second retailer is arbitrarily high and, therefore, customers visit just one retailer. Note that this high travel cost is about physically visiting the second retailer and not about the cost of price information. In that sense, customers are aware of the prices in each store as well as the product availability.

Utility Functions Without loss of generality, we place R_1 at origin and R_2 at the end of the unit line. In what follows, we derive the expected net utility of a representative customer visiting R_1 under each possible PMG policy; the expected

[6]Fill-rate of a product represents the long-run probability of finding it in stock for immediate purchase.

utility of visiting R_2 can be obtained similarly. Let $t_i(l)$ denote the cost of traveling to R_i for a customer located at point l on the unit line. Specifically, $t_1(l) = tl$ and $t_2(l) = t(1 - l)$. Furthermore, let $u_i^{\mathscr{P}_i}(l)$ denote the expected utility obtained at R_i under policy $\mathscr{P}_i \in \{C, PM, PMA\}$, $i \in \{1, 2\}$. Based on the above description, the expected utility of visiting R_1, when R_1 is not offering a PMG, is

$$u_1^C(l) = (Vq_1 - p_1)\delta_1 - t_1(l). \tag{8.1}$$

Consider now the case where R_1 offers PM. While estimating the utility of visiting R_1, customers will consider the price of R_2, too. On one hand, if R_1 is the high-priced retailer, then customers will ask for a price-match if they visit R_1. Each price-match request is going to be granted since the PMG is not contingent on availability. On the other hand, if R_1 is the low-priced retailer, then there is no need to ask for a price-match. Thus, the list price of R_1 can be replaced by the minimum of the two prices, $p_e \equiv \min\{p_1, p_2\}$, which we denote as the "effective price." The expected utility in this case is

$$u_1^{PM}(l) = (Vq_1 - p_e)\delta_1 - t_1(l). \tag{8.2}$$

Lastly, suppose that R_1 offers PMA. Similar to the PM case, customers ask for a price-match at R_1. However, the end result of their price-match requests depends on the availability of the product at R_2. Particularly, if the product is available at R_2, then customers receive the effective price and demand one unit of the product. If the product is not available at R_2, then they face the list price p_1, and demand one unit of the product as long as the price is less than their valuation. In other words, the price faced at R_1 is p_e with probability δ_2 and p_1 with probability $1 - \delta_2$, and the expected price is given by $\overline{p}_1 \equiv p_e\delta_2 + p_1(1 - \delta_2)$. The expected utility in this case is

$$u_1^{PMA}(l) = (Vq_1 - \overline{p}_1)\delta_1 - t_1(l) = (Vq_1 - p_e\delta_2 - p_1(1 - \delta_2))\delta_1 - t_1(l). \tag{8.3}$$

Demand Customers compare the above utilities and visit the retailer providing the higher expected net utility, as long as the net utility is also non-negative. For instance, if neither retailer offers a PMG, then each customer compares $u_1^C(l)$ and $u_2^C(l)$. The customers located at $l_{u_1} = (Vq_1 - p_1)\delta_1/t$ and $l_{u_2} = 1 - (Vq_2 - p_2)\delta_2/t$ are indifferent between not buying and visiting R_1 and R_2, respectively. The customer located at $l_{u_1-u_2} = [t + (Vq_1 - p_1)\delta_1 - (Vq_2 - p_2)\delta_2]/(2t)$ is indifferent between the two retailers.

Depending on the problem parameters and strategies of the retailers, we may have an uncovered market, meaning that there is a non-empty set of customers who cannot obtain a non-negative expected utility from any of the retailers and, therefore, do not visit any of them. In this case, the market share of R_i is going to be $((Vq_i - p_i)\delta_i)/t$ for $i = 1, 2$. We may also have a covered market where all customers obtain a non-negative expected utility from at least one of the retailers. In this scenario, the market shares of R_1 and R_2 are going to be $[t + (Vq_1 - p_1)\delta_1 - (Vq_2 - p_2)\delta_2]/(2t)$ and $1 - [t + (Vq_1 - p_1)\delta_1 - (Vq_2 - p_2)\delta_2]/(2t)$, respectively.

When we have an uncovered market scenario, each retailer acts as a local monopoly. This is an uninteresting case to analyze since retailers are essentially not competing for any of the customers in the market. Therefore, they do not have any incentive to offer a PMG. So, in the sequel, we assume that the problem parameters are such that the market is covered at equilibrium. A covered market equilibrium is guaranteed for a relatively low travel cost, i.e., retailers are close enough in terms of store characteristics and/or geographical proximity and compete directly for customers. Indeed, this assumption is consistent with the real-life implementation of PMGs, since they are valid only for the retail stores located in the same market area.

If we let $\mathscr{S} = (\mathscr{S}_i, \mathscr{S}_j)$ represent the strategy of the retailers, where $\mathscr{S}_i = (\mathscr{P}_i, p_i)$ for retailers $i = 1, 2$ and $j = 3 - i$, then we can derive the market shares of the two retailers for any combination of PMG offers as follows:

$$d_1(\mathscr{S}) = \frac{t + (Vq_1 - p_1^{\mathscr{P}_1})\delta_1 - (Vq_2 - p_2^{\mathscr{P}_2})\delta_2}{2t},$$

$$d_2(\mathscr{S}) = \frac{t - (Vq_1 - p_1^{\mathscr{P}_1})\delta_1 + (Vq_2 - p_2^{\mathscr{P}_2})\delta_2}{2t},$$

where $p_i^C = p_i$, $p_i^{PM} = p_e$, and $p_i^{PMA} = \overline{p}_i = p_e\delta_j + p_i(1 - \delta_j)$ for $i = 1, 2$ and $j = 3 - i$.

8.4 Verification of Availability as a Price-Matching Condition

In this section, we analyze the equilibrium price-matching policy of the retailers and identify how the verification of availability as a price-matching condition changes the pricing decisions.

The game consists of two stages. In stage one, retailers set PMG policies. In stage two, retailers choose prices. At each stage, involved parties make decisions simultaneously and we seek subgame perfect equilibrium by solving the game backwards. We consider two different models for the consumer behavior. We first assume that customers are cognizant of the availability clause as a price-matching condition; we refer to this case as the *attentive customers*. Specifically, customers make their store choice decisions based on the anticipated probability of being successful in their price-match requests in addition to other dimensions, such as price and anticipated probabilities of finding the product in stock at each retailer. In short, they are aware of the availability clause as a price-matching condition. Subsequently, we analyze a setting where customers ignore the availability clause as a price-matching condition, and only consider the price and the anticipated probability of finding the product in stock at each retailer; we call this case the *inattentive customers*. Note that our results would continue if customers only consider the price and ignore the issue of product availability and, therefore, ignore the availability clause completely, while making store choices.

In each case, there are nine subgames to be considered in stage two. For the sake of expositional clarity, we focus on the strategic decision of whether firms should offer a price-matching guarantee and report on the managerial findings of our analysis. All the technical details, the proofs of the propositions, as well as the expressions for equilibrium decisions are available from the authors upon request.

8.4.1 Attentive Customers

Profit Functions Suppose that R_1 is not offering any type of PMG. If customers decide to visit R_1, then they will face the list price, independent of the PMG policy offered by R_2, and receive one unit of the product as long as the item is available. Accordingly, the expected revenue in this setting is

$$\pi_1(C, p_1, \mathscr{S}_2) = p_1\delta_1 D_1(C, p_1, \mathscr{S}_2).$$

Suppose now that R_1 is offering PM. While making their store decisions, customers will look at the effective price instead of the list price because the price-matching offer at R_1 is not conditional on availability. Accordingly, the expected revenue function of R_1 is

$$\pi_1(PM, p_1, \mathscr{S}_2) = p_e\delta_1 D_1(PM, p_1, \mathscr{S}_2).$$

Recall that, when offering PMA, R_1 matches the price of R_2 only if the product is available at R_2 at the time of the request. In other words, some of the customers visiting R_1 will be able to get the effective price while others will receive the list price. We assume that the inventory is proportionally rationed in the event of a stock-out. Then, the expected revenue at R_1 is

$$\pi_1(PMA, p_1, \mathscr{S}_2) = [p_e\delta_2 + p_1(1-\delta_2)]\delta_1.D_1(PMA, p_1, \mathscr{S}_2).$$

Note that since customers are sensitive to product availability, the inventory decisions of the retailers have an effect on demand in all three scenarios. However, *only under PMA policy* the inventory decision of the competing retailer has a direct effect on the price charged to customers, i.e., $\overline{p}_i = p_e\delta_j + p_i(1-\delta_j)$ for $i = 1, 2$ and $j = 3 - i$.

Equilibrium Analysis For analytical tractability, we first report on the case where two retailers are symmetric in terms of the perceived quality and the fill-rate that they provide, specifically: $q_1 = q_2 = 1$ and $\delta_1 = \delta_2 = \delta < 1$. In Sect. 8.4.4, we extend our model and show that the managerial results of our framework continue to hold when we endogenize the fill-rate decisions of the two retailers.

Suppose that both retailers offer PM. At any solution with equal prices, R_1 is unable to attract more customers through price reduction since R_2 automatically matches the lower price because of its PMG. Also, R_1 has no incentive to increase the price, because it is also offering PM and has to match the price for all of its customers, i.e., the demand and the profit margin is going to be the same even if R_1 increases the price. In other words, offering a PM policy completely eliminates the price competition between the retailers, allowing them to collectively increase prices. As a result, when both retailers offer the guarantee, there exists a Pareto dominant price equilibrium where each retailer charges a price higher than the no-PMG case and also makes more revenue compared to the no-PMG case.

Suppose that both retailers offer PMA. If a retailer tries to attract more customers by cutting the price, then the competing retailer will match the price as long as the product is available. As such, similarly to the PM case, offering PMA policy also allows retailers to collectively increase prices and revenues compared to the no-PMG case.

A complete description of the game also requires the comparison between all three in order to identify the effects of verifying the availability before matching the price, which we provide in the following proposition.

Proposition 8.1 (Equilibrium PMG Policy) *Offering a simple price-matching guarantee (i.e., PM policy) is the (weakly) dominating equilibrium strategy for both retailers. At equilibrium, retailers tacitly collude, charge a higher price, and make more revenue compared to the no-PMG case.*

In terms of the availability verification clause, there exists a threshold for the unit travel cost t^* such that the following are true.

- *If the travel cost is lower than the threshold value, then verification of availability as a PMG condition lowers the price and decreases retail revenues.*
  Consequently, PMA policy is strictly dominated by PM policy.
- *If the travel cost is higher than the threshold value, then verification of availability as a PMG condition has no effect on the equilibrium decisions and revenues. Consequently, PMA policy is identical to PM policy.*

So, if customers are conscious about availability implications and take into consideration the probability of not receiving a price-match while making store choices, then retailers should *never* verify the availability as a PMG condition and should match the price of the competing retailer without any conditions (i.e., adopt PM).

For the sake of discussion, let us look at the benefits and costs of verifying the availability from the perspective of R_1. Verifying the availability allows R_1 to charge a higher price to its customers if a stock-out is observed at R_2. Suppose that the state of the demand is high and that the inventory of R_2 is exhausted. As soon as the last unit in R_2 is sold, R_1 can start declining the price-match requests via its availability verification clause. By declining the requests, R_1 charges a higher

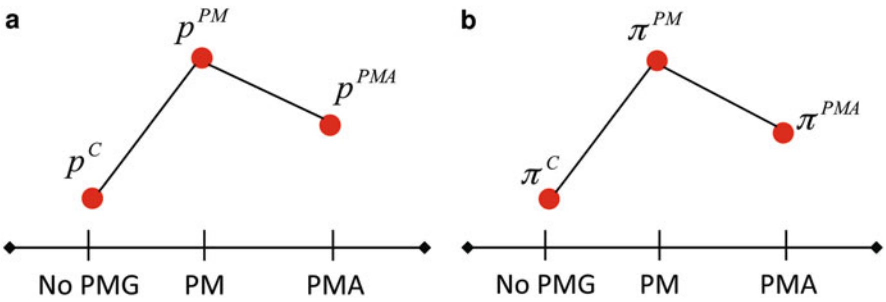

Fig. 8.3 Effects of PMGs with attentive customers. (**a**) Change in price. (**b**) Change in revenue

price to its customers when they are at the store asking for a price-match, i.e., once their travel cost to R_1 is sunk cost and there are no units available at R_2. Charging a higher price by declining requests is beneficial for R_1 since it extracts a higher margin from these customers. However, there is also a cost attached to such availability verification. Under PM policy, R_1 does not lose any customers to R_2, even if R_1 has a higher list price since it matches the lower price of R_2 for all its customers. However, when verifying the availability, R_1 is matching the price *only* if the product is available at R_2. That is, R_1 is matching the price not for *all* but *some* of its customers. As such, R_1 loses some customers to R_2. We observe both effects while deriving the best response functions of the players under PMA policy. However, in equilibrium, the cost associated with the verification condition turns out to be stronger than its benefits, thus creating an incentive for the retailers to reduce their prices compared to the PM case (Fig. 8.3).

Naturally, the benefit of a price reduction depends on the depth of the price cut, as well as the size of the customer segment that the retailers can attract. If the travel cost is low, then retailers can attract a significant number of customers by lowering prices. However, if the travel cost is high, then they need a much deeper price cut to attract the same number of customers. Therefore, a sufficiently low travel cost triggers competition leading to lower equilibrium prices under PMA policy compared to the PM case, and a high travel cost results in the same pricing and inventory decisions under both PM and PMA policies.

8.4.2 Inattentive Customers

In this section, we look into the case of customers who disregard the availability clause while making store choice decisions. That is, inattentive customers who effectively assume that they will always be able to execute the price-match successfully.

The utility functions for this case can be derived from Eqs. (8.1)–(8.3). The net utility of visiting R_1 from a customer's perspective for three different PMG policies can be written as $u_1^C(l) = (V - p_1)\delta_1 - tl$, $u_1^{PM}(l) = u_1^{PMA}(l) = (V - p_e)\delta_1 - tl$.[7] Clearly, there is no difference between PM and PMA policies from the customer's perspective. The market share and demand of each retailer can also be derived and substituted into the original expected revenue functions.

Similar to the previous case, we can prove the existence of pure strategy equilibrium under all possible PMG scenarios. Specifically, we can show that offering a PM policy dominates not offering a PMG, and leads to higher equilibrium prices and revenues, supporting that the tacit collusion outcome of deterministic PMGs *continues to hold* under PM policy in an uncertain demand environment with inattentive customers. But, we are more interested in identifying the effects of an availability verification clause; for this we need to compare PM and PMA policies. The following proposition summarizes the equilibrium of the game.

Proposition 8.2 (Equilibrium PMG Policy) *If customers are inattentive to the availability verification clause, then the following are true:*

- *Offering PMA policy is the equilibrium strategy for retailers.*
- *When both retailers offer PMA, there exist two equilibria that are mirror images of each other. In each equilibrium, one retailer charges a higher price and earns more, compared to the other retailer and the PM game. The other retailer sets its price to be identical to the PM equilibrium, and also earns the same revenue.*
- *Verification of availability clause leads to higher retail prices.*

Note that, when retailers offer PMA, there are two equilibria that are mirror images of each other. This dual outcome makes it impossible to favor one equilibrium based on commonly used selection criteria such as Pareto or Risk Dominance (Harsanyi and Selten 1988). Given the symmetry, the analysis and managerial insights of each equilibrium lead to the same conclusions. For this reason, we are not concerned about the equilibrium choice and we proceed with the implicit assumption that the players have some mechanism by which they arrive at one of these equilibria (Schelling 1980). Let the subscripts H and L represent the high- and low-priced retailers under PMA policy, respectively. Accordingly, the previous proposition says that in equilibrium R_H sets price as p_H^{PMA} and earns π_H^{PMA} and R_L sets price as p_L^{PMA} and earns π_L^{PMA} where $p_H^{PMA} > p_L^{PMA} = p^{PM}$ and $\pi_H^{PMA} > \pi^{PM} = \pi_L^{PMA}$ (Fig. 8.4).

Recall that customers now believe that they will always get the effective price. Thus, R_H is still able to attract customers based on the effective price even if it lists a higher price. When customers visit the store, R_H matches the price as long as

[7] An alternative modeling approach is as follows. While making store choice decisions, customers focus completely on price and ignore the possibility of a stock-out and, therefore, also the possible decline of their price-matching request based on availability. We have explored this approach in detail; the managerial and economic insights as to the effects of verifying the availability are the same as discussed in Sect. 8.4.2. In that sense, the two setups: (1) customers consider the possibility of stock-outs while making store choices but ignore the availability condition of PMGs, or (2) they ignore the possibility of stock-outs completely, are equivalent.

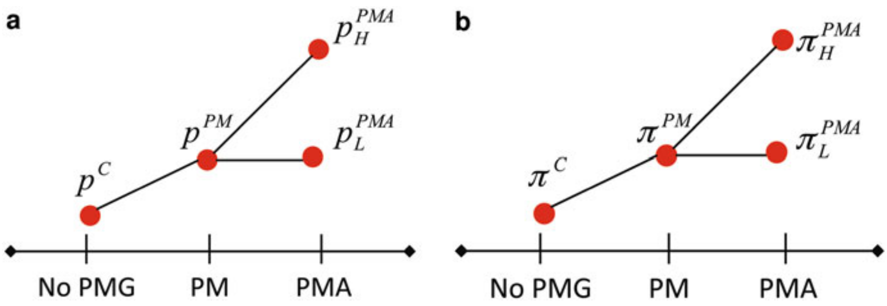

Fig. 8.4 Effects of PMGs with inattentive customers. (**a**) Change in price. (**b**) Change in revenue

R_L has the products available. If R_L faces a stock-out, then R_H can start declining the price-match requests and charge its list price since it is the sole provider of the product in the market place. This scenario provides an incentive for retailers to set a higher price than the competitor and to extract a premium from customers when there is scarcity in the market. Consequently, R_H declines their price-match requests to achieve a higher profit margin. As such, verification of availability leads to a higher margin as well as to higher revenues.

8.4.3 Comparison of the Two Cases: Attentive Customers Versus Inattentive Customers

Our analysis reveals that the impact of availability verification as a price-matching condition depends significantly on how consumers make their retail store choice decisions. Specifically, it depends on whether customers are attentive to the consequence of availability verification clause. In the following, we briefly summarize the impact of the availability verification clause on both the degree of competition between the retailers and on the consumer surplus. We measure consumers' surplus as the expected net utility obtained by all customers after visiting the retailers.

Whether with attentive or inattentive customers, offering a simple price-matching guarantee (PM) eliminates price competition, thus allowing retailers to collectively increase prices compared to the no-PMG scenario. This result shows that the tacit collusion results related to PMGs in a deterministic setting continue to hold even with less than perfect product availability.

More interesting is the effect of verification of availability captured in the PMA policy. If customers are cognizant about the possibility of product unavailability, then *verification of availability may re-institute price competition*, particularly if the travel cost is sufficiently low.

Note that the prices with availability verification are still higher than the case of no-PMG, that is, the price competition is not completely re-installed. As a result of the partially re-installed price competition, verification of availability actually decreases retail revenues.

In summary, from the retailers' point of view, verification of availability is not beneficial with attentive customers, and can actually be detrimental when the degree of horizontal differentiation between the retailers is high. Our analysis reveals that the verification of availability as a price-matching condition (weakly) increases the consumers' surplus if customers are attentive.

We can also show that verification of availability as a price-matching condition decreases the consumers' surplus if customers are inattentive to the availability clause. So, verification of availability is detrimental to customers if they ignore the price-matching conditions. In this scenario, availability-contingent guarantees not only strengthen the tacit collusion ability of the retailers but also allow them to increase prices and discriminate against customers based on availability. This means that verification of availability acts as a tool for discrimination and leads to price dispersion *even in a symmetric market when all customers are informed about prices*.

Note that, based on the existing literature on PMGs, we need information heterogeneity among customers in order to observe the price discrimination results (Png and Hirshleifer 1987). However, we establish that availability verification may lead to price dispersion and price discrimination in the market, even if all customers are informed about retail prices. From the retailers' perspective, verification of availability leads to higher revenues and is the equilibrium strategy.

8.4.4 Extended Discussions on Fill-Rate

In the above analysis, we assumed that the fill-rate of each retailer was exogenously set. We also assumed that the fill-rate of each retailer is identical, primarily for expositional brevity. However, these two assumptions do not restrict the generality of our findings. In particular, we have extended our analysis to allow retailers to make fill-rate decisions. The game will now consist of three stages. In the first stage, retailers simultaneously decide on their price-matching policy. In the second stage, retailers simultaneously decide on their fill-rate. In the third stage, retailers simultaneously decide on their price.

In order to ensure that the optimal fill-rate decisions are interior, we assume a quadratic cost function $\alpha\delta_i^2$ as the cost of retailer $i = 1, 2$. We again solve the game via backward induction. The equilibrium decisions of the retailers in terms of PMGs and the verification of availability are identical, even when the fill-rate decisions are involved. Specifically, PMGs allow retailers to collude and increase prices compared to the no-PMG case. Retailers also increase the fill-rate provided as the tacit collusion eliminates the price competition between them and intensifies the degree of service competition (fill-rate). If customers are attentive to the availability verification condition, then verification of availability re-institutes the price competition and, thereby, leads to lower prices and higher fill-rates compared to the simple PM case. If, however, customers are inattentive to the availability verification condition, then the verification clause leads to higher retail prices and lower fill-rates.

8.5 The Impact of Supplier-Owned Direct Channel on Price-Matching

Our goal in this section is to analyze the extent of channel coverage as a price-matching condition. For this purpose, we consider a dual-channel setting. We suppose that one of the retailers in our general model framework is owned and operated by a supplier that distributes its product via this direct channel, as well as by the retail channel. Specifically, we will keep retailer 1 as a reseller of the product and suppose that retailer 2 is owned by the supplier, which also sells its product through retailer 1. The game now consists of three stages. In the first stage, the retailer and the supplier simultaneously announce whether or not they are offering a price-matching guarantee. In the second stage, the unit wholesale price, w, is declared. In the third stage, the supplier announces the direct channel price, p_S, and the retailer announces the retail channel price, p_R.

The second stage of the game is essentially the specification of the wholesale price between the supplier and the retailer.

We are interested in understanding the role of PMGs in dual channels. PMGs are known to alter the demand of competing firms as well as the resulting equilibrium prices (Hay 1981; Salop 1986). Accordingly, they are expected to have an impact on the pricing decisions between the supply chain partners as well. Therefore, it is important to identify the role of the channel power between the retailer and the supplier.

In our stylized framework, we simplify the bargaining process by giving absolute power to one of the supply chain members. We investigate two distinct scenarios: a *supplier dominant chain*, in which the supplier decides the wholesale price for its product, and a *retail dominant chain*, in which the retailer dictates the wholesale price for the supplier's product. Channel power, in our model, through the ability of dictating the wholesale price, is artificially imposed and not self-driven through unique characteristics such as operational efficiency, excellence in quality, or forcible market power. However, it is also worthwhile to note that our model does not forego the advantage that a firm can enjoy through perceived quality and service characteristics. In fact, through a general model of horizontal and vertical differentiations, we allow either the retail channel or the direct channel to benefit from providing higher quality and service. Nonetheless, analyzing the scenarios of supplier and retail dominance is sufficient for our analysis to demonstrate that a shift in channel power provides a simple explanation as to why we observe price-matching guarantees, particularly by suppliers, in dual channels.

Note that, being an operational decision, product availability would not be considered the main criteria for the retailer and the supplier to make their price-matching decisions. Therefore, and also for analytical tractability, we suppose that both channels have perfect availability in the analysis of this section. The utility functions for this case can be derived from Eqs. (8.1)–(8.3) by setting $\delta_1 = \delta_2 = 1$. Moreover, we normalize the perceived quality of the retail channel as $q_R = 1$, and denote the perceived quality of the supplier channel as $q_S > 0$. In our analysis, we consider all cases: $0 < q_S \leq 1$ and $1 \leq q_S$. We can then derive the market shares of the two retailers for any combination of PMG offers as follows:

$$d_R(\mathscr{S}) = \frac{t + (V - p_R^{\mathscr{P}_R}) - (Vq_S - p_S^{\mathscr{P}_S})}{2t}$$

$$d_S(\mathscr{S}) = \frac{t + (Vq_S - p_S^{\mathscr{P}_S}) - (V - p_R^{\mathscr{P}_R})}{2t}$$

where $\mathscr{S} = (\mathscr{S}_i, \mathscr{S}_j)$ represent the strategy of the retailers with $\mathscr{S}_i = (\mathscr{P}_i, p_i)$, $\mathscr{P}_i \in \{C, PM\}$, $p_i^C = p_i$, $p_i^{PM} = p_e$, for $i = 1, 2$ and $j = 3 - i$.

The profits for the retailer and the supplier, respectively, are then given as

$$\pi_R(\mathscr{S}) = (p_R^{\mathscr{P}_R} - w) \cdot d_R(\mathscr{S}) \quad \text{and} \quad \pi_S(\mathscr{S}) = p_S^{\mathscr{P}_S} \cdot d_S(\mathscr{S}) + w \cdot d_R(\mathscr{S}).$$

Note that, as a result of the dual-channel approach, the suppler collects a margin—albeit different from each channel—from all the transactions in the market. The supplier's profit margin from the direct channel sales is p_S whereas the profit margin from the retail channel sales is w.

8.5.1 Traditional Chain

In the traditional chain, the supplier sets the product wholesale price. An important observation is that the supplier can decide the exact profit margin of the retailer when the retailer is offering a price-matching guarantee. The retailer's profit margin is $\min\{p_R, p_S\} - w$ when offering a price-matching guarantee. If the supplier sets the direct channel price lower than the retail channel price, then the profit margin for the retailer becomes $\min\{p_R, p_S\} - w = p_S - w$ and both p_S and w are set by the supplier. In fact, price-matching becomes detrimental for the retailer since the supplier eliminates the retailer's margin by setting $p_S \leq w$. The retailer is left with no profits regardless of the supplier's price-matching strategy. Therefore, offering price-matching is the dominant strategy for the retailer.

In a similar vein, the supplier also does not offer the guarantee. If the supplier offers a price-matching guarantee, then she cannot charge a higher price than the retailer due to her price-matching promise. All the customers of the direct channel request the supplier to match the lower price of the retailer. In addition, a segment of customers, who would otherwise shop at the retail channel attracted by the lower price, start shopping at the direct channel. As such, the retailer cannot attract customers by cutting the price. Since the demand is insensitive to the price, the retailer increases his price.

In terms of supplier profitability, we see that price-matching decreases the price and increases the demand of the direct channel. However, the supplier would be forced to provide a much deeper price cut to achieve an equivalent demand increase in the absence of price-matching. As such, direct channel profits are higher under price-matching.

We observe the opposite effect on the retail channel. Price-matching decreases the retail channel demand while increasing the profit margin of the supplier. Yet, the overall effect of the demand decrease is stronger, so the supplier is left with lower retail channel profits. In fact, the decrease in the retail channel profit is higher than the increase in the direct channel profit and, therefore, it is not profitable for the supplier to offer a price-matching guarantee. We formalize this result in the following proposition.

Proposition 8.3 *In the traditional chain, price-matching guarantees do not prevail in equilibrium.*

Neither the supplier nor the retailer offers a price-matching guarantee at equilibrium. This finding is particularly valuable since the supplier would benefit from offering the price-matching guarantee if the wholesale price is an exogenous problem parameter in our model framework. However, if the supplier can set the wholesale price for her product, then offering a price-matching guarantee is no longer a beneficial strategy. The main reason is that the supplier has the ability to capitalize on certain situations by adjusting the wholesale and the direct channel prices together and shifting the demand from one channel to the other. This ability is forfeited when the supplier offers a price-matching guarantee.

If the direct channel quality is sufficiently high, then the direct channel price will be higher than the retail channel in the absence of price-matching. This means that the supplier enjoys high margin-low demand in the direct channel and low margin-high demand in the retail channel. Offering a price-matching guarantee shifts demand from the retail channel to the direct channel. The supplier could facilitate this shift by adjusting the wholesale price in the second stage, even without price-matching in place—if it was profitable for the supplier. Offering a price-matching guarantee limits the supplier's ability to extract profits from the retail channel by shifting demand from the retail channel to the direct channel. Consequently, the supplier's profit loss from the retail channel is increasing in a convex fashion as the channel quality increases. As a result, the value of price-matching for the supplier is negative and decreasing in a convex fashion as the quality increases. Eliminating the price competition through promises to match the price is not beneficial for the supplier in the traditional chain because the ability to manipulate the price and demand in each channel through the wholesale price is more available.

8.5.2 Retail Dominant Chain

In the retail dominant chain, the wholesale price—i.e., the profit margin of the supplier from the retail channel—is decided by the retailer in the second stage of the game. This means that the retailer can set the supplier's margin for the direct channel if the supplier offers the price-matching guarantee. In particular, the supplier's margin in the direct channel would be $\min\{p_R, p_S\}$ and it would be decided by the retailer through price-matching when $p_R \leq p_S$. As such, dictating the profit margin of the opponent through the price-matching guarantee of the opponent is possible in the retail dominant chain as well.

If the retailer sets the wholesale price equal to zero, then the supplier does not make any profit from the retail channel and direct channel remains the only source of profit for the supplier. However, there may be other benefits of selling through the retail channel that are not quantified in our model setting; for instance, product awareness and brand equity. Therefore, we assume that the supplier agrees to sell through the retail channel even when the wholesale price is zero. As a result of this assumption, however, there is no incentive for the retailer to set a positive wholesale price. The only exception is that there are cases, depending on the problem parameters, where the profit function of the retailer is constant with respect to the wholesale price.

In these cases, the retailer can project the wholesale price on the retail price and maintain a constant level of profit. So, the retailer can either favor the supplier by providing a positive wholesale price or favor the customers by setting the wholesale price equal to zero. Our analysis shows that the managerial insights, with respect to the implications of price-matching guarantees, are robust with respect to the wholesale price that the retailer chooses among the alternatives. For the sake of expositional clarity, however, we proceed with the assumption that the retailer favors the customers and sets the wholesale price equal to zero.

Price-Matching by the Supplier Recall that offering a price-matching guarantee is not a profitable strategy for the supplier in a traditional chain. However, implications for the supplier are different in the retail dominant chain. Specifically, offering a price-matching guarantee is the preferred strategy by the supplier in the retail dominant chain. The price-matching guarantee essentially allows the supplier to shift demand from the retail channel to the direct channel. In the retail dominant chain, the supplier has limited influence on the wholesale price and the retail channel price. Therefore, the supplier has limited ability to allocate customers to the two channels via the direct channel price in the absence of price-matching. Therefore, shifting demand from the retail channel to the direct channel via the guarantee is valuable for the supplier.

Price-Matching by the Retailer Offering a price-matching guarantee has adverse consequences for the retailer in the traditional chain since it allows the supplier to extract all the profits of the retail channel. This is no longer a risk in the retail dominant chain because the retailer dictates the wholesale price in the second stage of the game.

The retailer is indifferent to offering a price-matching guarantee when the direct channel quality is high. However, there is a region—relatively high travel cost and relatively low quality—where the retailer strictly prefers not to offer a price-matching guarantee. If the direct channel quality is low, then the retail channel captures a significant portion of the customers. By offering a price-matching guarantee, the retailer reduces his profit margin and attracts more customers. But the increase in demand is insufficient to overcome the loss of the lower profit margin since the demand is already high without the price-matching in place. As a result, the profits are lower with the price-matching guarantee. On the other hand, if the quality is high, then the retailer does not attract a lot of customers in the absence of price-matching. Price-matching allows the retailer to increase the demand. In fact, the increase in the demand outweighs the loss due to the lower profit margin. Therefore, it is profitable for the retailer to offer a price-matching guarantee for relatively high quality.

Equilibrium We now present the equilibrium solution and investigate the impact of PMGS on the profitability of the retailer and the supplier compared to the case where price-matching is not allowed.

Proposition 8.4 *The equilibrium price-matching strategy for the retailer and the supplier in a retail dominant chain is as follows.*

Equilibrium strategy

$$= \begin{cases} (\bar{S}, \bar{R}) & \textit{if}\, q_S \le \min\{4t-2, (3t-1)/2\}, \\ (\bar{S}, R) & \textit{if}\, \min\{4t-2, (3t-1)/2\} \le q_S \le \max\{1-t, t+1/3\}, \\ (S, \bar{R}) & \begin{cases} \textit{if} & t \le 2/3 \textit{ and } \max\{1+t, 3-3t\} \le q_S \\ \textit{or if} & 2/3 \le t \le 1 \textit{ and } t+1/3 \le q_S, \end{cases} \\ (S, R) & \textit{if}\, \max\{1-t, t+1/3\} \le q_S \le \min\{1+t, 3-3t\}, \end{cases}$$

where

$(\bar{S}, \bar{R})$ = *Neither the supplier nor the retailer offers price-matching guarantee,*

$(\bar{S}, R)$ = *The retailer offers price-matching, but the supplier is indifferent,*

$(S, \bar{R})$ = *The supplier offers price-matching, but the retailer is indifferent,*

(S, R) = *Both the supplier and the retailer offer price-matching guarantee.*

A comparison of the equilibrium solution to the case of no PMGs reveals that there are regions where both the retailer and the supplier benefit from offering price-matching guarantees, or only one of the channel members benefit from price-

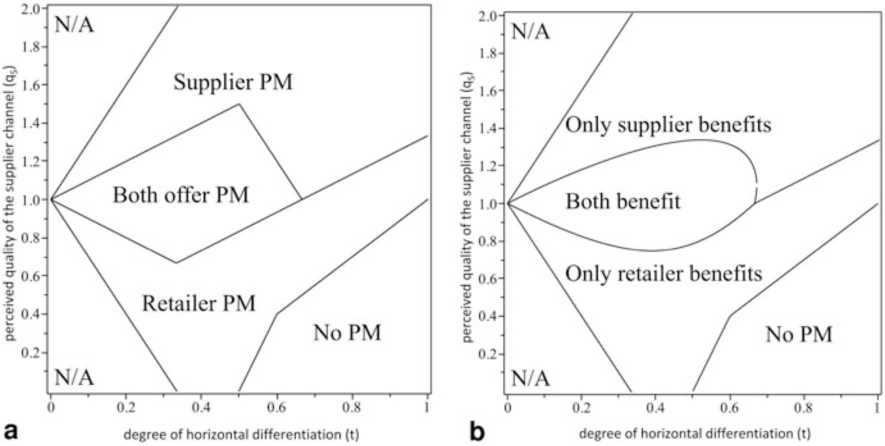

Fig. 8.5 Equilibrium in the retail dominant chain. (**a**) Equilibrium strategy. (**b**) Impact of price-matching on profitability

matching guarantees. In the former case, consumer surplus decreases compared to the no price-matching guarantee case.

The four equilibrium categories are also depicted in Fig. 8.5. In the first one, labeled *No PM*, the retailer does not offer a price-matching guarantee. The supplier is indifferent since the direct channel price is already lower than the retail channel. In this case, neither the retailer nor the supplier offers price-matching. In the second category, labeled *Retailer PM*, the retailer offers price-matching guarantee. The direct channel quality is relatively low in this region. As such, the supplier prefers to be in the low-priced channel, but the retailer's price-matching is sufficient to change the equilibrium prices. Therefore, given the retailer's price-matching guarantee, the supplier is indifferent between offering and not offering a price-matching guarantee. In the third category, labeled *Supplier PM*, we have the same phenomenon except the retailer and the supplier switch roles, i.e., the supplier offers price-matching guarantee and the retailer is indifferent. In the fourth and last category, labeled *Both offer PM*, both the retailer and the supplier offer price-matching guarantees.

There are cases in the retail dominant chain where the supplier is indifferent to offering price-matching provided the retailer is offering, or vice versa. But, until this point, we have overlooked how the price-matching offer by the retailer affects the profits of the supplier. Figure 8.5b reports on the impact of PMGs by comparing the equilibrium profits of the retailer and the supplier to the case where there are no price-matching guarantees. A closer look at the figure shows that the price-matching offer benefits both channels only in a particular region. In this region, the degree of both vertical and horizontal differentiation is low. So, the two channels are very similar to each other, and eliminating the price competition through PMGs benefits both. However, for the remaining set of problem parameters, PMGs by one of the channels hurt the profitability of the other channel. These are the cases where the two channels are sufficiently differentiated, either horizontally or vertically. For instance, suppose that the direct channel quality is low. The supplier would like to set a low price in order to attract customers to the direct channel. Offering a price-match benefits the retailer since it allows the retailer to capture customers who would otherwise visit the direct channel, which in turn hurts the supplier.

8.5.3 A Comparison of Traditional and Retail Dominant Channels

PMGs fail to prevail in a traditional dual channel. In a traditional chain, offering a price-matching guarantee becomes an important decision for the supplier only if the direct channel has a higher quality than the retail channel. In this case, in the absence of a price-matching guarantee, the supplier charges a higher price in the direct channel while encouraging the retailer to charge a low price through a low wholesale price. In essence, the supplier positions the two channels such that the direct channel enjoys the high profit margin while the retail channel enjoys the high volume. However, if the supplier offers a price-matching guarantee, she will effectively be shifting demand from the retail channel to the direct channel when indeed it is more profitable to keep demand at the retail channel. The ability to manipulate the channel prices is more valuable for the supplier than price-matching guarantees. For the retailers, offering a price-matching guarantee is also detrimental in this scenario. If the retailer offers a price-matching guarantee, then the supplier can squeeze the profit margin of the retail channel by inflating the wholesale price, and also by forcing the retailer to price-match by lowering the direct channel price.

In a retail dominant chain, all possible outcomes (both parties offer price-matching guarantee, neither offers, only the retailer offers, and only the supplier offers) are possible at equilibrium, depending on the degree of vertical and horizontal differentiation between the retail and the direct channel. Our findings suggest that, if the offerings of the two channels are alike (with a small degree of horizontal differentiation), then customers' channel (degree of vertical differentiation) preference becomes a determining factor in the benefits of price-matching guarantees. Managers of the preferred channel should capitalize on this opportunity by offering price-matching guarantees. If customers have no distinct preference over one channel, then both channels benefit from offering the guarantee.

Under these conditions, consumer welfare is negatively affected. PMGs by both channels soften the intensity of price competition and support high channel prices. As a result, the consumer surplus decreases compared to the no price-matching scenario. On the other hand, if the channel offerings are diverse (high degree of horizontal differentiation), then retail channel managers should be cautious in setting price-matching strategies. In particular, they should avoid offering PMGs when the retail channel is highly preferred, because, in this scenario, the retail channel already attracts a significant portion of the market demand.

8.6 Conclusion

Our goal in this chapter is to identify how price-matching guarantee conditions, such as verification of availability and the extent of channel coverage (traditional versus direct channels) change the value proposition of price-matching guarantees. Specifically, we shed light on the following two specific questions: (1) How does availability verification as a condition for price-matching affect the value of the guarantee for consumers and retailers? (2) What are the market conditions that can extend the coverage of price-matching guarantees to manufacturers, and does this wider coverage necessarily benefit consumers?

We find that the availability condition affects the equilibrium decisions and retail performance in completely different directions based on what factors customers take into account while making store choices. In particular, availability-contingent PMGs favor customers and decrease retail profits if customers pay attention to the *verification of availability clause*. In contrast, availability-contingent PMGs favor retailers and increase their profits while decreasing the consumers' surplus if customers fail to recognize the risks clause. In light of our findings, we conjecture that retailers adopt availability contingent PMGs when they think the majority of their customers are inattentive, i.e., they somehow ignore the possible consequences of the availability clause while making store choices.

The most interesting finding is that even though the verification of availability is seen to be a hassle for customers, it can indeed be beneficial for them if they are aware of the possibility of product unavailability and the effects thereof while making store choices. Such verification has pro-competitive effects since it increases consumers' surplus.

We also show that offering a price-matching guarantee is not beneficial and is not the equilibrium strategy in a traditional dual chain between a retailer and a supplier, even under conditions at which competing retail channels (non-dual channel) would benefit from price-matching.

Consequently, managers should be mindful of the fact that offering a price-matching guarantee is not profitable and is not the equilibrium strategy in industries where the negotiation power in pricing lies in the upper echelons of the value chain. On the contrary, a shift in the pricing power toward lower echelons, in particular to the retailer, may turn PMGs into a profitable strategy.

In a traditional chain, offering a price-matching guarantee has adverse consequences for the supplier. Promising to match the price of the retail channel limits the supplier's ability in optimally allocating the demand between the two channels. In the absence of a price-matching guarantee, the supplier, through the direct channel and wholesale prices, can transfer demand from one channel to another. Our analysis shows that, for the supplier, the benefits of allocating the demand between the two channels are higher than the benefits of offering a price-matching guarantee. Offering a price-matching guarantee in a traditional chain proves to be detrimental for the retailer. If the retailer offers price-matching, then the supplier sets the direct channel price equal to the wholesale price. The retailer is then left with zero profit margin since it cannot charge more than the direct channel price (which is equal to the wholesale price) because of price-matching. Therefore, the retailer does not offer a price-matching guarantee in a traditional chain.

Offering a price-matching guarantee allows the supplier to attract demand from the retail channel to the direct channel. In a retail dominant chain, the supplier has limited influence on the pricing of the retail channel and on the demand distribution across the two channels. Consequently, the supplier benefits from attracting customers to the direct channel via a price-matching guarantee. We show that, in a retail dominant chain, all possible outcomes (both parties offer price-matching guarantees, neither offers, only the retailer offers, and only the supplier offers) are possible at equilibrium, depending on the degree of vertical and horizontal differentiation between the retail and direct channels. Offering a price-matching guarantee, in this case, benefits the retailer and the supplier by softening the intensity of price competition. We also show that, under the condition of both parties offering price-matching, the consumer surplus is reduced compared to the no price-matching guarantee scenario.

A potential future research direction is to investigate the effects of PMGs from the supplier's point of view. Conventional wisdom suggests that suppliers enjoy enhanced competition between retailers because they would lead to lower prices and increase demand. Simple price-matching policies offered by retailers, however, inflate retail prices and decrease order quantity and, therefore, may not be the desired strategy for a supplier. On the other hand, higher retailer prices, via price-matching, may allow the suppler to increase its wholesale price. In addition, as we have shown in this paper, issues such as verification of availability and the extent of channel coverage as a price-matching condition changes the nature of interaction between supply chain members. Clearly, this stream of research can benefit from the construction of a model framework with active suppliers via endogenizing the decisions of upstream members in a distribution channel with competing retailers and manufacturers in the context of price-matching guarantees.

References

Arbatskaya M (2001) Can low-price guarantees deter entry? Int J Ind Organ 19(9):1387–1406

Arbatskaya M, Hviid M, Shaffer G (1999) Promises to match or beat the competition: evidence from retail tire prices. Adv Appl Microecon 8:123–138

Arbatskaya M, Hviid M, Shaffer G (2004) On the incidence and variety of low-price guarantees. J Law Econ 47(1):307–332

Arbatskaya M, Hviid M, Shaffer G (2006). On the use of low-price guarantees to discourage price cutting. Int J Ind Org 24(6):1139–1156

Arya A, Mittendorf B, Sappington DE (2007) The bright side of supplier encroachment. Market Sci 26(5):651–659

Aviv Y, Levin Y, Nediak M (2009) Counteracting strategic consumer behavior in dynamic pricing systems. In: Netessine S, Tang CS (eds) Consumer-driven demand and operations management models: a systematic study of information-technology-enabled sales mechanisms, vol 131. Springer, Berlin

Bell DR, Wang Y, Padmanabhan V (2006) An explanation for partial forward integration: why manufacturers become marketers. Technical report, The Wharton School, University of Pennsylvania

Belton T (1987) A model of duopoly and meeting or beating competition. Int J Ind Organ 5:399–417

Boyaci T (2005) Competitive stocking and coordination in a multiple-channel distribution system. IIE Trans 37(5):407–427

Butz D (1990) Durable-good monopoly and best-price provisions. Am Econ Rev 80(5):1062–1076

Cai GG (2010) Channel selection and coordination in dual-channel supply chains. J Retail 86(1):22–36

Cattani KD, Gilland WG, Swaminathan JM (2004) Coordinating traditional and Internet supply chains. In: Simchi-Levi DW, Shen M (eds) Supply chain analysis in the ebusiness era (International series in operations research and management science). Kluwer Academic Publishers, Dordrecht

Cattani K, Gilland WG, Heese HS, Swaminathan JM (2006) Boiling frogs: pricing strategies for a manufacturer adding a direct channel that competes with the traditional channel. Prod Oper Manag 15(1):40

Chen Z (1995) How low is a guaranteed-lowest-price. Can J Econ 28(3):683–701

Chen Z (2003) Dominant retailers and the countervailing-power hypothesis. RAND J Econ 34(4):612–625

Chen Y, Narasimhan C, Zhang Z (2001) Research note: consumer heterogeneity and competitive price-matching guarantees. Mark Sci 20(3):300–314

Chen KY, Kaya M, Özer Ö (2008) Dual sales channel management with service competition. Manuf Serv Oper Manag 10(4):654–675

Chiang WYK (2010) Product availability in competitive and cooperative dual-channel distribution with stock-out based substitution. Eur J Oper Res 200(1):111–126

Chiang WYK, Monahan GE (2005) Managing inventories in a two-echelon dual-channel supply chain. Eur J Oper Res 162(2):325–341

Chiang WYK, Chhajed D, Hess JD (2003) Direct marketing, indirect profits: a strategic analysis of dual-channel supply-chain design. Manag Sci 49(1):1–20

Choi SC (1991) Price competition in a channel structure with a common retailer. Mark Sci 10(4):271–296

Choi SC (1996) Price competition in a duopoly common retailer channel. J Retail 72(2):117–134

Cooper T (1986) Most-favored-customer pricing and tacit collusion. RAND J Econ 17(3):377–388

Corts K (1995) On the robustness of the argument that price-matching is anti-competitive. Econ Lett 47(3–4):417–421

Corts K (1997) On the competitive effects of price-matching policies. Int J Ind Organ 15(3):283–299

Coughlan A, Shaffer G (2003) Price-matching guarantees, retail competition, and product-line assortment, Simon Business School Working Paper No. FR 03-29

Dong Y, Shankar V, Dresner M (2007) Efficient replenishment in the distribution channel. J Retail 83(3):253–278

Doyle C (1988) Different selling strategies in Bertrand oligopoly. Econ Lett 28(4):387–390

Dugar S (2007) Price-matching guarantees and equilibrium selection in a homogenous product market: an experimental study. Rev Ind Organ 30(2):107–119

Dukes AJ, Gal-Or E, Srinivasan K (2006) Channel bargaining with retailer asymmetry. J Mark Res 43(1):84–97

Dukes AJ, Geylani T, Srinivasan K (2009) Strategic assortment reduction by a dominant retailer. Mark Sci 28(2):309–319

Edlin A (1997) Do guaranteed-low-price policies guarantee high prices, and can antitrust rise to the challenge. Harv L Rev 111:528–575

Fatas E, Mañez J (2007) Are low-price promises collusion guarantees? An experimental test of price matching policies. Span Econ Rev 9(1):59–77

Geylani T, Dukes AJ, Srinivasan K (2007) Strategic manufacturer response to a dominant retailer. Mark Sci 26(2):164–178

Harsanyi J, Selten R (1988) A general theory of equilibrium selection in games. MIT Press, Cambridge

Hay G (1981) Oligopoly shared monopoly and antitrust law. Cornell L Rev 67:439–481

Hess J, Gerstner E (1991) Price matching policies: an empirical case. Manag Decis Econ 12(4):305–315

Ho T, Camerer C, Chong J, Weigelt K, Chong J, Lim N (2004) A theory of lowest-price guarantees, Working paper, University of California, Berkeley, CA

Hviid M, Shaffer G (1999) Hassle costs: the Achilles' heel of price-matching guarantees. J Acad Mark Sci 8(4):489–521

Jain S, Srivastava J (2000) An experimental and theoretical analysis of price-matching refund policies. J Mark Res 37(3):351–362

Kadiyali V, Chintagunta P, Vilcassim N (2000) Manufacturer-retailer channel interactions and implications for channel power: an empirical investigation of pricing in a local market. Mark Sci 19(2):127–148

Kolay S, Shaffer G (2013) Contract design with a dominant retailer and a competitive fringe. Manag Sci 59(9):2111–2116

Kukar-Kinney M, Grewal D (2007) Comparison of consumer reactions to price-matching guarantees in internet and bricks-and-mortar retail environments. J Acad Mark Sci 35(2):197–207

Lai G, Debo L, Sycara K (2007) Impact of price matching policy on pricing, inventory investment and profit with strategic consumers. Working paper, Carnegie Mellon University, Pittsburgh, PA

Lee E, Staelin R (1997) Vertical strategic interaction: implications for channel pricing strategy. Mark Sci 16(3):185–207

Levin Y, McGill J, Nediak M (2007) Price guarantees in dynamic pricing and revenue management. Oper Res 55(1):75–97

Levy D, Gerlowski D (1991) Competition, advertising and meeting competition clauses. Econ Lett 37:217–221

Lin Y (1988) Price matching in a model of equilibrium price dispersion. South Econ J 55(1):57–69

Liu Q (2012) Tacit collusion with low-price guarantees. Manch Sch 81(5):828–854

Logan J, Lutter R (1989) Guaranteed lowest prices: do they facilitate collusion. Econ Lett 31(2):189–92

Lu Y, Wright J (2010) Tacit collusion with price-matching punishments. Int J Ind Organ 28(3):298–306

Mago SD, Pate JG (2009) An experimental examination of competitor-based price matching guarantees. J Econ Behav Organ 70(1–2):342–360

Mahar S, Bretthauer KM, Venkataramanan M (2009) The value of virtual pooling in dual sales channel supply chains. Eur J Oper Res 192(2):561–575

Manez J (2006) Unbeatable value low-price guarantee: collusive mechanism or advertising strategy? J Econ Manag Strat 15(1):143–166

Mamadehussene S (2019) Price-matching guarantees as a direct signal of low prices. J Mark Res 56(2):245–258

Moorthy KS, Fader P (2012) Strategic interaction within a channel. Retail Mark Channels (RLE Retail Distrib) 6:84

Moorthy S, Winter R (2006) Price-matching guarantees. RAND J Econ 37(2):449–465

Moorthy S, Zhang X (2006) Price matching by vertically differentiated retailers: theory and evidence. J Mark Res 43(2):156–167

Png I, Hirshleifer D (1987) Price discrimination through offers to match price. J Bus 60(3):365–383

Raju J, Zhang ZJ (2005) Channel coordination in the presence of a dominant retailer. Mark Sci 24(2):254–262

Salop SC (1986) Practices that (credibly) facilitate oligopoly co-ordination. In: Stiglitz J, Mathewson F (eds) New developments in the analysis of market structure. Palgrave Macmillan, London, pp 265–294

Sargent M (1992) Economics upside-down: low-price guarantees as mechanisms for facilitating tacit collusion. Univ Pa Law Rev 141:2055–2118

Schelling T (1980) The strategy of conflict. Harvard University Press, Cambridge

Sudhir K (2001) Structural analysis of manufacturer pricing in the presence of a strategic retailer. Mark Sci 20(3):244–264

Tirole J (1988) The theory of industrial organization. MIT Press, Cambridge

Tsay AA, Agrawal N (2004) Channel conflict and coordination in the e-commerce age. Prod Oper Manag 13(1):93–110

Useem J, Schlosser J, Kim H (2003) One nation under Wal-Mart. Fortune 147(4):65

Varian H (1980) A model of sales. Am Econ Rev 70(4):651–659

Zhang Z (1995) Price-matching policy and the principle of minimum differentiation. J Ind Econ 43(3):287–99

Chapter 9
Collaborative Micro-Retailing in Developing Economies

Luyi Gui, Christopher S. Tang, and Shuya Yin

Abstract In many developing countries, micro-retailers in remote rural places struggle to survive due to high inventory replenishment costs caused by lack of efficient infrastructure and distribution networks. Consumers, in turn, suffer from higher prices and limited accessibility of products provided by these micro-retailers. Some of these products are necessities, including food or medical items. In some cases, micro-retailers are the only sources for consumers to obtain these products. In order to help micro-retailers and the communities they serve, NGOs have been exploring different approaches aiming to coordinate the retailers' inventory replenishment strategy. This chapter explores two major types of collaborative strategies observed in practice, and studies their welfare implications for micro-retailers and local consumers. One strategy is based on an "open" cooperative where participating retailers jointly replenish inventories and share the travel cost incurred. The other strategy introduces an intermediary "non-profit" wholesaler who will consolidate retailers' orders and replenish on their behalf under a low service charge. The study unveils several key trade-offs associated with these collaborative strategies. In particular, when retailers' market entry is controlled and regulated, the cooperative strategy always leads to Pareto improvement. That is, retailers' profit improves and consumers are also better off. However, establishing a non-profit wholesaler improves retailers' profit at the expense of consumer welfare. This trade-off can be mitigated when retailers can freely enter the retail market. That is, the wholesaler strategy also leads to Pareto improvement under some general conditions. We further show that the cooperative strategy benefits the consumers more, while the non-profit wholesaler strategy is more effective in improving

L. Gui · S. Yin (✉)
Paul Merage School of Business, University of California, Irvine, CA, USA
e-mail: luyig@uci.edu; shuya.yin@uci.edu

C. S. Tang
UCLA Anderson School of Management, University of California, Los Angeles, CA, USA
e-mail: chris.tang@anderson.ucla.edu

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_9

retailers' profit and encouraging their market participation. We discuss the policy implications of these trade-offs for the deployment of collaborative replenishment for micro-retailers in practice.

Keywords Micro-retailers · Developing economies · Collaborative replenishment

9.1 Introduction

9.1.1 Micro-Retailing in Developing Countries

In developing countries, especially in remote and poor rural areas, the "last mile distribution" of life necessities such as food, water, grocery items, and medical products is oftentimes managed by small family owned stores or street carts known as the "micro-retailers." These retailers typically have less than four employees, stock a small selection of goods, and operate with a low monthly turnover, typically less than $2000 a month (Nieuwoudt 2015). Examples of micro-retailers are sari-sari stores in the Philippines and Kirana stores in India. They operate in locations close to the neighborhood and the villages, selling a variety of items such as small packets of snacks, cold drinks, cigarettes, small sachets of shampoos, and simple over-the-counter drug tablets. Many stores are simply opened inside the owners' homes, and many micro-retailers set up movable stands and travel through the community.

Micro-retailers play an important role for the economic and social development of the local area. First, they provide the daily needs of rural residents that cannot access formal markets due to remoteness and the lack of modern transportation infrastructure. For example, in India, 80% of the villages are not connected by well-constructed roads (Kumar and Gogoi 2017), and "68% of the rural market still lies untapped (by formal distribution channels) primarily due to inaccessibility" (Neuwirth 2012). In those areas, micro-retailers are indispensable. It is estimated that there are four million micro-retailers in 627,000 villages in India (Kumar and Gogoi 2017). In Africa, the distribution of medical products heavily relies on small private drug retailers in rural areas (Goodman et al. 2004; Yadav 2013). Second, micro-retailers can provide additional services such as phone and Internet access, card games and Karaoke, even coin laundry and haircut. Hence, these small stores often act as the social, communication, and recreational center of the community served. Third, micro-retailing brings in additional money for the store owners. It is considered a major component of the non-farm economy in poor rural areas and an important way to create jobs and business opportunities (International Fund for Agricultural Development 2001).

Despite the importance of micro-retailing in developing economies, micro-retailers earn meager profits and operate under prohibitive operating constraints. First, the micro-retailing sector is highly fragmented, and individual retailers have no bargaining power to negotiate better wholesale prices. Second, in the absence of formal distribution channels, micro-retailers have to travel afar by themselves to replenish stocks. For example, private drug retailers in the rural areas of Tanzania are reported to travel 150–400 km with few paved roads for replenishment from suppliers in a city (Goodman et al. 2004). This results in a high travel cost burden (including the imputed cost of time, labor, and inconvenience) on micro-retailers. Third, micro-retailers are typically cash-strapped and have no access to financial support (such as loans). Neither do they have the knowledge, technology support, or even enough storage space, to implement sophisticated strategies such as inventory planning. In addition, micro-retailers may need to survive in a harsh environment where they suffer from stealing and harassment from criminal groups. Far away from formal markets and distribution channels, oftentimes, they also face unfair competition from counterfeit products that are fake or unauthorized replicas of the genuine products and are sold at much cheaper prices.

It is evident that the survivalist model of micro-retailing in developing economies is in sharp contrast to the rocket-science model of modern retailing that the literature has concentrated on. The modern retailing literature is featured by big, powerful retailers, such as Walmart and Amazon that are equipped with state-of-the-art logistics/operational management capabilities and have access to good resources. Hence, the challenges faced by micro-retailers motivate a rich set of new research avenues. From an operational point of view, a major question is how to improve replenishment efficiency for micro-retailers. In the literature, replenishment/distribution channel strategies have been extensively discussed. For example, a widely studied strategy is the vendor-managed inventory (VMI) model where a manufacturer jointly replenishes orders at multiple retailers (Lee et al. 2000; Çetinkaya and Lee 2000; Mishra and Raghunathan 2004). Another popular strategy is the group-buy initiative among competing retailers (e.g., Chen and Roma 2011). However, despite the ample discussion on replenishment strategies, these studies focus on developing sophisticated solutions that may not be applicable to the resource-constrained micro-retailers. For example, the multi-period inventory planning model is often used for well-established and operationally strategic retailers that are quite different from the micro-retailers considered in this chapter. Moreover, the literature on replenishment strategies typically takes the perspectives of profit-maximizing entities, e.g., the manufacturer in the VMI model, whereas social benefit metrics such as consumer welfare and retailer's market participation are ignored. Another example is the stream of research that studies distribution channel structure of suppliers. This stream focuses on "direct sell" strategies of suppliers, where they establish their own distribution channel to vertically compete with common retailers, aiming at obtaining a higher profit (e.g., Cattani et al. 2004; Tsay and Agrawal 2004). However, such a "value extracting" perspective may not be desirable in the context of developing economies. In particular, inefficient replenishment in micro-retailing creates a vicious cycle: poor micro-retailers cannot earn much and few of them can survive in the market. This leads to higher retail prices that poor consumers cannot afford, which in turn further hurts micro-retailers profits (International Fund for Agricultural Development 2001; OECD 2013). To break this cycle, merely focusing on profit-maximization of micro-retailers is not enough, and a win–win solution that improves both retailer profit and consumer welfare (e.g., via encouraging market participation and reducing retail price) is highly beneficial. In other words, the developing economy context calls for a "value creation" standpoint, an important ingredient of the emerging stream of research on socially responsible operations at the bottom of the pyramid, e.g., those that develop sustainable energy and transportation solutions (e.g., Uppari et al. 2017; McCoy and Lee 2014; Chen et al. 2017). Yet similar research on micro-retailing remains limited. In this chapter, we provide an example that analyzes the use of collaborative strategies in rural areas to help micro-retailers reduce travel cost associated with replenishment trips and at the same time, improve consumer welfare.

9.1.2 Collaborative Replenishment Strategies

As aforementioned, high travel cost is a major burden for micro-retailers that hinders the development of remote rural areas. This has motivated various NGOs, non-profit organizations, and social enterprises to develop novel and effective replenishment strategies for micro-retailers. Many of these strategies involve encouraging collaboration among multiple micro-retailers to replenish together and reduce the travel cost. Two major examples stand out in practice: purchasing cooperatives and non-profit wholesalers. In a purchasing cooperative, participating retailers replenish their stocks jointly and share the travel cost. Take ShopRite, a major retailer cooperative in the USA, as an example. The ShopRite cooperative "buys, warehouses and transports" products through its own procurement arm, the Wakefern Food Corp., on behalf of its fifty members, all of which are individually owned and operated local stores (ShopRite 2017). The cooperative strategy has a long history in the retailing sector and is known for reducing purchasing and transportation costs. In Europe, purchasing cooperatives among independent retailers have become some of the largest organizations in retail sectors, e.g., Rewe and Edeka in Germany, Intermarche, E. Leclerc and System U in France, and El Corte in Switzerland (von Ravensburg 2011). Purchasing cooperatives have already been proven to be effective in developing countries, in particular, among individual farmers in the agriculture sector (e.g., Rabobank Foundation, Kilimanjaro Native Cooperative Union) and among individual local craft makers (e.g., Machakos District Co-op Ltd.). There also exist ample interest and opportunities to apply this concept to micro-retailing, according to the International Labor Organization (von Ravensburg 2011).

Next, the non-profit wholesalers are a variation of common wholesaler services. They serve as replenishment hubs by transporting goods to places that are easily accessible to micro-retailers. These wholesalers are established to help the rural population, and therefore charge low wholesale prices, just enough to sustain their operations, instead of profit maximization. Under this model, participating retailers "implicitly" share the travel cost by paying the surcharge in the wholesale price that jointly covers the travel cost. It is different from the purchasing cooperative strategy because the wholesaler is an independent intermediary. In practice, non-profit wholesalers are mostly concentrated for medicines and drugs, and are typically geared towards health care facilities, e.g., Action Medeor and MEDS that operate in African countries such as Kenya, Uganda, and Tanzania (MEDS 2017). In India, the Delhi Cooperative Societies Act has led to the establishment of the Delhi Consumer's Cooperative Wholesale Store Ltd that supplies consumer goods to affiliated stores and to the public at low prices (Govt. of NCT of Delhi 2017). In the private sector, Tata Global Beverages Ltd has launched a non-profit project called "Gaon Chalo Abhiyan" (Let's Go To the Villages) that provides tea products to local retailers at low costs in order to yield an attractive margin to shopkeepers (Tata Trust 2016). Some for-profit wholesalers have also organized low-price wholesaling services, tailored for remote rural areas. Examples include the rural distribution service provided by Drishtee, and the Hindustan Lever Limited's (HUL) project "Shakti" (Rangan and Rajan 2007). It is evident through these examples that non-profit wholesaler can be an effective way to help micro-retailers to overcome the travel cost hurdle that disconnects them, and therefore their rural communities, from the formal market.

In this chapter, we analyze the economic implications of the purchasing cooperative and the non-profit wholesaler replenishment strategies for remote rural areas in developing countries. The analysis is based on stylized operations management models that capture (1) the impact of travel cost on micro-retailers' profit and their survival in a resource-constrained operating environment, (2) market participation decisions of micro-retailers and the competition among them, (3) consumer choice among multiple micro-retailers and the resulting consumer welfare implications. The goal of the analysis is threefold. First, it outlines the key economic mechanism underlying these two strategies. Second, it enables an equilibrium study that quantifies the effectiveness of these strategies in improving micro-retailer profit and consumer welfare. Third, it helps identify market factors that critically affect the relative performance between the two strategies and, accordingly, provides policy insights as to the choice and design of collaborative replenishment programs in rural areas in developing countries.

The rest of the chapter is organized as follows. Section 9.2 introduces models of the two collaborative strategies discussed in Sect. 9.1.2 and also the benchmark strategy where there is no coordination among the micro-retailers. Section 9.3 analyzes the market equilibrium under these strategies and presents a comparative study. We conclude the chapter with managerial insights in Sect. 9.4.

9.2 Model Description

In practice, there may be multiple micro-retailers that serve the same community. Hence, we consider n micro-retailers in the market, each selling a product. We note a few key operational features of micro-retailers in practice. First, these retailers are typically substitutable because they sell similar products and use similar sales format, yet they also differentiate in certain ways. For example, micro-retailers often carry different brands of similar types of products, such as packaged snacks. They may also offer different services such as phone and Internet access, card games/Karaoke, and even services such as coin laundry, water refilling, and haircut—to attract customers and improve store demand. Second, micro-retailers are symmetric in many other aspects such as size, budget, and cost structure (Prahalad 2006). Third, micro-retailers usually compete on price. These retailers have the pricing power as they often have close connections with the local community, e.g., via services such as home delivery and store credits to customers (Fransoo et al. 2017) and act as "local monopolies." (There are also cases where micro-retailers act as price takers. We discuss those cases in a quantity competition extension in Sect. 9.4.) In view of these observations, we assume that, under any given inventory replenishment strategy, retailers set their own retail prices, denoted by $\mathbf{p} = \{p_1, p_2, \ldots, p_n\}$. Customers then decide on which retailer they will buy from and their purchasing quantities. This determines the demand of each micro-retailer, which we model by a demand function $\mathbf{d} = \{d_1, d_2, \ldots, d_n\}$. We follow a well-established framework in the economics literature that models consumer choices under price competition (e.g., Anderson et al. 1992; Dixit 1979; Singh and Vives 1984). In this framework, a representative consumer decides to buy the amount d_i from retailer i, for each $i \in \{1, 2, \ldots, n\}$, in order to maximize its overall welfare, given the retailers' prices and their substitution levels. A widely used consumer welfare function is expressed by

$$U(\mathbf{p}, \mathbf{d}, n) = \sum_{i=1}^{n}\left(d_i - \frac{d_i^2}{2}\right) - \tau \sum_{i=1}^{n} d_i \left(\sum_{j=1, j\neq i}^{n} \frac{d_j}{2}\right) - \sum_{i=1}^{n} p_i d_i, \tag{9.1}$$

where the parameter $\tau \in (0, 1)$ measures the substitution level between any two products sold in the market. Intuitively, for example, where there are two retailers i and j in the market, retailer i is τ-substitutable to retailer j if a unit price change at j results in a demand change at i that equals the fraction τ of that caused by i's own unit price change. The welfare function (9.1) will be used later to evaluate the consumer welfare under each replenishment strategy, which is one of the dimensions based on which we evaluate the performance of these strategies.

Consumer welfare maximization based on function (9.1) yields the following demand function:

$$d_i(n, \mathbf{p}) = a - bp_i + \theta \sum_{j=1, j\neq i}^{n} (p_j), \tag{9.2}$$

where $a = 1/[(n-1)\tau + 1]$, $b = [(n-2)\tau + 1]/[(1-\tau)((n-1)\tau + 1)]$ and $\theta = \tau/[(1-\tau)((n-1)\tau + 1)]$.

Knowing the demand function above, micro-retailers engage in price competition in the market to maximize their own profits that are contingent on the inventory replenishment strategy implemented. For example, without collaboration, individual retailers make their own independent trips to replenish inventories. We call this case the *independent strategy*, and use it as the benchmark when evaluating the performance of the two collaborative replenishment strategies. Note that when a trip made to the supplier site for inventory procurement, a fixed travel cost, denoted by K, and a variable unit procurement cost, denoted by c, are incurred. Since micro-retailers, in practice, share similar cost structures, we assume that these costs are the same for all retailers in our model. Moreover, for tractability, we also assume that the fixed travel cost remains the same, regardless of whether the trip is made by individual retailers, the purchasing cooperative, or the non-profit intermediary (wholesaler). The justifications for this simplification are discussed in Gui et al. (2019). That paper also extends the fixed travel cost to more general structures that increase in the number of participating retailers, since a bigger group of retailers might require a bigger vehicle to transport goods and hence incur a higher operating cost. Under these assumptions, each retailer i that replenishes independently has a profit function that can be expressed as follows:

$$\Pi_i(\mathbf{p}, n) = (p_i - c) \cdot d_i - K, \tag{9.3}$$

where the demand function is given in Eq. (9.2) above. Since individual retailers simultaneously set selling prices, the Nash concept is adopted to characterize the prices in equilibrium.

Collaborative replenishment strategies help reduce the retailers' travel cost. Essentially, under the *open cooperative strategy*, only one replenishment trip is needed to the supplier site for all member retailers' procurement needs and the fixed travel cost is incurred once. In practice, that travel cost depends on factors such as the number of participating retailers, the actual location of these retailers, and how routing choices are made. In this study, we assume that the travel cost to commute between retailers is negligible, compared to that from where these retailers are situated in the local remote community to the supplier site that is typically in large cities. This assumption is supported by the fact that the remoteness of micro-retailers from the main market where the suppliers are located is the major cause of their travel cost burden. Under this assumption, the travel cost that the cooperative incurs can be modeled as K as well. We assume that this fixed cost is equally shared among the members of the cooperative. This can be justified by the symmetry among the individual retailers due to the similarities of their business scales and the packaged product types (of possibly different brands) that they carry (Prahalad 2006). As a result, each retailer i's profit function can be expressed as follows:

$$\Pi_i(\mathbf{p}, n) = (p_i - c) \cdot d_i - \frac{K}{n}. \tag{9.4}$$

Similarly, the Nash concept is adopted to characterize the prices in equilibrium.

Under the *non-profit wholesaler strategy*, the wholesaler procures products on behalf of all the retailers and incurs the fixed travel cost K. An important feature of a non-profit wholesaler that differentiates itself from for-profit ones is that it aims to help local retailers and thus typically charges them a cheaper wholesale price, mainly to cover operational expenses such as the travel cost incurred in replenishment trips. To capture this feature, we assume in this study that the non-profit wholesaler charges a price premium on top of the unit purchasing cost c, just enough to break even and cover the travel cost K. We denote the resulting unit wholesale price as w. Based on this wholesale price w, individual retailers set their selling prices accordingly. The interaction between the wholesaler and the retailers is formulated as a two-stage Stackelberg game, where the wholesaler is the game leader and the retailers are the followers. Backward induction is adopted to solve for the equilibrium decisions. Specifically, in stage 2, given w, each retailer i's profit function can be expressed as follows:

$$\Pi_i(\mathbf{p}, n, w) = (p_i - w) \cdot d_i. \tag{9.5}$$

In this stage, the Nash approach can again be applied to derive the retailers' equilibrium selling prices as functions of the wholesale price, $p_i(w)$.

In stage 1, in anticipation of retailers' prices $p_i(w)$, the wholesaler sets the wholesale price to break even and just cover the fixed cost incurred during the procurement process. Technically, the wholesale price, w, is set to achieve the following equality:

$$(w - c)\sum_{i=1}^{n}[d_i(p_i(w))] = K. \tag{9.6}$$

The extension of the model to the case where the wholesale price is determined for a for-profit wholesaler can be found in Gui et al. (2019).

9.3 Model Analysis and Comparison

In this section, we first analyze the retailers' equilibrium prices that will subsequently facilitate the derivation of their individual profits in equilibrium and the consumers' overall welfare. The quantification of these terms would enable us to compare the three replenishment strategies from the perspectives of both the retailers and the consumers. The following result presents the retailers' selling prices in equilibrium. A detailed proof is available in Gui et al. (2019).

Proposition 9.1 (Equilibrium Retail Prices) *In a market with n competing micro-retailers:*

(1) *With either the independent replenishment strategy or the open purchasing cooperative, the equilibrium prices are as follows:*

$$p(n) = c + \frac{(1-c)(1-\tau)}{(n-3)\tau + 2}. \tag{9.7}$$

(2) *With the wholesaler strategy, the equilibrium prices are as follows:*

$$p(n) = w(n) + \frac{(1 - w(n))(1 - \tau)}{(n - 3)\tau + 2}, \tag{9.8}$$

where

$$w(n) = \frac{1}{2}(1 + c - \sqrt{(1 - c)^2 - 4L(n)}), \quad \textit{and}$$

$$L(n) = \frac{K[(n - 3)\tau + 2][(n - 1)\tau + 1]}{n[(n - 2)\tau + 1]}.$$

Based on the equilibrium prices identified in Proposition 9.1, the retailers' equilibrium profits and the overall consumer welfare (calculated by Eq. (9.1)) can be subsequently computed. We denote by $\Pi_i^*(n)$ the retailer's equilibrium profit when there are n retailers in the market.

Proposition 9.1 provides a few insights. First, we observe that, regardless of the inventory replenishment policy, the retailers always adopt a price markup on top of the unit purchasing costs (or the unit wholesale price in the non-profit wholesaler model) that they are charged. Second, it is straightforward to note that the cooperative strategy does not change the pricing game among retailers in the independent model and it only reduces the individual retailers' fixed travel cost from K in the independent model to K/n. So, the equilibrium prices in the independent and the cooperative models are identical.

Compared to the cooperative strategy, the wholesaler strategy presents a more sophisticated way to reduce the travel cost for the participating retailers and hence introduces model complexity which is reflected by its equilibrium price outcome. Specifically, note that, in equilibrium, $w(n) \geq c$ holds, which implies that the retailers pay a higher unit price under the wholesaler model than the regular unit purchasing cost under the cooperative and the independent strategies. We conclude that, even though the wholesaler (as an intermediary) is not-for-profit, its presence in the system increases the retailers' variable cost which we show later leads to higher retail prices.

Based on the equilibrium price characterization under different replenishment strategies in Proposition 9.1, we are now ready to evaluate the effectiveness of the two collaborative strategies in improving the retailers' profit and consumer welfare relative to the independent one. Various examples discussed in the introduction indicate that there are two practical settings with different setup complexities that could affect our comparative study. One setting is where government authorities regulate a fixed number of micro-retailers in the market. From a modeling perspective, this implies that the number of retailers, n, is exogenously chosen, typically to guarantee a certain level of individual retailers' profit. This setting is referred to as the *regulated market*. The other setting is where individual retailers can freely enter or exit the market as long as their profits meet a reservation value (i.e., the lowest profit that a micro-retailer is willing to stay in business). In other words, the value of n is endogenously determined through an individual rationality constraint on retailers' profit. This setting is referred to as the *unregulated market*. In what follows, we will separately analyze these two practical settings. For consistency, we assume that in both settings, the retailers' reservation value is v (≥ 0) above which they are willing to enter or stay in the market.

9.3.1 Regulated Market

Government may regulate the market entry of retailers through business licenses. For example, in many developing countries, such as China, the Philippines, or India, micro-retailers that sell food, tobacco, or liquor products are required to apply for specific permits before doing business and these permits can be strictly controlled (e.g., Shanxi Province Administration for Industry and Commerce 2007). Typically, government monitors the number of such permits (and subsequently the number of retailers in the market), so that the survival (or the participation) of the retailers can be ensured, i.e., their profits are at least equal to or above their reservation value v. Hence, the analysis in this section assumes that the model parameters (τ, c, K) meet the condition that the retailers' equilibrium profit under any of the three replenishment strategies is guaranteed to be not smaller than v. Proposition 9.2 summarizes the results from the comparison between the three strategies in terms of their equilibrium prices, retailers' profits, and consumer welfare.

Proposition 9.2 (Comparison of the Three Strategies in the Regulated Market) *Assuming that there are at least two retailers in the market, the following results hold.*

(1) *Retail Prices: The wholesaler strategy leads to the highest equilibrium prices compared to the cooperative and the independent strategies, which lead to the same prices.*
(2) *Retailers' Profits: Relative to the independent strategy, both cooperative and wholesaler strategies improve retailers' profits. Moreover, the wholesaler strategy benefits the retailers more than the cooperative strategy, if the substitution level τ is higher than a certain threshold (indicating that retail competition in the local area is sufficiently strong).*
(3) *Consumer Welfare: The wholesaler strategy leads to the lowest consumer welfare, compared to the cooperative and the independent strategies, which lead to the same level of consumer welfare.*

Proposition 9.2 provides a set of useful insights about the economic implications of collaborative replenishment strategies. In terms of the retail prices, it is not surprising to observe that the wholesaler strategy leads to the highest retail prices in equilibrium. This is due to the fact that, under such a strategy, the retailers need to pay a higher wholesale price in equilibrium than the unit purchasing cost that they need to pay under the independent or the cooperative strategy (see the discussion following Proposition 9.1). So, the retailers transfer some of this cost increase to their customers via higher retail prices. This effect subsequently leads to different retailer profit and consumer welfare implications of these replenishment strategies, which we discuss in the next paragraph.

In terms of the retailers' profits, relative to the independent strategy, the cooperative strategy benefits the retailers since it keeps their gross profits (without considering the fixed cost) at the same level, while reducing their individual fixed costs from K to K/n. On the contrary, the wholesaler strategy reduces the retailers' gross profits (due to increased retail prices) but relieves them from paying the fixed travel costs, and the net impact improves retailers' profit as compared to that in the independent strategy case overall. The question now is which one of the two cooperative strategies is more effective in improving the retailers' profits. To facilitate this comparison, let us first transform the retailers' profits in the wholesaler model as follows: We substitute the equilibrium retail prices in Eq. (9.8) in the retailers' profit functions in Eq. (9.5) and apply the wholesaler's break-even function in Eq. (9.6), which yields an alternative retailer profit function under the wholesaler strategy:

$$\Pi_i(\mathbf{p}, n) = (p_i - c) \cdot d_i - \frac{K}{n}. \tag{9.9}$$

This transformation implies that the wholesaler model essentially enables the retailers to "share" the fixed travel cost indirectly, just like in the open cooperative model. Hence, the comparison of the wholesaler and cooperative models reduces to the comparison of the retail prices and the realized demands in these two models. On one hand, Proposition 9.2(1) indicates that the wholesaler model leads to higher retail prices, and subsequently, a higher unit profit margin. On the other hand, it also implies that the cooperative strategy has an advantage on inducing higher demand (due to lower retail prices). As the degree of retail substitutability increases, the wholesaler model's profit margin advantage is amplified, while the cooperative model's demand advantage is reduced. As a result, the cooperative strategy works more effectively in improving the retailers' profits when the substitution level and the competition intensity in the market are low. When retailers compete more strongly, the wholesaler strategy will be preferred by the retailers. This explains Proposition 9.2(2).

From a practical point of view, Proposition 9.2(2) implies that retailers would favor the open cooperative framework if they are highly differentiated via, e.g., selling different products, or being located with a relatively long distance. Otherwise, they would prefer a non-profit intermediary. Further analysis in Gui et al. (2019) provides a complete characterization of the sufficient lower bound on the substitution level τ, under which retailers' preferences switch. It is shown that this lower bound is no larger than 0.3 and decreases quickly as the number of retailers in the market increases. For example, in India's rural market, where a village is served by six micro-retailers on average (Kumar and Gogoi 2017), the wholesaler strategy leads to higher retailer profit, as long as the substitution level is no smaller than 0.02, a practically plausible condition. This discussion indicates that the wholesaler strategy is likely to be preferred by the micro-retailers in practice.

Finally, in terms of the consumer welfare, the cooperative and the independent strategies are again equivalent because they lead to the same equilibrium retail prices. However, since the wholesaler strategy gives rise to the highest prices among the three strategies, it generates the lowest consumer welfare level and may be the least preferred from the perspectives of local consumers.

Following the discussions above, it is reasonable for us to conclude that in a regulated market, for the retailers, collaborative replenishment via a non-profit wholesaler is typically more beneficial than that via an open cooperative. However, the introduction of an intermediary wholesaler is detrimental to the consumer welfare, even if it is not-for-profit. This trade-off implies that the selection of the best collaborative replenishment strategy needs to factor in the policy makers' priorities. For example, in the pharmacy market, where the number of drug retailers is strictly controlled, the policy maker needs to be mindful that the wholesaler model can benefit the retailers the most, but it undermines the consumer welfare. In situations where the consumer welfare is the top concern, the alternative cooperative strategy, where a purchasing cooperative is formed among the existing retailers, should be considered and encouraged.

9.3.2 Unregulated Market

In the previous section, we have examined the effectiveness of collaborative replenishment strategies in improving retailers' profit and consumer welfare in a regulated market. Note that the regulated market setting applies to situation where the number of retailers in the market is strictly controlled by the government for a variety of reasons (e.g., Govt. of NCT of Delhi 2017; MEDS 2017). For example, in the case of micro drugstores, market entry restriction may be imposed to mitigate market competition so as to ensure the survival rate of these stores and hence to maintain a sustainable market. The market could also be regulated to facilitate quality control (see the discussion in the introduction section). However, there are many cases in practice where retailers can *freely* enter or exit the market depending on their own incentives. In particular, they would enter the market if they obtain a profit above their reservation value (e.g., the profit that could be achieved from other business alternatives, i.e., the opportunity cost for these retailers). On the other hand, existing retailers may choose to exit the market if there are too many competing retailers and their profit falls below the reservation value. In the free market entry setting, the number of retailers in the market is endogenously determined in equilibrium by the operating environment characterized by our model parameter specifications. The free market entry setting applies to situations where the retailing market is loosely monitored by the government, or where the rules and regulations that govern the market entry of the retailers are weakly enforced, as it is the case in many developing countries. There could also exist situations where the government intends to introduce more competition in the retailing market in the hope of bringing down retailing prices and benefiting the consumers. Allowing retailers' free market entry could be an applicable strategy to that end.

In the unregulated setting, it is natural for us to assume that the individual retailers will first decide whether or not to enter the market by taking into consideration their reservation profits. We model this decision as stage 1 of the game. After the number of participating retailers is determined, they would subsequently engage in price competition by setting their retail prices simultaneously and then their equilibrium profits will materialize. We model these events in stage 2 of the game. We adopt the classic backward induction to solve the equilibrium solution to the game. We first solve for the retailers' equilibrium profit functions $\Pi_i^*(n)$, assuming there are n participating retailers following the analysis discussed in Sect. 9.3. Next, we calculate the equilibrium value of n, defined as the highest possible *integer* value of n such that

$$\Pi_i^*(n) \geq v. \tag{9.10}$$

We then compare the three replenishment strategies, based on the equilibrium number of retailers they induce in the market.

Note that in the unregulated market, there is no need to compare the individual retailers' equilibrium profits, since, given how the equilibrium n is defined, the retailers' profits will equal their reservation values in equilibrium under all strategies. However, the free market entry setting gives rise to another welfare metric, i.e., the number of participating retailers in the market in equilibrium. The number of participating retailers in equilibrium can be an important performance measure of the micro-retailing market for developing economies. This is because more participating retailers imply more sources of additional income for the poor rural people and create business opportunities for the non-farm economy in the local area, which could be crucial for developing countries (International Fund for Agricultural Development 2001). Hence, in the unregulated market case, we also compare the equilibrium number of retailers in the market under the three replenishment strategies considered. Unfortunately, it is quite challenging to derive a closed-form solution to Eq. (9.10), especially in the wholesaler model. This is due to the complicated form of the best response functions of the retail and wholesale prices, given the number of participating retailers in the market (see Proposition 9.1(2)). This in turn creates difficulty in applying the floor operation to the solution to Eq. (9.10) to derive the integer value of the number of participating retailers. Good news is that part of the comparison can be done analytically based on an approximation method proposed in Gui et al. (2019). That is, the direct solution of n to inequalities (9.10), $\Pi_i^*(n) \geq v$, is used in the comparison. Accordingly, the following result can be derived.

Proposition 9.3 (Comparison of the Three Strategies in the Unregulated Market) *When the retailers can freely enter or exit the market, we have:*

(1) *Number of Participating Retailers: The open cooperative strategy always leads to more retailers participating in equilibrium than the independent strategy. Among all three strategies considered, the wholesaler strategy leads to the highest number of retailers in the market if*

$$\frac{(1-\tau)K^2 + (1+\tau-\tau^2)Kv + \tau v^2}{(1-\tau)K + \tau v} \leq \left(\frac{1-c}{2}\right)^2. \tag{9.11}$$

(2) *Retail Prices: Among all the three strategies, the open cooperative strategy leads to the lowest retail prices in equilibrium.*

(3) *Consumer Welfare: The open cooperative strategy always leads to higher consumer welfare in equilibrium than the independent strategy. Among all the three strategies, the cooperative strategy leads to the highest consumer welfare if condition (9.11) holds.*

Proposition 9.3(1) indicates that, relative to the base strategy, the open cooperative strategy always encourages more retailers to enter the market. This is not surprising since the cooperative strategy reduces the retailers' travel cost from K to K/n. However, it is not trivial to observe that the wholesaler strategy can lead to even more participating retailers than the open cooperative if condition (9.11) holds. Hence, it is helpful to understand the implications of that condition in detail. Note that in order to ensure a reasonable market environment, we set feasible ranges for model parameters, K, c, and v, so that there is at least one retailer who would be willing to enter the market when replenishing independently. Under this assumption, it can be shown that condition (9.11) holds in more than 85% of the feasible ranges of the model parameters. In the extreme case when $\tau \rightarrow 1$ (i.e., retailers are fully substitutable) or when $\tau \rightarrow 0$ (i.e., retailers are completely differentiated), condition (9.11) always holds. In addition, if the retailers' reservation profit is set at a zero level (e.g., when there are no alternative income options for the micro-retailer, which can be true for the poor residents in some developing countries), condition (9.11) always holds as well. In all, this analysis indicates that condition (9.11) is a practically plausible condition. Accordingly, it can be concluded that in general, the wholesaler strategy is more effective in encouraging more retailers to participate in the market. It is only when the reservation profit is sufficiently high or the fixed travel cost is sufficiently large that the condition (9.11) may be violated, in which case the open cooperative strategy may induce more participating micro-retailers. This result is in line with Proposition 9.2(2), as both suggest that the retailers are better off under the wholesaler strategy, compared to the cooperative strategy under general conditions in both the regulated and the unregulated markets.

Propositions 9.3(2) and 9.3(3) leads to a number of implications. First, the open cooperative strategy warrants the lowest retail prices among all the three strategies, and accordingly leads to the highest consumer welfare when condition (9.11) is satisfied. It is interesting to note that under the same condition, Proposition 9.3(1) indicates that the wholesaler model leads to the highest number of retailers in the market (which also implies the highest total profit for all the retailers, nv). This observation enriches the intuition obtained from the analysis of the regulated market. That is, when retailers are allowed to enter and exit the market freely, the wholesaler strategy continues to improve the retailers' profit the most, whereas the cooperative strategy continues to be the best for consumer welfare.

Second, Propositions 9.3(1) and 9.3(3) combined indicate that the cooperative strategy is Pareto improving compared to the independent strategy. A natural question is whether this holds true for the wholesaler strategy. To analyze this question, recall that Proposition 9.2 demonstrates that in a regulated market, the wholesaler strategy raises the retail prices and hence undermines the consumer welfare relative to the independent strategy. However, how these effects may change when retailers' free market entry is permitted is unclear. So, it would be meaningful to characterize the conditions under which the wholesaler strategy can be Pareto improving in an unregulated market. On the other hand, as discussed right before Proposition 9.3, it is difficult to obtain closed-form expressions for the retailer profit and consumer welfare in equilibrium for the wholesaler strategy under free market entry. To overcome that difficulty, the numerical study in Gui et al. (2019) shows a number of observations regarding the comparison between the independent and the wholesaler strategies.

First, for the set of instances where the number of participating retailers is higher under the wholesaler strategy, the consumer welfare tends to be higher as well. This positive correlation between the number of retailers and the consumer welfare suggests that the free market entry in the unregulated market can mitigate the supply chain inefficiency caused by the introduction of the intermediary wholesaler in the regulated market where the number of retailers is fixed.

Second, we observe that the wholesaler strategy improves both the number of participating retailers and the consumer welfare, relative to the independent strategy when $K + v$ is small, K/v is large, and τ is small. Indeed, following the discussion on Proposition 9.3, in particular, condition (9.11), we understand that the smaller values of $K + v$ and τ yield a higher number of participating retailers in equilibrium under the wholesaler strategy. In addition, it can be shown that for any given set of $(c, \tau, K + v)$, the equilibrium number of retailers in the market under the wholesaler strategy is increasing in K/v. This explains why the conditions under which the wholesaler strategy is Pareto improving hinges on the parameter specifications that are previously discussed.

9.4 Discussions

In the previous sections, we have made a number of simplifying assumptions. In this section, we will briefly discuss the implications of relaxing two main assumptions: the nature of retail competition and the uncertainty in demand.

First, we assumed that the micro-retailers are the price-setter and are engaged in price competition in the market. In practice, many of these micro-retailers have strong bounds with their local communities (Fransoo et al. 2017), which differentiates them from each other and grants them local monopolist power. However, if these micro-retailers sell commodity type of products which are very similar and they are located closely to each other, then it might be more reasonable to assume that the retailers engage in quantity competition. Our analysis of this extension leads to the following observations on the impact of a different type of retail competition (see Observations 2 and 3 in Gui et al. 2019, for details).

1. In the regulated market, the effect of the cooperative strategy on retailers' retail prices, profits, and consumer welfare (relative to the independent strategy) remains robust under either quantity or price competition. The positive impact of the wholesaler strategy that improves retailer profit and its negative impact that undermines consumer welfare under price competition is mitigated under quantity competition. In some cases, the introduction of a non-profit wholesaler may even benefit the consumers under quantity competition.
2. In the unregulated market, relative to price competition, quantity competition encourages even more participating retailers in the market when either an open cooperative is formed or an intermediary non-profit wholesaler is introduced into the system. However, the impact of quantity competition on the relative effectiveness between the two collaborative strategies is highly sensitive to the specific operating parameters.

An important take-away from the above two observations is that the type of market condition (regulated vs. unregulated) influences the implications of quantity vs. price competition on the effectiveness of collaborative replenishment strategies. In particular, in a regulated market the benefits of the wholesaler strategy relative to the independent strategy are *reduced* when the retailers compete on quantities instead of prices. This is because quantity competition gives retailers more "monopolist" power than price competition if the number of competing retailers is fixed (Singh and Vives 1984), and leads to a higher profit under any replenishment strategy. Accordingly, Gui et al. (2019) show that quantity competition and collaborative replenishment work as substitutes in improving retailers' profit. However, when retailers can freely enter the market, the benefits of the cooperative or the wholesaler strategy relative to the independent strategy are *reinforced* when retailers change from price to quantity competition. In that case, quantity competition and collaborative replenishment turn out to be complements in encouraging retailers' market participation.

Second, we assumed so far that the demand is deterministic for individual retailers. This setting is applicable to products with predictable and stable demand, e.g., everyday grocery items. For products such as drugs or medical items, uncertainty in demand might be unavoidable which, in turn, raises the issue of possible mismatch between demand and supply. Given this consideration, we revise our main model to incorporate demand uncertainty for micro-retailers. For consistency with the main model, the extension to demand uncertainty is conducted under the assumption that the retailers engage in price competition. This analysis yields the following observations on the impact of demand uncertainty (see Observations 4 and 5 in Gui et al. 2019, for details).

1. In the regulated market, the effect of the cooperative strategy on retailers' prices, profits, and consumer welfare (relative to the independent strategy) remains robust under demand uncertainty. However, similar to the quantity competition case, the positive impact of the wholesaler strategy that improves retailer profit and its negative impact that undermines consumer welfare under price competition is mitigated when demand becomes uncertain.
2. In the unregulated market, the introduction of demand uncertainty would weaken the effectiveness of the open cooperative or the wholesaler strategies in motivating more participating retailers in the market relative to the independent strategy. The way this uncertainty changes the effectiveness of these two collaborative strategies on consumer welfare is highly sensitive to the specific operating parameters.

To intuitively understand the above findings, we note that according to Bernstein and Federgruen (2005), for a given number of retailers competing in the market, the uncertainty in demand leads to mismatch between demand and supply and hence reduces retailer profit under any replenishment strategy. This poses an additional challenge for the retailers to be better off via an intermediary player, and the benefits of the wholesaler strategy (relative to the independent strategy) are weakened in the presence of demand uncertainty. The same effect is also observed in an unregulated market setting, with free entry through the equilibrium number of participating retailers. That is, the wholesaler strategy can become less effective in encouraging retailers to participate in the market when demand uncertainty is considered.

Contrasting the findings from these two major model extensions reveals the following insight: Quantity competition exhibits a similar impact on the effectiveness of collaborative replenishment strategies as demand uncertainty does in a regulated market. However, they exert the opposite impact in an unregulated market. This further characterizes the critical role of the market environment (regulated vs. unregulated, form of competition, demand uncertainty) in determining the effectiveness of collaborative replenishment strategies.

9.5 Conclusions

Without efficient infrastructure and distribution channels in rural areas of many developing countries, poor retailers incur high replenishment costs. As such, either poor micro-entrepreneurs not yet in the micro-retailing market cannot afford to participate or existing micro-retailers' on-going operations generate meager income. Consequently, consumers' product accessibility is hampered, micro-retailers' earning potential is stagnant, and manufacturers' growth in emerging markets is limited. Motivated by different collaborative replenishment programs developed by social

Table 9.1 Effect of cooperative replenishment strategies

	Regulated market	Unregulated market
Cooperative vs. independent	Cooperative strategy benefits retailers but does not change consumer welfare.	Cooperative induces more retailers in the market and also benefits consumers.
Wholesaler vs. independent	Wholesaler strategy benefits retailers but undermines consumer welfare.	Wholesaler induces more retailers in the market and can benefit consumers under general conditions.
Cooperative vs. wholesaler	Wholesaler is typically better for retailers but cooperative is better for consumers.	Wholesaler induces more retailers but cooperative is better for consumers under general conditions.

enterprises and not-for-profit companies, we develop economic models to examine the implications of two main types of strategies: (1) forming an open purchasing cooperative among all participating retailers that coordinate their replenishment, pooling their inventories and share cost; and (2) introducing a non-profit wholesaler that procures on the retailers' behalf under a low service charge. We analyze these collaborative strategies against a benchmark case (called the independent strategy) where individual retailers replenish on their own and no collaboration/coordination exists. By examining the equilibrium outcomes associated with each of these inventory replenishment strategies in both regulated and unregulated market settings, we have obtained a set of insights that are summarized in Table 9.1.

The findings from our analysis suggest that, in determining and implementing a suitable strategy to coordinate individual retailers' inventory replenishment, the policy maker needs to be mindful about various business conditions. For example, is the market entry strictly controlled by government (such as in the case of drugs or medical items) or is free market entry feasible where retailers can enter the market willingly as long as their profit is no less than a pre-specified reservation value? It is also important for the policy maker to keep in mind the priorities in adopting a coordinating strategy. For instance, in the regulated market setting, is it more important to consider the retailers' benefits and survival conditions, or is the consumer welfare a more serious issue to be taken care of? In the unregulated case, is the size of the retailer network (defined by the number of participating retailers) a critical consideration to maintain the sustainability of the market? In addition, other factors such as the nature of competition (i.e., whether retailers engage in price or quantity competition) and the demand characteristics (i.e., whether it is deterministic for, e.g., commodity products, or uncertain for, e.g., medical or fashion products) may also influence the effectiveness of the collaborative strategies. Contingent on these factors and the priorities in policy-making, the policy maker may choose to encourage an open cooperative formed by the participating members or to introduce an intermediary non-profit wholesaler to the system.

We close this chapter by discussing a few research avenues for collaborative micro-retailing. First, it is interesting to observe that purchasing cooperatives and non-profit wholesaling can be operationalized in different ways in practice. For example, under the Hapinoy program for local sari-sari stores in the Philippines (Hapinoy 2018), the wholesaling services are undertaken by one of the micro-retailers which is called the community store. Such dual role of the retailer creates additional complexity to the competition dynamics in the market and affects the retailer profit and consumer welfare. Studying the impact of such operational factors can generate important insights as to effective implementation of these collaborative strategies (see Zhang et al. 2017, for a relevant study). Second, as financial and technology support to micro-retailers increases and the operational constraints of these retailers loosen up (as it is indeed the case in many rural areas), more strategic decision-making tools are required. Examples include collaborative information and demand forecast sharing mechanisms (e.g., Chen and Tang 2015) and complex inventory planning strategies to mitigate frequent stockouts.

References

Anderson SP, De Palma A, Thisse JF (1992) Discrete choice theory of product differentiation. MIT Press, Boston

Bernstein F, Federgruen A (2005) Decentralized supply chains with competing retailers under demand uncertainty. Manag Sci 51(1):18–29

Cattani KD, Gilland WG, Swaminathan JM (2004) Coordinating traditional and internet supply chains. In: Simchi-Levi D, Wu SD, Shen ZM (eds) Handbook of quantitative supply chain analysis. Springer, Berlin, pp 643–677

Çetinkaya S, Lee CY (2000) Stock replenishment and shipment scheduling for vendor-managed inventory systems. Manag Sci 46(2):217–232

Chen RR, Roma P (2011) Group buying of competing retailers. Prod Oper Manag 20(2):181–197

Chen YJ, Tang CS (2015) The economic value of market information for farmers in developing economies. Prod Oper Manag 24(9):1441–1452

Chen L, Kim SH, Lee HL (2017) Enabling healthcare delivery through vehicle maintenance. Working paper, Cornell University

Dixit A (1979) A model of duopoly suggesting a theory of entry barriers. Bell J Econ 10(1):20–32

Fransoo JC, Blanco EE, Argueta CM (2017) Reaching 50 million nanostores: retail distribution in emerging megacities. CreateSpace Independent Publishing Platform, Scotts Valley

Goodman C, Kachur SP, Abdulla S, Mwageni E, Nyoni J, Schellenberg JA, Mills A, Bloland P (2004) Retail supply of malaria-related drugs in rural Tanzania: risks and opportunities. Trop Med Int Health 9(6):655–663

Govt of NCT of Delhi (2017) Delhi consumer's co-operative wholesale store Ltd. delhi.gov.in/wps/wcm/connect/doit_dccws/DOIT_DCCWS/Home. Accessed Jan 2019

Gui L, Tang C, Yin S (2019) Improving micro-retailer and consumer welfare in developing economies: replenishment strategies and market entries. Manuf Serv Oper Manag 21(1):231–250

Hapinoy (2018) We create opportunities for sari-sari stores to grow. https://www.hapinoy.com/. Accessed 17 Jan 2019

International Fund for Agricultural Development (2001) Markets for the rural poor. Rural poverty report 2001: the challenge of ending rural poverty, chap 5. Oxford University Press, Oxford, pp 161–187

Kumar B, Gogoi M (2017) Fast moving consumer goods industry in rural market of India: a case of mutual reinvigoration. USHUS J Bus Manag 12(4):51–65

Lee HL, So KC, Tang CS (2000) The value of information sharing in a two-level supply chain. Manag Sci 46(5):626–643

McCoy JH, Lee HL (2014) Using fairness models to improve equity in health delivery fleet management. Prod Oper Manag 23(6):965–977

MEDS (2017) Supply chain. Accessed Jan 2019. http://www.meds.or.ke/index.php/our-services/supply-chain

Mishra BK, Raghunathan S (2004) Retailer-vs. vendor-managed inventory and brand competition. Manag Sci 50(4):445–457

Neuwirth B (2012) Marketing channel strategies in rural emerging markets. Kellogg School of Management. http://www.kellogg.northwestern.edu/~/media/files/research/crti/marketing%20channel%20strategy%20in%20rural%20emerging%20markets%20ben%20neuwirth.ashx. Accessed Jan 2019

Nieuwoudt T (2015) Challenges of micro retailers in frontier markets. The Supply Chain Lab. https://thesupplychainlab.blog/2015/12/03/micro-retail-challenges-in-emerging-markets/. Accessed Jan 2019

OECD (2013) Global forum on competition: competition and poverty reduction. https://www.oecd.org/daf/competition/competition-and-poverty-reduction2013.pdf. Accessed Jan 2019

Prahalad CK (2006) The fortune at the bottom of the pyramid. Pearson Education. Prentice Hall, Upper Saddle River

Rangan VK, Rajan R (2007) Unilever in India: Hindustan Lever's project Shakti – Marketing FMCG to the rural consumer. Harvard Business School Case #9-505-056

Shanxi Province Administration for Industry and Commerce (2007) Nong cun xiao mai bu de jian guan [Chinese] Monitoring of rural micro-retailers. http://www.jcaic.gov.cn/news_read.asp?id=1072

ShopRite (2017) ShopRite today – Families serving families. http://www.shoprite.com/about-us/

Singh N, Vives X (1984) Price and quantity competition in a differentiated duopoly. RAND J Econ **15**:546–554

Tata Trust (2016) Trading in tea. http://tatatrusts.org/article/inside/himmotthan-uttarakhand-tata-himalaya

Tsay AA, Agrawal N (2004) Modeling conflict and coordination in multi-channel distribution systems: a review. In: Simchi-Levi D, Wu SD, Shen Z-J (eds) Handbook of quantitative supply chain analysis. Springer, Berlin, pp 557–606

Uppari BS, Popescu I, Netessine S (2017) Selling off-grid light to liquidity constrained consumers. Manuf. Serv Oper Manag Published online 27 Mar 2018. https://doi.org/10.1287/msom.2017.0673

von Ravensburg NG (2011) Economic and other benefits of the entrepreneurs cooperative as a specific form of enterprise cluster. International Labour Organization

Yadav P (2013) Value chain innovations or developing economies. Presentation, Supply Chain Thought Leaders Roundtable, UCLA

Zhang K, Tang CS, Zhou S (2017) Replenishment strategies for micro-retailers in developing countries. Prod Oper Manag 26(12):2207–2225

Chapter 10
The History and Progression of Sustainability Programs in the Retail Industry

Tiffin Shewmake, Adam Siegel, and Erin Hiatt

Abstract The retail industry's economic and environmental impact encompasses global supply chains and many thousands of facilities; it therefore has a significant impact on human health and the environment, and a responsibility for reducing this impact. This chapter begins with an overview of retail's most common and significant environmental impacts and examines the origin of sustainable planning and operations, the business case for sustainability programs, and the maturation of retail sustainability programs.

Companies start sustainability programs for different reasons—some internal to the company and some as a result of external pressures. These pressures continue to drive further action, advancing and maturing sustainability programs across the industry. Those programs tend to grow and mature around common dimensions, beginning with a focus on the basics, primarily complying with regulatory requirements, then growing to cover operational efficiencies that save costs, reducing reputational risks, and eventually innovating on the very core of their businesses. The chapter includes business-actionable steps for retail sustainability practitioners and ends by describing the critical programmatic components for a strong retail sustainability program.

10.1 Introduction

The retail industry's economic and environmental impact encompasses global supply chains and many thousands of facilities—material extraction to manufacture to consumer disposal of products after their purchase. As a result, retailing has a significant impact on human health and the environment. Retail is not unique in

T. Shewmake · E. Hiatt
Retail Industry Leaders Associations, Arlington, VA, USA

A. Siegel (✉)
Sustainability Entrepreneur and Consultant, Arlington, VA, USA

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1_10

this aspect; all organizations, in all economic sectors, contribute to environmental degradation. While some industries, such as mining and manufacturing, are more directly associated with environmental issues, all sectors including real estate, banking, telecommunications, medicine, entertainment, and more, use energy for their operations, require material inputs, produce waste, develop land, and generate greenhouse gases (GHG). The environmental consequences of these activities span geographies from local, such as the pollution of a pond or stream, to global, as in the case of climate change.

Just as every organization in this chain causes environmental damage, each is also responsible for reducing this damage. This is true not just for primary sectors such as manufacturing, but also governments who are responsible for setting and enforcing protective regulations, and consumers, whose purchase, use and disposal habits, all affect the potential environmental harm from a product. The retail industry is in a unique position to reduce impacts because of its influence on its supply chains, including product design, manufacturing, and logistics, as well as control of its more direct operations. However, retail's greatest influence potential comes from the unique relationship between retailers and their shoppers. Retail has the ability to educate consumers about the impact of their buying decisions and, in turn, can translate consumer's desire for "greener" products back through the supply chain.

10.1.1 What We Will Cover

This chapter covers the pursuit of sustainability by retail companies, with a focus on environmental sustainability. It begins with an overview of retail's most common and significant types of environmental impact, and examines the origin of sustainable planning and operations, the business case for sustainability programs, and the maturation of retail sustainability programs. Each section includes business-actionable insights and steps that retail sustainability practitioners can use to guide their strategic planning and operational activities.

The chapter also provides an overview of responsible sourcing, which is a primary element of a retailer's holistic corporate social responsibility program (CSR). Responsible sourcing has environmental elements such as pollution and deforestation as well as social areas such as forced labor. Some of the environmental aspects of responsible sourcing are covered here, as well as a history of related social aspects that are integral in the development of modern CSR programs.

10.1.2 History

There is a long history of concern over the environment, particularly following the industrial revolution of the late eighteenth and early nineteenth centuries. During this time, industrial advancements were essentially inseparable from environmental degradation, as the full consequences of environmentally intensive materials extraction techniques and carbon-based fuel sources were not yet known. In 1798, Thomas Malthus wrote his famous essay predicting dire consequences from a growing population and the stresses this would place on the earth's resources. In the USA, it would take nearly 200 years before concern over growing and visible pollution lead to major federal environmental legislation in the seventies such as the Clean Air Act and the Clean Water Act. At that time, companies focused their environmental efforts primarily on complying with the new laws and later with managing direct environmental damage such as pollution prevention initiatives.

There was no widespread societal expectation that businesses were responsible for anything more than complying with the law when operating in a capitalistic economy. In the late eighties, the concept of sustainability, that is to balance development and environmental degradation so that the latter does not inevitably lead to the prevention of the former, was brought to international recognition and articulated by the 1987 Brundtland Report (United Nations 1987). Although companies were still focused mostly on direct environmental issues, there was a shift to a new and broader concept, that of CSR.

CSR encompasses a range of areas in addition to the environment and is perhaps best characterized by the term "Triple Bottom Line," first coined by John Elkington in 1994, which refers to social, environmental, and financial aspects, or in short: people, planet, and profits (Elkington 1994). Today, corporate CSR programs cover issues ranging from sustainable operations to philanthropy, labor practices, supply chain, and volunteering. Because CSR is so broad-reaching, implementation of different aspects of a CSR program may be divided up among different parts of a company. For example, a CSR goal may be to reduce greenhouse gas emissions but it is the facility operations staff who implement the practices needed to reach the goal. CSR programs frequently have a focus on transparency, with companies committing to annual reporting of their efforts. In response, several organizations have developed sustainability and CSR reporting frameworks such as the Global Reporting Initiative's (GRI) Sustainability Reporting Standards.

Although a few retailers were early adopters of these new values, such as the 1987 launch of the Body Shop, CSR as business discipline is relatively new. The earliest programs in retail companies began in the mid-1990s, primarily as a result of media and activist attention on consumer products, focusing on human rights abuses in textile and footwear factories. The drive for low-cost manufacturing had created a global production economy with few regulations or oversight, and little transparency for consumers. Child labor and sweatshop-like conditions in factories like the ones producing the celebrity Kathie Lee Gifford's clothing line made the front-page news in major publications, catching the attention of consumers worldwide.

Initially, the pressure to change focused on consumer brands, the producers of retail products, rather than on the retailers themselves. Brands tended to take one of two divergent paths in response to the pressure. Some brands pushed back, claiming that the factories were not under their control, and that they had little influence in correcting these human rights challenges. These brands also characterized the problems as a natural consequence of globalized economic markets that push for the lowest possible manufacturing costs. On the other hand, some brands accepted the challenge, recognizing that their reputation was at stake. These companies stepped up by partnering with local experts, non-governmental organizations (NGO), and government agencies to test and scale solutions.

Nike is perhaps the most famous example, having been the target of numerous protests after Harpers published a 1992 exposé on working conditions in Indonesian factories (Ballinger 1992). While Nike initially attempted to dampen criticism, public outcry stretched from protests at the 1992 Barcelona Olympics to backlash following an insufficiently critical report from the diplomat and activist Andrew Young, peaking in 1997 when college students nationwide protested university contracts with the company. In 1998, Nike made a hard pivot, beginning with a speech by then-CEO Phil Knight, acknowledging and taking responsibility for the exposed practices. This was quickly followed by the creation of the Fair Labor Association, a rapid escalation of factory audits, factory contract transparency, and the start of regular, detailed reports on commitments, progress, and ongoing challenges. This was the strongest indicator the industry had seen to this point of the inevitable rapid push for transparency and accountability, with consumers demanding that companies face issues head on rather than shield the public from global capitalism's darker potential.

Over two decades later, the stigma from the initial headlines continues to have staying power, with many consumers still associating accused brands with child labor and sweatshop-like conditions. As a result, consumer brands, whose reputation accounts for a large portion of their market value, now see the need to address this threat and recognize the importance of a "social license to operate," whether or not they directly control their manufacturing. For example, after the 2013 Rana Plaza factory collapse in Bangladesh, where over 1000 garment factory workers died, many retailers became involved in initiatives designed to prevent similar tragedies. Some companies, including Gap Inc., J. C. Penney, Kohl's, Target, VF Corporation, and Walmart developed a new organization, the Alliance for Bangladesh Worker Safety, which is a "legally binding, 5-year commitment to improve safety in Bangladeshi ready-made garment (RMG) factories" (Alliance for Bangladesh Worker Safety 2018). The organization was unprecedented for its level of collaboration and accountability, bringing the US and Bangladeshi governments together with consumer brands and retailers, NGOs and labor groups. This modern example illustrates how major retailers moved from a head in the sand position relative to their supply chain to one of positive public engagement (Fig. 10.1).

In the early and mid-2000s, off the heels of fresh wins and a growing consumer consciousness of factory-related issues, activists, socially responsible investors, and the media expanded their coverage into other industries and beyond human rights issues. It was perhaps the exposure to shocking human rights issues that led consumers and investors to demand greater transparency in retail supply chains, expanding the scope of issues covered to include environmental problems. Retailers, whose product supply chains now stretched deep into the developing world and had influence—whether strong or not—on production facilities, had many layers of social and environmental impacts for which they could now be held

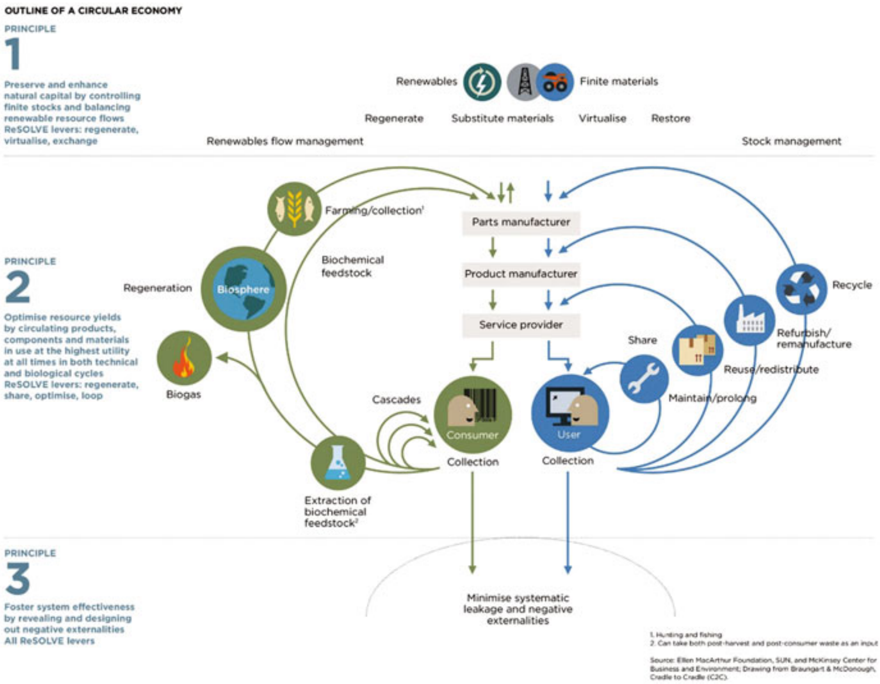

Fig. 10.1 Outline of a circular economy (graphic from The Ellen MacArthur Foundation https://www.ellenmacarthurfoundation.org/circular-economy/infographic)

accountable. The public and NGOs brought new focus to environmental issues such as deforestation from palm oil plantations, water pollution from textile manufacturing, and GHG emissions. As more consumer brands developed sustainability programs, often taking a proactive approach in improving their brands, retailers followed suit, building on their lessons learned. The main areas of focus for retail environmental sustainability programs align directly with their areas of influence, including facility operations, product lifecycles, and marketing to consumers—all subsequently detailed in this chapter.

Retail sustainability programs have matured since the earliest programs. The leading programs, like those of Apple, IKEA, Nike, and VF Corporation, take a holistic and systemic approach to addressing the environmental impacts of their businesses. The term "circular economy" has recently emerged as a vision for an alternative consumption model with significant implications for the role of retailers. The Ellen McArthur Foundation, a leading organization in the field, describes the circular economy as:

> Looking beyond the current "take, make, and dispose" extractive industrial model, the circular economy is restorative and regenerative by design. Relying on system-wide innovation, it aims to redefine products and services to design waste out, while minimizing negative impacts. Underpinned by a transition to renewable energy sources, the circular model builds economic, natural, and social capital (Ellen MacArthur Foundation 2017).

To conform to a circular economy, brands and retailers need to design and source products with recycled and recyclable materials, produce and distribute products with renewable energy, and provide opportunities for consumers to share, rent, take-back, reduce, reuse, or recycle products when they are finished using them.

The pursuit of a circular economy is aligned with the UN Sustainable Development Goal (SDG) 12, Responsible Consumption and Production. This goal, summarized most simply as "doing more and better with less," requires a systems-thinking approach with implications stretching from producer to consumer (United Nations 2015). It is becoming more common for companies, including retailers, to align their sustainability activities and the UN goals when setting their corporate objectives and in public reporting. Dave Lewis, CEO of United Kingdom grocer and merchandise behemoth, Tesco, is the Chairman of Champions 12.3, a group of executives from business, government, and NGOs dedicated to achieving SDG goal 12.3 to reduce food waste. According to Lewis, measuring waste is a first step: "We believe that what gets measured gets managed. Ultimately, the only way to tackle food waste is to understand the challenge—to know where in the supply chain food is wasted." In 2013, Tesco became the first UK retailer to report on its food waste and is over 70% of the way towards its goal that no food, safe for human consumption, goes to waste. Other retailers have also set goals aligned with SDG 12, such as Starbucks's vision for more sustainable coffee farms, which includes programs such as Coffee and Farmer Equity (CAFE) and the Sustainable Coffee Challenge.

Currently, retail operations, product lifecycles, and marketing are generally not aligned with the principles of a circular economy. Innovations in business models, improved product and process design, and reduced compliance obligations can all result from a circular approach to sustainability management. Coca-Cola, for example, set an ambitious goal in early 2018 to recycle the equivalent of 100% of the packaging they sell by 2030, ultimately aiming for a complete circular economy for their packaging. To achieve this, they will need to design products made from recyclable and recycled materials, and with materials that can be recycled in most locales, as well as invest in recycling infrastructure to improve recycling rates (Karas 2018). Other innovative companies are developing systems to take-back and reuse the components in products they sell. Apple is one example, creating a robot called Liam to disassemble 1.2 million iPhone 6 units per year. Liam will be able to retain a greater quality and value of the components relative to the current e-waste management process of shredding old electronics (Rujanavech et al. 2016).

10.2 Impact Areas: Facility Operations, Product Lifecycles, and Marketing

Retail operations impact the environment through three main mechanisms: *facility operations*, *product lifecycles*, and *marketing*. These operations and the most significant effects on the environment are detailed below.

10.2.1 Facility Operations

This section covers facility operations in the most common types of retail facilities, which are stores and distribution centers (DC). A minority of retailers also own and operate manufacturing facilities for their private-brand products, which are not covered here. The environmental areas associated with facility operations include air emissions, water and wastewater, and waste generation and recycling. The retail sector has several additional challenges in facility operations. One issue is the sheer number of facilities; a large retailer can have hundreds to even thousands of stores. While the number alone makes operations more difficult, this is compounded by the variations in different regions. For example, a retailer operating in the Edwards Aquifer area in Texas has to comply with significantly greater stormwater requirements to protect the groundwater than in other areas of the country, and retailers in flood prone areas must be prepared to protect their facilities and look for ways to reduce the cost, waste, and damage from flooding. In addition, store types vary widely from the large footprint of a standalone home improvement retailer with outside storage to a small store operating in a mall. The differences mean that the stores will have very different impact on the environment as well as different responsibilities and opportunities.

Retailers have the most direct influence for reducing environmental impact with their *facility operations*. Managing, for example, energy efficiency projects in stores is more direct, than managing the supply chain for a product that contains components from multiple companies in multiple countries across a wide geographic distance. Facility operations are typically carried out or managed by employees of the retail company and often involve facilities owned or managed directly by the company, giving the retailer a high degree of control. In facility operations that retail employees do not directly manage, for example, waste hauling, but where the retailer is the primary client, they can include contract language and provide oversight to help achieve their sustainability goals.

While retailers may build and own their facilities, typically using a contractor to construct the site to their specifications, most lease facilities from shopping center developers. For leased properties, retailers typically have control over the interior build-out, but depending on the size of the space and length of the lease, their influence on building operations may be limited. The interior build-out of leased space typically refers to the entire design and layout, including procurement for walls, floors, ceilings, lighting fixtures, signage, and technology systems.

Decisions about site development and store build-outs have significant implications on the immediate and long-term environmental footprint of the facility. For example, a store developed on a wetland site can have a significant long-term negative impact, while a store placed near a city center can have reduced footprint both in terms of the structure and how consumers travel to the store. A build-out using more energy efficient technology can reduce energy use and save money over the store's life. Some retailers have policies to implement green building standards for buildings and interiors. The best-known standard in the USA is the Leadership in Energy and Environment (LEED), which has two rating systems specifically for retail, one for new construction and one for interiors. In 2008, Kohl's committed to achieving LEED certification for all new stores and corporate facilities. Kohl's also operates stores using the LEED for Existing Buildings: Operations & Maintenance guidelines (USGBC 2017). According to Starbucks in 2014, "Each day in 18 countries around the world, an estimated 229,000 people visit a LEED-certified Starbucks location.... In fact, Starbucks LEED-certified stores actually use 30% less energy and 60% less water than non-certified locations" (USGBC 2014).

The rise in green building design and construction in retail is illustrated by the growth in retailers' LEED certifications and commitments. In 2001 there was only one LEED-certified retail project; by 2013 there were 905. Companies like Target, Starbucks, Kohl's, Yum! Brands, and others led the way with their commitments to green building design and the LEED program. Many retail companies are benefiting from the energy- and cost-savings potential of greener buildings by updating their standard store designs to LEED standards but without a formal certification (USGBC 2014).

The demand for DCs is growing to support e-commerce and customer expectations for fast shipping. DCs, which can be up to one million square feet, are not passive storage warehouses, but rather high-tech spaces that use automation, computers, and technology to optimize cost and speed. DCs have a significant impact on the environment, including impaired water quality from stormwater runoff, air pollution from trucks, GHG emissions from energy use, and waste. Many retailers are working to make DCs more sustainable by focusing on areas such as energy efficiency, renewable energy, and waste reduction and recycling. Leaders go further, for example, REI's Arizona DC is LEED platinum certified with net zero energy and zero waste. However, the reality is that the majority of DCs have a long way to go to match this performance.

Air emissions from store operations can be from direct sources such as vehicles, boilers (for heating larger stores and distribution centers), and emergency generators, or indirect from the use of electricity. Air pollution has significant environmental and health impacts including contributing to major health effects such as strokes, heart disease, and asthma, as well as to climate change. While direct sources may seem small individually, in urban areas the combined emissions can be significant. The extent of air pollution from electricity depends on the fuel source. Coal-fired power plants have the most health and environmental negatives from the coal mining to air pollution to coal ash waste. Air emissions from coal-fired power plants also contribute to water pollution and are a leading cause of mercury in water. Although considered cleaner, unburned natural gas is a potent GHG and fracking can cause there are significant environmental damage.

Energy efficiency is a cost-effective way to reduce the negative impact of energy use and GHG emissions. In retail, one of the challenges to implementing energy efficiency is that many of the changes, for example, lighting or refrigerated case design, affect customers. Retailers are naturally hesitant to introduce any change that may be perceived negatively by customers and therefore prefer to leave conditions as-is, rather than change existing systems. However, changes can also yield positive benefits. One study found that the use of LED lights in displays could reduce energy use by 50%, without sacrificing visual appeal or the ability to attract the attention of shoppers (Freyssinier 2006).

Another challenge is rolling out new technologies over hundreds or even thousands of facilities. The logistical and operational considerations in such large-scale implementations make even seemingly minor changes cost-prohibitive. Therefore, it is customary to introduce new, energy efficient technologies during routine maintenance and store systems upgrades. Performing rollouts in this way allows retailers to save on the cost of “truck rolls”—sending maintenance vehicles and building engineers to stores.

Leading retailers have set public targets for greenhouse gas reduction. Retailers with public science-based targets, i.e., targets based on reductions, need to keep warming below 2°C, include Best Buy, Gap, Target, CVS, and Walmart. As an example, CVS’s goal is “to reduce absolute scope 1 and 2 GHG emissions 36% by 2030 from a 2010 base-year. CVS Health also commits that 70% of its suppliers by emissions will set science-based emissions reduction targets on their scope 1 and 2 emissions by 2023” (Science Based Targets 2018).

Many retailers are also turning to renewable energy to reduce pollution and meet GHG emissions reduction goals. A number of major retailers participate in EPA’s Green Power Partnership, making commitments to use green power (solar, wind, or biogas) for some or all their electricity. Some retailers including IKEA, Estee Lauder Companies, H&M, Starbucks, Tesco, T-Mobile US, VF Corporation, and Walmart have made a commitment to 100% renewable energy through the RE100 global initiative. In 2016, Target, with 147.5 megawatts of solar capacity, passed Walmart as the US company with the greatest solar power capacity. Other retailers with large solar arrays are Costco, Kohl’s, and IKEA (Solar Energy Industries Association 2017).

Another air pollution source in retail are the common refrigerants such as chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs). These chemicals, used in freezers, air conditioning units, coolers, and vending machines, contribute to the destruction of atmospheric ozone and the resulting ozone hole, and can be very potent greenhouse gases. As a result, some types of refrigerants are banned and there are regulations to encourage the use of less harmful refrigerants and reduce leaks from equipment. Most of the regulations are at the federal level with the exception of California, which has state regulations designed to control refrigerant emissions and reduce the use of HFCs (California Air Resources Board 2018). Grocery retailers such as Albertsons, Aldi, BJs Wholesale Club, Food Lion, Giant Eagle, Meijer, Publix, and more participate in EPA’s GreenChill program. These companies make a “commitment to reduce their corporate refrigerant emissions by annually setting reduction goals, measuring corporate stocks and emissions, and reporting their data to EPA” (EPA 2018). However, ultimately to protect the ozone layer and meet climate change targets, companies will need to switch to “natural” refrigerants including hydrocarbons such as propane and iso-butane, carbon dioxide, ammonia, water, or air.

Another area, not usually associated with air pollution, are emissions from products that contain volatile organic compounds (VOC) such as adhesives, air fresheners, and solvents. VOCs can be unhealthy to breath and contribute to the formation of smog. As a result, there are federal and state regulations that limit the amount of VOCs allowed in products. Some retailers go beyond the regulatory requirements and implement internal policies to carry products with low or no VOC emissions. For example, in 2012 Home Depot transitioned to low-VOC paint tinting colorants, and currently all of their latex wall paints are zero or low VOC formulations. In addition, they offer other products such as flooring and carpet pads that are certified for very low chemical and VOC emissions (The Home Depot 2017).

Water pollution is another environmental concern with retail operations. Stormwater in developed areas flows faster and with greater force over impervious surfaces such as roofs, parking lots, and roads, than in natural areas. The force of the flow causes increased erosion and sediment, and the stormwater also picks up contaminants such as oil from parking lots, and fertilizers and pesticides. Brick and mortar stores have a lot of impervious surfaces from the roofs and pavement. While some retailers are implementing greener designs in new construction to reduce or even eliminate runoff, this is not an area where the retail industry is a leader, especially in regards to reducing impacts from existing structures.

Another environmental impact is the waste that retail operations generate from packaging and shipping material, food preparation, services such automotive care or photo printing, as well as unsalable or unusable products. The decomposition of waste in landfills releases methane, a potent GHG, and air emissions result from waste incineration. The transportation of waste and the lifecycle of disposed material is also a significant impact. Many retailers have extensive recycling programs and some have made commitments for zero waste facilities. Major retailers Kroger and Walmart both have zero waste aspirations, with Walmart's extending to their entire supply chain. There are challenges, however, as recycling infrastructure is limited in many communities and for some materials. Store design can also be a factor since recycling programs require space to store material prior to shipment, which is limited in many stores.

10.2.2 Product Lifecycle

While the facility operations described in the previous section have a significant environmental impact, in retail, it is really the products that have the largest and most far-reaching environmental impact. This includes the entire *product lifecycle:* design, natural resource extraction including agriculture, manufacture, transportation, use, and ultimately disposal. The role of retail is to sell goods, which encompasses a huge array of items including food, clothing, furniture, home improvement tools, school and office supplies, health and beauty products, pharmaceuticals, pet products, electronics, vehicles, and many, many more types of products. A vast global manufacturing and shipping infrastructure, which fuels the global economy and employs millions, is required to make and distribute these goods. According to a November 2016 McKinsey article, over 90% of the consumer sector's natural capital impact (e.g., impacting air, soil, land) and over 80% of consumer-goods' GHG emissions lie in the supply chain (Bove and Swartz 2016). The global production of paper and cardboard alone, for example, was approximately 407 million metric tons in 2014.

Products begin with design. Product designers determine the products' form and function, shapes, colors, materials and ingredients, packaging, and more based on market needs; i.e., what consumers want, creative elements, manufacturing and transportation considerations, cost, and other criteria. The decisions made at this stage influence all other stages of product development, including the product's environmental footprint. For example, the materials selected for a product's design will define the raw materials used as input into the product's manufacture. The size of a product and packaging determines the amount of natural resources used, the energy used in transportation, and the amount of waste generated. The design also determines the energy or water efficiency of a product, how long it will last, if it contains toxic or hazardous constituents, and if it can be reused or easily recycled. Therefore, it is critical to embed sustainability principles at the very beginning in the design of new products and reformulation of existing products. Gap Inc., parent to brands Banana Republic, Old Navy, Gap, Athleta, Intermix, and Hill City, describes their sustainable design process: "We educate our brand teams about how to design using more sustainable fibers, fabrics and manufacturing techniques that save water—and how to procure more responsible materials." In addition to training product design teams on making more informed design choices, the company addresses factors like the type and amount of raw materials used to reduce water consumption in all aspects of their product lifecycle (Gap Inc 2018a).

Retailers are increasingly using sustainability product guidelines and certifications for the products they sell to reduce the impact and to help customers select greener products. Staples offers a range of products with environmental certifications, including Forest Stewardship Council (FSC) paper products, EPA Safer Choice and Green Seal™ cleaners, Fair Trade and USDA Organic coffees, Level® furniture, and ENERGY STAR and EPEAT electronics. Even more, since 2012, Staples has partnered with the Rochester Institute of Technology on the Staples Sustainable Innovation Laboratory to find greener alternatives (Staples, Inc 2018).

Transportation is a significant component of a product's impact. Retailers can use a variety of strategies to reduce that impact. For example, companies can deploy the strategies promoted by the US EPA's voluntary SmartWay program to improve freight transportation efficiency, which promotes technologies like aerodynamic devices for trailers, low rolling resistance tire technologies, and idle reduction technologies. In 2018, SmartWay Excellence Awardees included Lowe's Companies, Inc., Meijer Inc., and Nordstrom, Inc. while previous retail awardees also have included Gap, Inc., Kohl's Department Stores, and The Home Depot

EPA 2018. This award "honors top shipping (retailers and manufacturers) and logistics company partners for superior environmental performance and additional actions to reduce freight emissions through effective collaboration, operational practices, a robust system for validating and reporting their SmartWay data, and communications and public outreach."

Eventually, products become waste. While some consumer products are recycled, the recycling infrastructure and culture in the USA are weak and most used products are sent to the landfill or incinerated. For example, less than 10% of plastic waste is recycled and it takes hundreds to thousands of years to decompose in a landfill, if it ever actually decomposes. Recycling products not only takes it out of the waste stream, but also reduces the need for virgin natural resources to make new products. Also, the manufacturing of recycled content products generally has a smaller environmental footprint. Some retailers have take-back programs for specific types of waste, such as Best Buy's e-waste program and the used plastic bags collected by grocery stores. However, the real, long-term solution to waste is a shift in consumer preferences so that consumers seek products made with recycled content, products that can be recycled, and view single-use or disposable products less positively. IKEA educates consumers to live healthier and more sustainably, including the energy and cost-savings benefits of using LED light bulbs, high efficiency refrigerators, induction cooktops, and insulating rugs; reducing tap water consumption and food waste; creating simple in-home recycling solutions; and more. The company also encourages consumers to use bicycles instead of cars for some trips, eat vegetarian, and grow produce locally (IKEA 2018).

Lifecycle Assessment (LCA) is used to measure all of a product's environmental impacts. LCAs look at the full lifecycle of a product, from production through use, and finally to disposal. The International Organization for Standardization (ISO) 14040:2006 lays out a framework for conducting LCAs (International Organization for Standardization 2006). LCAs are also used to compare the performance between different products. For example, one company compared wine packaging LCAs that included data from raw material extraction through to disposal, and found that cardboard packages have a significantly better environmental profile than other types of wine containers (Tetrapak 2018). Similar techniques are also used for the "social" lifecycle impact of products. The value of these tools is only as good as the underlying data and, while there is a growing body of research, more work is needed to assess product lifecycle impacts at scale.

10.2.3 Marketing

Surprisingly marketing has a significant influence on sustainability. Marketing can create demand for more sustainable products and awareness of important issues. For example, in 2014, Target launched a collection called "Made to Matter—Handpicked by Target" that featured more sustainable products across several categories including baby, beauty, and grocery. These products were marketed to consumers through a variety of channels and even given premium store real estate on aisle endcaps. Other retailers, such as The Body Shop, Patagonia, and Whole Foods, market sustainability to differentiate themselves and to educate customers. In some cases, customer concerns, often amplified by activist groups, convince brands and retailers to offer and then promote more sustainable products. For example, at the urging of animal welfare groups, most major grocery stores in the USA have made commitments to sell only cage-free eggs and sustainably sourced seafood, a commitment that they can then promote to their customers.

While companies with a core sustainability mission have been successful in translating their sustainability message to growth, like the sales bump from Patagonia's 2011 "don't buy this jacket" campaign, other retailers find it harder to use sustainability to drive sales (MacKinnon 2015). For example, one research article found that consumers' existing perceptions of retailers influenced their perception of the authenticity of sustainability efforts such that "More tangible aspects such as perceived customer service performance, create a halo effect that influences the authenticity of a sustainability program. If the goal of a sustainability effort is to salvage an ailing public image that is suffering from perception of bad service quality, then a sustainability program is not likely to provide a positive return on that goal" (Brockhaus et al. 2017). The message here may be that consumers trust brands that they believe in, and that companies without this sustainability core must work harder to develop trust and be more careful about the appearance of greenwashing.

Some retailers build sustainability into their brand and communicate this to their customers, who increasingly prefer to shop with more sustainable companies. Outdoor retailer REI and grocer Whole Foods Market provide product labels that give consumers information about the environmental impact of their products. For example, Whole Foods uses an Eco-Scale for cleaning products with orange, yellow, or green ratings based on criteria such as transparency, chemicals, and animal testing. For these retailers, their sustainability mission is aligned with their consumers.

In other cases, retailers not typically associated with sustainability are taking significant steps towards sustainability but the focus of their communications are different. In 2005, Walmart's then CEO H. Lee Scott, announced a far-reaching sustainability strategy for Walmart including the goal to sell more sustainable products. Unlike high end retailers REI and Whole Foods, Walmart, as a retailer dedicated to low prices, cannot pass additional costs along to consumers. As a result, Walmart's approach focused on moving all suppliers towards better environmental and social performance rather than depending on consumers to choose more sustainable products (Spicer and Hyatt 2017). Other companies direct some of their sustainability communications towards the environmental community. In 2017, Target set public goals around chemicals in products and committed to reporting results. Their reporting is not intended primarily for customers, but instead is for the environmental community through their CSR Report (Target Corporation 2017).

10.3 The Initiation of Retail Sustainability Program

Companies start sustainability programs for different reasons—some internal to the company and some as a result of external pressures. These pressures continue to drive further action, advancing and maturing sustainability programs across the industry. This section describes the most common pressures that result in the initiation of sustainability programs in retail companies.

10.3.1 Born with It: It is Who We Are

Some companies, albeit a relatively small segment of the retail industry, were founded on sustainable principles. This is often the result of an eccentric or dynamic founder, who has strong social values and an intrinsic interest in creating change in the consumption industry. Companies like Patagonia, Whole Foods, and Seventh Generation grew from their very start based on the sustainability motivations of their founders. More recently, retailers like TreeHouse, Warby Parker, and LUSH have been founded on sustainability principles to meet the growing demand for companies with positive social values.

10.3.2 Senior Management

Senior executives initiate sustainability programs for several reasons. Some because of corporate motivations like enhancing their company's reputation, reducing costs, or building their own legacy. Others are pressured by external motivations like the opinions of customers or the media, or a competition drive. However, some corporate leaders build a broader understanding of their company's environmental impact and respond by launching sustainability programs.

Executives Paul Polman of Unilever, Ray Anderson of Interface, and Mike Duke of Walmart all saw the importance of more sustainable operations. These corporate leaders became aware of, and concerned with, the consequences of their corporate operations on the global environment, or on their employees and consumers. They often talk about epiphanies from reading a book devoted to sustainability, touring an ecological disaster, or even from dealing with family health issues. These devotees become personally passionate about sustainability, often leading their company's sustainability programs themselves, or becoming public figureheads and influencing other corporate executives to follow in their footsteps.

Other corporate executives land on the value of sustainability based on cost-savings or the potential of new sales. Enterprising leaders recognize that sustainable operations can reduce the need for utility and material inputs like electricity, fuel, water, chemicals, or materials, can reduce wastes, and may reduce regulatory and other risks. These executives see sustainable thinking as a means for increasing efficiency and saving money. Entrepreneurial leaders also see the opportunity to go beyond sustainable operations to create new services or products to address sustainability challenges. For example, many retailers have created new product lines that are eco-friendly or health-conscious in response to growing demand from LOHAS (Lifestyles of Health and Sustainability) consumers.

Finally, a third segment of retail executives launch programs in response to competitive pressures; i.e., "everyone else is doing it." As with all corporate activities, certain industry practices grow in prevalence until they become the norm. Peer benchmarking, investor inquiries, and consumer insights all drive the dynamic of companies following their peers. According to a 2017 survey of 151 large companies by the nonprofit BSR, the CEO/C-suite leadership has the greatest influence on the corporate sustainability agenda, over that of stakeholders such as customers, investors and government. That sustainability is increasingly the norm is illustrated in the same survey by the response of companies to the USA pulling out of the Paris Agreement on climate change—75% of respondents indicated that this would either have no impact or strengthen their commitments to addressing climate change (BSR 2017).

10.3.3 Grass Roots: Employees

Interestingly, the human resources (HR) department may be one of the more valuable allies for promoting sustainability. Employees, especially younger people, want to work for companies with sustainable values. According to a 2016 survey, 76% of millennials consider social and environmental commitments when job hunting and 64% will not work for a company with poor CSR practices (Cone Communications 2016). As a result, HR must respond to new questions in areas such as CSR, diversity and inclusion programs, and philanthropic interests, and design recruiting programs that appeal to applicants who care about sustainability.

10.3.4 Investors and Customers

There are two major outside stakeholders that drive industry change: investors and customers. Customers exert direct influence on retail companies, as the driving factor in retail is acquiring and retaining customers. Retail companies are acutely aware of consumer trends, spending millions on consumer insights research each year. When consumers prefer eco-friendly, natural, organic, local, non-toxic, or other products with sustainable attributes, retailers are quick to respond by adding new product lines or improving the eco-friendliness of existing products.

Investors are increasingly factoring sustainability considerations into their investment criteria and rating companies on sustainability-related Key Performance

Indicators (KPI). Ranking and rating systems like the Dow Jones Sustainability Index (DJSI), which features a few major retailers (Best Buy, Canadian Tire, CVS Health, and Gap Inc.) further promote this trend. Investors have back-room conversations with executives in their portfolios, and when action is not taken, they may submit investor resolutions. Ceres, a sustainability nonprofit that works with many investors and corporations, maintains a database of shareholder resolutions since 2011. Over 150 were filed in 2017 alone (Ceres 2017). Highlighting this trend of investor engagement, Blackrock, the world's largest asset management company, recently communicated the importance of sustainability to companies, saying "To prosper over time, every company must not only deliver financial performance, but also show how it makes a positive contribution to society."

10.3.5 Activists and Media

Activists target retail companies because of their influence in the global economy and the consumption system. Activists may use exposés, documentaries, and hidden camera footage to highlight negative practices in consumer goods supply chains. Those efforts are intended to gain media attention and to spread via social media channels. Recent reports have focused on issues such as truckers in the USA, seafood produced in Thailand, and conflict minerals from the Democratic Republic of the Congo. Activists seek to raise awareness of these issues among the general population, hoping to change consumers' purchasing habits, company operations, or trigger new regulations. This indirect pressure influences investors and consumers, which further create a market environment that promotes the development of retail sustainability programs.

Because of these forces, more retailers are adopting sustainable practices and building sustainability programs and governance structures. The following section describes the typical maturity stages for a retail sustainability program, highlighting the progress companies make as their programs mature.

10.4 Stages in Retail Sustainability Program Maturity: (1) Compliance, (2) Efficiency, (3) Reputational Risks, and (4) Innovation and Circularity

No matter the origin, retail sustainability programs tend to grow and mature around common dimensions. They begin with a focus on the basics, primarily complying with regulatory requirements, then growing to cover operational efficiencies that save costs, reducing reputational risks, and eventually to innovating on the very core of their businesses. The following section outlines each of these four maturity stages, documenting the operational practices that retailers can undertake to address each stage.

10.4.1 Stage 1: Compliance—Beginning with What is Required

Environmental compliance programs are typically considered separate from sustainability programs. However, they can directly influence the success of a sustainability program and in turn, sustainability efforts can affect compliance efforts. Environmental compliance programs consist of the processes that a company uses to ensure compliance with regulatory and other requirements. The most broadly implemented program framework is the Environmental Management System (EMS), which is based on a plan-do-check-act approach to continual improvement. The formal EMS standard (International Institute of Standards (ISO) 14001:2015) also includes provisions related to sustainability.

Given the level of environmental requirements in retail, as compared to other sectors such as manufacturing, most retailers choose not to implement a full EMS. An exception is Best Buy, which has an ISO certified EMS. Interestingly, it was Best Buy's commitment to sustainability and subsequent decision to accept e-waste from consumers that helped drive their EMS implementation. Hazardous components in electronics trigger significant regulatory requirements for the management of hazardous waste and the EMS helps Best Buy stay in compliance.

A poor compliance program can damage a company's sustainability efforts in several ways. Environmental violations, especially major events, can erode the reputational benefits gained from a sustainability program and make even a well-intentioned company appear to be insincere and guilty of greenwashing. This is especially true for violations that customers perceive as affecting them, which is more likely in retail than in less consumer-facing industries. In 2013, four retailers paid fines of $1.26 million for selling fabric labeled as bamboo that was actually rayon, a violation of the Textile Products Identification Act and misleading to consumers because, while bamboo is viewed as renewable and environmentally friendly, rayon has a significant environmental impact.

Violations and environmental damage can also influence investor perceptions and stock prices. Lumber Liquidators stock collapsed after a 60 min story about laminate flooring from China containing formaldehyde over acceptable limits and further eroded when the company was charged with environmental violations related to importing wood from illegal sources. In addition to the impact on stock prices, sales were also negatively affected (Conniff 2016). This example illustrates how the supply chain is an area of significant risk for retailers. In another example, sales at the British company Tesco suffered from a 2013 scandal when beef products from several suppliers were found to contain significant amounts of horse and pig meat.

A good compliance program can help a company set and achieve their sustainability goals. Many compliance programs include a process to identify the most significant regulatory risks and environmental impacts. This process can be used to set sustainability goals that focus on reducing the organization's most negative environmental impacts. A company commitment, such as a public commitment to a specific GHG reduction goal, becomes a corporate requirement and can be included in the company's compliance program. This provides an established process for tracking progress, reporting, and achieving sustainability goals. For example, in addition to compliance with regulations, Best Buy uses their EMS to help track and achieve sustainability goals in areas such as recycling and GHG reduction.

Environmental regulations increasingly overlap with sustainability objectives. Several states have enacted legislation that bans or restricts certain chemicals in consumer products, especially children's products, or require labeling and reporting when the chemicals are used. California's Proposition 65 (1986 Safe Drinking Water and Toxic Enforcement Act) requires retailers to label products containing chemicals on the Prop 65 list to give consumers "clear and reasonable warnings" about possible carcinogens, or birth defects and other reproductive harm. Many retailers have responded to this type of legislation by reducing or eliminating chemicals that trigger the requirements from their products, either by reformulating store brands or asking suppliers to remove chemicals. In 2017, Target announced an ambitious plan to be more transparent about product ingredients, remove selected chemicals from products, and to fund efforts to find safer alternatives for potentially harmful chemicals.

Waste is another area where regulations can drive more sustainable operations. Many states and local jurisdictions ban certain items from landfills to promote recycling or reuse. This includes items such as used tires and appliances, as well as recyclables such as cans or cardboard. In California, Assembly Bill 1826 requires companies to compost organic waste such as food waste. Interestingly, these requirements are related to California's GHG emissions reduction efforts, as decomposing food waste in landfills generates methane. Collecting and composting food waste can be a challenge for many retailers who lack the space to set up a separate container and must control the smell of the organic waste and potential for attracting pests. One solution is to reduce the volume of waste. The EPA developed the Food Recovery Hierarchy with a focus on source reduction and finding other uses for food waste with landfilling at the bottom of the hierarchy (EPA 2017).

More sustainable operations have the potential to reduce compliance costs, regulatory requirements, and risk. Over the past few years, retailers have been fined millions of dollars for not complying with hazardous waste regulations. Under the regulations, common consumer products such as cosmetics, medicines, cleaning products and more may be considered hazardous waste when unsalable. Reducing the amount of waste by reducing breakage, expired products and unsold products, saves money as the disposal of hazardous waste is expensive and reduces the risk of noncompliance. Another example is the potential to reduce stormwater fees based on the amount of impervious surfaces by using green building approaches.

10.4.2 Stage 2: Efficiency—Reducing Short-Term Costs

Reducing utility bills is the next motivation for retail programs. Because utilities are already a clearly defined line item in a facility's operating expenses and affect a company's overall profitability, they are often the first sustainability area that businesses tackle. These projects are also easier to launch as projects with a guaranteed reduction in recurring expenses with direct financial returns are easier to sell internally and to evaluate. These business benefits are catalogued in case studies like those showcased in the Retail Industry Leaders Association (RILA) Resource Library (RILA 2017a), the Department of Energy (DOE) Better Buildings Solutions Database (Better Buildings: U.S. Department of Energy, Solutions 2018), and the Environmental Defense Fund (EDF) Climate Corps fellowship outcomes (EDF 2017).

Economic concepts like ecosystem services (i.e., benefits gained from the natural environment) and policy strategies like environmental excise taxes can be used to account for negative environmental externalities. The carbon taxes across Europe and in Canada are examples. As these frameworks improve and become normalized, they help assign costs to the previously unaccounted financial implications of environmental degradation. As these concepts become more popular with governments and corporations, environmental sustainability and profitability become increasingly intertwined.

Some retail companies already apply corrective pricing schemes to their operations in recognition of their absence in current markets and in anticipation of future legislation. Companies like Louis Vuitton, Moet Hennessy (LVMH), Puma, Disney, and Canadian Tire have all developed real (revenues are invested into carbon-reducing projects) or shadow (financial impact is included as a symbolic indication of carbon-intensity) internal prices on carbon. Many more companies plan to develop their own prices in the next few years (CDP 2016). However, the more intangible risks and benefits of sustainability will continue to evade monetization because of the inherent challenges to quantifying their importance and impact.

10.4.3 Stage 3: Managing Reputational Risk

Goodwill, the intangible corporate value that includes brand value, can be significant for a company, especially for consumer-focused retail. A public reputation for sustainability can help build goodwill; companies get the business benefit of their sustainability programs by differentiating them from their competitors and attracting customers concerned about the environment. However, this also means that companies open themselves to the risk of overpromising, underperforming, being perceived as misleading, or worse, dishonest. This can range from charges of greenwashing and even penalties as described in the previous section about textile labeling to a more serious loss of trust with the public. Consequences can also include the loss of sales, drop in share price, and damage to the company's reputation and goodwill.

Companies that embark on a sustainability program must be committed to the program and not simply using marketing to appear green. This is why it is critical for senior leadership to understand, support, and promote sustainability efforts. Volkswagen was promoting their cars as a greener alternative while implementing emissions override software, which was not just illegal but a breach of trust with their customers and cost the company billions of dollars. While the cause of the VW scandal likely has more to do with a breakdown in management, one underlying reason is that top management did not truly care about air pollution.

Sustainability or CSR reporting has become a standard practice for large and mid-cap companies. CSR reporting helps protect a company's reputation as it highlights their commitments while also showing transparency so that the company does not appear to be misleading. Public reporting also has internal benefits as it helps companies set and implement long-term goals and evaluate their performance. However, this is one area where retail has lagged other industries. According to a recent report on global CSR reporting, "more than two thirds of companies in all sectors except retail now report on their CR performance" (KPMG 2017).

Sourcing products that are more sustainable and ensuring the accuracy of sustainability claims can be a challenge for retailers, who often depend on the word of suppliers that their products conform with the company's sustainability guidelines. Walmart uses a Sustainability Index to evaluate supplier sustainability. Target launched a Sustainable Product Standard to evaluate products and plan to use the results in purchasing and placement decisions. In the broader world of CSR, especially around labor practices, many retailers conduct audits of overseas factories to ensure compliance with their standards. In this area, retail is one of the few industries that is expected to enforce compliance with US environmental and other regulations in their overseas supply chain. Retailers occasionally join collaborative organizations, typically nonprofits, like The Sustainability Consortium, Sustainable Apparel Coalition, the Sustainable Packaging Coalition, BSR, and others to jointly address environmental or human rights considerations. These organizations can pool resources, tap expertise that otherwise is not available to any one retailer, and validate the legitimacy of retailers' activities and claims.

Another approach to ensuring that "green" products are truly green is to sell products that meet third-party certifications. Most consumers are familiar with the "organic" label but there are established certifications for a variety of product types and issues from more environmental friendly practices such the Rainforest Alliance certification and the FSC, to the evaluation of products for specific attributes such as the Green Seal certification for products such as cleaners and paints, or US EPA's Safer Choice certification for products that are safer for human health and the environment. An advantage of this approach is that it is the responsibility of the third-party certifier to set and enforce standards. In addition, consumers can easily identify environmentally preferable products on the shelf.

Operating more sustainably and with greater transparency not only helps build goodwill, it can help reduce risk. Retailers that are leaders in this area are betting on the long-term benefits of more sustainable operations, as well as the business value that they will gain from consumers' preferring to shop at sustainable companies. Lagging companies, on the other hand, are taking on greater risk. They risk not only being viewed in a less positive light by consumers but have operational risks by not understanding or responding to environmental threats such as water shortages or climate change.

10.4.4 Stage 4: Innovating Through the Circular Economy

Innovation is the final stage of maturity for retail sustainability programs. Because sustainability requires creative solutions and cross-departmental collaboration, it is a natural source of innovation and can drive new value to the business.

Improved Packaging and Product Design They are core components of sustainable innovation. With the concepts of a circular economy in mind, retailers can uncover opportunities like using recycled materials, replacing non-recyclable materials with recyclable and natural alternatives, developing new product lines that promote health and nature, and transitioning to lighter, easier to open packages. Product design, sourcing, merchandising, and marketing teams often need to work together to initiate and implement these changes. One of the leaders in sustainable packaging is Lush with their "naked" line of products, which use minimal or recycled content packaging and encourage customers to return containers to the stores. Other retailers are working collaboratively in organizations such as the Sustainable Packaging Coalition to find solutions to packaging. However, this work is threatened to be overwhelmed by the growth in e-commerce and subsequent increase in cardboard boxes and shipping material. It is estimated that the 165 billion packages shipped in the USA every year use about 1 billion trees worth of cardboard, not even considering the plastic waste (Peters 2018).

Sustainable innovation can also drive more *resilient supply chains*. Since retailers operate global supply chains, they are intrinsically invested in the stability of the supply of materials and components that go into their products. Recent natural disasters, like Japan's major earthquake in 2011 and political disruptions like Indonesia's aggression toward Vietnamese fishermen, underscore the volatility of those product supply chains and spur companies to improve supply chain resiliency. Retailers are innovating to improve the resiliency of their supply chains by

- simplifying supply chains so that raw materials, components, and finished products have shorter transportation journeys;
- using more abundant and natural raw materials as input to products;
- from politically stable countries;
- moving production closer to the point of consumption.

New Business Models They are a critical form of sustainable innovation. Traditionally, retailers are seen simply as product distribution channels, and generate profits based almost exclusively on the volume and prices of products they sell. This definition of retail's core business model only further acts to perpetuate the "take, make, and dispose" consumption system. However, with societal changes such as the Internet and greater access to information, consumers are changing their lifestyles and consumption patterns. Companies like Zipcar, Airbnb, and Uber have pioneered rental and sharing models. A trend that retail companies like Rent the Runway and Bag Borrow or Steal have applied to consumer products. Traditional retailers are responding with new services like product repair, rental, sharing, and take-back that align customers' interests. These new channels of interaction, such as AutoZone's free Loan-A-Tool program, are expanding the business opportunities for brick-and-mortar stores.

10.5 Taking Action

As described earlier, retailers tend to follow a maturity path as their sustainability programs grow. This section describes the critical programmatic components for a strong retail sustainability program and is based on RILA's Retail Sustainability Leadership Model (RILA 2017b). The Leadership Model helps to identify pathways for retailers to implement strong environmental sustainability programs that fit their operations best. Though the elements presented below illustrate a progression, the reality is that every company is unique and will therefore take a different approach to initiating and growing its sustainability program. That is why the progression described below is only a suggested approach; each company should pick the elements and timing that make most sense for its culture and unique circumstances.

The first set of actions that a company takes is to focus on strategy and commitment. This helps a retailer develop the governance structures, strategies, goals, values, and incentives for a successful program. The strategy will change as the company's program matures and typically becomes more valuable as it becomes more refined.

As with any business function, oversight, accountability, also engaged executives are essential to the success of a sustainability program. Formal governance establishes Board oversight, which ensures that the Board understands the company's sustainability risks, sets corporate sustainability policy, and is accountable to implement that policy. As the definition of fiduciary duty evolves, Boards of Directors are taking responsibility to manage sustainability-related risks and opportunities. Engaging senior executives also means that companies move beyond isolated projects toward integration across business functions.

Executive Councils and Functional Councils are typically launched to help develop and implement sustainability strategy. Sustainability Executive Councils are one of the most effective ways to engage executives from across the organization. Retailers tend to have a two-pronged structure to Executive Councils, developing a (1) senior-most council and a (2) functional-level council.

Senior-most executive councils engage top executives from across the enterprise, namely the CFO, COO, Division Presidents, EVPs, and SVPs from critical functions like HR, Legal, Merchandising, and Private Brands. Their role is to ensure program alignment with business strategies and secure approvals for the company's approach to sustainability. Because of the seniority of these councils, they are likely to meet only three to four times a year. Functional councils consist of Directors, VPs, and SVPs who represent functional roles like Logistics, HR, and Store Operations. They are typically divided into working groups to drive progress on specific sustainability objectives and meet every 1–2 months.

With executive councils in place, the company can select a chief sustainability officer (CSO) to lead the councils and oversee strategy and execution. In the most mature programs, the CSO reports directly to the CEO, and together they partner to demonstrate the business relevance of sustainability program to investors and other stakeholders and communicate the support of the company. Retailers, brands, and others in the retailing supply chain, like 3M, Coca-Cola, General Mills, FedEx, IKEA, Kellogg, L'Oréal, Mars, Spectrum Brands, Tiffany's, Tyson Foods, Unilever, UPS, Walmart, and West Marine, have senior-level CSOs that are influential in their organizations (The Weinreb Group 2014).

Strategy is a critical step in building an effective sustainability program. Sustainability will only be seriously considered in business planning if its strategy aligns across departments and with the overall corporate strategy. The best programs go a step further and incorporate internationally recognized standards into the organization's overall long-term strategy, often using a triple bottom line approach for reviewing future strategies and corporate projects. For the strategy to be successful, the CEO must be a champion of the company's sustainability agenda, and regularly incorporate sustainability strategy in its meetings and corporate communications.

Companies perform materiality assessments to identify the issues that are critical to their business and to outside stakeholders. Materiality is a term of art in corporate finance but is increasingly used in the sustainability discipline. GRI, an organization that promotes the use of common sustainability reporting standards, defines material aspects as "those that reflect the organization's significant economic, environmental and social impacts; or that substantively influence the assessments and decisions of stakeholders" (GRI 2013). The best materiality assessments require quantitative analysis, qualitative assessment and discussion, involve individual departments (e.g., sourcing, operations) for their input and review, and use the same processes to review both sustainability risks and opportunities, and corporate risks and opportunities. For example, Walgreens Boots Alliance describes their materiality process in this way: "As part of our CSR materiality assessment process we mapped out potential material topics as well as issues raised by stakeholders. We considered the relevance of each topic to internal and external stakeholders, and the relevance of each topic to the Company. We also consider wider societal expectations and our influence on customers and suppliers to determine priority topics.

"Through the mapping, and as a result of the inputs described above, we confirmed that the issues most relevant to us are reflected in our 12 goals, which are grouped in four focus areas where we can have the most impact: Community, Environment, Marketplace and Workplace" (Walgreens Boots Alliance 2018).

Goal-setting is also a common step in developing a sustainability program. Goals help define the corporate-wide vision, drive action and alignment, and signal the company's commitment to sustainability with critical partners (e.g., suppliers, customers, investors) and the areas of most interest. Retailers set both short- and long-term sustainability goals, commonly beginning with a focus on facility operations, then reaching into their product lifecycles, and eventually into their consumer marketing and messaging. The leading goals are absolute reduction goals that address material issues. The best goals incorporate the global and local context of environmental conditions. For example, science-based targets, which are based on the reduction of GHG emissions needed to keep global temperatures below 2 °C or the UN SDGs. Some quantitatively link profit goals with sustainability goals or aim for zero impact. Companies like Delhaize, Eileen Fisher, Kroger, Nike, REI, and Walmart have set “big, hairy, audacious goals” for zero impact in at least one area of their operations.

Peter Drucker said that “culture eats strategy for breakfast.” This sums up why it is critical that the sustainability program’s strategy aligns with the corporate culture, or put another way, the company’s culture must reinforce its sustainability vision. The corporate values and/or mission should mention the importance of sustainability to the business, and that ethos needs to be reinforced in all major internal corporate communications, meetings, and events. Senior executives must regularly refer to the corporate values in terms of a sustainability and responsibility to set the tone for the corporate culture.

Gap’s CEO Art Peck connects the company’s corporate values with sustainability, “When Doris and Don Fisher opened the first Gap store in San Francisco in 1969, they did not expect to transform retail. They just Could not find a pair of good jeans that fit. From that single store to today’s global business, Gap Inc. is synonymous with equality, community and laid-back American style. Good business—the kind that puts people at its center—has the potential to change the world, no matter how small it starts. At Gap Inc., we still sell good jeans, and we still believe in good business. We are also part of a world that has changed a lot since 1969. Today, customers expect more from a product. They want to know the story behind it. Where did it come from? Who made it? Was it created in a fair, safe and environmentally responsible way? What was the impact on people and the planet? We owe it to our customers to ask ourselves those same questions. Some of the answers have brought us back to our core values, and some have compelled us to find new solutions and build new partnerships” (Gap Inc 2018b).

Related to culture are incentives, which define success for individuals and corporate departments by encouraging desired behaviors. Financial incentives are the most common; for example, companies can link executive compensation to sustainability performance. Leading retailers also invest in incentives to influence positive employee behaviors by providing benefits like charging stations for electric vehicles or bikes for employees to use.

Once the company’s house is in order, it can focus on executing its sustainability strategy. This entails engaging external stakeholders, partnering with the internal HR and communications teams, building funding mechanisms, and creating business innovation mechanisms.

Relevant stakeholders have influence on, or are influenced by, the enterprise. For retail, the major stakeholders are their customers, investors, activists, and the media. Retailers should identify the most relevant stakeholder groups and the most relevant KPIs for the business and their sustainability strategy. Then establish communication methods for each stakeholder group, finding ways to both provide updates and incorporate feedback from key stakeholders into the company’s sustainability strategy.

In addition to external stakeholders, there is a critical internal stakeholder: employees. To reach them, sustainability programs must partner with HR and communications teams. With HR, sustainability teams can provide a collaborative forum for high-initiative employees to receive recognition for their sustainability efforts while sharing best practices with colleagues. They can also develop employee orientation and ongoing training programs. Leading programs also regularly educate employees on the company's sustainability vision and business case to underscore relevance to their daily work. They also hold highly visible senior leadership meetings on sustainability where store employees, sourcing, merchants, logistics, and other staff are recognized. And finally, they train in-store employees to educate customers about the company's sustainability efforts.

The success of a business program often hinges on the investment of resources, most specifically funding. Some retailers provide a budget to key retail functions specifically earmarked for sustainability projects in those departments (e.g., distribution, sourcing, merchandising, store operations). Corporations with mature sustainability programs typically require that all corporate funding requests include sustainability metrics and increase the investment on sustainability-related programs over time.

As the retail industry evolves, so too must sustainability programs, making innovation a critical component of a strong program. Leading retailers have dedicated teams to create and invest in sustainable innovations, build an innovation fund to invest in new retail business models that promote sustainability, and leverage the Chief Innovation Officer's ability to place sustainability and continuous improvement into corporate innovation goals.

As with any business discipline, sustainability needs clear metrics to communicate priorities and measure success. Sustainability metrics should be focused on all material aspects of the business, including innovation, linking people and planet with profit and stakeholder concerns, and should be chosen from global frameworks. Metrics need to be tracked for the most relevant regions of the global business using automated measurement tools and IT systems.

While sustainability programs are valuable in their own right, their value is amplified when the company's goals and commitments, activities, and results are communicated to relevant stakeholders. In retail, this storytelling is done in several ways. The first is through reporting, with a best practice of using third-party standards (e.g., GRI), independent auditing, communicating via multiple channels like websites, product marketing and labeling, and advertising. Leaders go a step further, integrating their sustainability and financial reports, called "integrated reporting," and thereby telling a single, 360-degree story about their business.

Point-of-purchase consumer education channels are another opportunity for storytelling. Retailers can devote significant in-store signage or sections to products with advanced sustainability benefits, develop dedicated online storefronts to encourage consumers to select products with sustainability benefits, and engage consumers in other ways about products with sustainability benefits (e.g., catalogs, web filters, icons, online calculators, and product stories).

Since marketing campaigns are one of the most visible ways that retail brands interact with consumers, retailers should dedicate a sustainability-focused marketer within the marketing team, with a significant budget devoted for sustainability-related marketing and incorporate marketing effectiveness metrics into the ROI for that dedicated budget. Retailers can also create branded campaigns with themes that resonate with consumers. This ensures that the campaign is more than just one touchpoint that fizzles out. Retail companies can take a stand or have a clear position on certain issues (e.g., climate change, consumption) and connect their corporate sustainability stories to these issues.

Recognizing that an individual company cannot tackle the systemic issues of sustainability alone, most leading retailers join collaborative groups to increase their effectiveness. Leading retailers are premier members of one or more industry associations or multi-lateral groups focused on relevant sustainability issues; they partner with NGOs, governments, academia, or other institutions to identify improvement opportunities; and they take a leadership role in developing new tools or capabilities that will enable peers to improve their sustainability and supply chain sustainability performance.

Performing these steps, in the order appropriate for any specific retail enterprise, lays the foundation for a strong and self-perpetuating retail sustainability program.

References

Alliance for Bangladesh Worker Safety (2018) About the alliance for Bangladesh worker safety. http://www.bangladeshworkersafety.org/who-we-are/about-the-alliance. Accessed Feb 2018

Ballinger J (1992) The new free-trade heel. Harper's Magazine (1992, August)

Better Buildings (2018) U.S. Department of Energy, Solutions. https://betterbuildingsinitiative.energy.gov/solutions. Accessed 10 Feb 2018

Bove A-T, Swartz S (2016) Starting at the source: sustainability in supply chains (2016, November). https://www.mckinsey.com/business-functions/sustainability-and-resource-productivity/our-insights/starting-at-the-source-sustainability-in-supply-chains. Accessed 3 Feb 2018

Brockhaus S, Amos C, Fawcett AM, Knemeyer A, Michael, Fawcett SE (2017) Please clap! How customer service quality percepton affects the authenticity of sustainability initiatives. J Market Theory Pract 25(4):396–420

BSR (2017) The state of sustainable business 2017: results of the 9th annual survey of sustainable business leaders. Survey. https://www.bsr.org/reports/2017_BSR_Sustainable-Business-Survey.pdf. Accessed 11 Jul 2018

California Air Resources Board (2018) Refrigerant management program. https://ww2.arb.ca.gov/our-work/programs/refrigerant-management-program. Accessed 11 July 2018

CDP (2016) Embedding a carbon price into business strategy (2016, September). https://b8f65cb373b1b7b15feb-c70d8ead6ced550b4d987d7c03fcdd1d.ssl.cf3.rackcdn.com/cms/reports/documents/000/001/132/original/CDP_Carbon_Price_report_2016.pdf?1474899276. Accessed 10 Feb 2018

Ceres (2017) Shareholder resolutions. https://www.ceres.org/resources/tools/climate-and-sustainability-shareholder-resolutions-database. Accessed 7 Mar 2017

Cone Communications (2016) Millenial employee engagement study. (2 Nov 2016) Boston

Conniff R (2016) For logging's crimes, tougher punishments. The New York Times (19 Feb 2016)

EDF (2017) Project solutions database. http://edfclimatecorps.org/projects/results. Accessed 10 Feb 2018

Environmental Protection Agency (2018a) GreenChill partners. (7 Aug 2018). https://www.epa.gov/greenchill/greenchill-partners. Accessed 6 Nov 2018

Environmental Protection Agency (2018b) SmartWay excellence awardees (Oct 2018). https://www.epa.gov/smartway/smartway-excellence-awardees. Accessed 20 Oct 2018

Elkington J (1994) Towards the sustainable corporation: win-win-win business strategies for sustainable development. Calif Manag Rev 36(2):90–100

Ellen MacArthur Foundation (2017) What is a circular economy? https://www.ellenmacarthurfoundation.org/circular-economy. Accessed 27 Dec 2017

EPA (2017) Sustainable management of food: food recovery hierarchy. https://www.epa.gov/sustainable-management-food/food-recovery-hierarchy. Accessed 19 Feb 2017

EPA (2018) Sustainable management of food: food recovery hierarchy. https://www.epa.gov/sustainable-management-food/food-recovery-hierarchy. Accessed 10 Feb 2018

Freyssinier JP (2006) Reducing lighting energy use in retail display windows. Sixth International Conference on Solid State Lighting, Proceedings of SPIE. Rensselaer Polytechnic Institute, Lighting Research Center, Troy

Gap Inc (2018a) Global sustainability, product sustainability. https://www.gapincsustainability.com/product/product-sustainability. Accessed 20 Oct 2018

Gap Inc (2018b) Global sustainability, CEO letter. http://www.gapincsustainability.com/strategy/ceo-letter. Accessed 20 Oct 2018

GRI (2013) Glossary. https://g4.globalreporting.org/introduction/glossary/Pages/default.aspx#Material-Aspects. https://sftool.gov/learn/about/402/lca-example-light-emitting-diodes-leds. Accessed 10 Feb 2018

IKEA (2018) Sustainable life at home, make a difference without leaving your home. https://www.ikea.com/ms/en_US/this-is-ikea/people-and-planet/sustainable-life-at-home/. Accessed 20 Oct 2018

International Organization for Standardization (2006) ISO 14040:2006. Environmental management – life cycle assessment – principles and framework (2006, July). https://www.iso.org/standard/37456.html. Accessed 10 Feb 2018

International Organization for Standardization (2015) ISO 14000 family – Environmental management. https://www.iso.org/iso-14001-environmental-management.html

Karas B (2018) A vision of a circular economy: our packaging aspirations for the U.S. (18 Jan 2018). http://www.coca-colacompany.com/stories/a-vision-of-a-circular-economy-our-packaging-aspirations-for-the-u-s. Accessed 10 Feb 2018

KPMG (2017) The road ahead: the KPMG survey of corporate responsibility reporting 2017. KPMG. https://assets.kpmg/content/dam/kpmg/xx/pdf/2017/10/kpmg-survey-of-corporate-responsibility-reporting-2017.pdf. Accessed 10 Feb 10 2018

MacKinnon J (2015) Patagonia's anti-growth strategy. The New Yorker (21 May 2015). https://www.newyorker.com/business/currency/patagonias-anti-growth-strategy. Accessed 10 Feb 2018

Peters A (2018) Can online retail solve its packaging problem? Fast company. (20 Apr 2018). https://www.fastcompany.com/40560641/can-online-retail-solve-its-packaging-problem

RILA (2017a) Retail sustainability leadership model (22 Aug 2017). http://www.retailcrc.org/sustainability/Pages/Retail-Sustainability-Management-Leadership-Model.aspx. Accessed 10 Feb 2018

RILA (2017b) Sustainability resource library (29 Sep 2017). http://www.retailcrc.org/sustainability/Pages/Retail-Sustainability-Resource-Library.aspx. Accessed 10 Feb 2018

Rujanavech C, Lessard J, Chandler S, Shannon S, Dahmus J, Guzzo R (2016) Liam – An innovation story (Sep 2016). https://www.apple.com/environment/pdf/Liam_white_paper_Sept2016.pdf. Accessed 10 Feb 2018

Science Based Targets (2018) Companies taking action. https://sciencebasedtargets.org/companies-taking-action/. Accessed 6 Nov 2018

Solar Energy Industries Association (2017) Solar means business 2017. https://www.seia.org/research-resources/solar-means-business-2017

Spicer A, Hyatt D (2017) Walmart's emergent low-cost sustainable product strategy. Calif. Manag. Rev. 59(2):116–141. https://journals.sagepub.com/doi/abs/10.1177/0008125617695287?journalCode=cmra

Staples, Inc (2018) Staples sustainable products. https://www.staples.com/sbd/cre/marketing/about_us/corporate-responsibility/environment/sustainable-products/. Accessed 20 Oct 2018

Target Corporation (2017) Target announces new chemical strategy including policy and goals for its products and operations. (25 Jan 2017). https://corporate.target.com/article/2017/01/chemical-policy-and-goals. Accessed 12 Dec 2018

Tetrapak (2018) LCA examples. https://www.tetrapak.com/sustainability/environmental-impact/a-value-chain-approach/life-cycle-assessment/lca-examples. Accessed 10 Feb 2018

The Home Depot (2017) Responsibility report. https://corporate.homedepot.com/sites/default/files/image_gallery/PDFs/THD_Responsibility%20Report%202017.pdf

The Weinreb Group (2014) CSO back story III: the evolution of the chief sustainability officer. http://weinrebgroup.com/wp-content/uploads/2014/10/CSO-Back-Story-II.pdf. Accessed February 10, 2018

United Nations (1987) Our common future. United Nations, Oslo

United Nations (2015) Sustainable development goals (25 Sep 2015). http://www.un.org/sustainabledevelopment/sustainable-development-goals/. Accessed 26 Dec 2017

USGBC (2014) LEED in motion: retail. https://www.usgbc.org/sites/default/files/leed-in-motion-retail.pdf. Accessed 20 Oct 2018

USGBC (2017) LEED in motion: retail – Kohl's spotlight (12 Nov 2017). https://www.usgbc.org/articles/leed-motion-retail-%E2%80%94-kohls-spotlight. Accessed February 10, 2018

Walgreens Boots Alliance (2018) Materiality assessment. http://www.walgreensbootsalliance.com; http://www.walgreensbootsalliance.com/corporate-social-responsibility-report/our-approach-to-csr/materiality-assessment/. Accessed 10 Feb 2018

Walmart (2018) Walmart's sustainability index program. https://www.walmartsustainabilityhub.com/sustainability-index. Accessed 20 Oct 2018

Yasemin Y, Kor JP (2017) How large food retailers can help solve the food waste crisis. Harvard Business Review (19 Dec 2017). https://hbr.org/2017/12/how-large-food-retailers-can-help-solve-the-food-waste-crisis. Accessed 10 Feb 2018

Index

S. Ray, S. Yin (eds.), *Channel Strategies and Marketing Mix in a Connected World*, Springer Series in Supply Chain Management 9,
https://doi.org/10.1007/978-3-030-31733-1

S

T

U

V

W